Intron-mediated regulation of *LSM*s and splicing

A thesis presented for the degree of Doctor of Philosophy
by

Linda Palmisano

School of Biotechnology and Biomolecular Sciences
University of NSW
Australia
2006

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

(Signed) .....

Table of Contents

Abstract

One of the most unexpected results from the sequencing of all higher eukaryotic genomes completed so far is the abundance of transcribed non-coding DNA: in humans intronic sequences make up around 95% of all pre-mRNA transcripts. Aside from various effects on adjacent protein-coding regions, these sequences were long thought to be largely non-functional. However a number of recent studies have identified non-coding sequences within introns which are almost 100% conserved across species. Therefore, rather than being 'junk' DNA, the vast amount of sequence present in introns may play important roles in the cell, perhaps even forming part of a regulatory network controlling the expression of a number of distant genes. However, to date, little direct evidence for such regulation has been reported.

This study presents the first clear example of regulation by an intron *in trans*. A set of genes were identified in the model organism, *Saccharomyces cerevisiae,* that were coordinately regulated across a number of different environmental conditions. This regulation was at the level of transcriptional initiation and required the *LSM7* intron, but not the Lsm7 protein. The *LSM7* intron sequence was retained in all yeast species examined, including those that had lost many introns during their evolution. In addition, blocks of strong homology were identified across the introns of closely-related species. The exact mechanism underlying this regulation has yet to be fully elucidated, and several alternative models are presented to explain how it may be achieved.

Interestingly, both *LSM7* and the other genes in the co-regulated set code for proteins that form part of a complex at the core of the spliceosome. This complex stabilises U6snRNA, which is required to catalyse the first step of splicing. Both levels of U6snRNA, and changes in cellular splicing capacity correlated with expression of the co-regulated gene set. The *LSM7* intron was also essential for the changes in U6snRNA abundance and cellular splicing capacity in at least one of the conditions tested. This indicates that changes in *LSM* gene expression, U6snRNA abundance and splicing capacity may be functionally linked. Data from

splicing-specific microarray analyses indicate that *LSM7*-mediated regulation may affect the splicing of a wide range of transcripts and thus have significant effects on cell physiology.

Publications

Palmisano L.J., Kornfeld G.D., Mabbutt B.C., Dawes I.W. (2006). The intron of *LSM7* mediates transcriptional regulation of a set of genes encoding spliceosomal proteins. *manuscript submitted.*

Abbreviations

bp	base pairs
FISH	fluorescence *in situ* hybridisation
GTF	general transcription factor
IA	intron accumulation index
Inr	initiator sequence
LSM	"like-Sm"
miRNA	micro RNA
mRNA	messenger RNA
ncRNA	non-coding RNA
nt	nucleotides
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RNA Pol III	RNA polymerase III
RP	ribosomal protein
rRNA	ribosomal RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
TAF	TBP- associated factor
TBP	TATA-binding protein
UAS	upstream activator sequence
URS	upstream repressor sequence
UTR	untranslated region
YEPD	yeast extract peptone dextrose (complex medium)

Saccharomyces cerevisiae genes are referred to by a three-letter mnemonic, followed by a number. Gene names are italicised, dominant alleles capitalised and recessive ones given in lower case. Deletion is indicated by Δ. Protein gene product is named according to the gene name, in Roman type, with the first letter capitalised and a suffix –p.

Acknowledgements

So it has come to this. Honestly, I'm as surprised as you. That I made it this far is due in no small part to the contributions and support of so many people that thanking them all properly would fill a thesis by itself. So, with uncharacteristic restraint, I would like to very briefly express my sincere gratitude:

First and foremost to my supervisor, Ian, for giving me the opportunity to do this PhD in the first place, for his support, guidance and great patience in the face of extreme provocation (i.e. me),

To Geoff, for his generosity, his unflagging enthusiasm about all things splicing and for always being available to answer even the most inane of questions...or to steal all of my working pens,

To Manny Ares and lab, Anthony and Cristy, for generously sharing their data,

To Bronnie and Ruby, for joining me in a deep dislike of the ABI real-time PCR machine but allowing me to sneak in at strange hours to use it anyway,

To the various noble and long-suffering people who sat through all those half-formed sentences and random theories about Splicing Stuff! without obviously dozing off. I hope you realise just how helpful you were.

To the many wonderful people who were there with advice, empathy and humour and who made the whole thing worthwhile, particularly:

To Nazif, for always eerily knowing what it was I was trying to say, even if I didn't,

To Cristy, for her humour and unerring voice of reason,

To Emma, for endless reassurance and offers of alcohol,

To May, for teaching me The Moves (I tried, really I did),

To Nadia, for never laughing at my dress sense despite all the opportunities and for raising the bar for Class in the lab to levels never before seen,

To Matt, Gab and Anthony for bringing it back down again and for all the laughs,

To the rest of the Dawes lab, for their support and laughter (with me! WITH me!),

To the **multitude** of people who thought asking me if my thesis was finished yet at regular intervals was amusing and not going to result in me biting them,

To everyone who joined in to play the rather elaborate Let's Not Kill Linda with Ampicillin Just Yet! game. My airways thank you, and I live to hum another day.

Finally,

To my family, for their unique combination of heckling and unstinting support and generosity,

and

To Mark – your belief in me over the long life of this thesis has never failed to amaze, and your help in its final hours was nothing short of heroic.

THANKYOU

Chapter 1: Introduction

The structure of DNA, presented by Watson and Crick more than 50 years ago, indicated both the mechanism of DNA replication and the possibility of replication of RNA on DNA (Watson and Crick, 1953). This observation led to what was, for many years, the 'central dogma' of molecular biology:

DNA → RNA→ Protein.

It was postulated that unstable 'DNA-like' messenger RNA (mRNA) acted as an intermediary, moving from the nucleus to the translation machinery where it would become a template for protein production (Jacob and Monod, 1961; Brenner, *et al.* 1961). However analysis of cellular RNA composition indicated the majority of RNAs do not carry genetic information for protein synthesis. Indeed, recent estimates suggest that around 98% of all transcribed material in human cells is made up of non-protein-coding RNA (ncRNA) (Mattick, 2001).

1.1 Omnipotent RNA

Further analysis has identified ncRNAs with roles in a wide variety of processes in the cell. These include species involved in transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability, translation and protein degradation (summarised in Table 1.1). They range in size from around 21 nucleotides for micro RNAs (miRNAs) (Lagos-Quintana *et al.,* 2001; Lau *et al.,* 2001; Lee and Ambros, 2001), to more than 10,000 nucleotides for RNAs involved in gene silencing in higher eukaryotes (Sleutels *et al.,* 2002; Erdmann *et al.,* 2000).

The most abundant ncRNAs are the ribosomal RNAs (rRNAs), which associate with proteins to form ribonucleoprotein complexes (RNPs) involved in protein synthesis. Interestingly, isolated rRNAs are capable of forming stable compact three-dimensional structures which are highly similar to that of the assembled ribosomal snRNP complexes (Vasiliev *et al.,* 1978). Data from X-ray

crystallographic analyses indicate that it is these self-folded rRNAs that determine the morphology of the ribosomal subunits (Figure 1.1; Wimberly *et al.,* 2000; Schlünzen, *et al.,* 2000; Ban*et al.,* 2000). Indeed, many of the ribosomal functions required for protein synthesis are, at least in part, RNA-mediated (Moazed and Noller, 1991; Noller *et al.,* 1992; von Ahsen and Noller, 1995). Furthermore, prokaryotic ribosomes completely depleted of protein are still capable of catalysing reactions required for protein synthesis (Noller *et al.,* 1993).

Table 1.1: Functions of some different types of ncRNAs (Capara and Nilsen 2000; Spirin, 2002; Storz, 2002; Mattick and Makunin, 2005)

Function	Type of ncRNA
DNA replication	primer RNA , telomerase RNA
Regulation of transcription	6sRNA, 7SK RNA, SRA RNA, *Xist* RNA, *roX* RNA
mRNA degradation	siRNA, miRNA
mRNA processing	snRNA, self-splicing introns, gRNA
snRNA processing	snoRNA
rRNA processing	snoRNA, RNAse MRP
tRNA processing	RNase P
Protein synthesis	rRNA, tRNA
Regulation of translation	sRNA, miRNA, OxyS, asRNA, sRNA
Protein stability	tmRNA
Protein translocation	SRP 4.5sRNA, 7sRNA

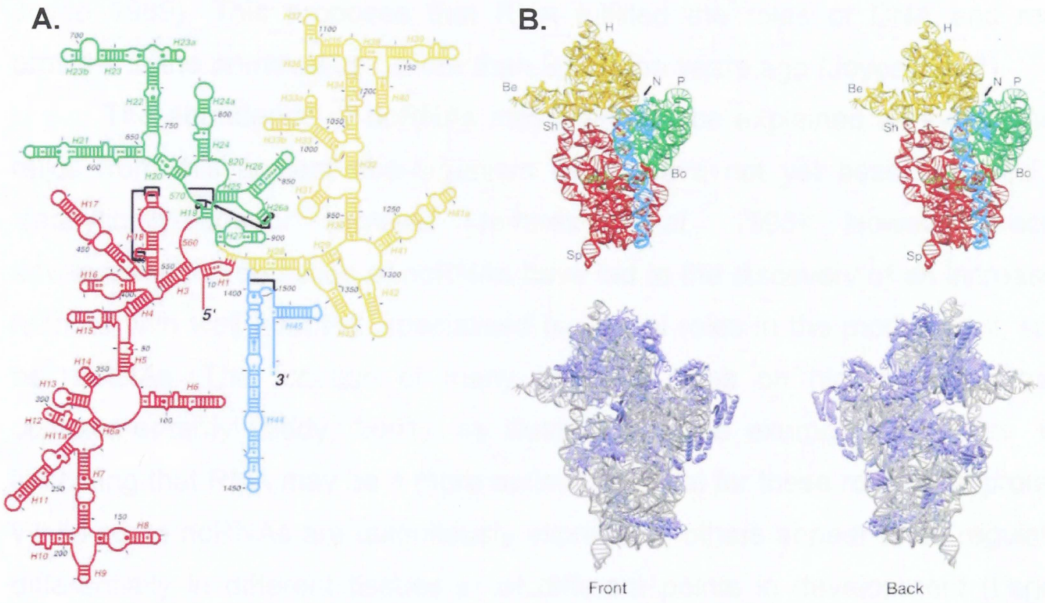


Figure 1.1: Secondary and crystal structure of the 16S rRNA subunit of the ribosome. This shows the original 2D (**A.**) and complex 3D (**B.**) rRNA structure that acts as the scaffold for ribosomal protein binding and protein synthesis. The colouring corresponds to: Red, 5' domain, green, central domain, orange 3' major domain, cyan 3' minor domain. Image reprinted from Wimberly *et al.* (2000).

A number of other ncRNAs have now been discovered or synthesised which act as 'ribozymes', catalysing a wide range of chemical reactions in a highly selective, sequence-specific manner (Kruger *et al.,* 1982; Guerrier-Takada *et al.,* 1983; Joyce, 1999). Indeed, RNA appears to be potentially capable of carrying out almost all of the functions in the cell currently performed by proteins.

In 2001 Johnston *et al.* successfully synthesized a ribozyme with *in vitro* RNA-dependent RNA polymerase activity. This indicates that ncRNA may also be capable of self-replication.

Therefore RNA has the unique ability to act as a catalyst, as well as carrying and replicating genetic information. Moreover, non-coding RNAs have

been identified in all species including Archea and eubacteria (Tang *et al.*, 2002; Selinger *et al.*, 2000). These discoveries have led to the theory of an 'RNA World' as an intermediate stage in the evolution of life (Gilbert, 1986; Joyce 1989). This proposes that RNA fulfilled the roles of DNA and most proteins in the primeval cell, more than 3.6 billion years ago (Joyce, 1991).

The abundance of ncRNAs may therefore be explained as evolutionary relics from this ancient ribo-organism which have not yet been replaced by 'catalytically-superior' proteins (Jeffares *et al.*, 1998). However, recent advances in the detection of ncRNAs have led to the discovery of an increasing number with well adapted, specialised biological roles in the modern cell, such as miRNAs. The function of many ncRNAs relies on highly-specific base complementarity (Eddy, 2001), as illustrated in the examples in Figure 1.2, indicating that RNA may be a more optimal material for these roles than protein. While some ncRNAs are ubiquitously expressed, others appear to be regulated differentially in different tissues or at different points in development (Lagos-Quintana *et al.*, 2002; Bashirullah *et al.*, 2003). Therefore, ncRNA may have continued to evolve following the end of the 'RNA World', to adopt or perfect many of the diverse roles they currently play in the eukaryotic cell.

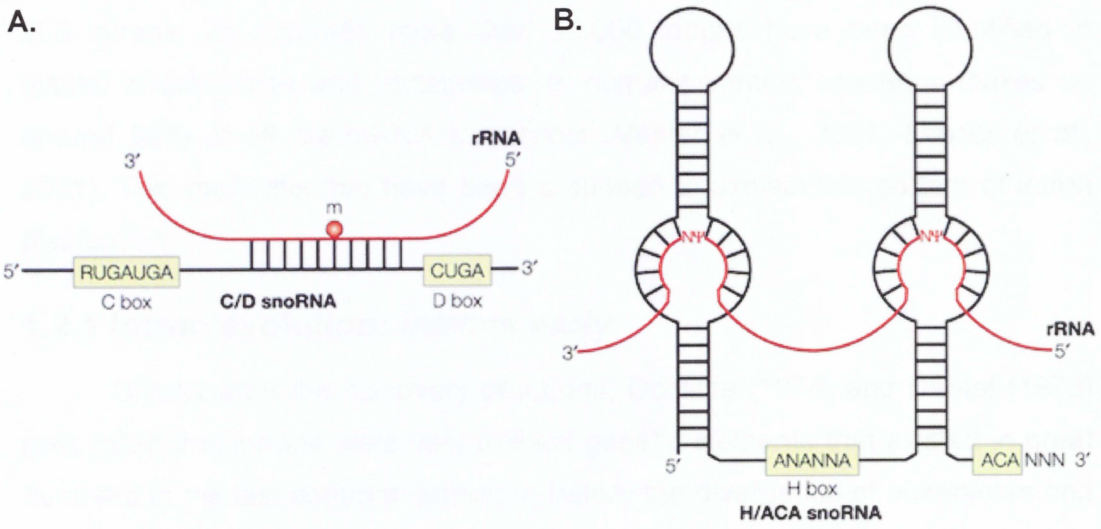


Figure 1.2. Examples of ncRNA functions dependent on specific-specific base complementarity. C/D Box (A) and H/ACA box (B) small nucleolar RNAs (snoRNAs) use antisense complementarity to target rRNA for 2'-*O*-ribose methylation (at site 'm'), or pseudouridylation (at site 'N'), respectively. R stands for A or G (a purine base). Image reprinted from Eddy (2001) and Kiss (2001).

Non-coding RNA species are difficult to identify and have only recently begun to be studied in any systematic way. Therefore it is likely that a great many other ncRNA species will be discovered in the future.

1.2 Evolution and function of nuclear introns

Introns are stretches of non-coding sequence that interrupt regions of coding sequence (exons) in DNA and must be removed from the primary mRNA transcript prior to translation. Most eukaryotic protein genes contain at least one intron and human genes generally contain around six to nine introns (Venter *et al.,* 2001, Sakarkar*et al.,* 2004). The average length of an intron is 2100 nucleotides (nt), often much larger than the adjacent exons, and introns more than 100,000 nt long have been identified. Due to the presence of multiple long introns, human genes can be more than a million nucleotides in length (Fedorova and Fedorov, 2003) and can take up to 16 hours to be transcribed (Tennyson, *et al.,* 1995).

More than 25 years have passed since introns were first identified, but their evolution and function remain controversial. Introns are not evenly distributed across species; a relatively small number are found in some eukaryotic organisms, such as *Saccharomyces cerevisiae*, which has less than 300 introns. By contrast, more than 10,000 introns have been identified in plants, invertebrates and vertebrates. In humans intronic sequence makes up around 96% of all pre-mRNA transcripts (Venter *et al.,* 2001; Lander *et al.,* 2001). Two main theories have been proposed to explain this pattern of intron distribution:

1.2.1 Intron evolution: *introns early*

Shortly after the discovery of introns, Doolittle (1978) and Darnel (1978) postulated that introns were very ancient genetic elements that existed in great numbers in the last common ancestor, before the divergence of eukaryotes and prokaryotes. This was extended in the exon theory by Gilbert (1987), who proposed that introns existed in early life in the RNA World. This view held that introns were subsequently extinguished from bacterial genomes due to selective pressure to streamline the genome. Unicellular eukaryotic cells such as *S. cerevisiae* may also have been under significant growth pressures and

thus lost many of their introns, although this loss was less dramatic. By contrast, higher eukaryotes generally divide more slowly, and would therefore be subject to less stringent selection against introns in the genome.

This theory is largely supported by cross-species comparisons of intron positions in orthologous genes, both within relatively closely-related groups (Bon *et al.,* 2003) and across much broader evolutionary distances (Federov *et al.,* 2002; Rogozin *et al.,* 2003). For example, orthologous genes from humans and the plant *Arabidopsis thaliana* share around 25% of intron positions (Rogozin *et al.,* 2003). Even genes from protists, which are believed to have diverged from multicellular eukaryotes early in evolution, share some introns with orthologous genes of plants and animals (Nixon *et al.,* 2002; Rogozin *et al.,* 2003).

The majority (90-95%) of these shared genes appear to be the result of evolutionary conservation and not parallel intron gain in these genes (Sverdlov *et al.,* 2005). Moreover, examination of elements of the general splicing machinery indicates that most have also been evolutionarily conserved across eukaryotes (Collins and Penny, 2005).

However, this theory does not explain the difference in intron architecture between *S. cerevisiae* and the distantly-related *Schizosaccharomyces pombe,* another single-celled eukaryote with a similar-sized genome. In *S. cerevisiae* only around 4% of genes contain introns and multiple introns in the one gene are rare (Deutsch and Long, 1999). However in *S. pombe* around 43% of nuclear genes have introns and each of these genes can contain up to 15 introns (Deutsch and Long, 1999; Wood *et al.,* 2002).

1.2.2 Intron evolution: *introns late*

In 1991, both Cavalier-Smith, and Palmer and Logsdon proposed an alternative hypothesis. They posited that introns had appeared only relatively recently in the genomes of eukaryotes, long after their divergence from prokaryotes. According to this theory, introns spread as mobile elements which insert themselves into genes at short (4-5 nt long) "proto-splice sites" (Long *et al.,* 1998). This theory was largely based on the discovery of introns with very limited phylogenetic distributions, suggesting recent gain (Logsdon *et al.,* 1995; Cho and Doolittle, 1997; Logsdon, 1998). However, while various mechanisms

have been suggested by which intron gain may have occurred (Palmer and Logsdon 1991; Cousineai *et al.,* 2000), only one example has been found that is clearly a result of the creation of an intron from a transposable event (Giroux *et al.,* 1994).

Taken together, this suggests that the most likely scenario is that both theories are, to some extent true: introns arose from a combination of conservation of ancestral intronic elements and a number of 'intron gain' events, after the evolutionary split from prokaryotic organisms. However, the exact contribution of each mechanism to the pattern of modern intron architecture has yet to be determined.

1.2.3 Function of nuclear introns

Four major classes of introns have been identified: group I self-splicing introns; group II self-splicing introns; Achaeal and/or tRNA introns; and spliceosomal introns in nuclear pre-mRNA (reviewed in Haugen *et al.,* 2005). Group I introns are widespread throughout the nuclear rDNA genes of protists, fungal mitochondria, bacteria and bacteriophages. Group II introns self-splice through a different mechanism and are found in the genomes of bacteria and organelles. Introns in tRNA are found in eukaryotic nuclei as well as in Archaea. Spliceosomal introns are the most common insertions found in eukaryotic nuclear pre-mRNA and are excised by a completely different mechanism from that of the tRNA introns. This review focuses on spliceosomal intron function.

Spliceosomal introns have long been regarded as evolutionary debris, without function in the cell. However they present three significant problems for the cell (Lynch and Richardson, 2002):

- They must be spliced out of precursor mRNAs (pre-mRNAs) with a high level of accuracy so that functional proteins can be produced
- The nucleotides contained within introns represent a large metabolic investment at the DNA and RNA levels; and
- Mutation of intronic sequences required for splicing (splice sites) increases the rate of null allele generation.

Given the prevalence of introns in higher eukaryotes, it is likely that spliceosomal introns must have some function in the cell to balance the potential cost of maintaining them in the genome.

Exon shuffling

The earliest function proposed for intronic sequence was to provide the opportunity for 'exon shuffling' (Gilbert 1987). This hypothesis suggests that introns act to break up the coding sequence of a gene into exonic 'pieces' which could then be rearranged via recombination within the intron regions to dramatically increase the speed of protein evolution. Known examples of this effect include the *lamin B* chain genes of *Drosophila melanogaster* and vertebrates (Vuolteenaho *et al.,* 1990; Chi *et al.,* 1991, Gow *et al.,* 1993), as well as the netrin receptor genes in humans and *Caenorhabditis elegans* (Cho *et al.,* 1994; Chan *et al.,* 1996). Examination of exon duplication, one type of exon shuffling, was found to account for at least 6% of all human exons (Fedorov *et al.,* 1998).

However, results from recent sequencing projects have shown a number of instances where sequences within introns may be very highly conserved across species (Bejerano *et al.,* 2004; Glazov *et al.,* 2005), more than might be expected if their only role was to increase recombination between exons.

Source of non-coding RNAs

Most snoRNAs characterised so far are found in the introns of nucleolar proteins, such as the hsc70 proteins and arise by independent transcription (Tycowski *et al.,* 1993). However some snoRNAs have been identified which originate instead from processing of the pre-mRNA in which they reside (Fragapane *et al.,* 1993; Kiss and Filipowicz 1993; Prislei *et al.,* 1992; Prislei *et al.,* 1993). Therefore spliceosomal introns are the source for at least some ncRNAs.

Source of *cis*-regulatory elements influencing gene expression

A number of elements have been found within introns of genes that are essential for the expression of those genes. For example, the second intron of

the apolipoprotein B gene in humans is required for expression of this gene in the liver (Brooks *et al.*, 1994).

It has become increasingly apparent that the presence of introns and their removal from pre-mRNA by the spliceosome can influence many other stages of gene expression, including transcription, polyadenylation, mRNA export, translational efficiency and mRNA decay. This occurs through an extensive network of coupling between the various complexes involved in mRNA metabolism (Figure 1.3). Elements required for the gene expression pathway interact both physically and functionally with other steps in the pathway (Maniatis and Reed, 2002). This ensures efficient processing of the mRNA. For example, coupling between transcription and splicing results in the tethering of splicing factors to an area immediately adjacent to where the nascent transcript is emerging (reviewed in Cramer *et al,* 2001). This can enhance transcriptional initiation through interactions between the spliceosomal U1 snRNP (bound to

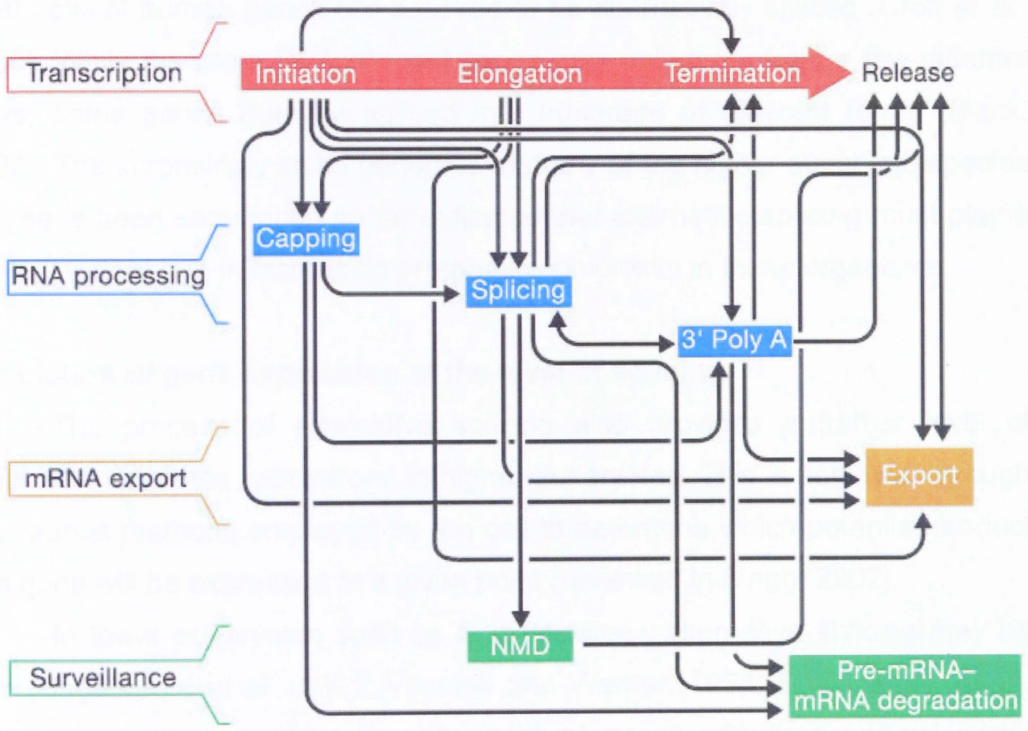


Figure 1.3: Schematic showing the network of coupled interactions involved in gene expression. The major steps of mRNA metabolism are shown on the left. 'Release' refers to the release of the mature mRNA from the transcription site. Black arrow indicate physical and/or functional coupling between two of the steps in gene expression. Taken from Maniatis and Reed (2002).

the 5' splice site) and the general transcription initiation factor TFIIH (Manley, 2002). Similarly, transcriptional fidelity may be improved for a given intron-containing transcript through spliceosomal-mediated recruitment of the protein TAT-specific factor1 (TAT-SF1), which ultimately affects the activity of the C-terminal domain of RNA polymerase II (Fong and Zhou, 2001).

The magnitude of intron-dependent effects can vary greatly between genes; from nearly nothing to a 400-fold increase in levels of mRNA (Buchman and Berg, 1988; Bourdon et al., 2001). The precise mechanisms underlying this coupling will not be discussed here in further detail but are reviewed in Le Hir et al., (2003) and Maniatis and Reed (2002).

Production of multiple proteins from a single gene

In eukaryotes, particularly higher eukaryotes, the presence of multiple introns within the one gene allows for the generation of an enormous amount of functional diversity through the process of alternative splicing of pre-mRNA. At least 35% of human genes are believed to be alternatively spliced (Croft *et al.,* 2000). While the pre-mRNA of most genes may only be spliced a few different ways, some genes may be spliced into hundreds of different forms (Black, 2000). The surprisingly small genomes of many of the higher eukaryote species that have been sequenced so far indicates that alternative splicing must play a very important role in increasing proteomic complexity in these organisms.

Regulation of gene expression at the level of splicing

The process of alternative splicing also provides a further level of regulation for genes with introns in higher eukaryotes. This is achieved through the various methods employed by the cell to determine which potential product of a gene will be expressed at a given point (reviewed in Singh, 2002).

In lower eukaryotes such as *S. cerevisiae* , alternative splicing may be rare (Engebrecht *et al.,* 1991; Vilardell and Warner, 1994; Davis *et al,* 2000). However, expression of some subgroups of genes with less optimal intron features are also controlled at the level of splicing. Such regulation is achieved by virtue of the requirement for specific ancillary splicing factors during the processing of these genes, to ensure splicing is successful (Spingola and Ares,

2000; Dahan and Kupiec, 2004). This is discussed in more detail in section 5.10.

1.3 Mechanism of splicing by the major spliceosome

The genomes of all eukaryotes contain introns which must be removed from nascent RNA transcripts by splicing. This is catalysed by the spliceosome, a large multicomponent complex comparable in size to the ribosome. Once assembled, the spliceosome is thought to be made up of 5 snRNAs and around 300 proteins (Zhou *et al.,* 2002; Jurica and Moore, 2003), making it possibly the most complex macromolecular machine in the cell (Nilsen, 2003).

Recently a second, low abundance spliceosome was identified in higher eukaryotes (Tarn and Steiz, 1997). This new 'minor' spliceosome was structurally and functionally analogous to the major-class spliceosome and acts to mediate the excision of a novel class of introns with non-canonical splice-site sequences (reviewed in Patel and Steitz 2003). However, since the work presented in this thesis focuses on regulation and splicing in yeast, the mechanism of the minor spliceosome will not be described in detail.

In the major class of introns, conserved sequences are generally found at either end of each intron (the 5' and 3' splice sites, respectively). A third sequence called the 'branchpoint', particularly highly conserved in *S. cerevisiae*, is found within the intron, usually between 20-40 nt upstream of the 3' splice site (Nilsen, 2003). The intron is removed from pre-mRNA and the remaining exons spliced together by two successive transesterification reactions, shown in Figure 1.4

The spliceosome is a highly dynamic complex which is comprised of five small nuclear ribonuclear protein particles (snRNPs) (U1, U2, U4, U5 and U6) and more than 60 non-snRNP proteins (reviewed in Burge *et al.,* 1999; Will and Lührmann, 2001). In turn, each snRNP is made up of a U-specific snRNA, core proteins and various U-specific proteins.

The process and components of pre-mRNA splicing are highly conserved from yeast to humans (Patel and Steitz, 2003). Accordingly, the spliceosome is capable of removing introns of many different lengths and with a variety of different splice site sequences. *In vitro*, spliceosome assembly and interaction

with the pre-mRNA substrate proceeds in an ordered fashion, shown in Figure 1.5 (reviewed in Brow, 2002).

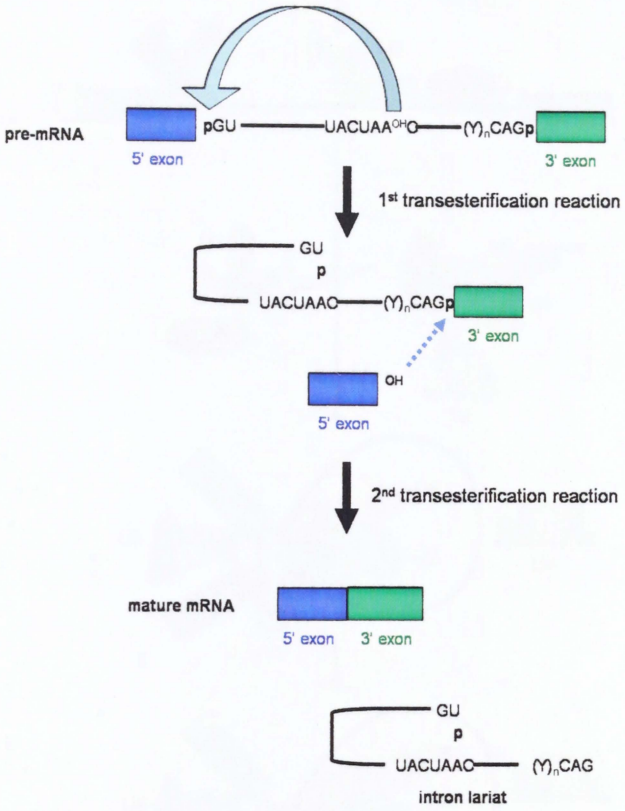


Figure 1.4: Mechanism of nuclear pre-mRNA splicing. In the first transesterification reaction the 2' hydroxyl of the branchpoint adenosine attacks the phosphodiester bond of the 5' splice site, cleaving the 5' exon-intron junction. In the second transesterification reaction the newly-freed 3' hydroxyl group on the 5' exon attacks the 3' intron-exon junction, displacing the lariat and ligating the remaining exons. Phosphate groups are shown as **'p'**. Splice site consensus sequences are taken from Lamond (1993).

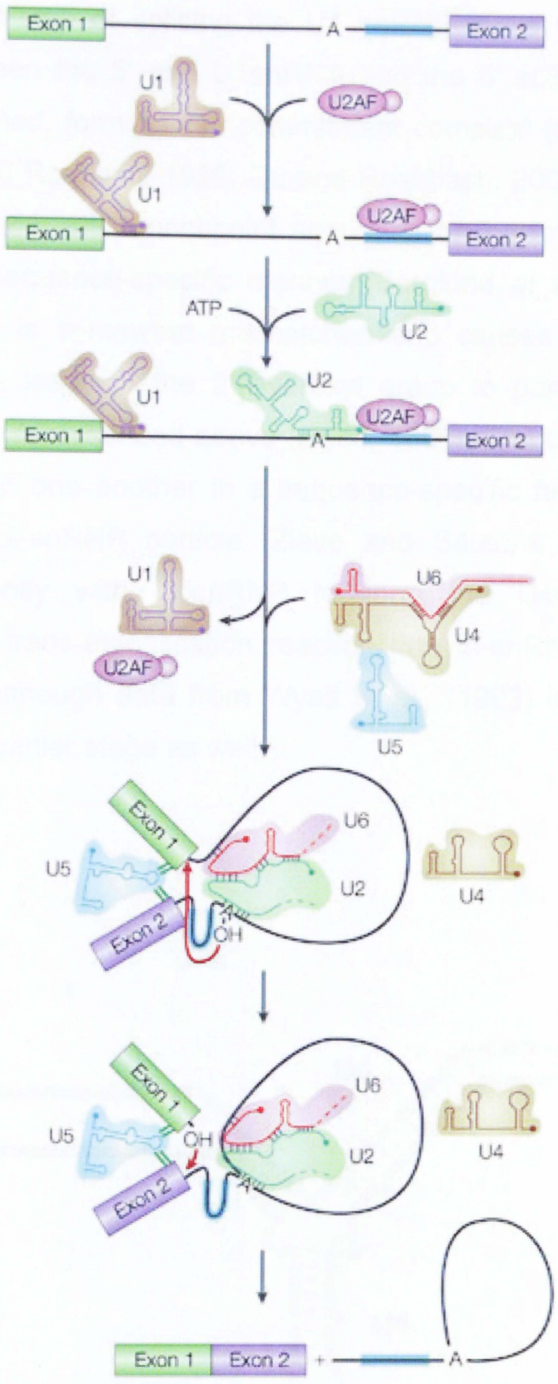


Figure 1.5: Mechanism of splicing by the major spliceosome. The two transesterification reactions are shown by red arrows. Each snRNP is shown as a small nuclear RNA (not to scale), with the 5' terminus indicated by a dot, with the surrounding shaded region representing U-specific proteins. The branchpoint sequence on the pre-mRNA is shown in blue. Figure adapted from Patel and Steitz, 2003.

In the initial step of splicing the U1 snRNP binds via complementary base-pairing between the 5' end U1snRNA and the 5' splice site of the pre-mRNA to be modified, forming the 'commitment complex' (Ruby and Abelson, 1988; Séraphin and Rosbash, 1989; Du and Roshbash, 2001). The U2snRNA can then interact with the branchpoint sequence of the intron (Parker *et al.,* 1987), also in a sequence-specific manner (Berglund *et al.,* 1997). The U2 branch-site duplex is somewhat mismatched and causes a 'bulging' of the adenosine residue, leading the 2' hydroxyl group to participate in the first nucleophilic attack, as described above (Reed, 1996). The U4 and U6 snRNAs are tightly bound to one another in a sequence-specific fashion (Figure 1.6), forming a single di-snRNP particle (Zieve and Sauterer, 1990). This then associates transiently with U5snRNP to form the U4/U6.U5 tri-snRNP. Following the first trans-esterification reaction, the tri-snRNP stably joins the pre-spliceosome, although data from Wyatt *et al.,* (1992) indicate U5 snRNP may interact at an earlier stage as well.

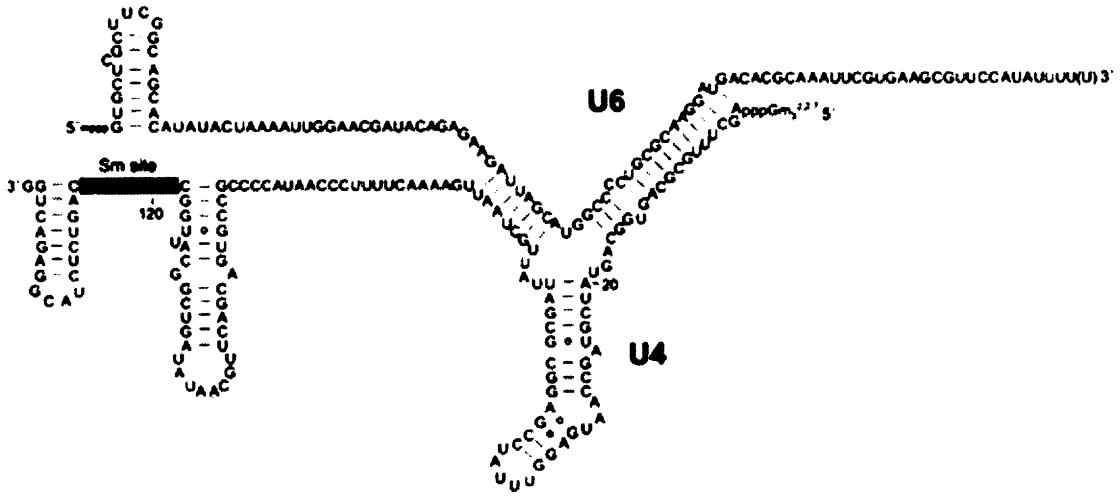


Figure 1.6: Secondary structure of U4/U6 snRNA duplex in humans, showing extensive base pairing. Adapted from Tycowski *et al.,* 2006.

The mature spliceosome is then rearranged in an ATP-dependent manner to form the catalytic core (Nilson, 1998). Base pairing between the U4 and U6 snRNAs is disrupted (Lamond, 1988), and U6 displaces U1 at the 5' splice site (Wasserman and Steitz, 1992). U6snRNP also interacts with a region of U2 snRNP near the U2 branch-site duplex. This brings the branchpoint adenosine into position for the first catalytic step of splicing, as shown in Figure 1.7 (Wu and Manley, 1991; Madhani and Guthrie, 1992). The U5 snRNP has been shown to interact weakly with splice-site sequences in both the 5' and 3' exons and is believed to position the ends of the two exons for the second step of splicing (Newman and Norman 1991; Newman and Norman 1992). The spliceosome is then rearranged to bring the 3' exon into the catalytic core for the second trans-esterification reaction, releasing both the ligated exons and intron lariat.

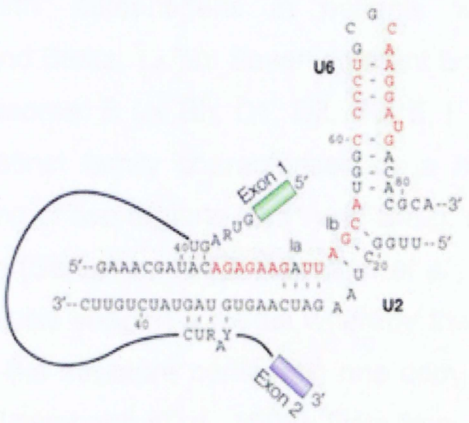


Figure 1.7: Catalytic core of the spliceosome. Highly conserved regions of human U2-U6 essential for catalysis are shown in red. Adapted from Patel and Steitz. 2003.

Following completion of splicing the spliceosome is disassembled - although this process is still poorly understood- and the spliceosomal components recycled for further rounds of splicing.

In vivo experiments have recently indicated that spliceosomal snRNPs may not bind the pre-mRNA substrate in a discrete, stepwise manner, but rather associate with one another prior to the splicing reaction (Stevens *et al.,* 2002; Nilsen, 2002)

Despite extensive research into the mechanism of pre-mRNA splicing, it is not yet known if the trans-esterification reactions are catalysed by the spliceosome snRNA or by spliceosomal proteins (Collins and Guthrie, 2000). However similarities between the mechanisms of intron removal in group II self-splicing introns supports the theory that the spliceosome is, in fact, a protein-assisted ribozyme (Fabrizio and Abelson, 1992; Yu *et al.,* 1995).

1.4 Structure of the spliceosome: Sm proteins

As mentioned previously, each spliceosomal snRNP is made up of U-specific RNA and proteins. These proteins can be divided into two groups: proteins specific to a particular snRNP (such as U1A or U2A), and core proteins (Lührmann *et al.,* 1990). In U1, U2, U4 and U5, these core proteins are the Sm proteins.

Sm proteins were originally discovered due to their reactivity with autoantibodies to 'Smith' autoantigens in patients with systemic lupus erythematosis (Lerner and Steitz, 1979). Seven different Sm proteins have been implicated in the spliceosome: B (or B'), D1, D2, D3, E, F and G proteins. The Sm proteins form a distinct family characterised by a highly conserved Sm sequence motif consisting of two regions (Sm1 and Sm2), separated by a linker region of variable length (Séraphin, 1995; Hermann *et al.,* 1995). Data from X-ray crystallographic analysis support a model whereby these Sm proteins form a heptameric doughnut-like structure containing one copy of each Sm protein, as shown in Figure 1.8 (Kambach *et al.,* 1999). Data from electron microscopic analysis of U1 snRNP also support this model, showing a ring-shaped main body large enough to accommodate the Sm ring structure, as well as regions corresponding to the size of known U1-specific proteins (Will and Lührmann, 2001). While the inner ring of the Sm complex has a high concentration of positive charges, as well as a central hole large enough for a single stranded RNA to fit (20 Å in diameter), the exact manner in which U1, U2,U4 or U5-specific snRNA associates with the Sm complexes has not yet been elucidated (Kambach *et al.,* 1999).

With the exception of U6 snRNA, all U-snRNAs are transcribed by RNA polymerase II. During transcription they also acquire a *N*7-methyl-guanosine (m^7G) cap structure. The U-snRNAs are then exported to the cytoplasm, where

the Sm proteins assemble around a conserved, uridyl-rich sequence present in U1, U2 U4 and U5 snRNAs, called the 'Sm site' (Hamm *et al.,* 1987, Raker *et al.,* 1999). Five Sm proteins form the D1.D2.E.F.G protein complex, which then binds to the Sm site of the U1snRNA (Raker *et al.,* 1996). When the D3.B dimer is finally added, a stable snRNP core is formed. Specificity of the Sm core for the Sm-binding site in the UnRNAs appears to be mediated by the 'survival of motor neurons' or SMN protein complex (Yong *et al.,* 2004)

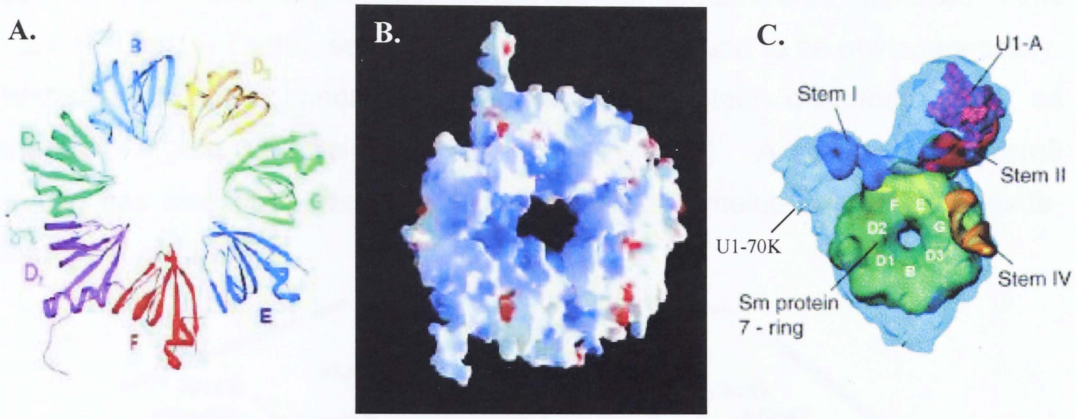


Figure 1.8: Structure of Sm spliceosomal complex. A. Ribbon diagram of the heptamer model showing the seven core proteins, with order determined by biochemical and genetic experiments (Kambach *et al.,* 1999). B. Surface representation of the heptameric ring model showing electrostatic potential (red is negative, blue is positive).C. Model of 3-D arrangement of RNA and proteins in the U1 snRNP. Precise positions of U1snRNA stem-loops I and II have not yet been confirmed. Also shown are the likely positions of U1-specific proteinsU1-70K and U1-A (A. and B adapted from Kambach *et al.,* 1999; C adapted from Will and Lührmann, 2001).

Once bound to the snRNA, Sm proteins trigger the hypermethlyation of the cap structure (Plessel *et al.,* 1994). The hypermethlated cap and the presence of the Sm core complex act as two independent nuclear localisation signals, via interactions of part of the Sm motif with importin β (Palacios *et al.,* 1997) and the m_3G cap with snurportin1 (Huber *et al.,* 1998). This allows movement of the assembled snRNP back into the nucleus to form part of the spliceosome.

1.5 Structure of the spliceosome: Lsm proteins

Unlike other spliceosomal snRNAs, U6 snRNA is transcribed by RNA polymerase III, acquires a 5' γ-monomethyl phosphate cap, and is thought to be retained in the nucleus (Boelens *et al.,* 1995). Additionally, U6 snRNA does not associate directly with any Sm protein. Instead, the core structure in the U6 snRNP is formed by a ring of 'Like-sm' or Lsm proteins (Achsel *et al.,* 1999; Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999).

Lsm proteins were originally identified by database searches aimed at finding proteins with homology to the Sm motif in *S. cerevisiae* (Salgado-Garrido *et al.,* 1999). Eight highly conserved Lsm proteins were identified in this manner (Lsm1 – Lsm8), seven of which were also found to be phylogenetically-related to the seven canonical spliceosomal Sm proteins described above, as shown in Figure 1.9 (Salgado-Garrido *et al.,* 1999). A yeast-specific Lsm9 protein has also been identified, but its function remains unknown (Salgado-Garrido *et al.,* 1999).

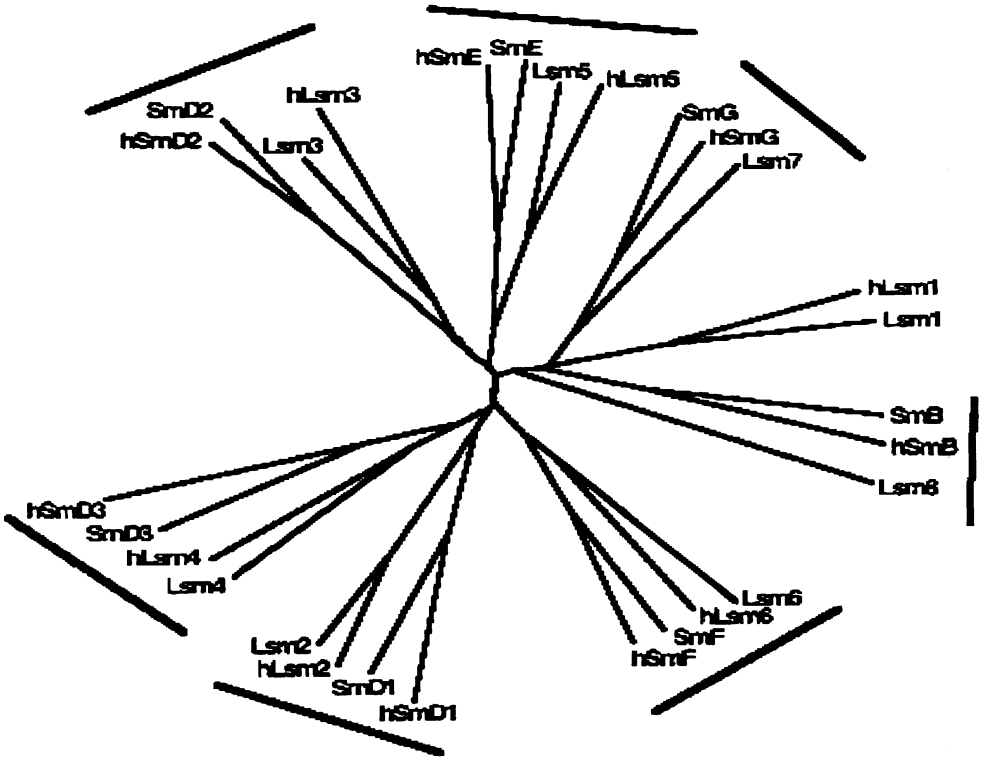


Figure 1.9: Phylogenetic tree of yeast, human and archaeal Sm and Sm-like proteins. This tree was based on multiple protein sequence alignments of the various Sm domains. Human proteins are labelled with an 'h' prefix. In each case it is apparent that the yeast Lsm or Sm proteins are more closely related to its human equivalent than any other gene. Also each Sm protein appears to be more similar to specific Lsm protein than it is to other Sm proteins. Human homologue for Lsm7p and Lsm8p are not shown as, when this study was performed, they had not had been confirmed. They have since been identified. Image taken from Salgado-Garrido *et al.* (1999).

Lsm2 - Lsm8 form a heptameric complex which binds specifically to a region of sequence at the 3' terminus of the U6 snRNA (Salgado-Garrido *et al.*, 1999). This stabilises the U6 snRNA and allows the assembly of a functional U6 snRNP (Mayes *et al.*, 1999; Salgado-Garrido *et al.*, 1999). The spliceosomal Lsm complex is additionally thought to be required for efficient regeneration of U4/U6 snRNAs duplexes following a round of splicing, as well as the formation of functional U4/U6.U5 tri-snRNPs (Verdone *et al.*, 2004).

The exact structure of this Lsm protein complex has yet to be experimentally confirmed. However, sequence similarity with the spliceosomal Sm proteins, along with electron microscopy data, suggest a doughnut-shaped structure, with each of the seven spliceosomal Lsms present in stochiometric amounts (Salgado-Garrido *et al.*, 1999; Achsel *et al.*, 1999; Kambach *et al.*, 1999). A model for the structure of the spliceosomal Lsm complex is presented in Figure 1.10. Unlike the Sm complex, the Lsm complex assembles entirely within the nucleus and does not require the presence of U6 snRNA (Achsel *et al.*, 1999).

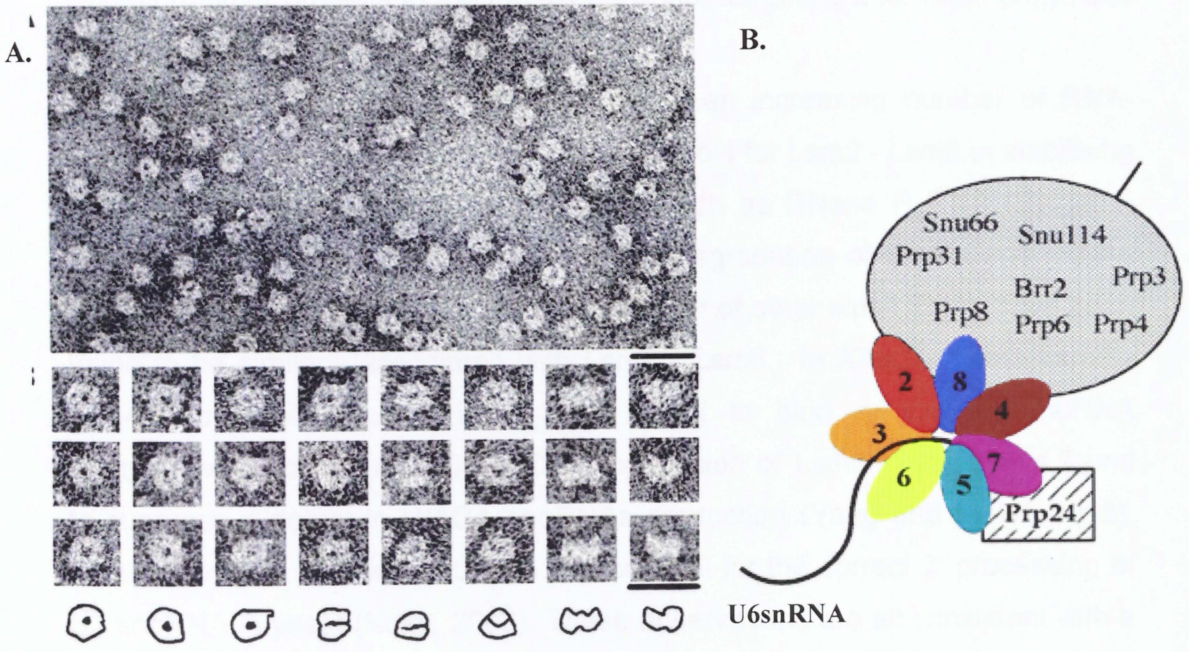


Figure 1.10: Structure of the spliceosomal Lsm protein complex. A. Electron microscopic analysis of human spliceosomal Lsm complexes. The bar at the top represents 25nm. The lower bar represents 20nm. Image taken from Achsel *et al.* (1999). **B.** Proposed structure of spliceosomal Lsm complex. The ring-shaped complex is shown binding to the 3' region of U6snRNA, and associating with other U-specific splicing factors. Image adapted from Bouveret *et al.* (2000).

1.6 Other roles for Lsm proteins

Lsm proteins have also been implicated in another heptameric ring complex involved in RNA processing: The Lsm1- 7 complex has been found to interact with mRNA decapping enzymes (Dcp1 and Dcp2) and a 5'-3' exoribonuclease (Xrn1) (Bouveret *et al.,* 2000; Tharun *et al.,* 2000). Together, these elements form part of a large complex in the cytoplasm which participates in deadenylation-dependent decapping and 5'-3' exonucleolytic degradation of mRNA following protein synthesis (Bouveret *et al.,* 2000; Tharun *et al.,* 2000). While very little is known about the structure of this Lsm degradation complex, it seems likely that it may also form a ring-shape, with Lsm1 replacing Lsm8 in the structure. The Lsm1 - Lsm7 complex binds mRNA directly, but the mechanism by which it acts to degrade the mRNA has not yet been elucidated (Tharun *et al,* 2000). It has been proposed that the Lsm1 - Lsm7 complex may promote ribonucleoprotein rearrangements, leading to the dissociation of the cytoplasmic cap-binding complex, and decapping of the mRNA (Tharun *et al.,* 2000; Schwartz and Parker 1999). Alternatively this Lsm complex could assist the binding of other proteins to the mRNA transcript, targeting it for rapid enzymatic cleavage (Bouveret *et al.,* 2000).

Lsm proteins have been implicated in an increasing number of RNA-processing events in the cell. These include a role for Lsm2 - Lsm8 in stabilising nascent polymerase III-transcribed genes, such as RNase P RNA (Salgado-Garrido *et al.,* 1999) or the processing and degradation of tRNAs and rRNAs (Kufel *et al.,* 2002; Kufel *et al.,* 2003). A number of other small RNAs have been found to be similarly associated with Lsm2 - Lsm8. In *Xenopus* both Lsm2 - Lsm4 and Lsm6 - Lsm8 have been found to bind to the U8 snoRNA (Tomascevic and Peculis, 2002). Over-expression of Lsm5 in yeast was found to suppress a defect in H/ACA snoRNAs production (Yang and Meier, 2003). Finally, Lsm2 - Lsm5 was found to be essential for the correct 3' processing of U3 snoRNA in yeast (Kufel, 2003). These observations are all consistent with a role for Lsm proteins as chaperones, stabilising and assisting in RNA-protein assembly (Kufel *et al.,* 2003)

Lsm proteins have been identified in all branches of life (Collins *et al.,* 2001; Moller *et al.,* 2002; Toro *et al.,* 2001). Even in Archaebacteria they form homohexameric ring structures which are remarkably similar to the Sm and Lsm complexes described above (Toro *et al.,* 2001; Toro *et al.,* 2002). This indicates these proteins may form part of an ancient family, stretching back into the last universal common ancestor (Khusial *et al.,* 2005). For example, bacterial Hfq shows striking similarity to Lsm and Sm proteins. It forms homohexameric ring-shaped complexes which affect multiple processes in the cell, including degradation of some mRNAs (via binding to the 3' region of the gene) and stabilisation of others (Beggs, 2005). In Achaeal genomes there are often two types of Lsm/Sm proteins present, indicating the Sm protein family may have evolved from a single ancestor which was present before Achaea and eukaryotes diverged. (Achsel *et al.,* 2001). As stated previously U6 snRNA binds to an Lsm complex and is the highly conserved catalytic component of the spliceosome. Therefore, of the two types of protein in the Sm family, it has been suggested that Lsm proteins may be the older form (Khusial *et al.,* 2005).

1.7 Transcriptional regulation in yeast

This section presents an overview of regulation of transcription in the yeast cell, to explain how multiple protein-coding genes may be potentially positively or negatively co-regulated under different conditions. This process has been extensively reviewed by Struhl (1995) and Hampsey (2000).

The initiation of mRNA transcription in eukaryotes requires the assembly and binding of a protein structure as large and complex as the ribosome (Struhl, 1999). This binding is reversible and may be either enhanced or repressed via the activities of a large number of proteins called transcription factors (TFs). In *S. cerevisiae*, more than 250 transcription factors are thought to be involved in the regulation of around 6000 protein-coding genes (Dolinski *et al.,* 2004).

Transcription in *S. cerevisiae* involves the activity of one of three possible DNA-dependent RNA polymerases: RNA polymerase I, II or III. Protein coding genes are transcribed by RNA polymerase II (RNAPol II). The mechanism of RNAPol II transcriptional initiation is highly conserved across eukaryotes (McKnight and Yamamoto, 1992) and is illustrated in Figure 1.11. Briefly, it proceeds as follows:

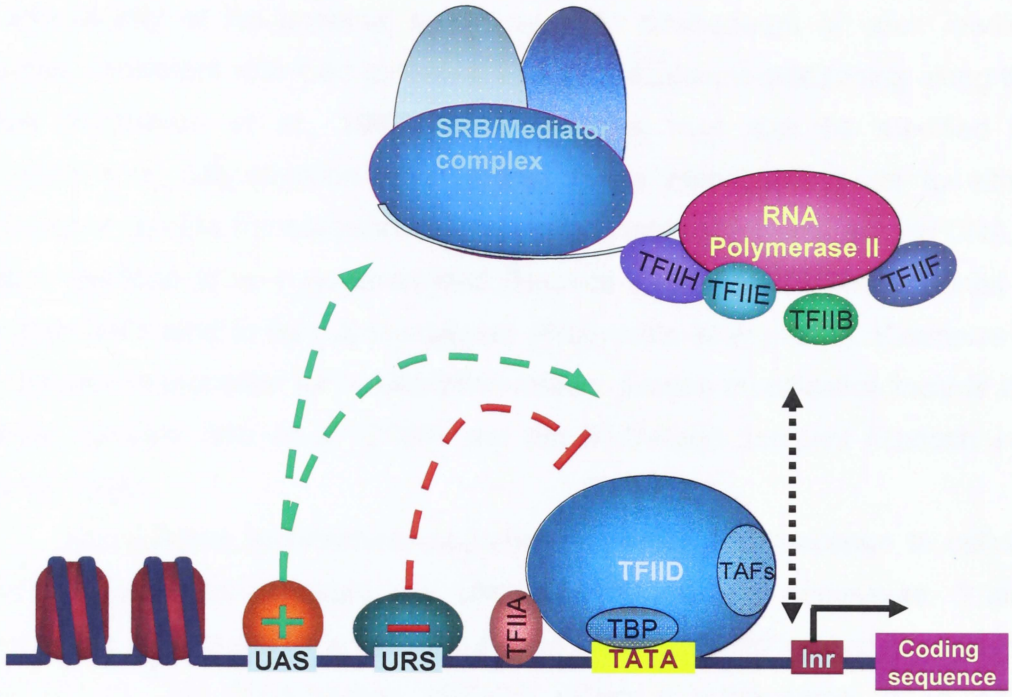


Figure 1.11: Schematic of a TATA-based promoter in yeast. Following the removal of nucleosomes from the area, elements of the transcription machinery bind to the TATA box and Inr to initiate transcription. General transcription machinery shown here includes RNA polymerase II and the general transcription factors (TBP, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH). The SRB/MED complex associates with RNA polymerase II through the C-terminal repeat domain of Rpb1, forming a holoenzyme complex capable of activation by other factors (Hampsey, 2000) Gene specific regulatory factors generally bind upstream of the TATA box to UAS or URS to enhance or repress transcription of the gene.

1.7.1 Chromatin modification

In the nucleus of all eukaryotic cells DNA is packaged into a dense nucleoprotein complex known as chromatin. The fundamental unit of chromatin is the nucleosome; in the yeast nucleus this consists of 146bp of DNA coiled around a core of 8 histone proteins and a nucleoprotein linker region (Zlatanova *et al.*, 1999; Kornberg and Lorch, 1999). Chromatin has a generally repressive

effect on transcription and nucleosomes must be displaced from around a gene in order for it to become transcriptionally active (Kuras and Struhl, 1999; Li *et al.,* 1999). Tracts of poly (dA-dT) elements are sometimes found in the regions immediately adjacent to genes and appear to enhance transcriptional activation, possibly by impairing local nucleosome stability (Iyer and Struhl, 1995). Indeed, genome-wide analysis of *S. cerevisiae* sequence reveals such tracts are found predominantly at nucleosomal length up- and downstream of open reading frames, consistent with their proposed role in nucleosome positioning along the DNA (Raghavan *et al.,* 1997). Nucleosomes may also be modified by acetylaton or deacetlyation of histones. These modifications act to either activate or repress transcription, respectively – histones associated with DNA at active loci tend to be hyperacetylated (Hebbes *et al.,* 1994), while histones at inactive sites tend to be hypoacetylated (Braunstein *et al.,* 1993). Examples of complexes responsible for transcription-related histone modification include the SAGA complex (Wu *et al.,* 2004) and the Sin3-Rpd3 complex (Kadosh and Struhl, 1998).

Nucleosome architecture may also be modified in response to cellular and environmental changes via chromatin remodelling complexes. These complexes are generally recruited to promoters by specific transcription factors and act to alter DNA-histone contacts within a nucleosome in an ATP-dependent manner (Martens and Winston, 2003). The main chromatin remodelling complexes in *S. cerevisiae* are SWI/SNF and the RSC complex. Both of these complexes can act to either activate or repress transcription of genes in different contexts (Imbalzano *et al.,* 1994; Martens and Winston, 2002; Damelin *et al.,* 2002).

1.7.2 Binding of transcriptional machinery to the core promoter region

The DNA sequence immediately adjacent to a gene is referred to as the promoter region and is made up of two functionally distinct elements: the core promoter region and regulatory sequences.

Once the core promoter region becomes accessible, the TATA-binding protein (TBP) binds to an AT-rich site called the TATA box (consensus TATAA; Chen and Struhl, 1988), which is located 25 - 30 bp upstream of the site of transcriptional initiation in most eukaryotes but, unusually, 40 -120 bp upstream

in *S. cerevisiae* (reviewed in Struhl, 1995). The TBP forms part of the TFIID general transcription factor (GTF), which also includes at least 14 other TBP-associated factors (TAF) (Buratowski, 1994; Hernandez, 1993).

TATA box elements are by no means universal – in *S. cerevisiae* they have only been identified in the promoter sequences of around 20% of genes (Basehoar *et al.,* 2004). In TATA-less promoters the TFIID instead interacts with alternative element located around 30 bp downstream of the transcription start site, called the downstream promoter element (DPE) (Burke and Kadonaga, 1996).

After the TFIID has bound, the other members of the basal transcription machinery begin to associate with it and each other to form the transcription preinitiation complex (PIC) (Roeder, 1996). This assembly proceeds in a highly ordered manner. TFIIA binds immediately 5' of TFIID, stabilising the interactions of TBP to the TATA box. TFIIB then binds specifically to the DNA 3' to TFIID (Nikolov *et al.,* 1995) and, along with the Rpb1 subunit of RNAPol II, selects the site of transcriptional initiation. Interestingly, when TFIIB and Rpb1 subunits from *S. cerevisiae* are substituted with homologues from *S. pombe* transcription initiation sites are selected at a distance typical of *S. pombe* genes instead (25-30 bp from the TATA box) (Li *et al.,* 1994). TFIIF, TFIIE, and TFIIH are then recruited to DNA, complexed to RNAPol II (Corden, 1993; Zawel and Reinberg, 1993). Through the activities of these GTFs, RNA polymerase II can bind to the initiator sequence (Inr – consensus TCG/AA and RRYRR), which encompasses the transcription start site (Furter-Graves and Hall, 1990; Weis and Reinberg, 1992). TFIIH contains three separate enzymatic activities thought to be involved in promoter activation and clearance by the polymerase: a DNA–dependent ATPase (Roy *et al.,* 1994); a helicase (Schaeffer *et al.,* 1993); and a CTD kinase (Feaver *et al.,* 1991).

Yeast RNAPol II is composed of 12 highly conserved subunits, some so highly conserved that human RNAPol II can functionally substitute for yeast homologues (Woychik and Young, 1994; McKune *et al.,* 1995). Recent studies show that the large two large subunits, Rpb1p and Rpb2p appear to form a positively charged cleft through which the negatively charged DNA-template is channelled (Chen *et al.,* 2004). As transcription progresses and the first RNA phosphodiester bonds are formed, the promoter is cleared. This occurs by disruption of the protein-protein and protein-DNA interactions of the initiation

complex through the enzymatic activities of TFIIH (Zurita and Merino, 2003). The nascent transcript is then extended by the complex process of elongation, reviewed in Shilatifard (2004).

RNA polymerase II (RNAPol II) requires the other GTFs (TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH) to initiate transcription, even at a basal level (Weil *et al.,* 1979). The binding of elements to the core promoter region is sufficient for accurate initiation of transcription *in vitro* (Struhl, 1999), although the amount of transcript produced in these cases is minimal (Ogbourne and Antalis, 1998).

1.7.3 Activators

Normal levels of transcription of most (possibly all) eukaryotic protein genes require the binding of activator factors (usually proteins) to specific sequences between 6-30 bp in length (Struhl, 1995). These sequences are gene-specific and are located at variable distances from the core promoter. In yeast these elements, called upstream activator sequences (UAS) may function in either orientation and typically lie between 50-500 bp upstream of the initiation site (Verdier, 1990). Exceptions to this rule include the UAS of the *HO* promoter, which may lie up to 1000 – 1400 bp from the Inr (Nasmyth, 1985) and the intragenic control elements of *LPD1* (Sinclair *et al.,* 1994). In higher eukaryotes these sequences are called enhancers, and may function either up- or downstream of the gene.

Upon binding to their cognate sites, activator proteins act to stimulate transcription in a variety of different ways. These include the recruitment of chromatin remodelling complexes to alter promoter accessibility, or enhancing the binding of elements of the transcription initiation apparatus (Kuo *et al.,* 1998; Kingston, 1999; Ptashne and Gann, 1997).

Transcription activator protein factors are usually comprised of two physically and functionally distinct domains: a DNA-binding domain and an activation domain (Ptashne, 1986). DNA-binding domains of transcription factors tend to have certain protein motifs which form secondary structures that facilitate binding to DNA. More than 30 different families of such motifs have been identified in transcription factors, including 'zinc-finger', 'helix-turn-helix' 'helix-loop-helix' and 'leucine zipper' structures, (reviewed in Luscombe *et al.,* 2001).

In yeast, the activation domain of transcription factors are often acidic, to enhance interactions with the basic RNAPol II complex. Activation domains in higher eukaryotes may also contain glutamine-rich or proline-rich regions (Struhl, 1993).

Activators often act via co-activator proteins which are not bound directly to the DNA, such as Nhp6p (Roeder, 1991; Muscat *et al,* 1998). This protein appears to modify the local chromatin architecture to aid the binding of specific transcription factors to their cognate UAS sequences (Travers, 2003).

A more general transcriptional co-activator is the mediator complex ('SRB-MED'), which consists of at around 21-24 separate subunits (Myers and Kornberg, 2000). RNA polymerase II has a unique set of tandem heptapeptide repeat sequences at the C-terminus of the largest subunit, called the C-terminal domain (CTD; reviewed in Young, 1991). Amongst other functions, these repeats allow RNA polymerase II to associate with the SRB-MED complex, forming the RNA polymerase II 'holoenzyme'. This holoenzyme is then capable of responding to other transcriptional activators (reviewed in Hampsey, 2000). The mediator complex is highly conserved in both structure and function (Bourbon *et al.,* 2004), and is required for the regulated transcription of nearly all RNA polymerase II-dependent genes in *S. cerevisiae* (Bjőrkland and Gustafsson, 2005).

1.7.4 Repressors

Conversely, transcription factors can also act to repress gene expression, directly, or via cofactor proteins or complexes. A wide variety of transcriptional repressors have been identified in eukaryotes, many of which are highly conserved across species (Lee and Young, 2000). Many of these repressors factors bind to specific regions within the promoter (upstream repressor sequences, or URS).

Repressors may function in either an active or passive manner. Passive repression arises when gene-specific repressors bind to activators or compete with them for coincident or overlapping binding sites on the DNA (Vincent and Struhl, 1992; Lee and Young, 2000). Repressors which bind between UAS and TATA elements may also inhibit transcription by interfering with the interaction between upstream activators and the basic transcription machinery (Brent,

1984). Alternatively repressors may actively interfere with assembly of transcriptional initiation machinery on the DNA, (Maldonado *et al.*, 1999, Lee and Young, 2000), For example, Mot1 binds to TBP-DNA complexes, causing the TBP to dissociate from the DNA in an ATP-dependent manner (Auble *et al.,* 1997).

Lastly, repressors can act on the chromatin state of the promoter region, either by preventing nucleosome displacement from the region or by recruiting chromatin-modifying proteins to increase local chromatin structure, as observed for *SIN3* and *UME6* in *S. cerevisiae* (Rundlett *et al.,* 1998; Kadosh and Struhl, 1998).

Regulated gene expression results from the balance of repressor and activator-mediated activities *in vivo*, and controls both the timing and level of transcription.

1.8 Co-regulation of ribosomal protein genes

One well-characterised example of transcriptional co-regulation is the controlled expression of the ribosomal protein (RP) genes in *S. cerevisiae*. In the rapidly-dividing yeast cell there are nearly 200,000 ribosomes and RP gene transcripts account for 20 of the 30 most abundant RNAs (Velculescu *et al.,* 1997; Warner *et al.,* 1999). In addition, since ribosomal proteins are highly charged and capable of binding nucleic acids, any excess of these proteins may be profoundly deleterious to the cell. Accordingly, the expression of RP genes is stringently regulated and exquisitely sensitive to changes in environmental conditions, such as carbon source (Morrow *et al.,* 1993; Neuman-Silbergerg *et al,* 1995) or nitrogen source (Kraig *et al.,* 1982) in the growth media, or exposure to a number of environmental stress conditions (Causton *et al.,* 2001)

Control of RP production is mostly exerted at the level of transcription (Raué and Planta, 1991). RP gene transcriptional co-regulation requires the activity of protein kinase A and the TOR pathway, although the exact mechanisms involved have not yet been elucidated (reviewed in Thomas and Hall, 1997).

In *S. cerevisiae* there are around 80 different ribosomal protein genes scattered across the chromosomes (Planter and Mager, 1998). The promoter regions of these genes share several distinct regulatory elements, shown in Figure 1.12: a T-rich region; binding sites for either the Rap1 or Abf1

transcription factors (also shared by non RP genes) and an IFHL site (TC (C/T)GCCTA) which is involved in the recruitment of Fhl1 and Ifhl transcription factors (Gonçalves *et al.,* 1995; Wade *et al.,* 2004). Rap1 cannot be the specific regulator for transcription of the RP genes as many Rap1-activated promoters are not coordinately regulated with RP genes. By contrast, the transcription factors for Fhl1 and Ifh1 associate almost exclusively with RP promoters (Li *et al.,* 2002; Wade *et al.,* 2004).

It has been proposed that the T-rich region and the Rap1 or Abf1 transcription factors act to modulate the chromatin around the RP genes, enhancing the ability of a TBP-associated factor, $yTAF_{II}145$ to bind the TBP and recruit other key elements into the TFIID complex (Reese *et al.,* 1994; Planta, 1997). The $yTAF_{II}145$ protein is not essential for the transcription of most yeast genes. However it is specifically required for the expression of genes involved in the cell cycle and growth control (Walker *et al.,* 1997).

Both Rap1 and Abf1 transcription factors have also been implicated in the transcriptional activation of genes involved in cellular growth rate (Buchman *et al.,* 1988; Gailus-Durner *et al.,* 1996) and metabolism of different carbon sources (Chambers *et al.,* 1990 Lieb *et al.,* 2001). Therefore, by altering Rap1, Abf1 or $yTAF_{II}145$ protein abundance or activity, the cell can coordinate the expression of a number of key elements required for cellular growth.

Specific control of the RP genes may be provided by the additional activities of Fhl1 and Ifh1 transcription factors. In particular, Ifh1 is thought to act as a specific (co)activator for transcription of these genes (Schawalder *et al.,* 2004). The current model for this specific co-regulation proposes that, under favourable growth conditions, the essential co-activator protein Ifh1 is recruited to RP gene promoters through an interaction with one of the domains of the promoter-bound Fhl1 protein (Schawalder *et al.,* 2004). Ifh1 binding then decreases under poorer nutrient conditions via a mechanism related to TOR pathway inhibition, leading to downregulation of the RP genes. One theory for how this may occur postulates that the TOR pathway (or protein kinase A) is required to maintain the phosphorylation of a putative Fhl1-binding domain in Ifh1 (Schawalder *et al.,* 2004). Ifh1 dissociation, coupled with the short half-lives of RP mRNA, would permit rapid regulation of RP gene expression. However Fhl1 and Ifh1 do not appear to associate with the promoters of some co-

ordinately regulated RP genes, indicating that alternative regulatory mechanism are also likely to exist (Wade *et al.*, 2004).

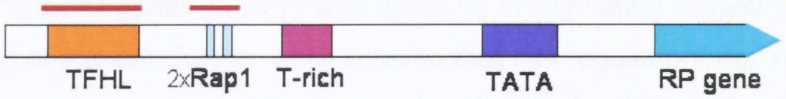


Figure 1.12 Schematic of a typical ribosomal protein promoter in yeast. Most ribosomal protein genes have two Rap1 binding sites in their promoters, along with a T-rich stretch and a TATA box. IFHL sites are usual found 100-200 bp 5' to the Rap1 sites (Tanay *et al.*, 2005) Expression of the ribosomal protein gene only occurs if neither of the regions under the red bars is blocked. Image from Warner (1999).

1.9 Aims

Splicing is an essential process in all eukaryotes. Both the mechanism of splicing and the elements involved have been highly conserved across evolution. However, while the assembly and function of the spliceosome has been examined in some detail, less is known about the regulation of synthesis of spliceosomal components.

The principle aim of this thesis was to examine the regulation of a group of *LSM* genes encoding spliceosomal proteins in *S. cerevisiae*. The introns in this organism are less prevalent than in higher eukaryotes, allowing greater scope for experimentation on elements of spliceosomal machinery. This is explored in more detail in section 5.6.

Lsm proteins form a complex at the core of the spliceosome that is required for cell viability. Therefore the initial aim of this work, described in Chapter 3, was to determine if the genes encoding these proteins were co-regulated at the level of transcription. *LSM* gene transcript levels were examined under a variety of conditions likely to induce changes in expression of a wide range of genes. To explore the extent of any possible co-regulation, the expression of genes encoding other spliceosomal proteins were also measured. This study determined that most of the spliceosomal *LSM* proteins were transcriptionally co-regulated in response to changing environmental conditions.

The abundance of some other key spliceosomal elements, such as U6 snRNA, changed in a similar manner, even though U6 snRNA is transcribed by RNA polymerase III.

Chapter 4 examined the mechanism by which the co-regulation of this set of *LSM* genes was achieved. In the course of this research it was found that one of the *LSM* genes *(LSM7)* was involved in regulating transcription of other members of the set. Further work showed that this regulation required the *LSM7* intron but was not dependent on the Lsm7 protein. This is the first reported example of an intron acting *in trans* to regulate other genes. The manner in which the *LSM7* intron mediates the transcriptional co-regulation of other *LSM* genes was also addressed.

The Lsm complex and U6snRNA are at the catalytic centre of the spliceosome. Accordingly, the aim of the work presented in Chapter 5 was to investigate the impact of regulating *LSM* genes on cellular splicing. These findings were related to the changes in the requirement for splicing in the cell under different growth conditions. *LSM7*-mediated changes in splicing capacity may affect the transcripts of a large proportion of genes with introns in this organism. Therefore the final aim of this work was to explore the wider ramifications of this regulation on cell physiology and growth.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 General reagents

Sources of reagents not noted in the specific description of the methods are shown in Table 2.1. General chemicals were purchased from Ajax Chemicals, BDH or Sigma Aldrich (Australia) and were of analytical grade.

2.1.2 General-use buffers

In all cases, water was purified using the MilliQ™ system (Millipore). Phosphate buffered saline (PBS) consisted of 0.8% sodium chloride, 0.144% disodium hydrogen phosphate, 0.024% potassium dihydrogen phosphate and 0.02% potassium chloride and had a pH of 7.4.

2.1.3 General procedures

Media, glassware, non-disposable plastic-ware and heat-stable solutions were sterilised by autoclaving for 15 min at 120°C (125 kPa). Heat labile solutions were sterilised by passage through a 0.22 μm sterile filter (Millipore).

DNases in general material were inactivated by autoclaving as above. RNases were inactivated by treatment with 0.1 % v/v diethylpyrocarbonate (DEPC) at 30°C overnight. Any remaining unreacted DEPC was then destroyed by autoclaving at 120°C (125 kPa) for 20 min.

Biological waste was autoclaved prior to disposal.

2.2 Strains and growth conditions

2.2.1 *Saccharomyces cerevisiae* strains

S. cerevisiae strains used in this study are shown in Table 2.2. Strains in each case were in the BY4742 background (Brachmann *et al.,* 1998).

Table 2.1: Materials and reagents used in this study

Material/Reagent	*Source*
4,6,diaminidino-2phenylindole dihydrochloride (DAPI)	Sigma-Aldrich, NSW
Agar type 1	Oxoid Ltd., NSW
Agar type 3	Oxoid Ltd., NSW
Agarose (DNA grade)	Progen Industries, Qld
Ammonium sulphate	Difco, NSW
Bacteriological peptone	Oxoid Ltd., NSW
Chloroform	APS Chemicals Ltd., NSW
D-glucose	Ajax FineChem, NSW
Diethyl pyrocarbonate (DEPC)	Sigma-Aldrich, NSW
Ethanol	Ajax FineChem, NSW
Ethidium bromide	Sigma-Aldrich, NSW
Ethylenediaminetetra-acetic acid di-sodium salt	APS Chemicals Ltd., NSW
Geneticin (G418)	Progen industries, Qld
Glycerol	APS Chemicals Ltd., NSW
Herring sperm DNA	Roche Diagnostics (NSW)
Lithium acetate	Sigma-Aldrich, NSW
Magnesium chloride	APS Chemicals Ltd., NSW
Magnesium sulphate	BDH Laboratory Supplies, Poole, UK
Media supplements (amino acids)	Sigma-Aldrich, NSW
Phenol	Wako Pure Chemical Industries Ltd., Japan
Polyethylene glycol 3350 (PEG)	BDH Laboratory Supplies, Poole, UK
Potassium acetate	Mallinckrodt Specialty Chemicals, USA
Potassium chloride	APS chemicals Ltd., NSW
Sodium chloride	Ajax FineChem, NSW
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich, NSW
Sodium hydroxide	APS Chemicals Ltd., NSW
Tris	Merck Pty. Ltd., NSW
Yeast extract	Oxoid Ltd., NSW
Yeast nitrogen base (without amino acids or ammonium sulphate)	Difco, NSW

2.2.2 Strain growth and maintenance

Yeast was routinely propagated on YEPD medium containing 2% (w/v) D-glucose, 2% (w/v) bactopeptone and 1% (w/v) yeast extract. For yeast propagation in anaerobic conditions, YEPD plates were supplemented with ergosterol and Tween 80 by addition of 5 ml/l of a stock solution (ergosterol at 4

mg/ml and Tween 80 40% (v/v) in ethanol). Cells grown in the presence of carbon sources other than D-glucose were grown on medium containing 2% (w/v) bactopeptone and 1% (w/v) yeast extract as well as 2% (w/v) D-raffinose, 2% (w/v) potassium acetate, 3% (v/v) glycerol or 3% (v/v) 72% sodium lactate.

Cells tested for various responses to anaerobic or aerobic growth conditions were grown on simple defined (SD) medium. SD medium was also employed at various steps in the construction of the various *lsm7* deletion mutants. SD media contained 2% (w/v) D-glucose, 0.5% (w/v) ammonium sulphate and 0.17% (w/v) Yeast Nitrogen Base. This medium was further supplemented with the following amino acids, purines and pyrimidines for selection according to their auxotrophic requirements.

Table 2.2: Strains used in this study

Strain	*Genotype*	*Source*	*Accession number*
BY4742 (wild type)	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*	Euroscarf	Y10000
prp17Δ	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; prp17::kanMX4*	Euroscarf	Y14201
lsm1Δ	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; lsm1::kanMX4*	Euroscarf	Y11301
lsm6Δ	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; lsm6::kanMX4*	Euroscarf	Y14214
lsm7Δ	*MATα; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; lsm7::kanMX4*	Euroscarf	Y17383
lsm7Δu	*MATα; his3Δ1; leu2Δ0; met15Δ0; lsm7::URA3*	This study	-
lsm7 intronless	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lsm7Δ*intron	This study	-
lsm7 intron only	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lsm7Δ*exon1*Δ*exon2	This study	-

Solid media contained either 2% (w/v) Oxoid agar type 1 (for SD media) or Oxoid type 3 (for rich media) supplemented as shown in Table 2.3. Selection of strains containing the *KANMX4* marker gene was carried out using solid YEPD medium with 200 μg/ml geneticin (G418). Selection against cells

containing the *URA3* gene was achieved by adding 0.1% 5-FOA to solid synthetic defined medium supplemented with all amino acids and uracil.

For long-term storage, cells were stored at -80°C in 15% (v/v) glycerol.

2.2.3 Cells grown under aerated conditions

Cells used to measure the effect of different carbon sources were grown in 100 ml of medium in 250 ml flasks at 30°C with shaking at 220 opm, until at least 5 doublings were obtained and an OD_{600} of 0.5 ±0.05 was achieved. Cells used for analysis of expression at different phases of growth were grown in 1 l of YEPD in 2.5 l flasks, also at 30°C with shaking. In both cases the media was originally inoculated with stationary-phase cells resulting from overnight growth in YEPD.

2.2.4 Anaerobic growth conditions

Anaerobic growth of cells was performed by Anthony Beckhouse (UNSW) using methods modified from that of Skoneczny and Rytka (2000), Panozzo *et al.,* (2002) and Passoth *et al.,* (2003). Liquid SD medium containing amino acids outlined in Table 2.3 was produced and supplemented with the redox indicator dye resazurin (2 mg/l). Media was then dispensed into 100 ml penicillin bottles and purged with high purity nitrogen gas for 20 min. Bottles were plugged with butyl-rubber stoppers, sealed with aluminium cap seals and autoclaved as outlined in section 2.2.2. Following this, media was stored at room temperature in dark conditions to minimise any photo-oxidisation and degradation of supplements.

At initiation of experiments, further supplementation was required for anaerobic growth. A fresh stock solution of ergosterol and Tween 80 was made (ergosterol 4 mg/ml; Tween 80, 40% (v/v) in ethanol) and 500 µl of the stock was added to each 100 ml bottle via a syringe through the rubber stopper. The reducing agent, sodium dithionite was added to a final concentration of 10 mg/l to remove any traces of oxygen.

For experiments where anaerobic cultures needed to be opened and closed in the absence of oxygen, a Modular Atmosphere Controlled System (DW Scientific, UK) anaerobic chamber was used. Anaerobic incubation of solid media was also performed in the anaerobic chamber at 30°C.

Table 2.3: Supplements added to SD medium

Supplement	Final concentration (mg/l)
adenine	10
arginine	50
asparagine	80
histidine	20
isoleucine	50
leucine	100
lysine	500
methionine	20
phenylalanine	50
threonine	100
tryptophan	50
tyrosine	50
uracil	20
valine	140

2.2.5 Construction of *lsm7* deletion mutants

Creation of the *intronless* and *intron only* mutants was achieved via a two-step process. Firstly the *lsm7*Δu mutant was constructed, in which the entire *LSM7* open reading frame is replaced by the *URA3* gene. An oligonucleotide cassette was created by PCR using the oligonucleotide primers *LSM7uraF* and *LSM7uraR* that allowed amplification of the entire *URA3* open reading frame plus 800 bp of promoter sequence, flanked at the 5'end by 45 bp of sequence immediately upstream of *LSM7* and at the 3' end by another 45 bp of 3' UTR sequence immediately downstream of the *LSM7* gene. This cassette was transformed into BY4742 by the lithium acetate method of Gietz and Schiestl (1995) and transformants screened on synthetic defined medium lacking uracil. Positive clones, where the DNA cassette has been integrated into the genome in place of the native *LSM7* gene, were confirmed by PCR of *LSM7* using *LSM7screen*F and *LSM7 screenR* primers and sequencing.

The intronless cassette, containing the entire *LSM7* coding sequence without the intron, flanked by 32 bp of sequence immediately upstream of *LSM7* and 1.3 kb of sequence immediately downstream of the gene, was created by two sequential PCRs. The intron-only cassette was also made by two sequential PCRs, with 38 bp of sequence immediately upstream of *LSM7*, followed by the start codon, the entire *LSM7* intron, the stop codon and finally 1.3 kb of sequence downstream of *LSM7.* These constructs were concentrated

by ethanol precipitation, gel purified and sequenced for verification. Each amplicon was transformed into an *lsm7*Δu mutant. Resulting transformants were screened on solid synthetic defined medium containing all amino acids, uracil and 5-FOA as described above and positive clones confirmed by PCR and sequencing.

2.3 General DNA protocols

2.3.1 General methods and reagents

All routine procedures including agarose gel electrophoresis, DNA and RNA quantification, ethanol precipitation and phenol-chloroform-isoamyl alcohol DNA purification were performed as described by Sambrook *et al.* (1989). DNA fragments were recovered from agarose gels using the Qiagen Gel extraction kit, according to the manufacturers' instructions. DNA size standards were from MBI Fermentas.

2.3.2 Genomic DNA isolation

Yeast genomic DNA was obtained from 8 ml of yeast cultures grown overnight to stationary phase in YEPD. DNA was isolated according to the method of Hoffman and Winston (1987). Typically this yielded around 20-30 μg genomic DNA in 50 μl volume.

Primers used for PCR reactions in this study are shown in Table 2.4. All primers were synthesised by Sigma-Aldrich.

2.3.4 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used for DNA amplification. PCR used for colony screening typically used 100 ng-1 μg - template DNA, 100 μM of each dNTP (Roche Diagnostics Australia), 0.25 μM $MgCl_2$, 0.2 μM of each primer, 1x PCR reaction buffer (Applied Biosystems, NSW) and 1 unit *Taq* DNA polymerase in a 50 μl reaction. Typical cycling conditions were: 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 1 min, annealing for 30 s (at 48-58 °C depending on the melting temperature of the oligo), extension at 72°C for 1 min per 1 kb of product and finally 10 min extension at 72°C.

Table 2.4: PCR primers

Primer Name	Sequence (5'-3')	Notes
LSM7ura F	GAGAGCAGCACTTGTTTACTACACAGAA CATTAACCAAAAAAACCAATAAAATCGA ATTCCAAC	Used to create knockout cassette for generation of *lsm7::URA3* mutant
LSM7ura R	TTCAACTGTAAGGAAGGGAGTTTATATG AGATTATATTATTAAACTACTGTTACTTG GTTCTGGCC	
LSM7delta F	TGTTTACTACACAGAACATTAACCAAAA AAACATGCATCAGCAACACTCCCAAAG GAAAAAATTCGAAGGCCCTAAAAGAGAA GCTATTCTGGATTTAGCG	Used to create knockout cassette for generation of an *lsm7* mutant with the intron region deleted
LSM7delta R	GATGAGAGAATTGGAATAAATTAGTTTG CCAGCAACACGTAATAAATGAAAGGGTA GATATTGATTTGACC	
LSM7 intron A	CTTTGTTTACTACACAGAACATTAACCA AAAAAACATGGTATGTTTCACTTCTTATT TTCTTCCGTGGCAATAACCTTCC	Used to create the first half of a knockout cassette for generation of an *lsm7* mutant with all coding regions deleted
LSM7 intron B	ATGAGATTATATTATTAAACCTACTGTTG TGGTTTGTTCTCTGATTTCTTAGTTCAAA AAGG	
LSM7 intron C	CAGAGAACAAACCACAACAGTAGGTTTA ATAATATAATCTCATATAACTCCCTTCCT ACAGTTGAAAACAAAATCAAATCACGGT ATGG	Used to create the second half of a knockout cassette for generation of an *lsm7* mutant with all coding regions deleted. This was then combined with the half above and amplified by PCR with A and D primers
LSM7 intron D	TCGAATGTAAGGGTGGTTCATCTTTCAT TTCCTGCGGAGGGAAAGGCG	
LSM7screen F	GACGTCTTCCACCTCGGCGGA	Used to screen potential positive clones
LSM7 screen R	GGGAGTTTATATGAGATTATATT	
*LSM7*seq F	CGCATGACGTGTATACTA	Primer in *LSM7* promoter region for sequencing possible positive mutants

For PCR used to create DNA cassettes for use in generating the *lsm7* deletant mutants, a 1:5 Platinum *Pfx* (Invitrogen) to *Taq* DNA polymerase enzyme mix was used in place of the *Taq* DNA polymerase alone. This allowed adequate proofreading to reduce the chance of errors in the sequence of the cassette. In these reactions $MgCl_2$ was replaced by 1 mM $MgSO_4$ and the extension temperature was changed to 68 °C.

2.3.5 DNA sequencing

DNA was sequenced using ABI Prism™ BigDye Terminator System (Applied Biosystems, NSW), according to the manufacturer's instructions. Samples were analysed on an Applied Biosystems Model 373 Automated DNA sequencing system or an ABI 3730 capillary sequencer at The Ramaciotti Centre for Gene Function Analysis (UNSW).

2.4 RNA-based protocols

2.4.1 RNA extraction and quantification

Samples were harvested into pre-chilled (-80ºC) 50 ml tubes containing 10 g ice. Following harvest, cells were pelleted by centrifugation at 4000 x *g* for 2 min, resuspended in 1 ml TRIzol® reagent (Life Technologies Inc.) and RNA was extracted as described by the manufacturer. RNA purity was checked by agarose gel electrophoresis with ethidium bromide staining, as well as by determination of the optical density at 260 nm and 280 nm. For long-term storage, RNA was ethanol precipitated and stored at -80ºC.

2.4.2 Treatment of RNA

Prior to real-time PCR or microarray analysis, RNA was treated to remove any residual DNA. For RNA to be used in real time PCR analysis, this was achieved using DNA-*free*™ reagents from Ambion Inc, following manufacturers' instructions. The absence of contaminating DNA was confirmed by comparison of replicate PCRs performed either with or without reverse transcriptase.

RNA to be used in microarray analysis was further purified following extraction using the Qiagen RNeasy™ kit, including DNase 1 treatment.

2.4.3 Primers for real time PCR analysis

Primers used in real time PCR analysis were designed using Primer Express version 1.5 (Applied Biosystems). As above, all primers were synthesised by Sigma-Aldrich.

Primers were designed to follow the following guidelines where possible:

- GC content of each amplicon was within 30-80% range.

- Primers forming primer-dimers with themselves or each other were avoided where possible.
- Runs of identical nucleotides, especially guanine, were avoided.
- Melting temperatures of primers were within the range 58-60ºC.
- Amplicons between 95 -110 bases in length.
- Primers contained no more than 2 cytosine and/or guanine bases within the last (3') five nucleotides.

Primers used for real time PCR analysis in this study are shown in Table 2.5. Primer concentrations were optimised to minimise primer-dimer formation and all measurements carried out above the melting temperature for the primer-dimer formation, but below that for the target amplicon.

Standard curves were generated for each primer pair using control RNA in two-fold serial dilutions over three orders of magnitude, and found to be linear over the sensitivity range required (R^2=0.99). Intra and inter-assay variations in results were found to be less than 4% in each case. The presence of only one product was confirmed using dissociation curve analysis or by gel electrophoresis. To correct for the varying amplification efficiency of different primer pairs, each result was transformed using the slope and y-intercept obtained from the standard curve for relevant primer pair, as described in ABI Prism 7700 User Bulletin 2 (1997). To confirm the product results from the target DNA and not some other source, 'no template' controls were run on each plate for each primer pair present.

2.4.4 Analysis of relative mRNA abundance

For each 25 μl real time PCR reaction, 5 μg RNA was added to 15 μl of a master mix containing: 0.5x 2xSYBR® Green master mix, 10 U RNase inhibitor, 6 U MultiScribe™ reverse transcriptase and 1.875 μl DEPC-treated water. All reagents except RNA and water were obtained from Applied Biosystems (NSW). Finally the specific primers were added, at the concentrations specified in Table 2.5. If multiple samples were run simultaneously, a master mix was made up for all samples and then 15 μl was aliquotted into each sample well as described above. The standard real time PCR reaction involved the following steps: 2 min at 50°C, 10 min at 95°C, then

40 repeated cycles of 15 sec at 95°C, 15 sec at cut-off temperature (specific for those particular primers and listed in Table2.5) then 1 min at 60°C.

Data were obtained from duplicate PCR on triplicate cultures and normalised relative to either the level of *SIR3* transcript present, or the total amount of RNA in each sample, as determined prior to the real time PCR by spectrophotometric analysis. In each case, cultures were grown and harvested at the same time, to minimise error that might arise due to variations in experimental conditions on different days. Following PCR analysis, the presence of a single PCR product resulting from the target DNA was confirmed by melt-curve analysis using Dissociation Curve Software (Applied Biosystems).

2.4.5 *In vivo* splicing capacity analysis

To measure any changes in the capacity of the native cell to splice under the different conditions examined, a novel *in vivo* splicing assay was developed. A real time PCR assay was employed as described in section 2.4.4, using primers specific for the 5' region of the gene and either an area within the intron, or a primer spanning the exon junction (Table 2.5). Splicing capacity was measured as a ratio of unspliced to total mRNA levels for that gene and this result confirmed using the ratio of spliced to total mRNA in each case. As for all real time PCRs, data presented were obtained from at least duplicate PCR on triplicate cultures. Intra-and inter-assay variations in results were found to be less than 4.5% in each case.

2.4.6 Assay for mRNA half-life determination

Yeast cultures were grown to mid-exponential phase in rich media containing D-glucose or sodium lactate as the carbon source, as described above. Upon reaching this density, 1, 10 phenanthroline (Sigma) was added to each culture to a final concentration of 100 µg/ml and immediately 50 ml of each culture was harvested by centrifugation at 3,350 x g for 2 minutes into pre-chilled 50 ml tubes containing 10 g ice. These samples were designated as time 0. Further 50 ml samples were taken at intervals of 5, 10 20 and 60 mins following addition of the transcriptional inhibitor. Relative RNA abundance was then determined for each sample as described above.

Table 2.5: Primers and conditions used for real time PCR

Primer Name	Temperature for measurement (°C)	Sequence (5'-3')	Concentration of primers (nM)	Notes
LSM1 F	76.9	AGGATGCGAAGCGACAACA	200	Used to examine *LSM1*
LSM1 R	76.9	TTACGGTCTACTGAGCTTACAATA GCA	200	
LSM2 F	71.9	AGGTACGTTTACTTGAATAAGAAC ATGGT	250	Used to examine *LSM2*
LSM2R	71.9	TGCATAATTCTTATTTTCTTTCAGT CATT	250	
LSM3F	78.3	CATCGTGCTGAGTGATGCAGTA	250	Used to examine *LSM3*
LSM3R	78.3	CACTGCGCCATCGTCATCT	250	
LSM4F	73.4	CCATACTCTCAAAACCGTCAATAC A	200	Used to examine *LSM4*
LSM4R	73.4	GACCTTTTGAGAAGAGCTGTT	200	
LSM5F	77.7	AAGACGCTGTCGAGTGGCTTAT	100	Used to examine *LSM5*
LSM5R	77.7	CTGGCACAGGATGGCAATA	100	
LSM6F	74.8	TCAATATAGCACAAGACGCCATAT	150	Used to examine *LSM6*
LSM6R	74.8	CTCCGTAGTAACGCTACCCTCTGT	150	
LSM7F	71.1	GATGATACAGTAGAATATATGTCT AATCCTGATG	250	Used to examine *LSM7*
LSM7R	71.1	CAAAATAGTACCTCTTATGACGGT CAAAC	250	
LSM8F	74.5	GCATTCGCCGTCAACTTTG	150	Used to examine *LSM8*
LSM8R	74.5	ACACCATACAAGAAGAACACACTA TCGT	150	
SMX2F	77.0	AGGAATTTTGCGAGGCTACGA	300	Used to examine *SMX2*
SMX2R	77.0	GCATCTAGAGCCTCTAGGGATATT ATG	300	

Primer Name	Temperature for measurement (°C)	Sequence (5'-3')	Concentration of primers (nM)	Notes
SMX3F	78.4	GTTGAACAGCACCGAATATAGAGG TA	300	Used to examine *SMX3*
SMX3R	78.4	CCTGATGTACAGCACGTTATTCG	300	
U6snRNA F	73.0	CGAAGTAACCCTTGTGGACAT	300	Used to examine U6snRNA
U6 snRNA R	73.0	CTCTTTGTAAAACGGTTCATCCTTA TG	300	
RPL25un F	-*	TGCTCCACGTCGAGGTATCA	100	Used to examine unspliced *RPL25* for splicing assay
RPL25un R	-*	TCTTAGCGGCAGAGCCTTAGCT	100	
RPL25t F	73.4	CCACATTACAACAGATTGGACTCA T	100	Used to examine total *RPL25* for splicing assay
RPL25t R	73.4		100	
RPL25spl F	77.0	GGCTCCATCTGCTAAGGCTACTG	200	Used to examine spliced *RPL25* for splicing assay
RPL25spl R	77.0	TGGTAGTCTGAAGGTAGCAGAAGT TCT	200	
LSM2unF	72.0	AAACTAACGGGTATCGTACATCAA TTT	300	Used to examine unspliced *LSM2* for splicing assay
LSM2un R	72.0	TGTACCTTTTATTTCAATGTCGTTT TTTA	300	
LSM2tF	-*	AATCTGAAACTAGACAACATATCA TGCA	300	Used to examine total *LSM2* for splicing assay
LSM2tR	-*	AACGTACCTGACTGTTGAACCTCT TA	300	

Primer Name	Temperature for measurement (°C)	Sequence (5'-3')	Concentration of primers (nM)	Notes
LSM2spl F	-*	GAAGTGGCGTAGAGTTAAAAAACG A	300	Used to examine spliced *LSM2* for splicing assay
LSM2spl R	-*	ATATGTTGTCAGTTTCAGATTCAAA AATTG	300	
LSM7unF	-*	CAGAGAACAAACCACAACAGCAA	300	Used to examine unspliced *LSM7* for splicing assay
LSM7un R	-*	ATTTGACGCGAATTTTAGAATCTTT ATACT	300	
LSM7tF	71.5	AGTATAAAGATTCTAAAATTCGCG TCAAAT	300	Used to examine total *LSM7* for splicing assay
LSM7tR	71.5	TGTATCATCAAGTACCAAGTTCAT CAGTT	300	
LSM7spl F	-*	CAACACTCCGATTTAGCGAAGTAT AA	300	Used to examine spliced *LSM7* for splicing assay
LSM7spl R	-*	TGATCATAGCCTTTTAGGACACCT ATAAC	300	
SIR3F	-*	TTGGACGGTTGGCAAGTTATC	75	Used to examine *SIR3*, as external PCR control
SIR3R	74.1	AAATACGTTTTCTCCTCCTCTTTTT CT	75	

Primers with a -* do not form primer-dimers under any conditions tested. Therefore these samples are measured at 60 °C.

2.5 Microarray analysis

2.5.1 Microarray analysis

Microarray analysis of RNA from cells grown on different carbon sources (glucose or lactate) was performed by Cristy Gelling (UNSW). Microarray analysis of cells grown under aerobic or anaerobic conditions was carried out by

Anthony Beckhouse (UNSW). In both cases the *S. cerevisiae* microarrays employed were obtained from the Ramaciotti Centre for Gene Function Analysis (UNSW) and were printed with 40-mer oligonucleotide probes for 6,250 yeast open reading frames (MWG Biotech). Slides were processed according to the manufacturers' instructions. Synthesis and labelling of cDNA were carried out according to a modified amino-allyl dye-coupling protocol (Hughes *et al.,* 2001). Briefly, cDNA was synthesised from 20 μg total RNA using Superscript II reverse transcriptase (Invitrogen) and a 2:1 ratio of 5-(3-aminoallyl)-uridine 5'triphosphate (Sigma) to dTTP.

The RNA was then hydrolysed using NaOH and the cDNA purified with QIA quick columns (Qiagen). Samples were labelled with N-hydroxy succinamide esters of Cy3 or Cy5, respectively (Amersham Biosciences). Hybridisation was carried out for 14-16 hours at 37°C in DIG Easy Hyb (Roche Diagnostics) containing 0.5 mg/ml *Escherichia coli* tRNA and 0.5 mg/ml herring sperm DNA. Slides were then washed 3 times in 1x SSC with 0.1% sodium dodecylsulphate at 50°C, rinsed in 1x SSC with 0.15 M NaCl and 0.015 M trisodium citrate, and dried by centrifugation and scanned using an Applied Precision ArrayWoRx E Biochip Reader.

To obtain RNA from cells growing in anaerobic or aerobic conditions, multiple anaerobic cultures (100 ml) were grown to an OD_{600} 1.0 and split into two aliquots. One aliquot was transferred to a pre-warmed flask and incubated in the presence of oxygen for 5, 10, 20, 30, 60 or 120 min. The other aliquot was harvested immediately as a zero time-point by centrifugation (3350 × g, 4°C, 1 min). The supernatant was removed and cells snap frozen in liquid nitrogen before storage at -80°C. Aerated samples were harvested in the same manner upon completion of the time-course.

For both carbon source and aerobic/anaerobic microarrays, each condition was represented by duplicate slides, labelled in a reciprocal fashion relative to the dyes used (dye swap). For the microarray analysis of cells growing in aerobic or anaerobic conditions, RNA for the zero (anaerobic) time point was pooled from all samples.

2.5.2 Splicing-sensitive microarray analysis

Splicing sensitive microarrays were performed by the Ares laboratory as described in the supplementary material of Clark *et al.* (2002). They were designed using oligonucleotides of 35 nucleotides (nt) in length which symmetrically spanned splice junctions and exon-intron junctions in the mRNA These oligonucleotides therefore allowed detection of spliced and unspliced mRNA, respectively. Each oligonucleotide was present in four copies on each array.

2.6 Microscopy

2.6.1 DAPI staining

Cells were fixed in ice-cold 70% (v/v) ethanol, washed with Tris-HCl (pH 8.0), stained with 4', 6-diamidino-2-phenylindole (DAPI) at a concentration of 0.01 mg/ml and observed under a phase-contrast Olympus BX-60 microscope. The cells were classified into three groups; unbudded cells; budded, neck (where all nuclear DNA is in the neck between the mother and daughter cell); budded, separate (where the nuclear DNA has been segregated into the mother and daughter cell. For each cell type examined, a minimum of 400 cells were counted. Standard deviations for all data were calculated based on a binomial distribution, as a percentage of the total number of cells counted.

2.6.2 Actin staining

For every 4 ml of culture to be stained, 1 ml of EGTA (pH 6.9) and 0.1 ml each of PMSF (17.5 mg/ml in methanol), leupeptin, pepstatin and aprotinin (1 mg/ml respectively) were added. Cells were then incubated for 5 min at room temperature. To this mixture 5 ml of PEM buffer (0.1 M PIPES-KOH pH 6.9, 5 mM EGTA, 5 mM $MgCl_2$) with 8% (w/v) formaldehyde and a 1:1000 dilution of the protease inhibitors listed above. Cells were fixed for 10 min and washed three times with PEM with inhibitors (again at 1: 1000 dilutions). Cells were permeabilised by the addition of 0.1% (v/v) TritonX-100 for 1 min, washed three times as previously, and stained with TRITC-phalloidin (4 μg/ml; Sigma) for 30 min in the dark, followed by at least another three washes. Samples were

observed under a phase-contrast Olympus BX-60 microscope, as for section 2.6.1. Cells were classified according to actin localisation patterns and cell division status into one of four groups: unbudded cell with delocalised actin; budded cells with delocalised actin; unbudded cell with actin at the bud site; and budded cells containing actin in the bud. As for section 2.6.1, for each sample a minimum of 400 cells were examined and standard deviations are shown as a percentage of cells counted.

2.7 Computational analysis

2.7.1 Microarray Data Analysis

Image analysis for each microarray slide was performed using GenePix Pro 3.0 (Axon Instruments). For the carbon source microarray the fluorescence intensities were normalised in GeneSpring 5.0 (Silicon Genetics) relative to the values for *ACT1* and *SIR3.* Along with the aerobic/ anaerobic microarray, data from the carbon source microarray were also normalised by the LOWESS normalisation method and dye-swap experiments were colour-reversed. Normalised data from the splicing-sensitive microarray was also analysed using Genespring 5.0. The splicing-sensitive microarray was normalised using "Norm" a custom written software application. Data for the splicing-sensitive microarrays were further analysed with a standard correlation (with a separation ratio of 0.5, minimum distance of 0.0001 to maximise the number of clusters and minimise false positives), using Genespring 5.0 software.

2.7.2 Other computational analyses

The programmes and uses for all other programs employed in this study that have not been previously mentioned are listed in Table 2.6.

Table 2.6: Programs used throughout this study

Program Name	*Use*	*Source*
ABI PRISM® 7700 Sequence Detection System	real time PCR detection and some analysis	Applied Biosystems (NSW)
Excel v.10.0	all statistical analysis – standard deviations, t-test significance, generation of graphs, processing of real time PCR data	Microsoft corporation (USA)

Program Name	Use	Source
RSA tools	identifying and displaying putative regulatory sites in promoter regions	Helden *et al.*, 2000
Motifsampler	identification of putative transcription factor binding motifs	Thijs *et al.*, 2002
Weblogo	display of putative transcription factor motif across various promoters	Crooks *et al.*, 2004
UCSC genome browser database	display of regions of conservation between the genomes of *S. cerevisiae, S.paradoxus, S. mikitae, S.kudriavzevii* and *S. bayanus*	Karolchik *et al.*, 2003
WuBlast2	sequences for *sensu stricto* and *sensu lato Saccharomyces* species	Gish *et al.*, 1996-2004
Clustal W	multiple sequence alignment	Thompson *et al.*, 1994
GeneDB	*LSM7* sequence for *S. pombe*	Hertz-Fowler *et al.*, 2004
Ashbya Genome Database	*LSM7* sequence for *A. gossypii*	Dietrich *et al.*,2004
Génolevures	all *LSM7* sequences for species not mentioned above	Dujon *et al.*, 2004
RNAfold	unspliced *LSM7* mRNA secondary structure prediction	Hofacker *et al.*, 2003
Molecular Toolkit	generation of hydrophobicity plot for putative protein-coding sequence within *LSM7* intron	University of Colorado (http://arbl.cvmbs.colostate.edu /molkit/hydropathy)
Aable	generation of 3-D bubble graph for splicing-sensitive microarray data	Gigawiz Ltd
GO ontologies	Classification of genes into functional categories	www.geneontology.org
Genespring GX	Microarray data analysis and presentation	Agilent technologies
Saccharomyces Genome Database	*S. cerevisiae* gene sequenes, function and transcript abundance	www.yeastgenome.org

Chapter 3: Co-regulation of *LSM* genes encoding spliceosomal proteins

3.1 Introduction

To survive and be competitive in their natural environment, single celled organisms such as *S. cerevisiae* must be able to adapt quickly and efficiently to a variety of environmental cues. Such adaptations can involve altering the expression of thousands of genes, allowing the cell to achieve optimum growth under each specific condition (Gasch *et al.,* 2000). Genes encoding proteins in a shared metabolic pathway tend to be co-expressed (Eisen *et al.,* 1998; Grigoriev, 2001; Ge *et al.,* 2001). Components of macromolecular molecules are often required in stoichiometric amounts for assembly of functional complexes (Papp *et al.,* 2004). Further, over-production of some protein components may be deleterious to the cell (Rottensteiner *et al*, 1997; Wu *et al.*, 2001). Accordingly, genes coding for proteins which form large complexes are generally also co-expressed over a variety of conditions, indicating that they may be transcriptionally co-regulated (Simonis *et al,* 2004; Teichmann and Babu, 2002).

More than 300 distinct proteins are thought to make up the spliceosome (Hartmuth, *et al*, 2002; Rappsilber *et al.,* 2002; Zhou *et al.,* 2002). However while many studies have investigated the structure and function of proteins involved in the spliceosome, none so far have explored the regulation of the genes that encode them.

3.1.1 Aim

Electron microscopic analysis and pull-down assays indicate that the core of the U6 snRNP in the spliceosome contains equimolar amounts of each of the seven Lsm proteins (Bouveret *et al.,* 2000). Therefore the aim of this study was to determine if the genes that encode these proteins are co-ordinately regulated at the level of transcription, in response to changes in the metabolic state of the cell. Expression of both spliceosomal and non-

spliceosomal *LSM* genes were examined under a variety of environmental conditions, along with other components of the spliceosome, to determine the extent, purpose, and trigger for any regulation observed. The half-life of each mRNA transcript was also investigated to determine the level at which regulation occurred.

3.1.2 Acknowledgements

The determination of cellular growth rate following a switch to anaerobic or aerobic conditions was carried out by Anthony Beckhouse (UNSW). RNA samples from cells at 0 and 120 minutes after the switch to anaerobic or aerobic conditions were also kindly provided by A. Beckhouse.

3.2 Levels of most LSM transcripts vary co-ordinately across different phases of growth

When inoculated into a medium rich in glucose, yeast transition from a stationary, quiescent state to one of exponential growth. This transition is called the 'lag' phase (Figure 3.1). During exponential growth ('log' phase) the yeast utilise their preferred carbon source, glucose, repressing all genes involved in the metabolism of other carbon sources. Growth during this phase is fuelled by fermentative metabolism, resulting in the production of ethanol. When glucose levels become limiting the yeast cells switch to respiratory metabolism, using ethanol as a source of carbon. The temporary decrease in growth rate as the organism shifts from fermentative growth to respiratory growth is called the 'diauxic shift' and involves significant changes in the expression of over 1700 genes (DeRisi *et al.,* 1997). Accordingly, *LSM* transcript levels were measured at different phases of growth in a batch culture to observe if any change in expression could be detected.

An analysis of growth kinetics was performed to determine the optical density of the strain, BY4742, at the different phases of growth. Duplicate cultures were inoculated with stationary-phase cells. Samples were then taken from each culture at intervals over a total of 42 hours and the optical density (OD) of the culture measured using a spectrophotometer.

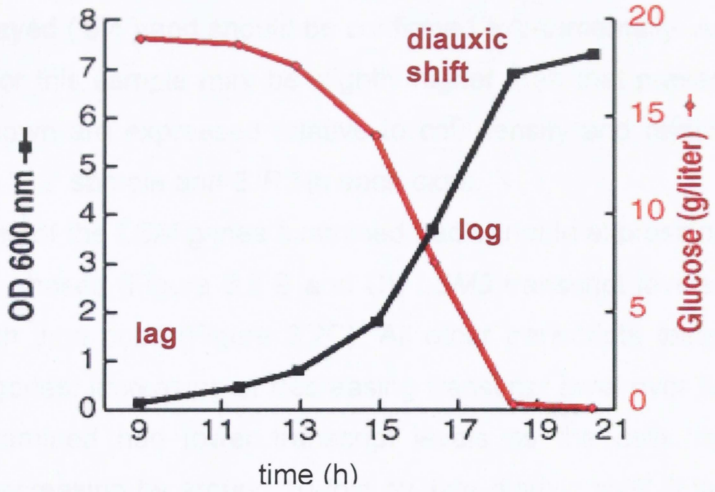


Figure 3.1 Changes in cell density and glucose concentration over time during cell growth in batch culture. Cell density was measured as the optical density of the culture at 600nm (adapted from DeRisi *et al.,* 1997)

While levels of carbon sources in the media were not directly measured, the pause in growth observed in the culture as the cells shifted from 'log' phase growth into growth employing respiratory metabolism will be referred to here as the 'diauxic shift', as shown on Figure 3.1. Six optical densities were selected, corresponding to different phases of cell growth; OD at 600 nm of 0.2 ('early log'), 0.4 ('mid log'), 0.8 ('late log'), 2 ('early diauxic shift'), 3.2 ('mid diauxic shift') and 6.4 ('late diauxic shift/early stationary phase') (Figure 3.2A).

Samples of cells at each of these densities were then harvested from triplicate batch cultures, RNA extracted and levels of *LSM* mRNA determined. Most *LSM* transcripts were of such low abundance that Northern analysis was not sufficiently sensitive to quantify the expression of these genes. Real time PCR assays were therefore developed to allow analysis of the mRNA transcripts down to a level of less than one copy per cell (see Appendix). As the total amount of RNA content of cells decreases dramatically during the diauxic shift (Ju and Warner, 1994), transcript levels were normalised relative to total RNA (to take into account human error) and cell density. This was based on the assumption that the density of each BY4742 cell remains relatively constant across the different phases of growth. As stationary phase cells have a thicker cell wall (Werner-Washburne *et al.,* 1996) and possibly altered vacuolar content

compared to cells in log phase, this assumption may not be valid for the last data point displayed ('6.4') and should be confirmed experimentally. Accordingly the real result for this sample may be slightly higher than that presented here. Thus results shown are expressed relative to cell density and relative to data obtained for the '0.2' sample and *SIR3* in each case.

All but one of the *LSM* genes examined had variable expression over the different growth phases (Figure 3.2 B and C). *LSM3* transcript levels remained constant at each time point (Figure 3.2C). All other transcripts tested fell into two broad categories: increasing or decreasing transcript level over time. Most *LSM* genes examined had lower transcript levels as the cells moved into 'diauxic shift', decreasing by around 10-fold by 'late diauxic shift' (Figure 3.2B). Indeed, the pattern of expression of each of these *LSM* genes (*LSM2, 4, 5, 7* and *8*) was extremely similar, indicating that they may be subject to coordinate regulation. The remaining transcripts both showed increased transcript levels over time – *LSM6* increased gradually, to around 10-fold by the 'mid-diauxic shift' point (Figure 3.2C). Most dramatic was the response of *LSM1,* whose protein is located in the cytoplasm and not involved in splicing. Levels of *LSM1* mRNA increased almost 70-fold by the 'early diauxic shift' point (Figure 3.2C). In both cases the increase appeared to be transient, with the mRNA levels returning to around that of the 'early log' sample by the final time point. Thus the coordinate mRNA transcript pattern observed for the genes encoding Lsm2p, Lsm4p, Lsm5p and Lsm7p over the different phases of growth is more likely to be related to the activity of these proteins in the splicing complex (along with Lsm8p), instead of any of the complexes involving Lsm1p.

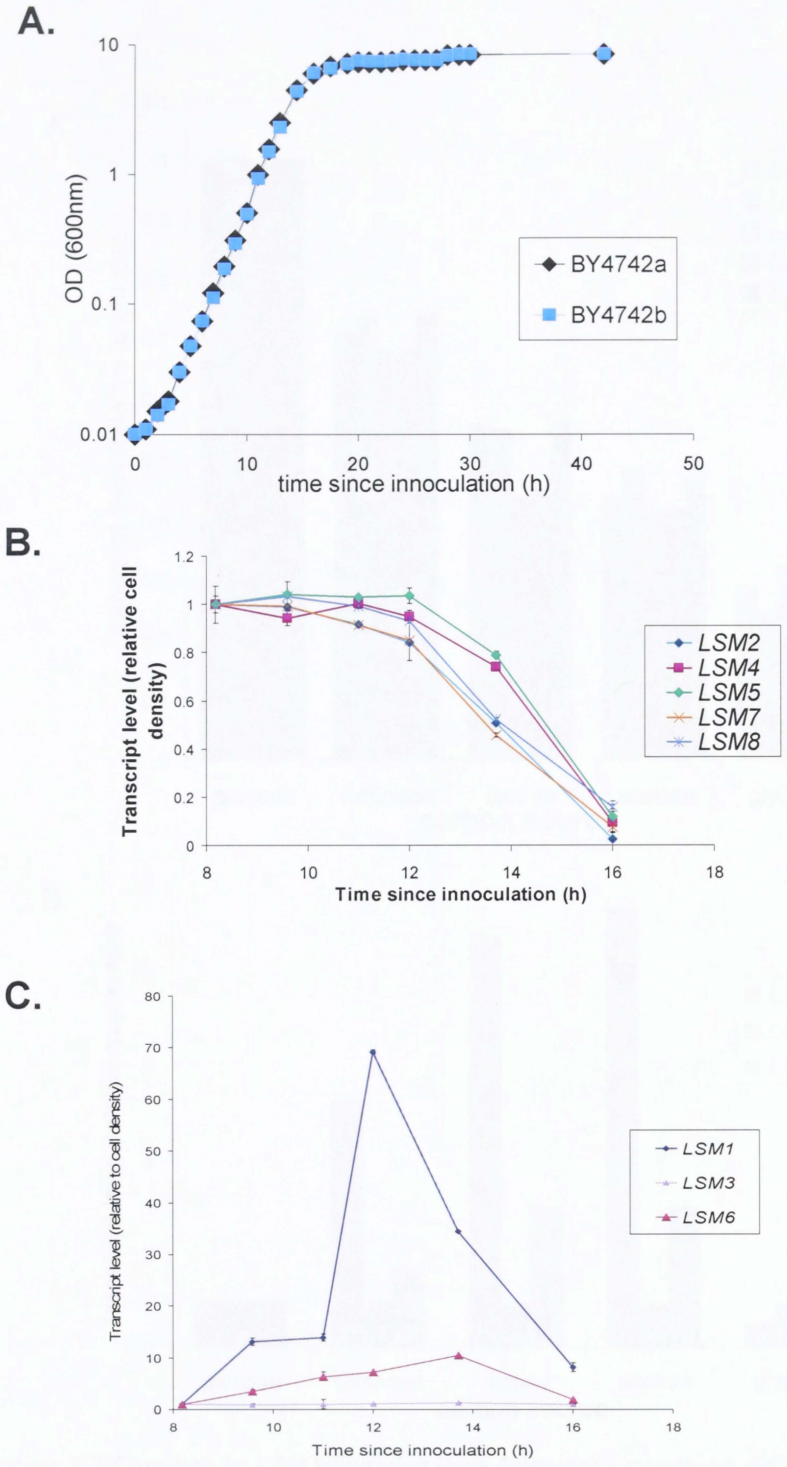


Figure 3.2 Changes in *LSM* transcript levels at different phases of growth

A. Wild-type cells were grown to stationary phase in YEPD. Equal numbers of cells were then transferred to fresh medium and optical density determined every 1-1.5 h for a further 30 h. To confirm stationary phase has been reached a final sample was taken 40 h after inoculation into fresh media. Cultures were performed in duplicate (BY4742a and BY4742b).

RNA was extracted from cells with OD_{600} of 0.2, 0.4, 0.8, 2, 3.2 and 6.4. Relative levels of *LSM* transcripts in these cells were determined, using *SIR3* as an external control. Data are also shown relative to the result obtained for the '0.2' sample in each case. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments. **B.** Transcripts from the *LSM* genes encoding spliceosomal proteins which exhibit a highly similar pattern of expression.

C. Transcripts from other *LSM* genes which showed different patterns of expression.

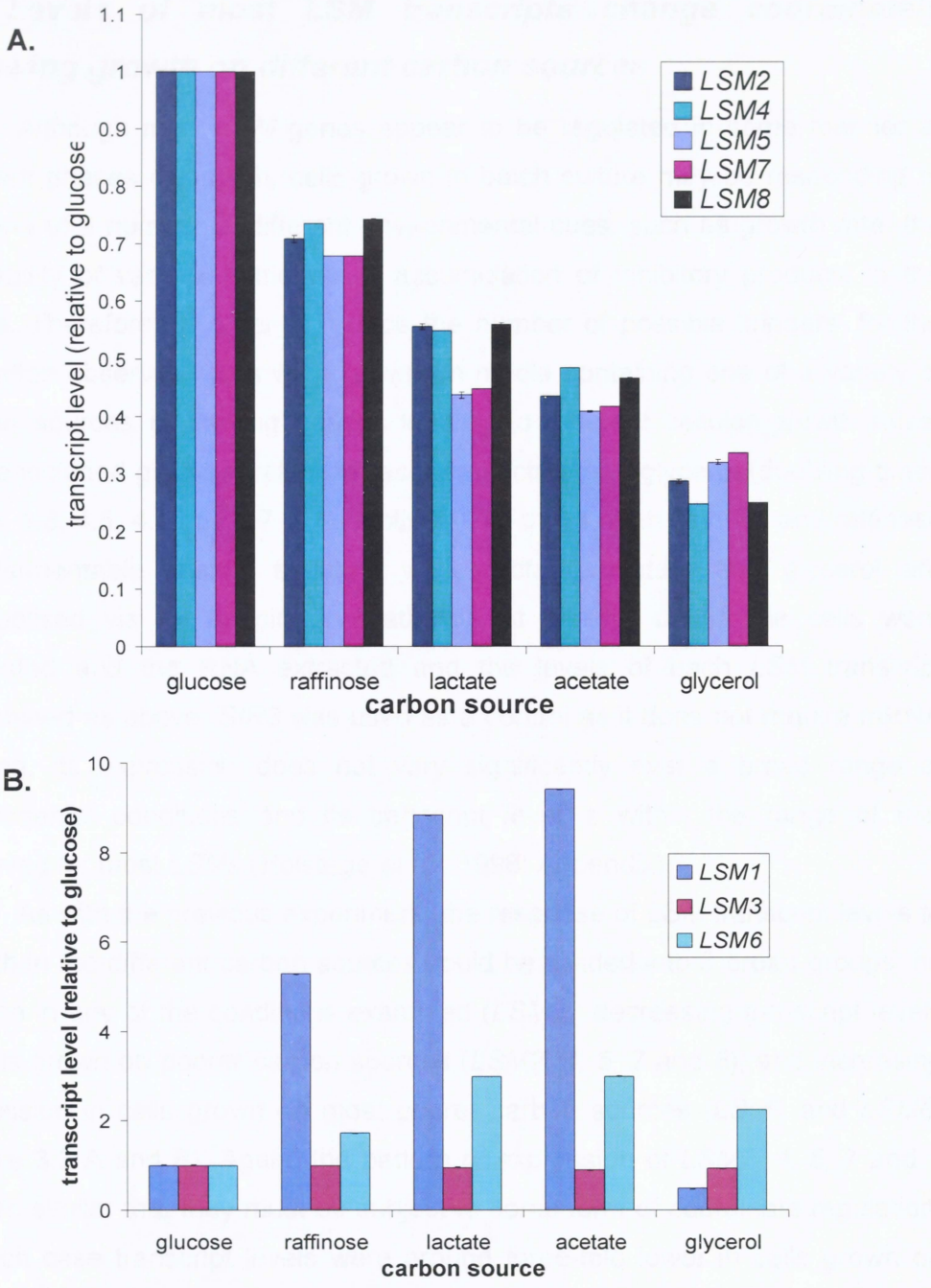


Figure 3.3 Changes in *LSM* transcript level following growth on different media. Wild-type cells were grown in rich media in the presence of the following different carbon sources; glucose, raffinose, lactate, acetate or glycerol. Relative levels of *LSM* transcripts in these cells were determined, using *SIR3* as an external control. Data are also presented relative to the result obtained for cells grown on glucose. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments. **A.** Transcripts from the *LSM* genes encoding spliceosomal proteins which exhibit a highly similar pattern of expression. **B.** Transcripts from other *LSM* genes which showed different patterns of expression. Raw data for these figures can be found in the Appendix.

3.3 Levels of most LSM transcripts change coordinately following growth on different carbon sources

Although most *LSM* genes appear to be regulated in some manner at different phases of growth, cells grown in batch culture may be responding to any one of a number of different environmental cues, such as growth rate, the availability of various nutrients, or accumulation of inhibitory products in the media. Therefore, in order to reduce the number of possible 'triggers' for the regulation observed, cells were grown on media containing one of a variety of carbon sources of varying quality, leading to different cellular growth rates. These included glucose, raffinose, acetate, lactate and glycerol (doubling times of 1.6, 1.8, 4.5, 4.7 and 5.7 h respectively). Of these, both glucose and raffinose are fermentable carbon sources, while lactate, acetate and glycerol are metabolised via the respiratory pathway. At mid-log phase the cells were harvested and the RNA extracted and the levels of each *LSM* transcript determined as above. *SIR3* was used as a control as it does not require mRNA splicing, its expression does not vary significantly over a broad range of experimental conditions and its transcript level is within the range of that observed for most *LSMs* (Holstege *et al.,* 1998; Appendix).

As with the previous experiment, the response of *LSM* transcript levels to growth in the different carbon sources could be divided into 3 broad groups: no change in any of the conditions examined (*LSM3*); decreasing transcript levels in cells grown on poorer carbon sources (*LSM2, 4, 5, 7* and *8*); and increasing expression in cells grown on most poorer carbon sources (*LSM1* and *LSM6*) (Figure 3.3 A and B). Again, the pattern of expression of *LSM2, 4, 5, 7* and *8* was so similar that they must be subject to some form of coordinate regulation. In each case transcript levels were around three-fold lower in cells grown on glycerol or lactate than in cells grown on glucose (Figure 3.3A). In contrast, levels of *LSM6* mRNA increased, also by around three-fold in cells grown on lactate or acetate compared to cells grown on glucose. *LSM1* changed more than any other gene examined – *LSM1* mRNA was more than nine times higher in cells grown on acetate than in cells grown on glucose (Figure 3.3B). Interestingly, both *LSM1* and *LSM6* exhibited decreased transcript levels in

glycerol, the poorest carbon source tested, with *LSM1* expression even lower in these cells than in cells grown on glucose (Figure 3.3B).

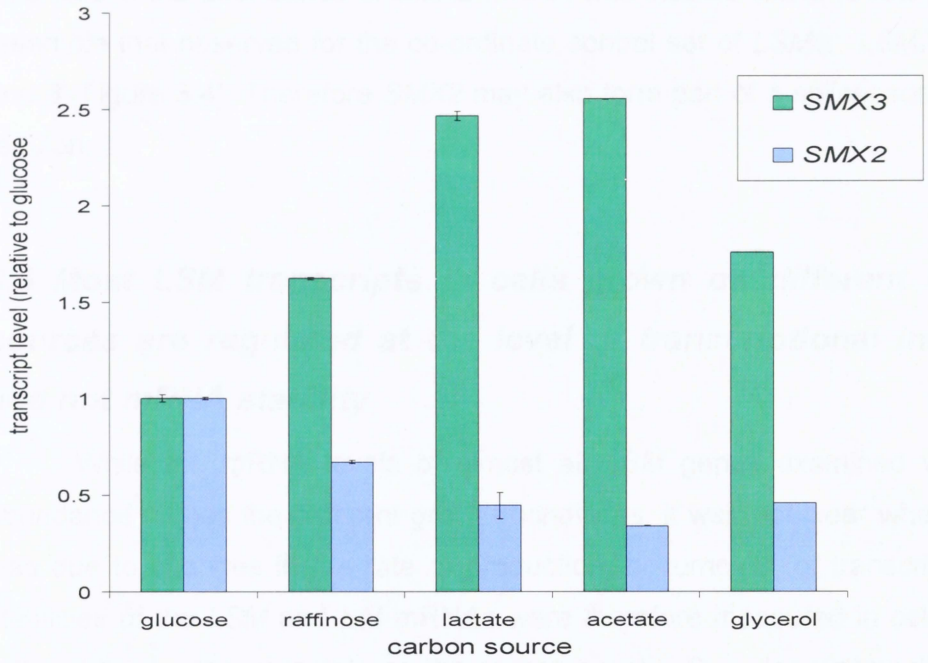


Figure 3.4 Changes in *SM* transcript levels following growth on different media. Wild-type cells were grown in rich media in the presence of a variety of different carbon sources; glucose, raffinose, lactate, acetate or glycerol. Relative levels of *SMX2* or *SMX3* transcripts in these cells were determined, using *SIR3* as an external control. Data are also presented relative to the result obtained for cells grown on glucose. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments.

3.4 Levels of other spliceosomal protein transcripts also change under these conditions

To ascertain if the discrepant pattern of transcript levels observed for *LSM6* correlated with other non-*LSM* genes involved in splicing, transcript levels for two spliceosomal *SM* genes, *SMX2* and *SMX3* were also determined in cells grown to mid-log phase on different carbon sources. As mentioned in Chapter 1, protein sequence comparisons indicate that *LSM6* and *SMX3* share a close

phylogenetic relationship, as do *LSM7* and *SMX2* (Salgado-Garrido *et al.*, 1999). Changes in the transcript levels of *SMX3* in the various media were in almost perfect concordance with that observed previously for *LSM6* (Figure 3.4). Thus the different levels of mRNA transcript for these two genes across the conditions tested may be a result of some shared regulatory mechanism. Changes in the abundance of *SMX2* mRNA was instead found to more closely resemble that observed for the co-ordinate control set of *LSMs, LSM2, 4, 5, 7* and *8* (Figure 3.4). Therefore *SMX2* may also form part of a spliceosomal gene regulon.

3.5 Most LSM transcripts in cells grown on different carbon sources are regulated at the level of transcriptional initiation and not mRNA stability

While the mRNA levels of almost all *LSM* genes examined varied in abundance across the different growth conditions, it was not clear whether this was due to changes in the rate of production, or turnover, of transcripts. The stabilities of the *LSM* and *SM* mRNAs were therefore measured in cells grown with either glucose or lactate as the carbon source. Samples were collected at various times following the addition of the transcriptional inhibitor, 1,10-phenanthroline (Brown and Sagliocco, 1996). Transcript levels were determined relative to the total amount of RNA present in each case. The half-lives of *RPL25* and *SIR3* transcripts were also examined in cells grown on glucose as controls and were found to be equivalent to those reported previously (Holstege *et al.*, 1998, Table 3.1). In all but one case, the *LSM* and *SM* transcripts examined showed no change in half-life in either growth condition (Table 3.1). Therefore the variation in abundance of these transcripts in cells grown in the different media is likely to be due to changes in initiation of transcription. The exception to this was *LSM7*, which was much more stable (with double the half life) in cells grown on glucose than in those grown on lactate (Table 3.1), indicating that expression of *LSM7* may be subject to additional regulation at the level of mRNA stability.

Table 3.1: Half-lives and Abundance of mRNA Transcripts in the Different Growth Media

mRNA transcript	Abundance (relative to *SIR3*)	Half-life in glucose (min)	Half-life in lactate (min)
LSM1	2.46 ± 0.03	17.2 ± 1.6	16.9 ± 0.8
LSM2	0.13 ± 0.02	16.0 ± 1.8	16.3 ± 1.5
LSM3	2.46 ± 0.10	17.9 ± 1.7	17.0 ± 1.1
LSM4	0.84 ± 0.03	14.2 ± 1.2	12.8 ± 0.6
LSM5	1.18 ± 0 .2	12.4 ± 0.5	11.8 ± 0.9
LSM6	0.18 ± 0.05	17.9 ± 1.4	18.4 ± 0.9
LSM7	0.01 ± 0.01	30.5 ± 1.3	14.7 ± 0.9
LSM8	4.26 ± 0.26	19.5 ± 0.6	20.2 ± 0.6
SMX2	0.38 ± 0.05	16.9 ± 1.2	15.4 ± 0.8
SMX3	6.29 ± 0.10	14.2 ± 0.6	13.4 ± 0.8
SIR3	1.00 ± 0.05	26.3 ± 0.9	24.8 ± 1.5
RPL25	18.99 ± 0.40	12.9 ± 1.6	18.6 ± 1.8

Abundance data indicates the relative levels of each *LSM* and *SM* transcript in wild-type cells grown on rich media containing glucose. Levels of *SIR3* and *RPL25* transcripts were also determined and were comparable to that found by other investigators using different methods. Half-life data was obtained from wild-type cells grown to exponential phase (OD600=0.5) on rich media with either glucose or lactate as the carbon source. Transcription was then inhibited through the addition of 1, 10-phenanthroline and RNA was obtained at various intervals thereafter. Levels of each gene examined were determined and the half-life calculated as indicated in the methods. Errors shown correspond to 2 standard deviations in each case.

3.6 LSM2 *and* LSM7 *transcripts are the least abundant and most likely to be limiting factors in Lsm protein complex formation*

Relative abundance was determined for each transcript in cells grown on glucose using real time PCR. *RPL25* and *SIR3* abundances were measured as controls representing medium-to-highly expressed genes or genes with low expression, respectively. Results for these two genes were similar to that reported by Holstege *et al.* (1998).

Significantly, *LSM7* was the least abundant transcript under these conditions, although levels of *LSM2, 4* and *5* and *SMX2* were also relatively rare (Table 3.1). Previous data above indicated that levels of these transcripts decreased in poorer carbon sources (Figure 3.2B). Taking these data together it

could be expected that regulation of *LSM2, 4, 5, SMX2* and particularly *LSM7* in this manner would lead to very low transcript abundance in cells grown on carbon sources of lower quality, which may therefore limit the rate of formation of the spliceosomal Lsm protein complex.

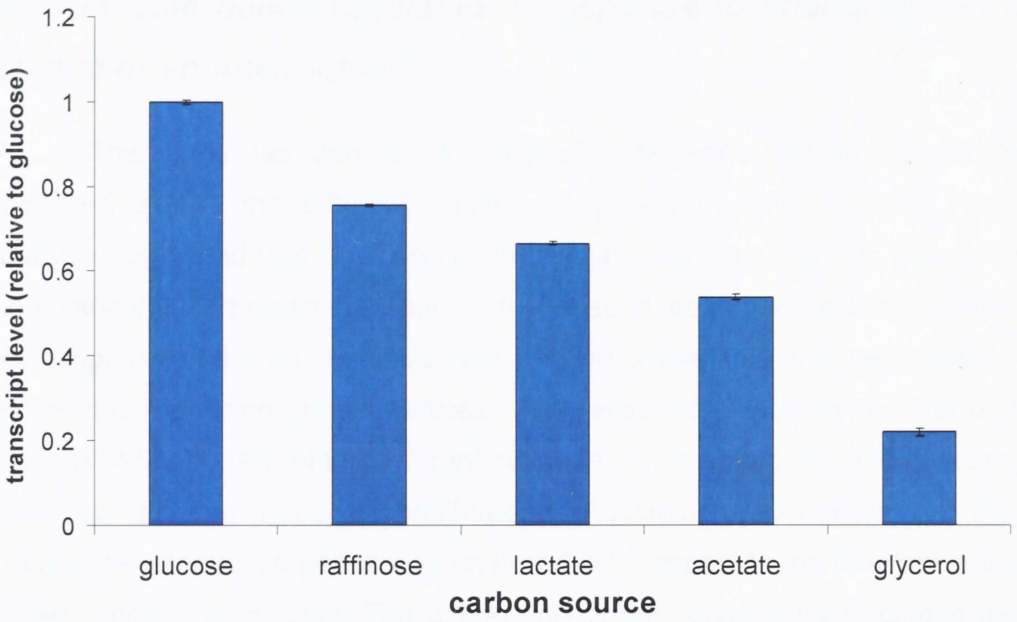


Figure 3.5 Changes in U6snRNA abundance following growth on different media. Wild-type cells were grown in rich media in the presence of a variety of different carbon sources; glucose, raffinose, lactate, acetate or glycerol. Relative levels of U6snRNA in these cells were determined, using *SIR3* as an external control. Data are also presented relative to the result obtained for cells grown on glucose. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments.

3.7 Levels of U6 snRNA correlated with the expression of co-ordinately regulated LSM genes

During splicing the Lsm protein complex (Lsm2p- Lsm8p) binds to a region at the 3' end of U6 snRNA, which stabilises the snRNA (Mayes *et al.*, 1999). Accordingly levels of U6 snRNA were also examined in cells grown on the various carbon sources, also using real time PCR. Levels of U6 snRNA were lower in cells grown on poorer carbon sources than in those grown on glucose (Figure 3.5). The pattern of change in U6 snRNA abundance over the various conditions was concordant with that found previously for the coordinately controlled set of spliceosomal *LSM* genes (Figure 3.3A and Figure

3.5). Further, the change in U6snRNA levels across the media tested is also in the range of the change in gene expression exhibited by the coordinate control gene set under these conditions (Figure 3.5).

3.8 Are LSM *genes regulated in response to changing carbon source or another signal?*

This study has identified a group of *LSM* genes that are coordinately regulated across the different phases of growth in batch culture. Further analysis suggested that this coordinate regulation is also found in cells grown on a variety of carbon sources and, in that case at least, operates at the level of transcriptional initiation. However the precise signal that the cell senses to trigger this regulation remains unclear. As mentioned previously, growth of the strain BY4742 in each of the different carbon sources employed in these assays leads to differing rates of proliferation. Accordingly expression of these coordinately-controlled genes was analysed with respect to the different cellular growth rates in each case. The expression of the coordinately-regulated gene set (represented by the results for *LSM7*) correlated very strongly with the changing growth rate of cells grown on the different carbon source (Figure 3.6). Therefore it is also possible that this regulation may be in response to varying cellular growth rate rather than through direct sensing of the different carbon sources in the media or within the cell.

3.9 Levels of most LSM *transcripts change coordinately in cells grown under aerobic or anaerobic conditions*

In order to identify the primary trigger for co-regulation of the *LSM* and *SM* genes, mRNA levels were examined in cultures growing in either aerobic or anaerobic conditions. As mentioned earlier, cells grown in batch culture in rich media will preferentially metabolise glucose by fermentation, using little or no oxygen (Dickinson, 1999; Macey and Miller, 1983). As glucose levels decrease the cell shifts to respiratory growth. While this switch from primarily anaerobic to

aerobic growth is largely in response to varying glucose levels, changes in oxygen availability can also have an effect (reviewed in Zitomer and Lowry, 1992). Comparison of transcript profiles of strictly anaerobic and aerobic cultures of *S. cerevisiae* show a greater than 3-fold change in more than 140 genes (Ter Linde *et al.*, 1999).

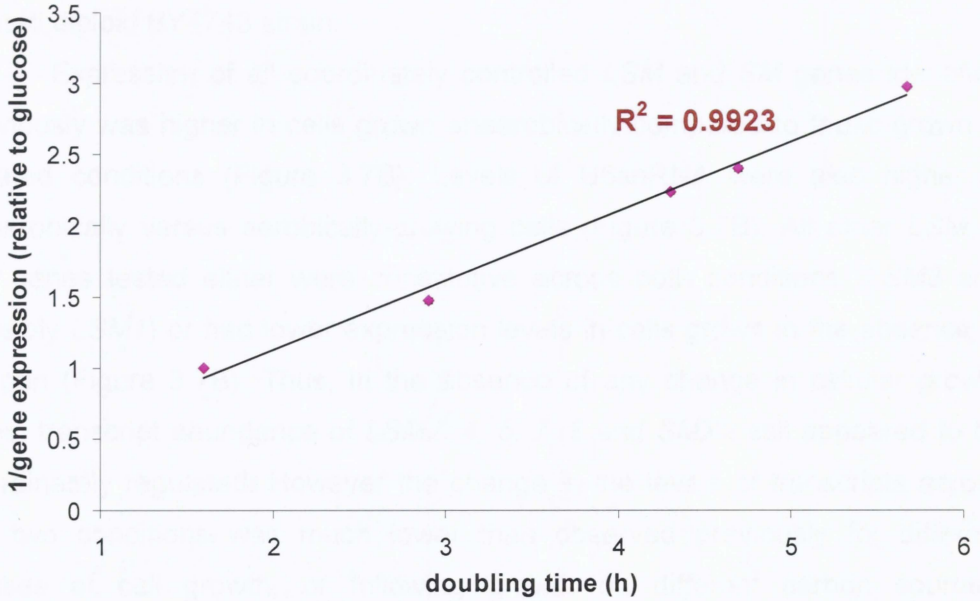


Figure 3.6 Correlation between growth rate and expression of coordinate control set genes. The doubling time of cells grown on different carbon sources was plotted against the inverse of the transcript level for each gene in the coordinate control set, here represented by *LSM7*. Correlation was measured by strength of linear association between two data sets, as represented by the correlation co-efficient (R^2).

Cells were grown anaerobically until they reached late-log phase (OD at 600nm of 1) and split into aerobic and anaerobic samples. Growth was then monitored over a period of 10 hours.

In accordance with previous reports, after around 8 hours of growth following the shift to different oxygen conditions the growth rate of the anaerobic culture began to decrease relative to that observed for aerobically growing cells (Figure 3.8A; Macy and Miller, 1983). Surprisingly, the growth rates immediately

following the shift in conditions until around 5 hours post-shift were the same for both cultures (Figure 3.7A).

In order to test if the *LSM* gene co-regulation observed in this study was in response to differential growth rate or some other factor, RNA was harvested from cells, both immediately before, and two hours after the shift in oxygen conditions. It should be noted that, while all previous experiments were conducted on the haploid BY4742 strain, these experiments employed the related diploid BY4743 strain.

Expression of all coordinately controlled *LSM* and *SM* genes identified previously was higher in cells grown anaerobically compared to those grown in aerated conditions (Figure 3.7B). Levels of U6snRNA were also higher in anaerobically versus aerobically-growing cells (Figure 3.7B). All other *LSM* or *SM* genes tested either were constitutive across both conditions (*LSM3* and possibly *LSM1*) or had lower expression levels in cells grown in the absence of oxygen (Figure 3.7B). Thus, in the absence of any change in cellular growth rates, transcript abundance of *LSM2, 4, 5, 7, 8* and *SMX2* still appeared to be coordinately regulated. However the change in the levels of transcripts across the two conditions was much lower than observed previously for different phases of cell growth, or following growth on different carbon sources. Therefore, while a change in cellular growth rate is not required to stimulate the co-ordinate regulation of these genes, cellular responses to differential cell growth may act to enhance this regulation under the various conditions tested above.

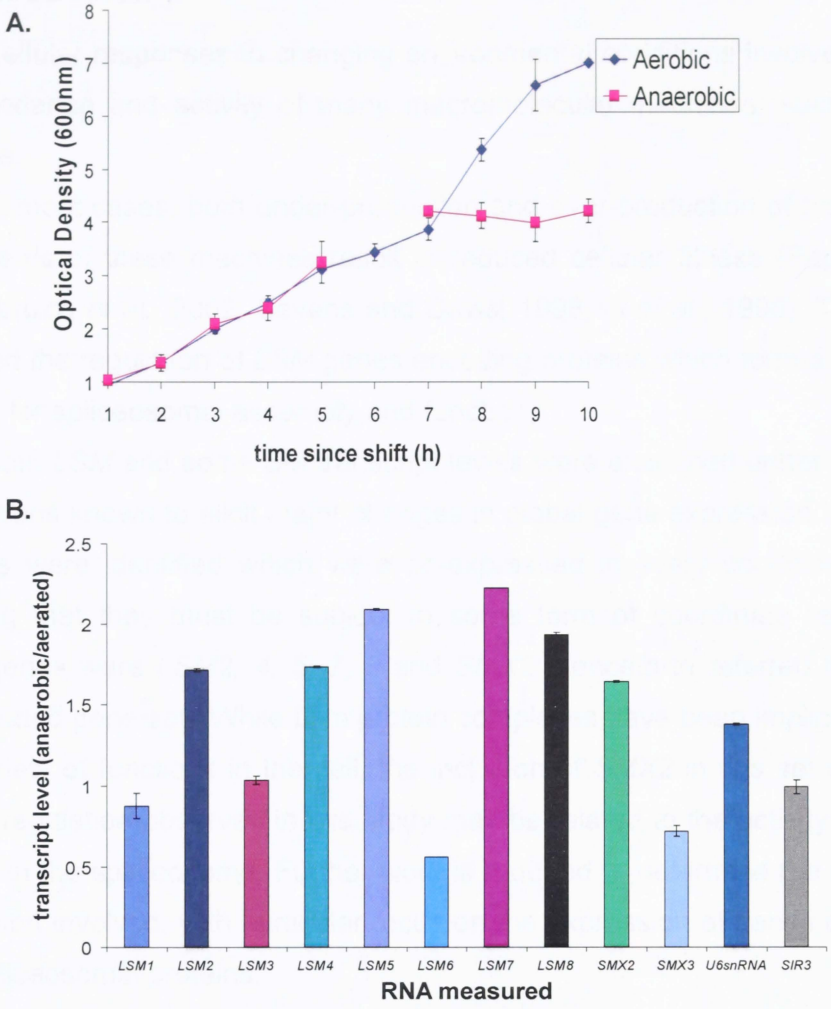


Figure 3.7 Change in transcript abundance of spliceosomal elements in cells grown under aerobic or anaerobic conditions. BY4743 cells were grown anaerobically until they reached an OD_{600} of 1. These cells were then split into two groups and grown either anaerobically or aerobically. **A.** Growth was monitored for 10 hours following the shift in oxygen conditions and optical density measured each hour. Error bars represent the standard deviation of the mean OD of triplicate cultures in each case. **B.** RNA was obtained from cells either immediately after or 2 h following the change in oxygen conditions. Relative levels of *LSM* and *SM* transcripts, as well as U6snRNA were determined, using *SIR3* as an external control. Data are also presented relative to the result obtained for aerated cells. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least three separate assays on samples from two independent experiments.

3.10 Discussion

Cellular responses to changing environmental conditions involve varying the abundance and activity of many macromolecular machines, such as the ribosome.

In most cases, both under-production and over-production of the protein components of these machines result in reduced cellular fitness (Papp *et al.,* 2003; Abruzzi *et al.,* 2002; Sevens and Davis, 1998; Li *et al.*, 1996). This work examined the regulation of *LSM* genes encoding proteins which form a complex required for spliceosomal assembly and function.

Both *LSM* and some *SM* transcript levels were examined under a variety of conditions known to elicit major changes in global gene expression. A group of genes were identified which were co-expressed in every condition tested, indicating that they must be subject to some form of coordinate regulation. These genes were *LSM2, 4, 5, 7, 8* and *SMX2*, henceforth referred to as the 'co-regulated gene set'. While Lsm protein complexes have been implicated in a wide variety of functions in the cell, the inclusion of *SMX2* in this set suggests that the regulation observed in this study may be related to the activity of these proteins in the spliceosome. Further work is required to determine the extent of the regulon involved, with particular focus on the expression of genes encoding other spliceosomal proteins.

During exponential growth in batch cultures containing glucose, global gene expression is relatively stable (DeRisi *et al.,* 1997). However, as glucose is depleted from the growth media the majority of transcripts in the cell decrease, some by more than 4-fold (DeRisi *et al.,* 1997). Indeed, the total RNA content of a given culture declines by nearly 50% around the diauxic shift (Ju and Warner, 1994). The expression of the co-regulated gene set also follows this pattern, decreasing dramatically as the cultures move towards 'late diauxic shift' and stationary growth. Interestingly, levels of most of the co-regulated gene set transcripts started to fall even while the culture was still growing exponentially, before levels of glucose would have become limiting. This pattern of expression is very similar to that reported for the majority of ribosomal protein genes in cells grown in batch culture (Ju and Warner, 1994).

The transcript abundance of the co-regulated gene set also followed a similar pattern to that of a large number of other genes in cells grown on media with different sugars; higher in cells grown in glucose than in those grown on poorer carbon sources (Kuhn *et al.,* 2001). This may reflect some trigger or mechanism of regulation that is shared between the co-regulated gene set and other genes with similar expression profiles. This will be discussed in more detail in Chapter 5.

The half lives of the majority of genes in the co-regulated gene set were the same in cells grown on either glucose or lactate. Therefore the change in transcript abundance in cells grown on different carbon sources is due to regulation at the level of transcriptional initiation. The average yeast mRNA half-life is about 24 min (Holstege *et al.,* 1998; Wang *et al.,* 2002). The transcript half-lives of most of the co-regulated gene set were relatively short, which would allow for a rapid mRNA response in the event of transcriptional repression. Although the levels of *LSM7* transcript abundance closely paralleled that of the rest of the genes in the co-regulated gene set, it was additionally regulated at the level of mRNA stability. Therefore *LSM7* or its products may be somehow involved in the regulation of the other genes in the set.

LSM7 was almost twice as stable in cells grown in glucose than in cells grown on lactate. A number of other genes have been identified which are more stable in glucose than in lactate, most notably the mRNAs involved in the biogenesis of both large and small ribosomal subunits (Yin *et al.,* 2003). Alternatively the change in mRNA stability may be due to increases in mRNA decapping and decay in response to different carbon sources, as occurs with transcripts of the *SDH1, SDH2* or *SUC2* genes (Prieto *et al.,* 2000; Scheffler *et al.,* 1998).

Lsm proteins are known to be expressed at low levels and one U6snRNP may be recycled over several consecutive splicing reactions (Moore *et al.,* 1993). However *LSM2* and *LSM7* mRNA were found to be so rare that it is likely that levels of Lsm2p and Lsm7p would determine the amount of functional Lsm complex available, even in the presence of significant translational regulation. Levels of Lsm proteins under the various conditions still need to be experimentally quantified and the influence of various types of post-

transcriptional regulation (such as control of mRNA export, translation or complex formation) assessed before this can be confirmed. Interestingly, *LSM8* is more abundant and has a longer half-life than any of the other *LSMs*, with the exception of *LSM7*, in cells grown on glucose. Therefore, assuming this trend is maintained up to the protein level, it is possible that Lsm8p may be present in slight excess and thus could immediately bind to the proteins adjacent to it in the Lsm spliceosomal complex (Lsm2p and Lsm4p) as they become available. This would sequester them (and the other proteins in the ring) for use in the splicing reaction instead of the Lsm decapping and decay complex (Lsm1p-Lsm7p). This presumes that chaperone proteins do not play any significant role in spliceosomal formation. Unfortunately testing such a model *in vivo* would be difficult.

Levels of U6snRNA also changed in a manner concordant with the expression of the co-regulated gene set in cells grown on different carbon sources. This could be the result of a number of different mechanisms. The spliceosomal Lsm complex associates directly with U6snRNA (Achsel *et al.,* 1999; Vidal *et al.,* 1999) and is required for its stability (Mayes *et al.,* 1999). Therefore lowering the expression of *LSM* genes encoding the majority of Lsm spliceosomal proteins may act to destabilise the U6 snRNA, leading to reduced levels in the cell.

Alternatively the decrease in U6snRNA abundance in media with poorer carbon sources may be due to separate but parallel regulation at the level of transcription. U6snRNA is transcribed by RNA polymerase III, unlike the other snRNA genes. (reviewed in Hernandez, 2001). RNA polymerase III has been found to partly dissociate from DNA during stationary phase growth when carbon source may be limiting, leading to a down-regulation in expression of Pol III genes. (Harismendy *et al.,* 2003).

Lastly, there may be some common regulatory element between the U6snRNA and the genes of the co-regulated set which, directly or indirectly, regulate the transcription of U6snRNA under these conditions. This will be discussed in more detail in Chapter 5.

Not all of the *LSM* genes were co-regulated in this manner. Levels of *LSM3* transcript were constitutive under all conditions examined. Furthermore,

levels of *LSM6* mRNA appeared to increase when that of the coordinate control gene set decreased. Accordingly it may not be necessary to coordinately regulate all elements in a complex in order to regulate the activity of that complex. Excess Lsm3p or Lsm6p could be avoided via translational regulation or, given that the transcripts for both proteins are of medium to low abundance, it may even be tolerated by the cell, as surplus protein production should impose little energetic cost. Alternatively, any Lsm3p or Lsm6p not employed in the splicing complex may instead be required for other cellular activities. In support of this theory Lsm3p has been found to form a stable homo-oligomeric ring complex *in vitro*, although to date no homo-oligomeric Lsm complexes have been observed *in vivo* (Collins *et al,* 2003).

The discrepant pattern of expression exhibited by the gene for Lsm6p was also shared by *SMX3,* which encodes a closely-related protein. The expression patterns of the gene encoding Lsm7p and *SMX2,* which codes for the evolutionarily-related protein SmG, were also highly similar. It would be interesting to explore this further by testing if the responses of genes encoding other Sm proteins are also shared with the genes encoding their related Lsm proteins, reflecting the common ancestry between the two protein groups.

While gene co-regulation was clearly observed across the various conditions tested, it was difficult to identify the specific environmental feature which was stimulating this response. Regulation of growth rate is very tightly linked to nutrient availability and it is difficult to distinguish the effects of each individually. Our results show that, for at least one condition (degree of aeration), *LSM* genes are coordinately regulated in the absence of any variation in growth rate. However further work is required to determine if this is the case in all conditions presented here, and, if so, what else the cell is sensing in each case. This could be addressed by examining *LSM* regulation in cells grown in continuous culture, where growth rate may be modulated under well-defined growth conditions, or by using other nutritional conditions, such as varying nitrogen sources, to alter growth rate.

In this chapter a group of *LSM* genes were identified which are co-regulated across a variety of conditions. The mechanism by which the cell achieves this regulation is explored in Chapter 4.

Chapter 4: Co-regulation of the *LSM* gene set is mediated by the *LSM7* intron

4.1. Introduction and Aim

In Chapter 3 a set of *LSM* genes was identified which were co-regulated under different growth conditions, most likely at the level of transcriptional initiation. The aim of the work detailed in this chapter was to determine the mechanism by which this regulation is achieved. DNA sequences surrounding and within the co-regulated genes were analysed to identify regulatory regions. Elements required for coordinate gene expression under these conditions were investigated through the use of deletion mutants. In the course of the analysis it was found that one of the *LSM* genes was involved in regulating transcription of other members of the set. This gene contained an intron and the role of this element in regulating the expression of the other genes in the set was explored.

4.1.1 Acknowledgements

Introns in the *LSM* and *SM* genes of yeast distantly related to *S. cerevisiae* were identified from various sequence databases by Prof. Ian W. Dawes (UNSW). Conservation of sequence in the promoter and across the gene of *LSM7* in five *sensu stricto* yeast (Figure 4.2B) was obtained from the UCSC Genome Browser database (Karolchik *et al.,* 2003).

4.2 Identification of putative regulatory motifs in LSM *promoters*

As mentioned above, control of transcriptional initiation requires assorted *cis*- acting sequences upstream of the core promoter region to which activator or repressor proteins may bind. Transcriptional co-regulation may be mediated by particular activator or repressor proteins acting on multiple, distant genes through common UAS or URS elements. Promoter sequences of *LSM* and *SM* genes were accordingly analysed *in silico* to identify known or putative regulatory motifs

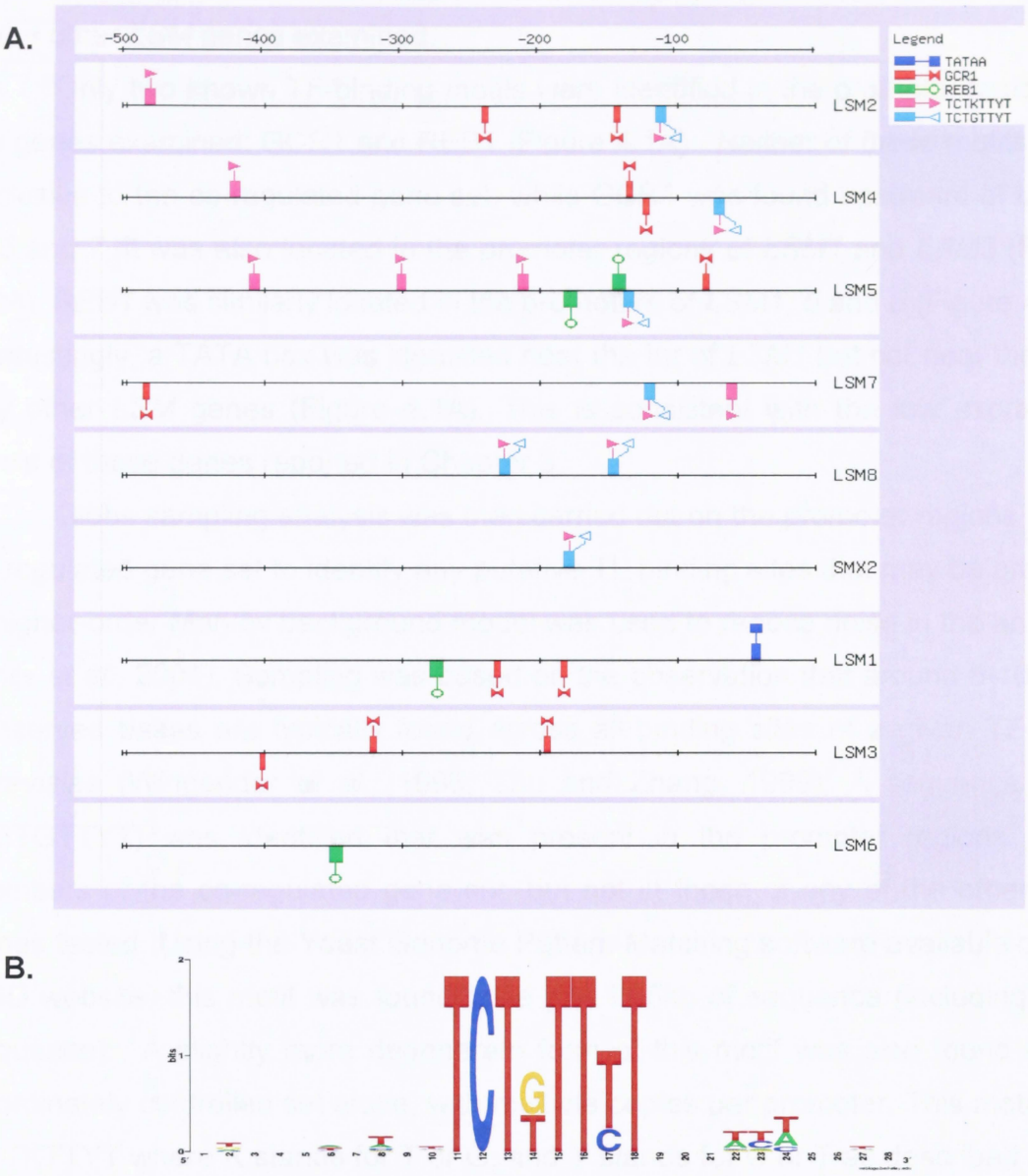


Figure 4.1:Transcription factor motifs in the promoters of *LSM* and *SM* genes
A. Feature map of predicted regulatory sites in *LSM* and *SM* promoters generated using RSA tools software (Helden *et al.*, 2000). Known protein-binding motifs (TATA box sequence and UAS sequence) were located in the promoters of *LSM* and an *SM* gene by matrix analysis also using RSA tools software (Helden *et al.*, 2000). Putative TF-binding motifs shared across the co-ordinate control set (TCTGTTYT and the more general TCTKTTYT) were identified through Gibbs sampler analysis using MotifSampler software (Thijs *et al.*, 2002). **B.** Weblogo display (Crooks *et al.*, 2004) of the consensus sequence motif located in the promoters. Each column of the alignment is represented by a stack of letters. The overall height of each stack indicates the sequence conservation at that position (ranging from 0-2 bits), whereas the height of letters within the stack reflects the relative frequency of the base at that position.

adjacent to all genes of the co-regulated set (*LSM2, 4, 5, 7, 8* and *SMX2)*, but not to the other *LSM* genes examined.

Only two known TF-binding motifs were identified in the promoter regions of the genes examined: GCR1 and REB1 (Figure 4.1A). Neither of these motifs were exclusive to the co-regulated gene set; while GCR1 was found upstream of *LSM2, 4*, *5* and *7,* it was also located in the promoter regions of *LSM1* and *LSM3* (Figure 4.1A). REB1 was similarly located in the promoters of *LSM1, 5* and *6* (Figure 4.1A). Interestingly, a TATA box was identified near the Inr of *LSM1* but not near those of any other *LSM* genes (Figure 4.1A). This is consistent with the low expression levels of these genes reported in Chapter 3.

Gibbs sampling analysis was then carried out on the promoter regions of the co-regulated gene set to identify any putative TF binding sites that may be present. A higher-order Markov background model was used to reduce noise in the analysis (Thijs *et al.*, 2001). Sampling was based on the observation that around 6-10 well-conserved bases are typically found across all binding sites of a given TF in *S. cerevisiae* (Wingender *et al.,* 1996; Zhu and Zhang, 1999). A sequence motif (TCTGTTYT) was identified that was present in the promoter regions of all members of the co-regulated gene set, but not in those of any of the other *LSM* genes tested. Using the Yeast Genome Pattern Matching software available on the SGD website, this motif was found once per 1.25kb of sequence (including ORF sequence). A slightly more degenerate form of this motif was also found in the coordinately controlled set alone, with multiple copies per promoter. This motif was TCTKTTYT where K stands for T or G, and Y stands for C or T as described in the IUPAC-IUB code (Figure 4.1A; Figure 4.1B).

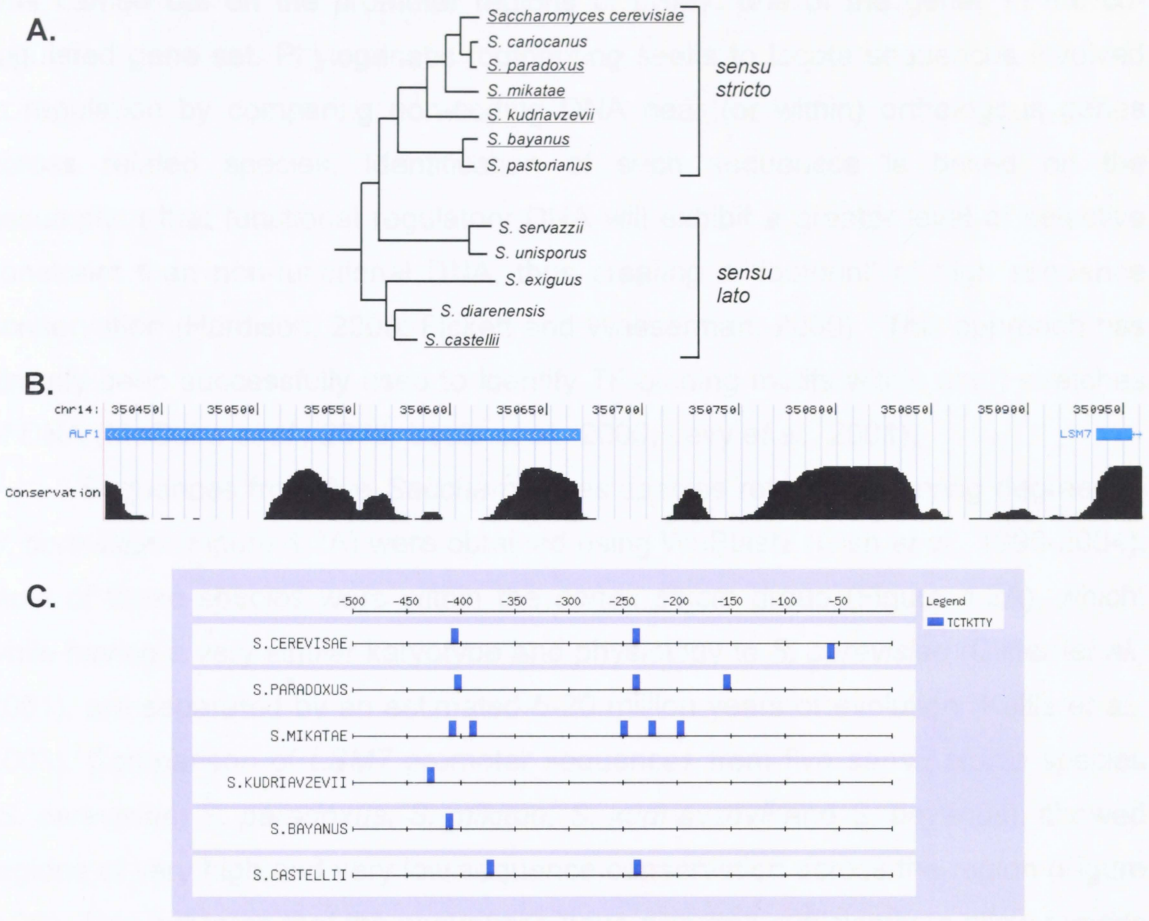


Figure 4.2:

A. Phylogeny of closely-related yeast species. Phylogenetic tree depicting both the physiologically similar *sensu stricto* and the more divergent *sensu lato* yeast species (adapted from Kurtzman and Robnett, 1998). Underlining indicates organisms examined in this study. **B.** Analysis of LSM7 promoter sequence conservation across four *sensu stricto* yeast species. Conservation is shown between regions of the genomes of *S. cerevisiae, S. paradoxus, S. mikitae, S. kudriavzevii* and *S. bayanus*. Data was extracted from the UCSC Genome Browser database (Karolchik *et al.*, 2003) and is based on sequence available from the *Saccharomyces Cerevisiae* Database. Chromosomal position in *S. cerevisiae* is shown at the top of the image, with coding regions and the respective direction of transcription indicated below in blue. **C.** Occurrence of the shorter putative TF motif in closely-related yeast species. Feature map of the shorter TF motif (TCTKTTY) in the *LSM7* promoters of *sensu stricto* (*S. cerevisiae, S. paradoxus, S. mikitae, S. kudriavzevii* and *S. bayanus*) and *sensu lato* (*S. castellii*) species **D.** Sequence conservation near the shorter putative TF motif in closely-related yeast species. Alignment of *LSM7* promoter sequences surrounding a conserved TF motif around 400 bp upstream of the start site. Sequence was obtained from *sensu stricto* (*S. cerevisiae, S. paradoxus, S. mikitae, S. kudriavzevii* and *S. bayanus*) and *sensu lato* (*S. castellii*) species using WuBlast2 (Gish *et al.*, 1996-2004) and aligned using Clustalw (Thompson *et al.*, 1994). The putative TF motif is shown in red. Shading indicates conservation in at least 4 of the sequences examined.. Sequences were obtained using WuBlast2 (Gish *et al.*, 1996-2004). The motif was identified within these regions and the image generated using RSA tools software (Helden *et al.*, 2000).

To examine the potential significance of this motif, phylogenetic footprinting was carried out on the promoter regions of *LSM7*, one of the genes in the co-regulated gene set. Phylogenetic footprinting seeks to locate sequences involved in regulation by comparing non-coding DNA near (or within) orthologous genes across related species. Identification of such sequences is based on the assumption that functional regulatory DNA will exhibit a greater level of selective constraint than non-functional DNA, thus creating a 'footprint' of high sequence conservation (Hardison, 2000; Fickett and Wasserman, 2000). This approach has already been successfully used to identify TF-binding motifs within short stretches of DNA (McGuire *et al.,* 2000; Loots *et al.,* 2000; Levy *et al.,* 2001).

Sequences from five *Saccharomyces* species related in varying degrees to *S. cerevisiae* (Figure 4.2A) were obtained using WuBlast2 (Gish *et al.,* 1996-2004). Most of these species were within the *sensu stricto* group (Figure 4.2A), which, while having a very similar karyotype and physiology to *S. cerevisiae* (Cliften *et al.,* 2001), are separated by an estimated 5-20 million years of evolution (Kellis *et al.,* 2003). Comparison of *LSM7* promoter sequences from five *sensu stricto* species (*S. cerevisiae, S. paradoxus, S. mikitae, S. kudriavzevii* and *S. bayanus*), showed regions of very high and very low sequence conservation across the region (Figure 4.2B). This indicates that the sequences have diverged sufficiently to allow regions of functional significance to be distinguished. In addition, no consensus TATA box motif could be found within 200 bp of the start of the coding sequence in any of the species.

The full TCTKTTYT sequence motif identified in the promoters of the co-regulated gene set in *S. cerevisiae* was not present in the *LSM7* promoters of many other closely-related yeast species. However a shorter form of this motif, TCTKTTY, was strongly conserved in the *LSM7* promoters of all of the species examined at around 400 bp upstream of the start of the *LSM7* coding sequence (Figure 4.2C). This sequence motif was even conserved at this position in the highly diverged *sensu lato* species, *S. castellii* (Figure 4.2C; Cliften *et al.*, 2001). Interestingly, this motif lies within one of the regions of high sequence conservation observed in the *LSM7* promoter across the five *sensu stricto* species, although it

should be noted that it also falls within the coding region of the divergent *ALF1* gene (Figure 4.2B and C). Closer analysis of the sequence in the different species around this site shows blocks of homology which gradually decrease in size further away from the TCTKTTY motif (Figure 4.2D), supporting the suggestion that it may have some functional role. However the biological significance of this motif has not yet been experimentally verified. Since *ALF1 is not an essential gene, s*equence comparisons of *ALF1* orthologues in closely-related yeast species may also provide some information regarding the importance of the motif in *LSM* gene regulation.

4.3 LSM7 is required for the carbon source response of the co-regulated LSM gene set

In Chapter 3 it was reported that *LSM7* displayed a pattern of mRNA abundance that strongly correlated with the rest of the *LSM* coordinate control gene set, but differed to some extent in its mechanism of regulation. Therefore *LSM7* may influence the transcriptional regulation of the other genes in the set. While most *LSM* genes encoding spliceosomal proteins are essential, mutants with the entire *LSM6* or *LSM7* gene deleted are viable, although they exhibit a growth defect and heat sensitivity (Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999*).* These mutants were obtained from the Euroscarf consortium (Brachmann *et al.,* 1998) and their status confirmed by PCR and growth on selective media (Figure 4.3A; data not shown). Levels of the various *LSM* transcripts were then determined in cells following growth on media containing glucose or lactate compared to results for wild-type cells. Data from *lsm6*Δ was analysed along with those from *lsm7*Δ to determine if the effects observed were due simply to the deletion of the gene for a protein involved in splicing.

As in previous assays, there was around a three-fold change in the expression of the coordinate control set (*LSM2, 4, 5, 7* and *8)* in wild-type cells grown on glucose versus those grown on lactate (Figure 4.3B). Relative to the wild type the change in expression of the co-regulated gene set was reduced in the *lsm6*Δ mutant (Figure 4.3B). However, the *lsm7*Δ mutant was no longer responsive

to changes in carbon source (Figure 4.3B). Therefore *LSM7,* or its products, must be involved in the regulation of the co-regulated gene set.

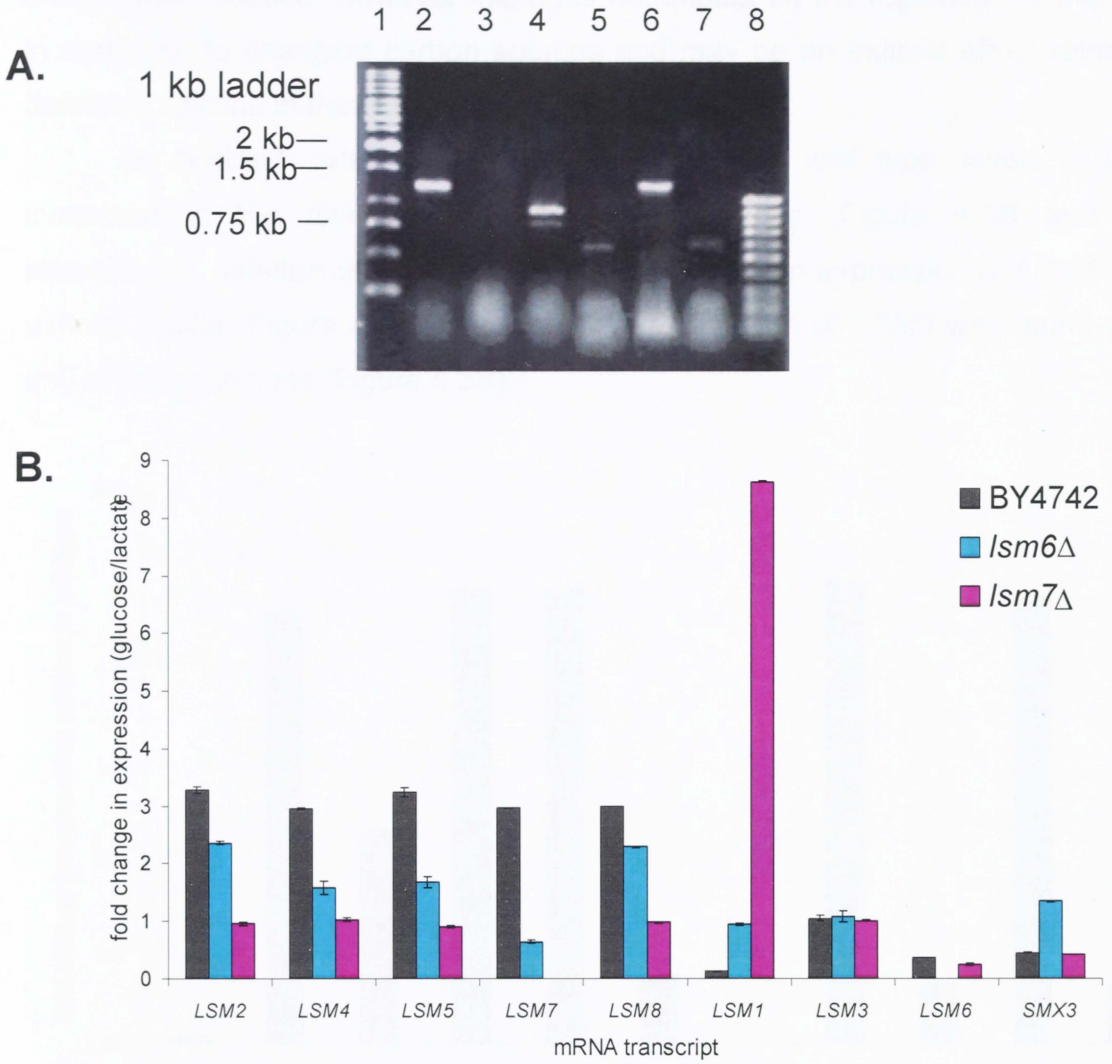


Figure 4.3 A **Confirmation of Euroscarf mutants**. Genomic DNA from BY4742 and *lsm1*Δ, *lsm6*Δ, and *lsm7*Δ mutants was amplified by PCR using the KanMXB primer specific KanMX, and a primer specific for sequence in the 5' promoter regions of *LSM1, LSM6, or LSM7*, respectively. PCR products were then separated by eletrophoresis on a 1.5% agarose gel. Lane 1 contains a 1 kb DNA ladder and lane 8 contains a 100 bp DNA ladder. Lanes 2-8 contained PCR products from: *lsm1*Δ using KanMXB and *LSM1* primer; wt using KanMXB and *LSM1* primer; *lsm6*Δ KanMXB using *LSM6* prime: wt using KanMXB and *LSM6* primer; *lsm7*Δ using KanMXB and *LSM7* primer; and wt using KanMXB and *LSM7* primer. Presence of bands in each case indicate the presence of the KanMX4 cassette. Bright bands in each case indicate the presence of the *KanMX4* cassette. The fainter bands observed are assumed to be PCR artifacts, since they were repeatedly found in negative controls for these assays.**B. Carbon-source response in mutants with *LSM6* or *LSM7* deleted**. Cells from the wild type or *lsm6*□ and *lsm7*□ mutants were grown on rich media in the presence of glucose or lactate. *LSM* or *SM* gene transcript levels were determined in each condition, with *SIR3* as an external control. Data shown represent the level of expression of each gene in cells grown on glucose relative to that in cells grown on lactate. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments.

Deletion of *LSM7* relative to wild type had little impact on overall changes in expression of *LSM6,* although in both media expression of this gene in the *lsm7*Δ mutant was reduced. However this does not impact on the regulation of this gene in response to changing carbon sources and may be an indirect effect related to defective splicing in this mutant.

By contrast, relative to that measured in the wild type, levels of *LSM1* increased in the *lsm7*Δ mutant grown on glucose (Figure 4.3B and 4.4). Interestingly, deletion of *LSM6* abolished any change in expression of *LSM1* in the various media (Figure 4.3B). No effect on expression of *LSM3* was found under any of the conditions (Figure 4.3B).

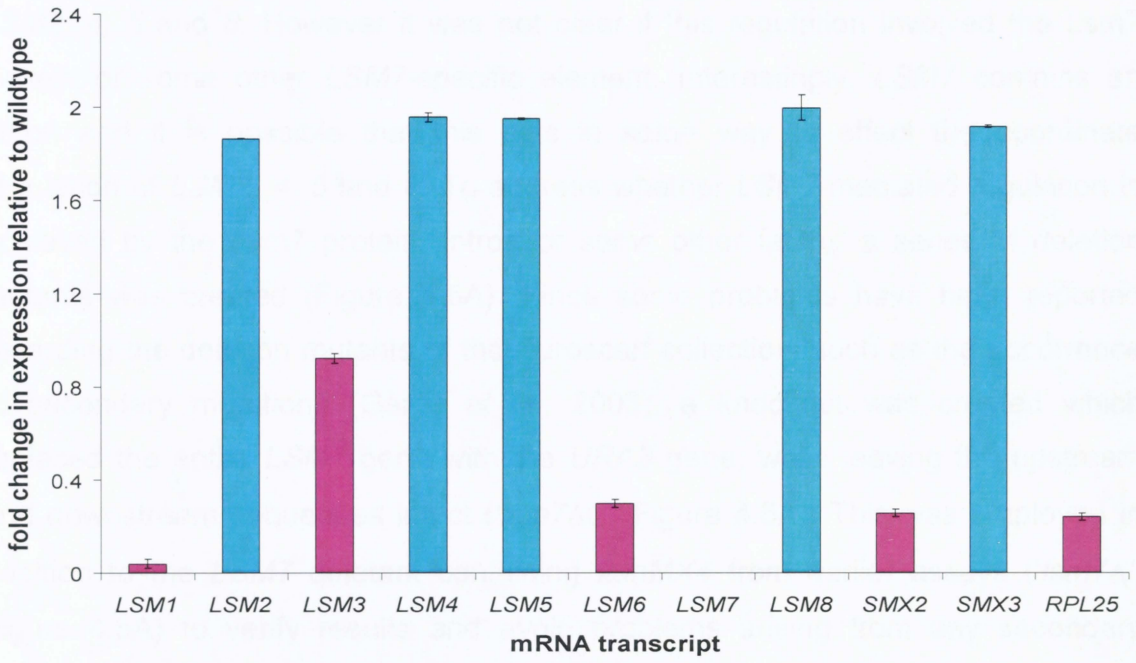


Figure 4.4: Gene expression in *lsm7*Δ cells relative to wild-type cells
Wild-type cells and *lsm7*Δ mutants were grown on rich media in the presence of glucose. *LSM* or *SM* gene transcript levels were determined in each condition, with *SIR3* as an external control. Data shown represent the level of expression of each gene in mutant cells grown on glucose relative to that obtained from wild-type cells under the same conditions. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments. Green indicates cells involved in the coordinate control set (*LSM2, 4, 5, 7, 8* and *SMX2*). Red shows the converse.

Following growth on glucose, all genes in the coordinate control set were expressed at a level almost two-fold higher in *lsm7*Δ than in wild-type cells (Figure 4.4). This trend was not observed in *lsm6*Δ cells (data not shown). This supports the hypothesis that *LSM7* acts to regulate the expression of these genes, probably by some repressive mechanism.

4.4 LSM7- dependent regulation requires the presence of the LSM7 intron

The above results showed that *LSM7* is required to regulate expression of *LSM2, 4, 5* and *8.* However it was not clear if this regulation involved the Lsm7 protein or some other *LSM7*-specific element. Interestingly, *LSM7* contains an intron and it is possible that this acts in some way to effect the coordinate regulation of *LSM2, 4, 5* and *8.* To address whether *LSM7*-mediated regulation is achieved by the Lsm7 protein, intron or some other factor, a series of deletion mutants was created (Figure 4.5A). Since some problems have been reported regarding the deletion mutants of the Euroscarf collection, such as the occurrence of secondary mutations (Game *et al.,* 2003), a knockout was created which replaced the entire *LSM7* gene with the *URA3* gene, while leaving the upstream and downstream sequences intact (*'lsm7*Δu', Figure 4.5A). This was employed in addition to the *LSM7* deletant containing *kanMX4* from earlier assays (*'lsm7*Δ*',* Figure 4.5A) to verify results and avoid problems arising from any secondary mutations that may exist elsewhere in the genome. The *lsm7*Δu strain was also used to generate an *LSM7* mutant with a functional *LSM7* gene that no longer contained an intron (*'intronless'*, Figure 4.5A). An alternative mutant was produced with all of the *LSM7* coding regions removed, leaving only sequences required for transcript initiation, splicing of the remaining intron and the stop codon (*'intron only'*, Figure 4.5A). Successful clones were confirmed through growth on selective media, PCR and sequencing (Figure 4.5B).

Each mutant was grown in rich media with either glucose or lactate as the carbon source and transcript levels subsequently measured using real time PCR. The amount of mutant *LSM7* transcript in the *intron only* mutant was measured by real time PCR analysis and found to be comparable to that of total *LSM7* mRNA in the wild-type cell (see Appendix). Further, spliced mutant mRNA could also be detected, although at levels lower than found for spliced native *LSM7* mRNA in the wild type (see Appendix).

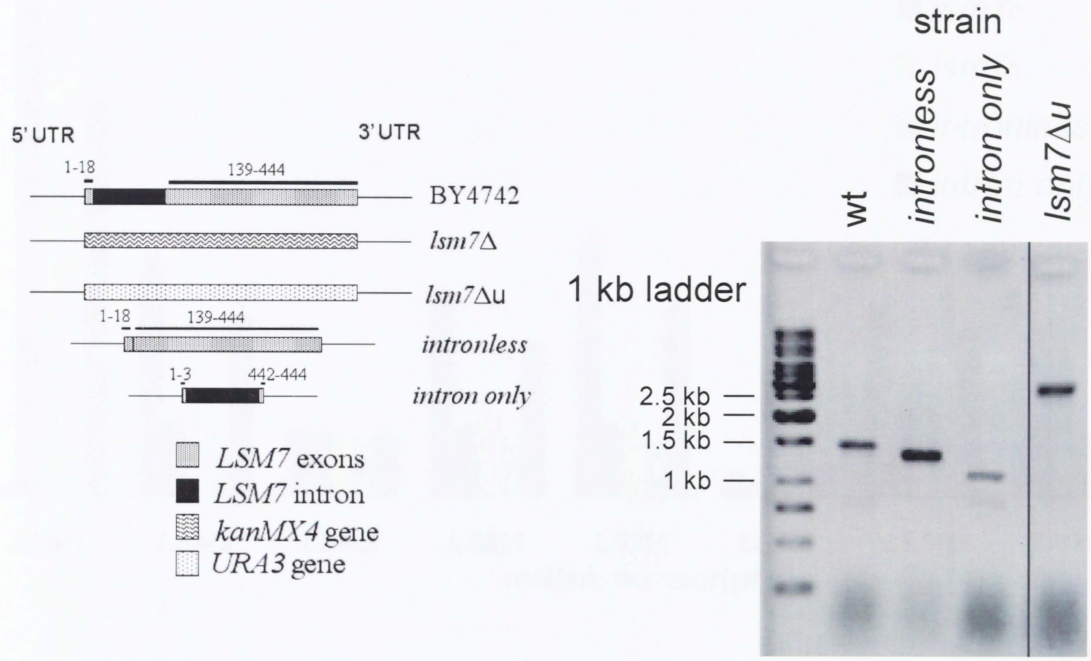


Figure 4.5: A. Creation of specific *LSM7* deletion mutants Five different strains were employed in this assay: the BY4742 wild-type strain; *lsm7*Δ, with the entire *LSM7* gene replaced by *kanMX4* (Brachmann *et al.*, 1998); *lsm7*Δu, with the entire *LSM7* gene replaced by *URA3* plus 800 bp of its promoter sequence; *intronless*, with the *LSM7* intron removed but the first (1-19) and second (139-444) exons present; and *intron only*, with the *LSM7* intron present but all of the exons except the start and stop codons for the gene deleted. Lines indicate *LSM7* exons. In each case the 5'UTR and 3'UTR of *LSM7* remained intact. **B: Confirmation of sucessful deletant mutants.** Genomic DNA from BY4742, *intronless, intron only* and *lsm7*Δu mutants was analysed by PCR using the *LSM* screening primers, which amplify the area across the 'gene' plus around 1 kb of *LSM7* promoter sequence. PCR products were then separated by eletrophoresis on a 1.5% agarose gel PCR product from wild-type should therefore be around 1.4 kb. PCR products indicating successful gene deletion should be around :2.6 kb for *lsm7*Δu, 1.3 kb for *intronless* and 1.1 kb for *intron only*.

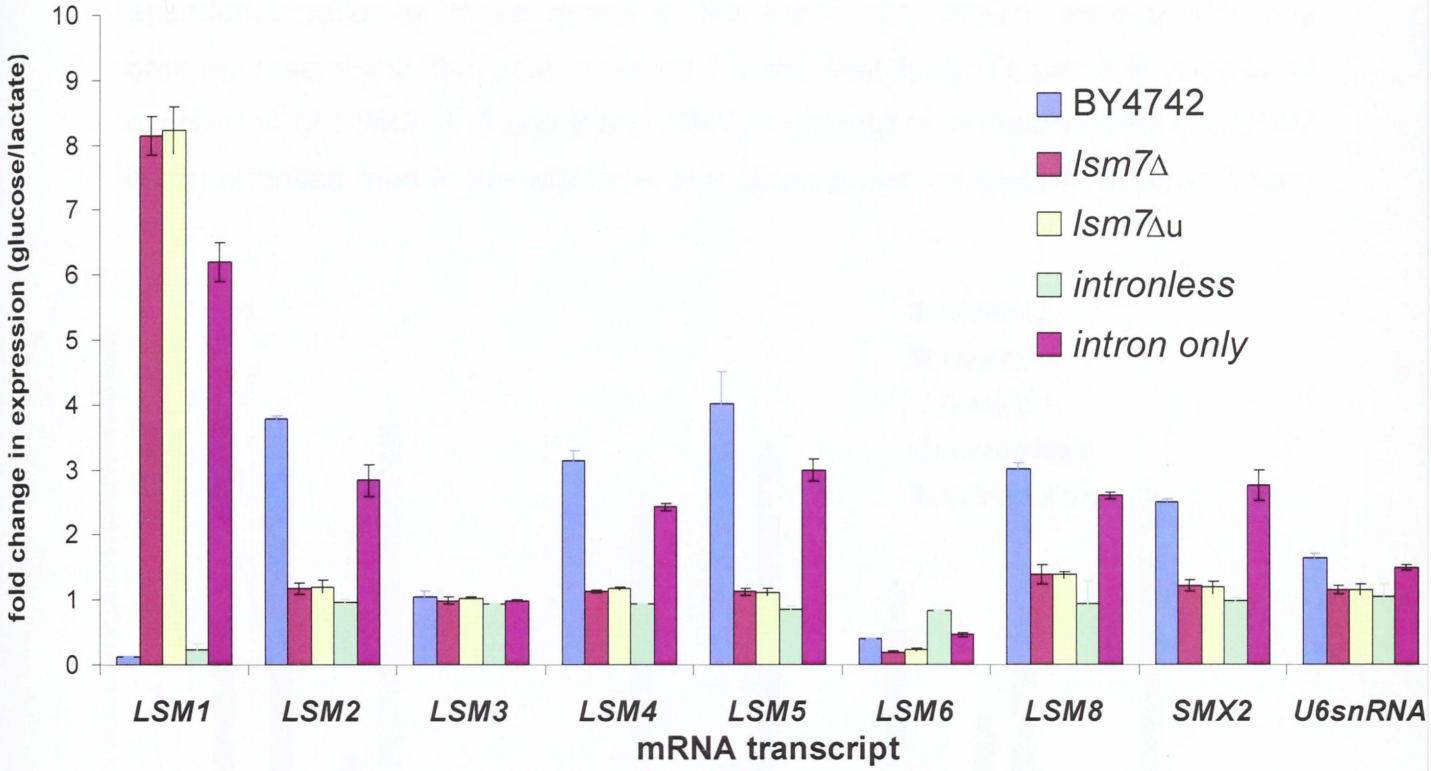


Figure 4.6: Carbon-source response of gene in mutants with different regions of *LSM7* deleted. Each strain was grown to exponential phase on rich media containing either glucose or lactate as the carbon source. Levels of *LSM* or *SM* mRNA, or U6snRNA were measured in each case with *SIR3* as an external control. Data shown represents the level of each transcript in cells grown on glucose relative to that obtained from cells grown on lactate. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments. Raw data from this experiment is shown in the Appendix.

In each case, results for both *lsm7Δ and lsm7Δu* were the same (Figure 4.6), confirming that there are unlikely to be secondary mutations in these strains that impact on the data measured. Changes in the ratio of transcript abundance for *LSM2, 4, 5, 8* and *SMX2* in the *intronless* mutant were similar to those measured for the complete gene knockouts, *lsm7Δ and lsm7Δu* (Figure 4.6), indicating a total loss of coordinate regulation. By contrast, the transcript abundance ratios of these genes in the *intron only* mutant were significantly greater, resembling the data obtained for the wild type (Figure 4.6). Levels of expression of *LSM2, 4, 5* and *8* and *SMX2* were higher in mutants with the *LSM7* intron removed than in the wild-type cells when grown on glucose (around 2-fold)

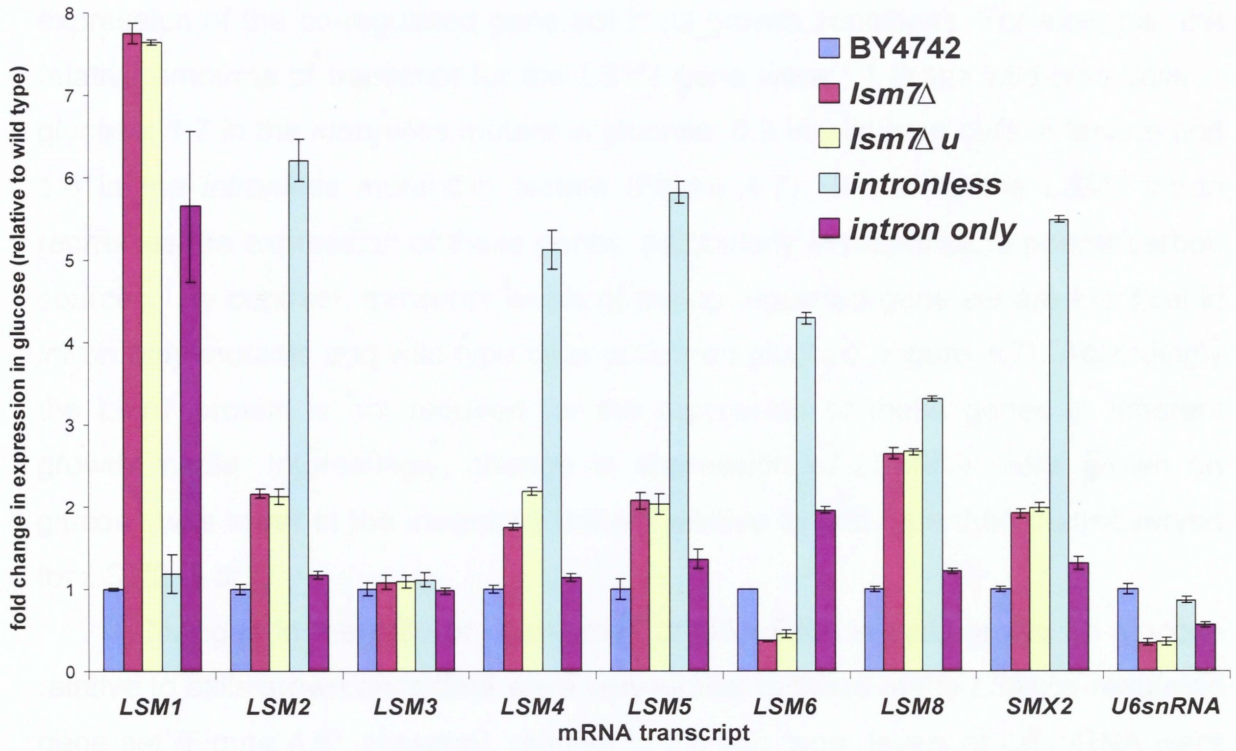


Figure 4.7: Gene expression in mutants with different regions of *LSM7* deleted relative to wild type
Each strain was grown to exponential phase on rich media containing glucose as the carbon source. Levels of *LSM* or *SM* mRNA, or U6snRNA were measured in each case with *SIR3* as an external control. Data shown represents the level of each transcript in mutant cells grown on glucose relative to that obtained from wild-type cells under the same conditions. Error bars represent the standard deviation of the mean transcript level for each gene, obtained from at least two separate assays on samples from three independent experiments.

or lactate (around 9-fold). The presence of the *intronless* construct alleviated the slow growth phenotype of *lsm7*Δu when grown on glucose, indicating that a functional Lsm7 protein is made in the *intronless* mutant (doubling times of 1.9 ±0.1 h and 3.2 ±0.2 h respectively). This effect was not observed in *intron only* mutants. These data indicate that the *LSM7* intron alone mediates most of the coordinate control of *LSM2, 4, 5, 8* and *SMX2* in response to different carbon sources.

While *LSM2* also contains an intron, it is unlikely that this plays an accessory or redundant role along with the *LSM7* intron in the regulation of *LSM2, 4, 5* and *8* and *SMX2* since the *intronless* mutant is completely unresponsive to changes in carbon source.

Strains with the *LSM7* intron removed had constitutively high levels of expression of the co-regulated gene set in all growth conditions. For example, the relative amounts of transcript for the *LSM4* gene were: 1 in the wild-type cells in glucose; 1.7 in the *intronless* mutant in glucose; 0.3 in wild-type cells in lactate and 1.6 in the *intronless* mutant in lactate (Figure 4.7). Therefore the *LSM7* intron represses the expression of these genes, particularly in response to poorer carbon sources. By contrast, transcript levels of the co-regulated gene set are identical in *intron only* mutants and wild-type cells grown on glucose (Figure 4.7). Accordingly the Lsm7 protein is not required for the repression of these genes in different growth media. Interestingly, change in expression of *LSM8* in cells grown on glucose was lower in the intronless mutant relative to wild type than that observed for *LSM2, 4* or *5*

Changes in the ratio of abundance of U6snRNA in cells grown on glucose relative to cells grown on lactate were very similar to those of the *LSM* co-regulated gene set (Figure 4.6). However, relative to the wild type, levels of U6snRNA were lower in mutants with either the *LSM7* intron or the entire *LSM7* gene removed, when grown on glucose (Figure 4.7). This indicates that, while elements of the co-regulated gene set influence the changes in U6 snRNA abundance, the effect may not be a direct result of transcriptional repression mediated by *LSM7* intron.

As with all previous assays, *LSM3* expression levels were identical in all strains tested (Figure 4.7). In contrast, the absence of Lsm7 protein led to

increased *LSM1* expression in glucose, indicating a possible role for this protein in *LSM1* regulation, although this may be an indirect effect (Figure 4.6 and 4.7). Neither the *LSM7* intron, nor the Lsm7 protein had a consistent impact on *LSM6* transcript levels compared to that observed in the wild type when grown on glucose. Therefore the changes in expression level observed in the different mutants are likely due to complex indirect effects (Figure 4.7). The effect of deleting either the *LSM7* intron or Lsm protein on expression of *LSM6* did not follow any consistent pattern, indicating that the changes observed are likely to be due to some indirect effect, such as changes in cell size or growth rate of the mutants relative to the wild- type.

4.5 The LSM7 intron sequence is conserved across yeast species

The data presented above indicate that the *LSM7* intron plays a regulatory role in the cell. Consequently phylogenetic footprinting was carried out on the *LSM7* intron sequence. This was performed to identify any areas of particularly high conservation which may contain important elements important to this regulation. Intron sequences from five *Saccharomyces* species related in varying degrees to *S. cerevisiae* were obtained using WuBlast2 as for section 4.2 (Gish *et al.,* 1996-2004). These were aligned and the surrounding sequence analysed to determine where each was located in the genome. In each case they were found to have been retained within the *LSM7* gene. *LSM7* exons as well as non-coding sequences in the promoter region of each gene were also aligned.

Lsm protein sequences are highly conserved across eukaryotes (Salgado-Garrido *et al.* 1999) and, as expected, the *LSM7* exons from the different yeast species were highly homologous, with 67% identity between exons from *S. cerevisiae* and the *sensu lato* yeast *S. castellii*. The *LSM7* intron was more highly conserved across the yeast species than other non-coding sequences 250 bp upstream of the gene (between *S. cerevisiae* and *S. castellii* there was 58% identity and 40% identity, respectively). As expected, both the 5'- and 3'- splice signals, as well as the branchpoint sequence, were conserved (Figure 4.8).

Moreover, blocks of strong conservation could be identified, particularly among the *sensu stricto* yeast, separated by 'hypervariable' sites where the bases diverged widely across the species (Figure 4.8). In *S. castellii* these blocks showed less conservation, although the intron length and splice site signals are still retained.

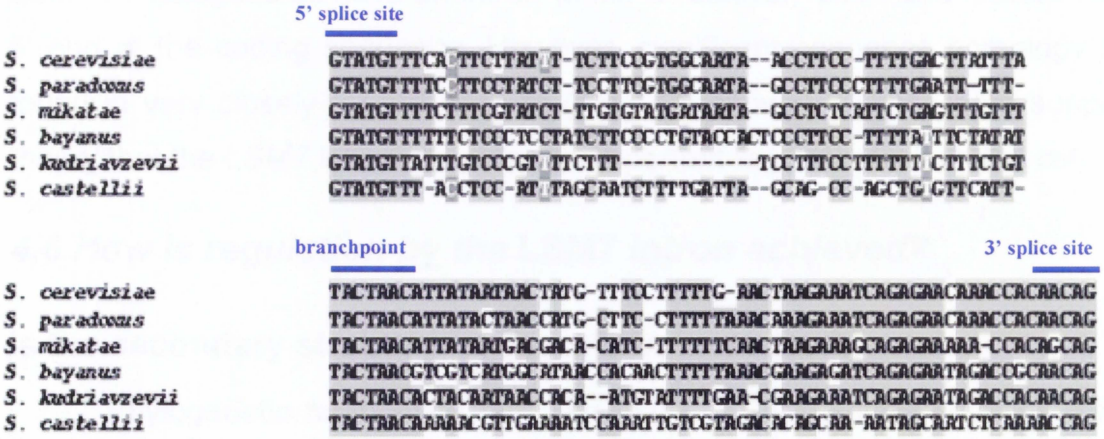


Figure 4.8: Conservation of the *LSM7* intron in closely-related yeast species

Alignment of *LSM7* intron sequences from five *sensu stricto Saccharomyces* species (*S. cerevisiae, S. paradoxus, S. mikitae, S. kudriavzevii* and *S. bayanus*) and a *sensu lato* species (*S. castellii)*. Sequences were obtained using WuBlast2 (Gish *et al.,* 1996-2004) and aligned using Clustalw (Thompson *et al.,* 1994). Splice sites and the branch point motif are indicated in blue. Shading indicates conservation in at least 3 of the sequences examined.

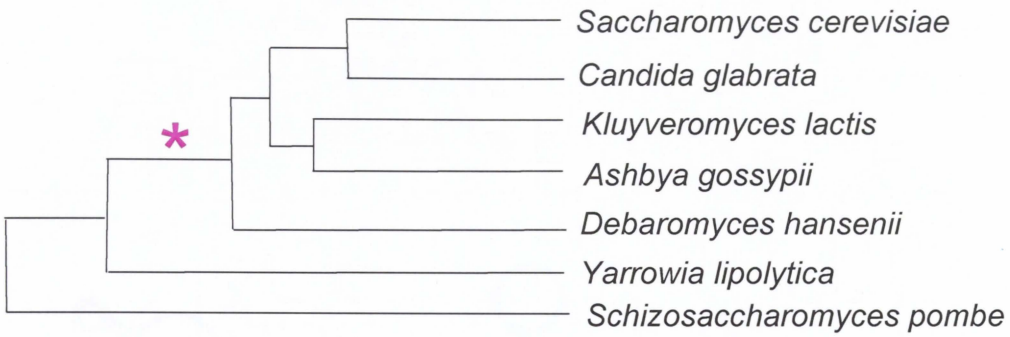
As illustrated in Figure 4.9A, introns were progressively lost during evolution from yeast species after the branch to *Y. lipolytica* (Bon *et al.,* 2003). Interestingly, while each *LSM* and *SM* gene contains an intron in *Yarrowia lipolytica* and *Schiz. pombe*, only the homologue of *LSM7* consistently retained its intron across all of the completely sequenced species (Figure 4.9B). The length of these introns and their position in the *LSM7* gene is highly similar, although by *K. lactis* most sequence conservation has been lost (data not shown). Furthermore, in the higher eukaryotes examined, including human, chimpanzee, mouse, rat, zebrafish, *Drosophila melanogaster*, mosquito, pufferfish and *Caenorhabditis elegans*, the *LSM7* homologues all have an intron which is relatively short and located near the 5' end of the coding sequence. However, significant sequence homology is only found in very closely-related species (data not shown). These data support the theory that the *LSM7* intron may play an important regulatory role in the cell.

4.6 How is regulation by the LSM7 *intron achieved?*

Is the secondary structure of the mRNA important?

Phylogenetic footprinting revealed that the *LSM7* intron contains blocks of sequence which are highly conserved, interrupted by short regions of highly variable sequence. Therefore the secondary structure of the intron of unspliced pre-mRNA may also be conserved. To investigate this possibility RNAfold (Hofacker *et al.,* 2003) was used to predict secondary structures of *LSM7* sequence from all *sensu stricto* species examined previously, as well as that from *K. lactis*, which diverged from *S. cerevisiae* more than 150 million years ago (Cliften *et al.,* 2001).

A.



B.

Gene	*S. cerevisiae*	*C. glabrata*	*K. lactis*	*D. hansenii*	*A. gossypii*	*Y. lipolytica*	*Sch. pombe*
LSM1	-	-	-	-	-	-	+
LSM2	+	+	-	-	+	+	+
LSM3	-	-	-	-	-	+	+
LSM4	-	-	-	+	-	+	+
LSM5	-	-	-	-	-	+	+
LSM6	-	+	-	-	-	+	+
LSM7	**+**	**+**	**+**	**+**	**+**	**+**	**+**
LSM8	-	-	-	-	-	?	+
SMD1	-	-	-	-	-	+	+
SMD2	+	+	?	+	+	+	+
SMD3	-	-	-	-	-	+	+
SME1	-	-	-	-	-	+	+
SMX3	-	-	+	+	-	+	+
SMX2	-	-	-	?	+	?	+
SMB1	-	-	-	+	-	+	+

Figure 4.9 Conservation of the *LSM7* intron in distantly-related yeast species
A. Phylogeny of distantly-related yeast species. Phylogenetic tree depicting the relationship between distantly related yeast species (adapted from Souciet *et al.*, 2000). The point at which introns are thought to have begun to be significantly depleted from the genome is indicated by '*' (Bon *et al.*, 2003). **B.** Patterns of intron retention and intron loss across distantly-related yeast species. Retention of introns in genes encoding *SM* or *LSM* genes across yeast species. Sequences were obtained from GeneDB (Hertz-Fowler *et al.*, 2004), the Ashbya Genome Database (Dietrich *et al.*, 2004) or Génolevures (Dujon *et al.*, 2004) and examined for the presence (+) or absence (-) of an intron. Instances where homologous genomic sequence was not identified are indicated by '?'.

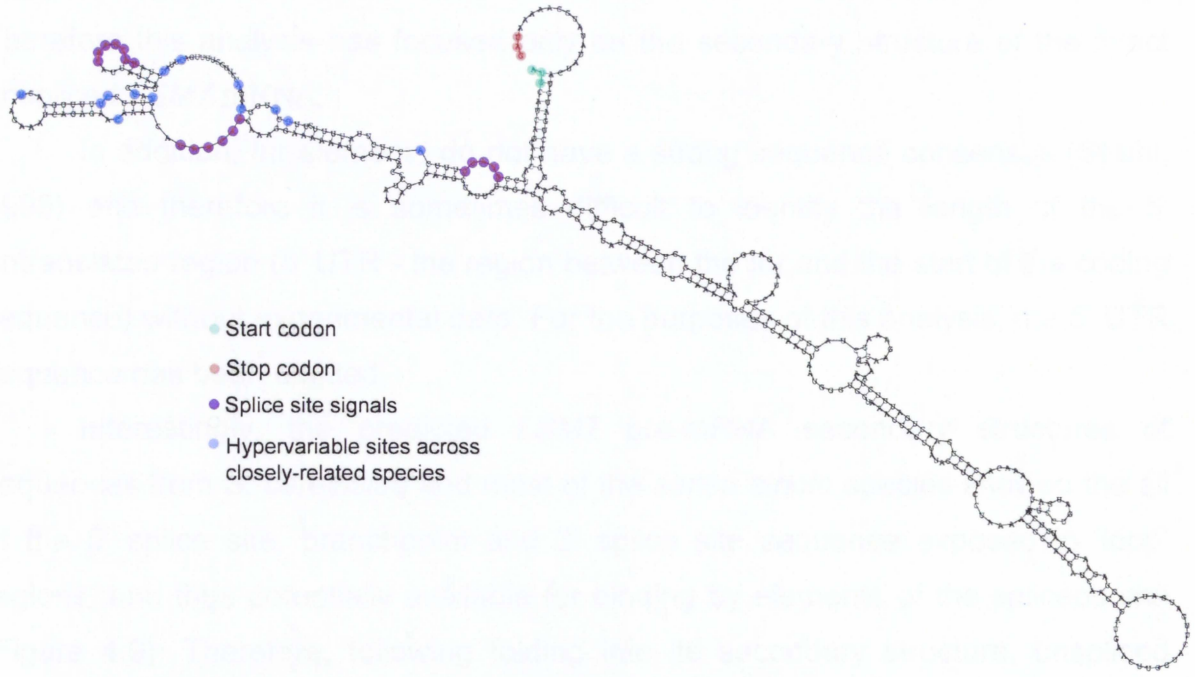


Figure 4.10: Secondary structure of unspliced *LSM7* mRNA
Unspliced *LSM7* mRNA secondary structure prediction for sequence from *S. cerevisiae,* adapted from analysis produced with RNAfold software (Hofacker *et al.*, 2003). This structure does not take into account any contribution from 5' UTR sequence. Start and stop codons are shown in green and red, respectively. Splice site signals (5' splice site, branchpoint and 3' splice site) are indicated in purple. Sites of highly variable nucleotide content across the yeast species examined are indicated in blue. The free energy of this structure at 30ºC is 117.23 kcal/mol.

It is unclear from the findings so far whether the spliced intron lariat structure or the entire unspliced *LSM7* mRNA is involved in the regulation of the coordinate control gene set under the various growth conditions. Unfortunately no *in silico* structural analysis programs are currently available which take into account the constraints of the lariat structure on intron secondary structure post-splicing. Therefore this analysis has focused only on the secondary structure of the intact unspliced *LSM7* mRNA.

In addition, Inr elements do not have a strong sequence consensus (Struhl, 1995) and therefore it is sometimes difficult to identify the length of the 5' untranslated region (5' UTR - the region between the Inr and the start of the coding sequence) without experimental data. For the purposes of this analysis, the 5' UTR sequence has been omitted.

Interestingly, the predicted *LSM7* pre-mRNA secondary structures of sequences from *S. cerevisiae* and most of the *sensu stricto* species showed the all of the 5' splice site, branchpoint and 3' splice site sequence exposed in 'loop' regions, and thus potentially available for binding by elements of the spliceosome (Figure 4.9). Therefore, following folding into its secondary structure, unspliced *LSM7* transcript may still be able to be spliced. Furthermore, regions of sequence hypervariability were also confined to loop regions, whilst the conserved blocks of sequence formed base-paired stem structures (Figure 4.10). Therefore mutation at the hypervariable sites would not interrupt the overall stem-loop structure. This supports the hypothesis that the secondary structure of the unspliced *LSM7* mRNA may indeed play an important role in the cell. The intron in *K. lactis* has not retained these blocks of homologous sequence, although the branchpoint sequence and the 3' splice site are still confined to 'loop' regions in the secondary structure of the mRNA (data not shown).

While some of the secondary structure in the unspliced mRNA is due to base-pairing solely within the intron, some of the stem structures predicted are created by base-pairing between both the intron and the second exon of the *LSM7* gene. Therefore, if the structure of the unspliced mRNA was important for

regulation, sequences within the second exon which may be required to interact with the intron sequence would be subject to more selective constraint and thus be more conserved. However, while the first exon is quite variable, even within the *sensu stricto* group, the entire sequence of the second exon is too highly conserved, to distinguish any potential regulatory regions from surrounding sequence.

Does the *LSM7* intron code for a regulatory protein?

Another way in which the *LSM7* intron may act to regulate the expression of the coordinate control set is by itself coding for some small regulatory protein. Indeed, a putative coding sequence, with in-frame stop codon, can be identified entirely within the intron of *S. cerevisiae* (Figure 4.11A). When translated, this forms a peptide sequence 19 bp in length which, although unusually short, may still potentially form a functional peptide (Figure 4.11B). These amino acids form highly hydrophobic domains which score more than 2 on the Kyte-Doolittle scale; a significant result given that highly hydrophobic membrane-spanning domains generally have a score of 1.6 or above. Thus it is theoretically possible that, if the spliced intron was exported out of the nucleus and translated, it could form a small, highly hydrophobic protein. This may be consistent with a possible role for this protien in transcriptional regulation, since hydrophobic residues appear important for transcription factor activity (Struhl, 1989). Analysis of possible open reading frames that span coding and intron-containing sequences in the unspliced *LSM7* sequence did not yield any long enough to encode other potential polypeptide sequences of sufficient length to be considered functional

If the production of a protein from intronic sequence is required for the carbon-source regulation of the *LSM* co-regulated gene set, it may be expected that the putative 'start' and 'stop' codons would be retained in related species. Both the proposed start and stop codons in the *LSM7* intron of *S.cerevisiae* lie within highly conserved splice site sequences (the 5' splice site and branchpoint sequence, respectively). Accordingly they are present in all of the *sensu stricto* species examined, as well as the *sensu lato* yeast, *S. castellii* (Figure 4.8).

However, with increasing evolutionary distance the stop codon is either rapidly lost or the putative protein sequence interrupted by a premature stop codon, resulting in a sequence which is too short to encode a functional protein. Furthermore, by *Y. lipolytica* the proposed start codon has also been lost.

A.

ATGCATCAGCAACACTCCGTATGTTTCACTTCTTATTTTCTTCCGTGGCAATAACCTTCCTTTTGACTT
ATTTATACTAACATTATAATAACTATGTTTCCTTTTTGAACTAAGAAATCAGAGAACAAACCACAACAG
CAAAGGAAAAAATTCGAAGGCCCTAAAAGAGAAGCTATTCTGGATTTAGCGAATGATAAAGATTCTAAA
ATTCGCGTCAAATTAATGGGTGGTAAATTAGTTATAGGTGTCCTAAAAGGCTATGATCAACTGATGAAC
TTGGTACTTGATGATACAGTAGAATATATGTCTAATCCTGATGATGAAAACAACACTGAACTGATTTCT
AAAAACGCAAGAAAGCTAGGTTTGACCGTCATAAGAGGTACTATTTTGGTCTCTTTAAGTTCCGCCGAA
GGTTCTGATGTACTATATATGCAAAAATAG

Translated sequence: MFHFLFSSVAITFLLTYLY

B.

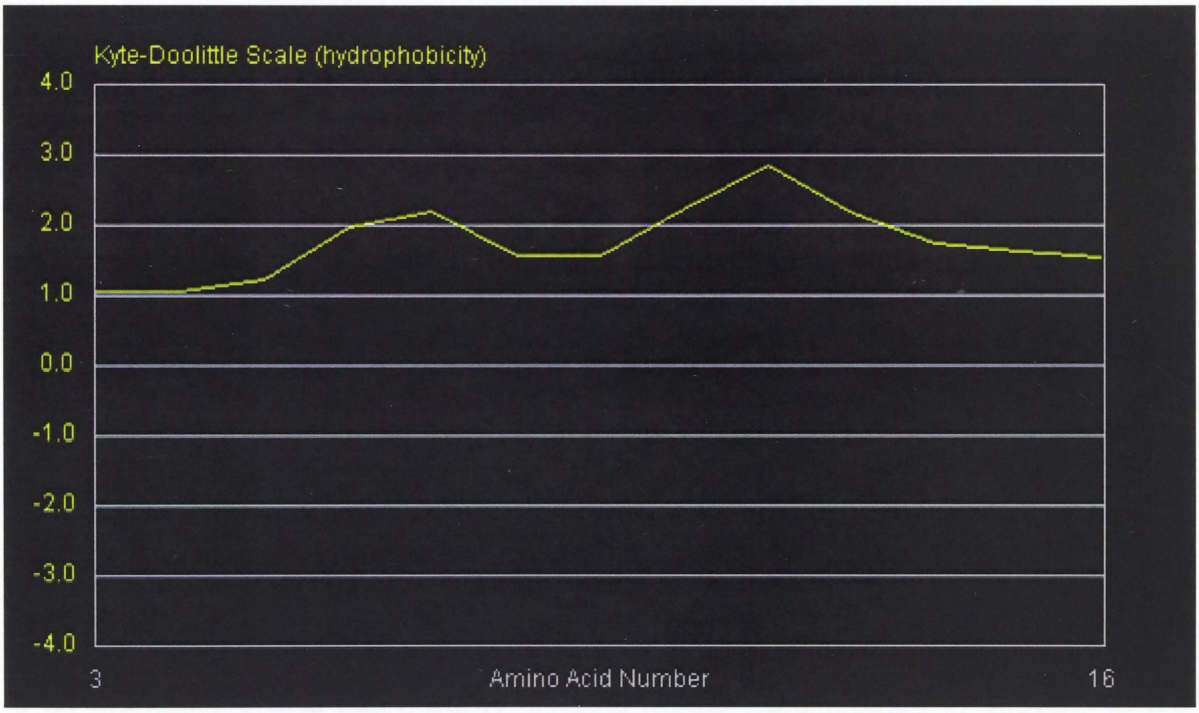


Figure 4.11:
A. Putative protein-coding sequence within the *LSM7* intron. Unspliced *LSM7* mRNA sequence from *S. cerevisiae* indicating the site of a putative small protein. Both confirmed and theoretical start and stop codons are shown in green and red, respectively. Splice site signals (5' splice site, branchpoint and 3' splice site) are shown in blue. The putative protein sequence in grey, and the translated polypeptide sequence presented below. **B.** Hydrophobicity plot for the putative protein within the *LSM7* intron. Degree of hydrophobicity of the residues making up the translated putative protein sequence within the *LSM7* intron. Image generated using Molecular Toolkit software from the University of Colorado (http://arbl.cvmbs.colostate.edu/molkit/hydropathy). Hydrophocity was measured on the Kyte Doolittle scale.

4.7 Discussion

Regulation of transcription in eukaryotes involves a complex interplay between activator and repressor factors. These factors can coordinate changes in expression of thousands of genes scattered across the chromosomes by modulating specific interactions with *cis*-acting promoter elements. This study explored potential mechanisms by which coordinate transcriptional regulation of *LSM2, 4, 5, 7, 8,* and *SMX2* may be achieved under different growth conditions.

Interestingly, *in silico* analysis of the promoters of these co-regulated genes found no consensus TATA boxes, despite the central role this motif is believed to play in specifically binding the basal transcription machinery of a number of genes in higher eukaryotes. The exact mechanism by which transcription is initiated in genes with TATA-less promoters has not been fully elucidated. However it is thought that, in the absence of the TATA box, proteins of the basal transcription machinery bind to other promoter elements, possibly in a non-specific manner (as discussed in section 1.7.2; Sharp, 1992; Struhl, 1995). In higher eukaryotes TATA-less promoters have traditionally been associated with weakly-transcribed genes (Robinson and Lopes, 2000) or genes coding for proteins involved in basic housekeeping functions requiring relatively mild regulation, such as protein synthesis (Dynan, 1983, Basehoar *et al.*, 2004). This group is also enriched for genes which are regulated in response to stimulation of cell proliferation (Azizkhan *et al.,* 1993). This is consistent with one of the proposed triggers for regulation of genes of the coordinate control set explored in Chapter 3 (Section 3.8).

Genes that are transcriptionally co-regulated often contain common TF-binding motifs within their promoters. However analysis of the promoter regions of genes in the co-regulated set did not reveal any known TF-binding motifs common to all members of the set. Two TF-binding motifs were located in subsets of the genes examined: GCR1 and REB1.

GCR1, found in the promoters of the co-regulated genes *LSM2, 4, 5* and *7*, has been shown to modulate the expression of a wide array of genes, principally those encoding glycolytic enzymes (Scott and Baker, 1993). REB1, located in the promoters of *LSM1, 5* and *6*, has been shown to play a role in the regulation of

both ribosomal protein genes as well as a number of genes involved in the glycolytic pathway (Yagi *et al.*, 1989; Kulkens *et al.,* 1992; Scott and Baker, 1993) and is often located in promoters which lack a consensus TATA box (Liaw and Brandl, 1994). It has been proposed that REB1 acts to modulate chromatin structure resulting in the removal of nucleosomes from the region immediately upstream of its binding site, thus allowing other TFs to bind to this area (Fedor *et al.,* 1988; Brandl and Struhl, 1990).

Matrix analysis of the promoters in the co-regulated gene set detected a putative sequence motif common to the promoters of all genes within the coordinate control set, but not the promoters of the other *LSM* genes examined. Non-coding regions which are conserved across the promoters of closely-related species are enriched for TF-binding sites when compared with background sequence (Levy *et al.,* 2001). A slightly shortened form of this putative motif was found within such an area of comparatively high sequence homology in the *LSM7* promoter of four of the closely-related *sensu stricto* species. Alignment of sequence around the putative motif in this region from both *sensu stricto* and *sensu lato* yeast showed that this motif was indeed part of an unusually highly conserved block of sequence. However conservation of sequence does not necessarily imply functionality and the biological significance of this motif has yet to be verified experimentally, through mutation or deletion analysis. Given the complex regulatory role of *LSM7* outlined in the second part of this chapter, it may have been more appropriate to perform phylogenetic footprinting on different genes from the co-regulated gene set.

Analysis of deletion mutants confirmed that the pattern of regulation exhibited by the co-regulated gene set under the different conditions depends on the presence of *LSM7,* which is itself one of the members of this gene set. Comparisons between expression data obtained from *lsm6*Δ and *lsm7*Δ mutants indicate that this regulation is specific to *LSM7* and not merely an indirect effect of deleting a protein involved in various RNA-processing complexes such as the spliceosome. Use of various *lsm7* deletion mutants further defined the regulatory element required for this response. Mutants with the coding region of the gene

deleted, but the intron intact, behaved in a manner similar to that of the wild-type cell. By contrast, cells with only the *LSM7* intron removed exhibited complete loss of coordinate regulation. Therefore the *LSM7* intron, and not the Lsm7 protein, mediates the coordinate control of *LSM2, 4, 5, 8, SMX2,* and probably *LSM7* under these conditions. This is the first clear example of an intron acting in *trans* to regulate the expression of other genes. This result demonstrates that introns may have important regulatory functions in the cell, well beyond those currently reported. This will be discussed in more detail in Chapter 6.

Cells with the intron deleted (*lsm7*Δ or *intronless*) exhibited a higher level of expression of *LSM2, 4, 5* and *8* following growth on either glucose or lactate than that found in wild-type cells under glucose conditions. Therefore the *LSM7* intron represses the expression of these genes, especially in response to poorer carbon sources.

Levels of U6snRNA were also measured in the various *LSM7* mutants and followed the same pattern as transcripts from the *LSM* co-regulated gene set across the different media. This further supports the premise put forward in Chapter 3, that the activity or abundance of members of the co-regulated gene set impact on the amount of U6snRNA available. This is in accordance with reports from Verdone *et al.* (2004) suggesting that reconstitution of *LSM7*Δ mutants with Lsm7p does not lead to a significant change in U6 snRNA levels. However, relative to the wild type, levels of U6 snRNA were lower in mutants with the *LSM7* intron removed. This is also consistent with previous work by Verdone *et al.* (2004) and Mayes *et al.* (1999) on *LSM7*Δ. Taken together, this suggests that any regulatory effect is unlikely to be a direct result of repression of RNA polymerase III-mediated transcription by the *LSM7* intron. Therefore regulation of U6snRNA under these conditions must involve a different primary mechanism than that of the *LSM* co-regulated gene set.

Cells with *LSM6* deleted no longer responded to growth media by altering levels of *LSM*. In addition, data from the assorted *LSM7* deletion mutants indicate that removal of the Lsm7 protein leads to a de-repression of *LSM1* in cells grown

on glucose. Given the complexity of these responses, further work must be carried out to determine if these effects are direct or due to some indirect mechanism.

While *LSM2* also contains an intron in *S. cerevisiae*, it does not appear to play an important or even redundant role in the regulation described here since this control was abolished by the deletion of the *LSM7* intron. However the potential activity of the *LSM2* intron under different conditions has not yet been explored.

Inspection of *LSM7* intron sequence alignment across closely-related species, as well as *S. castellii*, showed blocks of highly conserved sequence interspersed with hypervariable sites where mutations appear to have accumulated over evolutionary time. Predicted secondary structure of unspliced *LSM7* message from a number of *Saccharomyces* species suggest that these hypervariable sites fall exclusively within 'loop' regions. Therefore variations in nucleotide composition at these sites in the different species would leave the base-paired stems intact. This suggests that the pre-mRNA adopts a highly specific stem-loop structure in which the stems are functionally important. Interestingly all of the splice site signals are also within loop regions of the predicted structure, where they should be available to interact with components of the spliceosome. However *in silico* structure prediction cannot take into account all physiological influences present *in vivo*, and these results should be experimentally confirmed.

In current models of transcription of genes containing introns, snRNPs are recruited to the nascent transcript by elements of the transcription machinery (Fong and Zhou, 2001). This would inhibit any *LSM7* unspliced message from forming the secondary structure illustrated above. However a study by Eperon *et al.,* (1988) suggests that structures containing relatively short lengths of intronic sequence (110 nucleotides or less) may have time to form before snRNPs can bind. Indeed the effect of pre-mRNA secondary structure *in vivo* has been invoked to explain a number of different findings regarding the expression of intron-containing genes, including the production of defective proteins in human disease (Takahashi *et al.*, 1998; Varani *et al.,* 1999; Buratti and Baralle, 2004;).

There are several alternative reasons why the secondary structure of the unspliced *LSM7* transcript may have been conserved across yeast species. One

possibility is that formation of stem-loop structures across this region enhances the splicing efficiency of *LSM7* pre-mRNA, as has been observed for other genes previously in both yeast and higher eukaryotes (Goguel and Roshbash, 1993; Chen and Stephan, 2003). Alternatively, double-stranded stem-loop structures may trigger degradation of accumulated unspliced *LSM7* transcript by Xrn1p and Rapt1p exonucleases and the nuclear exosome, via the activity of Rnt1p (Danin-Kreiselman *et al.,* 2003). This possibility may be particularly likely if accumulation of this structure was deleterious to the cell under some conditions.

It should be noted that it has not yet been determined if the conserved intronic sequences in *LSM7* are related to the structure of the entire unspliced transcript or the just the spliced intron lariat. However Rnt1p also initiates degradation of double-stranded intron lariats in the yeast nucleus (Danin-Kreiselman, *et al*, 2003).

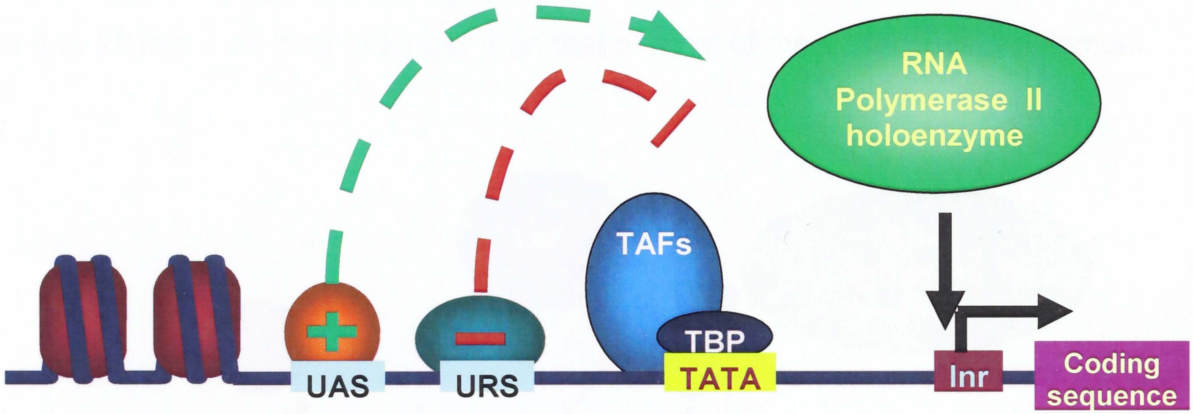


Figure 4.12 Simplified schematic of typical eukaryotic promoter
This model shows the mechanism by which regulation of transcriptional initiation usually occurs. For simplicity, the RNA polymerase II holoenzyme is shown as a completely assembled complex and TBP-associated factors have been omitted. For a more detailed illustration, see Figure 1.11.

In this chapter it was shown that the *LSM7* intron partially represses the expression of *LSM2, 4, 5* and *8, SMX2,* and probably *LSM7* itself, in cells grown on poor carbon sources. Figure 4.12 shows an overview of regulation at a typical eukaryotic promoter. Based on all of these findings I propose the following four different models to explain how such regulation may be achieved:

'Counteract' Model: The intron encodes a repressor protein

In this model mRNA derived from the intron sequence is exported into the cytoplasm where it is translated to form a functional protein. This protein can then move back into the nucleus where it specifically binds to URS of the genes in the co-regulated gene set, as illustrated in Figure 4.13. Once bound it can then act to inhibit binding of activators to the UAS or binding of the basal machinery to the DNA, possibly via other cofactor proteins. Alternatively this repressor protein could interact with activators directly to alter their activity.

Since it is not yet known if the *LSM7* intron needs to be spliced to mediate repression, the mRNA in this model may be derived from either the intron lariat, post-splicing, or from *de novo* transcription of the putative coding sequence in the DNA. Until recently it was generally assumed that, following excision, most introns are rapidly degraded at their site of origin in the nucleus (Padgett *et al.,* 1986). However initial studies addressing the issue show that intron half-lives can vary from 6 to 29 min - as long or longer than that of most of the *LSM* mRNAs (Clement

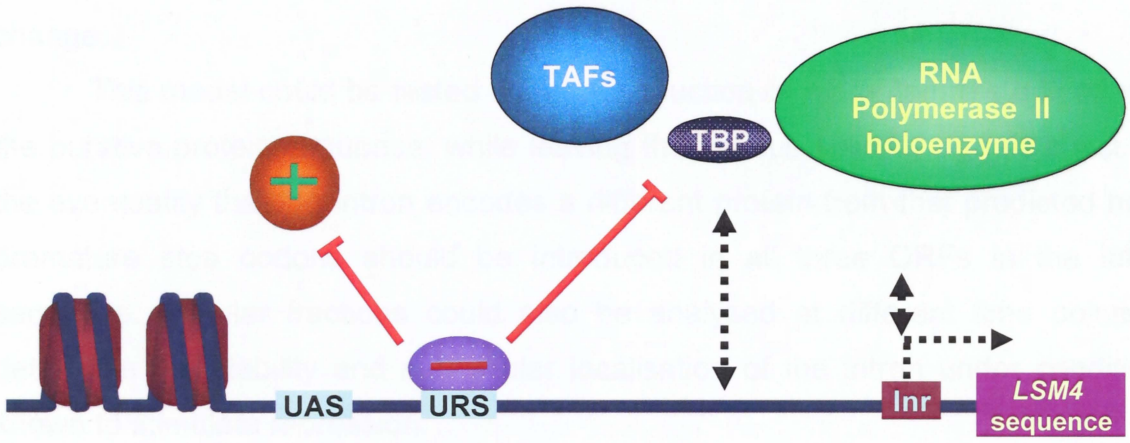


Figure 4.13: 'Counteract' model of regulation.
This illustrates one way in which the *LSM7* intron may mediate transcriptional repression of the co-regulated gene set (*LSM2, 4, 5, 8, SMX2* and possibly *LSM7* itself). The *LSM7* intron is hypothesised to be transported out into the cytoplasm, where it is translated. The small protein produced could then re-enter the nucleus and act to repress transcription by binding to URS of the regulated gene (here represented by *LSM4*), and blocking binding of activator proteins or the transcriptional machinery directly. Dotted lines represent reduced binding or transcription at that point.

et al.,2001; Figure 3.1) Further, there is some evidence that some introns are exported into the cytoplasm (Clement *et al.*,2001). However, as the 'start' codon for this putative protein is within the 5'-splice site of the intron, it is likely that formation of the lariat structure during splicing (where the 5'-splice site is bound to the branchpoint sequence; Section 1.3) would block any initiation of translation. Therefore *de novo* transcription from within the intron in the DNA would be a more likely to produce translatable mRNA. This could be confirmed using an *intron only* mutant with the *LSM7* sequence immediately upstream of the intron altered to remove regions normally required for the transcription of the entire *LSM7* gene, so that no unspliced transcript would be produced.

Data presented in this study indicate that the regulation mediated by the *LSM7* intron does not completely repress transcription of the co-regulated gene set, but rather reduces expression of these genes by a factor of around three. Therefore the repressor protein in this model must have a relatively low affinity for the DNA or proteins with which it interacts, in order for some transcription to occur even when the protein is present. Such a protein must also have a relatively short half-life, so that the genes can be efficiently de-repressed if growth conditions change.

This model could be tested by the introduction of a premature stop codon in the putative protein sequence, while leaving the various splicing signals intact. In the eventuality that the intron encodes a different protein from that predicted here, premature stop codons should be introduced in all three ORFs in the intron sequence. Cellular fractions could also be analysed at different time points to determine the stability and subcellular localisation of the intron under conditions known to stimulate repression.

'Cling' Model: The mRNA interacts with the promoter to block transcription

A. Formation of an RNA-DNA triple helix

This model is illustrated in Figure 4.14A and involves the formation of transient triple helices between the RNA from *LSM7* (either unspliced mRNA or just the intron lariat) and DNA within the promoter region. Depending on the binding site for the RNA, this would directly block binding of activator proteins or various components of the transcription machinery to the DNA.

DNA-RNA or DNA-DNA triple helices occur when tracts of either pyrimidine or purine bases on the RNA (or single DNA) strand bind to a purine stretch in the DNA duplex (Letai *et al.*, 1988; Fosella *et al.*, 1993). The third strand lies in the major groove of the DNA duplex, with each base binding to the purines of the duplex by two Hoogsteen hydrogen bonds (Radhakrishnan and Patel, 1994). DNA-DNA triple helices have been observed in human interphase nuclei, where sections of single-stranded DNA transiently interact with polypurine tracts in DNA from other parts of the genome, leaving the remaining single DNA strand DNA free to interact with other factors and leading to a change in the spatial organisation of the nucleus (reviewed in Frank-Kamenetskii and Mirkin 1995). DNA-RNA triple helices have so far only been observed using synthetic oligonucleotides. However, under such conditions triplex formation within the promoters has been shown to block binding by activator proteins and inhibit gene activation (Conney *et al.,* 1988; Noonberg *et al.,* 1994; Kim and Miller, 1995). Further, synthetic RNA strands have been used in mammalian cell culture to repress transcription of a variety of genes (Ing *et al.,* 1993; Kovacs *et al.,* 1996; Bailey and Weeks, 2000).

Interestingly, although it is shorter than the oligonucleotides used in the *in vitro* experiments mentioned above (typically 15 -18bp), a short pyrimidine tract is present in the *LSM7* intron (TTTCCTTTTT). Moreover, this sequence lies within the largest loop region of the secondary structure predicted for unspliced *LSM7* mRNA (Figure 4.10). This model could therefore be tested by mutating this region of the intron to form a more heterogeneous sequence. Pyrimidine-motif triple helices are

thought to be relatively unstable at physiological pH, which would be consistent with the partial repression of gene expression observed in this study (Hoyne *et al.*, 2000). Although this may make detection somewhat difficult, it could be possible to detect the formation of any triple helices on the promoter regions of the co-regulated gene set using a combination of fluorescence *in situ* hybridisation (FISH) and labelled anti-triplex antibodies in a manner similar to that reported by Ohno *et al.*, (2002).

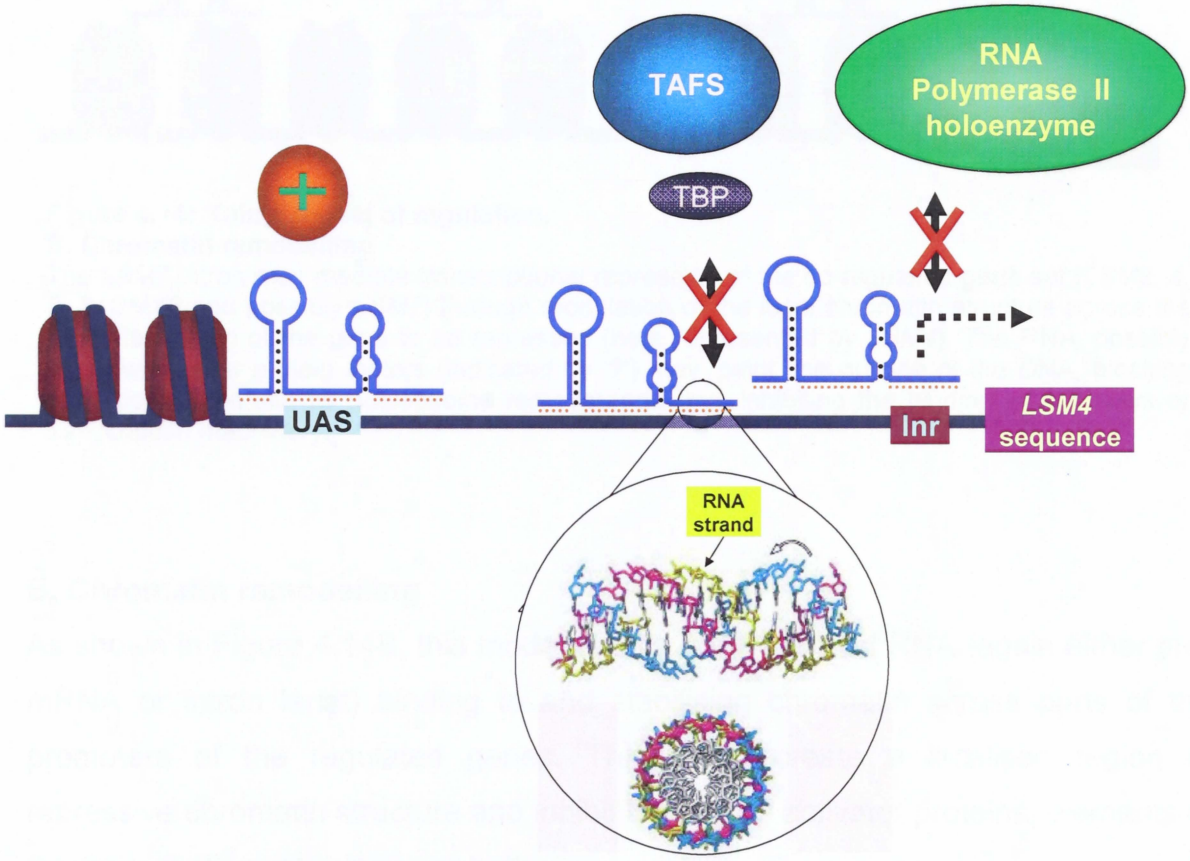


Figure 4.14: 'Cling' model of regulation.
A. Formation of an RNA-DNA triple helix
This depicts another way in which the *LSM7* intron may mediate transcriptional repression of the co-regulated gene set (*LSM2, 4, 5, 8, SMX2* and possibly *LSM7*). The *LSM7* RNA (either in unspliced form or intron lariat) is proposed to form transient triple helixes with duplex DNA at various possible points across the promoter region of the regulated gene (here represented by *LSM4*). These structures may then block the binding of activator proteins or the transcriptional machinery to the DNA, depending on the site of RNA binding. Image of DNA/RNA triple helix adapted from Lodish *et al.*(2003).

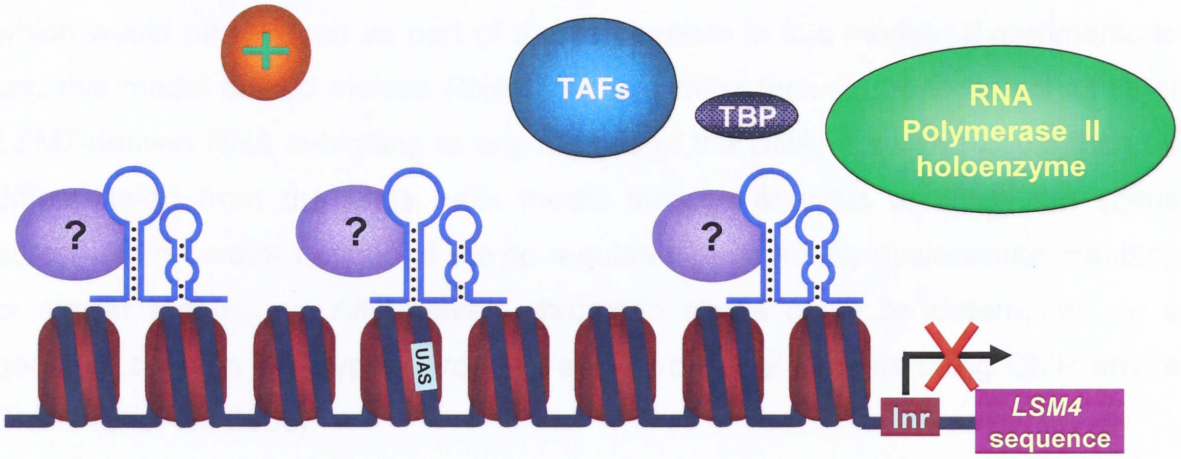


Figure 4.14: 'Cling' model of regulation.
B. Chromatin remodelling
The *LSM7* intron may mediate transcriptional repression of the co-regulated gene set (*LSM2, 4, 5, 8, SMX2* and possibly *LSM7*) through modulation of the local chromatin structure across the promoter region of the gene to be repressed (here represented by *LSM4*). The RNA, possibly along with other protein factors (indicated by '?') may 'paint' the outside of the DNA, blocking activation of the site by nucleosome removal, and thus inhibiting the binding of activators or transcription machinery.

B. Chromatin remodelling

As shown in Figure 4.14B, this model entails *LSM7*-derived RNA (again either pre-mRNA or intron lariat) binding to and stabilising chromatin across parts of the promoters of the regulated genes. This would create a localised region of repressive chromatin structure and inhibit binding by activator proteins, elements of the transcription machinery, or both.

Chromatin stabilisation may be achieved through the formation of a ribonucleoprotein complex with auxiliary chromatin remodelling enzymes, as occurs in dosage compensation of the X chromosome in mammals and *Drosophila* (reviewed in Kelly and Kurdora, 2000). In these cases a non-coding RNA is expressed which recruits a protein complex and binds to chromatin near the site of synthesis, eventually spreading in *cis* to coat the entire chromosome and

eventually renders it transcriptionally inactive. The end result of this process in the fully developed animal is permanent silencing of the X chromosome from which the non-coding RNA was transcribed. However early chromatin remodelling events appear to be relatively unstable (Silva *et al.,* 2003; Plath *et al.,* 2003), a feature which would be required as part of the mechanism in this model. Experiments to test this model should include RNA FISH or similar techniques to detect whether *LSM7*-derived RNA is binding to any regions of the DNA. If so, this could then be differentiated from the triple helix model through analysis of chromatin status across the promoter regions of the co-regulated gene set by nucleosome mapping or similar techniques. Alternatively chromatin status could be determined on a genomic scale in wild type, *intronless* and *intron only* mutants using ChIP arrays (Ballestar *et al.,* 2003).

'Distract' model: The RNA competes with the UAS for binding to the activator protein

In this model part of the secondary structure of *LSM7*-derived RNA binds in a sequence-specific manner to the activator protein, reducing the amount available to bind to the UAS of the regulated gene (Figure 4.15). As in the previous two models, the RNA structure involved could either be made up of the entire unspliced *LSM7* mRNA, or merely the intron lariat. Competition by this structure with the UAS would be most likely result in a relatively mild reduction in expression, such as that observed in this study.

In at least one instance a protein has been identified which binds specifically to the secondary structure of an unspliced mRNA (Vilardell and Warner, 1994; Li *et al.,* 1995). In addition, transcription factors have been identified which specifically bind RNA or RNA-DNA hybrids (Shi and Berg, 1995; Ladomery, 1997; Lie and Macdonald, 1999). To verify this model the RNA secondary structure could be disrupted by swapping regions of the intron around. Alternatively, large amounts of unspliced *LSM7* mRNA or intron lariats could be produced, and used in pull-down assays, two-hybrid analysis, or protein arrays under conditions where the

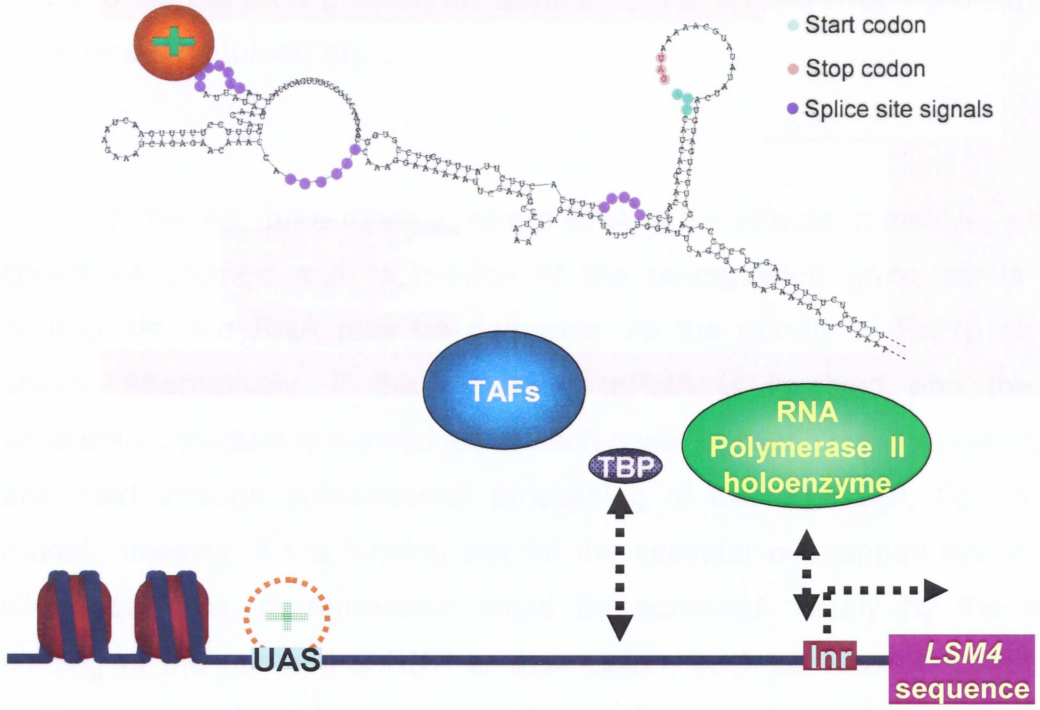


Figure 4.15: 'Distract' model of regulation.
LSM7 intron-mediated repression of transcriptional initiation of the co-regulated gene set (*LSM2, 4, 5, 8, SMX2* and possibly *LSM7*) may also be achieved by the *LSM7* RNA competing with the UAS of regulated genes (*LSM4* in this case) for binding with the activator protein. If the TF binding site on the folded RNA molecule overlapped one of the splice site sequences, this repression could be relieved through subsequent binding by components of the spliceosome (such as the U1 or U2 snRNP), displacing the activator and freeing it to bind to the UAS sequence in the DNA. The unspliced *LSM7* mRNA secondary structure fragment shown was adapted from output generated by RNAfold software (Hofacker *et al.*, 2003). This structure does not take into account any contribution from 5' UTR sequence. Start and stop codons are shown in green and red, respectively. Splice site signals (5' splice site, branchpoint and 3' splice site) are indicated in purple. The mechanism of repression hypothesised in this model could also be mediated by the *LSM7* intron lariat alone.

secondary structure of the RNA remains intact, to identify any binding by specific nuclear proteins. The yeast 3-hybrid system may be also used to test for interactions between the *LSM7* mRNA structure and the hypothesised activator protein. However, in order to use this technique the activator protein must first be identified by other means, as the three-hybrid system would require that this protein is hybridised with the transcription activation domain of a Gal4 transcription

factor or another similar element. The other protein in this system would consist of separate DNA- and RNA-binding domains. Thus when the *LSM7* mRNA structure binds to the activator protein, (in addition to the second hybrid protein) a reporter gene would be turned on.

In the last three models, where RNA is the effector molecule, when growth conditions change and repression of the co-regulated gene set is no longer appropriate, the RNA may be degraded via the activity of Rnt1p as described above. Alternatively, if the unspliced mRNA is involved and the predicted secondary structure is maintained in each case, transcriptional repression could be alleviated through spliceosomal processing of the transcript. For the 'Distract' model, however, if the binding site for the activator overlapped one of the splice site sequences, de-repression could be achieved merely by the high-affinity binding of the relevant snRNP to that region, such as when U1 binds to the 5' splice site, or U2 binds to the branchpoint. This would displace the activator from the RNA structure, freeing it to bind once again to the UAS and stimulate transcription of the co-regulated gene set.

Many other experiments must still be carried out before the mechanism of regulation by the *LSM7*intron can be fully elucidated. These include:

- Transforming the *intronless* mutant with a plasmid containing the *intron only* construct and some *LSM7* promoter sequence, to verify that gene responsiveness is restored.
- Over-expression of the unspliced *LSM7* transcript, or the *LSM7* intron alone.
- Mutate the TACTAAC box to test if splicing or recruitment of splicing factors is required.
- If so, insert the *LSM7* intron into another gene in the *lsm7*Δ mutant to explore the role of flanking DNA sequence in this regulation.
- Replace the *LSM7* intron with another functional intron (such as that of *LSM2* or *SMD2,* or an intron from an unrelated gene) to confirm the *LSM7*-specific intron effect, and test for possible redundancy

- Sequential deletion or window deletions of the promoters of at least one of *LSM2, 4, 5, 8* and *SMX2*, to identify regulatory sequences required for their co-ordinate control under different growth conditions.

Both *LSM7* and the other genes it regulates code for proteins which form part of the core of the spliceosome. The effects of altering the expression of these genes on splicing, as well as the wider implications for the cell as a whole, are explored in Chapter 5.

Chapter 5: Effect of *LSM* gene regulation on splicing and cellular physiology

5.1 Introduction and Aim

Lsm proteins form part of an ancient family stretching back into Archaea (Collins *et al.,* 2001; Moller *et al.,* 2002; Toro *et al.,* 2001). In the eukaryotic cell they have been implicated in a number of key RNA modification complexes. Principle among these is their role in the spliceosome, as discussed in Section 1.5. Spliceosomal Lsm proteins bind specifically to a region of sequence at the 3' terminus of U6snRNA (Salgado-Garrido *et al.,* 1999), stabilising it and allowing the assembly of a functional U6snRNP (Mayes *et al.,* 1999). During splicing the binding of the U4/U6 heterodimer must be disrupted to allow the U6snRNA to form part of the catalytic centre of the spliceosome (Stanley and Guthrie, 1998). Spliceosomal Lsm proteins have been found to affect this U4/U6 base pairing (Achsel *et al.*1999, Verdone *et al.,* 2004). In addition, cells with these Lsm protein genes deleted are incapable of recycling the spliceosome for subsequent rounds of splicing (Verdone *et al.,* 2004).

In previous chapters the majority of genes encoding spliceosomal Lsm proteins were shown to be co-expressed under different environmental conditions. Further, these changes in expression were regulated by the *LSM7* intron. Accordingly, the work in this chapter set out to determine the effect of this regulation on splicing in the cell. The possibility of autoregulation via the splicing of *LSM7* was also investigated. Abundance of transcripts requiring splicing was measured and found to be lower under conditions where splicing is impaired. Finally the physiological impact of altering the abundance of spliceosomal *LSM* transcripts was explored.

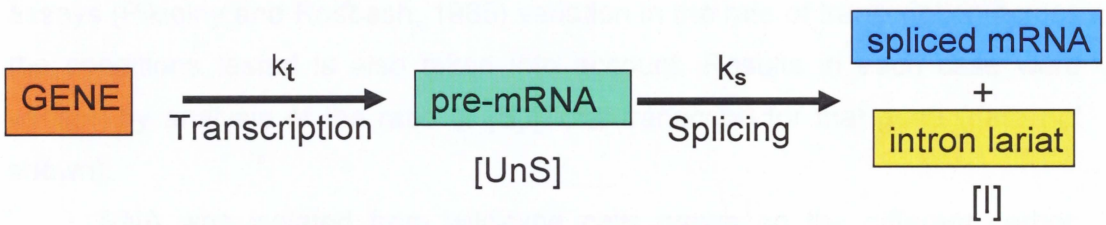
5.1.1 Acknowledgements

Raw data from the microarray analysis of gene expression in wild-type cells grown on glucose or lactate was obtained from Cristy Gelling (UNSW). Normalised data from the microarray analysis of splicing in *lsm1*Δ, *lsm6*Δ and *lsm7*Δ compared to wild type in cells grown on glucose was kindly provided by

the laboratory of Manny Ares (University of California Santa Cruz). Processed data from the microarray analysis of cells growing under aerobic or anaerobic conditions is presented courtesy of Anthony Beckhouse (UNSW). The three-dimensional bubble graph (Figure 5.11) was generated in collaboration with Geoffrey Kornfeld (UNSW).

5.2 Splicing capacity changes following growth on different carbon sources

In order to address whether changes in transcript levels of the co-regulated gene set affects the ability of the cell to splice pre-mRNA, a sensitive *in vivo* splicing assay was developed. This employed an intron-containing gene (*RPL25*) in *S. cerevisiae* that is expressed at a high level and is present in only one copy in the genome. In addition, *RPL25* has all of the intron features shared by the majority of genes with introns in this organism, such as size, location in the gene and conserved splice site sequences and position (Spingola *et al.*, 1999). Real time PCR was used to accurately quantify levels of unspliced and total transcript using primer pairs with either one primer within the intron region, or both primers in the second exon of the gene, as shown in Figure 5.1A. Abundance of spliced transcript was similarly measured using a primer pair with one primer spanning the exon junction.



Where k_t = the rate of transcription of a given gene

k_s= rate of splicing of that gene,

[UnS] = concentration of the unspliced pre-mRNA

[S] = concentration of spliced transcript

and [I] concentration of the intron lariat

then $$[UnS] = \frac{k_t}{k_s}$$

(for mathematical proof see Appendix)

Therefore, if the rate of transcription is proportional to the total amount of transcript produced to time t, then

$$k_s \propto \frac{[Total]}{[UnS]}$$

Where [Total] = [UnS] + [S] or [UnS] + [I],

For this assay k_s was used as a measure of the cells capacity to splice under a given set of conditions. Since [UnS] may be extremely small under some circumstances, changes in splicing capacity in cells grown under different conditions were expressed as the amount of unspliced transcript ([UnS]) relative to the total amount of transcript for that gene ([Total]).

This assumes that none of these elements are being degraded. This model additionally assumes that [UnS] does not affect the rate of transcription and neither [S] nor [I] affect the rate of splicing. By studying changes in the [UnS]/[Total transcript] ratio instead of the [UnS]/[S] ratio used in classic splicing assays (Pikielny and Rosbash, 1985) variation in the rate of transcription across the conditions tested is also taken into account. Results in each case were verified by analysis of the ratio of [S]/[Total transcript] for that gene (data not shown).

RNA was isolated from wild-type cells grown on the different carbon sources as described previously (Section 3.3) and levels of unspliced, spliced and total *RPL25* transcript determined. Splicing capacity in the wild-type cell was found to be lower in poorer carbon sources (around 2% of *RPL25* transcript was unspliced) than in cells grown on glucose (less than 0.1% of *RPL25* mRNA was unspliced) (Figure 5.1B).

Changes in splicing capacity correlated in an exponential manner with the pattern of expression of *LSM2, 4, 5,* 7 and *8* under the same conditions (represented by *LSM7* transcript in Figure 5.1C, $R^2 = 0.99$). However it is

unclear if there is a direct functional link between these two events or if this is the result of separate, parallel regulatory mechanisms.

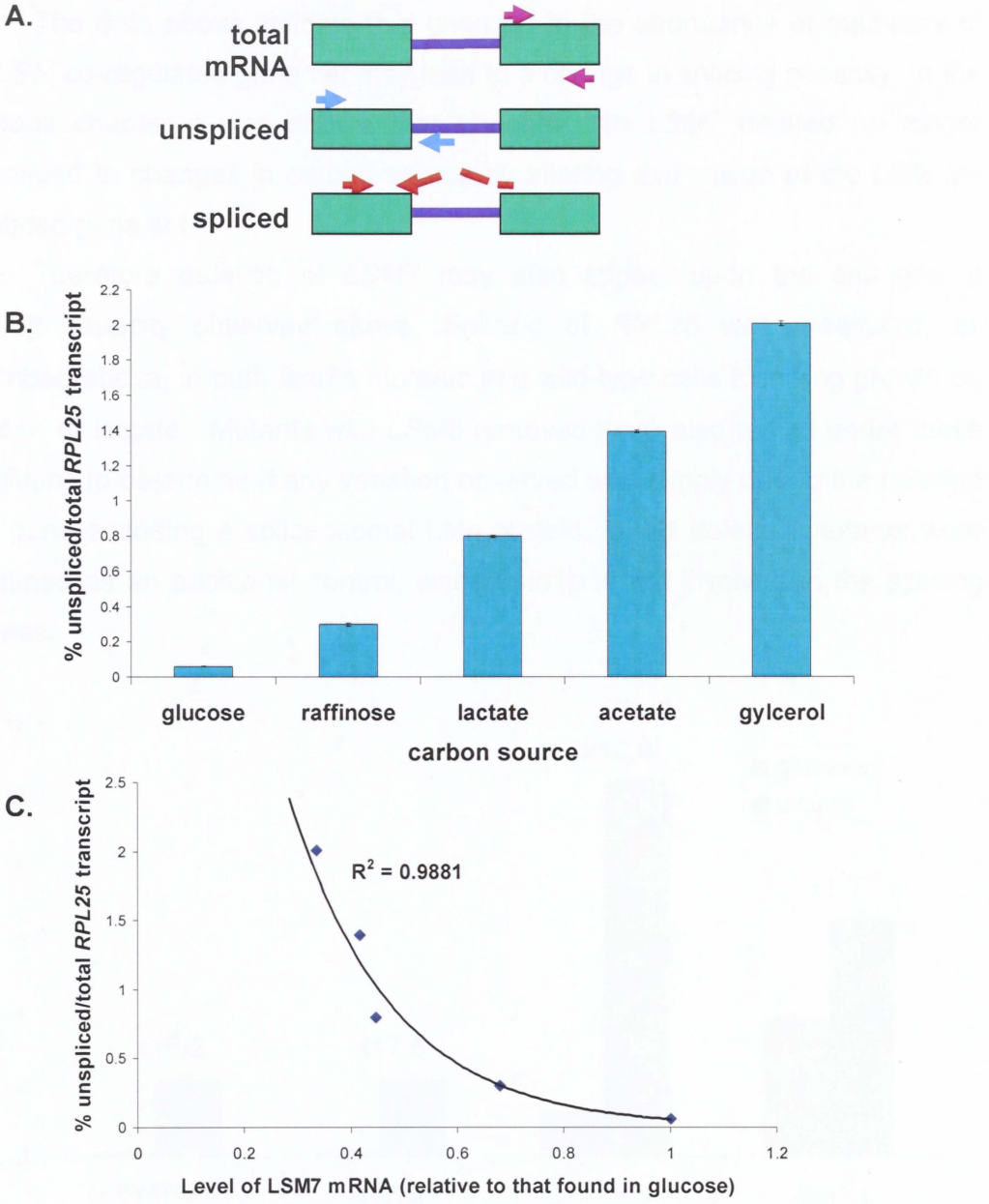


Figure 5.1: Changes in Splicing Capacity of the Cell. A. Schematic depicting the design of the *in vivo* splicing assay. Levels of total RNA were measured using primers specific to sequences in the second exon alone. Unspliced transcript was measured using one primer specific to the first exon and one to a region within the intron. Spliced transcript was measured by using a primer specific to the first exon and a second primer which spans the exon junction. **B.** Effect of growth on different carbon sources on cellular splicing capacity. Wild-type cells were grown on rich media with a variety of different carbon sources; glucose, raffinose, lactate, acetate or glycerol. Levels of unspliced *RPL25* mRNA and total *RPL25* mRNA were measured in each case. Error bars represent the standard deviation of the mean transcript level in each case, obtained from at least two separate assays on samples from three independent experiments. **C.** Correlation of *LSM7* expression and changes in splicing capacity.

5.3 LSM7 is required for the change in splicing capacity in cells grown on different carbon sources

The data above indicate that changes in the abundance of members of the *LSM* co-regulated gene set may lead to a change in splicing capacity. In the previous chapter it was shown that mutants with *LSM7* deleted no longer responded to changes in carbon source by altering expression of the *LSM* co-regulated gene set.

Therefore deletion of *LSM7* may also impact upon the changes in splicing capacity observed above. Splicing of *RPL25* was measured, as described above, in both *lsm7Δ* mutants and wild-type cells following growth on glucose or lactate. Mutants with *LSM6* removed were also tested under these conditions to determine if any variation observed was simply due to the deletion of a gene encoding a spliceosomal Lsm protein. *LSM1* deletion mutants were examined as an additional control, since Lsm1p is not involved in the splicing process.

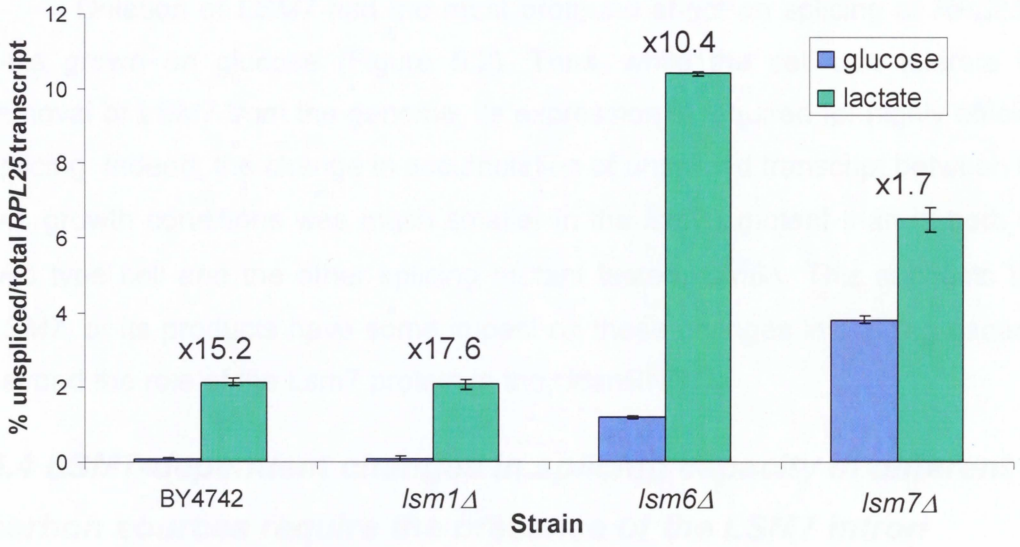


Figure 5.2: Splicing of *RPL25* in mutants with *LSM1, LSM6* or *LSM7* deleted. Wild-type cells (BY4742) or *lsm1Δ*, *lsm6Δ*, or *lsm7Δ* mutants were grown on rich media containing glucose or lactate. Levels of unspliced *RPL25* and total *RPL25* were determined as described previously. Error bars represent the standard deviation of the mean transcript level in each case, obtained from at least two separate assays on samples from three independent experiments. Fold change in splicing capacity between the growth conditions is shown above the column for each strain

As expected, changes in splicing capacity for *RPL25* in both the wild type and *lsm1*Δ strains across the different media were equal to that observed previously for wild-type cells, within error (Figure 5.1B and Figure 5.2).

Deletion of *LSM6* resulted in lower splicing capacity than in wild-type cells when grown on glucose (Figure 5.2) indicating the impact of removing an element of the spliceosomal Lsm complex on splicing. This is in accordance with reported results from other researchers based on *in vitro* splicing assays (Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999; Verdone *et al.,* 2004). Furthermore, relative abundance of unspliced *RPL25* in *lsm6*Δ mutants only increased around 10-fold in cells grown on lactate as compared that obtained after growth on glucose. Therefore deletion of *LSM6* may impact on the change in splicing capacity across the two conditions. Alternatively it is possible that unspliced transcript may be degraded at a higher rate if present above a particular threshold level. This would result in a lower apparent accumulation of unspliced *RPL25* transcript in *lsm6*Δ in lactate than present due to changes in splicing capacity alone.

Deletion of *LSM7* had the most profound effect on splicing of *RPL25* in cells grown on glucose (Figure 5.2). Thus, while the cell can tolerate the removal of *LSM7* from the genome, its expression is required for highly efficient splicing. Indeed, the change in accumulation of unspliced transcript between the two growth conditions was much smaller in the *lsm7*Δ mutant than in both the wild type cell and the other splicing mutant tested, *lsm6*Δ. This suggests that *LSM7*, or its products have some impact on these changes in splicing capacity beyond the role of the Lsm7 protein in the U6snRNP.

5.4 LSM7-dependent changes in splicing capacity in different carbon sources require the presence of the LSM7 intron

To identify which *LSM7*-specific element is involved in regulating splicing capacity under these conditions, the splicing capacities of mutants with various regions of *LSM7* deleted (described in Chapter 4) were also determined. Each mutant was grown in media containing either glucose or lactate as the carbon source and levels of unspliced and total *RPL25* mRNA determined by real time PCR.

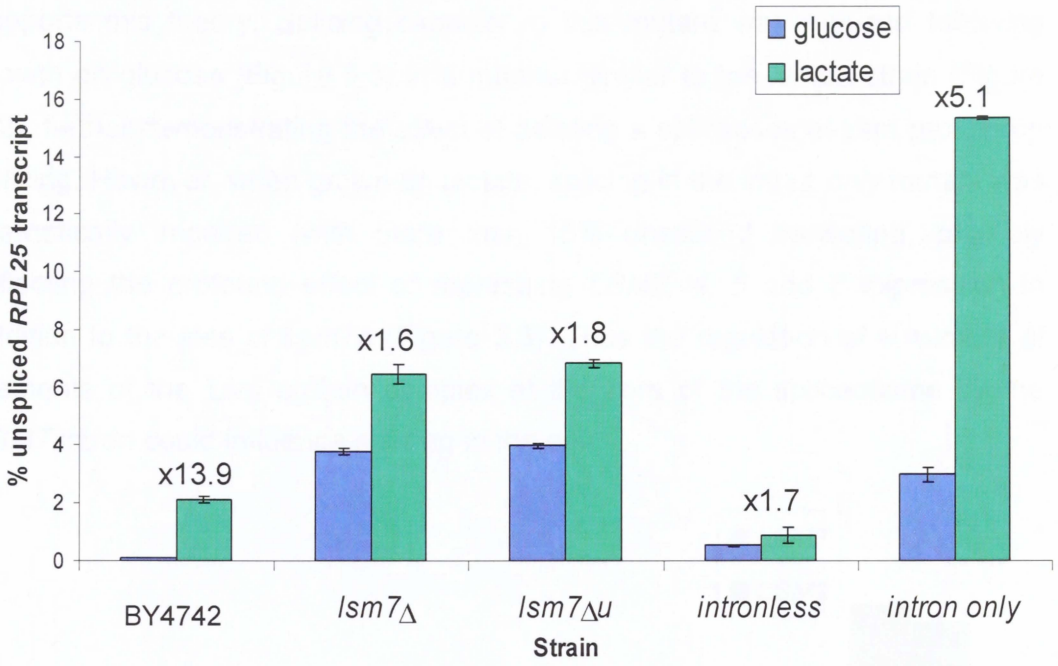


Figure 5.3: Splicing of *RPL25* in mutants with regions of *LSM7* deleted. Wild-type cells (BY4742), *lsm7Δ*, *lsm7Δu*, *intronless* or *intron only* mutants were grown on rich media containing glucose or lactate. Levels of unspliced *RPL25* mRNA and total *RPL25* mRNA were measured and expressed as described previously. The numbers above the data presented for each strain indicate the fold change in results for that strain across the two conditions examined.

As in the previous experiment, mutants with the entire *LSM7* gene deleted (*lsm7Δ* and *lsm7Δu*) had a substantially smaller change in splicing capability across the different media compared to wild-type cells (less than 2-fold for *lsm7Δ* and *lsm7Δu* mutants compared to 14-fold in the wild type, (Figure 5.3). The *intronless* mutant, which produces a functional Lsm7 protein, was also unresponsive to changes in carbon source. Thus it is unlikely that the Lsm7 protein plays a significant role in this regulation (Figure 5.3). Conversely, the *intron only* mutant exhibited a much greater change in splicing capacity under the different conditions, although less than the wild type, possibly due to the absence of the *LSM7* protein (Figure 5.3). Therefore the *LSM7* intron is required to regulate splicing capacity in response to different carbon sources.

This regulation may be achieved via the repression of *LSM2, 4, 5* and *8* expression observed in this study, particularly in cells grown on poorer carbon sources, leading to a reduced number of functional spliceosomal Lsm protein complexes. The behaviour of the *intron only* mutant, with the Lsm7p deleted,

supports this theory. Splicing capacity in this mutant was reduced following growth on glucose (Figure 5.3) in a manner similar to the *lsm6Δ* strain (Figure 5.2), further demonstrating the effect of deleting a spliceosomal Lsm protein on splicing. However, when grown on lactate, splicing in the *intron only* mutant was dramatically impaired (with more than 15% unspliced transcript), possibly reflecting the profound effect of repressing *LSM2, 4, 5* and *8* expression in addition to the loss of Lsm7p (Figure 5.3). Thus the regulation of a number of elements of the Lsm protein complex at the core of the spliceosome by the *LSM7* intron could influence splicing in the cell.

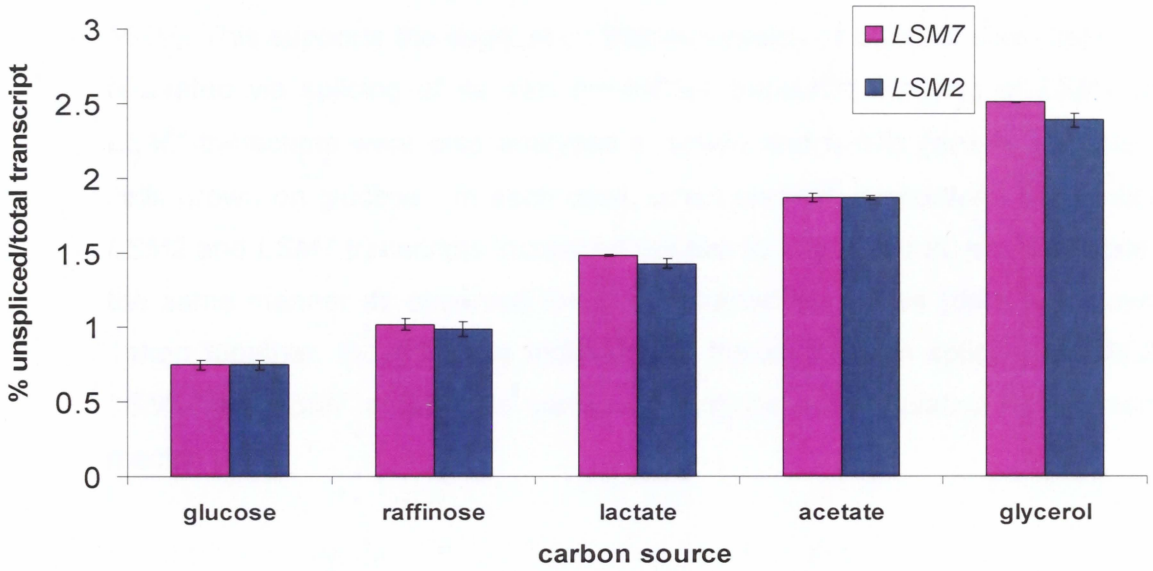


Figure 5.4: Autoregulation of *LSM2* and *LSM7* transcripts. Wild-type cells were grown on rich media with a variety of different carbon sources; glucose, raffinose, lactate, acetate or glycerol. Levels of unspliced mRNA and total mRNA were measured for both *LSM2* and *LSM7*. Error bars represent the standard deviation of the mean transcript level in each case, obtained from at least two separate assays on samples from three independent experiments.

5.5 Are LSM7 and LSM2 auto-regulated at the level of splicing?

Both *LSM7* and *LSM2* genes contain introns themselves which share all general features (size, position, etc.) with *RPL25* (Grate and Ares, 2002). Therefore it is possible that expression of these genes may be additionally auto-regulated at the level of splicing. To explore this possibility splicing assays were developed to measure levels of unspliced, spliced and total *LSM7* or *LSM2* mRNA as described for *RPL25* previously (Figure 5.1A). RNA from wild-type

cells grown on rich media containing glucose, raffinose, lactate, acetate or glycerol were harvested and analysed by real time PCR.

As observed with *RPL25* mRNA above, the capacity of the cell to splice *LSM2* and *LSM7* pre-mRNA transcripts decreased in cells grown on poorer carbon sources, with around 2.5% of transcript being unspliced in each case (Figure 5.4). Furthermore, removal of the Lsm7 protein led to impaired splicing (Figure 5.2 and 5.3). Therefore it is likely that expression of the *LSM7* gene is additionally auto-regulated at the level of splicing. *LSM2* is essential for cell viability and cells depleted of Lsm2 through use of temperature-sensitive mutants exhibit a splicing defect (Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999). This supports the suggestion that expression of *LSM2* is also likely to be regulated via splicing of its own pre-mRNA transcript. Splicing of *LSM2* and *LSM7* transcripts were also analysed in *lsm6*Δ and *lsm7*Δ (where possible) in cells grown on glucose. In each case, when present, proportions of unspliced *LSM2* and *LSM7* transcripts increased relative to that found in wild-type cells in the same manner as observed for *RPL25* transcripts above (data not shown). Taken together, these results indicate that the changes in splicing of *RPL25*, *LSM2* and *LSM7* across the various conditions are regulated by the same mechanism.

5.6 Most genes containing introns in S. cerevisiae fall into a relatively small number of functional groups

As described in Chapter 1, less than 4% of nuclear genes in *S. cerevisiae* contain introns (Spingola *et al.* 1999). Moreover, the majority of introns that are present share a number of features, including position (predominantly near the 5' end of the gene), length (most are around 300 nt in length) and conservation of splice site sequence (Bon *et al,* 2003), leading some investigators to suggest that these introns have been retained in this species to fulfil some functional role in the cell (*Lander* et al., 2001; Bon, *et al.,* 2003). To further explore potential functions for introns in this organism, genes with introns were classified into groups based on the function or functions of the gene product. A complete list of genes with introns in *S. cerevisiae* was

obtained from the Ares lab Yeast Intron Database (Version 3; Grate and Ares, 2002). These introns were then annotated based on Gene Ontology (GO) Process data to identify if they are over-represented in any particular biological process in the cell (Ashburner *et al.,* 2000). It should be noted that one gene product may be associated with a number of different cellular processes.

The majority of introns in *S. cerevisiae* were found to fall into relatively few broad functional groups: ribosomal subunits, components and assembly (98 genes); protein transport and secretion (57 genes); cytoplasm organisation and biogenesis (46 genes); cell cycle (24 genes); regulation of translational initiation and fidelity (16 genes); ubiquitin-dependent protein catabolism (10 genes); mRNA processing (9 genes); regulation of transcription from a polymerase III promoter (7 genes); glycogen metabolism (3 genes) and mitochondrial electron transport (2 genes). Interestingly, more than half of the genes with introns in this organism are involved in protein metabolism (142 out of 238 genes with introns). As some of these groups contain a large number of genes, not all of these classifications were found to be statistically significant. However, genes with introns were significantly enriched in subsets of each of these groups. Gene lists and probability data for each of these classifications are shown in the Appendix.

The majority of genes with introns in *S. cerevisiae* are subunits of, or involved in, the assembly and maintenance of the ribosome, such as *RPS11B* or *RPL6B* (Figure 5.5B). Relative to the number of genes with introns in this species, only a small proportion fall into each of the other remaining functional categories identified by GO analysis (Figure 5.5B).

Genes with introns involved in ribosomal subunit composition, assembly and maintenance make up less than 10% of the total number of genes present in this category (Figure 5.5A). However the majority of genes involved in the assembly of the ribosome, particularly of the small ribosomal subunit (8 out of 13 possible genes), were found to have introns (Figure 5.5A). Interestingly, regulation of translational fidelity was also highly enriched for genes containing introns, such as *RPS23B* (7 of 8 possible genes, Figure 5.5A).

A number of small functional groups involved in the cell cycle were also enriched for intron-containing genes. Around 10% of genes containing introns

are involved in the M phase of the cell cycle (Figure 5.5B). These genes make up 21 of a possible 205 genes in this category (Figure 5.5A). More significantly, two out of three possible tubulin genes involved in homologous chromosome segregation (*TUB1* and *TUB3*) also have retained introns (Figure 5.5A).

Two processes involved in protein degradation are significantly enriched for genes with introns. Out of six possible genes involved in protein neddylation, three contain introns (DCN1, RUB1, UBC12), at least two of which physically interact (Liakopoulos *et al.,* 1998). Less than 10% of genes involved in cytoplasm organisation and biogenesis have introns (Figure 5.5B). However genes with introns are overrepresented in actin-related processes, particularly in the polymerisation or depolymerisation of actin (2 out of a possible 7 genes, Figure 5.5A).

Interestingly, of the 238 genes with introns in this organism, 37 genes have nucleic acid-binding properties. Moreover, a number of genes with introns are involved in transcription polymerase II and polymerase III; three out of a possible eight genes involved in transcriptional regulation of RNA polymerase III genes have introns (Figure 5.5A). However further information is required regarding the exact nature and extent of influence of each of these DNA-binding elements to clarify the role of intron-containing gene products in these processes.

5.7 Splicing capacity varies in a manner similar to the abundance of transcripts requiring splicing: Supply and Demand

This study has shown that the capacity of the cell to splice some pre-mRNA transcripts may vary under different conditions. In each case examined, these pre-mRNA transcripts contained intron sequences with features shared by the majority of genes with introns in this organism. Therefore it is possible that the same mechanism that regulates the splicing of pre-mRNA transcripts from these genes may also affect the splicing of transcripts from a number of

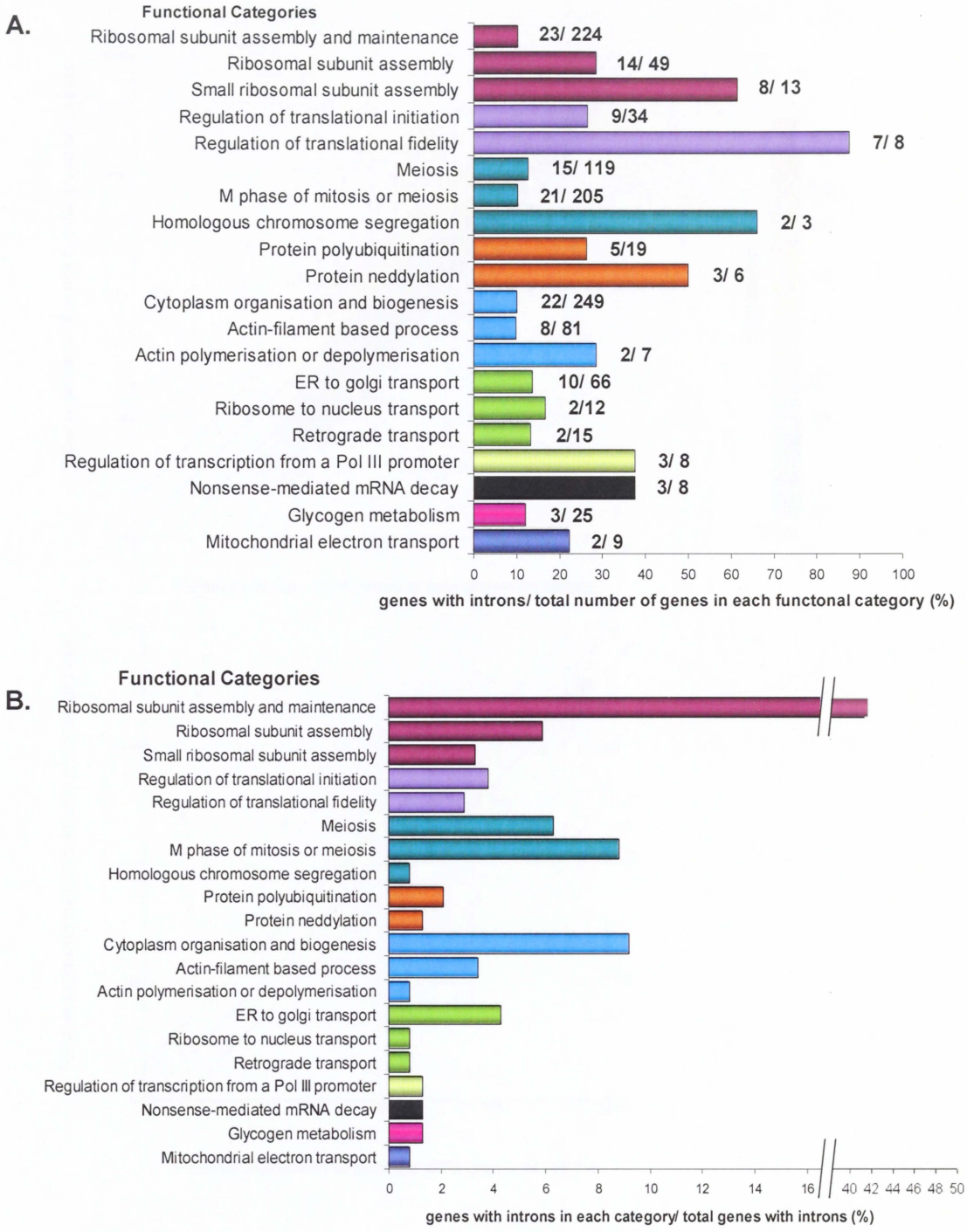


Figure 5.5: Functional classifications of genes with introns in *S. cerevisiae*. A. The proportion of genes with introns in each functional category, relative to the total number of genes in that category in *S. cerevisiae*. Each result is shown both as a bar on the graph and as a numerical ratio. **B.** The proportion of genes with introns in each category relative to the total number of genes with introns in *S. cerevisiae*. In both cases, the numbers of genes in the genome in each category was obtained from GO ontologies (http://www.geneontology.org/). The total number of introns was obtained from the Ares Intron database (Grate and Ares,2002).

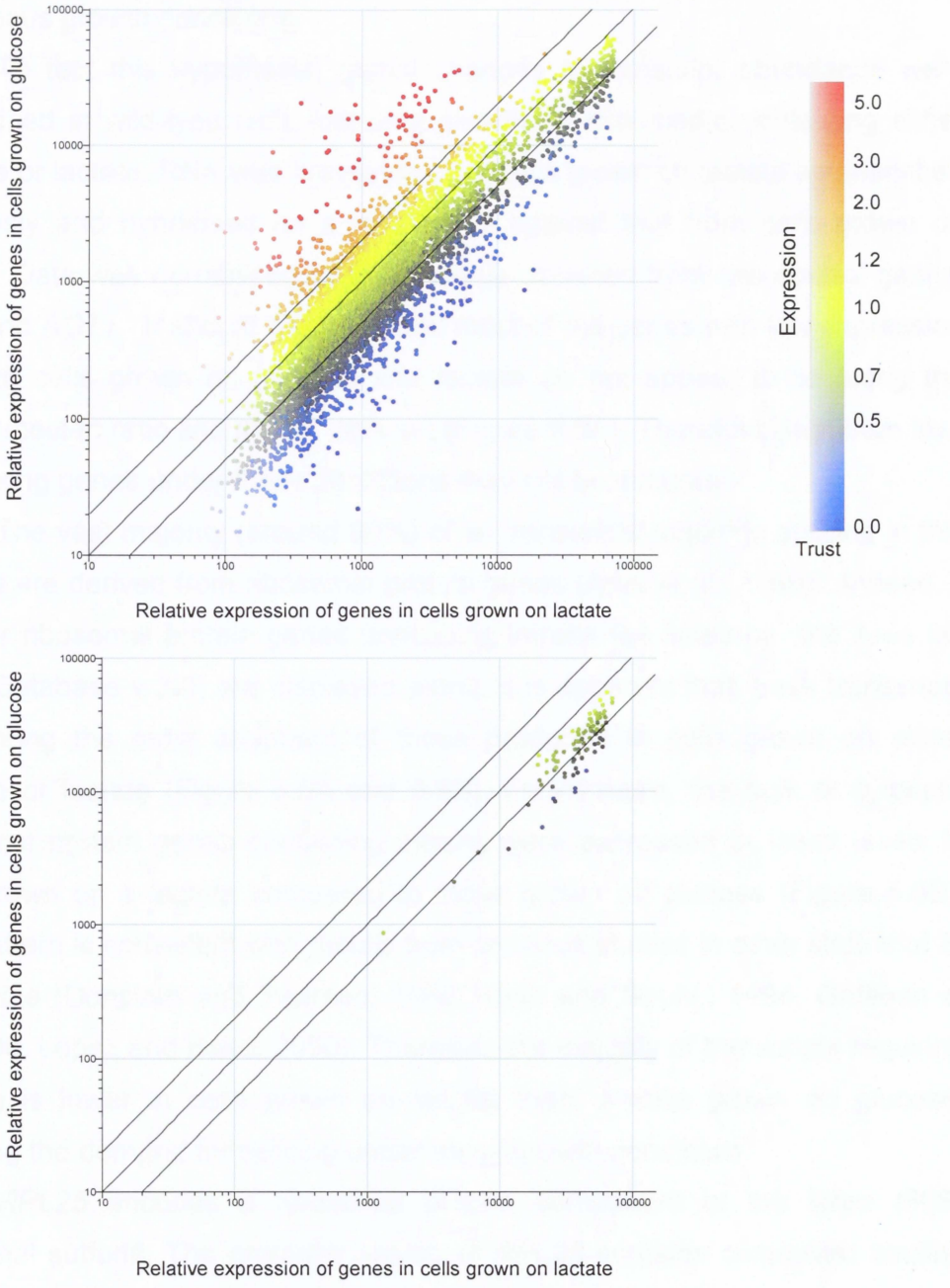


Figure 5.6: Global change in expression of intron-containing genes in cells grown on glucose or lactate. Wild-type cells were grown on rich media in the presence of either glucose (y axis) or lactate (x axis). A. Scatter plot of relative expression of all genes on the array under each growth condition. B. Only results from ribosomal protein genes which contain introns. In each scatter plot the dark central line indicates the line of equivalence and the two light lines around it show a two-fold change in expression. Graphs generated using Genespring GX.

other genes. Such regulation may reflect differential requirements for splicing in the various growth conditions.

To test this hypothesis, global changes in transcript abundance were determined in wild-type cells following growth on rich media containing either glucose or lactate. RNA was harvested from cells grown on lactate as described previously and hybridised on a microarray against that from cells grown on lactate. Data was normalised relative to data obtained from two control genes: *SIR3* and *ACT1*. It should be noted that most of the genes with low expression levels in cells grown on glucose and lactate do not appear to lie along the central median ratio line for the data set (Figure 5.6A). Therefore data from low-expressing genes under these conditions may not be accurate.

The vast majority (around 90%) of all transcripts requiring splicing in this species are derived from ribosomal protein genes (Ares *et al.,* 1999). Indeed, if data for ribosomal protein genes containing introns (as listed by the Ares lab Intron Database v 3.0) are displayed alone, it is apparent that these transcripts are among the most abundant of those produced in cells grown on either glucose or lactate (Figure 5.5A and 5.5B). Furthermore, the bulk of cytosolic ribosomal protein genes containing introns were expressed at lower levels in cells grown on a lactate compared to those grown on glucose (Figure 5.6B). This pattern is consistent with results from previous studies in other strains of *S. cerevisiae* (Donovan and Pearson, 1986; Klein and Struhl., 1994, Griffioen *et al.,* 1996, Lopez and Baker 2000). Therefore the majority of transcripts requiring splicing is lower in cells grown on lactate than in cells grown on glucose, reducing the demand for splicing under these growth conditions.

RPL25 encodes a ribosomal protein component of the large (60S) ribosomal subunit. The promoter region of *RPL25* contains conserved binding sites for a number of different transcription elements known to coordinate expression of most ribosomal protein genes under different environmental conditions (Woudt *et al.,* 1986; Moehle and Hinnebusch, 1991). Expression of *RPL25* was therefore measured by real time PCR in cells grown on the different carbon sources, to verify its use in future assays of this type to represent the transcriptional response of the majority of ribosomal protein genes under

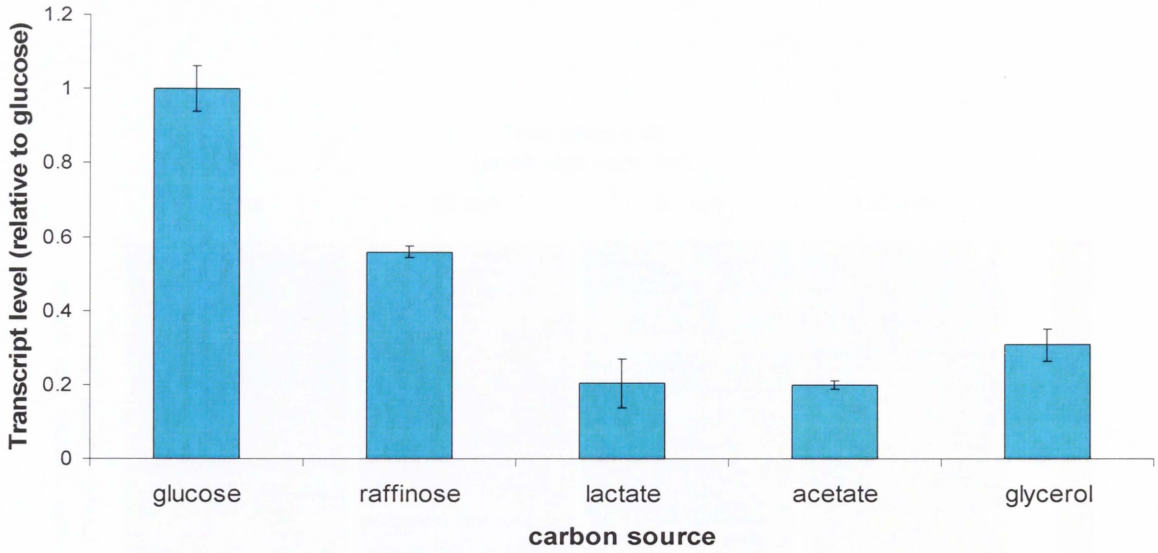


Figure 5.7: *RPL25* expression in wild-type cells grown on different carbon sources. Wild-type cells were grown on rich media in the presence of glucose. Levels of the ribosomal protein gene *RPL25* were determined in each condition relative to *SIR3* as an external control. Error bars represent the standard deviation of the mean transcript level in each case, obtained from at least two separate assays on samples from three independent experiments.

various conditions. As observed in the microarray data, transcript levels of *RPL25* measured by real time PCR were lower in cells grown on lactate than in cells grown on glucose (Figure 5.7). Indeed the general pattern of expression of *RPL25* across the different carbon sources corresponded with that of the co-regulated *LSM* gene set across the same conditions: decreasing expression in progressively poorer carbon sources (Figure 3.3A and 5.7). However the decrease is more rapid for *RPL25* than for the *LSM* genes.

The capacity of the cell to splice *RPL25* and the abundance of intron-containing transcripts were also measured in cells grown in aerobic or anaerobic conditions. Wild-type cells were grown under anaerobic conditions until late-log phase. These cells were then shifted to aerobic conditions and samples taken at 0, 20, 50 and 120 min after this point. Samples at time 0 and

120 min following the shift to aerobic conditions were selected to test the splicing capacity for *RPL25*. All samples were then hybridised to microarray slides against a time 0 reference sample.

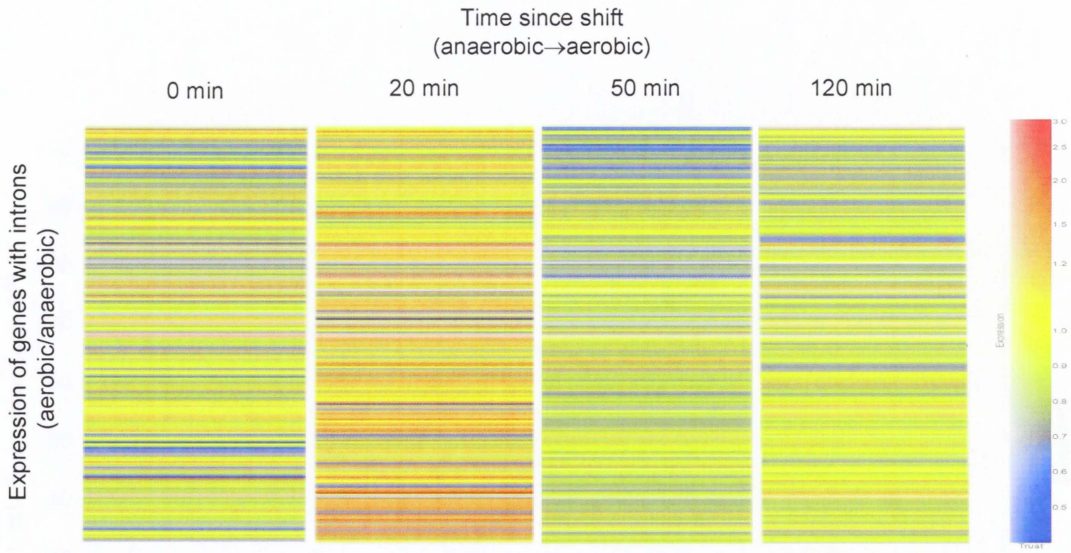


Figure 5.8: Expression of genes with introns in cells grown under aerobic or anaerobic conditions. Wild-type cells were grown under anaerobic conditions to an OD_{600} of 1 and then split into two aliquots, one which was taken as a 'time 0' sample, and one which was switched to aerobic growth for 0, 20, 50 and 120 min. Expression data for all genes with introns in this organism are shown as a ratio of aerobic / anaerobic as indicated by colour. Raw data for this assay can be found in the Appendix. Image generated using Genespring GX

Interestingly, 120 minutes following the shift to anaerobic conditions, around 9% of total *RPL25* transcript was unspliced. This was around 18-fold higher than capacity for *RPL25* found in the time 0 sample (0.5% of *RP25* transcript is unspliced) and more than 4-fold higher than found in wild-type cells in any other condition tested so far. Data from microarray analysis of samples taken 0, 20, 50 and 120 minutes following the shift to aerobic conditions show a complex pattern of response. The majority of genes with introns (defined as for Figure 5.6B) exhibited an increase in expression 20 minutes after a switch to aerobic conditions (Figure 5.8). However by 50 minutes the expression of these genes had again decreased relative to the reference sample (Figure 5.8). Due

to the complexity of this response, it is difficult to determine if changes in the splicing capacity of the cell were in concordance with the abundance of transcript requiring splicing under these conditions. This may be resolved if the splicing capacity was also examined at 20 and 50 minutes after the change in oxygen conditions.

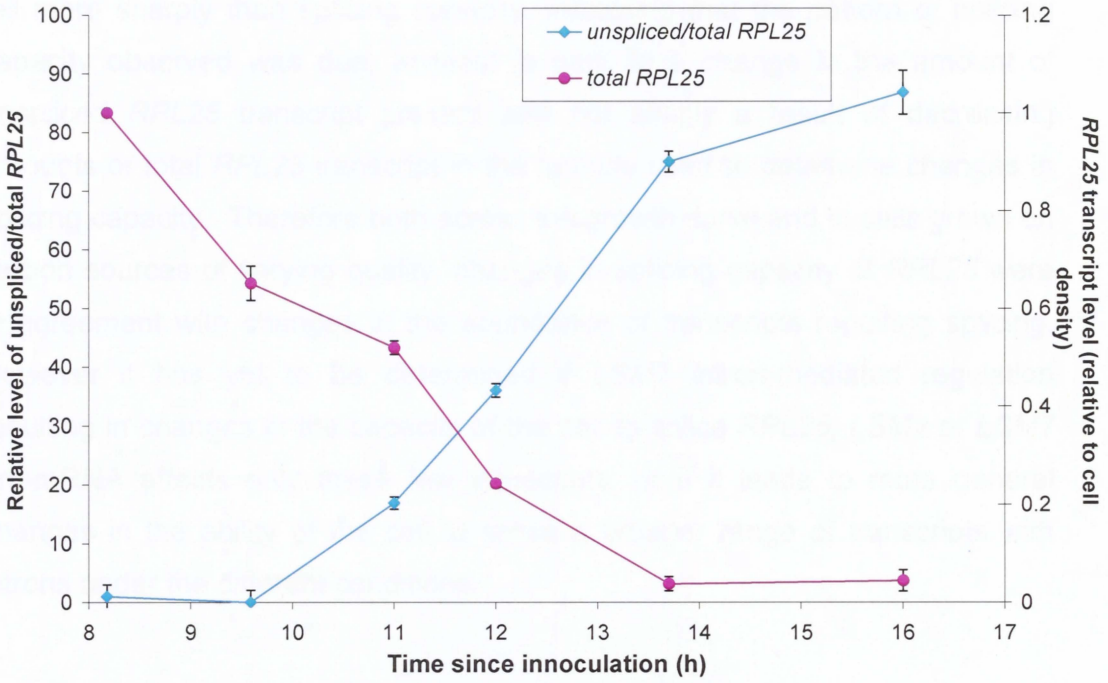


Figure 5.9: Relative expression and splicing of *RPL25* in wild-type cells at different stages of the growth curve. RNA was harvested from BY4742 cells at six different points across the growth curve and analysed by real time PCR for levels of unspliced and total *RPL25*. This data was further used to test changes in splicing capacity, by analysis of the unspliced *RPL25* / total *RPL25* ratios. Total levels of expression of *RPL25* relative to the cell density of the culture at each point is shown on the right hand axis. Error bars represent the standard deviation of the mean transcript level in each case, obtained from at least two separate assays on samples from three independent experiments. Growth data for this experiment can be found in Figure 3.2.

A simpler system was therefore used to test for any correlation between the splicing capacity of the cell and the abundance of transcript requiring splicing. RNA obtained previously from different points throughout growth in batch culture (Section 3.2) was used to measure both the abundance of total *RPL25* transcript as well as the changes in the splicing capacity for *RPL25* transcripts. Interestingly, both abundance of ribosomal protein gene transcripts (represented by *RPL25*) and capacity of the cell to splice *RPL25* decreased from relatively early in growth (Figure 5.9). However levels of *RPL25* expression fell more sharply than splicing capacity, indicating that the pattern of splicing capacity observed was due, at least in part, to a change in the amount of unspliced *RPL25* transcript present and not simply a result of decreasing amounts of total *RPL25* transcript in the formula used to determine changes in splicing capacity. Therefore both across the growth curve and in cells grown on carbon sources of varying quality, changes in splicing capacity of *RPL25* were in agreement with changes in the abundance of transcripts requiring splicing. However it has yet to be determined if *LSM7* intron-mediated regulation resulting in changes in the capacity of the cell to splice *RPL25, LSM2* or *LSM7* pre-mRNA affects only these few transcripts, or if it leads to more general changes in the ability of the cell to splice a broader range of transcripts with introns under the different conditions.

5.8 Deletion of LSM7 affects the splicing of the majority of genes with introns

To explore potential effects of *LSM7*-mediated regulation on global changes in splicing capacity, splicing-sensitive microarrays were used to analyse RNA from mutants with *LSM1, LSM6* or *LSM7* deleted, following the method of Clark *et al.* (2002). Oligonucleotides which bound to the sequence within the intron, across the splice junction or in the second exon of each gene were printed onto glass slides to allow detection of unspliced transcript, spliced transcript and the total of both unspliced and spliced transcripts, respectively (Figure 5.10) Data for each gene was normalized relative to that obtained from the coding regions of seven 'stoic' genes (*SLY1, SEC4, VPS45, TAF145, RSC2*

and *YAP1*) which were identified as having constant levels of expression across 80 different microarray experiments, a broad range of transcript abundance, and no role in mRNA processing (Table 5.2; Clark *et al.*, 2002).

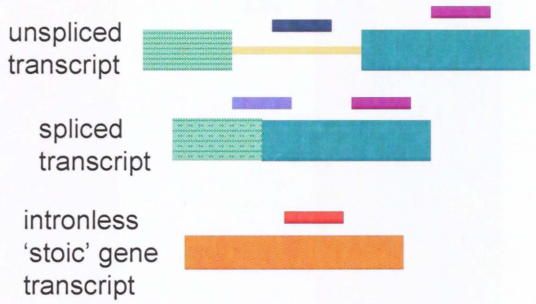


Figure 5.10: Schematic of microarray assay designed to measure global changes in splicing. Probes were designed to be specific for: within the intron and the second exon (for determination of unspliced transcript accumulation (IA) relative to the total amount of that transcript), spanning the exon junction, and in the second exon (determination of spliced transcript accumulation, as for IA). Data from seven 'stoic' genes were used as controls.

Name	Function
Sly1	vesicle trafficking between ER and golgi
Sec4	golgi to plasma membrane transport
Vps45	golgi to vacuole transport
Pex4	ubiquitin conjugating enzyme
Taf145	general RNA Pol II TF
Rsc2	chromatin remodelling
Yap1	jun-like transcription factor

Table 5.2: Common names and functions of proteins produced by the 'stoic' genes used as controls for normalisation of splicing-specific microarray data

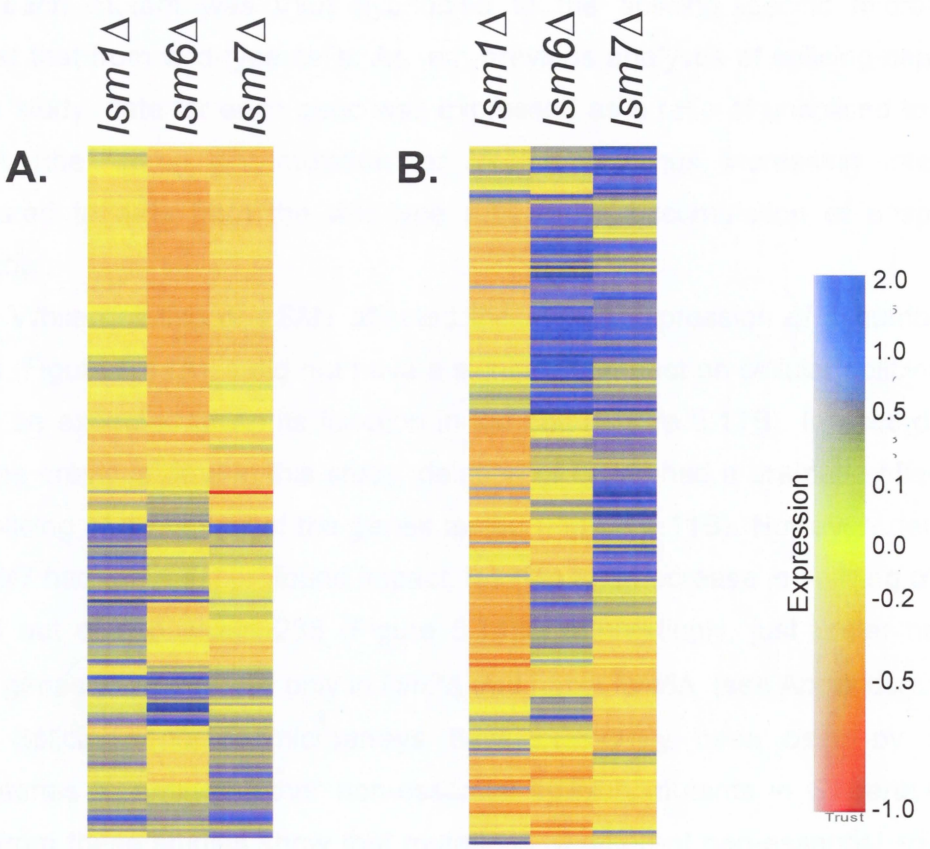


Figure 5.11: Global changes in expression and splicing of intron-containing genes in *S. cerevisiae*. Data from the splicing-sensitive microarrays were analysed for expression of genes containing introns (A) or IA index (B) in *lsm1*Δ, *lsm6*Δ or *lsm7*Δ, relative to data obtained from wild type cells. In all cases, cells were grown on glucose. Changes in this ratio are indicated by colour. Data shown has been hierarchically clustered using Genespring GX. Raw data is presented in the Appendix.

Each cell type was grown on rich medium with glucose as the carbon source. RNA was harvested from cells grown to mid-log phase on rich media containing glucose as described by Clark *et al* (2002). Polyadenylated RNA from each mutant was then hybridised to the splicing-specific microarray against that from wild-type cells. As with previous analyses of splicing capacity in this study, data for each gene was expressed as a ratio of unspliced to total mRNA (the 'intron accumulation' or IA index). Thus increasing intensity compared to data from the wild-type reflects an accumulation of unspliced transcript.

While deletion of *LSM1* affected the overall expression of a number of genes (Figure 5.11A), it did not have a significant impact on cellular splicing, as would be expected given its function in the cell (Figure 5.11B). In accordance with the previous data in this study, deletion of *LSM6* had a dramatic effect on the splicing of a number of the genes tested (Figure 5.11B). However, deletion of *LSM7* had the most profound impact, leading to a decrease in splicing of 155 genes out of a possible 238 (Figure 5.11B). Interestingly, just under half of these genes were affected only in *lsm7*Δ and not in *lsm6*Δ (see Appendix).

Splicing-sensitive microarrays have previously been used by other laboratories to analyse other non-essential splicing mutants in *S. cerevisiae*. Data from these studies show that mutations of different non-essential splicing factors affect different subsets of introns with shared features such as length or branchpoint position (Clark *et al.,* 2002; Burckin *et al.,* 2005; Sapra *et al.,* 2004). However no such association could be identified between genes with high IA, lariat length, unusual intron position (such as in the 5'UTR or near the 3' end of the gene), or unusual splice-site sequences. Transcripts with higher IA in *lsm7*Δ than in the wild type appeared to be enriched for genes with smaller introns, although this was not found to be a statistically significant association (Figure 5.12, Appendix). Genes exhibiting the highest IA in the *lsm7*Δ mutant tended to be of low abundance (Figure 5.12; Holstege *et al.,* 1998). However the converse was not true (Figure 5.12). This may reflect the mechanism by which splicing is modulated in the *lsm7*Δ mutant. Previous studies have found that mutations of splicing factors which function at the same stage of the splicing process tend to affect the same groups of genes (Clark *et al.,*2002;

Burckin *et al.*, 2004). However no association was found between genes with high IA in *lsm7Δ* and those identified as having a high IA in the non-essential splicing mutants *nam8Δ*, *msl1Δ*, *snu66Δ*, *prp18Δ*, *prp17Δ*, or *brr1Δ* (Clark *et al.*, 2002). These mutants involve deletions of genes implicated in the U1snRNP, U2snRNP, U4/U6.U5 tri-snRNP, U5snRNP, the second step of splicing, and snRNP recycling, respectively (as identified on the Stanford Genome Database). Therefore the process of splicing modulation in the *lsm7Δ* mutant may be different to that in any of these other non-essential splicing mutants.

GO analysis of genes with a high IA index in *lsm7Δ* mutants found few instances where these genes were significantly over-represented in particular functional groups. Of the gene processes which were identified as being enriched for high IA genes, the most significant were regulation of translational fidelity (with 6 out of 8 possible genes in the genome) and assembly of ribosomal subunits (Table 5.1; Appendix).

Functional Classification	Proportion relative to total number of genes with introns in that category	Proportion relative to total number of genes in that category	p value
ribosomal small subunit assembly	5 out of 8, 62.5%	5 out of 13	8.98×10^{-6}
ribosomal subunit assembly	9 out of 12, 75%	9 out of 52	1.72×10^{-6}
regulation of translational fidelity	6 out of 7, 85.7%	6 out of 8	4.49×10^{-7}
ER to golgi transport	7 out of 10	7 out of 66	0.00088
cytoplasm organisation and biogenesis	14 out of 22	14 out of 249	0.00027
M phase of meiosis or mitosis	7 out of 21	7 out of 205	0.00081
actin filament-based process	6 out of 8	6 out of 81	0.02208

Table 5.1: Genes whose transcripts exhibited reduced splicing in *LSM7Δ* relative to wild type fall into a broad range of functional groups. Groups were determined based on Gene Ontology (GO) data (see Appendix for more detail).

Some of the genes tested (43) were not affected by the deletion of *LSM7* but had impaired splicing in *lsm6*Δ. These included many genes with introns very close to the start of the gene. A group of about 20 genes was identified which were unaffected by deletion of either *LSM6* or *LSM7*. These included genes with introns involved in vital cellular processes such as sporulation (such as *MEI4* and *UBC4*), DNA repair (UBC13 and RFA2), and all five genes with introns implicated in protein polyubiquitination (Appendix). These data confirm the importance of *LSM7* in splicing transcripts of the majority of genes with introns but may also provide an explanation as to why the mutants with this gene deleted are still viable. While deletion of *LSM7* has significant impacts on the splicing of many genes, these data show than not all transcripts are affected equally – some important transcripts appear to be 'protected' from this effect.

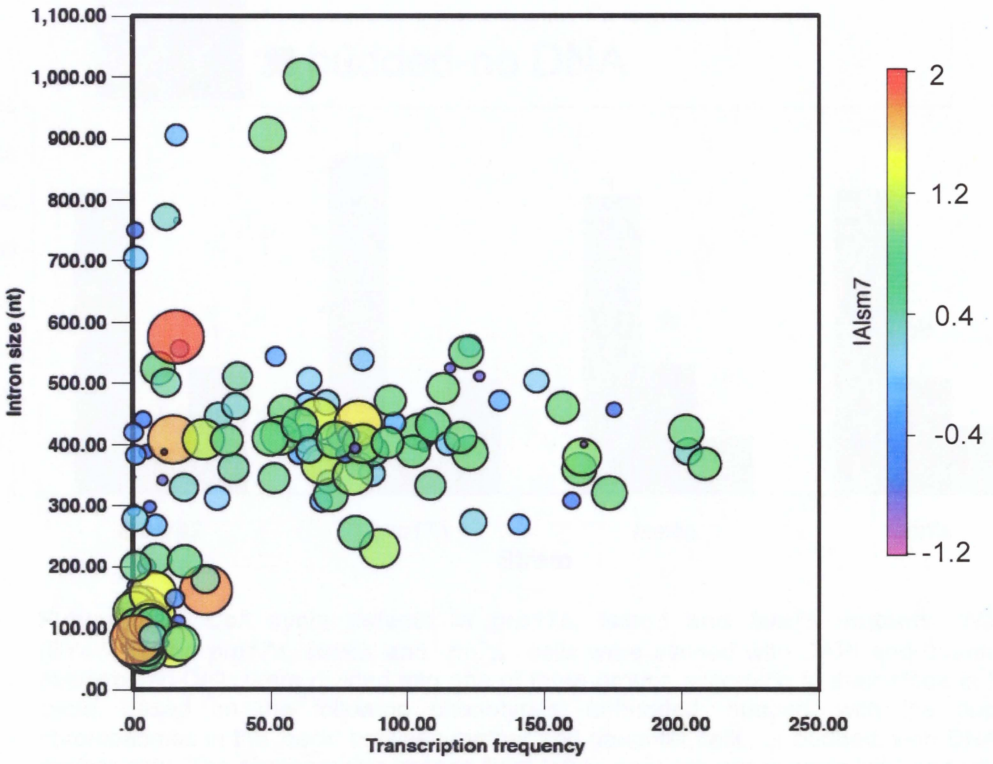


Figure 5.12: IA index in *lsm7*Δ mutants relative to transcription frequency or intron size. Data for transcription frequency and intron size were obtained from the Saccharomyces Genome Database or the Ares Intron Database (Grate and Ares,2002). The graph was generated using Aable software. Raw data used to generate this graph may be found in the Appendix.

5.9 Deletion of LSM7 has significant effects on cell physiology

Further evidence of the effect of *LSM* gene expression on cellular growth is provided from DAPI staining of cells from *prp17*Δ, *lsm6*Δ and *lsm7*Δ mutants as compared to the wild-type. Cells were grown to mid-log phase on rich media containing glucose and then fixed, stained and examined microscopically.

Prp17 protein (also called Cdc40) is required for the second trans-esterification reaction during splicing (Jones *et al.,* 1995; Vijayraghavan *et al.,* 2000). It has also been implicated in progression through the G1/S phase of the cell division cycle (Boger-Nadjar *et al.,* 1998). At the restrictive temperature,

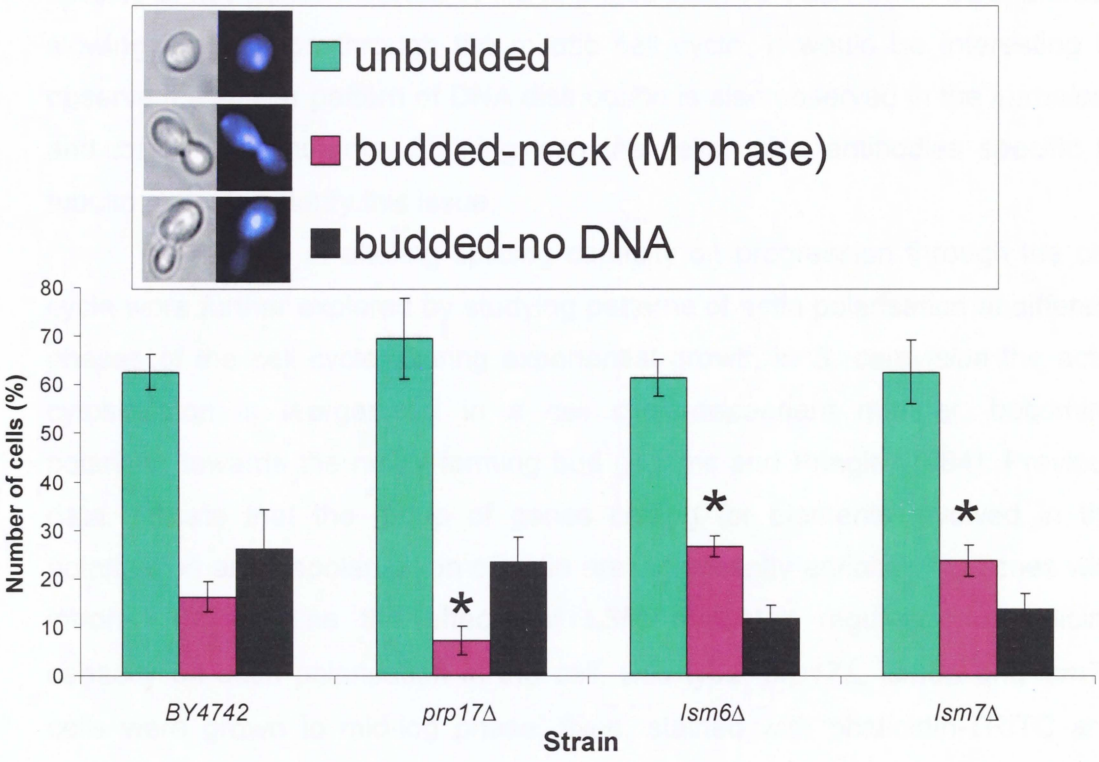


Figure 5.13: Cell cycle defects in *prp17*Δ, *lsm6*Δ and *lsm7*Δ mutants. Wild type (BY4742) and *prp17*Δ, *lsm6*Δ and *lsm7*Δ cells were stained with DAPI and quantified by microscopy. Cells were divided into one of three groups according to their stage in the cell cycle, based on the following phenotypes: unbudded; budded, with the duplicated chromosomes in the 'neck' between mother and daughter cells; or budded, with DNA in the mother only. The photographic images from left to right are under white light and under UV light, respectively. For each cell type a minimum of 400 cells were examined. Stars indicate results which diverged significantly from data obtained from the wild-type, according to a two-tailed t-test.

prp17 strains arrest as large unbudded cells with an undivided nucleus containing 2N levels of DNA (Vaisman *et al.,* 1995; Ben-Yehuda *et al.,* 2000). Therefore mutants with *PRP17* deleted were used as a positive control in this analysis.

As expected the *prp17*Δ mutant showed an increased number of unbudded cells and a significant concomitant decrease in the abundance of cells with DNA in the neck between the two cells. By contrast, both *lsm6*Δ and *lsm7*Δ mutants showed a significant increase in the proportion of cells with the DNA in the neck between the mother and budding daughter cell (Figure 5.13; $P<0.01$ in both cases). This effect may result from the impact of reduced splicing of key genes involved in the M phase such as *TUB1* and *TUB2*, thereby slowing progression through the mitotic cell cycle. It would be interesting to observe if a similar pattern of DNA distribution is also observed in the *intronless* and *intron only* mutants. Staining with the cells using antibodies specific to tubulin may also clarify this issue.

The effects of altering splicing capacity on progression through the cell cycle were further explored by studying patterns of actin polarisation at different phases of the cell cycle. During exponential growth in *S. cerevisiae* the actin cytoskeleton is reorganised in a cell cycle-dependent manner, becoming polarised towards the newly-forming bud (Adams and Pringle, 1984). Previous data indicate that the group of genes coding for elements involved in the polarisation and depolarisation of actin are significantly enriched for genes with introns. To examine the effects of *LSM7*-mediated regulation of splicing capacity on actin polarisation in the cell, wild type, *prp17*Δ, *lsm6*Δ and *lsm7*Δ cells were grown to mid-log phase, fixed, stained with phalloidin-TRITC and examined by fluorescence microscopy. Cells were then classified according to actin localisation patterns and cell division status into one of four groups: unbudded cell, delocalised actin; budded cell, delocalised actin; unbudded cell, actin localised primarily at the potential bud site; and budded cell, actin localised in the bud.

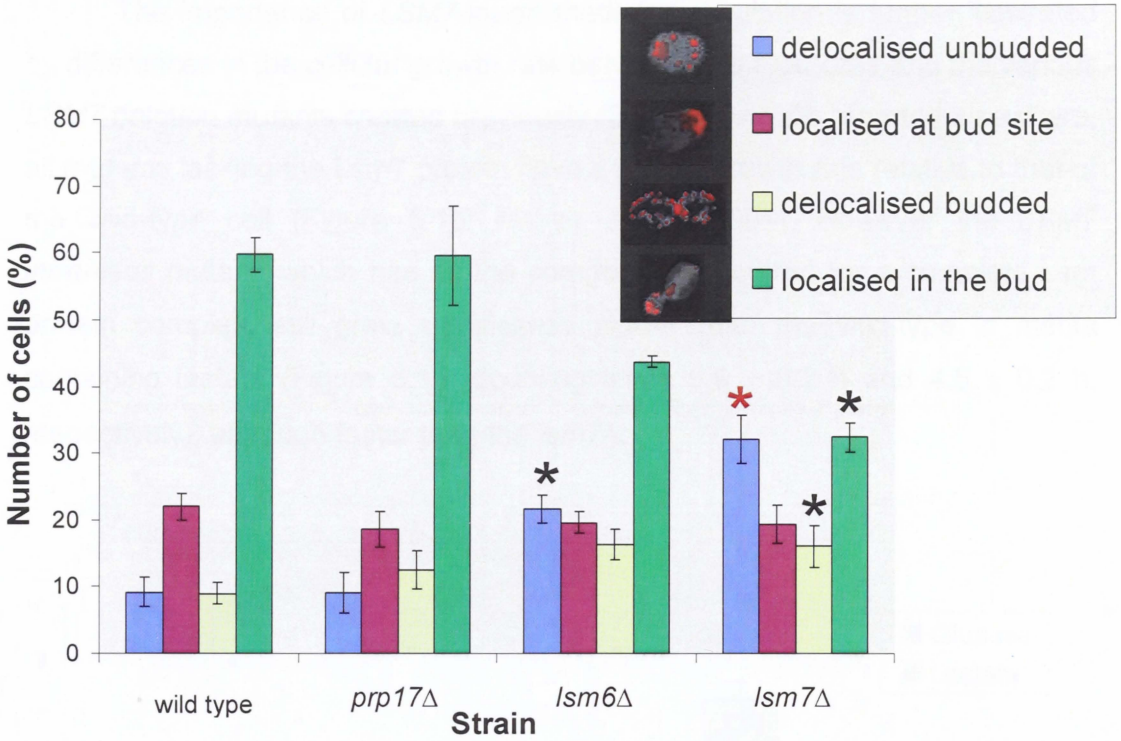


Figure 5.14: **Actin depolarisation in *prp17Δ*, *lsm6Δ* and *lsm7Δ* mutants**. Wild type (BY4742) and *prp17Δ*, *lsm6Δ* and *lsm7Δ* cells were stained with TRITC-phalloidin and quantified by microscopy. Cells were divided into one of four groups according to actin polarisation patterns: unbudded, delocalised actin; unbudded, actin at bud site; budded delocalised actin; and budded actin located primarily in the bud. The photographic images from left to right overlay images from UV light and white light for each cell. For each cell type a minimum of 400 cells were examined. Stars indicate results which diverged significantly from data obtained from the wild-type, according to a two-tailed t-test. The red star indicates a highly significant difference.

Patterns of actin localisation in the *prp17Δ* mutant did not differ significantly from that of the wild type cell (Figure 5.14). By contrast, in both *lsm6Δ* and *lsm7Δ* mutants, the proportion of cells with delocalised actin was higher than in the wild type. This difference was most pronounced between *lsm7Δ* mutants and wild type cells, with a significantly higher proportion of both budded and unbudded cells with delocalised actin (Figure 5.14). Mutants with *LSM7* deleted also had a significantly lower proportion of dividing cells with the actin localised within the bud region (Figure 5.14). These data indicate that deletion of *LSM7* can have significant effects on key elements of cellular physiology and growth.

The importance of *LSM7*-intron mediated regulation is further illustrated by differences in the cellular growth rate between wild-type cells and the various *LSM7* deletion mutants created previously (Section 4.4). As reported elsewhere, all mutants lacking the Lsm7 protein have a slower growth rate relative to that of the wild-type cell (Figure 5.15; Mayes *et al.*, 1999). However the *LSM7 intronless* mutant, which has all the components required for a complete Lsm protein complex, still grew significantly slower than the wild-type in media containing lactate (Figure 5.15; doubling times 5.9 ± 0.2 h and 4.5 ± 0.2 h, respectively), although faster than the *lsm7*Δ.

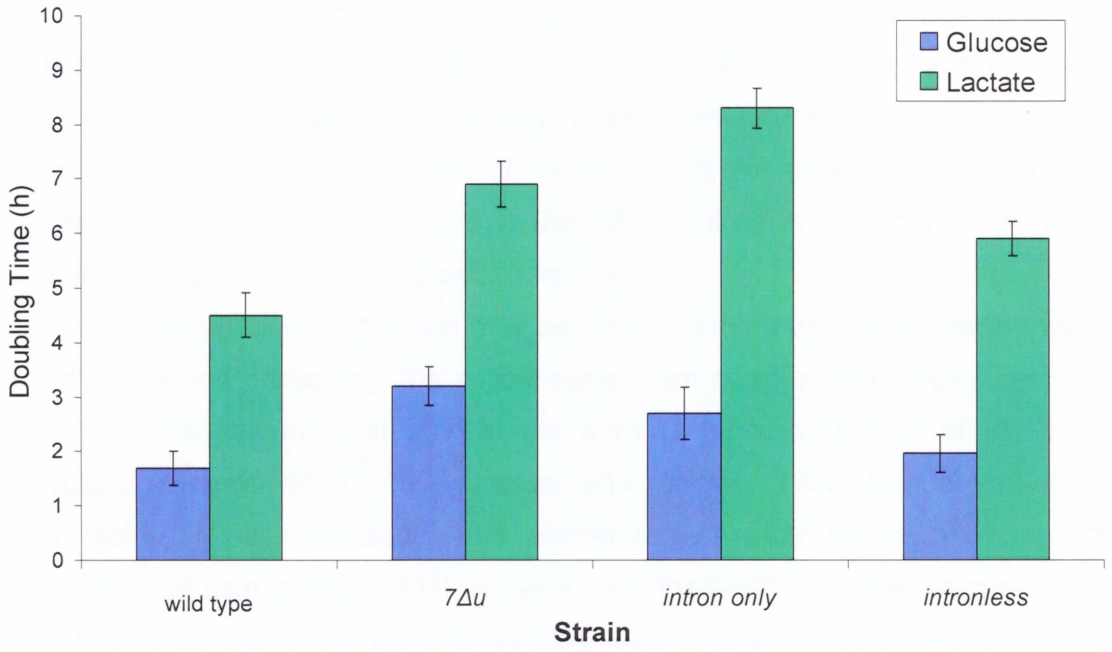


Figure 5.15: Growth rates of different *LSM7* mutants in media containing glucose or lactate. Both wild type and *LSM7* mutant cells (*lsm7*Δu, *intronless* and *intron only)* were grown for at least 5 doublings in media containing either glucose or lactate as the carbon source, and the growth rates determined based on the density of the culture at OD_{600}nm.

5.10 Discussion

LSM2, 4, 5, 7 and *8* all encode proteins which form part of an essential complex required for splicing in the cell. Data from previous chapters show that transcription of the genes encoding these proteins is repressed in a co-ordinate fashion under different growth conditions. Such regulation may result in a change in the capacity of the cell to splice pre-mRNA. Accordingly, *in vivo* splicing capacity was measured in the various growth conditions by real time PCR analysis of a representative gene, *RPL25.* In each case, the capacity for splicing *RPL25* pre-mRNA changed in a manner concordant with the change in expression of the *LSM* co-regulated gene set. This indicates that these two phenomena may be functionally related.

Furthermore, this response was not found in mutants with functional Lsm7p but lacking the *LSM7* intron. Conversely, splicing capacity in the *intron only* mutant decreased dramatically when it was grown on poorer carbon sources, such that around 15% of the *RPL25* mRNA detected was unspliced. These data strongly indicate that the *LSM7* intron is required to modulate splicing under the different growth conditions.

This splicing modulation may be achieved in a number of different ways. As mentioned previously, the spliceosomal Lsm complex stabilises U6snRNA, which catalyses the first step in the splicing process (Mayes *et al.,* 1999; Salgado-Garrido *et al.*, 1999; Lesser and Guthrie, 1993; Kandel-Lewis and Séraphin, 1993). Regulating the expression of *LSM* genes encoding the majority of Lsm spliceosomal proteins may therefore alter the amount of U6 snRNA available for the splicing reaction, leading to a change in cellular splicing capacity. Indeed, deletion of some *LSM* genes involved in the spliceosome is known to result in a lower level of U6snRNA than found in wild-type cells (Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999; Verdone *et al.,* 2004).

Levels of U6snRNA may also be altered in a slightly more indirect manner. GO analysis indicated that a group of genes involved in the positive regulation of RNA polymerase III genes (such as U6snRNA) are enriched for genes with introns in this organism. Further, a number of genes with introns encode proteins which form part of the basal machinery required for RNA polymerase III transcription. Perhaps the most interesting of these is β-actin.

This protein is involved in a large number of activities both in the cytoplasm and the nucleus. Highly purified RNA polymerase III contains associated β-actin (Hue *et al.,* 2004). Further to this, β-actin has been located in the promoter region of actively transcribed U6 snRNA (Hu *et al.,* 2004). Several groups of genes implicated in processes related to actin were also significantly enriched for genes with introns. Therefore initial effects on splicing by some other mechanism may reduce the amount of Pol III transcriptional activators or appropriately processed actin available, and thus lead to a down-regulation of Pol III-mediated transcription. This is consistent with reports by other investigators suggesting that RNA polymerase III partly dissociates from DNA during stationary-phase growth, when splicing and expression of the *LSM* co-regulated gene set are both reduced (Harismendy *et al.,* 2003).

An alternative way in which regulation of *LSM2, 4, 5* and *8* may impact on splicing capacity is through the interactions of their respective proteins and Prp24. This is an essential splicing factor involved in spliceosome assembly and recycling (Shannon and Guthrie, 1991; Jandrositz and Guthrie 1995). The Lsm 2-8p complex strongly enhances the binding of Prp24 to the telestem region of U6snRNA, promoting the reassembly of the U4-U6 di-snRNP and U4-U6•U5 tri-snRNP complexes following an initial splicing reaction (Ryan *et al.,* 2002; Verdone *et al.,* 2004). Extracts from cells with *LSM7* deleted have been found to be incapable of supporting a second round of splicing *in vitro* (Verdone *et al.,* 2004) and thus cells with reduced levels of *LSM7* transcript may also have an impaired ability to recycle the spliceosome through successive reactions, leading to a reduced overall splicing capacity.

Since the *LSM7* intron modulates splicing, expression of the Lsm7 protein is likely to include at least one autoregulatory step. Reduction in levels of either Lsm2p or Lsm7p has been shown, both in this study and elsewhere, to result in impaired splicing (Mayes *et al.,* 1999; Salgado-Garrido *et al.,* 1999). The capacity of the cell to splice either *LSM7* or *LSM2* pre-mRNA transcripts varied in the same way as that for *RPL25* in cells grown on different carbon sources. Thus both *LSM2* and *LSM7* may be additionally auto-regulated at the level of splicing.

Several examples of genes which are negatively auto-regulated at the level of splicing have been identified in *S. cerevisiae,* including *YRA1*, *RPL32* and *DBP2* (Dabeva and Warner 1993; Barta and Iggo, 1995; Preker *et al.,* 2002). However none of the genes discovered so far encodes a protein involved in splicing.

Since the *LSM7* intron also regulates transcription of the *LSM* co-regulated gene set, changes in the ability of the cell to splice *LSM7* under the different growth conditions may also lead to more complex patterns of autoregulation by other members of this set. This will be discussed in more detail in Chapter 6.

While alternative splicing in *S. cerevisiae* is rare, a number of examples have been found where the splicing of a particular transcript or small group of transcripts can vary under different conditions. The most well studied of these is the splicing of *MER2, MER3* and *SPO70* transcripts following the production of the meiosis-specific splicing factor Mer1p (Engebrecht *et al.,* 1991, Nakagama and Ogawa, 1999, Spingola and Ares, 2000). In each of the cases identified so far this variation in splicing is believed to affect a relatively small subset of introns, most of which have unusual intron features affecting size, secondary structure, location within the gene or splice-site signals (Eng and Warner, 1991; Dahan and Kupiec, 2004; Chawla *et al.,* 2003; Burns *et al.,* 2002). Change in the splicing status of these transcripts is thought to be mediated by the production of specific splicing factors which alter the processing of these abnormal introns. In this study, however, changes in splicing capacity appeared to depend on the production of elements at the core of the spliceosome. The gene used to measure these changes in splicing capability (*RPL25*) has none of the unusual intron features usually found in variably spliced transcripts. Such regulation could potentially impact on the splicing of a wide array of genes with introns in this organism.

While only a small number of genes in this organism contain introns, genes with introns are over-represented in a number of small functional groups with key roles in the cell. Deletion of *LSM7* was found to affect the splicing of the majority of genes with introns, including ribosomal protein genes and genes encoding proteins involved in translational regulation and the cell cycle.

Interestingly, some genes with introns were not affected by deletion of *LSM7*. This could occur if pre-mRNA transcripts from these genes had particularly high affinities for elements of the spliceosome, and thus could successfully compete with other intron-containing transcripts for the U6snRNP even if levels were limiting. However further experiments are required to determine which genes are most affected by *LSM7* intron-mediated regulation of splicing capacity in wild-type cells.

Although they constitute only about 2% of the genes encoded in the *S. cerevisiae* genome, almost 90% of mRNA transcripts containing introns arise from genes encoding ribosomal proteins (Warner 1999; Ares *et al.,* 1999). As discussed in section 1.8, ribosomal protein genes are regulated at multiple levels to ensure a tight balance between production and cellular requirement, so that no pool of free ribosomal protein is ever present (reviewed in Mager, 1988 and Warner, 1999). Expression of ribosomal protein genes containing introns correlated with changes in the splicing capacity. Regulation at the level of splicing may therefore act in concert with both transcriptional and translational regulation to reduce the amount of free ribosomal proteins present under these conditions. Further, by influencing the splicing of genes involved in both the ribosome and transcriptional regulation, this regulation may act along with other factors to down-regulate translation in the cell.

A number of genes have been identified which have functions in both splicing and the cell cycle, suggesting that the two processes are linked (Ben-Yehuda *et al.,* 2000; Russell *et al.,* 2000). Mutations in some of these splicing factors have been shown to lead to cell cycle arrest, which can be suppressed by the deletion of the intron in one or both of the genes encoding α-tubulin (Burns *et al.,* 2002; Dahan and Kupiec, 2002). Analysis of DNA distribution in *lsm7*Δ mutants showed a significant number had accumulated with the DNA in the neck between the mother and daughter cell (around the 'M' phase of mitosis). Genes with introns are statistically over-represented in the 'M phase' gene group in the GO database. Therefore the accumulation of *lsm7*Δ cells in M phase may be a result of impaired splicing in this mutant. This could be further investigated by deletion of the introns in some 'M phase' genes in the *lsm7*Δ

mutant, particularly those encoding α-tubulin, to see if this alleviates the aberrant cell cycle phenotype.

Deletion of *LSM7* also led to a significant increase in cells with depolarised actin in the cytoskeleton. To determine if this phenotype is due to a lower splicing capacity in the *lsm7*Δ mutant, the intron of the actin gene should also be removed and actin-staining patterns in the resulting mutant compared with both *lsm7*Δ and wild-type cells.

It has been estimated that during each hour of active growth, *S. cerevisiae* must catalyse the removal of at least ten thousand introns (Ares *et al.,* 1999). Therefore even the small changes in splicing capacity observed for individual transcripts in the wild type cell (from < 0.1% to around 2%) may have a profound overall effect. Smaller differences in splicing capacity were observed between *LSM7* mutants and wild type cells in each of the growth conditions tested. However in every case, these mutants were found to have a significantly impaired growth rate, illustrating the major effects modulating splicing may have on cellular physiology. This regulation may act as an important additional level of regulation for many genes with introns in this organism. Such a multi-tiered approach to regulation would provide a highly coordinated and more rapid response to important changes in cellular conditions.

In each of the cases presented in this study, it has been assumed that changes in splicing capacity values in cells in the different growth conditions were at least partly due to changes in the rate of splicing of pre-mRNA. However it is possible that the changes measured are instead due to differences in the rate of degradation of the unspliced transcript. To test this, splicing assays should be carried out on strains with mutations in genes involved in degradation of unspliced mRNA (such as Rnt1p or Xrb1p), under the various growth conditions used in this study.

Future work should also include analysis of the effect of coordinate regulation of *LSM* genes on Lsm protein complexes in the cell other than the spliceosome. The effect of altering splicing on other elements requiring spliceosome activity, but not present on the splicing-specific microarray used here, should also be considered, such as abundance of snoRNAs. Lastly, by modulating splicing, the transcription and translation of many genes with introns

may also be altered, as a result of functional links between these processes (as discussed in section 1.2.3). This possibility is discussed in more detail in Chapter 6.

Chapter 6: Summary and Perspectives

This study set out to explore the regulation of a group of genes which encode proteins that form an essential complex in eukaryotes. Most, but not all, of these genes were found to be co-regulated at the level of transcription initiation in response to a variety of environmental conditions. Expression of these co-regulated genes changed in the same manner as the apparent activity of the complex (as measured by splicing capacity). Two genes in the co-regulated gene set were found to be expressed at very low levels in the growth conditions tested, and thus are most likely to encode the limiting factors in the final assembly of the protein complex. Levels of protein for each gene should be quantified under the different growth conditions to verify these findings, however, to ensure no significant antagonistic regulation occurs at the level of translation. These data indicate that, for low-abundance proteins at least, it may not be necessary to co-regulate the expression of all members of a complex to control the activity of that complex. This assumes the proteins encoded by genes which are not part of the co-regulated gene set are not deleterious to the cell if present in slight excess.

Analysis of various deletion mutants indicated that co-expression of the *LSM* co-regulated gene set (*LSM2, 4, 5, 7* and *8)* was a result of a repressive mechanism, particularly in cells grown on poorer carbon sources. Interestingly, this regulation was mediated, at least in part, by the intron of *LSM7* itself. This is the first reported example of a regulatory intron acting *in trans* in any eukaryote.

The existence of *trans*-acting regulatory elements within introns has been previously proposed based on a number of lines of evidence (Mattick and Gagen 2001). In higher eukaryotes small nucleolar RNAs (snoRNAs) are involved in the site-specific modification of rRNA and snRNA (Tycowski *et al.,* 1998) and are derived by a complex process from within the intron of unspliced pre-mRNA or from the excised intron lariat (reviewed in Bachellerie *et al.,* 2002).

Non-coding regulatory RNAs from intergenic regions of DNA ('micro' or miRNA and 'short interfering' or siRNA), such as *lin-4* and *let-7,* influence a wide variety of developmental processes in plants and higher eukaryotes

(reviewed in Ambros, 2004; Bartel 2004, Mattick and Makunin, 2005). They have also been implicated in chromatin modification, imprinting and DNA methylation (Mattick, 2003).

Finally, non-coding sequences which are almost 100% conserved across species have been identified in the introns of various higher eukaryotes (Dermitazakis *et al.,* 2003; Bejerano *et al.,* 2004; Glazov *et al.,* 2005). This indicates that these sequences may be under stabilising selection and therefore likely to be functional, although this does not differentiate between *trans*-acting and *cis*- acting effects, such as those described in section 1.2.3

The *LSM7* intron is also required to mediate the changes in U6snRNA levels and splicing capacity in wild-type cells grown on different carbon sources. Further, these changes correlated with the change in expression of the co-regulated gene set observed previously, indicating that these three events may be functionally linked. A model describing how this could occur is presented in Figure 6.1.

No experiments so far have directly addressed whether the repression mediated by the *LSM7* intron is enhanced, relieved or unaffected by the splicing of the *LSM7* transcript. Work is currently underway to create an *LSM7* mutant with the splice sites destroyed to answer this question. However, in this study, members of the *LSM* co-regulated gene set were found to be repressed under conditions where splicing of *LSM7* pre-mRNA was also found to be impaired. Therefore regulation by the *LSM7* intron is more likely to be mediated by the unspliced *LSM7* mRNA, than the intron lariat.

Repression of transcription of these genes may lead to a reduction in the levels of most spliceosomal Lsm proteins available to form the Lsm spliceosomal complex. This would then affect U4:U6 hetero-dimer formation and U6snRNA stability, leading to a reduction in cellular splicing capacity. Since the U6 snRNP is at the core of the spliceosome, the regulation described here may affect the splicing of a broad range of transcripts, acting as an additional level of regulation in the expression of those genes. Results from splicing-sensitive microarrays performed on *lsm7*Δ (compared to wild type) support this suggestion, since many genes, with a variety of different intron features, are

affected by deletion of *LSM7.* In contrast, fewer genes with introns are affected by the deletion of *LSM6.*

Genes with introns in *S. cerevisiae* are overrepresented in a number of different processes in the cell. The most significant of these include functions related to translation (such as ribosomal assembly and maintenance, or regulation of translational fidelity), cell cycle and regulation of polymerase III-mediated transcription. This last process may form part of an autoregulatory loop by inhibiting transcription of U6snRNA.

Other autoregulatory loops may also exist at different stages of this regulation, as shown in Figure 6.1. For example, if the *LSM7* intron (in an unspliced or lariat form) represses the expression of *LSM7* itself, as well as other genes, this may form a negative autoregulatory loop (represented by the blue line on Figure 6.1). It should be noted that, since the data indicates that splicing is modulated, rather than turned off altogether, there must be mechanisms which interrupt these repressive autoregulatory loops and thus prevent the down-regulation of splicing from becoming too severe. One way to stop the regulation from turning off splicing altogether would be if unspliced *LSM7* repressed the expression of its own transcript more strongly than those of the other members of the co-regulated gene set. This may account for the very low abundance of *LSM7* transcripts, relative those of the other *LSM* genes. Therefore if levels of unspliced *LSM7* got too high, the transcription of the *LSM7* gene may be practically turned off, eventually allowing the expression of the other members of the co-regulated gene set to increase. This assumes some unspliced *LSM7* pre-mRNA is eventually degraded by the activity of enzymes such as Rnt1p (Danin-Kreiselman *et al.,* 2003).

Considerable evidence exists to suggest that the process of splicing is either directly or indirectly coupled to other upstream and downstream processes, such as transcription, translation, or mRNA export, as discussed in section 1.2.3. (reviewed in Maniatis and Reed, 2002; Le Hir *et al.,* 2003). Therefore altering splicing may also impact upon many other stages of gene expression. Due to the complex array of interacting factors, it would be difficult to comprehensively address this issue. However, analysis of transcript abundance and protein expression in double mutants lacking both the *LSM7*

intron and the introns from another gene, compared to those of the *intronless* mutant alone, may give some indication.

Interestingly, a number of different steps in this model may lead to a repressive effect on translation – either specifically, by disruption of the functional coupling between splicing and translation for genes with introns, or more generally, as an effect of reduced expression of genes involved in key steps of the translation process (Figure 6.1). Moreover, as mentioned in Section 1.6, in addition to their other roles in the cell, Lsm proteins are required for the correct processing of pre-tRNAs (Kufel *et al.,* 2002). Therefore regulation of *LSM* gene transcription may have quite a profound effect on translation in the cell.

Many questions remain regarding the intron-mediated regulation reported in this study, in addition to those mentioned previously. As discussed in section 4.7, the precise manner by which the *LSM7* intron mediates this regulation has yet to be determined, although sequence conservation findings indicate it may be related to the secondary structure of the unspliced RNA.

What is the 'trigger' or sensing mechanism which turns this repression on? *LSM7* was the only *LSM* gene examined which was additionally regulated at the level of mRNA stability. Therefore changes in the cell resulting in increase or decrease in *LSM7* stability could potentially act as a type of switch. Alternatively, as mentioned in section 4.7, modulation of some other spliceosomal element, such as a factor involved in the binding of the U1 or the U2 snRNP to this transcript, could cause a decrease in splicing, leading to an increase in the abundance of the unspliced regulatory *LSM7* and its assorted effects. When the conditions change an increase in the abundance of this unknown spliceosomal factor could allow it to bind unspliced *LSM7*, targeting it for processing, or perhaps directly displacing an activator protein (section 4.7). This could promptly alleviate the repression of transcription of the various *LSM7* intron-regulated genes. The type of initial environmental signal (or signals) which causes this effect has not yet been identified, although either changes carbon source or growth rate seem to be good candidates.

What is the purpose of this regulation? One role for the *LSM7* intron in the cell is to affect the synthesis of components of the spliceosome, through a

process likely to involve at least one autoregulatory step. Modulation of splicing may also form an additional level of regulation for genes involved in various elements of translation, as discussed in Chapter 5. Regulation of splicing may also impact on transcripts from genes involved in the cell cycle. Ribosomal protein genes are known to be exquisitely sensitive to growth potential of the cell, as indicated by environmental factors including carbon source (Ju and Warner, 1994). Ribosome biogenesis has also been linked to cell cycle regulation in both yeast and *Drosophila melanogaster* (Montange, *et al.,* 1999; Thomas, 2000). The regulation examined in this thesis may fulfil a similar purpose, reducing cellular growth under less favourable environmental conditions, such as poor carbon sources.

The regulon of genes affected by *LSM7*-intron mediated regulation has not yet been fully elucidated, and may provide some further clues as to its cellular function. This study has already shown that *LSM7*-intron mediated regulation affects one non-*LSM* gene (*SMX2)* and the regulation of other remaining 5 spliceosomal *SM* genes should also be explored. In addition, this study did not examine the effects of *LSM7*-intron mediated regulation on the other Lsm complexes in the cell.

Does this system exist in higher eukaryotes? Most splicing research in higher eukaryotes has focussed on peripheral splicing factors involved in regulation of alternative splicing. However general splicing factors in *D. melanogaster* do appear to be developmentally regulated. Interestingly, a number of snoRNAs in higher eukaryotes have been located within the introns of genes encoding proteins involved in the ribosome or the cell cycle, indicating they may also play some regulatory role related to translation or cell growth (Filipowicz *et al.,* 1999; Rebane *et al.,* 1998).

These results represent a key advance in our understanding of how the complex pattern of gene expression is achieved in the eukaryotic cell under different conditions. This has important implications for how we perceive much of the sequence that makes up the human genome and that of other higher eukaryotes.

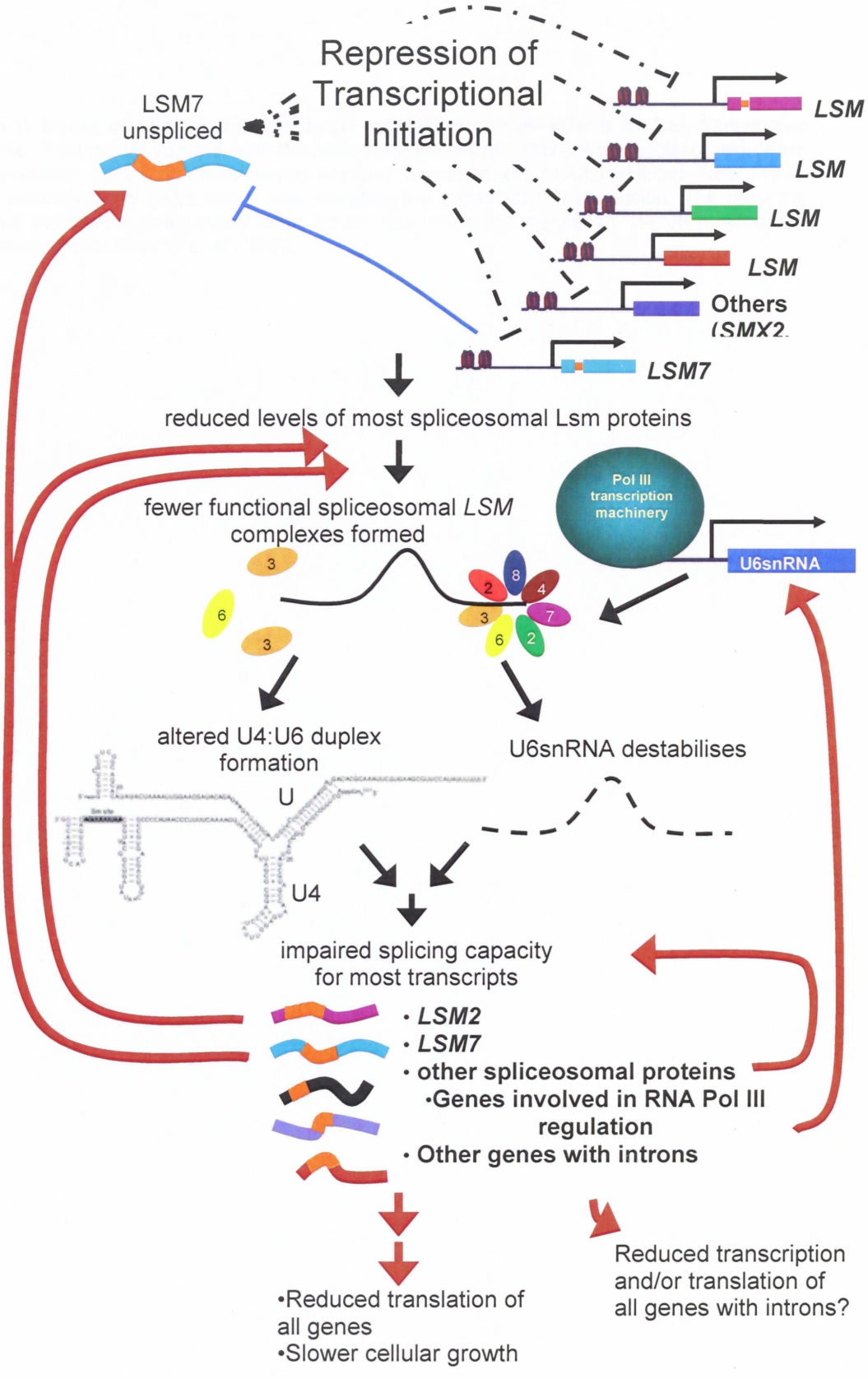

Repression of Transcriptional Initiation
LSM7 unspliced
LSM
LSM
LSM
LSM
Others (SMX2
LSM7
reduced levels of most spliceosomal Lsm proteins
fewer functional spliceosomal LSM complexes formed
Pol III transcription machinery
U6snRNA
altered U4:U6 duplex formation
U6snRNA destabilises
U
U4
impaired splicing capacity for most transcripts
• LSM2
• LSM7
• other spliceosomal proteins
•Genes involved in RNA Pol III regulation
• Other genes with introns
•Reduced translation of all genes
•Slower cellular growth
Reduced transcription and/or translation of all genes with introns?

Figure 6.1: Model of *LSM7*-intron mediated regulation and its effects in *Saccharomyces cerevisiae.* This model displays both the immediate effects of LSM7-intron mediated regulation and downstream effects on spliceosomal complex formation and cellular splicing. Red arrows indicate autoregulatory loops which may increase the impact of this regulation. The blue line represents potential autoregulatory loops which may inhibit this regulation. U4:U6 heterodimer image adapted from Nottrott *et al.,* 2002.

References

Abruzzi, K. C., A. Smith, *et al.* (2002). "Protection from free beta-tubulin by the beta-tubulin binding protein Rbl2p." Molecular & Cellular Biology 22(1): 138-47.

Achsel, T., H. Brahms, *et al.* (1999). "A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro." EMBO Journal 18(20): 5789-802.

Achsel, T., H. Stark, *et al.* (2001). "The Sm domain is an ancient RNA-binding motif with oligo(U) specificity." Proceedings of the National Academy of Sciences of the United States of America 98(7): 3685-9.

Adams, A. E. and J. R. Pringle (1984). "Relationship of actin and tubulin distribution to bud growth in wild-type and morphogenetic-mutant Saccharomyces cerevisiae." Journal of Cell Biology 98(3): 934-45.

Adams, C. C. and D. S. Gross (1991). "The yeast heat shock response is induced by conversion of cells to spheroplasts and by potent transcriptional inhibitors." Journal of Bacteriology 173(23): 7429-35.

Ambros, V. (2004). "The functions of animal microRNAs." Nature 431(7006): 350-5.

Ares, M., Jr., L. Grate, *et al.* (1999). "A handful of intron-containing genes produces the lion's share of yeast mRNA." Rna-A Publication of the Rna Society 5(9): 1138-9.

Ashburner, M., C. A. Ball, *et al.* (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nature Genetics 25(1): 25-9.

Auble, D. T., D. Wang, *et al.* (1997). "Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA." Molecular & Cellular Biology 17(8): 4842-51.

Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." Nature Reviews Genetics 2(1): 59-67.

Azizkhan, J. C., D. E. Jensen, *et al.* (1993). "Transcription from TATA-less promoters: dihydrofolate reductase as a model." Critical Reviews in Eukaryotic Gene Expression 3(4): 229-54.

Bachellerie, J. P., J. Cavaille, *et al.* (2002). "The expanding snoRNA world." Biochimie 84(8): 775-90.

Bailey, C. and D. L. Weeks (2000). "Understanding oligonucleotide-mediated inhibition of gene expression in Xenopus laevis oocytes." Nucleic Acids Research 28(5): 1154-61.

Ban, N., P. Nissen, *et al.* (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.[see comment]." Science 289(5481): 905-20.

Barta, I. and R. Iggo (1995). "Autoregulation of expression of the yeast Dbp2p 'DEAD-box' protein is mediated by sequences in the conserved DBP2 intron." EMBO Journal 14(15): 3800-8.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell 116(2): 281-97.

Basehoar, A. D., S. J. Zanton, *et al.* (2004). "Identification and distinct regulation of yeast TATA box-containing genes.[erratum appears in Cell. 2004 Jun 11;117(6):847]." Cell 116(5): 699-709.

Bashirullah, A., A. E. Pasquinelli, *et al.* (2003). "Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis." Developmental Biology 259(1): 1-8.

Beggs, J. D. (2005). "Lsm proteins and RNA processing." Biochemical Society Transactions 33(Pt 3): 433-8.

Bejerano, G., M. Pheasant, *et al.* (2004). "Ultraconserved elements in the human genome." Science 304(5675): 1321-5.

Ben-Yehuda, S., I. Dix, *et al.* (2000). "Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae." Genetics 156(4): 1503-17.

Berglund, J. A., K. Chua, *et al.* (1997). "The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC." Cell 89(5): 781-7.

Bjorklund, S. and C. M. Gustafsson (2005). "The yeast Mediator complex and its regulation." Trends in Biochemical Sciences 30(5): 240-4.

Black, D. L. (2000). "Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology." Cell 103(3): 367-70.

Boelens, W. C., I. Palacios, *et al.* (1995). "Nuclear retention of RNA as a mechanism for localization." Rna-A Publication of the Rna Society 1(3): 273-83.

Boger-Nadjar, E., N. Vaisman, *et al.* (1998). "Efficient initiation of S-phase in yeast requires Cdc40p, a protein involved in pre-mRNA splicing." Molecular & General Genetics 260(2-3): 232-41.

Bon, E., S. Casaregola, *et al.* (2003). "Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns." Nucleic Acids Research 31(4): 1121-35.

Bourbon, H.-M., A. Aguilera, *et al.* (2004). "A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II." Molecular Cell 14(5): 553-7.

Bourdon, V., A. Harvey, *et al.* (2001). "Introns and their positions affect the translational activity of mRNA in plant cells." EMBO Reports 2(5): 394-8.

Bouveret, E., G. Rigaut, *et al.* (2000). "A Sm-like protein complex that participates in mRNA degradation." EMBO Journal 19(7): 1661-71.

Brachmann, C. B., A. Davies, *et al.* (1998). "Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications." Yeast 14(2): 115-32.

Brandl, C. J. and K. Struhl (1990). "A nucleosome-positioning sequence is required for GCN4 to activate transcription in the absence of a TATA element." Molecular & Cellular Biology 10(8): 4256-65.

Brauer, M. J., A. J. Saldanha, *et al.* (2005). "Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures." Molecular Biology of the Cell 16(5): 2503-17.

Braunstein, M., A. B. Rose, *et al.* (1993). "Transcriptional silencing in yeast is associated with reduced nucleosome acetylation." Genes & Development 7(4): 592-604.

Brenner, S., Jacob, F. and Meselson, M. (1961). Nature 3: 576-581.

Brent, R. and M. Ptashne (1984). "A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene." Nature 312(5995): 612-5.

Brooks, A. R., B. P. Nagy, *et al.* (1994). "Sequences containing the second-intron enhancer are essential for transcription of the human apolipoprotein B gene in the livers of transgenic mice." Molecular & Cellular Biology 14(4): 2243-56.

Brow, D. A. (2002). "Allosteric cascade of spliceosome activation." Annual Review of Genetics 36: 333-60.

Brown, A. J. and F. A. Sagliocco (1996). "mRNA abundance and half-life measurements." Methods in Molecular Biology 53: 277-95.

Buchman, A. R. and P. Berg (1988). "Comparison of intron-dependent and intron-independent gene expression." Molecular & Cellular Biology 8(10): 4395-405.

Buchman, A. R., N. F. Lue, *et al.* (1988). "Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein." Molecular & Cellular Biology 8(12): 5086-99.

Buratowski, S. (1994). "The basics of basal transcription by RNA polymerase II." Cell 77(1): 1-3.

Buratti, E. and F. E. Baralle (2004). "Influence of RNA secondary structure on the pre-mRNA splicing process." Molecular & Cellular Biology 24(24): 10505-14.

Burge, C. B., Tuschi, T. and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the spliceosomes. Cold Spring Harbor, CSHL Press.

Burke, T. W. and J. T. Kadonaga (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." Genes & Development 10(6): 711-24.

Burns, C. G., R. Ohi, *et al.* (2002). "Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae." Molecular & Cellular Biology 22(3): 801-15.

Caprara, M. G. and T. W. Nilsen (2000). "RNA: versatility in form and function. [Review] [33 refs]." Nature Structural Biology 7(1072-8368): 831-833.

Causton, H. C., B. Ren, *et al.* (2001). "Remodeling of yeast genome expression in response to environmental changes." Molecular Biology of the Cell 12(2): 323-37.

Cavalier-Smith, T. (1991). "Intron phylogeny: a new hypothesis." Trends in Genetics 7(5): 145-8.

Chambers, A., C. Stanway, *et al.* (1990). "ARS binding factor 1 binds adjacent to RAP1 at the UASs of the yeast glycolytic genes PGK and PYK1." Nucleic Acids Research 18(18): 5393-9.

Chan, S. S., H. Zheng, *et al.* (1996). "UNC-40, a C. elegans homolog of DCC (Deleted in Colorectal Cancer), is required in motile cells responding to UNC-6 netrin cues." Cell 87(2): 187-95.

Chawla, G., A. K. Sapra, *et al.* (2003). "Dependence of pre-mRNA introns on PRP17, a non-essential splicing factor: implications for efficient progression through cell cycle transitions." Nucleic Acids Research 31(9): 2333-43.

Chen, B.-S., S. S. Mandal, *et al.* (2004). "High-resolution protein-DNA contacts for the yeast RNA polymerase II general transcription machinery." Biochemistry 43(40): 12741-9.

Chen, W. and K. Struhl (1988). "Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein." Proceedings of the National Academy of Sciences of the United States of America 85(8): 2691-5.

Chen, Y. and W. Stephan (2003). "Compensatory evolution of a precursor messenger RNA secondary structure in the Drosophila melanogaster Adh gene." Proceedings of the National Academy of Sciences of the United States of America 100(20): 11499-504.

Chi, H. C., D. Juminaga, *et al.* (1991). "Structure of the Drosophila gene for the laminin B2 chain." DNA & Cell Biology 10(6): 451-66.

Cho, G. and R. F. Doolittle (1997). "Intron distribution in ancient paralogs supports random insertion and not random loss.[erratum appears in J Mol Evol 1997 Aug;45(2):206]." Journal of Molecular Evolution 44(6): 573-84.

Cho, K. R., J. D. Oliner, *et al.* (1994). "The DCC gene: structural analysis and mutations in colorectal carcinomas." Genomics 19(3): 525-31.

Clark, T. A., C. W. Sugnet, *et al.* (2002). "Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays." Science 296(5569): 907-10.

Clement, J. Q., S. Maiti, *et al.* (2001). "Localization and stability of introns spliced from the Pem homeobox gene." Journal of Biological Chemistry 276(20): 16919-30.

Clement, J. Q., L. Qian, *et al.* (1999). "The stability and fate of a spliced intron from vertebrate cells." Rna-A Publication of the Rna Society 5(2): 206-20.

Cliften, P. F., L. W. Hillier, *et al.* (2001). "Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis.[see comment]." Genome Research 11(7): 1175-86.

Collins, B. M., L. Cubeddu, *et al.* (2003). "Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF." Journal of Biological Chemistry 278(19): 17291-8.

Collins, B. M., S. J. Harrop, *et al.* (2001). "Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs." Journal of Molecular Biology 309(4): 915-23.

Collins, C. A. and C. Guthrie (2000). "The question remains: is the spliceosome a ribozyme?" Nature Structural Biology 7(10): 850-4.

Collins, L. and D. Penny (2005). "Complex spliceosomal organization ancestral to extant eukaryotes." Molecular Biology & Evolution 22(4): 1053-66.

Cooney, M., G. Czernuszewicz, *et al.* (1988). "Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro." Science 241(4864): 456-9.

Corden, J. L. (1993). "RNA polymerase II transcription cycles." Current Opinion in Genetics & Development 3(2): 213-8.

Cousineau, B., S. Lawrence, *et al.* (2000). "Retrotransposition of a bacterial group II intron.[see comment][erratum appears in Nature 2001 Nov 1;414(6859):84]." Nature 404(6781): 1018-21.

Cramer, P., A. Srebrow, *et al.* (2001). "Coordination between transcription and pre-mRNA processing." FEBS Letters 498(2-3): 179-82.

Croft, L., S. Schandorff, *et al.* (2000). "ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome." Nature Genetics 24(4): 340-1.

Crooks, G. E., G. Hon, *et al.* (2004). "WebLogo: a sequence logo generator." Genome Research 14(6): 1188-90.

Dabeva, M. D. and J. R. Warner (1993). "Ribosomal protein L32 of Saccharomyces cerevisiae regulates both splicing and translation of its own transcript." Journal of Biological Chemistry 268(26): 19669-74.

Dahan, O. and M. Kupiec (2002). "Mutations in genes of Saccharomyces cerevisiae encoding pre-mRNA splicing factors cause cell cycle arrest through activation of the spindle checkpoint." Nucleic Acids Research 30(20): 4361-70.

Dahan, O. and M. Kupiec (2004). "The Saccharomyces cerevisiae gene CDC40/PRP17 controls cell cycle progression through splicing of the ANC1 gene." Nucleic Acids Research 32(8): 2529-40.

Damelin, M., I. Simon, *et al.* (2002). "The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress." Molecular Cell 9(3): 563-73.

Danin-Kreiselman, M., C. Y. Lee, *et al.* (2003). "RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns." Molecular Cell 11(5): 1279-89.

Darnell, J. E., Jr. (1978). "Implications of RNA. RNA Splicing in Evolution of Eukaryotic Cells." Science 202(4374): 1257-1260.

Davis, C. A., L. Grate, *et al.* (2000). "Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast." Nucleic Acids Research 28(8): 1700-6.

de Boinod, A. J. (2005). The Meaning of Tingo. London, Penguin Books.

DeRisi, J. L., V. R. Iyer, *et al.* (1997). "Exploring the metabolic and genetic control of gene expression on a genomic scale." Science 278(5338): 680-6.

Dermitzakis, E. T., A. Reymond, *et al.* (2003). "Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs).[see comment]." Science 302(5647): 1033-5.

DeSilva, U., L. Elnitski, *et al.* (2002). "Generation and comparative analysis of approximately 3.3 Mb of mouse genomic sequence orthologous to the region of human chromosome 7q11.23 implicated in Williams syndrome." Genome Research 12(1): 3-15.

Deutsch, M. and M. Long (1999). "Intron-exon structures of eukaryotic model organisms." Nucleic Acids Research 27(15): 3219-28.

Dickinson, J. (1999). Carbon metabolism. In: The Metabolism and Molecular Physiology of Saccharomyces cerevisiae. Philadelphia, PA, Taylor & Francis.

Dolinsky, T. J., P. M. J. Burgers, *et al.* (2004). "SPrCY: comparison of structural predictions in the Saccharomyces cerevisiae genome." Bioinformatics 20(14): 2312-4.

Donovan, D. M. and N. J. Pearson (1986). "Transcriptional regulation of ribosomal proteins during a nutritional upshift in Saccharomyces cerevisiae." Molecular & Cellular Biology 6(7): 2429-35.

Doolittle, W. F. (1978). "Genes in pieces: were they ever together?" Nature 272(5654): 581-582.

Doolittle, W. F. (1999). "Lateral genomics." Trends in Cell Biology 9(12): M5-8.

Du, H. and M. Rosbash (2001). "Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing." Rna-A Publication of the Rna Society 7(1): 133-42.

Dynan, W. S. and R. Tjian (1983). "The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter." Cell 35(1): 79-87.

Eddy, S. R. (1999). "Noncoding RNA genes." Current Opinion in Genetics & Development 9(6): 695-9.

Eddy, S. R. (2001). "Non-coding RNA genes and the modern RNA world." Nature Reviews Genetics 2(12): 919-29.

Eisen, M. B., P. T. Spellman, *et al.* (1998). "Cluster analysis and display of genome-wide expression patterns." Proceedings of the National Academy of Sciences of the United States of America 95(25): 14863-8.

Eng, F. J. and J. R. Warner (1991). "Structural basis for the regulation of splicing of a yeast messenger RNA." Cell 65(5): 797-804.

Engebrecht, J. A., K. Voelkel-Meiman, *et al.* (1991). "Meiosis-specific RNA splicing in yeast." Cell 66(6): 1257-68.

Eperon, L. P., I. R. Graham, *et al.* (1988). "Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?" Cell 54(3): 393-401.

Erdmann, V. A., M. Szymanski, *et al.* (2000). "Non-coding, mRNA-like RNAs database Y2K." Nucleic Acids Research 28(1): 197-200.

Fabrizio, P. and J. Abelson (1992). "Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing in vitro." Nucleic Acids Research 20(14): 3659-64.

Feaver, W. J., N. L. Henry, *et al.* (1994). "Yeast TFIIE. Cloning, expression, and homology to vertebrate proteins." Journal of Biological Chemistry 269(44): 27549-53.

Fedor, M. J., N. F. Lue, *et al.* (1988). "Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast." Journal of Molecular Biology 204(1): 109-27.

Fedorov, A., L. Fedorova, *et al.* (1998). "Influence of exon duplication on intron and exon phase distribution." Journal of Molecular Evolution 46(3): 263-71.

Fedorov, A., A. F. Merican, *et al.* (2002). "Large-scale comparison of intron positions among animal, plant, and fungal genes." Proceedings of the National Academy of Sciences of the United States of America 99(25): 16128-33.

Fedorov, A., S. Roy, *et al.* (2003). "Mystery of intron gain." Genome Research 13(10): 2236-41.

Fedorova, L. and A. Fedorov (2003). "Introns in gene evolution." Genetica 118(2-3): 123-31.

Fickett, J. W. and W. W. Wasserman (2000). "Discovery and modeling of transcriptional regulatory regions." Current Opinion in Biotechnology 11(1): 19-24.

Filipowicz, W., P. Pelczar, *et al.* (1999). "Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification." Acta Biochimica Polonica 46(2): 377-89.

Fong, Y. W. and Q. Zhou (2001). "Stimulatory effect of splicing factors on transcriptional elongation." Nature 414(6866): 929-33.

Fossella, J. A., Y. J. Kim, *et al.* (1993). "Relative specificities in binding of Watson-Crick base pairs by third strand residues in a DNA pyrimidine triplex motif." Nucleic Acids Research 21(19): 4511-5.

Fragapane, P., S. Prislei, *et al.* (1993). "A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA." EMBO Journal 12(7): 2921-8.

Frank-Kamenetskii, M. D. and S. M. Mirkin (1995). "Triplex DNA structures." Annual Review of Biochemistry 64: 65-95.

Furter-Graves, E. M. and B. D. Hall (1990). "DNA sequence elements required for transcription initiation of the Schizosaccharomyces pombe ADH gene in Saccharomyces cerevisiae." Molecular & General Genetics 223(3): 407-16.

Gailus-Durner, V., J. Xie, *et al.* (1996). "Participation of the yeast activator Abf1 in meiosis-specific expression of the HOP1 gene." Molecular & Cellular Biology 16(6): 2777-86.

Game, J. C., G. W. Birrell, *et al.* (2003). "Use of a genome-wide approach to identify new genes that control resistance of Saccharomyces cerevisiae to ionizing radiation." Radiation Research 160(1): 14-24.

Gasch, A. P., P. T. Spellman, *et al.* (2000). "Genomic expression programs in the response of yeast cells to environmental changes." Molecular Biology of the Cell 11(12): 4241-57.

Ge, H., Z. Liu, *et al.* (2001). "Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae.[see comment]." Nature Genetics 29(4): 482-6.

Gilbert, W. (1986). "Origin of life: The RNA world." Nature 319(6055): 618-618.

Gilbert, W. (1987). "The exon theory of genes." Cold Spring Harbor Symposia on Quantitative Biology 52: 901-5.

Giroux, M. J., M. Clancy, *et al.* (1994). "De novo synthesis of an intron by the maize transposable element Dissociation." Proceedings of the National Academy of Sciences of the United States of America 91(25): 12150-4.

Gish, W. R. (1996-2004).

Glazov, E. A., M. Pheasant, *et al.* (2005). "Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing." Genome Research 15(6): 800-8.

Goguel, V. and M. Rosbash (1993). "Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast." Cell 72(6): 893-901.

Goncalves, P. M., G. Griffioen, *et al.* (1995). "Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p." Nucleic Acids Research 23(9): 1475-80.

Gow, C. H., H. Y. Chang, *et al.* (1993). "Analysis of the Drosophila gene for the laminin B1 chain." DNA & Cell Biology 12(7): 573-87.

Grate, L. and M. Ares, Jr. (2002). "Searching yeast intron data at Ares lab Web site." Methods in Enzymology 350: 380-92.

Griffioen, G., R. J. Laan, *et al.* (1996). "Ribosomal protein gene transcription in Saccharomyces cerevisiae shows a biphasic response to nutritional changes." Microbiology 142(Pt 8): 2279-87.

Grigoriev, A. (2001). "A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae." Nucleic Acids Research 29(17): 3513-9.

Guerrier-Takada, C., K. Gardiner, *et al.* (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell 35(3 Pt 2): 849-57.

Hamm, J., M. Kazmaier, *et al.* (1987). "In vitro assembly of U1 snRNPs." EMBO Journal 6(11): 3479-85.

Hampsey, M. (2000). "RNA polymerase comes into focus." Trends in Genetics 16(1): 20.

Hampsey, M. (2001). RNA polymerase II transcriptional machinery. Oxford, Oxford University Press.

Hardison, R. C. (2000). "Conserved noncoding sequences are reliable guides to regulatory elements." Trends in Genetics 16(9): 369-72.

Hare, M. P. and S. R. Palumbi (2003). "High intron sequence conservation across three mammalian orders suggests functional constraints." Molecular Biology & Evolution 20(6): 969-78.

Harismendy, O., C.-G. Gendrel, *et al.* (2003). "Genome-wide location of yeast RNA polymerase III transcription machinery." EMBO Journal 22(18): 4738-47.

Hartmuth, K., H. Urlaub, *et al.* (2002). "Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method." Proceedings of the National Academy of Sciences of the United States of America 99(26): 16719-24.

Haugen, P., D. M. Simon, *et al.* (2005). "The natural history of group I introns." Trends in Genetics 21(2): 111-9.

Heard, E. (2004). "Recent advances in X-chromosome inactivation." Current Opinion in Cell Biology 16(3): 247-55.

Hebbes, T. R., A. L. Clayton, *et al.* (1994). "Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain." EMBO Journal 13(8): 1823-30.

Hermann, H., P. Fabrizio, *et al.* (1995). "snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions." EMBO Journal 14(9): 2076-88.

Hernandez, N. (1993). "TBP, a universal eukaryotic transcription factor?" Genes & Development 7(7B): 1291-308.

Hernandez, N. (2001). "Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription." Journal of Biological Chemistry 276(29): 26733-6.

Herruer, M. H., W. H. Mager, *et al.* (1987). "Transcriptional control of yeast ribosomal protein synthesis during carbon-source upshift." Nucleic Acids Research 15(24): 10133-44.

Hofacker, I. L. (2003). "Vienna RNA secondary structure server." Nucleic Acids Research 31(13): 3429-31.

Hoffman, C. S. and F. Winston (1987). "A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli." Gene 57(2-3): 267-72.

Holstege, F. C., E. G. Jennings, *et al.* (1998). "Dissecting the regulatory circuitry of a eukaryotic genome." Cell 95(5): 717-28.

Hoyne, P. R., A. M. Gacy, *et al.* (2000). "Stabilities of intrastrand pyrimidine motif DNA and RNA triple helices." Nucleic Acids Research 28(3): 770-5.

Hu, P., S. Wu, *et al.* (2004). "A role for beta-actin in RNA polymerase III transcription." Genes & Development 18(24): 3010-5.

Huber, J., U. Cronshagen, *et al.* (1998). "Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure." EMBO Journal 17(14): 4114-26.

Hughes, T. R., M. Mao, *et al.* (2001). "Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer." Nature Biotechnology 19(4): 342-7.

Imbalzano, A. N., H. Kwon, *et al.* (1994). "Facilitated binding of TATA-binding protein to nucleosomal DNA.[see comment]." Nature 370(6489): 481-5.

Ing, N. H., J. M. Beekman, *et al.* (1993). "In vivo transcription of a progesterone-responsive gene is specifically inhibited by a triplex-forming oligonucleotide." Nucleic Acids Research 21(12): 2789-96.

Iyer, V. and K. Struhl (1995). "Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure." EMBO Journal 14(11): 2570-9.

Jacob, F. M., J. (1961). "Genetic regulatory mechanisms in the synthesis of proteins." Journal of Molecular Biology(3): 318-356.

Jandrositz, A. and C. Guthrie (1995). "Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs." EMBO Journal 14(4): 820-32.

Jeffares, D. C., A. M. Poole, *et al.* (1997). "Relics from the RNA world. [Review] [164 refs]." Journal of Molecular Evolution 46(0022-2844): 18-36.

Jeffares, D. C., A. M. Poole, *et al.* (1998). "Relics from the RNA world." Journal of Molecular Evolution 46(1): 18-36.

Johnston, W. K., P. J. Unrau, *et al.* (2001). "RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension.[see comment]." Science 292(5520): 1319-25.

Jones, M. H., D. N. Frank, *et al.* (1995). "Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast." Proceedings of the National Academy of Sciences of the United States of America 92(21): 9687-91.

Joyce, G. F. (1989). "RNA evolution and the origins of life." Nature 338(6212): 217-24.

Joyce, G. F. (1991). "The rise and fall of the RNA world." New Biologist 3(4): 399-407.

Joyce, G. F. (1999). Reactions catalyzed by RNA and DNA enzymes. New York, Cold Spring Harbor Lab Press.

Joyce, G. F. (2002). "The antiquity of RNA-based evolution." Nature 418(6894): 214-21.

Ju, Q. and J. R. Warner (1994). "Ribosome synthesis during the growth cycle of Saccharomyces cerevisiae.[erratum appears in Yeast 1994 Jul;10(7):979-80]." Yeast 10(2): 151-7.

Jurica, M. S. and M. J. Moore (2003). "Pre-mRNA splicing: awash in a sea of proteins." Molecular Cell 12(1): 5-14.

Kadosh, D. and K. Struhl (1998). "Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo." Molecular & Cellular Biology 18(9): 5121-7.

Kambach, C., S. Walke, *et al.* (1999). "Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs." Cell 96(3): 375-87.

Kandels-Lewis, S. and B. Seraphin (1993). "Involvement of U6 snRNA in 5' splice site selection.[see comment][erratum appears in Science 1994 Feb 11;263(5148):739]." Science 262(5142): 2035-9.

Karolchik, D., R. Baertsch, *et al.* (2003). "The UCSC Genome Browser Database." Nucleic Acids Research 31(1): 51-4.

Kelley, R. L. and M. I. Kuroda (2000). "The role of chromosomal RNAs in marking the X for dosage compensation." Current Opinion in Genetics & Development 10(5): 555-61.

Kellis, M., N. Patterson, *et al.* (2003). "Sequencing and comparison of yeast species to identify genes and regulatory elements.[see comment]." Nature 423(6937): 241-54.

Khusial, P., R. Plaag, *et al.* (2005). "LSm proteins form heptameric rings that bind to RNA via repeating motifs." Trends in Biochemical Sciences 30(9): 522-8.

Kim, H. G. and D. M. Miller (1995). "Inhibition of in vitro transcription by a triplex-forming oligonucleotide targeted to human c-myc P2 promoter." Biochemistry 34(25): 8165-71.

Kingston, R. E. (1999). "A shared but complex bridge.[comment]." Nature 399(6733): 199-200.

Kiss, T. (2001). "Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs." EMBO Journal 20(14): 3617-22.

Kiss, T. and W. Filipowicz (1993). "Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1." EMBO Journal 12(7): 2913-20.

Klein, C. and K. Struhl (1994). "Protein kinase A mediates growth-regulated expression of yeast ribosomal protein genes by modulating RAP1 transcriptional activity." Molecular & Cellular Biology 14(3): 1920-8.

Kooi, E. A., C. A. Rutgers, *et al.* (1994). "Mutational analysis of the C-terminal region of Saccharomyces cerevisiae ribosomal protein L25 in vitro and in vivo demonstrates the presence of two distinct functional elements." Journal of Molecular Biology 240(3): 243-55.

Kornberg, R. D. and Y. Lorch (1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome." Cell 98(3): 285-94.

Kovacs, A., J. C. Kandala, *et al.* (1996). "Triple helix-forming oligonucleotide corresponding to the polypyrimidine sequence in the rat alpha 1(I) collagen promoter specifically inhibits factor binding and transcription." Journal of Biological Chemistry 271(3): 1805-12.

Kraig, E., J. E. Haber, *et al.* (1982). "Sporulation and rna2 lower ribosomal protein mRNA levels by different mechanisms in Saccharomyces cerevisiae." Molecular & Cellular Biology 2(10): 1199-204.

Kruger, K., P. J. Grabowski, *et al.* (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." Cell 31(1): 147-57.

Kufel, J., C. Allmang, *et al.* (2003). "Lsm Proteins are required for normal processing and stability of ribosomal RNAs." Journal of Biological Chemistry 278(4): 2147-56.

Kufel, J., C. Allmang, *et al.* (2003). "A complex pathway for 3' processing of the yeast U3 snoRNA." Nucleic Acids Research 31(23): 6788-97.

Kufel, J., C. Allmang, *et al.* (2002). "Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p." Molecular & Cellular Biology 22(14): 5248-56.

Kuhn, K. M., J. L. DeRisi, *et al.* (2001). "Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source." Molecular & Cellular Biology 21(3): 916-27.

Kulkens, T., C. A. van der Sande, *et al.* (1992). "A system to study transcription by yeast RNA polymerase I within the chromosomal context: functional analysis of the ribosomal DNA enhancer and the RBP1/REB1 binding sites." EMBO Journal 11(12): 4665-74.

Kuo, M. H., J. Zhou, *et al.* (1998). "Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo." Genes & Development 12(5): 627-39.

Kuras, L. and K. Struhl (1999). "Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme." Nature 399(6736): 609-13.

Ladomery, M. (1997). "Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes." Bioessays 19(10): 903-9.

Lagos-Quintana, M., R. Rauhut, *et al.* (2001). "Identification of novel genes coding for small expressed RNAs.[see comment]." Science 294(5543): 853-8.

Lagos-Quintana, M., R. Rauhut, *et al.* (2002). "Identification of tissue-specific microRNAs from mouse." Current Biology 12(9): 735-9.

Lamond, A. I. (1993). "The spliceosome." Bioessays 15(9): 595-603.

Lamond, A. I., M. M. Konarska, *et al.* (1988). "Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein." Proceedings of the National Academy of Sciences of the United States of America 85(2): 411-5.

Lander, E. S., L. M. Linton, *et al.* (2001). "Initial sequencing and analysis of the human genome.[see comment][erratum appears in Nature 2001 Aug 2;412(6846):565 Note: Szustakowki, J [corrected to Szustakowski, J]]." Nature 409(6822): 860-921.

Lau, N. C., L. P. Lim, *et al.* (2001). "An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans.[see comment]." Science 294(5543): 858-62.

Le Hir, H., A. Nott, *et al.* (2003). "How introns influence and enhance eukaryotic gene expression." Trends in Biochemical Sciences 28(4): 215-20.

Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in Caenorhabditis elegans.[see comment]." Science 294(5543): 862-4.

Lee, T. I., N. J. Rinaldi, *et al.* (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae.[see comment]." Science 298(5594): 799-804.

Lee, T. I. and R. A. Young (2000). "Transcription of eukaryotic protein-coding genes." Annual Review of Genetics 34: 77-137.

Lerner, M. R. and J. A. Steitz (1979). "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." Proceedings of the National Academy of Sciences of the United States of America 76(11): 5495-9.

Lesser, C. F. and C. Guthrie (1993). "Mutations in U6 snRNA that alter splice site specificity: implications for the active site.[see comment]." Science 262(5142): 1982-8.

Letai, A. G., M. A. Palladino, *et al.* (1988). "Specificity in formation of triple-stranded nucleic acid helical complexes: studies with agarose-linked polyribonucleotide affinity columns." Biochemistry 27(26): 9108-12.

Levy, S., S. Hannenhalli, *et al.* (2001). "Enrichment of regulatory signals in conserved non-coding genomic sequence." Bioinformatics 17(10): 871-7.

Li, B., J. Vilardell, *et al.* (1996). "An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness." Proceedings of the National Academy of Sciences of the United States of America 93(4): 1596-600.

Li, H., S. Dalal, *et al.* (1995). "Characterization of the pre-mRNA binding site for yeast ribosomal protein L32: the importance of a purine-rich internal loop." Journal of Molecular Biology 250(4): 447-59.

Li, X. Y., A. Virbasius, *et al.* (1999). "Enhancement of TBP binding by activators and general transcription factors." Nature 399(6736): 605-9.

Li, Y., P. M. Flanagan, *et al.* (1994). "RNA polymerase II initiation factor interactions and transcription start site selection." Science 263(5148): 805-7.

Li, Z., A. G. Paulovich, *et al.* (1995). "Feedback inhibition of the yeast ribosomal protein gene CRY2 is mediated by the nucleotide sequence and secondary structure of CRY2 pre-mRNA." Molecular & Cellular Biology 15(11): 6454-64.

Liakopoulos, D., G. Doenges, *et al.* (1998). "A novel protein modification pathway related to the ubiquitin system." EMBO Journal 17(8): 2208-14.

Liaw, P. C. and C. J. Brandl (1994). "Defining the sequence specificity of the Saccharomyces cerevisiae DNA binding protein REB1p by selecting binding sites from random-sequence oligonucleotides." Yeast 10(6): 771-87.

Lie, Y. S. and P. M. Macdonald (1999). "Apontic binds the translational repressor Bruno and is implicated in regulation of oskar mRNA translation." Development 126(6): 1129-38.

Lieb, J. D., X. Liu, *et al.* (2001). "Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association.[see comment][erratum appears in Nat Genet 2001 Sep;29(1):100]." Nature Genetics 28(4): 327-34.

Logsdon, J. M., Jr. (1998). "The recent origins of spliceosomal introns revisited." Current Opinion in Genetics & Development 8(6): 637-48.

Logsdon, J. M., Jr., M. G. Tyshenko, *et al.* (1995). "Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory." Proceedings of the National Academy of Sciences of the United States of America 92(18): 8507-11.

Long, M., S. J. de Souza, *et al.* (1998). "Relationship between "proto-splice sites" and intron phases: evidence from dicodon analysis." Proceedings of the National Academy of Sciences of the United States of America 95(1): 219-23.

Loots, G. G., R. M. Locksley, *et al.* (2000). "Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons.[see comment]." Science 288(5463): 136-40.

Lopez, M. C. and H. V. Baker (2000). "Understanding the growth phenotype of the yeast gcr1 mutant in terms of global genomic expression patterns." Journal of Bacteriology 182(17): 4970-8.

Luhrmann, R., B. Kastner, *et al.* (1990). "Structure of spliceosomal snRNPs and their role in pre-mRNA splicing." Biochimica et Biophysica Acta 1087(3): 265-92.

Luscombe, N. M., S. E. Austin, *et al.* (2000). "An overview of the structures of protein-DNA complexes." Genome Biology 1(1): REVIEWS001.

Lynch, M. and A. O. Richardson (2002). "The evolution of spliceosomal introns." Current Opinion in Genetics & Development 12(6): 701-10.

Macy, J. M., Miller, M.W. (1983). "Anaerobic growth of Saccharomyces cerevisiae in the absence of oleic acid ergosterol." Academy of Microbiology(134): 64-67.

Madhani, H. D. and C. Guthrie (1992). "A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome." Cell 71(5): 803-17.

Mager, W. H. (1988). "Control of ribosomal protein gene expression." Biochimica et Biophysica Acta 949(1): 1-15.

Mager, W. H. and R. J. Planta (1990). "Multifunctional DNA-binding proteins mediate concerted transcription activation of yeast ribosomal protein genes." Biochimica et Biophysica Acta 1050(1-3): 351-5.

Maldonado, E., M. Hampsey, *et al.* (1999). "Repression: targeting the heart of the matter." Cell 99(5): 455-8.

Maniatis, T. and R. Reed (2002). "An extensive network of coupling among gene expression machines." Nature 416(6880): 499-506.

Manley, J. L. (2002). "Nuclear coupling: RNA processing reaches back to transcription.[comment]." Nature Structural Biology 9(11): 790-1.

Maquat, L. E. and G. G. Carmichael (2001). "Quality control of mRNA function." Cell 104(2): 173-6.

Martens, J. A. and F. Winston (2002). "Evidence that Swi/Snf directly represses transcription in S. cerevisiae." Genes & Development 16(17): 2231-6.

Martens, J. A. and F. Winston (2003). "Recent advances in understanding chromatin remodeling by Swi/Snf complexes." Current Opinion in Genetics & Development 13(2): 136-42.

Mattick, J. S. (2001). "Non-coding RNAs: the architects of eukaryotic complexity." EMBO Reports 2(11): 986-91.

Mattick, J. S. (2003). "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms." Bioessays 25(10): 930-9.

Mattick, J. S. and M. J. Gagen (2001). "The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms." Molecular Biology & Evolution 18(9): 1611-30.

Mattick, J. S. and I. V. Makunin (2005). "Small regulatory RNAs in mammals." Human Molecular Genetics 14 Spec No 1: R121-32.

Mayes, A. E., L. Verdone, *et al.* (1999). "Characterization of Sm-like proteins in yeast and their association with U6 snRNA." EMBO Journal 18(15): 4321-31.

McGuire, A. M., J. D. Hughes, *et al.* (2000). "Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes." Genome Research 10(6): 744-57.

McKnight, S. L., Yamamoto, K.R. (1992). Transcriptional Regulation. Cold Spring Harbor, New York, Cold Spring Harbor Lab Press.

McKune, K., P. A. Moore, *et al.* (1995). "Six human RNA polymerase subunits functionally substitute for their yeast counterparts." Molecular & Cellular Biology 15(12): 6895-900.

Milne, A. A. (2002). The Complete Winnie-the-Pooh. London, Dean.

Moehle, C. M. and A. G. Hinnebusch (1991). "Association of RAP1 binding sites with stringent control of ribosomal protein gene transcription in Saccharomyces cerevisiae." Molecular & Cellular Biology 11(5): 2723-35.

Moller, T., T. Franch, *et al.* (2002). "Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction." Molecular Cell 9(1): 23-30.

Montagne, J., M. J. Stewart, *et al.* (1999). "Drosophila S6 kinase: a regulator of cell size.[see comment]." Science 285(5436): 2126-9.

Moore, M. J., Query, C.C., Sharp, P.A. (1993). Splicing of precursors to mRNA by the spliceosome. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Morrow, B. E., S. P. Johnson, *et al.* (1993). "The rRNA enhancer regulates rRNA transcription in Saccharomyces cerevisiae." Molecular & Cellular Biology 13(2): 1283-9.

Moser, H. E. and P. B. Dervan (1987). "Sequence-specific cleavage of double helical DNA by triple helix formation." Science 238(4827): 645-50.

Muscat, G. E., L. J. Burke, *et al.* (1998). "The corepressor N-CoR and its variants RIP13a and RIP13Delta1 directly interact with the basal transcription factors TFIIB, TAFII32 and TAFII70." Nucleic Acids Research 26(12): 2899-907.

Myers, L. C. and R. D. Kornberg (2000). "Mediator of transcriptional regulation." Annual Review of Biochemistry 69: 729-49.

Nakagawa, T. and H. Ogawa (1999). "The Saccharomyces cerevisiae MER3 gene, encoding a novel helicase-like protein, is required for crossover control in meiosis." EMBO Journal 18(20): 5714-23.

Nasmyth, K. (1985). "At least 1400 base pairs of 5'-flanking DNA is required for the correct expression of the HO gene in yeast." Cell 42(1): 213-23.

Neuman-Silberberg, F. S., S. Bhattacharya, *et al.* (1995). "Nutrient availability and the RAS/cyclic AMP pathway both induce expression of ribosomal protein genes in Saccharomyces cerevisiae but by different mechanisms." Molecular & Cellular Biology 15(6): 3187-96.

Newman, A. and C. Norman (1991). "Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage." Cell 65(1): 115-23.

Newman, A. J. and C. Norman (1992). "U5 snRNA interacts with exon sequences at 5' and 3' splice sites." Cell 68(4): 743-54.

Nikolov, D. B., H. Chen, *et al.* (1995). "Crystal structure of a TFIIB-TBP-TATA-element ternary complex." Nature 377(6545): 119-28.

Nilsen, T. W. (2002). "The spliceosome: no assembly required?[comment]." Molecular Cell 9(1): 8-9.

Nilsen, T. W. (2003). "The spliceosome: the most complex macromolecular machine in the cell?" Bioessays 25(12): 1147-9.

Nixon, J. E. J., A. Wang, *et al.* (2002). "A spliceosomal intron in Giardia lamblia.[see comment]." Proceedings of the National Academy of Sciences of the United States of America 99(6): 3701-5.

Noller, H. F. (1993). "Peptidyl transferase: protein, ribonucleoprotein, or RNA?" Journal of Bacteriology 175(17): 5297-300.

Noller, H. F., V. Hoffarth, *et al.* (1992). "Unusual resistance of peptidyl transferase to protein extraction procedures.[see comment]." Science 256(5062): 1416-9.

Noonberg, S. B., G. K. Scott, *et al.* (1994). "Inhibition of transcription factor binding to the HER2 promoter by triplex-forming oligodeoxyribonucleotides." Gene 149(1): 123-6.

Nottrott, S., H. Urlaub, *et al.* (2002). "Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins." EMBO Journal 21(20): 5527-38.

Ogbourne, S. and T. M. Antalis (1998). "Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes." Biochemical Journal 331(Pt 1): 1-14.

Ohno, M., T. Fukagawa, *et al.* (2002). "Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies." Chromosoma 111(3): 201-13.

Orphanides, G. and D. Reinberg (2002). "A unified theory of gene expression." Cell 108(4): 439-51.

Padgett, R. A., P. J. Grabowski, *et al.* (1986). "Splicing of messenger RNA precursors." Annual Review of Biochemistry 55: 1119-50.

Palacios, I., M. Hetzer, *et al.* (1997). "Nuclear import of U snRNPs requires importin beta." EMBO Journal 16(22): 6783-92.

Palmer, J. D. and J. M. Logsdon, Jr. (1991). "The recent origins of introns." Current Opinion in Genetics & Development 1(4): 470-7.

Papp, B., C. Pal, *et al.* (2003). "Dosage sensitivity and the evolution of gene families in yeast." Nature 424(6945): 194-7.

Papp, B., C. Pal, *et al.* (2004). "Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast." Nature 429(6992): 661-4.

Parker, R., P. G. Siliciano, *et al.* (1987). "Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA." Cell 49(2): 229-39.

Patel, A. A. and J. A. Steitz (2003). "Splicing double: insights from the second spliceosome." Nature Reviews Molecular Cell Biology 4(12): 960-70.

Paulovich, A. G., J. R. Thompson, *et al.* (1993). "Molecular genetics of cryptopleurine resistance in Saccharomyces cerevisiae: expression of a ribosomal protein gene family." Genetics 135(3): 719-30.

Pikielny, C. W. and M. Rosbash (1985). "mRNA splicing efficiency in yeast and the contribution of nonconserved sequences." Cell 41(1): 119-26.

Planta, R. J. (1997). "Regulation of ribosome synthesis in yeast." Yeast 13(16): 1505-18.

Planta, R. J. and W. H. Mager (1998). "The list of cytoplasmic ribosomal proteins of Saccharomyces cerevisiae." Yeast 14(5): 471-7.

Planta, R. J. and H. A. Raue (1988). "Control of ribosome biogenesis in yeast." Trends in Genetics 4(3): 64-8.

Plath, K., J. Fang, *et al.* (2003). "Role of histone H3 lysine 27 methylation in X inactivation." Science 300(5616): 131-5.

Plessel, G., U. Fischer, *et al.* (1994). "m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain." Molecular & Cellular Biology 14(6): 4160-72.

Poole, A. M., D. C. Jeffares, *et al.* (1998). "The path from the RNA world. [Review] [121 refs]." (0022-2844).

Preker, P. J., K. S. Kim, *et al.* (2002). "Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism." Rna-A Publication of the Rna Society 8(8): 969-80.

Prieto, S., B. J. de la Cruz, *et al.* (2000). "Glucose-regulated turnover of mRNA and the influence of poly(A) tail length on half-life." Journal of Biological Chemistry 275(19): 14155-66.

Prislei, S., A. Michienzi, *et al.* (1993). "Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes." Nucleic Acids Research 21(25): 5824-30.

Prislei, S., S. Sperandio, *et al.* (1992). "The mechanisms controlling ribosomal protein L1 pre-mRNA splicing are maintained in evolution and rely on conserved intron sequences." Nucleic Acids Research 20(17): 4473-9.

Proudfoot, N. J., A. Furger, *et al.* (2002). "Integrating mRNA processing with transcription." Cell 108(4): 501-12.

Ptashne, M. (1986). "Gene regulation by proteins acting nearby and at a distance." Nature 322(6081): 697-701.

Ptashne, M. and A. Gann (1997). "Transcriptional activation by recruitment." Nature 386(6625): 569-77.

Radhakrishnan, I. and D. J. Patel (1994). "DNA triplexes: solution structures, hydration sites, energetics, interactions, and function." Biochemistry 33(38): 11405-16.

Raghaven, S., Burma, P.K. & Brahmachari (1997). "Positional preferences for polypurine/polypyrimidine tracts in *Saccharomyces cerevisiae* genome: implications for *cis* regulation of gene expression." Journal of Molecular Biology 45: 485-498.

Raker, V. A., K. Hartmuth, *et al.* (1999). "Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner." Molecular & Cellular Biology 19(10): 6554-65.

Raker, V. A., G. Plessel, *et al.* (1996). "The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro." EMBO Journal 15(9): 2256-69.

Raue, H. A. and R. J. Planta (1991). "Ribosome biogenesis in yeast." Progress in Nucleic Acid Research & Molecular Biology 41: 89-129.

Rebane, A., R. Tamme, *et al.* (1998). "A novel snoRNA (U73) is encoded within the introns of the human and mouse ribosomal protein S3a genes." Gene 210(2): 255-63.

Reed, R. (1996). "Initial splice-site recognition and pairing during pre-mRNA splicing." Current Opinion in Genetics & Development 6(2): 215-20.

Reese, J. C., L. Apone, *et al.* (1994). "Yeast TAFIIS in a multisubunit complex required for activated transcription.[see comment]." Nature 371(6497): 523-7.

Robinson, K. A. and J. M. Lopes (2000). "The promoter of the yeast INO4 regulatory gene: a model of the simplest yeast promoter." Journal of Bacteriology 182(10): 2746-52.

Roeder, R. G. (1991). "The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly." Trends in Biochemical Sciences 16(11): 402-8.

Roeder, R. G. (1996). "The role of general initiation factors in transcription by RNA polymerase II." Trends in Biochemical Sciences 21(9): 327-35.

Rogozin, I. B., Y. I. Wolf, *et al.* (2003). "Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution." Current Biology 13(17): 1512-7.

Rottensteiner, H., A. J. Kal, *et al.* (1997). "A heterodimer of the Zn2Cys6 transcription factors Pip2p and Oaf1p controls induction of genes encoding peroxisomal proteins in Saccharomyces cerevisiae." European Journal of Biochemistry 247(3): 776-83.

Roy, R., L. Schaeffer, *et al.* (1994). "The DNA-dependent ATPase activity associated with the class II basic transcription factor BTF2/TFIIH." Journal of Biological Chemistry 269(13): 9826-32.

Ruby, S. W. and J. Abelson (1988). "An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly." Science 242(4881): 1028-35.

Rundlett, S. E., A. A. Carmen, *et al.* (1998). "Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3." Nature 392(6678): 831-5.

Russell, C. S., S. Ben-Yehuda, *et al.* (2000). "Functional analyses of interacting factors involved in both pre-mRNA splicing and cell cycle progression in Saccharomyces cerevisiae." Rna-A Publication of the Rna Society 6(11): 1565-72.

Ryan, D. E., S. W. Stevens, *et al.* (2002). "The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region." Rna-A Publication of the Rna Society 8(8): 1011-33.

Sakharkar, M. K., V. T. K. Chow, *et al.* (2004). "Distributions of exons and introns in the human genome." In Silico Biology 4(4): 387-93.

Salgado-Garrido, J., E. Bragado-Nilsson, *et al.* (1999). "Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin." EMBO Journal 18(12): 3451-62.

Sambrook, J., Fritsch, F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Schaeffer, L., R. Roy, *et al.* (1993). "DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor.[see comment]." Science 260(5104): 58-63.

Schawalder, S. B., M. Kabani, *et al.* (2004). "Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1." Nature 432(7020): 1058-61.

Scheffler, I. E., B. J. de la Cruz, *et al.* (1998). "Control of mRNA turnover as a mechanism of glucose repression in Saccharomyces cerevisiae." International Journal of Biochemistry & Cell Biology 30(11): 1175-93.

Scherer, S. W., J. Cheung, *et al.* (2003). "Human chromosome 7: DNA sequence and biology." Science 300(5620): 767-72.

Schluenzen, F., A. Tocilj, *et al.* (2000). "Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution." Cell 102(5): 615-23.

Schwartz, D. C. and R. Parker (2000). "mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E." Molecular & Cellular Biology 20(21): 7933-42.

Scott, E. W. and H. V. Baker (1993). "Concerted action of the transcriptional activators REB1, RAP1, and GCR1 in the high-level expression of the glycolytic gene TPI." Molecular & Cellular Biology 13(1): 543-50.

Scott, M. P., P. Matsudaira, *et al.* (2003). Molecular Cell Biology. New York, W H Freeman & Co.

Selinger, D. W., K. J. Cheung, *et al.* (2000). "RNA expression analysis using a 30 base pair resolution Escherichia coli genome array.[see comment]." Nature Biotechnology 18(12): 1262-8.

Seraphin, B. (1995). "Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs." EMBO Journal 14(9): 2089-98.

Seraphin, B. and M. Rosbash (1989). "Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing." Cell 59(2): 349-58.

Shannon, K. W. and C. Guthrie (1991). "Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs." Genes & Development 5(5): 773-85.

Sharp, P. A. (1992). "TATA-binding protein is a classless factor." Cell 68(5): 819-21.

Shi, Y. and J. M. Berg (1995). "Specific DNA-RNA hybrid binding by zinc finger proteins." Science 268(5208): 282-4.

Shilatifard, A. (2004). "Transcriptional elongation control by RNA polymerase II: a new frontier." Biochimica et Biophysica Acta 1677(1-3): 79-86.

Shore, D. (1994). "RAP1: a protean regulator in yeast." Trends in Genetics 10(11): 408-12.

Silva, J., W. Mak, *et al.* (2003). "Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes.[see comment]." Developmental Cell 4(4): 481-95.

Simonis, N., J. van Helden, *et al.* (2004). "Transcriptional regulation of protein complexes in yeast." Genome Biology 5(5): R33.

Sinclair, D. A., G. D. Kornfeld, *et al.* (1994). "Yeast intragenic transcriptional control: activation and repression sites within the coding region of the Saccharomyces cerevisiae LPD1 gene." Molecular & Cellular Biology 14(1): 214-25.

Singh, R. (2002). "RNA-protein interactions that regulate pre-mRNA splicing." Gene Expression 10(1-2): 79-92.

Sleutels, F., R. Zwart, *et al.* (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature 415(6873): 810-3.

Souciet, J., M. Aigle, *et al.* (2000). "Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies." FEBS Letters 487(1): 3-12.

Spingola, M. and M. Ares, Jr. (2000). "A yeast intronic splicing enhancer and Nam8p are required for Mer1p-activated splicing." Molecular Cell 6(2): 329-38.

Spingola, M., L. Grate, *et al.* (1999). "Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae." Rna-A Publication of the Rna Society 5(2): 221-34.

Spirin, A. S. (2002). "Omnipotent RNA." FEBS Letters 530(1-3): 4-8.

Staley, J. P. and C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell 92(3): 315-26.

Steinmetz, L. M., C. Scharfe, *et al.* (2002). "Systematic screen for human disease genes in yeast." Nature Genetics 31(4): 400-4.

Stevens, R. C. and T. N. Davis (1998). "Mlc1p is a light chain for the unconventional myosin Myo2p in Saccharomyces cerevisiae." Journal of Cell Biology 142(3): 711-22.

Stevens, S. W., D. E. Ryan, *et al.* (2002). "Composition and functional characterization of the yeast spliceosomal penta-snRNP.[see comment]." Molecular Cell 9(1): 31-44.

Storz, G. (2002). "An expanding universe of noncoding RNAs." Science 296(5571): 1260-3.

Struhl, K. (1989). "Molecular mechanisms of transcriptional regulation in yeast." Annual Review of Biochemistry 58: 1051-77.

Struhl, K. (1993). "Yeast transcription factors." Current Opinion in Cell Biology 5(3): 513-20.

Struhl, K. (1995). "Yeast transcriptional regulatory mechanisms." Annual Review of Genetics 29: 651-74.

Struhl, K. (1999). "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." Cell 98(1): 1-4.

Sverdlov, A. V., I. B. Rogozin, *et al.* (2005). "Conservation versus parallel gains in intron evolution.[erratum appears in Nucleic Acids Res. 2005;33(17):5713]." Nucleic Acids Research 33(6): 1741-8.

Tanay, A., A. Regev, *et al.* (2005). "Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast." Proceedings of the National Academy of Sciences of the United States of America 102(20): 7203-8.

Tang, T.-H., J.-P. Bachellerie, *et al.* (2002). "Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus." Proceedings of the National Academy of Sciences of the United States of America 99(11): 7536-41.

Tarn, W. Y. and J. A. Steitz (1996). "Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns.[see comment]." Science 273(5283): 1824-32.

Tarn, W. Y. and J. A. Steitz (1997). "Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge." Trends in Biochemical Sciences 22(4): 132-7.

Teichmann, S. A. and M. M. Babu (2002). "Conservation of gene co-regulation in prokaryotes and eukaryotes." Trends in Biotechnology 20(10): 407-10; discussion 410.

Tennyson, C. N., H. J. Klamut, *et al.* (1995). "The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced." Nature Genetics 9(2): 184-90.

ter Linde, J. J., H. Liang, *et al.* (1999). "Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of Saccharomyces cerevisiae." Journal of Bacteriology 181(24): 7409-13.

Tharun, S., W. He, *et al.* (2000). "Yeast Sm-like proteins function in mRNA decapping and decay." Nature 404(6777): 515-8.

Thijs, G., M. Lescot, *et al.* (2001). "A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling." Bioinformatics 17(12): 1113-22.

Thijs, G., K. Marchal, *et al.* (2002). "A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes." Journal of Computational Biology 9(2): 447-64.

Thomas, G. (2000). "An encore for ribosome biogenesis in the control of cell proliferation." Nature Cell Biology 2(5): E71-2.

Thomas, G. and M. N. Hall (1997). "TOR signalling and control of cell growth." Current Opinion in Cell Biology 9(6): 782-7.

Tomasevic, N. and B. A. Peculis (2002). "Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence." Molecular & Cellular Biology 22(12): 4101-12.

Toro, I., J. Basquin, *et al.* (2002). "Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus." Journal of Molecular Biology 320(1): 129-42.

Toro, I., S. Thore, *et al.* (2001). "RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex." EMBO Journal 20(9): 2293-303.

Travers, A. A. (2003). "Priming the nucleosome: a role for HMGB proteins?" EMBO Reports 4(2): 131-6.

Tycowski, K. T., Kolev, N.G., Conrad, N.K., Fok, V. and Steitz J.A. (2006). The Ever-Growing World of Small Nuclear Ribonucleoproteins. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Tycowski, K. T., M. D. Shu, *et al.* (1993). "A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3." Genes & Development 7(7A): 1176-90.

Tycowski, K. T., Z. H. You, *et al.* (1998). "Modification of U6 spliceosomal RNA is guided by other small RNAs." Molecular Cell 2(5): 629-38.

Vaisman, N., A. Tsouladze, *et al.* (1995). "The role of Saccharomyces cerevisiae Cdc40p in DNA replication and mitotic spindle formation and/or maintenance." Molecular & General Genetics 247(2): 123-36.

van Helden, J., B. Andre, *et al.* (2000). "A web site for the computational analysis of yeast regulatory sequences." Yeast 16(2): 177-87.

Varani, L., M. Hasegawa, *et al.* (1999). "Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17." Proceedings of the National Academy of Sciences of the United States of America 96(14): 8229-34.

Vasiliev, V. D., O. M. Selivanova, *et al.* (1978). "Specific selfpacking of the ribosomal 16 S RNA." FEBS Letters 95(2): 273-6.

Velculescu, V. E., L. Zhang, *et al.* (1997). "Characterization of the yeast transcriptome." Cell 88(2): 243-51.

Venter, J. C., M. D. Adams, *et al.* (2001). "The sequence of the human genome.[see comment][erratum appears in Science 2001 Jun 5;292(5523):1838]." Science 291(5507): 1304-51.

Verdier, J. M., R. Stalder, *et al.* (1990). "Preparation and characterization of yeast nuclear extracts for efficient RNA polymerase B (II)-dependent transcription in vitro." Nucleic Acids Research 18(23): 7033-9.

Verdone, L., S. Galardi, *et al.* (2004). "Lsm proteins promote regeneration of pre-mRNA splicing activity." Current Biology 14(16): 1487-91.

Vidal, V. P., L. Verdone, *et al.* (1999). "Characterization of U6 snRNA-protein interactions." Rna-A Publication of the Rna Society 5(11): 1470-81.

Vilardell, J. and J. R. Warner (1994). "Regulation of splicing at an intermediate step in the formation of the spliceosome." Genes & Development 8(2): 211-20.

Vincent, A. C. and K. Struhl (1992). "ACR1, a yeast ATF/CREB repressor." Molecular & Cellular Biology 12(12): 5394-405.

von Ahsen, U. and H. F. Noller (1995). "Identification of bases in 16S rRNA essential for tRNA binding at the 30S ribosomal P site." Science 267(5195): 234-7.

Vuolteenaho, R., L. T. Chow, *et al.* (1990). "Structure of the human laminin B1 chain gene." Journal of Biological Chemistry 265(26): 15611-6.

Wade, J. T., D. B. Hall, *et al.* (2004). "The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes." Nature 432(7020): 1054-8.

Walker, S. S., W. C. Shen, *et al.* (1997). "Yeast TAF(II)145 required for transcription of G1/S cyclin genes and regulated by the cellular growth state." Cell 90(4): 607-14.

Warner, J. R. (1999). "The economics of ribosome biosynthesis in yeast." Trends in Biochemical Sciences 24(11): 437-40.

Wassarman, D. A. and J. A. Steitz (1992). "Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing." Science 257(5078): 1918-25.

Watson, J. D. a. C., F.H.C. (1953). Nature(171): 964-967.

Watson, J. D. a. C., F.H.C. (1953). Nature(171): 737-738.

Weil, P. A., D. S. Luse, *et al.* (1979). "Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA." Cell 18(2): 469-84.

Weis, L. and D. Reinberg (1992). "Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes." FASEB Journal 6(14): 3300-9.

Werner-Washburne, M., E. L. Braun, *et al.* (1996). "Stationary phase in Saccharomyces cerevisiae." Molecular Microbiology 19(6): 1159-66.

Will, C. L. and R. Luhrmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Current Opinion in Cell Biology 13(3): 290-301.

Wimberly, B. T., D. E. Brodersen, *et al.* (2000). "Structure of the 30S ribosomal subunit.[see comment]." Nature 407(6802): 327-39.

Wingender, E., P. Dietze, *et al.* (1996). "TRANSFAC: a database on transcription factors and their DNA binding sites." Nucleic Acids Research 24(1): 238-41.

Wong, G. K., D. A. Passey, *et al.* (2001). "Most of the human genome is transcribed." Genome Research 11(12): 1975-7.

Wood, V., R. Gwilliam, *et al.* (2002). "The genome sequence of Schizosaccharomyces pombe.[see comment][erratum appears in Nature 2003 Jan 2;421(6918):94 Note: Cerrutti L [corrected to Cerutti L]]." Nature 415(6874): 871-80.

Woudt, L. P., A. B. Smit, *et al.* (1986). "Conserved sequence elements upstream of the gene encoding yeast ribosomal protein L25 are involved in transcription activation." EMBO Journal 5(5): 1037-40.

Woychik, N. A., and Young, R.A. (1994). Exploring polymerase II structure and function. Transcription: mechanisms and regulation. R. C. Conway, and Conway J.W. New York, Raven Press: 227-42.

Wu, J. A. and J. L. Manley (1991). "Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA." Nature 352(6338): 818-21.

Wu, P.-Y. J., C. Ruhlmann, *et al.* (2004). "Molecular architecture of the S. cerevisiae SAGA complex." Molecular Cell 15(2): 199-208.

Wu, X., H. Hart, *et al.* (2001). "Characterization of Gac1p, a regulatory subunit of protein phosphatase type I involved in glycogen accumulation in Saccharomyces cerevisiae." Molecular Genetics & Genomics: MGG 265(4): 622-35.

Wyatt, J. R., E. J. Sontheimer, *et al.* (1992). "Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing." Genes & Development 6(12B): 2542-53.

Yagi, S., K. Yagi, *et al.* (1994). "The UAS of the yeast GAPDH promoter consists of multiple general functional elements including RAP1 and GRF2 binding sites." Journal of Veterinary Medical Science 56(2): 235-44.

Yang, Y. and U. T. Meier (2003). "Genetic interaction between a chaperone of small nucleolar ribonucleoprotein particles and cytosolic serine hydroxymethyltransferase." Journal of Biological Chemistry 278(26): 23553-60.

Yin, Z., S. Wilson, *et al.* (2003). "Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs." Molecular Microbiology 48(3): 713-24.

Yong, J., L. Wan, *et al.* (2004). "Why do cells need an assembly machine for RNA-protein complexes?" Trends in Cell Biology 14(5): 226-32.

Young, R. A. (1991). "RNA polymerase II." Annual Review of Biochemistry 60: 689-715.

Yu, Y. T., P. A. Maroney, *et al.* (1995). "U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone." Rna-A Publication of the Rna Society 1(1): 46-54.

Zawel, L. and D. Reinberg (1993). "Initiation of transcription by RNA polymerase II: a multi-step process." Progress in Nucleic Acid Research & Molecular Biology 44: 67-108.

Zhou, Z., L. J. Licklider, *et al.* (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature 419(6903): 182-5.

Zhu, J. and M. Q. Zhang (1999). "SCPD: a promoter database of the yeast Saccharomyces cerevisiae." Bioinformatics 15(7-8): 607-11.

Zieve, G. W. and R. A. Sauterer (1990). "Cell biology of the snRNP particles." Critical Reviews in Biochemistry & Molecular Biology 25(1): 1-46.

Zitomer, R. S. and C. V. Lowry (1992). "Regulation of gene expression by oxygen in Saccharomyces cerevisiae." Microbiological Reviews 56(1): 1-11.

Zlatanova, J., S. H. Leuba, *et al.* (1999). "Chromatin structure revisited." Critical Reviews in Eukaryotic Gene Expression 9(3-4): 245-55.

Zurita, M. and C. Merino (2003). "The transcriptional complexity of the TFIIH complex." Trends in Genetics 19(10): 578-84.

Appendix

The Appendix of this thesis may be divided up into six sections: development of the real time PCR assay; validation of the splicing assay and splicing of the *LSM7 intron only* mutant; *GO* processes analysis of genes with introns in *S. cerevisiae*; data for individual genes and gene lists for anaerobic microarrays, details of splicing-sensitive microarray analysis (GO processes, normalised data, controls, gene lists and intron feature correlations, etc) and raw real time PCR data.

Section 1: Development of the real time PCR assay

Raw real time PCR data is presented as a plot of fluorescence vs cycle number. For accurate comparisons, measurements must be taken during exponential phase - before reagents become limiting and PCR inhibitors a greater effect. This assay relies on the use of external standards. One-step real time PCR was used to minimise human error in the assay, and general real time PCR reagents were initially aliquotted into each sample to be analysed from a master mix to further reduce this error.

One step real time PCR reactions proceed in three steps only as shown below (Figure 1 image derived using ABI Sequence Detector v1.7a software output). ROX is used as a passive internal reference dye, as it stays constant over time, accounts for variations in pipetting and removes noise and fluoresces at a different wavelength from SYBR green 1(605-610nm). SYBR Green 1 binds to minor groove of DS DNA and, on binding, fluoresces at 540nm. It does not inhibit PCR, but amplifies ANY DS DNA present in the sample. It is therefore necessary to optimise the reaction to avoid the production of primer-dimers over the course of the PCR reaction.

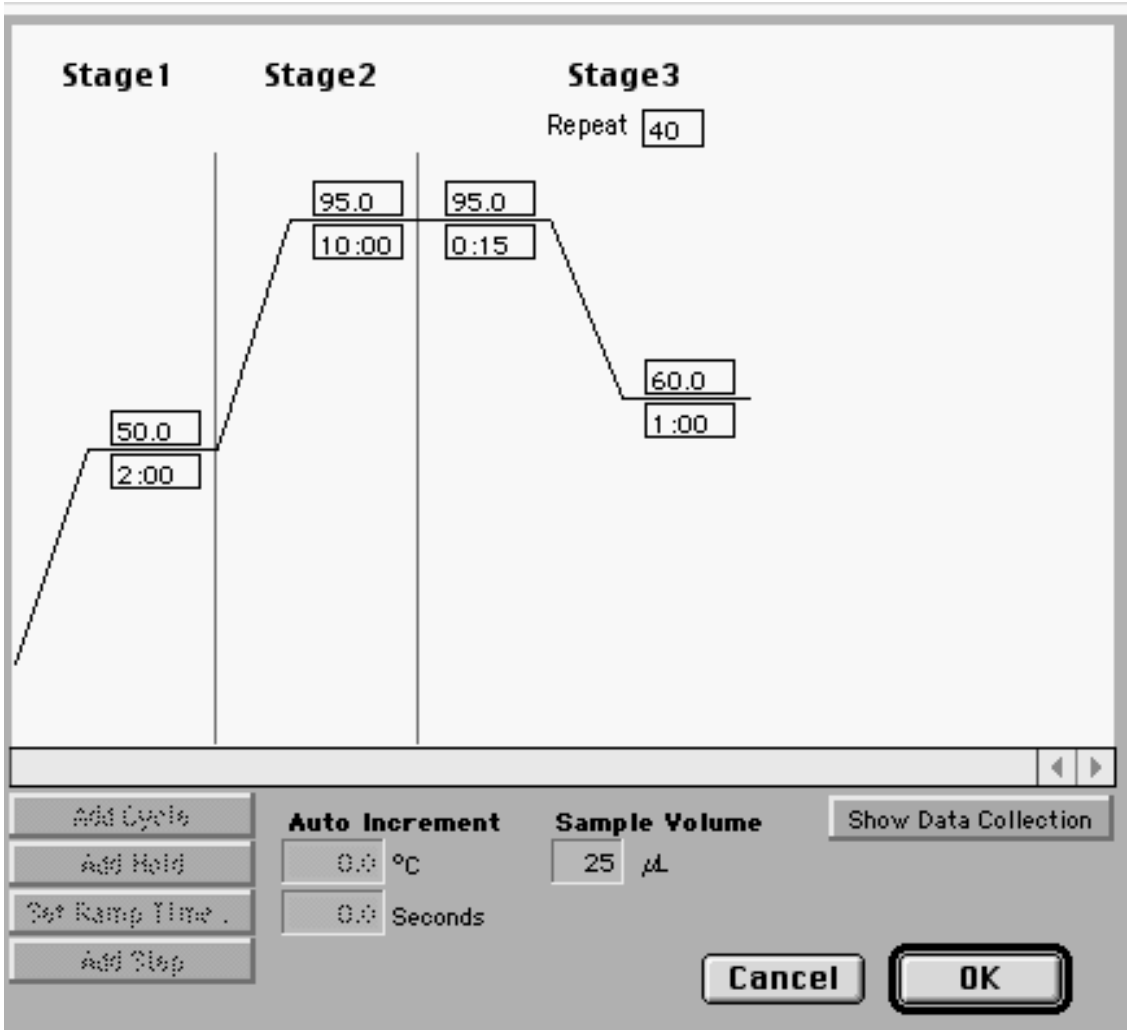
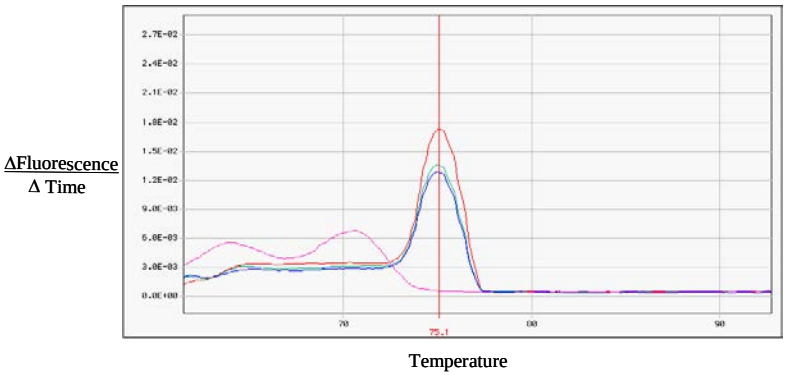


Figure 1: The different steps of typical One Step real time PCR analysis.

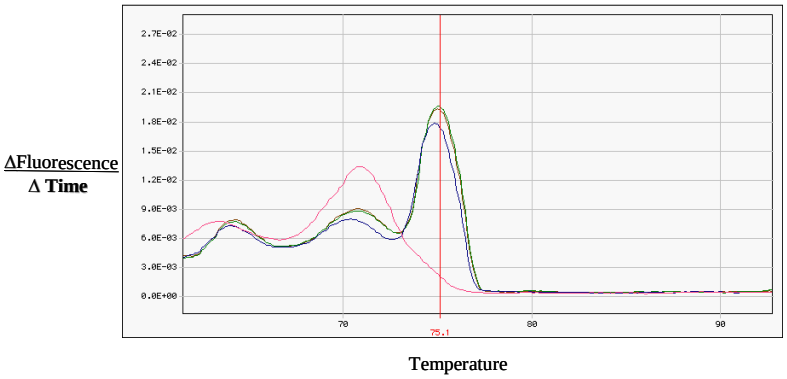
Prior to using a set of primers, two assays were carried out: determination of optimal primer concentration, using different concentrations of primers but a static concentration of RNA to test for the highest primer concentration that does not produce primer dimers during the PCR reaction, as shown by melt-curve analysis using Dissociation Curves software (ABI Biosystems). An example of this is shown below.

However at when the template RNA is very rare, primer dimers may still form in the during the PCR reaction. In all cases examined, primer dimers were found to be shorter (and therefore have the two strands dissociate from each other at a lower temperature) than the product from the template RNA.

Therefore the concentration of all samples was measured during the real time PCR analysis only at a temperature above that at which the primer-dimers would have 'melted' apart. For the example illustrated in Figure 2, this would have been around 72ºC.



LSM4 150nM



LSM4 300nM

Figure 2: Dissociation curves software output. This depicts the accumulation of primer dimers (the lower peak on the curve) with increasing concentration of PCR primers for *LSM4.* In both instances the pink line represents the 'No template' control (ie all PCR product present results from primer-dimer formation). From these two controls, 150nm would be the ideal concentration of the *LSM4* primers to amplify this particular concentration of RNA template.

Order of addition of substrates to the PCR reaction tube was also seen to impact on the formation of primer-dimers in solution, as illustrated in Figure 3. Therefore in each case, reagents were added to tubes on ice in the following order: PCR master mix (see methods), RNA, primers.

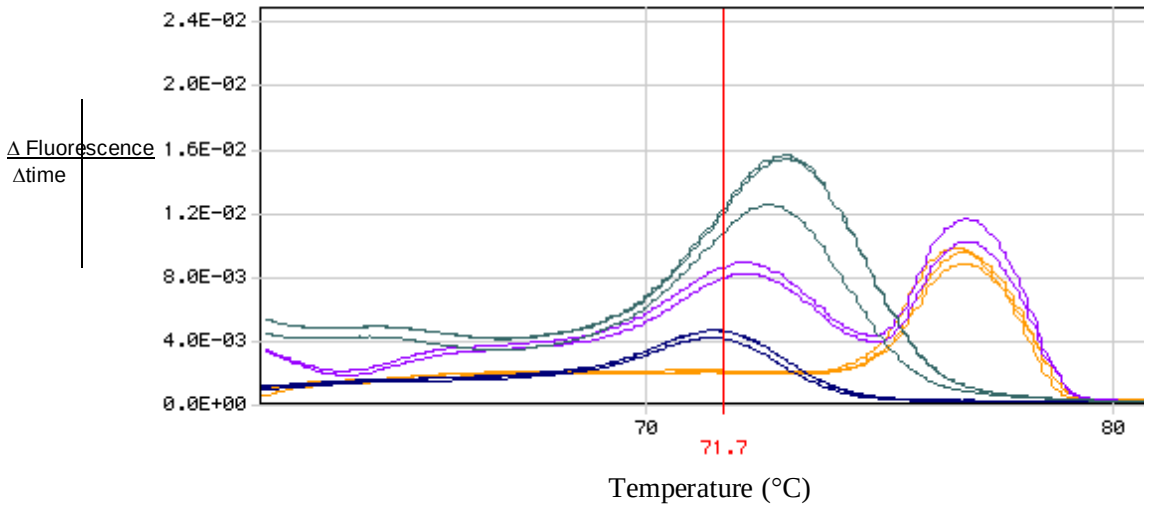


Figure 3: Effect of adding different PCR reagents in different orders. No template control is shown in green. In samples where primers are added to the master mix up to 30 minutes before the addition of the RNA , the levels of primer-dimers produced (the 72º C peak) increased dramatically (shown in purple) compared to samples where r master mix, RNA were mixed prior to the addition of primers (yellow line). Results from samples where RNA was added to the tube up to 30 minutes before the addition of either master mix or primers are shown and probably reflect the degradation of the RNA template by RNases in the reaction tube.

The second assay performed for each primer pair was analysis of PCR efficiency. Data from this last assay is used to analyse all data arising from the use of this primer set using the 'standard curve' technique. This was used in preference to the ΔΔCt method, as this assumes that each primer pair is 100% efficient at performing PCR, and in our experience this was not the case. Examples of the efficiencies of various real time PCR primer sets are shown in Table 1.

Table 1: PCR efficiencies for some of the primers used in this study, as determined by the standard curve. The number three represents 100% efficiency of a primer set.

Primer pair (F and R primers) designed to measure transcripts from	Efficiency
LSM1	**3.0262**
LSM2	**3.4763**
LSM3	**3.2165**
LSM4	**3.8626**
LSM5	**3.165**
LSM6	**3.2463**
LSM7	**3.309**
LSM8	**3.2116**
SIR3	**3.5886**
SMF	**3.1772**
SMG	**2.9354**
RPL25	**3.8984**

SIR3 was used as an external control for all assays except those relating to the half-life of the transcript, as levels of the transcript remained largely unchanged under a variety of environmental conditions (as determined by analysis of microarray data from other researchers) (Figure 4). In addition, its abundance was found to be similar to that of the *LSM* genes, supporting its use in assays optimised for relatively rare transcripts (Figure 5)

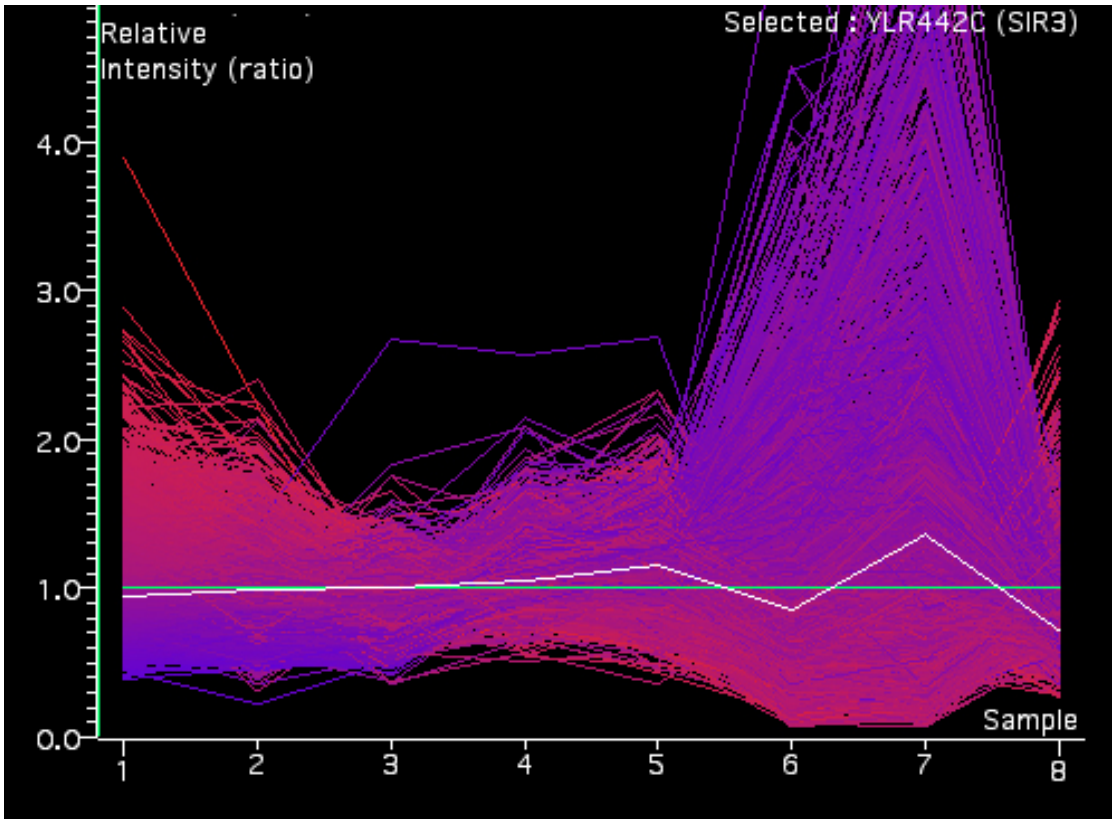


Figure 4: Microarray analysis of RNA from different points across the diauxic shift (de Risi *et al.,* 1997). *SIR3* transcript data is shown in white. Image generated using Genespring software.

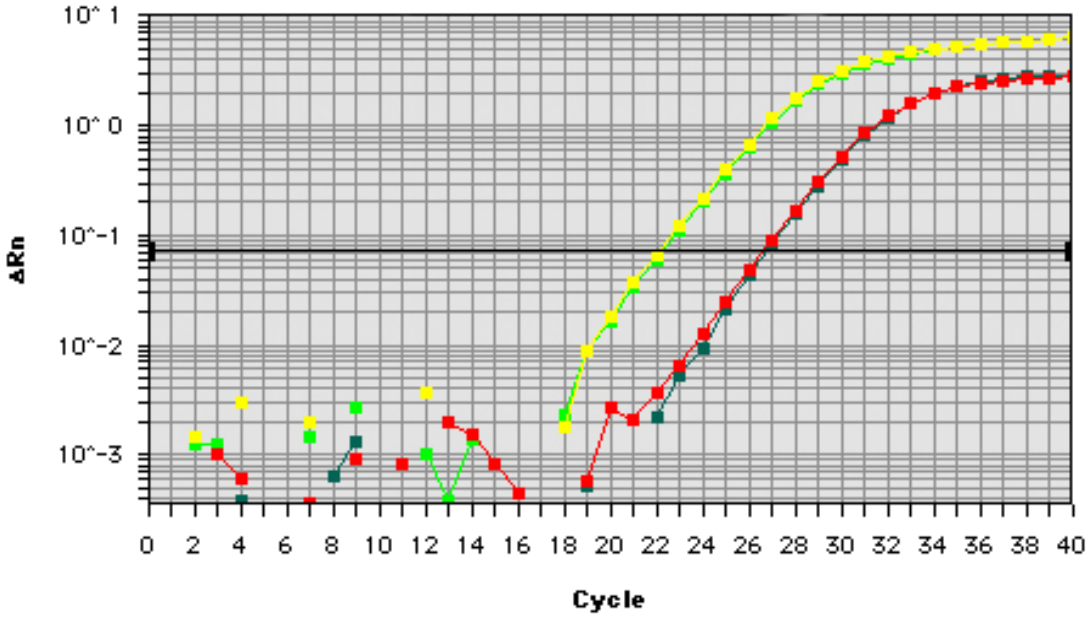
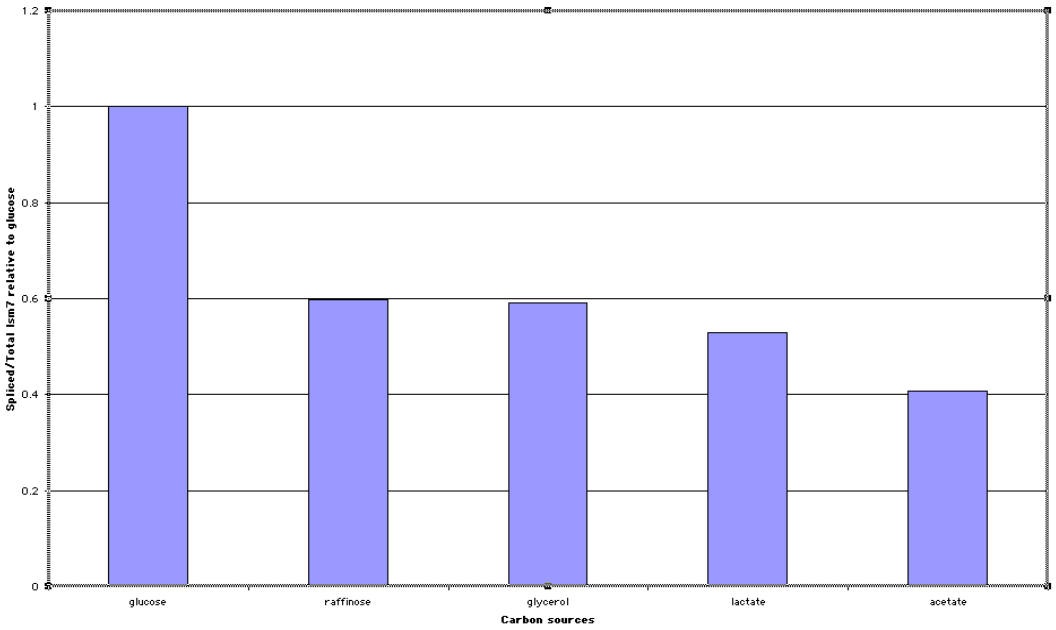


Figure 5: Real time PCR analysis showing duplicates of amplification of *SIR3* (yellow and light green lines) and *LSM4* (red and dark green lines).

Section 2:A. Validation of the splicing assay

The splicing assay employed in this study relied on the assumption that the ratio of unspliced RNA relative to the total concentration of a given transcript (measured by real time PCR) reflected the accumulation of that unspliced due to changes in the splicing capacity of the cell. To test this, levels of spliced transcript were also measured, relative to the total transcript for each gene tested. In every case, levels of spliced/total transcript were found to decrease as unspliced/total transcript increased. Examples are shown in Figure 6

A.



B.

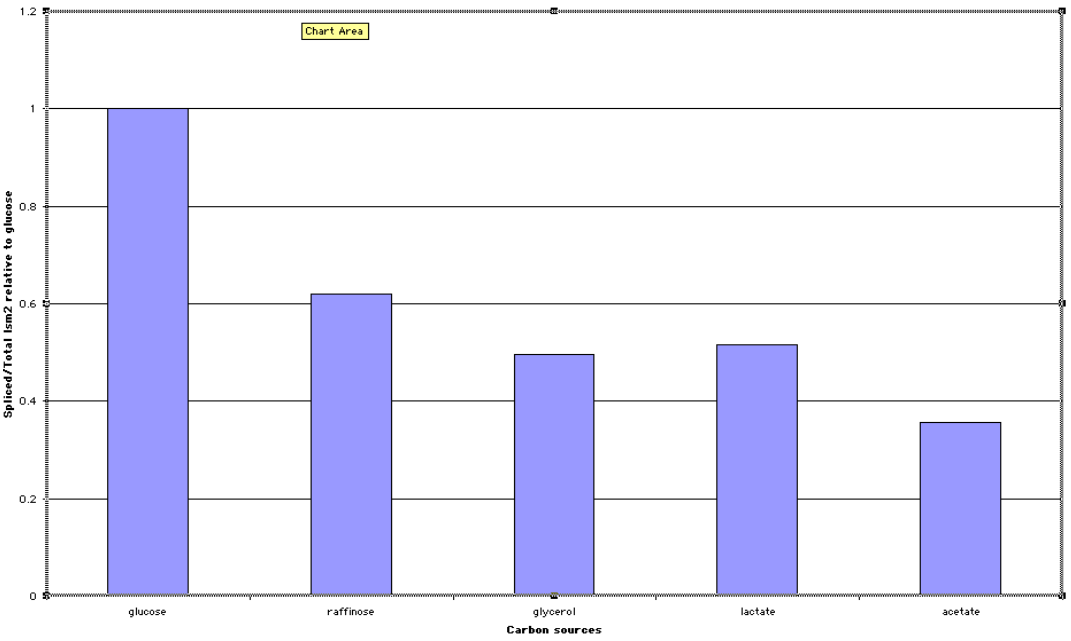


Figure 6: Spliced/Total transcript ratios for *LSM7* (A) and *LSM2* (B) transcripts in cells grown in the presence of varying carbon sources.

Section 2:B. Splicing assay for LSM7 intron only *transcript*

A direct measure of splicing transcript could not be carried out in the usual manner for *LSM7* in the *LSM7 intron only* mutant as the mRNA transcript did not contain sufficient sequence to ensure the transcript could be differentiated from the other possible forms. Accordingly an indirect method was developed. Primers were designed to measure

- unspliced RNA – one primer spanning the intron-exon junction and one primer within the intron
- total and spliced: primers immediately upstream and downstream of intron within the mRNA transcript.

Both primer concentrations and temperatures for primer-dimer cut-off analysis were carefully determined. The spliced transcript could then be differentiated from the total transcript (under conditions where no primer dimer forms) based on the size of the resulting PCR product

This differentiation is based technique similar to that employed to differentiate between signal derived from primer-dimers and signal derived from amplification of the template DNA described above. Levels of 'total +spliced' transcript were measured at 60ºC and also at a temperature above which the spliced product would have melted apart. Assuming no primer dimer was formed (since, due to careful primer design- and perhaps a little luck – no primer dimers formed in no template control samples at any time), the difference between the these two measurements would be an indication of the presence of spliced transcript. Melt curve analysis shows the presence of a smaller, spliced product along with a larger product representing the total *LSM7* transcript in the *intron only* mutant (Figure 7). Exact values for this produced were not determined from the meltcurve analysis, however, as in practice it is the algorithm used by this program does not appear to take into account error inherent in the measuring system or heat block used in the ABI prism real time PCR machine.

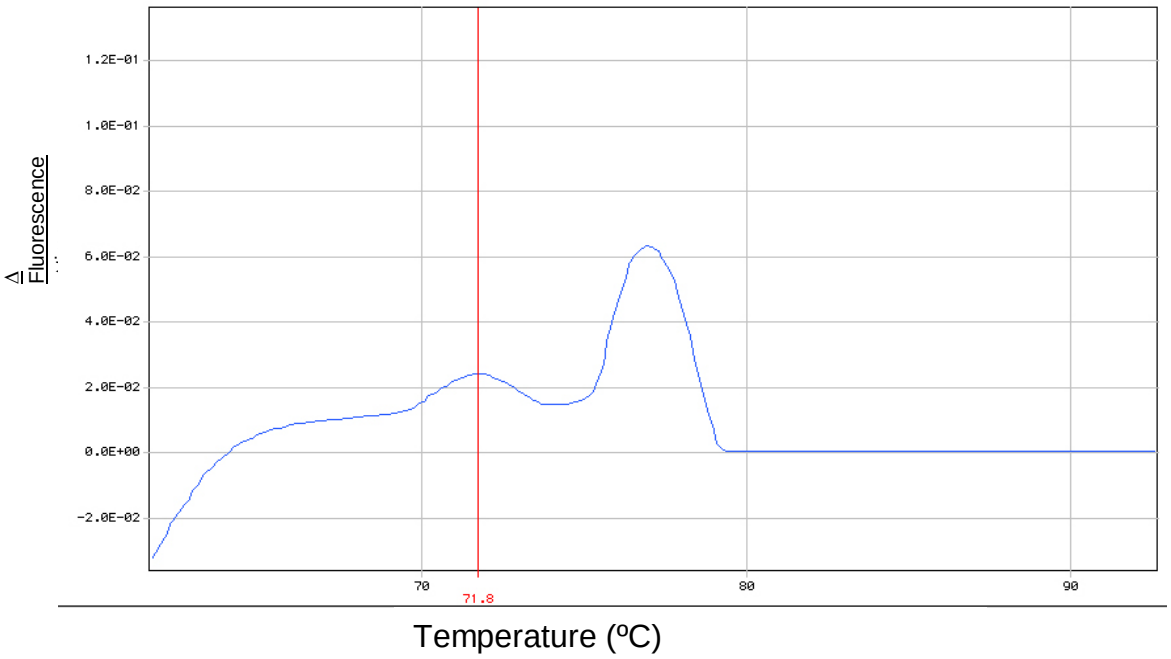


Figure 7:**Meltcurve analysis of products resulting from the real time PCR amplification** using the 'total' primers described above. Two distinct products can be seen. The smaller product (at 71ºC) is taken to represent the 'spliced' product.

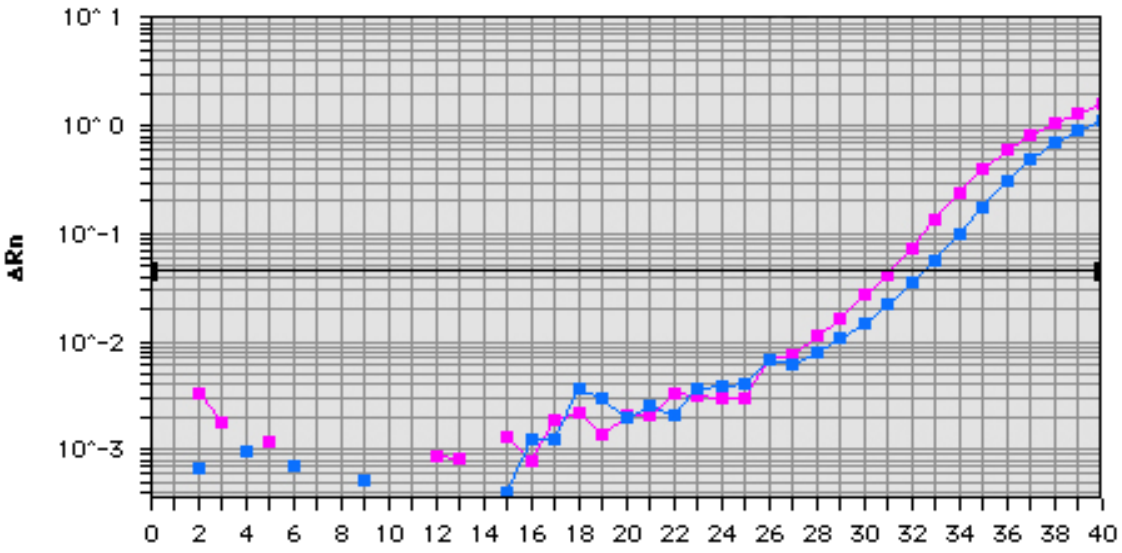


Figure 8: Relative amounts of product resulting from measurement of product using "total" primers at 60ºC (pink line) and 74ºC (blue line). The difference between the two is taken to be the a measure of the amount of 'spliced' product present

Section 2:C. Mathematical derivation of relationship between [Unspliced], and rate of splicing and transcription

From the previous model (Section 5.1) $\frac{d[UnS]}{dt} = k_t - k_s\,[UnS]$

$$\int d\ln\left|[UnS] - \frac{k_s}{k_t}\right| = -\int k_s dt$$

therefore, if both sides are integrated using a constant 'A'

$$\ln\left([UnS] - \frac{ks}{k_t}\right) = -k_s t + A$$

So

$$[US] - \frac{ks}{kt} = e^{A} \text{ x } e^{-k_s t}$$

If A' is set as equal to e^{A}

then

$$[UnS] = \frac{kt}{k_s}$$

Section 3: GO analysis of genes containing introns in S. cerevisiae

This data is presented in the excel spreadsheet accompanying this document.

Section 4: Anaerobic Microarray data

Patterns for individual genes generated by analysis of raw data in each case using Genespring GX. Data was normalised as for that presented in Section5.7 and Figure 5.8. This folder also contains gene lists for groups that

were identified based on pattern of expression across the timecourse (analysed by ANOVA using Genespring GX) and a list of for all genes which were significantly upregulated between time 0 and time 120 in these arrays.

Section 5: Analysis of splicing-sensitive microarray data

Files detailing the normalised Cy3 and Cy4 intensities for each data point are accompanying this document (in the 'normalised microarray data' file.). Lists of genes which all behave in the same manner across these assays (i.e all have inhibited – blue -, unaffected-yellow -, or enhanced – red - splicing in the *LSM7* or *LSM6* deletion mutants) and GO analysis of these varioua gene lists can be found in the 'splicing sensitive microarray data folder accompanying this document. Excel files detailing comparisons between some intron features (eg length) and IA are also provided.

Clark *et al.,* 2002 defines the **Splice Junction Index** as:

(SJmut/SJwt) / (E2mut/E2wt)
or
$\log_2$(SJmut/SJwt) – $\log_2$(E2mut/E2wt)

where:

SJmut is the signal intensity of the splice junction spot hybridised with cDNA from the mutant
SJwt is the signal intensity of the splice junction spot hybridised with cDNA from the wild type
E2mut is the signal intensity of the exon 2 spot hybridised with cDNA from the mutant
E2wt is the signal intensity of the exon 2 spot hybridised with cDNA from the wild type

Similarly, Ares defines the **Intron Accumulation Index** as:

(INmut/INwt) / (E2mut/E2wt)
or
$\log_2$(INmut/INwt) – $\log_2$(E2mut/E2wt)

where:

INmut is the signal intensity of the intronic spot hybridised with cDNA from the mutant

SJwt is the signal intensity of the intronic spot hybridised with cDNA from the wild type.

Table 2: Negative controls for the splicing-sensitive microarray

ORF	Type	LR	N	Mean A	Cy3	*Cy5*
BioB	E_coli	-0.30775	3	8.068104	111.3947	114
BioC	E_coli	-0.23969	3	8.017455	108.2759	103.2593
BioD	E_coli	-0.39251	5	8.2323	151.2051	136.8788
cheY	E_coli	NA	0	NA	55.8	24.66667
Cre	E_coli	-0.29407	10	8.462403	226.325	189.0526
fliM	E_coli	NA	0	NA	63.33333	115
htrE	E_coli	NA	0	NA	28.5	17.5
lamB	E_coli	-0.11747	7	8.090575	180.8158	153.9167
mdoG	E_coli	-0.5046	2	8.825189	178.75	198.6
mdoH	E_coli	NA	0	NA	131.6	74.5
mrcA	E_coli	-0.39308	2	7.288711	128.5517	112.8667
mrdA	E_coli	NA	0	NA	57.6	13
mrdB	E_coli	-0.05558	1	8.196963	68	60
nirD	E_coli	0.165509	2	8.117305	132.7778	145.8
ompA	E_coli	0.130828	2	10.20283	139.5556	157.9333
ompB	E_coli	-0.0016	3	8.646331	130.8	140.4828
ompC	E_coli	-0.625	2	8.851863	216.5714	154.5714
ompF	E_coli	NA	0	NA	124.5714	83.5
pal	E_coli	-0.21	1	9.667113	97.39474	94.1875
pgpA	E_coli	-0.21314	1	9.569376	153.5714	318
pgpB	E_coli	-1.04877	1	8.441936	94.28571	116
tolB	E_coli	-0.14635	1	8.692156	130.3333	71
tolC	E_coli	-0.6864	3	8.188585	156.625	195.4
YCONTROL02	neg_control	NA	0	NA	79.5	94
YCONTROL04	neg_control	NA	0	NA	93.57143	74.2
YCONTROL19	neg_control	NA	0	NA	64.28571	19.16667
YCONTROL57	neg_control	-0.10915	3	9.460058	737.875	593.5714
YCONTROL60	neg_control	NA	0	NA	66.375	38
YCONTROL63	neg_control	NA	0	NA	67.375	16
YCONTROL77	neg_control	0.155931	2	8.982619	187.1429	220.4286
YCONTROL91	neg_control	-0.06762	5	9.484697	912.125	817
YCONTROLAF	neg_control	-0.82416	1	8.345763	93.46667	70
Printing_Buffer	print_buff	-0.0031	34	8.963856	136.9195	138.7661

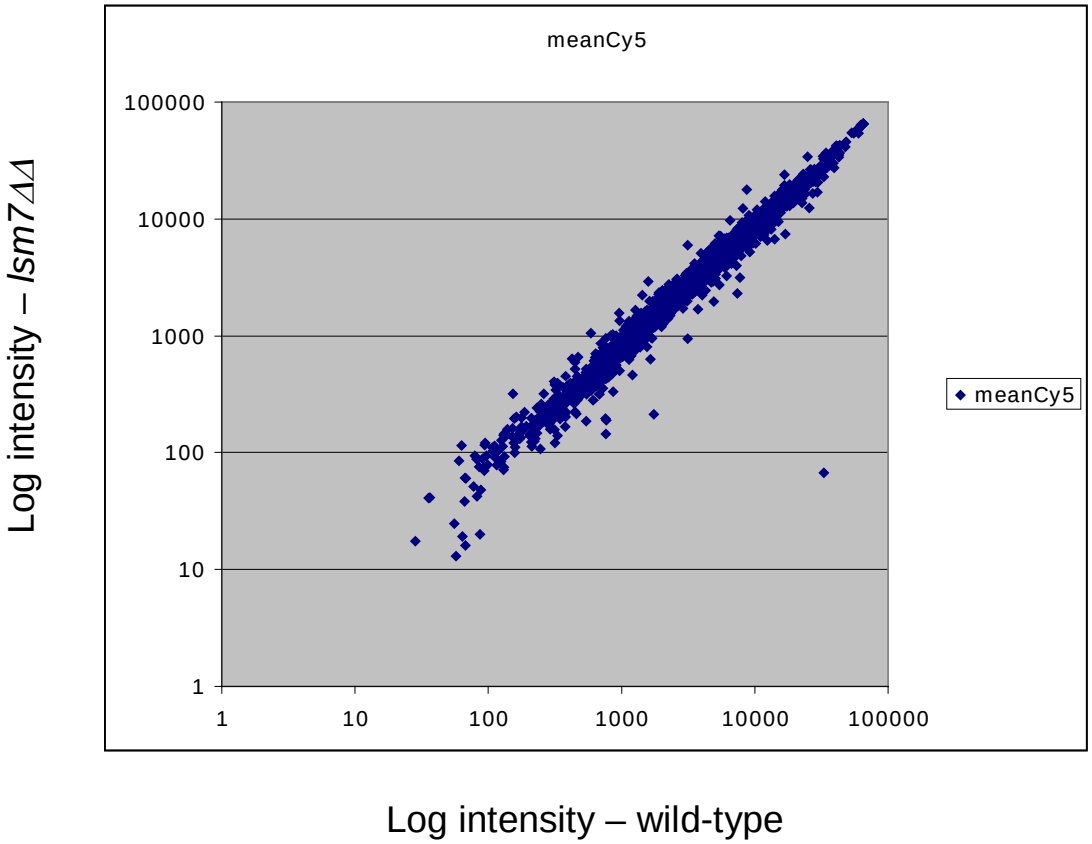


Figure 9:Scatter plots of probe intensities for *lsm7Δ*

Table3: Genes exhibiting the highest IA in *lsm7*Δ

Gene	Annotation	Phenotype	IALSM7	Transcriptional frequency	Intron Size (bp)
TNC3	Largest of six subunits of the RNA polymerase III transcription initiation factor complex (TFIIIC)	essential	0.955	1.1	90
SNC1	Vesicle membrane receptor protein (v-SNARE) involved in the fusion between Golgi-derived secretory vesicles with the plasma membrane;	non-essential	0.85	2.3	113
RIM1	Single-stranded DNA-binding protein essential for mitochondrial genome maintenance; involved in mitochondrial DNA replication	non-essential	10.6	83	
ARP2	Essential component of the Arp2/3 complex, which is a highly conserved actin nucleation center required for the motility and integrity of actin patches; involved in endocytosis and membrane growth and polarity	essential	0.83	7.1	123
RUB1	Ubiquitin-like protein with similarity to mammalian NEDD8; conjugation (neddylation) substrates include the cullins Cdc53p, Rtt101p, and Cul3p; activated by Ula1p and Uba3p (E1 enzyme pair); conjugation mediated by Ubc12p (E2 enzyme)	non-essential	0.85	14.2	73
EPT1	sn-1,2-diacylglycerol ethanolamine- and cholinephosphotranferase	non-essential	0.82	67.2	441

Gene	Annotation	Phenotype	IALSM7	Transcriptional frequency	Intron Size (bp)
RPS4B	Protein component of the small (40S) ribosomal subunit; identical to Rps4Bp and has similarity to rat S4 ribosomal protein	non-essential	0.89	3.9	131
RPL16A	N-terminally acetylated protein component of the large (60S) ribosomal subunit, binds to 5.8 S rRNA; has similarity to Rpl16Bp, E. coli L13 and rat L13a ribosomal proteins; transcriptionally regulated by Rap1p	0.92	82.1	82.1	434
RPL39	Protein component of the large (60S) ribosomal subunit, has similarity to rat L39 ribosomal protein; required for ribosome biogenesis; exhibits genetic interactions with SIS1 and PAB1	non-essential	1.01	14.7	408
ARP9	Actin-related protein involved in transcriptional regulation; subunit of the chromatin remodeling Snf/Swi complex	essential	0.88	7.3	156
YNL050c	'Hypothetical' gene, protein interacts with Lsm8p	non-essential	1.2	15.7	576
SAR1	GTPase, GTP-binding protein of the ARF family, component of COPII coat of vesicles; required for transport vesicle formation during ER to Golgi protein transport	essential	1.04	0.8	80

Section 6: Raw data for real time PCR assays

The CT numbers and subsequence analysis of data arising from each real time PCR point are found in the excel spreadsheets accompanying this document. These have been provided for key figures presented in this these,

that have also been submitted for publication – specifically, for Figure 3.3, 4.3. 4.6, 5.1B and 5.3.

References from Appendix

DeRisi JL. Iyer VR. Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. [Journal Article] Science. 278(5338):680-6, 1997 Oct 24.

Other parts of the appendices for this thesis can be found in digital format on the attached CD-ROM fron the print copy