THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Identifying novel targets for the snoRNA class of stable non-coding RNAs
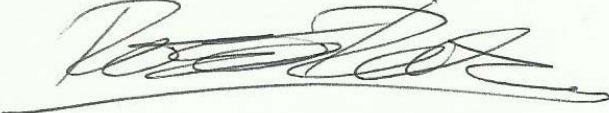
Rosie Peters

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2018

DECLARATION

I declare that this thesis was composed by myself. The research in this thesis is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

Rosie Peters   1/10/2018

Acknowledgements

I would first like to thank my supervisor, David, for his help and guidance throughout the PhD project. Many thanks to Tanya, for her supervision in the lab and her advice throughout, and also to Hywel, for all of his work and assistance in bioinformatics. I'd like to thank everyone in the Tollervey lab - past and present - for their advice, and whose work provided the foundation on which to develop my research.

I would also like to thank my thesis committee - Jean Beggs, Sander Granneman and Bill Earnshaw - for their support and guidance over the years. In addition, I thank David Barrass, for his invaluable help and instruction during certain experiments, and Alastair Kerr for his statistical expertise. I am very grateful to the Wellcome Trust for funding this four year PhD Programme in Cell Biology, and to Karen Traill for the excellent work she does. Furthermore, I'd like to thank members of the Wellcome Centre for Cell Biology, for their encouragement and know-how.

I would like to thank my close friends for their unwavering support and humour, throughout times which were not stress-free. Their friendship facilitated my work on this project. Finally, I'd like to thank my parents and my brother, Dan. Without their constant support and encouragement, I could not have completed this PhD, and I am incredibly grateful to them.

ABSTRACT

Non-coding RNAs (ncRNAs) are a subset of RNAs that do not code for protein. They are divided into a number of different groups based on their function and targets. Small nucleolar RNAs (snoRNAs) are ncRNAs that have long been known to function as guides for ribosomal RNA (rRNA) modifying enzymes. They are classified into two major groups: box C/D snoRNAs and box H/ACA snoRNAs. Most box C/D snoRNAs direct the 2'-*O*-methylation of rRNA substrates, but some lack known targets and are therefore termed 'orphan snoRNAs'. Studies have implicated orphan snoRNAs in pre-mRNA processing and stability, but the functional consequence of snoRNA binding to mRNAs has not been fully determined.

Saccharomyces cerevisiae had two orphan snoRNAs, snR4 and snR45, with no known function in ribosome synthesis. This project aimed to determine the targets of these snoRNAs, and investigate the effects of snoRNA binding to non-canonical target RNAs, as well as the underlying mechanism. Synthetic gene array screens with deletions of the *SNR4* and *SNR45* genes identified multiple positive and negative genetic interactions. In particular, deletion of either snoRNA gene was synthetic-lethal with mutation of the snoRNA-associated methyltransferase, Nop1 (Fibrillarin in humans), demonstrating that both have important functions. CLASH analyses of RNA-RNA interactions showed that these snoRNAs bind multiple mRNAs, while RNA sequencing and RT-qPCR revealed that snoRNA deletion altered mRNA abundance. Both orphan snoRNAs were well conserved between fungi, with a region of high conservation indicating a potential binding site. Associations were identified between snR4 and snR45 and multiple sequences within rRNA, including two recently identified sites of 18S rRNA acetylation. Work elsewhere showed that snR4 and snR45 function as guides for the acetyltransferase Kre33 using the region of high conservation, removing their 'orphan' status.

Orphan snoRNAs have been implicated in human diseases, such as Prader Willi Syndrome and cancers. The work discussed in this thesis helps to elucidate the RNA interactions of yeast orphan snoRNAs. It has provided a greater understanding of the mechanisms involved, and may inform future work in combatting human disease.

LAY SUMMARY

Genetic information is encoded in the sequences of long regions of DNA. RNA is similar to DNA and carries a copy of a region of the genetic information. The best known class of RNA is messenger RNA (mRNA), which carries the information used to programme protein synthesis. However, there are numerous other classes of RNA, which do not encode protein, named 'non-coding RNA' (ncRNA). Every cell needs to make a vast number of proteins, and these are made by cellular machines called 'ribosomes'. The main component of this machine is a large ncRNA called 'ribosomal RNA' (rRNA). To make the ribosome, the rRNA first needs to be synthesised, processed and modified. Small ncRNAs called 'snoRNAs' have long been known to direct these modifications, but it has more recently been discovered that snoRNAs can also bind to mRNA. The effect of snoRNA binding to mRNAs is currently unknown, though evidence suggests that snoRNAs may affect the quantity or processing of target mRNAs. Recent studies have revealed links between abnormal amounts of individual snoRNAs in the cell and cancer, indicating the importance of researching these interactions. Furthermore, the lack of one type of snoRNA in humans is directly linked to the genetic disease Prader Willi Syndrome, due to the loss of interactions with mRNA.

In this study, individual snoRNAs that lacked a known function in ribosome synthesis were deleted from the DNA in baker's yeast (a model organism for human cells). The project aimed to investigate how well or poorly the yeast grew without the snoRNAs, whether the quantity of mRNA changed, and whether other functional interactions could be discovered. Lack of these snoRNAs did not clearly affect how well the yeast grew, but did result in sickness or cell death when combined with other particular mutations. The snoRNAs were shown to bind to specific mRNAs, and deletion of the snoRNAs affected the amount of mRNAs present in the cell. Furthermore, these snoRNAs were found to guide a type of modification on rRNA that was different from that directed by other snoRNAs. This research has provided a better understanding of the role snoRNAs play in yeast, which can potentially help understand the mechanistic basis of human diseases.

CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

The canonical function and targets for the small nucleolar RNA (snoRNA) class of non-coding RNAs in ribosome biogenesis have been well documented in the last two decades. However, the function and targets of 'orphan' snoRNAs, with no known role in ribosome synthesis, are still poorly understood. Uncovering these mechanisms is imperative, taking into consideration the increasing association of orphan snoRNAs with human disease, such as cancer. The data presented in this work provides a deeper understanding of the novel targets of orphan snoRNAs and the functions governing these interactions, which may aid future work in treating human diseases.

1.1: Ribosome biogenesis

Ribosome biogenesis is a well-studied process, the correct coordination of which is necessary for growth and efficient cellular protein production. Ribosomal RNA (rRNA) is transcribed in a non-membrane-bound substructure called the nucleolus. This is a dense region formed in the nucleus around the genes encoding ribosomal DNA (rDNA) (**Sillevis-Smitt *et. al.,* 1973; reviewed in Hernandez-Verdun, 1991; Boisvert *et. al.,* 2007**). The rDNA consists of ~150 tandem repeats on chromosome XII. RNA polymerase I (Pol I) is responsible for the transcription of an rRNA precursor (35S pre-rRNA in *Saccharomyes cerevisiae,* 45S in higher eukaryotes). This contains a 5' and 3' external transcribed spacer (5' and 3' ETS), two internal transcribed spacers (ITS1 and ITS2), and the rRNAs: 18S, 5.8S and 25S. The 35S pre-rRNA undergoes many modification and processing steps in order to generate the mature 18S, 5.8S and 25S rRNAs (**reviewed in Venema and Tollervey, 1999; Woolford and Baserga, 2013**). 5S rRNA is independently transcribed by Pol III and 3' processed by the Rex1p exonuclease (**van Hoof *et. al.,* 2000**). These four rRNAs assemble with many proteins to form the 40S and 60S ribosomal subunits (Figure 1.1.1).
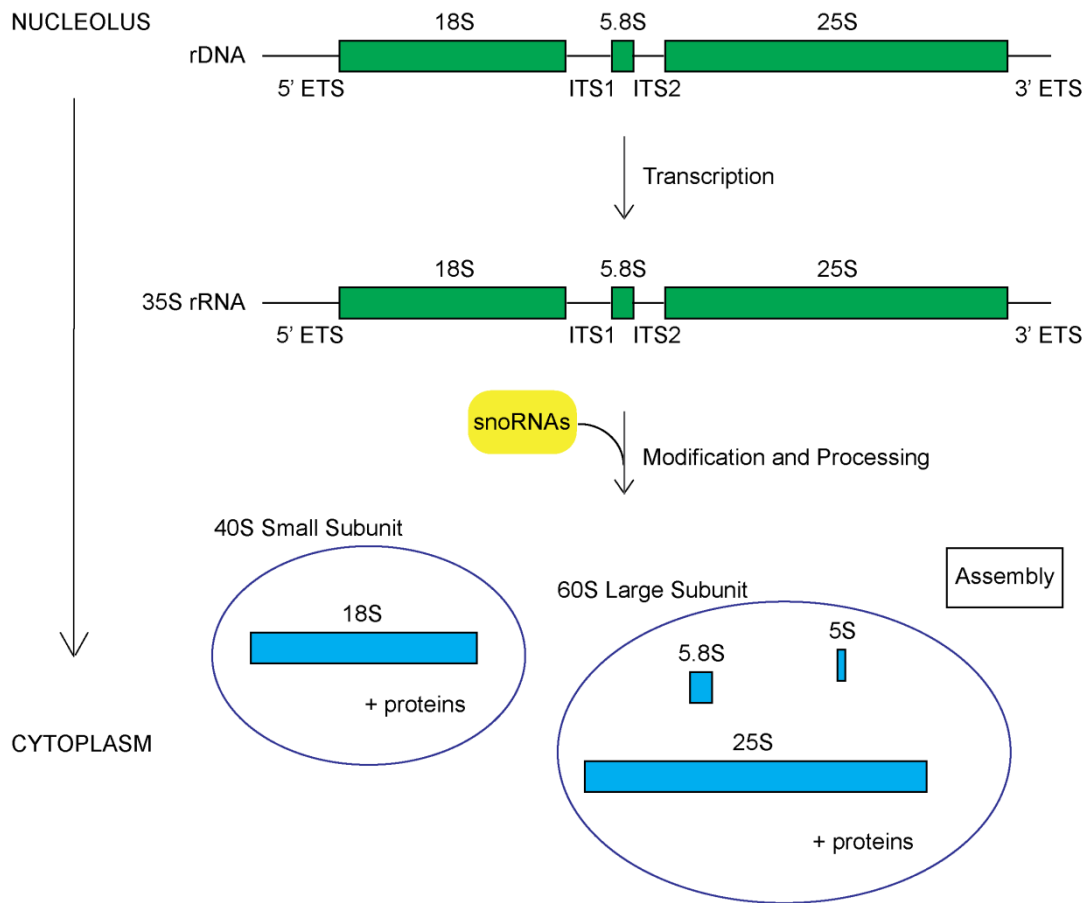
Figure 1.1.1 Ribosome biogenesis pathway overview in *Saccharomyces cerevisiae*. rDNA loci containing coding regions for 18S, 5.8S and 25S, external and internal transcribed spacers (ETS and ITS, respectively) are transcribed by RNA Pol I in the nucleolus into the precursor rRNA (35S in *S. cerevisiae*, 45S in higher eukaryotes). This undergoes modification, processing and multiple cleavage steps involving snoRNAs to generate individual rRNA species (18S, 5.8S and 25S). With the addition of 5S rRNA and multiple ribosomal proteins, these assemble into the 40S and 60S subunits of the ribosome, at which point they are in the cytoplasm.

Initially, the 3' ETS sequence is cleaved by Rnt1p at B_0 (**Kufel *et. al.,* 1999**). Cleavages cotranscriptionally occur at sites A_0, A_1 and A_2, which segment the 35S pre-rRNA into 20S and $27SA_2$ species (Figure 1.1.2). Canonically, snoRNAs guide the modification of rRNA (discussed in chapter 1.2); however, four yeast snoRNAs are required for pre-rRNA processing. The U3 box C/D snoRNP (snoRNA ribonucleoprotein) is required for cleavage at sites A_0, A_1 and A_2, along with a number of proteins (**Hughes and Ares, 1991; Jansen *et. al.,* 1993; Dunbar *et. al.,***

1997; Wiederkehr et. al., 1998; Lee and Baserga, 1999; Venema et. al., 2000). Overall, a high number of U3-associated proteins have been identified, that form a large RNP complex named the small subunit (SSU) processome (**Dragon et. al., 2002; Bernstein et. al., 2004**). All proteins within this complex are nucleolar, co-immunoprecipitate with both U3 snoRNA and the U3-specific protein Mpp10p, and are required for pre-18S rRNA processing. In total, the SSU processome comprises at least 40 proteins (**Bernstein et. al., 2004**). Furthermore, the t-UTP (transcription-required U3 proteins) subset of the SSU processome forms a sub-complex in the absence of the SSU processome, and is required for optimal transcription of rDNA (**Gallagher et. al., 2004**). The U14 box C/D snoRNP (also known as snR128) and the box H/ACA snoRNPs, snR30 and snR10, are also essential for production of 18S rRNA, and depletion of any of these four snoRNAs results in aberrant ribosome synthesis (**Li et. al., 1990; Morrissey and Tollervey, 1993; Tollervey, 1987**). Once these nuclear cleavages have occurred and the rRNA has folded, the 20S pre-rRNA is then exported to the cytoplasm and cleaved at site D by Nob1p to generate mature 18S rRNA (**Fatica et. al., 2003**).

Processing of the $27SA_2$ species occurs by one of two pathways. In the major pathway, site A_3 is cleaved by RNase MRP (mitochondrial RNA processing, chapter 1.2) in the nucleus, and nucleotides up to site B_{1S} are exonucleolytically digested by Rat1p and Rrp17p to form $27SB_S$ pre-rRNA (**Schmitt and Clayton, 1993; Chu et. al., 1994; Henry et. al., 1994; Lygerou et al., 1996; Oeffinger et. al., 2009**). In the minor pathway, site B_{1L} is cleaved by an unknown endonuclease, leaving a slightly longer form of 5.8S rRNA. Site B_2 is processed by Rex1p in both pathways (**Kempers-Veenstra et. al., 1986; van Hoof et. al., 2000**). Subsequently, cleavage occurs at site C_2 by an unknown nuclease followed by exonuclease digestion by Rat1p-Rrp17p to C_1, to separate mature 25S rRNA and a 7S pre-rRNA (**Geerlings et. al., 2000; Oeffinger et. al., 2009**). The 7S pre-rRNA is exonucleolytically digested from site C_2 by the exosome (a multi-subunit complex with exonuclease activity) and Rrp6p to form 6S pre-rRNA, before export to the cytoplasm (**Mitchell et. al., 1996; Mitchell et. al., 1997; Briggs et. al., 1998**). Rex1p/Rex2p, followed by Ngl2p exonucleolytically shorten the 6S pre-rRNA to site E to generate mature $5.8S_S$ or $5.8S_L$ rRNA (**van Hoof et. al., 2000; Faber et. al., 2002; Thomson and Tollervey, 2010**).
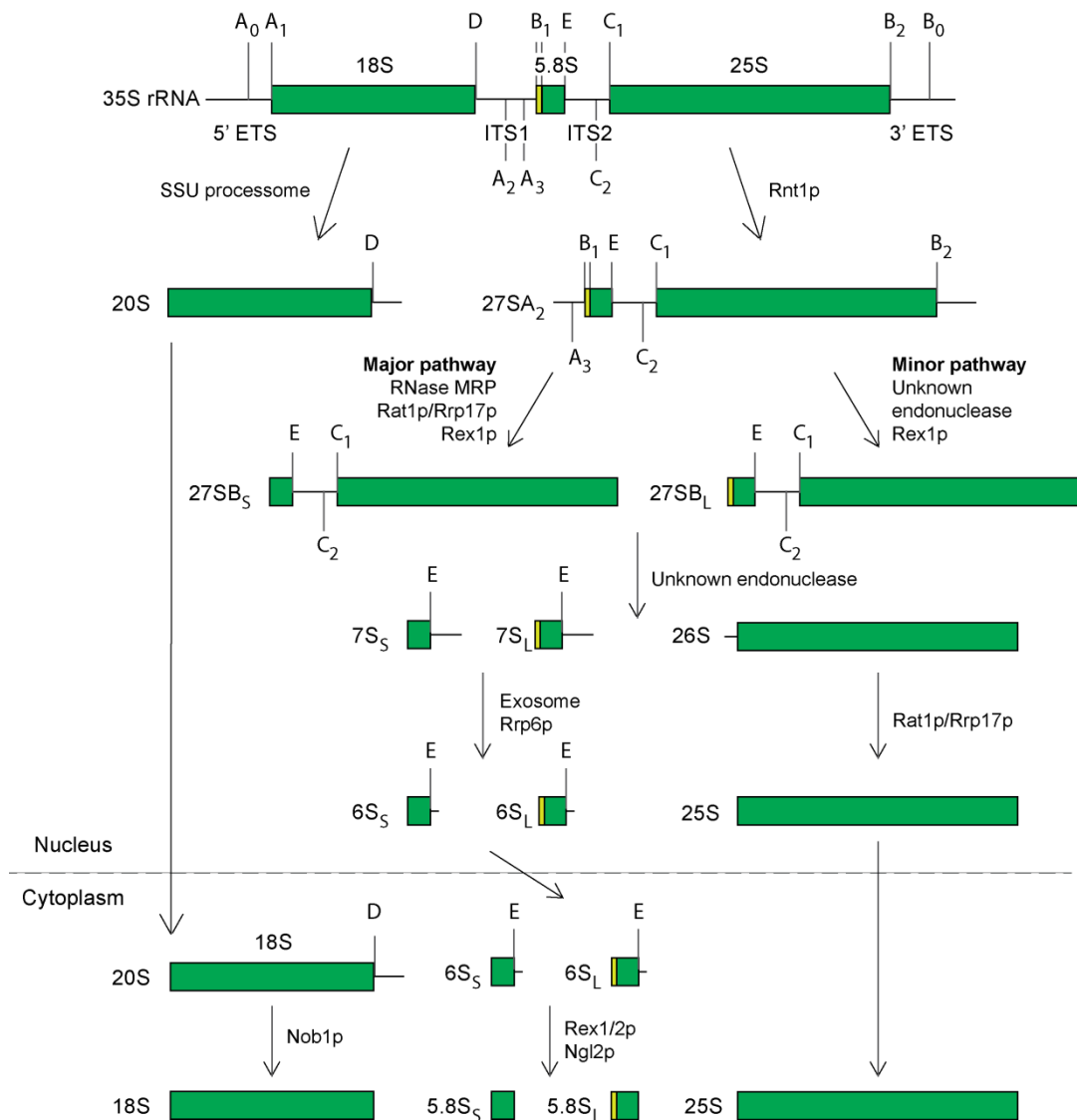
Figure 1.1.2 35S rRNA cleavage sites during ribosome biogensis. 35S rRNA is transcribed from rDNA. Cleavage sites are cleaved and processed sequentially by a number of endo- and exonucleases to produce the mature rRNAs: 18S, 5.8S and 25S. Green rectangles indicate the mature rRNAs (pale green indicates the additional nucleotides of $5.8S_L$ rRNA). Grey lines linked to letters indicate cleavage sites. Arrows indicate either processing by an endo/exonuclease or transport from the nucleus to the cytoplasm. Dashed line indicates the nuclear membrane. Adapted from Turowski and Tollervey, (2015).

18S rRNA associates with 33 ribosomal proteins (r-proteins) to form the 40S SSU, while 5.8S, 25S and 5S rRNAs associate with 46 r-proteins to form the 60S large subunit (LSU). These r-proteins directly interact with rRNA in mature ribosomes. R-

proteins can stabilise rRNA folds, or act as chaperones to alter misfolded rRNA. Most r-proteins are exposed on the ribosome surface, providing binding sites for other assembly factors (**Ben-Shem *et. al.,* 2011**). Numerous studies have been performed on individual r-proteins to determine their function in ribosome assembly. More recently, systematic analyses have been performed on r-proteins, showing that deficiency of any single r-protein results in a specific pre-rRNA processing step defect (**Ferreira-Cerca *et. al.,* 2005; Ferreira-Cerca *et. al.,* 2007; Pöll *et. al.,* 2009; Ohmayer *et. al.,* 2013**). Notably, their position within the ribosome is linked to their function in pre-rRNA processing. Ribosome biogenesis also requires the association of >200 assembly factors (**reviewed in Woolford and Baserga, 2013; Kressler *et. al.,* 2010**). These each have a specific function at a particular step in pre-rRNA processing or assembly. They include the endo- and exonucleases described above involved in pre-rRNA processing, along with enzymes such as RNA helicases, ATPases, GTPases and kinases. For the ribosomes to mature in the cytoplasm, assembly factors must be released from the pre-40S and pre-60S subunits, the remaining r-proteins must be incorporated, and the pre-ribosome scanned for correct assembly. Both assembly factors and r-proteins work together to prevent premature translation initiation. They coordinate this by overlapping binding sites of translation initiation factors or preventing subunit association (**Strunk *et. al.,* 2011; Gartmann *et. al.,* 2010**). Furthermore, the subunits undergo 'proof-reading' to test for correct translational functioning (**Bussiere *et. al.,* 2012; Lebaron *et. al.,* 2012; García-Gómez *et. al.,* 2014**). Final maturation steps occur, allowing the binding of 60S subunits to 40S subunits, and the initiation of translation.

Formation of the mature ribosome requires 76 different snoRNAs, the majority of which serve to guide modification of the pre-rRNA. This takes place cotranscriptionally, and results in 45 sites of pseudouridylation and 55 sites of 2'-*O*-ribose methylation (**Ofengand, 2002; reviewed in Sharma and Lafontaine, 2015**). These modifications provide structural stability, impose conformational constraints and alter the potential for RNA interactions (**Rife and Moore, 1998; reviewed in Helm, 2006; Charette and Gray, 2000**). Pseudouridylation increases the number of hydrogen bond donors (due to its additional NH residue relative to uridine), while methylation reduces hydrogen bonding potential. For the majority of ribosomal modifications, removal of individual modifications has no clear effect on ribosome biogenesis or cell viability but the lack of multiple or all modifications negatively affects growth rate, translational fidelity, and accumulation/stability of ribosomes

(**Parker *et. al.,* 1988; Tollervey *et. al.,* 1993; King *et. al.,* 2003; Liang *et. al.,* 2007; Liang *et. al.,* 2009; Jack *et. al.,* 2011**). Similarly, methylation directed to non-natural sites can heavily impact ribosome function and stability, and impair cell growth (**Liu and Fournier, 2004; Liu *et. al.,* 2008**).

Once the mature ribosomal subunits have been assembled, translation can commence. mRNAs contain codons, which are sets of three nucleotides that code for an amino acid. The ribosome scans from the mRNA cap for the start codon (AUG) to initiate translation (**reviewed in Jackson *et. al.,* 2010**). The SSU facilitates interactions between the codons of the mRNA and the anticodons of the tRNA. This determines the order of amino acids in the protein to be translated. The LSU contains the peptidyl-transferase centre (PTC). This catalyses the formation of peptide bonds in the protein to be translated. Between the SSU and LSU lie three binding sites for tRNAs: the A-site, P-site and E-site. The A-site binds aminoacyl-tRNAs, which must have correct codon-anticodon binding with the mRNA. The P-site is bound to peptidyl-tRNA, which is connected to the elongating peptide chain. This chain is transferred to the A-site tRNA when the peptide bond between the two amino acids is formed. The E-site (exit) contains a de-acylated tRNA, which is released upon translocation along the mRNA, stimulated by GTPase activity. This translocation causes the de-acylated tRNA in the P-site to move to the E-site, the peptidyl-tRNA in the A-site to move to the P-site, and the next aminoacyl-tRNA to bind to the A-site. Elongation of the peptide chain terminates when the ribosome reaches a stop codon in the mRNA, causing release of the protein through the peptide exit tunnel (**reviewed in Steitz, 2008; Wilson and Doudna Cate 2012**).

1.2: The different classes of snoRNA

There are two main types of snoRNA: box C/D and box H/ACA (Figure 1.2.1), which guide methylation and pseudouridylation, respectively (**reviewed in Bratkovic and Rogelj, 2011**). Generally, they are 60-300nt in length. Both types of snoRNA are transcribed by Pol II either from their own independent promoters, or excised from the introns of protein-coding transcripts. They are found either monocistronically or polycistronically, where they are processed by Rnt1p and trimmed by Rat1p/Xrn1p and the exosome (**Chanfreau *et. al.,* 1998; Petfalski *et. al.,* 1998; Allmang *et. al.,* 1999**). The majority of snoRNAs in *S. cerevisiae* are transcribed from their own

promoters (8 of the 77 *S. cerevisiae* snoRNAs are intronic, **Yoshihama *et. al.,* 2013**). However, in humans the majority of snoRNAs are excised from introns. Notably, a high frequency of these snoRNAs are found within the introns of ribosomal protein genes (RPGs) or nucleolar proteins, allowing co-regulation.

Box C/D snoRNAs contain two consensus motifs: RUGAUGA (R = A or G) denoted as box C, and CUGA, denoted as box D (Figure 1.2.1A). Most box C/D snoRNAs contain a second copy of each of these elements - denoted boxes C' and D', respectively – with more divergent sequences (**van Nues *et. al.,* 2011**). Stems are situated between the boxes. The antisense elements within box C/D snoRNAs are located immediately upstream of the D or D' boxes, binding to the target RNA with a 10-21nt target sequence. This guides the transfer of a methyl group from S-adenosyl-methionine to the 2'-hydroxyl group of the ribose on the target RNA nucleotide, positioned 5bp upstream of the 5' boundary of the D or D' box (figure 1.2.1C, denoted by the yellow star in figure 1.2.1A) (**Kiss-László *et. al.,* 1996**). Box H/ACA snoRNAs contain two hairpin domains connected by a hinge region (Figure 1.2.1B). They also contain two consensus motifs: ANANNA (N = any nucleotide) denoted as box H, and ACA, denoted as box ACA (**Ganot *et. al.,* 1997a; Balakin *et. al.,* 1996**). The latter is situated 3nt upstream of the 3' terminus. Either one or both of the hairpin stems includes large bulges containing an antisense loop 9-13nt long that binds to the target RNA, guiding the isomerisation of uridine to pseudouridine (pseudouridylation, figure 1.2.1D) (**Ganot *et. al.,* 1997b**). The pseudouridylated nucleotide is positioned 14-16nt upstream of the H and/or ACA box, denoted by the yellow star in figure 1.2.1B.

The biogenesis of these snoRNAs requires cotranscriptional recruitment of specific binding proteins to form functional snoRNPs. Box C/D snoRNAs are bound by four proteins: Nop1p (Fibrillarin in humans), Nop56p, Nop58p, and Snu13p (15.5K in humans). Nop1p is the protein catalysing transfer of the methyl group to the target nucleotide (**Tollervey *et. al.,* 1993; Niewmierzycka and Clarke, 1999; Wang *et. al.,* 2000; Galardi *et. al.,* 2002**). Snu13p binds directly to the stem formed between boxes C and D, providing the critical first step in box C/D snoRNP assembly (**Watkins *et. al.,* 2000; Watkins *et. al.,* 2002; Szewczak *et. al.,* 2002**). Its binding results in a structural rearrangement of the snoRNA, causing a kink-turn within the box C/D stem, which catalyses binding of the other snoRNP proteins and is homologous to the 5' stem loop of U4 snRNA. Nop58p binds at the C/D box motifs

and one copy of Nop1p binds at each of the D and D' boxes, then the C'/D' box motifs are bound by Nop56p (**Cahill _et. al.,_ 2002; van Nues _et. al.,_ 2011; Lafontaine and Tollervey, 2000**). Snu13p and Nop58p - but not Nop56p - are required for stability of box C/D snoRNAs (**Lafontaine and Tollervey, 1999**). Box H/ACA snoRNPs comprise four proteins: Cbf5p (Dyskerin in humans), Gar1p, Nhp2p and Nop10p. Cbf5p is the protein catalysing isomerisation of the target uridine base (**Lafontaine _et. al.,_ 1998; Zebarjadian _et. al.,_ 1999**). Nhp2p and Nop10p bind to Cbf5p in a ternary complex, which contacts the RNA. Gar1p binds last to Cbf5p, to form mature and functional snoRNPs (**Wang and Meier, 2004; reviewed in Massenet _et. al.,_ 2017**). Cbf5p, Nhp2p and Nop10p – but not Gar1p - are required for stability of box H/ACA snoRNAs (**Henras _et. al.,_ 2001; Watkins _et. al.,_ 1998; Lafontaine _et. al.,_ 1998**).

The third class of snoRNA present in the cell is the RNA component of RNase MRP. It was first discovered as an endonuclease that cleaves RNA primers for mitochondrial DNA replication, but has since been identified as the endonuclease that cleaves site A_3 in ITS1 of pre-rRNA, localised to the nucleus (described above, **Chang and Clayton, 1987; Schmitt and Clayton, 1993; Chu _et. al.,_ 1994; Lygerou _et al.,_ 1996**). RNase MRP consists of one RNA molecule and nine protein components and shows similarity to RNase P (**Chamberlain _et. al.,_ 1998**). This is the only member of its class, while both box C/D and box H/ACA snoRNAs comprise large families.
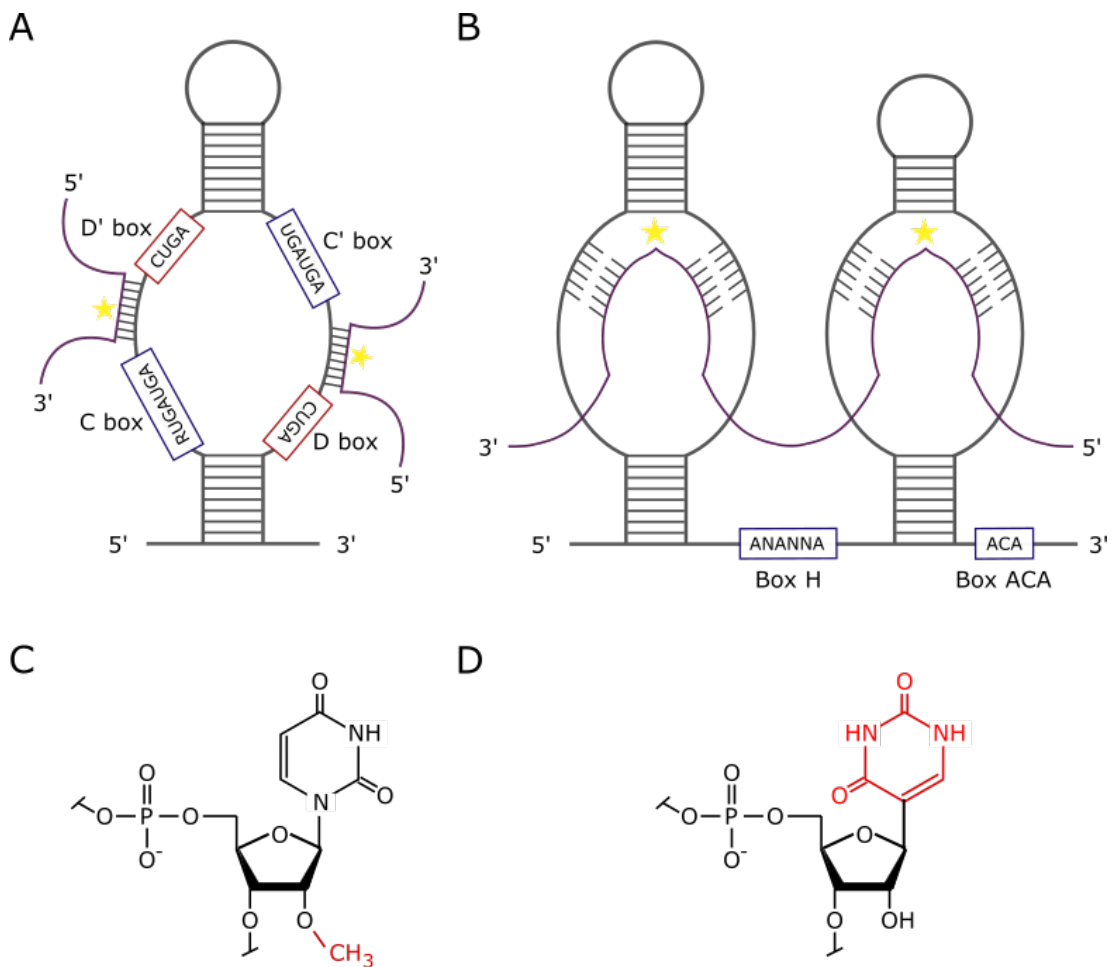
Figure 1.2.1 The two main types of snoRNA: box C/D and box H/ACA. A) Model of the structure of box C/D snoRNAs, showing target RNA in purple, the methyl group attached to the 5th nucleotide from the D or D' box (yellow star), and the complementary binding to the antisense elements between the C and D' boxes and the C' and D boxes. B) Model of the structure of box H/ACA snoRNAs, showing target RNA in purple, the isomerised uridine to pseudouridine (yellow star) and complementary binding to the anti-sense elements. Consensus sequences are written inside the labelled boxes. Horizontal lines indicate Watson-Crick base-pairing. C) The methylation of the 2'-oxygen of the ribose group, showing the added methyl group in red. Uridine is used as an example base. D) The converted uridine to pseudouridine (red) attached to the ribose. Figure adapted from Bratkovic and Rogelj, (2011).

scaRNAs (small Cajal body-specific snoRNAs) are snoRNAs that specifically localise to Cajal bodies (Jády and Kiss, 2001; Darzacq *et. al.,* 2002). Cajal bodies are sub-nuclear organelles that contain high concentrations of small nuclear RNPs

(snRNPs, functioning in pre-mRNA splicing) and play a role in the assembly and modification of snRNPs (**reviewed in Bohmann *et. al.,* 1995; Gall, 2000; Nizami *et. al.,* 2010**). scaRNAs accumulate and reside in Cajal bodies due to the presence of a Cajal body-specific sequence motif (Cajal body box, CAB) that results in their localisation (**Richard *et. al.,* 2003**). This localisation is in contrast to the box C/D and box H/ACA snoRNAs, which reside in the nucleolus. The first scaRNA discovered was shown to direct the 2'-*O*-methylation and pseudouridylation of the Pol II-transcribed U5 snRNA (**Jády and Kiss, 2001**). Six additional putative scaRNAs were then characterised, predicted to direct both the 2'-*O*-methylation and pseudouridylation of the Pol II-transcribed U1, U2, U4 and U5 snRNAs (**Darzacq *et. al.,* 2002**). While these scaRNAs were shown to contain the elements of both box C/D snoRNAs and box H/ACA snoRNAs, other scaRNAs only contain the elements of one type of snoRNA.

Recent investigations have reported the accumulation of short snoRNA fragments. These processed snoRNAs (psnoRNAs), snoRNA-derived RNAs (sdRNAs) or sno-microRNAs (sno-miRNAs) identified in higher eukaryotes purportedly generate miRNAs, function similarly to miRNAs or associate with non-canonical proteins. They were initially identified in a screen for RNA binding partners of the miRNA-associated Argonaute proteins: Ago1p and Ago2p. A box H/ACA snoRNA processed to 20-25nt was found associated with the Ago proteins, and a target mRNA was identified. It was found to be processed *in vitro* by the miRNA-associated Dicer complex. Other box H/ACA snoRNAs were also identified as having this potential miRNA function (**Ender *et. al.,* 2008**). An evolutionary analysis was then performed between miRNAs and snoRNAs to determine the similarity between the two types of regulatory RNA, which identified a number of miRNAs that showed genomic and functional similarities to box H/ACA snoRNAs (**Scott *et. al.,* 2009**). A further two studies reported sdRNAs from box C/D snoRNAs with miRNA-like functions (**Brameier *et. al.,* 2011; Ono *et. al.* 2011**). Further to this, a mouse box C/D snoRNA, MBII-52 – homologue of a human box C/D snoRNA linked to Prader Willi Syndrome (PWS, discussed in chapters 1.3 and 1.4) - was reported to be processed into smaller RNA forms that did not associate with canonical box C/D–associated proteins. These psnoRNAs were postulated to affect alternative splicing of mRNAs (**Kishore *et. al.,* 2010**). However, in response to this, another group performed their own investigation that found no evidence for these species (**Bortolin-Cavaillé and Cavaillé, 2012**). It instead argued that these psnoRNAs are

more likely to represent metabolically stable degradation intermediates, which may possibly develop a functional role under stress conditions. Whether these fragmented snoRNA species are indeed precursors for miRNA-like RNAs, affect alternative splicing or are misidentified stable degradation intermediates remains unclear. Further studies will be needed to determine their existence, biogenesis and function.

The focus of this study is the targets of box C/D snoRNAs. Canonical targets of box C/D snoRNAs and their enzymes are rRNAs or snRNAs. The strict positioning requirement of box C/D snoRNAs allows *in silico* predictions of the targets for methylation by each guide snoRNA. These predictions must fit stringent criteria. Firstly, the hybrid must contain ≥12 base-paired nucleotides with the 5th nucleotide upstream of the D/D' box base-paired exactly with rRNA. In addition, one or more nucleotides between the D/D' box and D+5 nucleotide must base-pair exactly with rRNA; and either a stretch of eight nucleotides from 1-14nt upstream of the D/D' box base-pairs exactly with rRNA (no mismatches) or a stretch of 10 nucleotides from 1-16nt upstream of the D/D' box base-pairs exactly with rRNA (at most one mismatch). A predicted folding energy of ≤-12ΔG can also be used as a cut-off to further filter hybrids for validity (**Kudla *et. al.,* 2011**).

Three yeast box C/D snoRNAs do not meet these criteria: the highly conserved U3 snoRNA – required for pre-rRNA processing in all eukaryotes tested (described above) - and the orphan snoRNAs, snR4 and snR45. The latter two lacked any known targets or function despite binding to the box C/D proteins. Only two orphan snoRNAs have been identified in yeast, but >200 have been identified in humans. Furthermore, snoRNAs that also contain a D' box contain two possible methylation guide sequences, and so can have two binding sites. They may therefore possess orphan binding sites, and thus have a dual-function. Examples of this are the seven snoRNAs processed from a polycistronic transcript in yeast, named snR72-78. These contain methylation guide sequences, but also possess orphan binding sites, and thus may have dual-function. Similarly, the box C/D snoRNA, snR190, guides methylation of a site in 25S rRNA through its D' box, but shows no complementary base-pairing to rRNA or snRNA upstream of its D box. Therefore, a large number of snoRNAs that are still retained in the genome have unknown targets and functions.

1.3: Potential targets of orphan snoRNAs

A number of studies have demonstrated interactions between orphan snoRNAs and mRNAs. One such study indicated the involvement of human snoRNAs in alternative splicing of mRNA (**Kishore and Stamm, 2006**). Multiple repeated copies of the human box C/D snoRNA HBII-52 - now designated SNORD115 - or its mouse homologue MBII-52 show complementarity to exon Vb of the serotonin receptor 5-HT$_{2C}$R. Inclusion of this exon encodes for the fully functional receptor. This exon is subject to adenosine-to-inosine editing, which promotes its inclusion, but causes an amino acid change resulting in a less functional receptor (**Burns *et al.,* 1997**). Experiments have indicated that binding of SNORD115 influences alternative splicing, promoting exon Vb inclusion in the mRNA by partially blocking a silencer (**Kishore and Stamm, 2006**). Other studies, however, predicted that the snoRNA can inhibit A-to-I editing by potentially methylating one of the edited sites, and showed that lack of SNORD115 altered RNA editing but not alternative splicing (**Vitali *et al.,* 2005; Doe *et. al.,* 2009**). This underlines the limited understanding of how orphan snoRNAs function, and the variety of processes they may affect.

Supporting the hypothesis that orphan snoRNAs may affect the alternative splicing of pre-mRNAs, another study reported the interaction of SNORD27 and an alternatively spliced exon of the *E2F7* transcription factor (**Falaleeva *et al.,* 2016**). SNORD27 is a dual-function snoRNA, as it guides the methylation of site A27 in 18S rRNA, but contained predicted complementary binding to pre-mRNA. Knockdown of SNORD27 reduced alternative exon skipping in *E2F7*. Consistent with the work by Kishore and Stamm, *E2F7* was not methylated at the SNORD27 binding site. It was also shown to affect the alternative splicing of a number of other pre-mRNAs. However, this study also reported that SNORD27 was detected in a fraction where Fibrillarin was not detected. In another study, a change in the alternative splicing of the fibroblast growth factor receptor, *FGFR3,* was observed upon overexpression of a human box C/D snoRNA, HBII-180C (**Scott *et al.,* 2012**). However, this was reported to be a box C/D derived sdRNA, rather than a *bona fide* snoRNA. While the human genome contains many introns, the yeast genome has few introns, which are generally present at only one per gene (**Spignola *et al.,* 1999**). This may indicate that the participation in alternative splicing of the yeast orphan snoRNAs is less likely.

Another study demonstrated that snoRNAs can be altered so that they target specific mRNAs for degradation (**Ono _et al.,_ 2010**). snoMEN vectors (<u>sno</u>RNA <u>m</u>odulator of gene <u>e</u>xpressio<u>n</u>) were developed, which replaced the 'M box' sequence present in a subset of snoRNAs with sequences complementary to target RNAs, delivered within a vector. The M box is a sequence of ~20nt located between the C and D boxes in these snoRNAs that showed almost perfect complementarity to endogenous pre-mRNA sequences. These altered snoRNAs reduced expression of the RNAs to which they were targeted. Similarly, knockdown of a human box C/D snoRNA, SNORD83B, resulted in an increase in steady-state levels of three mature target mRNAs (**Sharma _et al.,_ 2016**). This further indicates a function of orphan snoRNAs in mRNA stability.

To date, no study has demonstrated a general mechanism or function for snoRNA-mRNA interactions. It remains unclear whether binding occurs during mRNA transcription, splicing, translation or degradation. The highly regulated turnover of mRNAs suggests that snoRNAs may be involved in mRNA degradation, or protection from such degradation. This might result from interactions with the exosome nuclease complex or with its cofactors. However, canonical function of snoRNAs may suggest a role in mRNA translation, linking this to ribosome assembly/function.

1.4: Orphan snoRNAs are implicated in human disease

Studies have emerged indicating the involvement of orphan snoRNAs in human disease. The box C/D snoRNA, SNORD115 (discussed in chapters 1.2 and 1.3) affects the alternative splicing of the serotonin receptor 5-HT$_{2C}$R. PWS (discussed in chapter 1.2) is a disease caused by loss of paternally expressed genes at a maternally imprinted locus, and results in muscular hypotonia and slow development (**Nicholls _et. al.,_ 1998; Sahoo _et al.,_ 2008**). Multiple tandem copies of SNORD115 are found at this locus (**Cavaillé _et al.,_ 2000**). Patients who lack the paternally expressed genes do not express SNORD115, and develop PWS. Knowledge of the mechanism linking SNORD115 and the serotonin receptor will allow for the development of potential therapeutics. This highlights the importance of understanding the interactions of snoRNAs with respect to human disease.

miRNA expression levels are altered in cancer cells, and thus have been used for diagnosis and prognosis of cancer patients (**reviewed in Zhang *et. al.,* 2007; Paranjape *et. al.,* 2009**). Recently, these applications have been demonstrated with snoRNAs as well (**Mannoor *et. al.,* 2012**). The first study to show a link between snoRNAs and carcinogenesis demonstrated that a human box H/ACA snoRNA, h5sn2, was highly expressed in normal brain tissue, but showed heavily reduced expression in meningioma tissue (**Chang *et al.,* 2002**). Another study analysed the deletion of chromosome 6 in prostate cancers. It postulated that this region may contain a tumour suppressor gene, and identified the box C/D snoRNA, U50, within this region. U50 was predicted to guide the methylation of C2848 and G2863 in 28S rRNA, using its D and D' boxes, respectively (**Kiss-László *et. al.,* 1996**). Prostate cancer was correlated with either chromosomal loss of U50, or a homozygous 2bp deletion within U50 - which was not present in normal cells and abolished its function of suppressing colony formation (**Dong *et al.,* 2008**). The same snoRNA has been identified with the same 2bp deletion in cell lines and primary tumours from breast cancer (**Dong *et al.,* 2009**). In this case, heterozygous deletion showed correlation with breast cancer. For both prostate cancer and breast cancer, U50 showed significant transcriptional downregulation.

This knowledge has been applied to determine the prognosis of patients. Gee *et. al.* analysed the expression level of supposed housekeeping snoRNA genes in breast cancer and head and neck squamous cell carcinoma (HNSCC). Expression levels of three box C/D snoRNAs (RNU43, RNU44 and RNU48) showed high variability in cancer cells, invalidating their use as housekeeping genes in cancerous tissue. RNU43 and RNU44 expression level showed correlation with tumours. Furthermore, RNU44 maps to an intronic region of the growth arrest specific 5 gene (*GAS5*), which encodes multiple snoRNAs, but no protein product. Downregulation of both RNU44 and *GAS5* was associated with poor prognosis for both cancer types (**Gee *et al.,* 2011**). A similar effect was also observed in hepatocellular carcinoma cells (HCC), where the box C/D snoRNA 113-1 (SNORD113-1) was significantly downregulated in HCC tissues, correlated with decreased survival rate of patients (**Xu *et al.,* 2014**). Furthermore, the promoter of SNORD113-1 showed CpG island hypermethylation - which has been shown to be common in human cancers - and overexpression of SNORD113-1 inhibited cell viability and cancer proliferation (**Ferreira *et al.,* 2012**).

Further studies have indicated a diagnostic potential for human snoRNAs. A recent study identified six human box C/D snoRNAs that were overexpressed in the surgical tissue of non-small cell lung cancer patients (NSCLC) as compared to non-cancerous lung tissue (**Liao *et al.,* 2010**). Three of these – SNORD33, SNORD66 and SNORD76 - showed significantly higher expression level in the blood plasma of NSCLC patients compared to healthy individuals, as measured by RT-qPCR (reverse-transcription quantitative polymerase chain reaction). Analysis of the expression of these three snoRNAs may soon be used for early detection of NSCLC. Similarly, snoRNAs were found to have diagnostic potential for Peripheral T-cell lymphoma (PTCL; **Valleron *et. al.,* 2012**). U75 expression (another snoRNA from a *GAS5* intron) allowed the distinction between anaplastic large cell lymphoma (ALCL) and other PTCLs. Within ALCL, patients may either have or not have the anaplastic lymphoma kinase gene rearrangement (ALK$^+$ or ALK$^-$). U3 expression levels distinguished between ALK$^+$ and ALK$^-$ subtypes. Correct diagnosis of these different subtypes of PTCL is important for proper treatment and prognosis. Within the same study, it was found that high expression of SNORD HBII-239 in non-ALCL PTCL was associated with good prognosis. These studies further emphasise the significance of identifying orphan snoRNA functions and the phenotypes associated with them.

1.5: Methods of identifying the RNA targets of orphan snoRNAs

To determine a function for orphan snoRNAs, targets must first be identified. A number of methods have been developed to elucidate protein-RNA and RNA-RNA interactions, and are discussed below.

RIP-Seq (RNA immunoprecipitation) was developed by Zhao *et. al.,* to identify RNAs associated with the Polycomb repressor complex (**Zhao *et al.,* 2010**). The native protein-RNA complex is immunoprecipitated with antibodies for the protein of interest. The RNA is digested and reverse-transcribed to synthesise cDNA. Linkers are ligated and the RNA is PCR-amplified and gel-purified, before Illumina sequencing. This allows for single base-pair resolution of the protein binding site. However, it requires specific antibodies for the protein of interest, and contains no stabilisation or cross-linking step to preserve the interaction, leading to false negatives.

RPL (RNA proximity ligation) was developed to provide data about RNA secondary and tertiary structures (**Ramani *et al.,* 2015**). RNAs are partially digested, then treated with RNA ligase. This generates chimeras from any RNA strands with intra-molecular links, which are then deep sequenced. RPL also does not utilize crosslinking to identify stable interactions and thus is mostly limited to mapping intramolecular RNA interactions, rather than trans interactions.

CLIP (cross-linking and immunoprecipitaion, **Ule *et al.,* 2003**) was combined with high-throughput sequencing to generate HITS-CLIP (high-throughput sequencing of RNA isolated by CLIP; also known as CLIP-Seq, **Licatalosi *et al.,* 2008**). This method includes a cross-linking step missing from RIP-Seq and RPL. Cells are UV cross-linked then lysed and treated with RNase T1. The protein of interest is immunoprecipitated followed by gel purification, proteinase K digestion, linker ligation and RT-PCR. Illumina sequencing allows visualisation of the protein-RNA interaction site. The cross-linking step allowed for stabilisation of protein-RNA interactions. However UV cross-linking was not optimally efficient, requiring close protein-RNA interactions, and allowed the introduction of artefacts into the dataset. This led to the development of PAR-CLIP (photoactivatable-ribonucleoside-enhanced CLIP, **Hafner *et al.,* 2010**). PAR-CLIP incorporates 4-thiouridine (4SU) or 6-thioguanosine (6SG) into the transcripts of cultured cells. Cells are then UV cross-linked, immunoprecipitated and treated with RNase T1. RNA is then radiolabelled, gel purified, treated with proteinase K and a cDNA library prepared. Deep sequencing provides accurate mapping of protein-RNA interactions. Use of 4SU or 6SG improved cross-linking efficiency, and the protein binding site could be precisely identified by the presence of mutations in the cDNA sequence. However, photoreactive nucleosides can be cytotoxic, and high concentrations of 4SU can inhibit ribosome synthesis and activate the nucleolar stress response (**Burger *et al.,* 2013**). Furthermore, this method is limited to cell culture and *in vitro* work.

Also derived from CLIP is iCLIP (individual-nucleotide resolution CLIP, **König *et al.,* 2010**). CLIP could not resolve binding sites <30nt long, and cDNAs would prematurely truncate 5' to nucleotides that were cross-linked to residual amino acids. Therefore, iCLIP was developed to acquire single nucleotide resolution utilising the truncated cDNAs. Cells are UV cross-linked and immunoprecipitated. An RNA adapter is then ligated to the 3' end and the protein is digested by proteinase K. Transcripts are reverse-transcribed, using a random barcode sequence plus two

cleavable adapter regions as a primer. The cDNA is circularised to allow for sequencing of the barcode followed by the last nucleotide added during reverse transcription, followed by the remaining truncated cDNA. This is then linearised, PCR-amplified and sequenced. The random barcode allows for discrimination between unique cDNA products and PCR duplicates, and the first nucleotide 3' to this identifies where the cDNAs truncated during reverse-transcription. This method allows nucleotide resolution of binding sites, whilst avoiding the use of nucleases. However, artefacts can be introduced at the circularisation step.

All of these methods identify protein binding sites on RNA. However, to identify targets of RNAs, hybrids must be obtained. hiCLIP (RNA hybrid and individual-nucleotide resolution CLIP) was developed to identify RNA duplexes complexed with ribosome binding proteins (**Sugimoto *et al.,* 2015**). Cells are UV cross-linked, RNAs are partially digested, and the complex immunoprecipitated. Linkers are ligated to the 5' end of each RNA, then a ligation performed between the two RNAs via one of the linkers. The protein is digested, a cDNA library prepared and the library sequenced. The chimeric sequences can then be mapped to the genome, with the internal linker allowing for differentiation between the two reads.

CLASH (cross-linking, ligation and sequencing of hybrids) is a similar technique that also allows for hybrid identification (**Kudla *et al.,* 2011,** Figure 1.5.1). This technique was adapted from CRAC (cross-linking and analysis of cDNAs) and involves UV cross-linking RNAs bound by a tagged protein (**Granneman *et al.,* 2009**). For these studies, an HTP (His-Tev-Protein A) tag is used. The bait protein is UV cross-linked *in vivo* to any associated RNAs and undergoes a two-step purification. Cross-linked RNAs are ligated with barcoded linkers. At this step, the ends of any RNAs that are in duplexes associated with the bait protein can also be ligated together to form a chimeric molecule. Ligation is done at a low temperature overnight to stabilise hybrids between the two bound RNAs, as compared to higher temperatures in CRAC which favour linker ligation to the separate RNAs. Following protein digestion, RT-PCR and sequencing of the resulting cDNAs gives a list of RNAs bound by the bait proteins. Mapping of the cDNAs to the genome can identify hybrid cDNAs, and locate the target RNAs of an RNA of interest. This method has the same advantages as hiCLIP: it maps protein-RNA-duplex interactions *in vivo,* it provides binding site resolution and reveals target RNA identity. However, very short RNA strands are constrained with respect to their ligation, indicating that one of the two

RNA strands must be flexible enough for ligation to occur. An advantage CLASH has over hiCLIP is that if the bait protein is known and can be tagged, this removes the use of antibodies and the non-specific targets they may introduce to the datasets.
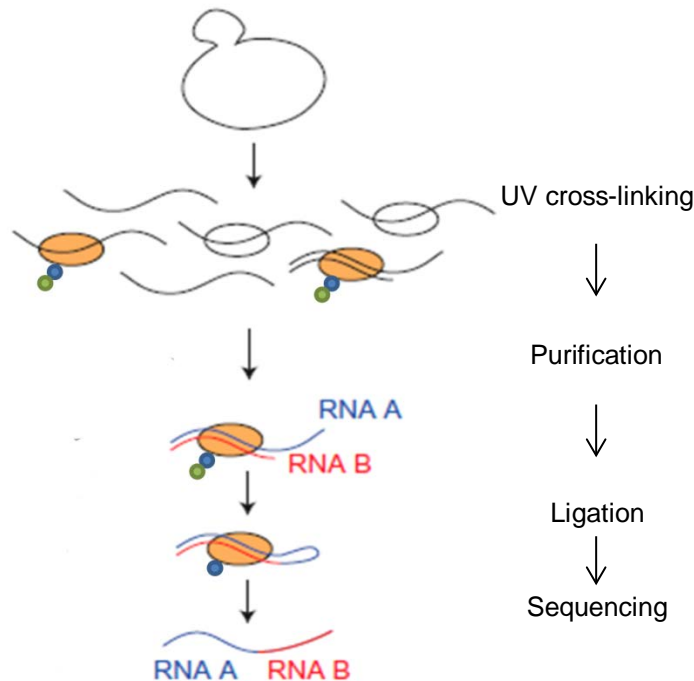
Figure 1.5.1 CLASH schematic. *S. cerevisiae* proteins of interest are tagged and cross-linked *in vivo*. They undergo a two-step purification, ligation, protein degradation and sequencing. The result is a range of chimeras of the RNAs that the protein binds to, giving a list of targets. Orange ovals represent the tagged protein. Green circles represent the Protein A tag, and blue circles represent the His tag. The blue RNA represents one RNA sequence, while the red RNA represents its RNA binding partner. Adapted from Kudla *et al.,* (2011).

Recent studies have further developed methods used to obtain RNA hybrids. SPLASH (sequencing of psoralen cross-linked ligated and selected hybrids) uses biotinylated psoralen to cross-link RNAs *in vivo* (**Aw *et al.,* 2016**). The RNA is extracted, fragmented and enriched on beads, before proximity ligation and reverse crosslinking. Adapters are ligated and the RNA reverse-transcribed, followed by cDNA circularisation (as in iCLIP) and PCR amplification. The circularisation protocol was found to capture chimeras more efficiently than independent 5' and 3' adapters, and allows for reorientation of the cDNA before PCR amplification, as in

iCLIP. The cDNA library is then deep sequenced. Use of psoralen allows for efficient cross-linking to stabilise the RNA interactions, while the biotinylation allows for purification of the duplexes on beads, removing the need for antibodies. This method also uses a fragmentation step instead of using RNases - which must be carefully controlled. Furthermore, it does not require use of a bait protein or knowledge of either RNA. However, psoralen shows a sequence bias, and there are limits on the intracellular concentrations that can be used.

Two more methods were developed within the same month, both using a psoralen derivative, AMT (4′-aminomethyltrioxsalen). PARIS (psoralen analysis of RNA interactions and structures) uses AMT to *in vivo* cross-link RNA duplexes (**Lu *et al.,* 2016**). These are partially RNase-digested and treated with proteinase, followed by gel purification. Samples undergo proximity ligation, are reverse cross-linked, reverse-transcribed and then sequenced. This method can ascertain long-range structures, single RNA structures and complex structures such as pseudoknots. LIGR-seq (ligation of interacting RNA followed by high-throughput sequencing) also uses AMT to cross-link RNA duplexes (**Sharma *et al.,* 2016**). Cells are lysed and the RNAs partially digested using S1 endonuclease. Ligation occurs using circRNA ligase, and is followed by treatment with an exoribonuclease to remove uncross-linked, linear RNAs. Samples are then reverse cross-linked, and a sequencing library prepared. AMT cross-linking is efficient, which reduces the rates of false-positive and –negative detection. A control sample (-AMT) can also be prepared alongside the library, for normalisation purposes. Similar to SPLASH, these methods also do not require use of a bait protein or knowledge of either RNA. However, again psoralen cross-linking shows a sequence bias.

The multitude of methods for capturing protein-RNA interactions listed above all possess advantages and disadvantages. The work to be undertaken in this study will investigate the RNA targets of a known subset of RNAs – snoRNAs – with known sequences. Furthermore, the well-studied biogenesis of snoRNPs has identified a number of bait proteins that may be used. Previous CLASH on snoRNAs showed that the function of the small Snu13p protein is impaired by tagging. However, Nop1p, Nop56p and Nop58p are easily tagged and appear to be fully functional. Therefore, a method that uses a bait protein is possible, and gives more targeted specific results for that RNP. CLASH has the advantage over hiCLIP by not requiring use of antibodies, and also that the sequence of one RNA within the

duplex is known. Moreover, expertise in the lab provides optimised protocols, and more efficient analysis of bioinformatics.

1.6: Aims

S. cerevisiae is used as the model organism in this study. Containing ~6,000 genes, yeast possess a relatively compact genome compared to humans, which contain at least 20,000 genes. However, there is significant overlap between the essential processes in both yeast and humans, meaning many genes are conserved. Furthermore, as yeast is one of the simplest eukaryotic organisms, it is easy to grow and manipulate, and experimental effects are much easier to study. >200 orphan snoRNAs have been detected in humans, whereas only two have been identified in yeast. Investigating the function and phenotype of the two orphan snoRNAs in yeast may expose a common function for the human orphan snoRNAs and elucidate mechanisms through which they act.

In light of the prior work described above, unresolved questions remain. Firstly, the consequence of deletion of yeast orphan snoRNAs is currently unknown. No general mechanism has thus far been proposed for the consequences of snoRNA-target interactions that do not direct RNA methylation. Studies on human data present evidence for alternative splicing; however, only 5% of the genes in *S. cerevisiae* contain introns and >95% of those intron-containing genes possess only a single intron, leaving few targets to study. The most likely starting hypothesis was that yeast orphan snoRNAs might have an effect on mRNA degradation or steady state levels. Were such an effect identified, the mechanisms that underpin these interactions could be elucidated, and the factors involved in the process uncovered.

Three main aims were thus proposed:-

- Aim 1: Identify a phenotype associated with deletion of orphan snoRNAs

- Aim 2: Characterise the effect of snoRNA knockdown on target mRNAs in yeast

- Aim 3: Determine the mechanisms by which orphan snoRNA binding alters target RNA pathways, and other factors involved in these interactions

The subsequent chapters will address each of these aims in turn. This was achieved by investigating the growth phenotype of snoRNA deletion strains and performing a genome-wide screen for synthetic genetic interactions (see chapter 3), analysing the non-canonical interaction with mRNAs by use of high-throughput and small scale experiments to reveal the effect of snoRNA deletion on mRNA steady state levels (see chapter 4), and finally by examining conservation of yeast orphan snoRNAs and revealing a novel function with respect to rRNA (see chapter 5).

2 MATERIALS AND METHODS

2.1: Materials

YPD solid medium was prepared by mixing 1% w/v yeast extract, 2% w/v peptone, and 2% w/v agar with H_2O up to 90% final volume. After autoclaving, glucose (D) was added to 2% w/v final concentration. Antibiotics were added as needed, using a concentration of 400µg/ml G418 (Geneticin), 200µg/ml Nourseothricin or 600µg/ml Hygromycin. Sugar was altered as necessary. For media using multiple sugars, the volume of water was adjusted accordingly.

Synthetic solid medium was prepared using 0.69% w/v Formedium Yeast Nitrogen Base (YNB), 2% w/v agar, Formedium Complete Supplement Mixture (CSM) as directed per mixture (790 mg/L Complete, 770mg/L –Ura, 770mg/L –His, 740mg/L –Trp) and H_2O up to final volume. After autoclaving, 2% w/v sugar and antibiotics were added as required. Liquid media were made by the same solutions without agar addition.

For nutrient shifts, a 10x solution of nitrogen shift medium was made by dissolving 19g/L Formedium YNB without amino acids without ammonium sulphate, 0.86g/L of Histidine, 0.86g/L of Uracil, 0.86g/L of Methionine, 1.72g/L of Leucine and 10g/L of Proline in H_2O.

Fast lysis buffer was made of 2% v/v Triton, 1% w/v Sodium Dodecyl Sulfate (SDS), 100mM NaCl, 1mM Ethylene Diamine Tetraacetic Acid (EDTA), and 10mM Tris pH8.

GTC was made by dissolving 100g of Guanidium thiocyanate in 100ml H_2O, with 10.6ml 1M Tris pH8, 4.24ml 0.5M EDTA pH8 and 20% w/v sarkosyl.

10X BPTE was prepared by mixing 30g/L PIPES (100mM), 60g/L Bis-Tris (300mM), 20ml/L 0.5M EDTA pH8 (10mM) and H_2O up to final volume.

20X SSC was prepared by mixing 175.3g/L NaCl (3M) with 88.2g/L sodium citrate (0.3M) and H_2O up to final volume, pH7.

Name	Description	Sequence
snoRNA deletions		
snr4 5'	SNR4 deletion with pFA6a-KanMX6 F	TAGTTTTTTGTCATTGATCTTTCATTTTTTATTTCAAAATCCCCATCGGATCCCCGGGTTAATTAAG
snr4 3'	SNR4 deletion with pFA6a-KanMX6 R	ACCCAGGTGAGACTGGATCGTCCATAGATTCCAAGATTTACGTAAGAATTCGAGCTCGTTTAAAC
snr4 CF	Check SNR4 deletion F	AACTATTACAGTCGATGAGG
snr4 CR	Check SNR4 deletion R	AAGTCATGCGTTACCAGAGC
snr45 5'	SNR45 deletion with pFA6a-KanMX6 F	ATTTTACCAAGACGCGTTAGATTTTCTTTGGAGGTTTCATTTAAAGAAACGGATCCCCGGGTTAATTAAG
snr45 3'	SNR45 deletion with pFA6a-KanMX6 R	CAGATGAGATGACTACTCCCAAGTCAAAAACAACAAAAGAATGAAAGAATTCGAGCTCGTTTAAAC
oRP-005	Check SNR45 deletion F	ACCCGTAGAGAAGTGCTGAA
oRP-006	Check SNR45 deletion R	TCGCTCCGAGAAGAATTGTTC
snr72-snr78 F	SNR72-78 deletion F	TACAAACTCTATCTTTCATAGAATACGTCCTTCGATACCATTTCACCAGAATTCGAGCTCGTTTAAAC
snr72-snr78 R	SNR72-78 deletion R	AATTAGCCTGTGATAATGACAAAAGACATATCTAAAAACTAACTTTCATTTTGAGATCCGGGTTTT
snr72-78 CF	Check SNR72-78 deletion F	AATTCTAATATTTGGCTTAC
snr72-78 CR	Check SNR72-78 deletion R	AATTAGCCTGTGATAATGAC
MX4-6F	Check MX6 insert F	CCTCGACATCATCTGCCCAGAT
MX4-6R	Check MX6 insert R	TGCAGCGAGGAGCCGTAAT
oRP-083	snR4F chunk	GCACCTTGTAAGCAGTCGAG
oRP-084	snR4R chunk	CATGCGTTACCAGAGCCAAC
oRP-085	snR45F chunk	GGATTACCCGTAGAGAAGTGC
oRP-086	snR45R chunk	GGCACGGTTAAAGGACACA
oRP-109	snR4F chunk2	TGCGGCATGTAACTTCGTTCA
oRP-110	snR4R chunk2	ACGATAATATTGCCCCATTTCCT
oRP-112	snR45F chunk2	TCATCTCATCTCATCAAAACCGT
oRP-113	snR45R chunk2	GACAGACCCCAGGCGTATTT
GAL1 Promoter tagging of snoRNAs		
oRP-063	PGAL1-SNR4 F	GAAGTGTTCTTGTGCGATTGAAGAAGGTTTATTTCGTTTCGCTGTGTGAATTCGAGCTCGTTTAAAC
oRP-064	PGAL1-SNR4 R	ATGTTACACATGAACTATAACCTATCCTCGACTGTAATAGTTGATACTTTTATTACATTTGAATAAGAAG
oRP-065	PGAL1-SNR45 F	CGATTTTCATTTAACAGGTATAAAAAGCGAAACACTCGGTACATAGAATTCGAGCTCGTTTAAAC
oRP-066	PGAL1-SNR45 R	ACGCAAATAATATCATCGTATTCTTTAAAAAACTTGGAAGGTCATGATACTTTATTACATTTGAATAAGAAG
oRP-067	Check PGAL1-SNR4 promoter F	GAAGGTTTATTTCGTTTCGCTGT
oRP-068	Check PGAL1-SNR4 promoter R	ATCCACATCGACCCAGGAAA
oRP-069	Check PGAL1-SNR45 promoter F	ACCCGTAGAGAAGTGCTGAA
oRP-070	Check PGAL1-SNR45 promoter R	AGCGTCCTTGTCATTCTTAAGA

Name	Description	Sequence
Creation of *nop1* mutants		
oRP-103	*nop1* mutants F	GTTATTCTTCGAGAAACAATTAG
oRP-104	*nop1* mutants R	ATATGGCGAATACGTACAATAC
qPCR		
oRP-013	*FBA1* F	TAAGCAGCACCACCGTTAGA
oRP-014	*FBA1*R	CGGTGTCATCGTTGGTGAAG
oRP-025	*CCW12* F2	CTTCAGTGGTCAATGGGCAC
oRP-026	*CCW12* R2	CGCTGCTAACGTTACCACTG
oRP-037	*TAF10* F	GCTAACAACAGTCAGGCGAG
oRP-038	*TAF10* R	GAGCCCGTATTCAGCAACAG
oRP-041	*PDC1* F	AGCAGACAATTCACCGACAC
oRP-042	*PDC1* R	CTTACGCCGCTGATGGTTAC
oRP-043	*TDH3* F	TGCTTGTCATCGTGGGAAAC
oRP-044	*TDH3* R	CGACCCATTCATCACCAACG
oRP-045	*RPL28* F	CCCAGAAGACAAGAGAGACCA
oRP-046	*RPL28* R	TTCAACAACACCACCAGCAG
oRP-049	*RPS20* F	GGTCAAGAAGGGTCCAGTCA
oRP-050	*RPS20* R	TTCGACATCCACACCAGGTT
oRP-051	*SCR1* F	GCAAGGTAGTTCTGGGTCCT
oRP-052	*SCR1* R	GTAAATCCTGATGGCACCGC
oRP-073	*RPS20* F	GGTCAAGAAGGGTCCAGTCA
oRP-074	*RPS20* R	TTCGACATCCACACCAGGTT
oRP-075	*TMA7* F	CCAAGGTGGTAAGATGAAGCC
oRP-076	*TMA7* R	GATACCACCACCCCACTAAAGG
oRP-091	*ARG4* F	GTGAACTAATTGGCCGCGAT
oRP-092	*ARG4* R	CTCACCTTCGGCCCTCTTAA
oRP-093	*ASN1* F	TAATCGCTTCCATTGCTGCC
oRP-094	*ASN1* R	CCGATGGCAAAGGAATGGAG
oRP-105	*RCK1* F	TGTGGAGTATTGGCTGCGTA
oRP-106	*RCK1* R	CATGGCGCCAAGAATTCGTA
oRP-107	*ALD6* F3	CGTCAACATCGTTCCAGGTC
oRP-108	*ALD6* R3	CGACAGCAACACTCTTACCG

Name	Description	Sequence
Target GFP tagging		
oRP-114	*ALD6*-GFP F	GTCTACCATGCATACACTGAAGTAAAAGCTGTCAGAATTAAGTTGCGGATCCCGGGTTAATTAAG
oRP-115	*ALD6*-GFP R	GAAAGTATTTTGTGTATATGACGGAAAAGAAATGCAGGTTGGTACAGAATTCGAGCTCGTTTAAAC
oRP-116	*ALD6*-GFP check F	GAAGAAGGTGTCGAAATGGCT
oRP-117	*ALD6*-GFP check R	TGACGGAAAGAAATGCAGGT
oRP-118	*ASN1*-GFP F	GATCCTTCTGGTAGATATGCCCAAATTCATGAAAAACATATCGAACGGATCCCGGGTTAATTAAG
oRP-119	*ASN1*-GFP R	TCTATAAGATTAATCCATAATTCTTTTCTATTTTTAATGTTATGAATTCGAGCTCGTTTAAAC
oRP-120	*ASN1*-GFP check F	GATTAAAAGATACCGCCGAAGC
oRP-121	*ASN1*-GFP check R	GATTTGCATCAAGATCGTTCGA
oRP-126	*RNR2*-GFP F	TCGACTAAGCAAGAAGCCGGGTGCTTTCACCTTCAACGAAGACTTTCGGATCCCGGGTTAATTAAG
oRP-127	*RNR2*-GFP R	GAGATTGAAGAGACTGCGTAAAAAGAAATATATAGAGAGATACTCGAATTCGAGCTCGTTTAAAC
oRP-128	*RNR2*-GFP check F	TGCTAGGTATGAACGCTGACT
oRP-129	*RNR2*-GFP check R	AAGAGATTGAAGAGACTGCGT
oRP-151	*ACO1*-GFP F	AAATATGGTTCTGCCTTAAATAAAATTAAGGCCGATGAGAGAAAACGGATCCCGGGTTAATTAAG
oRP-152	*ACO1*-GFP R	TGAAAACATTGTTATAATCTTTTAAAGGTTATTATTTGTCGAATTCGAGCTCGTTTAAAC
oRP-153	*ACO1*-GFP check F	CTGGGTCTAGCTGAATTGGC
oRP-154	*ACO1*-GFP check R	TCTTGACTTCTCCAGCCGAA
CLASH		
miRCat-33 linker	3' linker	AppTGGAATTCTCGGGTGCCAAG/ddC/
L5Aa	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrUrNrNrUrArArGrC-OH
L5Ab	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrArUrArGrC-OH
L5Ac	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrGrCrGrCrArGrC-OH
L5Bb	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrGrUrGrArGrC-OH
L5Bc	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrCrArCrUrArGrC-OH
L5Bd	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrUrCrUrCrUrArGrC-OH
L5Ca	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrCrUrArGrC-OH
L5Cb	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrGrArGrC-OH
L5Cc	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrArCrTrCrArGrC-OH
L5Cd	5' linker	invddT-ACACrGrArCrGrCrUrCrGrArUrCrGrArUrCrUrNrNrNrGrArCrTrTrArGrC-OH
miRCat-33 primer	RT primer	CCTTGGCACCCGAGAATT
P5	PCR Primer F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE-miRCat	PCR Primer R	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGGCCTTGGCACCCGAGAATTCC

Table 2.1.1 Table of primers used in this study.

2.2: Strain generation

The *snr4Δ*, *snr45Δ* and *snr72-78Δ* deletion strains were already in use in the Tollervey lab and were created by replacing the snoRNA sequence in BY4741 (*MAT*a; *his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) with the kanamycin resistance gene (*KanMX6*). The 50nt sequence immediately upstream of the snoRNA coding sequence was fused to sequence F1 from Longtine *et. al.* (1998) (*snr4/45-5'/snr72-78*F), and the reverse complement of the 50nt region downstream of the snoRNA coding sequence was fused to R1 (*snr4/45-3'/snr72-78R*). These primers were used to amplify the plasmid pFA6-KanMX6. The resulting strains were *snr4::KanMX6*, *snr45::KanMX6* and *snr72-78::KanMX6* (T. Dudnakova).

Transformation was carried out by growing the strain to be transformed in ~15ml YPD from OD_{600} 0.12 to OD_{600} ~0.5. The culture was then centrifuged for 3min at 1,940xg at 4°C, the media discarded and the pellet resuspended in 5ml H_2O. This was centrifuged again, resuspended in 1ml H_2O, and split into two 1.5ml Eppendorfs. The samples were centrifuged up to 9,250xg and the supernatant discarded. For initial strain generations, one 50µl Phusion-PCR reaction was used per transformation (performed exactly as instructed in the NEB Phusion High-Fidelity Polymerase protocol). In later constructs, three 50µl Phusion-PCR reactions were pooled, and ethanol added to a final concentration of 70% v/v with the addition of 100mM NaOAc pH5.2. Samples were incubated at -20°C for 30min, then centrifuged for 15min at 16,110xg at 4°C and the supernatant discarded. 70% v/v ethanol was added, the samples vortexed and centrifuged for 5min at 16,110xg at 4°C. The supernatant was removed and samples left to air-dry for 30min at room temperature (22°C). Samples were then resuspended in H_2O for transformation. PEG-3350 solution was added to the yeast pellet to a final concentration of 33.3%, followed by 34µl PCR product of either plasmid DNA or isolated genomic DNA, followed by 100µg boiled Salmon sperm DNA and LiAc to a final concentration of 100mM. The mixture was vortexed for resuspension, before incubation for 60-80min at 42°C. The samples were then centrifuged for 30s at 16,750xg and the supernatant discarded. Samples were resuspended in 80µl H_2O and plated onto agar plates before incubation overnight at either 25°C or 30°C. Colonies were then replica-plated onto selective media plates, before incubation at either 25°C or 30°C.

Strains for SGA analysis were created by replacing the *SNR4* or *SNR45* sequence in Y7092 starter strain (*MATα can1Δ::STE2pr-Sp_his5 lyp1Δ ura3Δ0 leu2Δ0 his3Δ1 met15Δ0*) with nourseothricin resistance gene (*NatMX6*) using homologous recombination of flanking regions. Primers *snr4*-5' and *snr4*-3', or *snr45*-5' and *snr45*-3' were used to PCR-amplify plasmid pFA6-NatMX6, which was transformed into the above starter strain as described above. The resulting strains contained *snr4::NatMX6* and *snr45::NatMX6.*

Double mutants were created with one snoRNA gene deleted and the other placed under *GAL1* transcriptional control. To achieve this, primers were designed with a 45nt region positioned 50bp upstream of the snoRNA coding sequence, fused to sequence F4 from Longtine *et. al.* (**1998**) (oRP-063 and oR-065), and the reverse complement of the first 45nt of the snoRNA coding sequence, fused to the reverse complement of 23 nucleotides upstream and including the transcription start site of *GAL1* gene, plus five nucleotides downstream (oRP-064 and oRP-066). These primers were used to PCR-amplify pFA6a-*His3MX6-PGAL1,* which was transformed into the reciprocal snoRNA deletion strain using the above method, to create *HIS3MX6-PGAL1-SNR4   snr45::KanMX6,* and *HISMX6-PGAL1-SNR45 snr4::KanMX6.*

To initially validate mutants identified in the SGA analysis, primers oRP-083 and oRP-084 were used to amplify a region from 62nt upstream of the *GAL1* promoter inserted at *SNR4* to 75nt downstream of the mature snR4 sequence, and oRP-085 and oRP-086 to amplify 50nt upstream of the *GAL1* promoter inserted at *SNR45* to 149nt downstream of the mature snR45 sequence. These were transformed into the TS mutant strains provided by the Andrews lab as described above. The resulting strains were *HIS3MX6-PGAL1-SNR4 ts::KanMX4* and *HIS3MX6-PGAL1-SNR45 ts::KanMX4*. Subsequently, primers oRP-109 and oRP-110 were used to amplify the region from 202nt upstream to 291nt downstream of the *NatMX6* insert in the *snr4Δ* SGA query strain mutants, and oRP-112 and oRP-113 to amplify 194nt upstream to 245nt downstream of the *NatMX6* insert in the *snr45Δ* SGA query strain mutants*.* The resulting strains were *snr4::NatMX6 ts::KanMX4* and *snr45::NatMX6 ts::KanMX4.*

NOP1 mutants were recreated from Tollervey *et. al.* (**1993**). Constructs were made containing a genomic flanking sequence at either end of the construct, to allow for homologous recombination. Wild-type (WT) *NOP1* coding sequence was codon

optimised using Thermo Fisher GeneOptimizer tool. The mutations were designed using the substitutions from the above paper, with a FLAG tag 3' to the coding sequence for detection. The *NOP1* terminator sequence followed the FLAG tag, 5' to the *URA3* promoter and gene. This construct was provided within a pMK-RQ (kanR) plasmid by Thermo Fisher GeneArt Gene Synthesis Service. The transformation method described above was used on this construct with primers oRP-103 and oRP-104, and transformed into either *snr4::KanMX6* or *snr45::KanMX6.*

To validate the qPCR results, the test genes and reference genes were GFP tagged. This used the same transformation method above, using primers oRP-114 and oRP-115 (*ALD6* tagging), oRP-151 and oRP-152 (*ACO1*), oRP-118 and oRP-119 (*ASN1*), and oRP-127 and oRP-128 (*RNR2*). The forward primer took 45nt at the 3' end of the coding sequence – without the stop codon – fused to sequence F2 from Longtine *et. al.* (**1998**), while the reverse primer used the reverse complement sequence of 45nt after the stop codon, fused to sequence R1.

Control	nop1 mutant strains	GFP-tagged mRNA strains
BY4741	NOP1-FLAG-URA3	ALD6-GFP-KanMX6 snr4::NatMX6
	nop1-2-FLAG-URA3	ACO1-GFP-KanMX6 snr4::NatMX6
snoRNA deletion strains	nop1-3-FLAG-URA3	ASN1-GFP-KanMX6 snr45::NatMX6
snr4::KanMX6	nop1-4-FLAG-URA3	RNR2-GFP-KanMX6 snr45::NatMX6
snr45::KanMX6	nop1-5-FLAG-URA3	
snr72-78::KanMX6	nop1-7-FLAG-URA3	**CLASH strains**
snr4::NatMX6	NOP1-FLAG-URA3 snr4::KanMX6	Nop1-HTP
snr45::NatMX6	nop1-2-FLAG-URA3 snr4::KanMX6	Kre33-HTP
HIS3MX6-PGAL1-SNR4 snr45::KanMX6	nop1-3-FLAG-URA3 snr4::KanMX6	NAT10-HTP
HIS3MX6-PGAL1-SNR45 snr4::KanMX6	nop1-4-FLAG-URA3 snr4::KanMX6	
	nop1-5-FLAG-URA3 snr4::KanMX6	
SGA mutant strains	nop1-7-FLAG-URA3 snr4::KanMX6	
yhc1-1::KanMX4	NOP1-FLAG-URA3 snr45::KanMX6	
snr4::NatMX6 yhc1-1::KanMX4	nop1-2-FLAG-URA3 snr45::KanMX6	
snr45::NatMX6 yhc1-1::KanMX4	nop1-3-FLAG-URA3 snr45::KanMX6	
mps1-3796::KanMX4	nop1-4-FLAG-URA3 snr45::KanMX6	
snr4::NatMX6 mps1-3796::KanMX4	nop1-5-FLAG-URA3 snr45::KanMX6	
snr45::NatMX6 mps1-3796::KanMX4	nop1-7-FLAG-URA3 snr45::KanMX6	
mps1-417::KanMX4		
snr45::NatMX6 mps1-417::KanMX4		
orc6-ph::KanMX4		
snr4::NatMX6 orc6-ph::KanMX4		
snr45::NatMX6 orc6-ph::KanMX4		

Table 2.2.1 Table of strains used in this study.

2.3: Cell growth and harvesting

Growth and handling of *S. cerevisiae* were carried out according to standard procedures. For growth curves, three independent transformants each of *snr4Δ*,

snr45∆ and *snr72-78∆* were grown in YNB supplemented with Formedium CSM complete and 2% w/v glucose. BY4741 was grown as a control. Cultures were grown from OD_{600} ~0.1 until they reached OD_{600} ~0.5. Samples were then diluted to either OD_{600} 0.05 or 0.1 in room temperature medium and grown at 25°C or 30°C for 22h. To test for phenotypes under stress conditions, three transformants of *snr4∆*, *snr45∆* and *snr72-78∆* were grown in the above medium with BY4741 as a control. Cultures were grown as above then split into three separate cultures and diluted to OD_{600} 0.05 in room temperature minimal medium. These cultures were incubated at 30°C for 1.5h. One set of cultures was incubated for a further 1h at 30°C, one set incubated for 1h at 18°C and one set incubated for 1h at 4°C. The recovery of all strains was then analysed by growth in triplicate cultures in a 96-well plate, grown at 30°C for 22h in the Tecan Sunrise plate readers, using Magellan software.

Cultures for carbon shift were grown as above and pelleted by centrifugation for 3min at 1,940xg at 4°C. They were then washed and diluted to OD_{600} 0.1 in YNB supplemented with CSM complete and 2% v/v ethanol plus 2% v/v glycerol (carbon shift medium), and analysed by growth in triplicate cultures in a 96-well plate, grown at 30°C for 22-44h. An additional experiment used this same method, but incubated the strains at room temperature with 1X PBS (phosphate-buffered saline) for 20min, before dilution to OD_{600} 0.1 in carbon shift medium. A similar experiment was performed using this method but with incubation at 4°C in 1X PBS instead of room temperature, and resuspension in 4°C carbon shift medium. Nitrogen shift cultures were grown as described for carbon shift, but resuspended in nitrogen shift medium as described in chapter 2.1 at a 1X concentration, and with 2% w/v glucose or 2% v/v ethanol plus 2% v/v glycerol.

The snoRNA double mutant phenotype was analysed by growth of *HISMX6-PGAL1-SNR45 snr4∆, HISMX6-PGAL1-SNR4 snr45∆, snr4∆, snr45∆* and BY4741 strains in YNB supplemented with Formedium CSM complete and 2% w/v galactose (G) plus 2% w/v sucrose (S). Cultures were grown from OD_{600} ~0.1 to OD_{600} ~0.5, at which point all samples were pelleted by centrifugation for 3min at 1,940xg at 4°C. Samples were then washed and diluted to OD_{600} ~0.1 in YNB supplemented with Formedium CSM complete and 2% w/v glucose, and grown to OD_{600} ~0.5. Samples were then diluted in this medium to OD_{600} 0.1 and analysed by growth in triplicate cultures in a 96-well plate, grown at 30°C for 22h.

For SGA analysis, *snr4::NatMX6 ts::KanMX4* and *snr45::NatMX6 ts::KanMX4* were grown with *snr4::NatMX6, snr45::NatMX6* and *ts::KanMX4* from OD_{600} ~0.1 to OD_{600} ~0.5 in YNB supplemented with Formedium CSM complete with 2% w/v glucose. Strains were diluted to OD_{600} 0.1 in the above medium and analysed by growth in triplicate cultures in a 96-well plate, grown at 30°C for <44h.

To validate SGA double mutant phenotypes on solid media, *HIS3MX6-PGAL1-SNR4 ts::KanMX4* and *HIS3MX6-PGAL1-SNR45 ts::KanMX4* strains were initially streaked for single colonies on YPGS/D and Synthetic Complete (SC: YNB supplemented with CSM complete or -His) GS/D plus G418 plates, and incubated at 18°C, 25°C, 30°C and 37°C until colonies appeared. Subsequently, *snr4::NatMX6 ts::KanMX4* and *snr45::NatMX6 ts::KanMX4* were grown with *snr4::NatMX6, snr45::NatMX6* and *ts::KanMX4* to OD_{600} ~0.5 in YNB supplemented with Formedium CSM complete with 2% w/v glucose, before dilution to OD_{600} 0.05, 0.01 and 0.005. These were then plated using a 48-prong replicator onto YPD or SCD solid medium, and incubated at 18°C, 25°C, 30°C and 37°C until colonies appeared.

For DNA or RNA analysis, three transformants each of *snr4Δ*, *snr45Δ* and *snr72-78Δ* were grown in YNB supplemented with Formedium CSM (complete or –Trp) and 2% w/v glucose. BY4741 was grown as a control. Cultures were grown from OD_{600} ~0.1 until they reached OD_{600} ~0.5, at which point all samples were pelleted by centrifugation for 3min at 1,940xg at 4°C. "Standard" samples were resuspended in 1ml 1X PBS and centrifuged, then the supernatant discarded and the pellets frozen. "Ice" samples were resuspended in 10ml 4°C 1X PBS and incubated for 20min on ice before being pelleted and frozen. Pellets were then processed as described in chapter 2.4 or 2.5 as necessary.

2.4: DNA isolation

400µl Fast lysis buffer (chapter 2.1) was added to frozen yeast culture pellets (following growth as described in chapter 2.3) along with 150µl 0.5mm Zirconia/silica beads and 400µl DNA Phenol:Chloroform:Isoamyl Alcohol (IAA) (25:24:1), pH8. Samples were vortexed for 5min at room temperature, then centrifuged for 5min at 16,750xg. The upper phase was transferred into 400µl chloroform:IAA (24:1), samples vortexed briefly and centrifuged for 5min as above. The upper phase was

then transferred into ethanol at a final concentration of 70% v/v with 100mM NaOAc pH5.2, then incubated at -20°C for 30min. Samples were then centrifuged for 15min at 16,110xg at 4°C and the supernatant discarded. 70% v/v ethanol was added, the samples vortexed and centrifuged for 5min at 16,110xg at 4°C. The supernatant was removed and samples left to air-dry for 30min at room temperature. Samples were then resuspended in 100µl H$_2$O.

2.5: RNA isolation

200µl GTC:RNA Phenol pH4 (1:1) (chapter 2.1) and 200µl 0.5mm Zirconia/silica beads were added to frozen yeast culture pellet (following growth as described in chapter 2.3) then vortexed for 5min at room temperature. 1ml GTC:RNA Phenol pH4 (1:1) was added to each sample, then all samples were vortexed for 5min before incubation at 65°C for 5min. Samples were then cooled on ice before adding 100mM NaOAc pH5.2 and 600µl chloroform:IAA (24:1). The mix was vortexed, then centrifuged for 20min at 16,750xg. The upper phase was transferred into 600µl RNA Phenol:Chloroform:IAA (25:24:1), pH4, the sample vortexed and centrifuged for 5min at 16,750xg. The upper phase of this was transferred into 550µl chloroform, the sample vortexed and centrifuged as before. The upper phase was then transferred into ethanol at a final concentration of 70% v/v, vortexed and incubated at -20°C for 30min. All samples were centrifuged for 15min at 16,110xg at 4°C, then the supernatant discarded. The pellet was then resuspended in 70% v/v ethanol and centrifuged for 5min at 16,110xg at 4°C and the supernatant discarded. Pellets were left to air-dry at room temperature for 30min, then dissolved in 100µl H$_2$O or 10mM Tris pH7.8. To proceed for RNA sequencing, samples were DNase-treated with Promega 10X RQ1 buffer, 1 unit (U) RQ1 DNase and 1U of Promega RNasin and incubated at room temperature for 30min. The reaction was stopped by adding 50mM EDTA and incubating on ice, followed by addition of 10mM Tris pH7.8 and 100mM NaOAc. Samples were transferred into 600µl RNA Phenol:Chloroform:IAA (25:24:1), pH4, and the remainder of the RNA isolation protocol continued from this step.

2.6: RNA sequencing library preparation

RiboMinus Preparation

125µl RiboMinus Magnetic Bead Suspension were washed with sterile water and resuspended in 100µl Hybridisation buffer, as per the Invitrogen RiboMinus Transcriptome Isolation Kit protocol. 10µg total RNA from each of "Standard" and "Ice" samples, as measured by Thermo Scientific NanoDrop Spectrophotometer, were incubated with 400pmol Yeast RiboMinus probe and 100µl Hybridisation buffer at 37°C for 5min, then placed on ice. This mixture was then added to the RiboMinus Magnetic Bead Suspension and incubated at 37°C for 15min, mixing occasionally. The mixture was placed in a magnetic stand and left for 1min, then the supernatant transferred to a fresh tube. Ethanol was added to a final concentration of 70% v/v, with the addition of 100mM NaOAc pH5.2, and incubated at -80°C for 30min. The sample was centrifuged for 10min at 16,110xg at 4°C, the supernatant aspirated and the pellet resuspended in 70% v/v ethanol. The sample was centrifuged again and the supernatant aspirated, then the pellet dried and resuspended in 10.5µl H_2O.

Poly(A)+ selection preparation

To select for poly(A)-tailed RNAs, the NEBNext poly(A) mRNA Magnetic Isolation Module kit was used, and the protocol followed. 20µl Oligo d(T)$_{25}$ beads per sample were washed twice with 100µl 2x RNA Binding Buffer. 2µg DNased total RNA from "Standard" and "Ice" samples (chapter 2.5), as measured by Agilent Bioanalyzer RNA chip, was diluted to a total of 50µl in nuclease-free water. The beads were resuspended in 2x RNA Binding Buffer and added to the RNA samples. The samples were heated at 65°C for 5min then cooled to 4°C. This was then resuspended, incubated at room temperature for 5min, resuspended a second time and incubated a second time. The samples were placed on a magnetic rack and the supernatant discarded. Each sample was washed twice with Wash Buffer, then resuspended in 50µl Tris Buffer and mixed. The samples were heated at 80°C for 2min then cooled to 25°C and diluted with 2x RNA Binding Buffer. Samples were subsequently incubated at room temperature for 5min, resuspended, and incubated again. These were then placed on the magnetic rack, and the supernatant discarded. Samples were washed with Wash Buffer, and all supernatant thoroughly

removed and discarded. mRNA was eluted from the beads by adding 17µl 10mM Tris pH7.8 and incubating at 80°C for 2min then held at 25°C. Samples were placed on the magnetic rack, and the supernatant transferred into a fresh tube. RNA concentration was measured by Agilent Bioanalyzer RNA chip and Thermo Fisher Qubit RNA HS (high sensitivity) Assay kit.

RNA library preparation for Illumina sequencing

RiboMinus-depleted RNA (10µg starting material) or approximately 50ng poly(A)+ selected mRNA was incubated with 5X NEBNext First Strand Synthesis Reaction Buffer and 1µl NEBNext Random Primers in 10µl total volume. The samples were incubated for 15min at 94°C, then cooled on ice. Added to this was 0.5µl Murine RNase Inhibitor, 0.1µg Actinomycin D, 1µl Protoscript II Reverse Transcriptase and 8.5µl nuclease-free H_2O. The samples were then incubated at 25°C for 10min, 42°C for 15min, then 70°C for 15min. 10x Second Strand Synthesis buffer, 4µl Second Strand Synthesis Enzyme mix and H_2O were added to the samples to a final volume of 80µl, and the tubes were incubated for 1h at 16°C. Samples were then purified using a QIAquick PCR purification kit as follows: 5 volumes of PB buffer were added to 1 volume of PCR reaction and mixed. The mixture was applied to a QIAquick column and centrifuged for 1min at 16,250xg. The flow-through was discarded. 750µl PE buffer was then added to the column and the column centrifuged as above, then the flow-through discarded. The column was then transferred to a fresh 1.5ml Eppendorf and left with lid open for 5min. This was centrifuged for 2min at 16,250xg. The column was then transferred to a fresh 1.5ml Eppendorf, 58µl 10mM Tris pH7.8 pipetted into the centre of the column and the column left to stand for 1min. Samples were eluted by centrifugation for 1min at 16,250xg and stored at -20°C overnight.

10x NEBNext End Repair Reaction Buffer and 3µl NEBNext End Prep Enzyme Mix were added to the thawed purified double-stranded cDNA. The samples were incubated at 25°C then 65°C for 30min each, before cooling to 4°C. 15µl Blunt/TA Ligase Master Mix and 1.5µM NEBNext Multiplex Adapter were added directly to the End Prep reaction mix along with nuclease-free water to make a total volume of 83.5µl. Samples were mixed and incubated for 15min at 20°C. The reactions were

then purified using the QIAquick PCR purification kit as above. Samples were eluted in 20µl 10mM Tris pH7.8.

3µl NEBNext USER enzyme, NEBNext Q5 2x Hot Start HiFi PCR Master Mix, 2.5µl Universal PCR Primer and one 2.5µl Index Primer per PCR reaction (1-9 for samples 1-9) were added to the 20µl cDNA and mixed. Samples were subjected to PCR by the following method: 37°C for 15min, 98°C for 30s, *98°C for 10s followed by 65°C for 75s* for 12 cycles, 65°C for 5min then held at 4°C. The PCR reactions were then purified using the QIAquick PCR purification kit as above. Samples were eluted in 18µl 10mM Tris pH7.8. Each sample was purified on a 3% w/v MetaPhor agarose 1X TBE gel (Fisher Scientific 10X Tris/Borate/EDTA solution) with a Fisher Scientific exACTGene 50bp Mini ladder and 1:10,000X Invitrogen SYBR safe DNA gel stain, until bromophenol blue had migrated the length of the gel. The band ranging between 150-200bp was extracted. This was purified using the QIAquick Gel Extraction Kit as follows: 6 volumes of QG buffer were added to 1 volume of gel. This mix was incubated at 50°C for 10min, vortexing occasionally. 1 volume of isopropanol was added and mixed by inversion. This mix was applied to a MinElute column and centrifuged for 1min at 16,250xg. The flow-through was discarded, and the remainder of the mix applied and centrifuged as above. The column was then washed with 500µl QG buffer and the column centrifuged as above, with flow-through discarded. 750µl PE buffer was then used to wash the column, followed by centrifugation of the column as above and transfer to a fresh 1.5ml Eppendorf. The column was left to dry for 5min, then centrifuged for 3min at 16,250xg. In a fresh 1.5ml Eppendorf, 16µl 10mM Tris pH7.8 was added to the centre of the column, left to incubate for 2min, then the cDNA eluted by a 1min centrifugation. The quality of library was assessed by Agilent Bioanalyzer DNA chip and Thermo Fisher Qubit DNA HS Assay kit.

30ng of each sample were mixed together according to the conditions they were grown in, in a total volume of 10.5µl for "standard" conditions and 16.5µl for "ice" conditions, and sent to Edinburgh Genomics for RNA Illumina sequencing.

2.7: Quantitative PCR

RNA isolated as in chapter 2.5 was measured by Thermo Scientific NanoDrop Spectrophotometer and DNase-treated as per the Ambion Turbo DNase protocol as follows: 10X TURBO DNase buffer and 2U TURBO DNase were added to 10µg RNA diluted in 45µl H_2O. This was incubated at 37°C for 30min, before adding 10X DNase Inactivation Reagent. This mix was incubated at room temperature for 5min, mixing occasionally. The mix was then centrifuged for 1.5min at 9,250xg, before transferring the supernatant to a fresh tube. 0.5µg of this was then used in RETROscript's Reverse Transcriptase (RT) kit. Random decamer primers or oligo(dT) primers were added to a final concentration of 5µM (after all reagents added) and nuclease-free H_2O added to a final volume of 12µl. The sample was mixed and heated at 80°C for 3min, then incubated briefly on ice. 10X RT buffer was added, together with 4µl dNTP mix (2.5mM per dNTP), 10U RNase Inhibitor, and 100U Reverse Transcriptase. The sample was mixed and incubated at 42°C for 1h, followed by 10min at 92°C. The sample was then diluted to either 1ng/µl or 0.1ng/µl. 4µl of this was mixed with 6µl of: Takara Bio SYBR *Premier Ex Taq* 2x mix, 50X ROX reference dye and 10µM of forward and reverse primers (final concentrations 1X, 1X and 0.2µM, respectively). This was then amplified by qPCR by the Agilent Stratagene Mx3005P, using the following method for SYBR Green with dissociation curve: 95°C for 1min, *95°C for 15s, 55°C for 15s, and 72°C for 15s* for 40 cycles, followed by 95°C for 1min, 55°C for 30s and 95°C for 30s. Primer sets were tested with a standard curve of genomic DNA before use with samples, and each primer set tested with a no-template mix in the qPCR. Each sample was run with a no-RT control of the same concentration to ensure DNA contamination was minimal. For a subset of qPCRs, RNA isolated from *Schizosaccharomyces pombe* (as described in chapter 2.5) was added at a concentration 10-fold lower than the isolated *S. cerevisiae* RNA and samples were DNase-treated. RNA from this was then used in the RT reaction, using random decamer primers or oligo(dT) primers. qPCR was carried out as described above.

Analysis of qPCR data was performed either by the $2^{-\Delta\Delta Ct}$ method (**Livak and Schmittgen, 2001**) or by a variation on this. The $2^{-\Delta\Delta Ct}$ for each experiment was determined as follows: the difference between the C_t (threshold cycle) of the test gene and the housekeeping gene (eg. *SCR1)* in the deletion strain was calculated, as was the difference between the C_t of the test gene and the housekeeping gene in

the WT strain. The latter value was subtracted from the former to give the $\Delta\Delta C_t$ value. The exponential of $-\Delta\Delta C_t$ gave the relative expression of the test gene in the deletion strain. The average of $2^{-\Delta\Delta Ct}$ replicates for each gene was taken to determine fold change (FC). A one-sample t-test was performed to test for statistical significance. qPCRs analysed by the variation on this were analysed as follows: each technical replicate for both the test gene and *TAF10* was normalised to the corresponding replicate of *S. pombe ACT1,* to account for experimental variance between samples. The test gene was then normalised to *TAF10,* and the average of each biological replicate taken. These were then normalised to the average of each WT biological replicate, and a two-tailed homoscedastic t-test applied.

2.8: Northern blotting

5µg (or 10µg) of RNA samples (as isolated in chapter 2.5) were diluted to the same volume with H_2O and mixed with the recommended volume of glyoxal as per the Ambion NorthernMax-Gly Kit protocol. Samples were then electrophoresed on a 1.2% w/v agarose 1x BPTE (10mM PIPES; 30mM Bis-Tris; 1mM EDTA) gel at 50V overnight at 4°C in 1x BPTE buffer, then the voltage increased and the gel run until bromophenol blue had migrated at least 75% of the length of the gel. The gel was then scanned to reveal ethidium bromide staining of the RNA. The gel was washed briefly with H_2O, then treated for 20min with 75µM NaOH, followed by a wash for 20min in 0.5M Tris pH7.5 plus 1.5M NaCl. It was then washed with 6X SSC (chapter 2.1) for 20min. Each wash was followed by a short rinse with H_2O. The gel was transferred onto a GE Healthcare Hybond-N+ membrane overnight at room temperature by capillary transfer. RNA was immobilised to the membrane by UV cross-linking at 120mJ/cm^2. The membrane was stored at 4°C until use.

For hybridisation, 15ml of 20X SSC (chapter 2.1) was mixed with 2.5ml 100X Denhardt hybridisation buffer (2% w/v Ficoll 400, 300mM NaCl, 2% w/v Polyvinylpyrrolidone, 2% w/v BSA) and 1.25ml 20% w/v SDS in a total volume of 50ml. This was incubated at 50°C and filter-sterilised. This solution was added to the membrane in a plastic box, and incubated for 1h at 37°C with shaking. To make the probe, 10U of T4 polynucleotide kinase (PNK) was mixed with 10µM oligo, 10X PNK buffer, 9µl H_2O and 2.5µl γ-ATP (^{32}P). This was incubated at 37°C for 40min, then purified on a Roche Mini Quick Spin Oligo column by centrifugation for 1min at

1,000xg. The labelled probe was added to fresh hybridisation buffer and incubated with the membrane overnight at 37°C. The membrane was subsequently washed with 6X SSC plus 0.1% w/v SDS at 37°C, a total of three times for 10min each. The membrane was then exposed to a phosphor screen for 1-3 days, depending on strength of signal. The signal was detected using a fluorescent imaging analyser (FLA-5000 scanner, Fujifilm). To dehybridise the membrane, 0.1X SSC plus 0.1% w/v SDS was boiled and poured on the membrane at room temperature for 20min, with a total of three washes. The dried membrane was then stored at 4°C until next use.

2.9: Sanger sequencing

DNA was isolated from cells as described in chapter 2.4. Dried pellets were dissolved in 40µl TE buffer (10mM Tris, 1mM EDTA, pH8) plus 10µg RNase A (0.25µg/µl final concentration). Samples were briefly vortexed and incubated at 42°C for 15min. DNA concentration was measured using Thermo Fisher Qubit DNA BR (broad range) Assay kit. 50ng DNA were PCR-amplified using NEB Phusion High-Fidelity Polymerase, performed exactly as instructed in the protocol. PCR products were electrophoresed on a 1% w/v agarose 1X TBE gel in 1X TBE buffer with Invitrogen TrackIt 1kb Plus DNA ladder, until bromophenol blue had migrated the length of the gel. If only one PCR product was identified per lane on the gel, the PCR reaction was purified using the QIAquick PCR purification kit as described earlier. If multiple products were observed, the PCR reaction was purified by gel electrophoresis and the correct product extracted. This was then purified using the QIAquick Gel Extraction Kit as described above. Samples were eluted in 17.5µl EB (elution buffer, 10mM Tris pH8.5) and the DNA concentration measured using Thermo Fisher Qubit DNA BR Assay kit. 20ng of each sample were diluted with 6.4pmol of primer and sent to Edinburgh Genomics for Big Dye Reaction for Sanger sequencing.

2.10: Cross-linking, Ligation and Sequencing of Hybrids (CLASH)

HTP-tagged yeast strains in use in the lab were grown in a 50ml preculture of YNB supplemented with Formedium CSM -Trp and 2% w/v glucose at 30°C overnight. Cells were diluted to OD_{600} ~0.1 in fresh 30°C medium and grown to OD_{600} ~0.5 at 30°C, then washed in warm 1X PBS and UV cross-linked at 254nm, for 100s at 400mJ/cm^2. Cultures were then pelleted for 15min at 2,700xg at 4°C, resuspended in 30ml 4°C 1X PBS and transferred to fresh 50ml Falcon tubes. Cultures were pelleted for 5min at 1,940xg at 4°C, the supernatant discarded and the pellet stored at -80°C. Pellets were lysed in 1V TMN buffer (50mM Tris, 150mM NaCl, 2mM $MgCl_2$, 0.4% v/v NP-40 and 1 mM DTT, pH7.8) plus Roche complete EDTA-free protease inhibitor, 4µl Promega RNasin and 1V 0.5mm Zirconia/silica beads by vortexing for 1min followed by incubation on ice for 1min, five times. Lysates were incubated on ice for 5min, then centrifuged for 10min at 16,110xg at 4°C. The supernatant was transferred to fresh 2ml Eppendorfs and DNase-treated with 15µl Promega RQ1 DNase, by incubating the sample at room temperature for 10min with gentle inversion to mix.

80µl GE Healthcare IgG Sepharose 6 Fast Flow bead suspension was added to each sample (washed in 9ml 1X PBS plus 1ml TMN) and nutated at 4°C for 1h. Samples were centrifuged to 92xg at 4°C, the supernatant was removed, 1ml HS-WB (50mM Tris, 600mM NaCl, 2mM $MgCl_2$ and 0.4% v/v NP-40, pH7.8) was added and samples were nutated for 5min at 4°C. This step was repeated. The same step was repeated twice with 1ml TMN buffer, and twice with 1ml PNK buffer (50mM Tris, 50mM NaCl, 10mM $MgCl_2$, and 0.5% v/v NP-40, pH7.8). Samples were centrifuged as above, the supernatant removed and the beads resuspended in 100µl of PNK buffer plus 0.1U of Stratagene RNace-IT. This was incubated at room temperature for 10min with manual mixing. 1.2ml of Ni-WB (50mM Tris, 300mM NaCl, 6M $GuCl_2$, 10mM Imidazole and 0.4% NP-40, pH7.8) was added to the reaction and briefly vortexed twice, with 1min incubation at room temperature in between. Samples were centrifuged to 92xg at 4°C and the supernatant transferred to 80µl Qiagen Ni-NTA bead suspension equilibrated in Ni-WB. Samples were nutated for 1h at 4°C.

Samples were transferred to Fisher Scientific Pierce Snap Cap Spin columns and washed twice with 1ml Ni-WB, followed by two washes with 1ml TMN buffer, and two washes with 400µl PNK buffer, with centrifugation to 92xg followed by

discarding of flow-through between each wash. To dephosphorylate RNA fragments, 80µl of the following mix was added: 4U Promega TSAP (<u>t</u>hermo<u>s</u>ensitive <u>a</u>lkaline <u>p</u>hosphatase), 10X MultiCore buffer and 2µl Promega RNasin. Samples were incubated at room temperature for 45min. These were then washed with 400µl Ni-WB, followed by two washes with 600µl TMN buffer and two washes with 600µl PNK buffer, with centrifugation to 92xg followed by discarding of flow-through between each wash.

Phosphorylation of 5' ends was performed by resuspending beads in a 60µl mix of 10X NEB PNK buffer, 30U T4 PNK, 1µl Promega RNasin and 2µl γ-ATP (^{32}P), and incubating for 45min at 20°C. 100mM NEB ATP was then added to the sample, and incubated for 20min at 20°C. These were then washed with 400µl Ni-WB, followed by three washes with 1ml Ni-WB, three washes with 1ml PBS-Tween plus 0.01% v/v Triton, two washes with 600µl TMN buffer, and two washes with 600µl PNK buffer, with centrifugation to 92xg followed by discarding of flow-through between each wash. Beads were resuspended in an 80µl mix of 10X NEB PNK buffer, 40U T4 RNA ligase 1, 10µM IDT miRCat-33 3' linker and 2µl Promega RNasin, and incubated overnight at 16°C.

Samples were washed twice with 600µl Ni-WB, twice with 600µl PBS-Tween plus 0.01% v/v Triton, twice with 600µl TMN buffer and twice with 600µl PNK buffer, with centrifugation up to 92xg, followed by discarding of the flow-through between each wash. Beads were resuspended in a 75µl mix of 10X PNK buffer, 40U T4 RNA ligase 1, 100mM ATP and 2µl Promega RNasin, plus 100µM IDT 5' linker, and ligated at room temperature for 3h. The beads were then washed twice with 600µl Ni-WB, twice with 600µl PBS-Tween plus 0.01% v/v Triton, twice with 600µl TMN buffer and twice with 600µl PNK buffer, with centrifugation to 92xg followed by discarding of flow-through between each wash. Samples were eluted from the beads using 70µl Elution Buffer (2x NuPage Sample buffer, 400mM Imidazole and 100mM DTT) by centrifugation up to 92xg, then the eluent passed through the column again. This eluent was electrophoresed on a 1.5mm thick Life Technologies NuPAGE 4-12% gradient Bis-Tris protein gel in NuPAGE MOPS SDS running buffer at 120V for 1.5h. Samples were then transferred onto GE Healthcare Hybond-C extra nitrocellulose membrane via wet transfer using NuPAGE transfer buffer plus 10% v/v MeOH for 2h at 100V on ice. This membrane was exposed to Kodak BioMax MS film overnight at -80°C.

Products of the correct size were extracted from the membrane and incubated for 2h at 55°C in 400µl of the following mix: 50mM Tris pH7.8, 50mM NaCl, 1% w/v SDS, 5mM EDTA, 5mM DTT and 100µg Proteinase K per sample. 100mM NaOAc pH5.2 was added to the sample, then 500µl RNA Phenol:Chloroform:IAA mix (25:24:1), pH4. Samples were briefly vortexed and centrifuged for 15min at 16,750xg at room temperature. The aqueous phase was extracted and mixed with 500µl isopropanol plus 1µl Ambion glycogen, then incubated at -80°C for 30min. Samples were centrifuged for 10min at 16,110xg at 4°C, washed with 70% v/v ethanol and left to air-dry at room temperature for 30min. The pellet was resuspended in 10µM IDT miRCat-33 RT oligo plus 10mM dNTP mix and H_2O. The mixture was incubated at 80°C for 3min then on ice for 5min. 5X first strand buffer of Invitrogen Superscript III Reverse Transcriptase kit was added, plus 100mM DTT and 40U RNasin, and the reaction incubated at 50°C for 3min. 200U Superscript III was added to the reaction, and incubated for 1h at 50°C, followed by incubation at 65°C for 15min. 10U NEB RNase H was then added to the mixture and incubated for 30min at 37°C.

2µl of this reaction was PCR-amplified, with the addition of 10X Takara LA *Taq* buffer, 10µM P5 and PE-miRCat primers, 10mM dNTP mix and 2.5U LA *Taq* polymerase. PCR amplification was performed by the following method: 95°C for 2min, *98°C for 20s, 52°C for 20s, and 68°C for 20s* for 25 cycles, followed by 72°C for 5min. The reaction was purified using the QIAquick PCR purification kit as described above. Samples were electrophoresed on a 2.5% w/v MetaPhor agarose gel in 1X TBE buffer with a Fisher Scientific exACTGene 50bp Mini ladder and 1:10,000X Invitrogen SYBR safe DNA gel stain at 80V, 4°C until bromophenol blue had migrated the length of the gel. The band ranging between 150-200bp was extracted. This was purified using the QIAquick Gel Extraction Kit as described above. Concentration of the cDNA libraries was measured by Thermo Fisher Qubit DNA HS Assay kit, and the libraries sent to Edinburgh Genomics for high-throughput sequencing. Bioinformatics was performed by H. Dunn-Davies, as described in Travis *et. al.* (**2014**).

3 PHENOTYPIC ANALYSIS OF snoRNAs

3.1: Introduction

snR4 and snR45 are non-essential snoRNAs that lacked any known targets or function (**Parker *et. al.,* 1988, Kudla *et. al.,* 2011**). This chapter investigates the phenotype associated with deletion of these two snoRNAs, and identifies pathways they may be involved in.

Synthetic Genetic Array (SGA) is a method that identifies genes with functionally interacting products by screening for genetic interactions (**Baryshnikova *et al.,* 2010**). A query mutation is mated in pairwise crosses with every strain from collections that each carry a single deletion of one of the ~5000 non-essential yeast genes, or mutation of one of the ~1000 essential genes. The fitness of the double mutant is compared to fitness of each single mutant, which determines whether the interaction between the two genes is negative (and shows synthetic lethality) or positive (and shows suppression). Figure 3.1.1 shows the methodology used. The query mutation and array mutation are mated and allowed to sporulate. Meiotic progeny are selected, followed by additional rounds of selection for the double mutants. Use of this array provides a high-throughput analysis, identifying potential interaction partners or pathways for genes of unknown function.

Successful deletions of *SNR4, SNR45* and the *SNR72-78* cluster had previously been constructed in the lab (unpublished work, T. Dudnakova). Therefore, experiments were designed to probe the phenotypes and interactions associated with these strains, as discussed in this chapter.
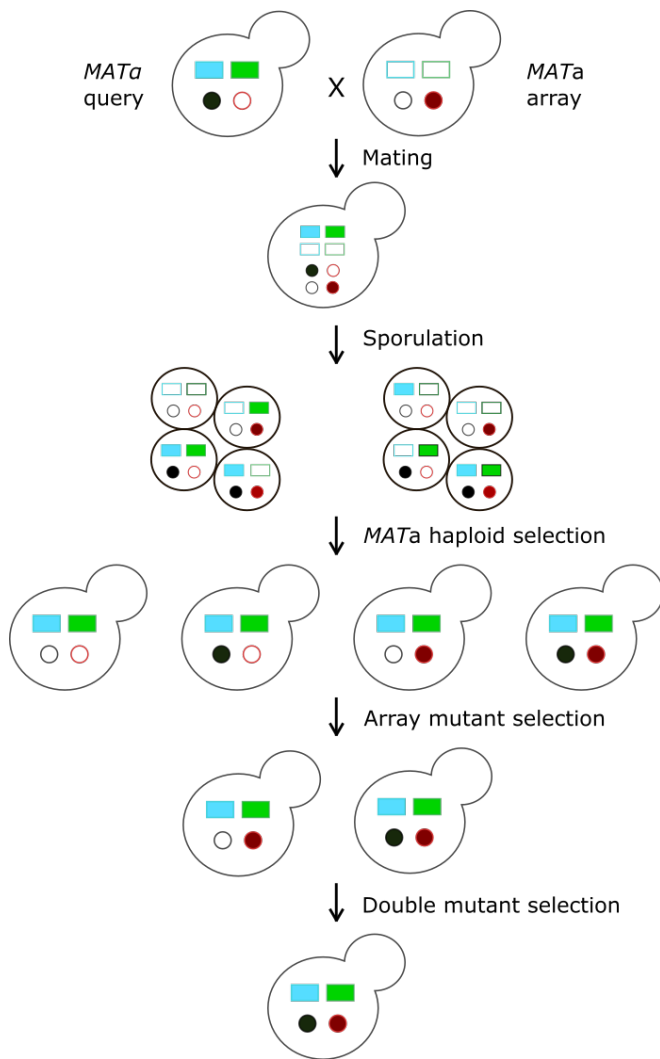
Figure 3.1.1 Schematic of SGA screen. Strains of the opposite mating type, carrying the query mutation and the array mutation are mated and allowed to sporulate. Query mutations contain the nourseothricin resistance gene (*NatMX6,* filled black circle), as well as the recessive *can1*Δ allele that confers resistance to the toxic arginine analogue, canavinine (filled blue rectangle), and the recessive *lyp1*Δ allele that confers resistance to the toxic lysine analogue, thialysine (filled green rectangle). Array mutants contain the kanamycin resistance gene (*KanMX6,* filled red circle). Following sporulation, diploids are counter-selected using canavinine and thialysine, which selects for meiotic progeny. Additional rounds select for the array mutant, through growth on G418 (kanamycin), followed by selection for double mutants through growth on both G418 and nourseothricin. Unfilled circles and rectangles represent wild-type genes. Adapted from Baryshnikova *et. al.* (2010).

3.2: Growth phenotyping of deletion strains

To investigate whether deletion of *SNR4, SNR45* or the *SNR72-78* cluster had any effect on cell growth, *snr4Δ, snr45Δ, snr72-78Δ* and BY4741 (wild-type, WT) were grown in minimal medium containing 2% w/v glucose at 30°C as described in chapter 2.3. Cultures in mid-log phase were diluted to 0.1 OD_{600} and grown in triplicate in a Tecan Sunrise plate reader, with data analysis using Magellan software. The OD_{600} of cultures within 96-well plates was measured every 15min for 22h, at either 30°C or 25°C. However, no clear differences in growth rate were observed between the mutant strains compared to the WT at either temperature. Due to machine limitations, the temperature could not be regulated accurately at lower temperatures. As no difference in the strains was observed at these two temperatures, 30°C was chosen for future experiments.

Although snR4, snR45 and snR72-78 appeared dispensable during normal growth, it remained possible that they would be more important during stress conditions. snR4 had previously been identified in hybrid sequences following cold shock (**Kudla *et. al.*, 2011,** and unpublished work, T. Dudnakova). It therefore seemed feasible that growth defects in strains lacking the snoRNAs would be more evident following cold shock treatment. The *snr4Δ, snr45Δ, snr72-78Δ* and WT strains were grown as above to mid-log phase, then split into three separate cultures and diluted to 0.05 OD_{600} in room temperature (22°C) minimal medium. These three sets of cultures were incubated for 1.5h at 30°C. One set of strains was incubated for a further 1h at 30°C, the second set of strains for 1h at 18°C, and the third set of strains for 1h at 4°C. The recovery of all sets of strains was then analysed by growth in triplicate cultures in a 96-well plate, grown at 30°C for 22h.

Figure 3.2.1 shows the results of these growth curves. Time in hours is shown along the x axis, while OD_{600} or $log_{10}(OD_{600})$ is shown on the y axis. At 30°C, none of the strains showed an obvious growth defect (Figure 3.2.1A). Strain *snr72-78Δ* peaked at a lower OD compared to the other strains. However, analysis of $log_{10}(OD_{600})$ showed no visible difference in growth rate (Figure 3.2.1B). This was also observed for 18°C treatment, where *snr72-78Δ* peaked at a lower OD, but $log_{10}(OD_{600})$ showed no visible difference in growth rate (Figures 3.2.1C and D). Following stronger cold shock at 4°C, the growth curve for *snr72-78Δ* was more distinct from the other strains. However, comparison with $log_{10}(OD_{600})$ indicated that the

difference was due to the strain having a lower starting OD than the others. These results show that deletion of these snoRNAs had no effect on the growth of these strains under different temperature conditions. Longer temperature 'shocks' may result in a visible growth defect; however these results did not give any indication of a stress defect.
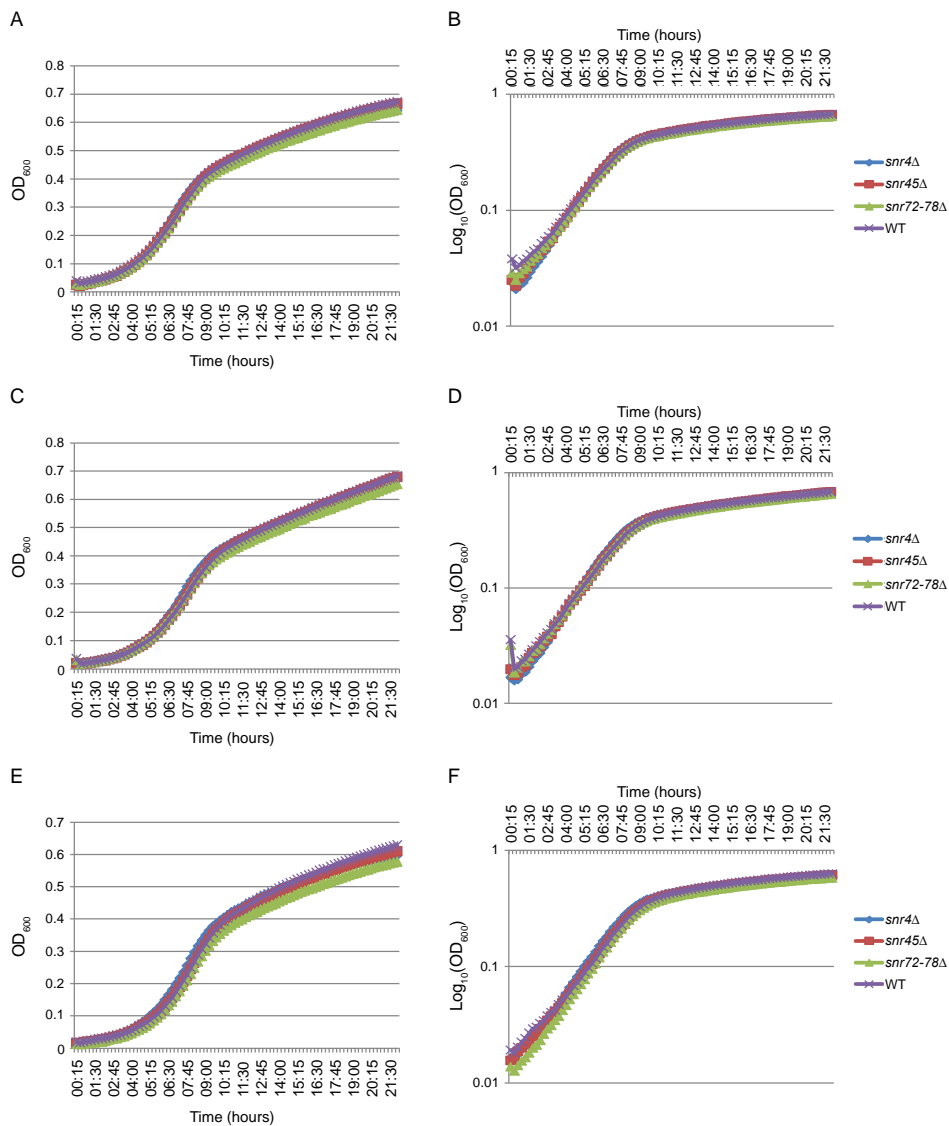
Figure 3.2.1 Growth curves of snoRNA deletion strains following different temperature shocks. The growth curve (OD_{600}) and growth rate ($\log_{10}(OD_{600})$) of *snr4Δ, snr45Δ, snr72-78Δ* and WT following incubation at 30°C (A and B), 18°C (C and D) and 4°C (E and F) for 1h. Time in hours is shown across the x axis, with OD_{600} or $\log_{10}(OD_{600})$ on the y axis. All cultures were grown at 30°C for 22h. Three biological replicates were grown, with three technical replicates analysed for each. A key indicating strains is shown to the right of each panel.

Another commonly used stress condition for budding yeast is transfer from glucose-containing medium to ethanol plus glycerol, to mimic the naturally occurring "diauxuic shift" - a major metabolic change from fermentative growth on glucose to aerobic respiration of ethanol. Strains were grown in a preculture medium containing 2% w/v glucose to mid-log phase, then shifted to medium containing 2% v/v ethanol plus 2% v/v glycerol as carbon source, and grown for 44h at 30°C. Figures 3.2.2A and B show the same strains diluted into the 2% w/v glucose medium for a control comparison. As shown in Figure 3.2.1A, the deletion strains showed no clear growth differences. Figures 3.2.2C and D show strains following growth in ethanol-glycerol medium. *snr72-78Δ* peaked at a lower OD in comparison to the other strains, whereas *snr4Δ* continued to grow to a higher OD. However, $\log_{10}(OD_{600})$ shows that the growth rate of all four strains was the same. Therefore, strains were 'shocked' before this carbon shift by incubation at room temperature with PBS (phosphate-buffered saline) for 20min. As PBS contains no carbon source this would highlight any stress-responding pathways in the cell. However, as shown in Figures 3.2.2E and F, growth rates between strains were the same, with only a slight defect observed in *snr45Δ.* None of the strains reached an OD higher than 0.3, showing that the stress did affect the cell, but no obviously increased defect was observed in the deletion strains. It was possible that in order to observe a phenotype, the cell had to be stressed even further. Therefore a cold shock was combined with the PBS incubation. Strains were incubated in cold PBS for 20min on ice before dilution in the carbon shift medium and grown for 44h at 30°C. Figure 3.2.2G shows that *snr4Δ* had a slower growth rate than the other strains, all of which showed little difference to WT. However, Figure 3.2.2H shows that this may again be due to an effect of dilution, rather than a growth defect.

Figure 3.2.2 Growth curves of snoRNA deletion strains in medium containing poor carbon sources. A) Growth curve (OD$_{600}$) of *snr4Δ, snr45Δ, snr72-78Δ* and WT following growth in minimal medium containing 2% w/v glucose. B) As in

(Figure 3.2.2 cont.) 'A' but showing growth rate ($log_{10}(OD_{600})$). C) Growth curve (OD_{600}) following shift from minimal medium containing 2% w/v glucose to growth in minimal medium containing 2% v/v ethanol plus 2% v/v glycerol. D) As in 'C' but showing growth rate ($log_{10}(OD_{600})$). E) Growth curve (OD_{600}) following incubation in room temperature PBS for 20min, then a shift into minimal medium containing 2% v/v ethanol plus 2% v/v glycerol. F) As in 'E' but showing growth rate ($log_{10}(OD_{600})$). G) Growth curve (OD_{600}) following incubation in cold PBS on ice for 20min, then a shift into minimal medium containing 2% v/v ethanol plus 2% v/v glycerol. H) As in 'G' but showing growth rate ($log_{10}(OD_{600})$). Time in hours is shown across the x axis, with OD_{600} or $log_{10}(OD_{600})$ on the y axis. Strains in panels C and D were grown at 30°C for 36h. All remaining cultures were grown at 30°C for 22h. Three biological replicates were grown, with three technical replicates analysed for each. A key indicating strains is shown to the right of each panel.

No obvious phenotype had been observed with altered carbon sources, but it remained possible that the snoRNAs function in an alternative pathway that would require a different stress to highlight it. Another stress condition for budding yeast is growth on a poor nitrogen source. As one of the key nutrients consumed by yeast, limitation of nitrogen availability results in reduced cell growth and reduced ribosome synthesis (reviewed in **Broach, 2012**). Therefore, strains were grown to mid-log phase in minimal medium containing 2% w/v glucose with ammonium as the sole nitrogen source, then washed and resuspended in minimal medium containing 2% w/v glucose with proline as the sole nitrogen source, as described in chapter 2.3. Proline is a poor nitrogen source, so a shift to this as the sole nitrogen source would stress the cells. Strains were grown for 44h at 30°C. Figure 3.2.3A shows the growth curve following shift to poor nitrogen medium. A lower peak of OD was once again observed in *snr72-78Δ*, though this was not observed when analysing $log_{10}(OD_{600})$, shown in Figure 3.2.3B. Neither *snr4Δ* nor *snr45Δ* showed any phenotype. Therefore, a carbon shift and nitrogen shift were combined by resuspending strains to 0.05 OD_{600} in medium with proline as the sole nitrogen source and 2% v/v ethanol plus 2% v/v glycerol as the carbon source. All snoRNA deletion strains appeared to grow faster than WT. However, strains grew to <0.12 OD_{600} after 44h of growth, and at these concentrations it is unreliable to determine whether this is a valid growth defect. Furthermore, analysis of $log_{10}(OD_{600})$ showed minimal differences between

growth rates of the different strains, suggesting that the differences observed in Figure 3.2.3C are unreliable (Figure 3.2.3D).
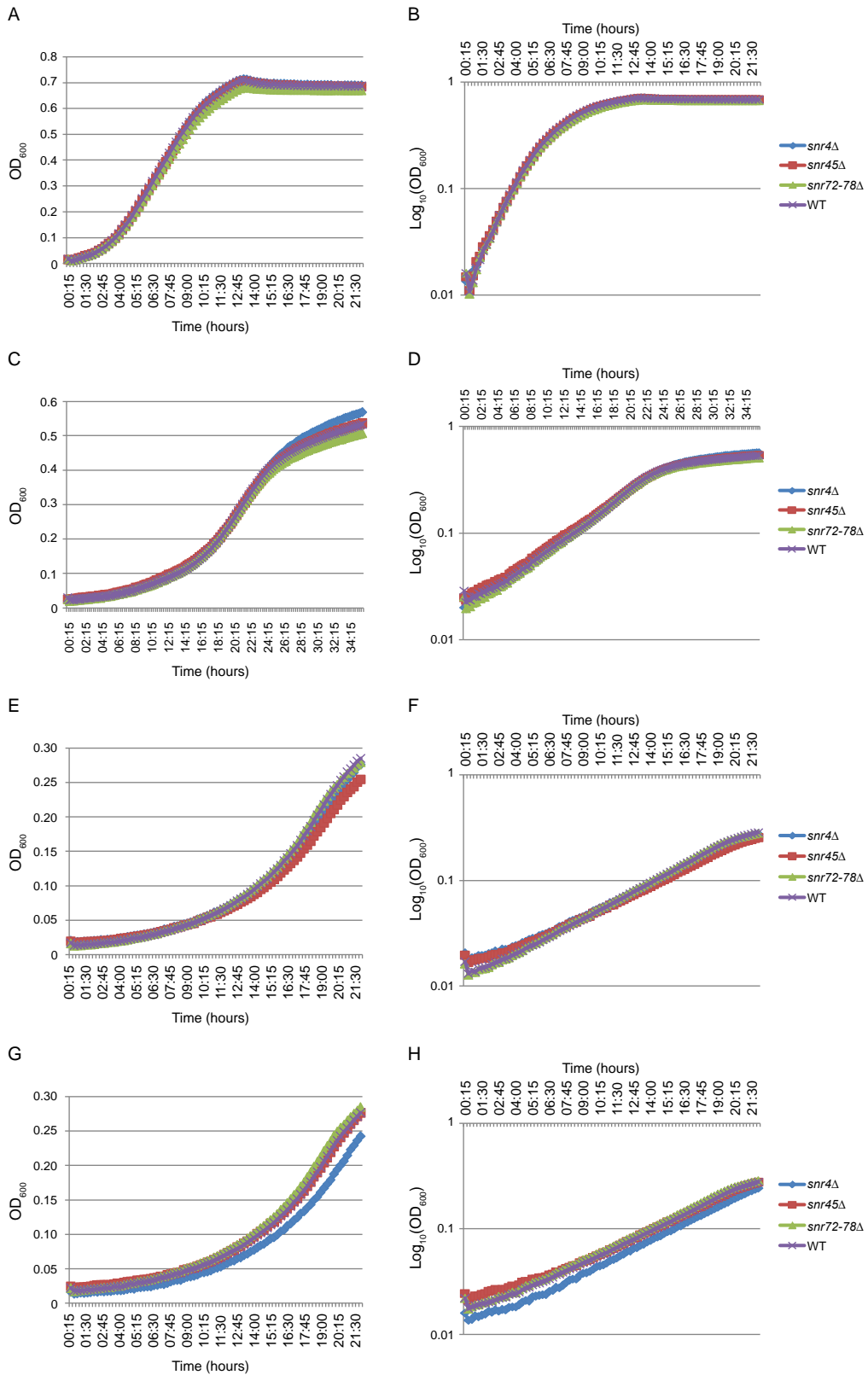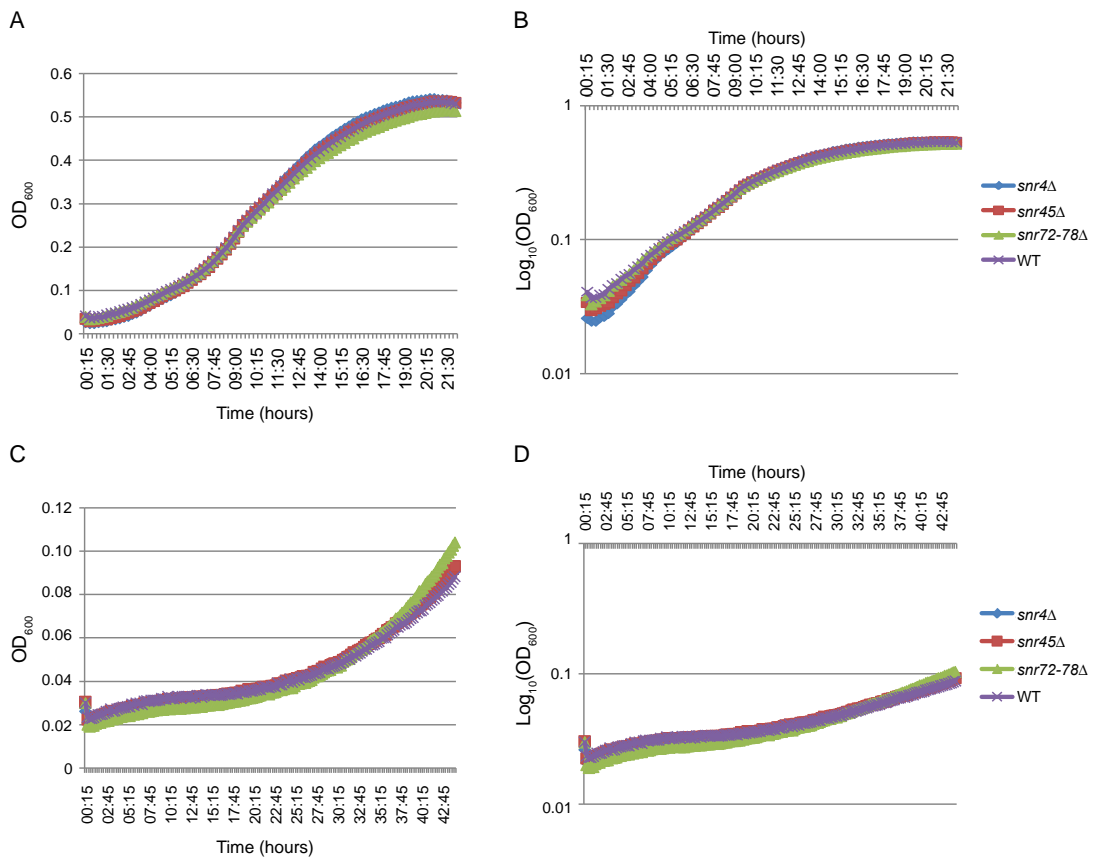

Figure 3.2.3 Growth curves of snoRNA deletion strains in medium containing poor nitrogen sources. A) Growth curve (OD$_{600}$) of *snr4Δ, snr45Δ, snr72-78Δ* and WT following shift from minimal medium containing ammonium (rich nitrogen source) to growth in minimal medium containing proline (poor nitrogen source). B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). C) Growth curve (OD$_{600}$) of *snr4Δ, snr45Δ, snr72-78Δ* and WT following shift from minimal medium containing ammonium with 2% w/v glucose to growth in minimal medium containing proline, with 2% v/v ethanol plus 2% v/v glycerol. D) As in 'C' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown across the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. Strains in panels 'A' and 'B' were grown at 30°C for 22h, whereas C and D were grown for 44h. Three biological replicates were grown, with three technical replicates analysed for each. A key indicating strains is shown to the right of each panel.

During the course of this project, strains were created replacing *SNR4* and *SNR45* with the gene for nourseothricin resistance, instead of kanamycin resistance as shown above. These were created for use in the SGA screen (chapters 3.1 and 3.3). The strains were grown to mid-log phase in a 2% w/v glucose minimal medium. Strains were diluted to 0.1 OD_{600} into a 96-well plate, with three technical replicates for each of three biological replicates, and grown for 22h at 30°C. Figure 3.2.4A shows the average of the three biological replicates plotted with error bars. Notably, both *snr4Δ* and *snr45Δ* with nourseothricin resistance grew much slower than WT. This is in contrast to the lack of phenotype observed above. A graph showing $log_{10}(OD_{600})$ shows that this is not due to a dilution error, but a reduction of growth rate (Figure 3.2.4B). The gene for antibiotic resistance replacing the gene of interest should not affect phenotype alone, but rather the deletion of the gene of interest causing a growth defect. It is thus unclear why *snr4::NatMX6* and *snr45::NatMX6* should display a phenotype when *snr4::KanMX6* and *snr45::KanMX6* do not.
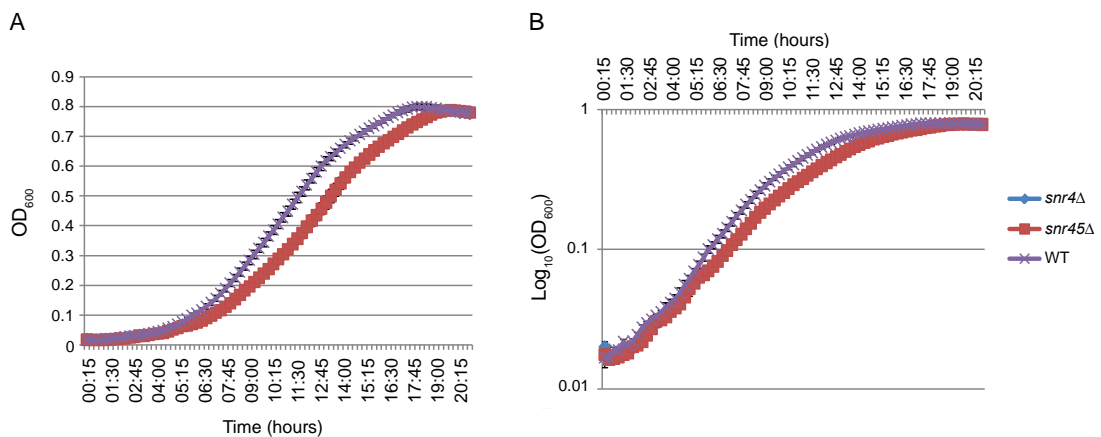
Figure 3.2.4 Growth curves of snoRNA deletion strains with nourseothricin resistance. A) Growth curve (OD_{600}) of *snr4::NatMX6, snr45::NatMX6,* and WT at 30°C for 22h. B) As in 'A' but showing growth rate ($log_{10}(OD_{600})$). Time in hours is shown across the x axis, with OD_{600} or $log_{10}(OD_{600})$ on the y axis. Three biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of each panel.

snr4Δ, snr45Δ and *snr72-78Δ* showed no phenotype under the different stresses tested. Given that *SNR72-78* is from a polycistronic cluster, it was considered that any phenotype that could be observed for this deletion might be more complex to interpret, compared to the monocistronic deletions of *SNR4* and *SNR45.* Therefore it was not included in any subsequent phenotypic experiments. It seemed possible that snR4 and snR45 were redundant in function, as they are the only two orphan yeast box C/D snoRNAs and neither showed a phenotype. Therefore, a double mutant strain was constructed. The *GAL1* promoter was inserted 5' to *SNR4* or *SNR45* in the reciprocal deletion strain, with the His selection marker 5' to this. *HISMX6-PGAL1-SNR45 snr4Δ, HISMX6-PGAL1-SNR4 snr45Δ, snr4Δ, snr45Δ* and WT strains were grown in minimal medium with 2% w/v galactose plus 2% w/v sucrose to mid-log phase, then shifted into minimal medium with 2% w/v glucose. Strains were again grown to mid-log phase, before dilution to 0.1 OD_{600} in 96-well plates. These were incubated at 30°C for 22h for phenotypic analysis. As controls, each *GAL1* promoter strain was grown in galactose plus sucrose medium and diluted onto the same plate. Figure 3.2.5A shows the growth curves of all strains grown in 2% w/v glucose medium. No growth defect was observed. This is supported by the $log_{10}(OD_{600})$ graph showing no difference in growth rate (Figure 3.2.5B). These graphs were segregated into strains containing *snr4Δ* and strains containing *snr45Δ,* to better observe smaller changes. Figures 3.2.5C and D show that depletion of snR45 in *snr4Δ* shows no difference in growth compared to both *snr4Δ* and WT. snR45 depletion in *snr4Δ* grown in galactose-sucrose medium showed much slower growth; however this is due to a slower cell growth rate based on carbon source, rather than being due to depletion of both snoRNAs. The same was observed in *snr45Δ* strains: snR4 depletion in *snr45Δ* showed no difference in growth compared to *snr45Δ* and WT when grown in glucose medium, but a slower growth rate was observed when grown in galactose medium (Figures 3.2.5E and F). These graphs clearly show that snR4 and snR45 do not share a redundant essential function, as depletion of both snoRNAs from the cell did not result in a growth phenotype. The same result was observed in drop-tests on solid agar plates (data not shown).
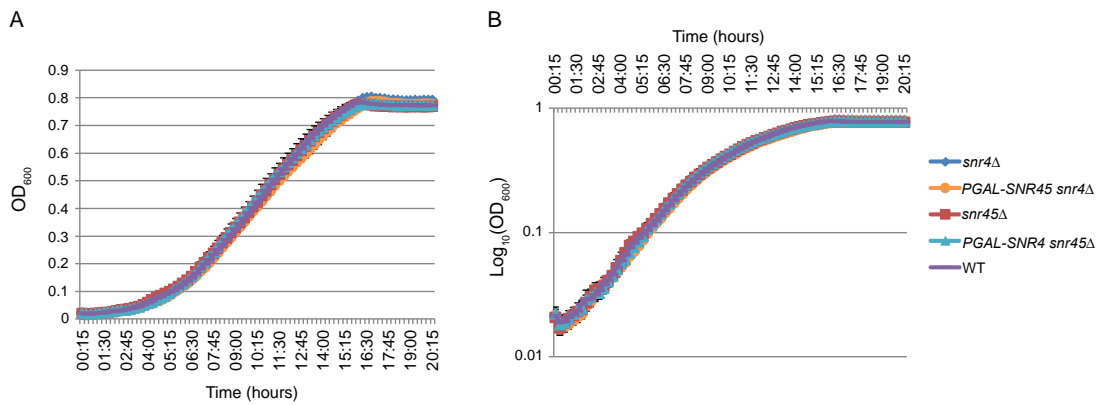
Figure 3.2.5 Growth curves of snoRNA double mutants. A) Growth curve (OD$_{600}$) of *snr4Δ, snr45Δ, HISMX6-PGAL1-SNR45 snr4Δ, HISMX6-PGAL1-SNR4 snr45Δ* and WT in minimal medium containing 2% w/v glucose. B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown across the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. All strains were grown at 30°C for 22h. Three biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of each panel.

3.3: Data analysis of SGA screen

snR4 and snR45 are non-essential snoRNAs, that show no clear phenotype upon deletion under different stresses. However, given the small size of the yeast genome, genes that have no function are unlikely to be retained as there is no advantage to keeping them, and thus no selection against losing them. As both snoRNAs are retained in the genome, they are likely to have a function. To identify this function, an SGA screen was performed. Deletion strains were constructed, replacing *SNR4* and *SNR45* with a gene encoding nourseothricin resistance (*NatMX6*) in a starter background strain provided by the Andrews/Boone lab, as described in chapter 2.2. The screen was performed by the Andrews/Boone lab, as described in chapter 3.1 (**Baryshnikova *et al.,* 2010**). This mated the *snr4Δ* and *snr45Δ* strains pairwise with mutations of the ~1000 yeast essential genes (Temperature Sensitive Array, or TSA) or deletion of the ~5000 non-essential yeast genes (Deletion Mutant Array, or DMA). The fitness of the double mutants was compared to fitness of *snr4Δ* or *snr45Δ* single mutants and fitness of array

mutations, giving a score for the strength of interaction for each double mutation. This determined whether the interaction between the two genes was synergistically negative (synthetic lethality) or positive (suppression).

Two replicates of each strain were assayed for each screen. A p value was reported showing an estimate of variance, determined based on local variability of replicate colonies (four colonies per plate for each double mutant) and variability of double mutants sharing the same mutation. Furthermore, the screen corrects for systematic experimental effects. The default cut-off for genetic interactions consists of a combination of p value <0.05 and SGA score >±0.08. The results of this screen were then analysed, and only the genes that appeared in both replicates were retained. The average score (strength of interaction) for the mutant from both replicates was taken, then ranked. The top 12 genes were extracted, including some weaker interactions that were functionally related. These were then grouped by common function, and are shown in the tables below.

Table 3.3.1 shows the top scoring genes identified in the TSA screen for *snr4Δ*. The *nop1-2* double mutant showed the strongest interaction of the screen, with a score of -0.956. A range of interaction was not provided, so the exact degree of growth impairment of this double mutant was unknown, but is presumed to be essentially synthetic lethal. For reference, a growth defect with a score of >±0.25 is visible by eye. Nop1p catalyses the methyl-transferase reaction performed by all canonical box C/D snoRNAs. Two other gene mutants that function in ribosome synthesis showed growth defects: *nop2-3* and *rsp5-3.* The high scores of these double mutants indicate that *SNR4* deletion has a reproducibly strong effect on a subset of genes involved in ribosome synthesis. A number of mutants were found with functions related to actin. The a*ct1-2* and *act1-3* double mutants showed synergistic negative phenotype, whereas the three *arp3* mutants showed suppression. Arp3p is an essential component of the Arp2/3 complex involved in actin nucleation. Similar interactions were uncovered involving microtubules. Mutations in *MPS1* combined with *snr4Δ* resulted in a suppression phenotype. Mps1p is required for spindle pole body duplication and spindle checkpoint function. It is bound by Mob1p and Mob2p (Mps one binder). Mob1p is required for cytokinesis, while Mob2p is involved in regulating cell polarity and morphogenesis. 13 *mob2* mutants each showed synergistic negative interactions with *snr4Δ,* but only the three highest scoring alleles are shown in Table 3.3.1. A *stu2* mutant was also synthetic sick. Stu2p

regulates microtubule dynamics during spindle orientation. Deletion of *SNR4* therefore showed clear interaction with the microtubule pathway.

Further double mutants were linked by involvement in the cell cycle. Cdc25p is required for progression through G1, Cdc10p is required for cytokinesis, Ycs4p is a component of the condensin complex, required for chromosome condensation and segregation, and Cdc21p expression is induced during G1 and is required for pyrimidine biosynthesis. In total, 16 *cdc* mutants showed synergistic negative phenotypes, while 10 *cdc* mutants were identified with suppression phenotypes. Two genes identified in the screen have functions in splicing: *YHC1,* and *SMD1.* Yhc1p is a component of the U1 snRNP complex. Double mutation with *snr4Δ* showed the strongest suppression phenotype within the screen. This is potentially a very important interaction. Smd1p is an Sm protein, contributing to the snRNPs. The final common function was translation. Mutants of *ded1* showed non-hierarchical heterogeneity, where one mutation resulted in a synergistic negative phenotype, but another mutation resulted in a suppression phenotype. Ded1p is a DEAD-box helicase required for translation initiation of all yeast mRNAs.

Down (synergistic negative)			Up (suppression)			
Gene	Allele	Score	Gene	Allele	Score	Function
NOP1	*nop1-2*	-0.956	*NOP2*	*nop2-3*	0.508	Ribosome synthesis/modificat on
RSP5	*rsp5-3*	-0.618				Ribosome synthesis
ACT1	*act1-2*	-0.410	*ARP3*	*arp3-31*	0.614	Actin
ACT1	*act1-3*	-0.306	*ARP3*	*arp3-f306g*	0.376	Actin
			ARP3	*arp3-d11a*	0.321	Actin
MOB1	*yil106w-ph*	-0.271	*MPS1*	*mps1-3796*	0.591	Microtubules
MOB2	*mob2-34*	-0.505	*MPS1*	*mps1-417*	0.586	Microtubules
MOB2	*mob2-11*	-0.503				Microtubules
MOB2	*mob2-20*	-0.462				Microtubules
STU2	*stu2-10*	-0.467				Microtubules
CDC25	*cdc25-1*	-0.518	*CDC10*	*cdc10-4*	0.527	Cell cycle
			YCS4	*ycs4-1*	0.452	Cell cycle
			CDC21	*cdc21-ts*	0.372	Cell cycle
			YHC1	*yhc1-1*	0.907	Splicing
			SMD1	*smd1-1*	0.508	Splicing
DED1	*ded1-f144c*	-0.480	*DED1*	*ded1-95*	0.530	Translation
QCR2	-	-0.603				Mitochondria
ORC6	*orc6-ph*	-0.519				DNA replication
GLC7	*glc7-12*	-0.446				Glycogen metabolism/ Transcription/Cell cycle
			HRP1	*hrp1-1*	0.358	mRNA processing

Table 3.3.1 Highest scoring double mutants for *snr4Δ* in the TSA screen. Double mutants were categorised by a negative SGA score (synergistic negative phenotype) or a positive SGA score (suppression phenotype). Double mutants were grouped by related functions. The table lists the gene, mutant allele, SGA score and related function. Only reproducible double mutants were analysed.

Four genes showed unrelated functions, but scored highly in the screen. Qcr2p is part of the cytochrome bc1 complex in mitochondria and showed the third strongest negative interaction. Orc6p is a subunit of the origin of replication complex, which directs DNA replication. This showed the fourth strongest negative phenotype of this screen. Glc7p is the catalytic subunit of type 1 protein phosphatase and is involved in a large number of cellular functions, including glycogen metabolism, transcription, translation, actin organisation and chromosome segregation in mitosis. Finally, Hrp1p is required for cleavage and polyadenylation of mRNAs and also functions in nuclear RNA surveillance.

Analysis of the *snr45Δ* TSA screen identified a high degree of overlap with the *snr4Δ* TSA screen (Table 3.3.2). The majority of the strongest interacting synergistically negative mutants for *snr45Δ* were identified in *snr4Δ*, while all but one of the strongest interacting suppression mutants for *snr45Δ* were identified in *snr4Δ*. The *nop1-2* double mutant again showed the strongest interaction within the screen, with an average score of -0.992. This shows that *nop1-2* is severely impaired in growth when paired with either *snr4Δ* or *snr45Δ*. The *nop2-3* double mutant also showed a suppression phenotype with *snr45Δ*. An *rsp5* mutant was identified for *snr45Δ* but with a different mutant allele, and a weaker interaction. *snr45Δ* showed similar interactions with mutations of genes in the actin pathway, with the addition of a weak interaction with *arp2-14*. The same two *mps1* mutants showed similar strength of suppression interaction with *snr45Δ*, and mutants of both *mob1* and *mob2* also showed a similar strength of synergistic negative interaction. In total, 13 *mob2* mutant alleles were identified as showing significant phenotypic differences with *snr45Δ*. Both *cdc10-4* and *cdc21-ts* showed suppression to a similar magnitude with *snr45Δ* as with *snr4Δ*, but neither *cdc25-1* nor *ycs4-1* was identified. However, a *cdc12* mutant showed synergistic negative interaction. This has a very similar function to Cdc10p, as a component of the septin ring, required for cytokinesis. 17 *cdc* mutants showed synergistic negative phenotypes, while 10 *cdc* mutants were identified with suppression phenotypes. The remaining mutants were all identified in the *snr4Δ* screen to a similar magnitude. These two snoRNAs are therefore likely to share related functions, despite the lack of a clear growth defect in the snoRNA double mutant strains.

Down (synergistic negative)			Up (suppression)			
Gene	Allele	Score	Gene	Allele	Score	Function
NOP1	nop1-2	-0.992	NOP2	nop2-3	0.469	Ribosome synthesis/modificaton
ACT1	act1-2	-0.425	ARP3	arp3-31	0.601	Actin
			ARP3	arp3-f306g	0.405	Actin
			ARP3	arp3-d11a	0.304	Actin
			ARP2	arp2-14	0.112	Actin
MOB1	yil106w-ph	-0.241	MPS1	mps1-3796	0.607	Microtubules
MOB2	mob2-11	-0.482	MPS1	mps1-417	0.583	Microtubules
MOB2	mob2-34	-0.482				Microtubules
MOB2	mob2-38	-0.451				Microtubules
MOB2	mob2-20	-0.442				Microtubules
STU2	stu2-10	-0.622				Microtubules
CDC48	cdc48-1	-0.446				Microtubules
CDC12	cdc12-td	-0.426	CDC10	cdc10-4	0.534	Cell cycle
			CDC21	cdc21-ts	0.399	Cell cycle
			YHC1	yhc1-1	0.922	Splicing
			SMD1	smd1-1	0.485	Splicing
DED1	ded1-f144c	-0.508	DED1	ded1-95	0.513	Translation
QCR2	-	-0.559				Mitochondria
ORC6	orc6-ph	-0.517				DNA replication
			HRP1	hrp1-1	0.350	mRNA processing

Table 3.3.2 Highest scoring double mutants for *snr45Δ* in the TSA screen. Double mutants were categorised by a negative SGA score (synergistic negative phenotype) or a positive SGA score (suppression phenotype). Double mutants were grouped by related functions. The table lists the gene, mutant allele, SGA score and related function. Only reproducible double mutants were analysed.

It was noted that the DMA screen identified genes transcribed from regions adjacent to *SNR4* and *SNR45* as showing apparent synergistic negative phenotypes with either *snr4Δ* or *snr45Δ,* respectively. This is an expected artefact of the genetic screen, which relies on recombination between the snoRNA gene and the test gene to generate the double mutant haploid. The presence of these genes in the screen demonstrates that the overlap in targets observed in TSA screens between *snr4Δ* and *snr45Δ* was not due to contamination or the use of incorrect strains.

Table 3.3.3 shows the top scoring genes identified in the DMA screen for *snr4Δ*. The top 12 genes were analysed; however any double mutant scoring <±0.25 (visibility

cut-off) was removed from the table unless in a related pathway. Globally, double mutants with non-essential genes showed a weaker phenotype than those with temperature-sensitive essential mutants. Furthermore, the top ranking genes were less functionally related in DMA screens. However, similar pathways were still identified. Two mutants were related to cell cycle: *dse1* and *whi5*. *DSE1* deletion has been shown to affect cell separation after division. Whi5p is a repressor of transcription during G1. Two genes were involved in meiosis: *PMD1* and *SPO23*. Pmd1p negatively regulates early meiotic gene expression. Spo23p has an unknown function, but associates with the meiosis-specific protein, Spo1p. Three genes had functions related to vacuoles: *VFA1, AVT6* and *YCK3*. Vfa1p has a role in vacuolar sorting and stimulates activity of another vacuolar protein. Avt6p exports aspartate and glutamate from vacuoles, and Yck3p is a casein kinase localised in the vacuolar membrane, which regulates vacuolar fusion during stress. Two other genes are also involved in stress response. Sho1p is a transmembrane osmosensor involved in the HOG pathway and filamentous growth. Wwm1p has an unknown function, but is involved in regulation of hydrogen peroxide-induced apoptosis.

Three genes that function in ribosome biogenesis were identified in this screen: *RPS26B, KNS1* and *RPL8B*. Rps26bp is a small ribosomal subunit protein, while Rpl8bp is a large ribosomal subunit protein, so both are directly related to the ribosome. Kns1p, however, negatively regulates PolIII transcription, which controls production of 5S rRNA and tRNAs. This supports the interactions identified in the TSA screen that functionally implicate snR4 in ribosome biogenesis. Two of the identified genes function within transcription. Rtr1p dephosphorylates serine 5-P in the C-terminal domain (CTD) of Rpo21p. Cbf1p, however, has multiple functions, mediating transcriptional activation/repression, affecting nucleosome positioning, and chromosome segregation. This also relates to pathways highlighted in the TSA screen. Syc1p is required for cleavage and polyadenylation of mRNAs, similar to *HRP1* identified in the TSA screen. These double mutants showed suppression of a similar magnitude. The highest scoring interaction was a double mutant with *SAK1*, with a score of -0.764. Sak1p is a kinase involved in glucose metabolism. Finally, Yme1p is a protease located in the mitochondrial inner membrane, while Scs2p is an endoplasmic reticulum (ER) membrane protein that regulates phospholipid metabolism. Deletion of these two genes showed synergistic negative effects.

Down (synergistic negative)			Up (suppression)		
Gene	Score	Function	Gene	Score	Function
DSE1	-0.502	Cell cycle	WHI5	0.305	Cell cycle
PMD1	-0.532	Meiosis	SPO23	0.341	Meiosis
VFA1	-0.606	Vacuolar sorting			
AVT6	-0.501	Vacuolar transporting			
YCK3	-0.494	Vacuolar casein kinase			
SHO1	-0.382	Stress response			
WWM1	-0.348	Stress response			
RPS26B	-0.574	Ribosome biogenesis	KNS1	0.320	Ribosome biogenesis
			RPL8B	0.233	Ribosome biogenesis
RTR1	-0.459	Transcription	CBF1	0.262	Transcription
			SYC1	0.309	mRNA processing
SAK1	-0.764	Glucose metabolism			
YME1	-0.447	Mitochondria			
SCS2	-0.400	Phospholipid metabolism			

Table 3.3.3 Highest scoring double mutants for *snr4Δ* in the DMA screen. Double mutants were categorised by a negative SGA score (synergistic negative phenotype) or a positive SGA score (suppression phenotype). Double mutants were grouped by related functions. The table lists the gene mutated, SGA score and related function. Only reproducible double mutants with scores >±0.25 were analysed.

Analysis of the DMA screen for *snr45Δ* showed a number of pathways that were similar to those detected for *snr4Δ,* also identifying some of the same genes. *WHI5, SPO23, PMD1, YME1* and *KNS1* were all identified in both screens. Interestingly, *PMD1* showed a suppression phenotype with *snr45Δ,* but a synergistic negative phenotype with *snr4Δ. SYC1* and *CBF1* were also identified, but with a score below the cut-off. Two other genes involved in the cell cycle were identified in the *snr45Δ* DMA screen. Ctf4p is required for sister chromatid cohesion, whereas Kar3p localises to the spindle pole body. Kar3p is a microtubule motor, which also links it to the microtubule functions distinguished in the TSA screens. *snr45Δ* affected a higher number of mitochondrial related pathways in the double mutant than *snr4Δ.* The *taz1* double mutant showed the second highest score within this screen. Taz1p is an acyltransferase, required for normal phospholipid content of mitochondrial membranes. Sue1p degrades unstable forms of cytochrome c, located in the mitochondria, while Nca2p regulates expression of mitochondrial ATP synthase subunits. Deletion of these four genes (including *YME1*) resulted in synergistic

negative phenotype with *snr45Δ*. However, deletion of *GCV1* resulted in suppression phenotype. This showed a weak interaction, but is a subunit of the mitochondrial glycine decarboxylase complex.

The remaining genes had functions that were not clearly related, although a number of these overlapped with pathways detected in the other screens. Urn1p is a protein of unknown function, but was implicated as a splicing factor. A double mutant with *asn1* showed the largest score of the DMA screen. Asn1p is an asparagine synthetase that catalyses the synthesis of asparagine from aspartate. Mep3p is a cytoplasmic membrane protein that transports ammonium. Nce102p is another protein of unknown function, but is a transmembrane protein implicated in protein secretion. Pin3p negatively regulates the activity of an actin nucleation-promoting factor. This links to the genes involved in actin pathways detected in the TSA screens. Tpo3p is a polyamine transporter. All of these double mutants showed synergistic negative phenotypes with *snr45Δ*. Double mutants showing suppression phenotypes revealed much weaker interactions than the synergistic negative mutants. Ssd1p is a translational repressor with a role in polar growth. Pfk26p is a phosphofructokinase, thus functioning in glycolysis. Finally, Pho86p guides the exit of a phosphate transporter from the ER.

Down (synergistic negative)			Up (suppression)		
Gene	Score	Function	Gene	Score	Function
KAR3	-0.473	Microtubules/Cell cycle	*WHI5*	0.395	Cell cycle
CTF4	-0.459	Cell cycle			
			SPO23	0.349	Meiosis
			PMD1	0.315	Meiosis
TAZ1	-0.803	Mitochondria	*GCV1*	0.276	Mitochondria
SUE1	-0.601	Mitochondria			
NCA2	-0.493	Mitochondria			
YME1	-0.350	Mitochondria			
URN1	-0.369	Splicing	*SSD1*	0.254	Translation
ASN1	-0.885	Amino acid biogenesis	*PFK26*	0.435	Glycolysis
MEP3	-0.527	Ammonium transport	*PHO86*	0.261	ER
NCE102	-0.502	Protein secretion	*KNS1*	0.260	Ribosome biogenesis
PIN3	-0.485	Actin			
TPO3	-0.374	Polyamine transport			

Table 3.3.4 Highest scoring double mutants for *snr45Δ* in the DMA screen. Double mutants were categorised by a negative SGA score (synergistic negative phenotype) or a positive SGA score (suppression phenotype). Double mutants were grouped by related functions. The table lists the gene mutated, SGA score and related function. Only reproducible double mutants with scores >±0.25 were analysed.

To analyse the level of overlap between these conditions, the SGA scores from both replicates for each snoRNA were compiled. Double mutants scoring <±0.25 were removed, due to the high numbers of mutants showing weak interaction. The remaining mutants were then scored based on the number of replicates in which they were identified. If the double mutant was identified >±0.25 in both replicates of both snoRNAs, it was given a score of '4'. If it was identified in both replicates of one snoRNA, but only one replicate of the other snoRNA, it was scored a '3'. If the double mutant was identified in both replicates of one snoRNA, but no replicates of the other snoRNA, it was scored '2.1'. However, if it was identified in both snoRNAs but in only one replicate, it was scored '2.2'. Double mutants identified in only one replicate for one snoRNA were scored a '1'. Table 3.3.5 shows the percentages for each screen, separated by whether the phenotype was synergistic negative or suppressive. The majority of TSA double mutants with synergistic negative interaction were identified in both replicates for both snoRNAs. For TSA double mutants with suppression interactions, this was the second highest score. The

majority of double mutants in this screen were identified in only one replicate of one snoRNA. This highlights the importance of filtering for reproducible phenotypes. The majority of double mutants were only identified in one replicate of one snoRNA for both phenotypes in the DMA screen. For suppression phenotypes, all other categories had very small numbers of double mutants with strength of interaction >±0.25. However, for synergistic negative phenotypes, the second highest majority scored 2.1. This means both replicates from one snoRNA contained the double mutant scoring >±0.25. A number of these mutants are from areas surrounding the deleted snoRNA in the genome, and thus are likely to be artefacts. Analysis of all significant SGA scores (>±0.08) showed that the percentages in table 3.3.5 represent similar distributions for all significant SGA scores.

The majority of DMA suppression phenotypes were not reproducible. However, only 77 DMA suppression double mutants gave an SGA score >±0.25. The majority of DMA suppression interactions were therefore weak. 148 double mutants showed synergistic negative DMA interactions >±0.25. For the TSA screen, a similar ratio was observed, with only 107 double mutants showing a suppression phenotype, but 195 showing a synergistic negative phenotype. Therefore, double mutants with synergistic negative phenotypes globally showed stronger interactions.

Replicate score	4	3	2.1	2.2	1
TSA Synergistic Negative	40.8%	14.1%	15.5%	8.5%	21.1%
TSA Suppression	34.1%	13.6%	9.1%	4.5%	38.6%
DMA Synergistic Negative	3.2%	2.1%	41.5%	2.1%	51.1%
DMA Suppression	7.3%	7.3%	1.8%	1.8%	81.8%

Table 3.3.5 Percentages of SGA hits in four categories for each screen. Categories represent the number of replicates identified for each double mutant. 4 = identified in both replicates for both snoRNA deletion strains; 3 = identified in both replicates for one snoRNA deletion strain, and one replicate of the other snoRNA deletion strain; 2.1 = identified in both replicates for one snoRNA deletion strain; 2.2 = identified in one replicate for both snoRNA deletion strains; 1 = identified in one replicate of one snoRNA deletion strain. Screens were classified by TSA or DMA, and synergistic negative phenotype or suppression phenotype.

Only three genes overlapped between the DMA synergistic negative screen and DMA suppression screen: *PMD1, RPL23B* and *RPS26B.* All three genes showed synergistic negative phenotypes in the *snr4Δ* double mutants, but suppression phenotypes in the *snr45Δ* double mutants. All other DMA double mutants showed a single direction of phenotype. Analysing the overlap of genes between synergistic negative and suppression phenotypes of TSA double mutants identified 10 genes showing non-hierarchical heterogeneity in *snr4Δ* (18 >±0.08)*,* 14 genes showing non-hierarchical heterogeneity in *snr45Δ* (20 >±0.08)*,* and 9 genes that were identified in both snoRNAs (13 >±0.08).

3.4: Evaluation of TSA growth phenotypes

The SGA screen identified a number of pathways that snR4 and snR45 may be involved in, and highlighted which double mutants result in a visible phenotype. However, the SGA screen may include false positives, and thus each strain needs to be validated before further experimentation. Single mutant TSA strains from tables 3.3.1 and 3.3.2 were received from the Andrews lab. However, a number of these no longer possessed temperature-sensitive phenotypes when tested, thus they were not continued with experimentation. As the exact phenotype of the double mutants was unknown, *SNR4* and *SNR45* were initially tagged with a promoter from the *GAL1* gene in strains, to ensure clear transformation success. None of the double mutants tested were synthetic lethal. Therefore, the snoRNA genes were deleted in these strains. Each strain (*snr4::NatMX6 ts::KanMX6* or *snr45::NatMX6 ts::KanMX6)* was grown with *snr4::NatMX6 (snr4Δ)* or *snr45::NatMX6 (snr45Δ)* and the original TS mutant strain (**Li *et al.,* 2011**). Strains were grown in 2% w/v glucose minimal medium to mid-log phase, then diluted to 0.1 OD_{600} in 96-well plates and analysed in triplicate in the plate reader at 30°C for <44h.

Figure 3.4.1 shows the growth curves of *yhc1-1, snr4Δ yhc1-1, snr45Δ yhc1-1, snr4Δ* and *snr45Δ.* Three technical replicates for each strain were analysed, with two biological repeats, shown by error bars. Time in hours is shown along the x axis, with OD_{600} or $log_{10}(OD_{600})$ on the y axis. The SGA scores of the double mutants were 0.907 and 0.922 for *snr4Δ* and *snr45Δ,* respectively. This means the double mutant would be expected to grow much faster than the single mutants alone. *yhc1-1* grew at the same rate as both *snr4Δ* and *snr45Δ.* Both double mutants also grew

at the same rate, showing no phenotype. The same result was observed in drop-tests on rich medium and minimal medium plates (data not shown).
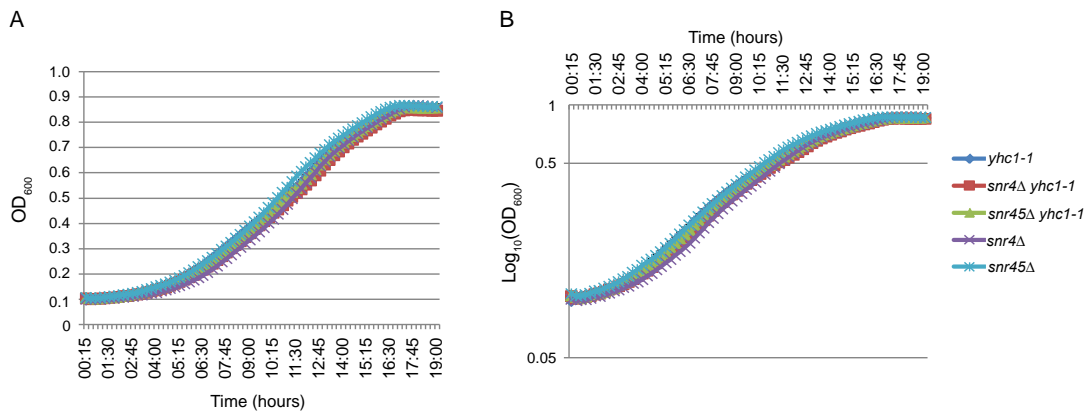
Figure 3.4.1 Growth curves of *yhc1-1* mutants with snoRNA deletions. A) Growth curve (OD$_{600}$) of *yhc1-1, snr4Δ yhc1-1, snr45Δ yhc1-1, snr4Δ* and *snr45Δ* at 30°C for 19h. B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown on the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. Two biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of the panel.

The *mob1* and *mob2* mutants were among those that did not display TS phenotypes when tested. However, the *mps* mutants were TS. This allowed for validation of one side of the pathway. *mps1-3796* was grown with *snr4Δ mps1-3796* and *snr45Δ mps1-3796.* Three technical replicates for each strain were analysed, with three biological repeats, shown by error bars. The SGA scores of the double mutant were 0.591 and 0.607 for *snr4Δ* and *snr45Δ*, respectively. This means the double mutant would be expected to grow faster than the single mutants alone. Figure 3.4.2A shows that both double mutants grew slightly faster than *mps1-3796,* with *snr45Δ mps1-3796* growing faster than *snr4Δ mps1-3796.* However, this phenotype was mild, and log$_{10}$ representation of the growth curves showed that the growth rates did not obviously differ (Figure 3.4.2B). No significant phenotype was observed in drop-tests on rich medium and minimal medium plates (data not shown).
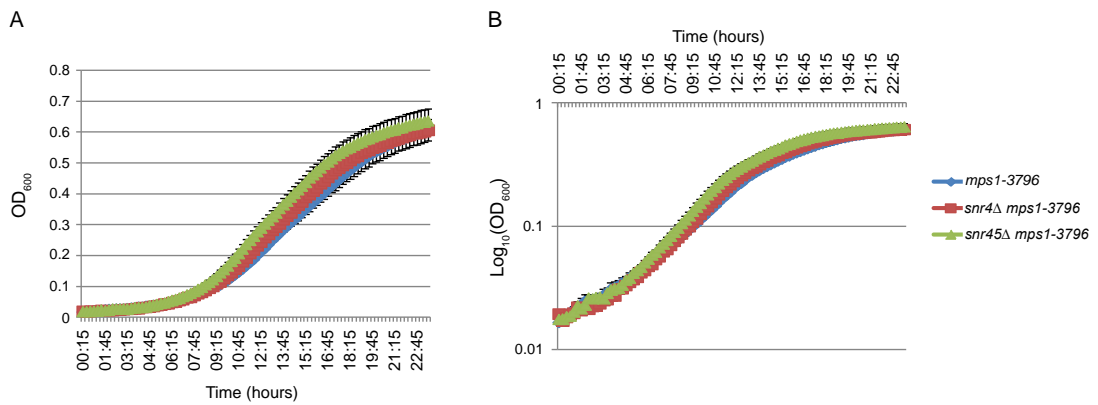
Figure 3.4.2 Growth curves of *mps1-3796* mutants with snoRNA deletions. A) Growth curve (OD$_{600}$) of *mps1-3796, snr4Δ mps1-3796* and *snr45Δ mps1-3796* at 30°C for 23h. B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown on the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. Three biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of the panel.
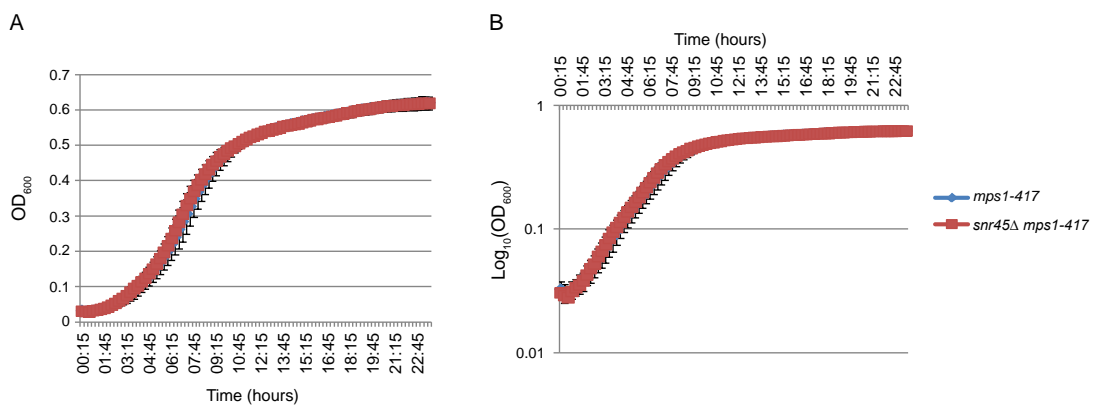
Figure 3.4.3 Growth curves of *mps1-417* mutants with snoRNA deletions. A) Growth curve (OD$_{600}$) of *mps1-417* and *snr45Δ mps1-417* at 30°C for 23h. B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown on the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. Three biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of the panel.

Transformation of the snoRNA deletion cassette for *snr45Δ* was successful for *mps1-417,* but no transformants were obtained for *snr4Δ mps1-417.* Therefore *mps1-417* was only compared to *snr45Δ mps1-417.* Three technical replicates for each strain were analysed, with three biological repeats, shown by error bars. The

SGA score for the double mutant was 0.586. This means the double mutant would be expected to grow faster than the single mutants alone. Figure 3.4.3 shows no observable difference between growth rates of *snr45Δ mps1-417* and *mps1-417* single mutant. The same result was observed in drop-tests on solid media plates (data not shown).
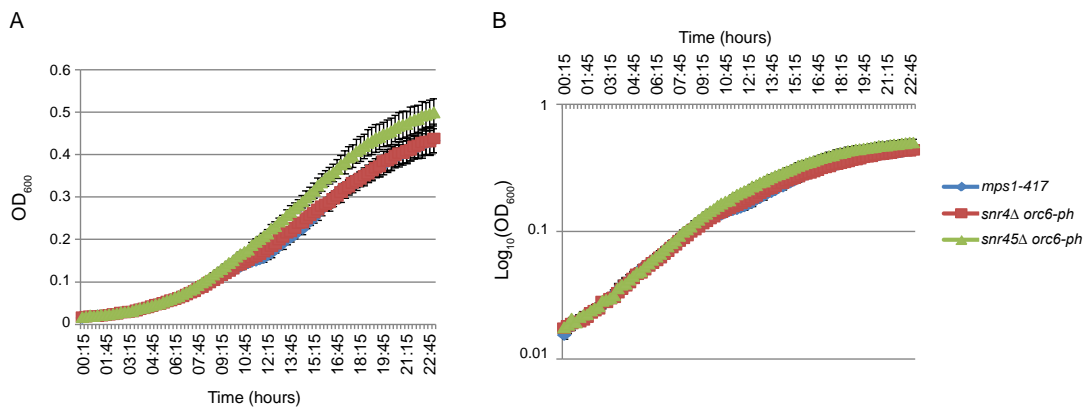
Figure 3.4.4 Growth curves of *orc6-ph* mutants with snoRNA deletions. A) Growth curve (OD$_{600}$) of *orc6-ph, snr4Δ orc6-ph* and *snr45Δ orc6-ph* at 30°C for 23h. B) As in 'A' but showing growth rate (log$_{10}$(OD$_{600}$)). Time in hours is shown on the x axis, with OD$_{600}$ or log$_{10}$(OD$_{600}$) on the y axis. Three biological replicates were grown, with three technical replicates analysed for each. Error bars represent biological variance. A key indicating strains is shown to the right of the panel.

Finally, *orc6-ph* and the double mutants, *snr4Δ orc6-ph* and *snr45Δ orc6-ph*, were analysed. The SGA scores for the double mutants were -0.519 and -0.517, respectively, meaning that the double mutant would be expected to grow slower than the single mutants. However, the growth curve in Figure 3.4.4A shows that *snr45Δ orc6-ph* grew faster than either *snr4Δ orc6-ph* or *orc6-ph.* The error bars of *snr45Δ orc6-ph* did not overlap the error bars of *snr4Δ orc6-ph* nor *orc6-ph* during part of exponential growth, and the mean final OD$_{600}$ of *snr45Δ orc6-ph* was 0.067 higher than the single mutant alone. Analysis of the growth rate by log$_{10}$(OD$_{600}$) showed *snr45Δ orc6-ph* did have a faster growth rate than either *snr4Δ orc6-ph* or *orc6-ph,* but the difference was marginal (Figure 3.4.4B). Therefore, more experiments will be needed to show if *snr45Δ orc6-ph* reproducibly grows faster than *snr4Δ orc6-ph* and *orc6-ph,* and how strong this phenotype is. No visible difference was observed in drop-tests on rich and minimal medium plates (data not shown).

A number of single mutant strains that showed TS phenotype were not tested: *rsp5-3, ycs4-1, cdc25-1* and *cdc10-4.* Double mutants were not constructed with these strains due to time limitations. Therefore, the phenotypes of the double mutants of these strains are yet to be validated. This may highlight whether snR4 and snR45 function within the cell cycle pathway.

3.5: Evaluation of Nop1 mutant interactions

The *nop1-2* double mutant showed the strongest phenotype of both TSA screens. Nop1p is the methyltransferase canonically bound to all box C/D snoRNAs. Furthermore, this mutant has been described in a previous study (**Tollervey *et al.,* 1993**). In that study, mutations were made in *NOP1,* five of which showed TS phenotype: *nop1-2, nop1-3, nop1-4, nop1-5* and *nop1-7.* Further investigation showed that these could be classed by the processes they affected. *nop1-2* and *nop1-5* affected cleavage and processing of the rRNA transcript, *nop1-3* affected methylation of target sites, and *nop1-4* and *nop1-7* affected ribosome subunit assembly. However, the mechanisms of these interactions were unclear. Upon *SNR4* and *SNR45* deletion in the SGA screen, *nop1-2* showed potential synthetic lethality. This indicates that snoRNAs may be involved in the rRNA processing defects observed.

Therefore, *SNR4* and *SNR45* were tagged with the *GAL1* promoter in the strains listed in the above paper. However, the transformations proved unsuccessful. Constructs were then designed to mutate the WT *NOP1* sequences in *snr4Δ* and *snr45Δ.* WT *NOP1* coding sequence was optimised, to ensure the mutations would replace the WT sequence. The mutations were designed using the substitutions from the above paper, with a FLAG tag 3' to the coding sequence for detection, and the *URA3* gene for selection purposes. Successful transformants were Sanger sequenced, as describe in chapter 2.9. However, due to time constraints, further experiments were not possible.

Future work with these strains will hopefully elucidate snR4 and snR45 involvement in the defects caused by these mutations. It would be interesting to observe whether the potentially lethal phenotype is validated, and whether this phenotype is also identified for *snr4Δ* and *snr45Δ* with the other *nop1* mutants, particularly *nop1-5.* The

phenotype may be specific to the function of rRNA processing, or may be specific only to *nop1-2.* The exact mechanism would be exciting to uncover.

3.6: Discussion

The aim of this chapter was to phenotypically characterise *snr4Δ* and *snr45Δ,* and identify pathways that may be affected by these deletions. Growth curves revealed that neither *snr4Δ* nor *snr45Δ* had a growth defect. This was performed under a number of different stress conditions, none of which highlighted any phenotype. None of the stresses used involved the use of chemicals. Inhibition using different drugs might thus uncover a specific phenotype. Notably, strains containing the gene for kanamycin resistance replacing *SNR4* and *SNR45* showed no phenotype, but strains containing the gene for nourseothricin resistance replacing *SNR4* and *SNR45* showed a mild growth defect (Figure 3.2.4). The primers used to make the replacement were the same for both deletions, and the gene encoding for antibiotic resistance should have no effect on the growth rate of the cell. Therefore, it is unclear how this could be anything but an artefact. Depletion of one snoRNA in the deletion strain for the other snoRNA also showed no growth phenotype, indicating that the two snoRNAs are not functionally redundant and the cell can survive without both. Applying stress conditions to this double mutant may identify a phenotype.

Analysis of the SGA screen revealed that double mutations with TS mutants generally showed stronger phenotypes than double deletions with non-essential genes, as might be expected. TS double mutants were also more functionally related than those with non-essential deletions. In both screens, synergistic negative interactions globally had a stronger phenotype than those with suppression interactions. Furthermore, the overlap between TS genes identified for *snr4Δ* and *snr45Δ* was much higher than the overlap between non-essential genes identified.

Within the TSA screen, a number of double mutants with high scores had functions related to actin. Double mutants with *ACT1* showed synergistic negative phenotypes. *ACT1* codes for actin, which is a major cytoskeletal protein involved in cell polarisation and other processes. The Arp2/3 complex is an actin nucleation complex that binds to existing actin filaments. It allows the formation of a new actin filament branched 70° from the existing filament (**Volkmann *et. al.,* 2001**). Double

mutants containing an *ACT1* mutation combined with deletion of *SNR4* or *SNR45* showed reduced growth compared to single mutants. However, combination of *ARP3* (and to a lesser extent, *ARP2)* mutations with deletion of the snoRNA showed increased growth compared to single mutants. Mutations of multiple different alleles also replicated this effect. This is strong evidence that both snR4 and snR45 play a role in actin nucleation or cytoskeletal processes. A possible hypothesis is that snR4 and snR45 facilitate polymerisation of actin, which regulates cell polarisation. Cells can survive without snR4 or snR45, but a combination of this with particular actin mutants may result in the inability to polymerise and form filaments. However, snR4 and snR45 may negatively regulate the branching of actin filaments via interaction with the Arp2/3 complex. Mutation of these proteins combined with *SNR4* or *SNR45* deletion may result in increased branching of actin filaments, which could result in an increased growth rate. It would be advantageous to experimentally validate these interactions in the lab.

A number of TSA double mutants with high scores also had functions related to microtubules. Mps1p is a kinase that phosphorylates Mob1p (**Luca and Winey, 1998**). It acts to coordinate spindle pole body (SPB) duplication and M-phase checkpoints (**Weiss and Winey, 1996**). Mob1p is part of the Mitotic Exit Network (MEN) which coordinates a number of processes involved in the exiting of mitosis, such as cytokinesis (**Luca and Winey, 2001**). It shares sequence similarity to Mob2p and both show similar functions (**Luca and Winey, 1998, Weiss *et. al.,* 2002**). Mob2p is also involved in maintenance of polarised growth. Mutation of *MPS1* paired with deletion of *SNR4* or *SNR45* resulted in a higher growth rate compared to single mutants alone. However, mutation of *MOB1* or *MOB2* paired with snoRNA deletion resulted in a lower growth rate. A high number of *mob2* mutants resulted in this phenotype. This strongly suggests snR4 and snR45 interaction within this pathway. However, experimental validation of the *mps1* double mutants could not verify the effects observed in the SGA screen. It would be valuable to discern whether *mob1* and *mob2* double mutants could replicate the effect seen in the SGA screen. One possible hypothesis is that snR4 and snR45 interact with Mps1p to coordinate SPB duplication and M-phase checkpoints. If *MPS1* is mutated, snR4 and snR45 would still ensure checkpoints and SPB duplication; however, when *SNR4* and *SNR45* are deleted, these processes are no longer regulated, meaning the cell can replicate faster and thus the strain grows quicker. Similarly, snR4 and snR45 may interact with Mob1p and Mob2p to

coordinate cytokinesis and exiting of mitosis. However, the mutations in *MOB1* and *MOB2* may hinder their function, which is fully inhibited when *SNR4* and *SNR45* are also deleted. This would mean the cell could not successfully complete mitosis and may die, or become very sick.

The double mutants from the SGA screen that were reconstructed in the lab did not replicate the phenotypes observed in the screen. The majority of these showed mild or no phenotype, and the *snr45Δ orc6-ph* double mutant showed the opposite direction of growth defect. Recreating double mutants from the DMA screen would indicate which of these related pathways were valid. However, as these genes are non-essential, a phenotype may be harder to determine than for essential gene mutations. Alternatively, as cells lacking non-essential genes are generally less sick than those with mutations of essential genes, a phenotype for a double mutant may be more obvious to identify.

Double mutants from both screens for both snoRNAs identified genes involved in ribosome biogenesis. snoRNAs canonically function during ribosome biogenesis, so involvement of snR4 and snR45 would be logical. The highest scoring phenotype from both screens was the double mutant containing *nop1-2*, showing a potentially synthetic lethal phenotype. The hypothesis would be that snR4 and snR45 ameliorate the defects caused by *nop1-2,* which become lethal upon deletion of either snoRNA. The snoRNAs may function only with the *nop1-2* mutation, or may be involved in the processing defects observed with both *nop1-2* and *nop1-5.* The constructs created from chapter 3.5 will be vital in determining the mechanics of this interaction. Future work will involve analysing the growth curves of the double mutants compared to the single mutants, to validate the effects seen in the SGA screen, and northern blot analysis of rRNA. This would identify which species of pre-rRNA accumulates and which is depleted in the double mutant, thus refining potential mechanisms.

4 Non-canonical Interactions between snoRNAs and mRNAs

4.1: Introduction

The orphan snoRNAs snR4 and snR45 do not meet the criteria of canonical box C/D snoRNAs. Despite binding to the box C/D proteins, they lacked any known targets or function. A number of studies demonstrated an interaction between orphan snoRNAs and mRNAs (discussed in chapter 1.3), such as snoRNA involvement in mRNA alternative splicing (**Kishore and Stamm, 2006**) or targeting specified mRNAs for degradation (**Ono *et al.,* 2010**). However, no study has yet provided a general mechanism or function for interactions between orphan snoRNAs and mRNA. Whilst there has been evidence for snoRNAs functioning in mRNA splicing and degradation, it is unclear whether they may also function during transcription or translation. As discussed in chapter 1.6, the most likely starting hypothesis was that yeast orphan snoRNAs might have an effect on mRNA degradation or steady state levels. In order to probe this interaction, the mRNA targets of snoRNAs must be uncovered.

CLASH was performed on the canonical snoRNA proteins, Nop1p, Nop56p and Nop58p, to identify snoRNA-RNA hybrids (unpublished work, T. Dudnakova). A large number of mRNA interactions were found in the preliminary CLASH data. The two yeast orphan snoRNAs were amongst those that bind mRNAs. Experiments were therefore designed to identify mRNA targets of these two snoRNAs, and investigate the consequence of snoRNA knockdown on target mRNAs.

4.2: RNA sequencing analysis

To determine the effect orphan snoRNAs have on mRNAs, sequencing libraries were created from strains in which these snoRNAs were deleted from the genome. snR4, snR45 and the snR72-78 cluster previously showed interactions with mRNAs under stress conditions (unpublished data). It was possible that these interactions

would be more important during stress conditions. Therefore, sequencing libraries were prepared for both standard conditions and stress conditions. As the project initially focussed on snR4 and the snR72-78 cluster, a sequencing library was not prepared for *snr45Δ*. s*nr4Δ, snr72-78Δ* and WT were grown in minimal medium containing 2% w/v glucose at 30°C as described in chapter 2.3. Half the culture was harvested (standard condition), while the other half was incubated on ice for 20 minutes in ice cold PBS solution, before harvesting (ice condition). RNA-seq on total RNA is dominated by rRNA, limiting the data that can be obtained for other RNA species. Recovery of rRNA can be reduced by oligonucleotide-mediated depletion using a RiboMinus kit, or full-length mRNAs can be enriched using oligo (dT) selection. The isolated RNA was thus treated with a RiboMinus kit as described in chapter 2.6, libraries were prepared and RNA sequencing was performed. RNA expression levels in the two strains carrying snoRNA deletion were compared to the WT.

The *snr4Δ* strain showed significant changes in the expression of numerous RNA species in both the standard and ice conditions, relative to WT. In contrast, the *snr72-78Δ* strain showed no significant increase or decrease in any mRNAs (data not shown), supporting the conclusion that differences seen in the absence of snR4 are meaningful. Figure 4.2.1 shows volcano plots of the RNA sequencing data for the *snr4Δ* strain. A negative value on the x axis signifies a decrease in relative RNA expression upon *SNR4* deletion, while a positive value signifies an increase. The y axis gives the $-\log_{10}$(p value), so a higher positive value on this axis indicates a more significant result. The red line indicates a p value of 0.05, thus anything above this line has a p value <0.05. In the standard conditions, nine RNAs were significantly (p<0.05) increased or decreased compared to the WT control (Figure 4.2.1A). Of these, seven were mRNAs, and two were snoRNAs. Notably, a much larger change in differential expression was observed in the ice condition, where 118 RNAs were significantly decreased in their differential expression upon loss of snR4, of which 71 were mRNAs (Figure 4.2.1B).
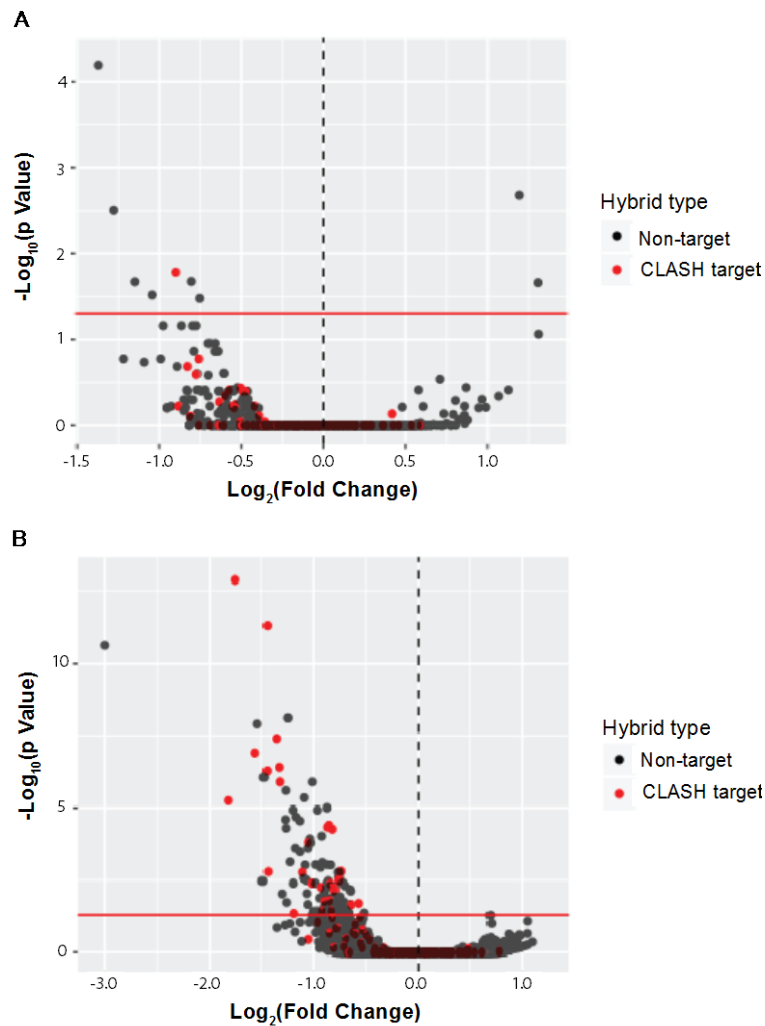
Figure 4.2.1. Volcano plots showing differential expression of RNAs upon *SNR4* deletion, using RiboMinus treated samples. A) Differential expression of total RNAs following growth in standard conditions in *snr4Δ* compared to WT. B) As in 'A', but harvested in the ice condition. X axis shows $\log_2$ of normalised RNA fold change. The dotted line at 0 signifies normalised WT RNA expression level. Negative values show a decrease in RNA expression, while positive values signify an increase in RNA expression, upon *SNR4* deletion. Y axis shows $-\log_{10}$(p value). Red line indicates p=0.05. Red dots indicate a direct target from CLASH data. Plots produced using R.

A list of CLASH targets for snR4 was created from hybrid analysis (CLASH experiments performed by T. Dudnakova). Comparison of the RNA sequencing analysis to this list indicated potential direct RNA targets (red dots on Figure 4.2.1). Under standard conditions, the only CLASH target that was significantly altered was another snoRNA: snR13. In the ice treated strains, 28 of the 118 significantly

reduced RNAs were snR4 targets identified by CLASH. These included seven mRNAs: *PDC1, TDH3, CCW12, FBA1, RPL28, RPS20,* and *RPL15A*. The *RPL15A* and *RPL15B* genes are highly homologous, making it unclear whether one or both mRNAs are targeted.

To generate a more comprehensive list of mRNA targets, RNA sequencing was performed following selection for poly(A)-tailed RNAs. The *snr4Δ* and *snr45Δ* strains were grown and harvested as above and the isolated RNA was treated with a poly(A)+ selection kit before library creation, as described in chapter 2.6. Figure 4.2.2 shows the distribution of differential RNA expression upon snoRNA deletion, compared to WT. The box and whisker plots show the upper and lower quartiles of the data, represented by the upper and lower boundaries of the boxes. The black line through the box represents the median, while the 'whiskers' represent the range of the data. Black dots signify individual outliers >1.5-fold of the upper quartile or <1.5-fold of the lower quartile. The red dotted line indicates the normalised expression level of all poly(A)-tailed RNAs in WT strain; therefore, a fold change (FC) greater than 1 indicates an increased differential RNA expression, while a fold change below 1 signifies a decrease in differential RNA expression, upon snoRNA deletion. All RNAs shown have a p value <0.05.

In Figure 4.2.2A, the median value for each condition was greater than 1, indicating that among poly(A)-tailed RNAs that are significantly altered, the majority increased in expression level upon deletion of *SNR4* or *SNR45*. To clarify the distribution of RNA abundance, a cut off of ±20% WT expression level was used, removing values within this range, and the data split by RNAs increasing in expression level and RNAs decreasing in expression level (Figure 4.2.2B). RNAs had a larger change in differential expression in *snr4Δ* compared to *snr45Δ* as there were a high number of outliers, meaning more snR4 targets lay outside 1.5-fold of the upper or lower quartiles. Additionally, more RNAs had an increased differential expression in standard conditions of *snr4Δ* than in ice condition.

To further stratify the distribution of RNA, a cut off of ±0.5 $\log_2(FC)$ of WT expression level was used (Figure 4.2.2C). At this level of significance, it was clear that the majority of RNAs in *snr4Δ* standard condition increased in expression level, while the majority of RNAs in *snr4Δ* ice condition decreased in expression level. Furthermore, the fold change of RNAs in *snr45Δ* ice condition was much higher than those in standard glucose condition.
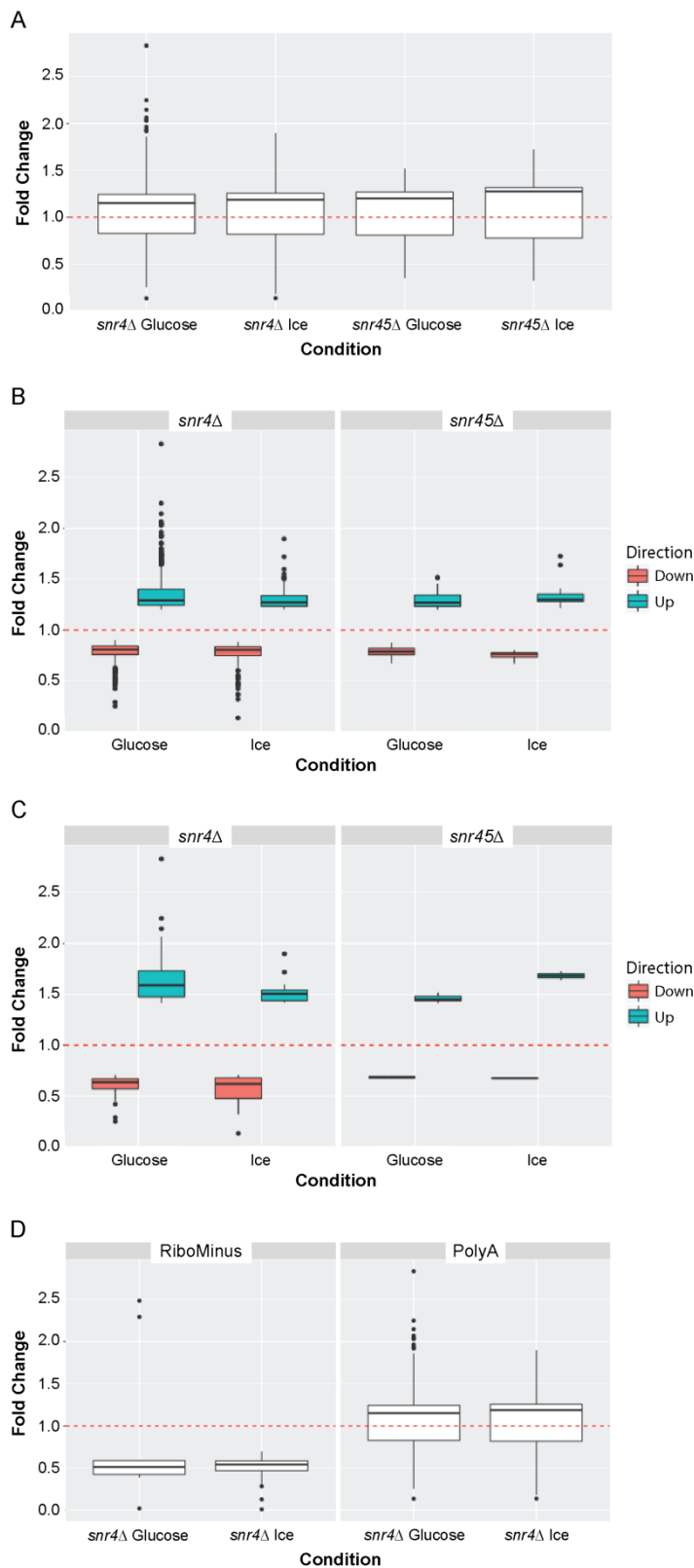
Figure 4.2.2 Box and whisker plots showing differential expression of RNAs upon snoRNA deletion, using poly(A)+ selected samples. A) Fold change of poly(A)-tailed RNAs following growth in standard and ice conditions in both

(Figure 4.2.2 cont.) *snr4Δ* and *snr45Δ*. **B)** As in 'A' but removing values ±20% WT expression, and showing separate box and whisker plots for values >1 and values <1. **C)** As in 'B', but with a cut-off of ±0.5 log$_2$(FC). **D)** RiboMinus treated-RNA sequencing of *snr4Δ* compared to poly(A)+ selected RNA sequencing of *snr4Δ* in both standard and ice conditions. The box represents Inter Quartile Range (IQR), the black line through the box represents the median, the 'whiskers' of the box represent the upper and lower range of the data, and the dots represent outliers >1.5-fold of the upper quartile or <1.5-fold of the lower quartile. Red dotted line signifies normalised WT RNA expression level. Below this line indicates a decrease in RNA expression, while above the line indicates an increase in RNA expression, upon snoRNA deletion. Plots produced using R.

Figure 4.2.2D shows the comparison between samples prepared for sequencing using the RiboMinus kit and poly(A)+ selection kit. In both standard and ice conditions for RiboMinus prepared sequencing, all but two of the RNAs decreased in differential expression compared to WT of the respective conditions. However, as demonstrated in Figure 4.2.2A, the majority of RNAs increased in differential expression compared to WT when selected for poly(A) tails. The mRNA-specific differential expression of the two RNA preparations was examined (data not shown). However, analysing the data to represent the same subsection of total RNA did not alter the result of the comparison. As the two sets of sequencing used different protocols, they cannot be directly compared. However, individual mRNAs that show similar differential expression in both sequencing preparations may indicate valid targets (discussed in chapter 4.4).

The overlap between different conditions was then analysed (Figure 4.2.3). 61% of the targets identified as significantly altered in *snr4Δ* ice were identified in *snr4Δ* standard conditions, showing substantial overlap (Figure 4.2.3A). Only 27% of *snr45Δ* ice targets were identified in *snr45Δ* standard conditions (Figure 4.2.3B). A high degree of overlap was shown between the *snr4Δ* and *snr45Δ* TSA arrays in chapter 3.3. Analysing the overlap between RNA sequencing data showed 87% of targets in *snr45Δ* standard conditions were identified in *snr4Δ* standard conditions (Figure 4.2.3C). 20% of *snr45Δ* ice targets were identified in *snr4Δ* ice conditions (Figure 4.2.3D). However, this representation differs when analysing the overlap in the reverse direction. While 61% of *snr4Δ* ice targets were found in *snr4Δ* standard

conditions, this comprises only 9% of targets in *snr4Δ* standard conditions. Similarly, 87% of *snr45Δ* standard condition targets were identified in *snr4Δ* standard conditions, but this comprises only 12% of *snr4Δ* standard condition targets. Notably, some targets appeared in both *snr4Δ* ice and *snr45Δ* standard, or *snr45Δ* ice and *snr4Δ* standard.
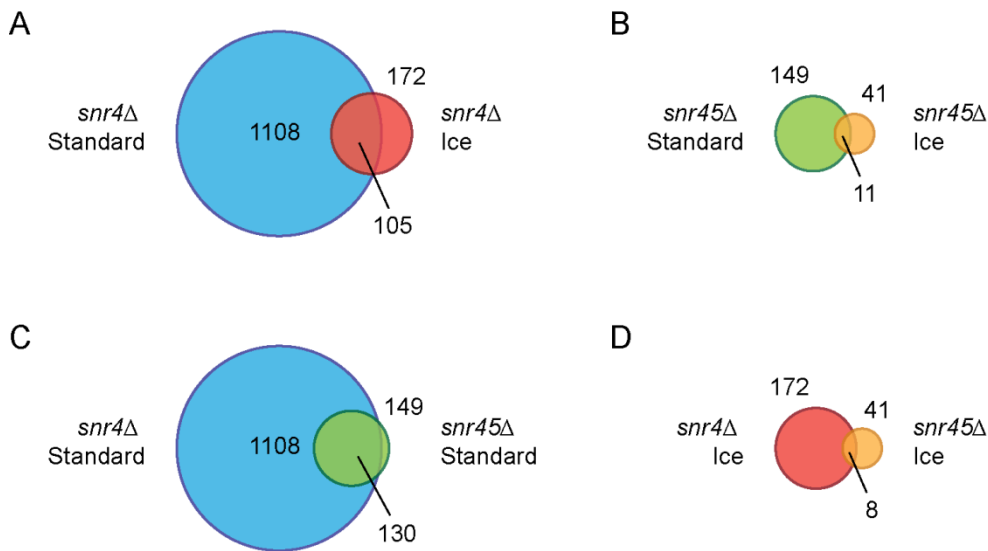
Figure 4.2.3 Venn diagrams showing overlap of RNA targets identified in different conditions upon snoRNA deletion, using poly(A)+ selected samples. A) Overlap between targets identified in *snr4Δ* standard conditions and those identified in *snr4Δ* ice conditions. B) As in 'A', but for *snr45Δ*. C) Overlap between targets identified in *snr4Δ* standard conditions and those identified in *snr45Δ* standard conditions. D) As in 'C', but for ice conditions. Colouring of circles indicates the condition. Number of targets identified for each condition is indicated either inside or above the circle. Number of targets overlapping the conditions is indicated below the circle, linked with a black line.

The targets that were identified in multiple conditions were further investigated. All but two RNAs in both *snr4Δ* standard and ice conditions decreased in expression in both conditions. RNAs in the ice condition identified in both *snr4Δ* and *snr45Δ* also all decreased in expression. RNAs overlapping *snr4Δ* and *snr45Δ* in standard conditions did not show a global direction of fold-change. However, each gene changed in the same direction in both conditions, as is true for those RNAs overlapping standard and ice conditions in *snr45Δ*. All *snr4Δ* ice targets identified in *snr45Δ* standard conditions decreased in expression in both conditions, excluding three. Targets identified in both *snr4Δ* standard and *snr45Δ* ice conditions showed

both increase and decrease in expression, but again each gene changed in the same direction in both conditions. To provide an indication of how snR4 and snR45 may both function, GO term analysis was performed on the overlapping targets found in *snr4Δ* and *snr45Δ* standard conditions. This showed enrichment for small molecule biosynthetic processes, including alpha-amino acid biosynthetic processes and arginine metabolic processes.

To better demonstrate the proportion of mRNAs with differential expression and highlight the number of CLASH targets, Figure 4.2.4 shows volcano plots of poly(A)+ selected RNA sequencing. Globally, in *snr4Δ* standard condition, mRNAs both increased and decreased in differential expression and did not show a clear directional change (Figure 4.2.4A). However, mRNAs in *snr4Δ* ice condition showed a clearer directionality, with the majority of mRNAs showing a decrease in expression level upon snoRNA deletion. This is comparable to the global trend of decreased mRNA expression level in Figure 4.2.1B.

RNAs in *snr45Δ* standard conditions showed a similar distribution as in *snr4Δ* standard conditions, with a higher number of mRNAs increasing in abundance than decreasing. It is noteworthy that the scale on the plot is smaller than for *snr4Δ,* with the x axis >2-fold smaller and the y axis approximately 4-fold smaller. This shows that in *snr45Δ,* fold changes were overall much smaller than for *snr4Δ,* and correspondingly less significant. This is further true for *snr45Δ* ice conditions, where the x axis is 2-fold smaller than *snr4Δ* ice conditions*,* and the y axis is 2.5-fold smaller than in *snr45Δ* standard condition. A higher number of mRNAs increased in expression level than decreased in *snr45Δ* ice condition.
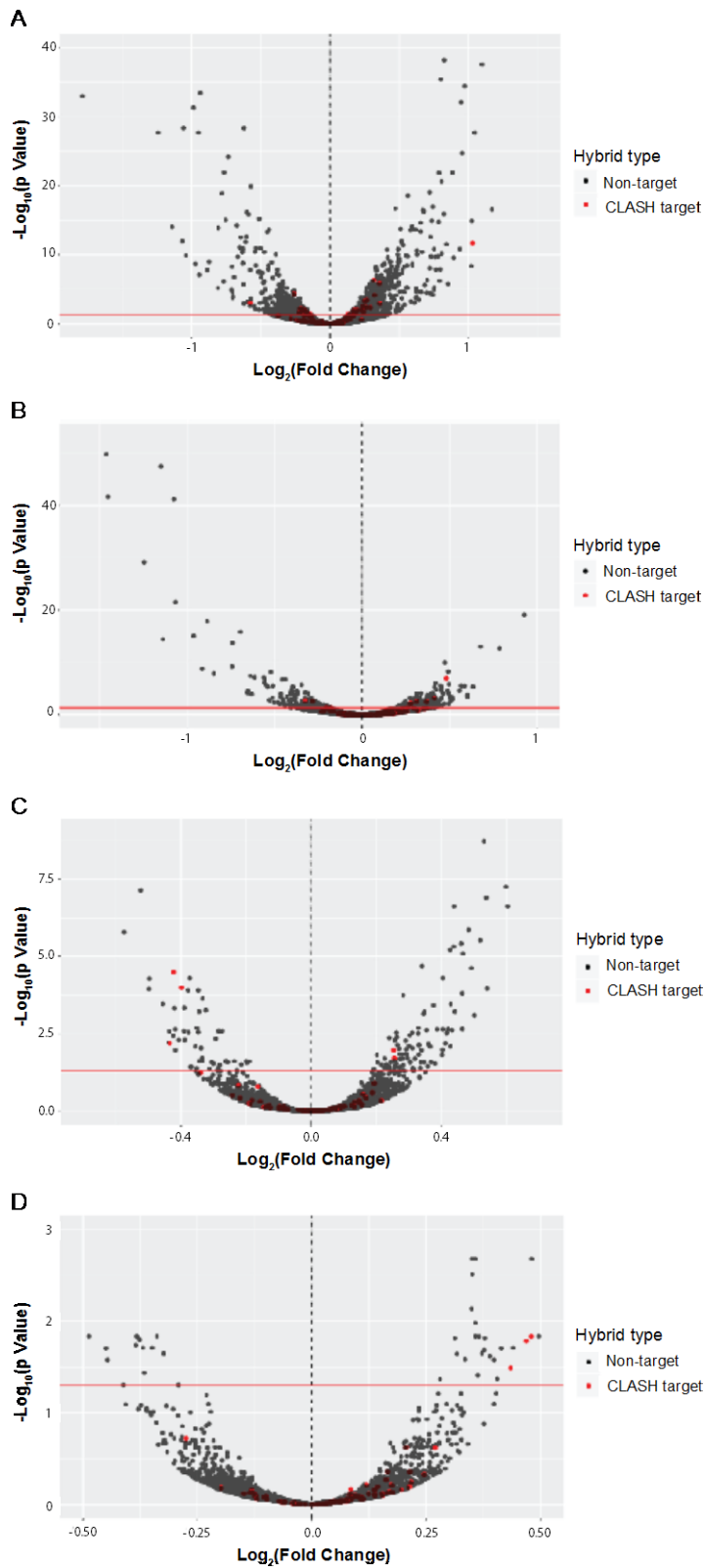
Figure 4.2.4 Volcano plots showing differential expression of RNAs upon snoRNA deletion, using poly(A)+ selected samples. A) Differential expression of poly(A)-tailed RNAs following growth in standard conditions in *snr4Δ*,

(Figure 4.2.4 cont.) compared to WT. B) As in 'A', but harvested in the ice condition. C) As in 'A' but for *snr45Δ*. D) As in 'C' but harvested in the ice condition. X axis shows log$_2$ of normalised RNA fold change. The dotted line at 0 signifies normalised WT RNA expression level. Negative values show a decrease in RNA expression, while positive values signify an increase in RNA expression, upon snoRNA deletion. Y axis shows −log$_{10}$(p value). Red line indicates p=0.05. Red dots indicate a direct target from CLASH data. Plots produced using R.

As with Figure 4.2.1, red dots indicate targets identified in CLASH. The CLASH target with the largest fold change in *snr4Δ* standard condition was *ALD6,* showing a fold change of 2. This was the largest fold change of a CLASH target within all four datasets. In comparison to typical effects seen on miRNA bindings, this fold change is not insubstantial. In total, 30 CLASH targets had significant differential expression in *snr4Δ* standard conditions, compared to WT. 11 CLASH targets showed a significant differential expression in *snr4Δ* ice conditions. Five CLASH targets showed significant differential expression in *snr45Δ* standard conditions and three in ice conditions. Although the poly(A)+ selection kit was used as recommended by the manufacturer, a number of non-poly(A) tailed RNAs appear in the datasets. These included ncRNA CLASH targets that apparently changed significantly. Some caution may therefore be needed in interpreting the results. In all four analyses, the mRNAs with the highest fold changes were not direct targets found by CLASH. Furthermore, a large number of CLASH hits were found to have fold changes that did not meet the threshold for significance (p ≤0.05). This indicates that snR4 and snR45 may not primarily affect mRNA synthesis or stability, but might alter translation instead.

A

snr4Δ Standard	CLASH+	CLASH-	Total
Changed	30	1079	1109
	15	**1094**	
Unchanged	97	8212	8309
	112	**8197**	
Total	127	9291	9418
p Value	1.14x10^{-4}		

B

snr4Δ Ice	CLASH+	CLASH-	Total
Changed	11	387	398
	5.4	**392.6**	
Unchanged	116	8923	9039
	121.6	**8917.4**	
Total	127	9310	9437
p Value	2.22x10^{-2}		

C

snr45Δ Standard	CLASH+	CLASH-	Total
Changed	5	145	150
	1.7	**148.3**	
Unchanged	102	9162	9264
	105.3	**9158.7**	
Total	107	9307	9414
p Value	2.97x10^{-2}		

D

snr45Δ Ice	CLASH+	CLASH-	Total
Changed	3	39	42
	0.5	**41.5**	
Unchanged	104	9319	9423
	106.5	**9316.5**	
Total	107	9358	9465
p Value	1.17x10^{-2}		

Table 4.2.1 Fisher's exact test of poly(A)+ selected RNA sequencing vs CLASH. A) Table showing the changed and unchanged RNAs observed in snr4Δ poly(A)+ selected RNA sequencing vs snR4 CLASH targets (CLASH+) or non-CLASH targets (CLASH-) in standard conditions. B) As in 'A' but for ice conditions. C) As in 'A' but for snR45 CLASH targets in snr45Δ. D) As in 'C' but for ice conditions. Numbers in red show the expected values based on the proportions of total RNAs identified in each category. Fisher's exact p value is shown in green.

To gauge further the statistical significance of these results, a Fisher's exact test was performed (Table 4.2.1). This test analyses the number of targets observed in each condition experimentally, compared to the number of targets expected for each condition if the conditions are not related. The p value generated from the test signifies whether the observed number of targets is significantly more or less than the expected number, thus accepting or rejecting the null hypothesis that the conditions are not related. Table 4.2.1 shows the number of CLASH targets that changed in RNA sequencing (CLASH+, Changed), versus the number of CLASH targets that did not change (CLASH+, Unchanged), compared to the number of RNAs that changed in RNA sequencing that were not CLASH targets (CLASH-, Changed), versus the number of RNAs that did not change that were not CLASH targets (CLASH-, Unchanged). The top left condition (CLASH+, Changed) for each table is the condition of interest, as this condition identifies potential direct RNA

targets with differential expression. Expected values (numbers in red) are determined by calculating the ratio between row totals and absolute total, then applying this to the column totals. Thus, to calculate the number of CLASH targets with differential expression expected if there were no relationship between the conditions, the ratio between Changed total and absolute total would be applied to the CLASH+ total, to give the expected number of 'CLASH+, Changed' targets. Applying the Fisher's Exact Test function in R to the tables generates a p value that determines if the number of CLASH targets with differential expression (CLASH+, Changed) was higher or lower significantly in the observed results compared to the expected.

In all mutant growth conditions (*snr4Δ* standard and ice, and *snr45Δ* standard and ice) the observed number of CLASH targets that changed in expression level was statistically significant compared to the expected number. A higher number was observed than the number expected from this set of data if there was no correlation. However, as you apply different cut-offs to these values, the significant observations decreased. At a cut-off of 0.2 $\log_2$(FC), only *snr4Δ* standard condition targets had a p value lower than the 0.05 threshold for significance, but at a cut-off of 0.3 $\log_2$(FC), only *snr45Δ* ice had a significant p value of 0.011. At a cut-off of 0.4 $\log_2$(FC), none of the conditions showed a statistically significant enrichment for CLASH targets.

The targets from poly(A)+ selected RNA sequencing were then compared to the SGA screen targets in chapter 3.3. Table 4.2.2 shows the common targets between the SGA screen and RNA sequencing for both essential (TSA) and non-essential genes (DMA). When paired with deletion of the snoRNA, the genes that showed decreased colony fitness upon mutation compared to single mutation alone are listed in blue. Those that increased in colony fitness in the double mutant are shown in orange. Beside each gene in the table is the fitness score determined by the SGA screen, and the mRNA fold change recorded in poly(A)+ selected RNA sequencing. The same fold change of each gene was listed if multiple alleles of the same gene were present in the SGA screen. A cut-off of ±0.25 was used for SGA score in table 4.2.3, as this is the value at which the majority of growth defects are visible by eye.

A

snr4Δ -TSA				snr4Δ -DMA		
Gene	**Allele**	**SGA score**	**Fold change**	**Gene**	**SGA score**	**Fold change**
NOP1	*nop1-2*	-0.956	1.153	*DSE1*	-0.502	0.859
RSP5	*rsp5-3*	-0.618	1.275	*AVT6*	-0.501	1.182
ORC6	*orc6-ph*	-0.519	0.838	*COX10*	-0.278	1.240
DED1	*ded1-f144c*	-0.480	1.480	*PTC2*	-0.276	1.173
GLC7	*glc7-12*	-0.446	0.879	*RPL23B*	-0.275	0.856
ACT1	*act1-2*	-0.410	1.127	*KNS1*	0.320	1.592
DED1	*ded1-95*	0.530	1.480			
NOP2	*nop2-3*	0.508	1.248			
HRP1	*hrp1-1*	0.358	1.315			
STT3	*stt3-1*	0.326	1.130			

B

snr45Δ -TSA				snr45Δ -DMA		
Gene	**Allele**	**SGA score**	**Fold change**	**Gene**	**SGA score**	**Fold change**
DED1	*ded1-f144c*	-0.508	1.325	*ASN1*	-0.885	0.747
RSP5	*rsp5-sm1*	-0.375	1.216	*TDA6*	-0.267	0.763
DED1	*ded1-95*	0.513	1.325	*KNS1*	0.260	1.353
HRP1	*hrp1-1*	0.350	1.193			

Table 4.2.2 Table showing gene overlap between targets identified in poly(A)+ selected RNA sequencing and targets identified in the SGA screen (chapter 3.3). A) Target overlap for genes affected in both *snr4Δ* RNA sequencing, and *snr4Δ* TSA and DMA screens. B) As in 'A', but for *snr45Δ*. Colour of the gene indicates rate of growth of the double mutant compared to the single mutant in the SGA screen. Blue indicates decreased colony fitness of the double mutant compared to single mutant, orange indicates increased colony fitness of the double mutant. The SGA score and fold change from RNA sequencing for each gene is listed.

10 TSA targets changed significantly in both RNA sequencing and the SGA screen upon *SNR4* deletion (Table 4.2.2A). Six of these showed synthetic negative interactions, while four showed suppression. *NOP1, ACT1* and *STT3* all showed a small increase in RNA abundance. The other seven genes, however, showed a larger change in differential expression, with the highest change in expression reported for *DED1*. Notably, the direction of mRNA fold change of *ORC6* and *GLC7* matched the direction of colony fitness, as did that of all the suppression mutations. *DED1* showed non-hierarchical heterogeneity, meaning the mutation in one allele caused one effect on its growth (*ded1-f144C* resulted in synthetic sickness) but the

mutation in a different allele had an opposite effect (*ded1-95* resulted in suppression). The effect of snoRNA deletion on *DED1* mRNA accumulation appeared to be ice-specific. Its mRNA showed differential expression upon both *SNR4* deletion and *SNR45* deletion, with its fold change higher in *snr4Δ* (ice) than *snr45Δ* (ice). All four genes that overlapped between the *snr45Δ* TSA screen and RNA sequencing were identified in the overlap for *snr4Δ* (Table 4.2.2B). Both *ded1* alleles showed similar magnitude of effect on growth in both directions, and for both snoRNA strains. *hrp1* double mutants also showed a similar magnitude of fitness with *snr45Δ* as with *snr4Δ*, though its fold change in RNA abundance was minimal. Interestingly, *HRP1* was identified as an snR45 CLASH target, but not as an snR4 CLASH target; however, its fold change was smaller in the *snr45Δ* strain. While both *SNR4* and *SNR45* deletion had a similar effect on *RSP5* mRNA, the *snr4Δ* double mutant had almost twice the growth impairment score of the *snr45Δ* double mutant.

Analysis of the overlap between DMA targets and RNA sequencing targets revealed six *snr4Δ* targets (Table 4.2.2A). *AVT6* and *PTC2* showed minimal increase in fold change. *DSE1, RPL23B* and *KNS1* all showed a direction of colony fitness matching their mRNA fold change. *KNS1* in *snr4Δ* also had the highest fold change throughout all the targets, though its SGA score was relatively mild. Analysis of DMA targets for *snr45Δ* revealed three targets with altered mRNA abundance. *KNS1* is a target of both snoRNAs, though its SGA score and fold change were of lower magnitude under *SNR45* deletion than *SNR4* deletion. *ASN1* and *TDA6* showed no overlap, and the direction of colony fitness matched the direction of mRNA fold change. Furthermore, *ASN1* is an snR45 CLASH target. Deletion of *ASN1* had a strong synthetic negative growth phenotype when paired with *SNR45* deletion, and the fold change of its mRNA was appreciable. This target will therefore be interesting for further study.

The majority of overlap in targets was seen in *snr4Δ* standard conditions for TSA genes. This may be related to the fact that the largest number of RNAs showing differential expression was also found in this condition. Few of the SGA targets overlapped with CLASH hits. None of the *snr4Δ* targets that were identified in both the SGA screen and CLASH target list showed significant changes in RNA sequencing. However, the SGA screen identifies pathways in which synergistic interactions have clear effects on cell fitness. This indicates that the SGA screen highlights the pathways that are altered by snoRNA deletion, but does not

necessarily indicate specific gene targets of the snoRNAs. Therefore, while overlap between RNA sequencing and SGA targets may indicate specific targets of snR4 and snR45, a high correlation may not be observed.

4.3: CLASH analysis of snoRNA-mRNA interactions

To better understand the interactions between snoRNAs and mRNAs, CLASH sequences were used to map exact binding sites within the RNAs. The snR4 and snR45 sequences of each hybrid were mapped onto the full *SNR4* or *SNR45* sequence, respectively, to visualise regions of the snoRNA involved in mRNA binding (Figure 4.3.1). The x axis shows the nucleotide position over the gene, and the y axis represents the number of independent CLASH hybrids identified for each nucleotide. The different colours indicate where the snoRNA boxes are positioned. As snR4 has no clear D' box consensus sequence, no D' box is indicated, however it would be expected to be positioned around +100.

The three main peaks of mRNA binding on *SNR4* were centred on +40 (at the C box), +90 (between the C box and C' box), +150 (adjacent to the C' box) and +190 (adjacent to the D box) (Figure 4.3.1A). The peak in *SNR45* was centred on +130, adjacent to the C' box (Figure 4.3.1B). The sequences located between the C' and D boxes, and between the C and D' boxes, are termed the guide sequences. This is where snoRNAs canonically interact with ribosomal RNA, guiding Nop1 in the methyltransferase reaction. As snR45 does not form hybrids directly before the D box, this indicates that it is unlikely to guide methylation of target mRNAs. The boxes are necessary for snoRNA structure and stability. Therefore, a low number of hybrids over the boxes would show that these are still necessary for structure in the mRNA interaction. snR4 has a large number of hybrids over the C' and D boxes. However, as canonical box C/D proteins would protect the regions of bound RNA from RNases, during the CLASH protocol, fragments longer than the cross-linked site would be recovered. Therefore, these may be the ends of the sequences cross-linked next to the boxes, rather than hybrids binding to the boxes. It may also indicate that the C' box in snR4 is not necessary for structure during RNA interactions, especially given the lack of a reliable D' box. Comparison of the distribution of mRNA binding between the two snoRNAs shows that while snR45 appeared to have one main binding site, snR4 binding was at multiple sites on the

gene. The number of hybrids identified for snR4 at each site was also lower than for snR45, possibly because hybrids are more highly distributed across the gene.

A

B
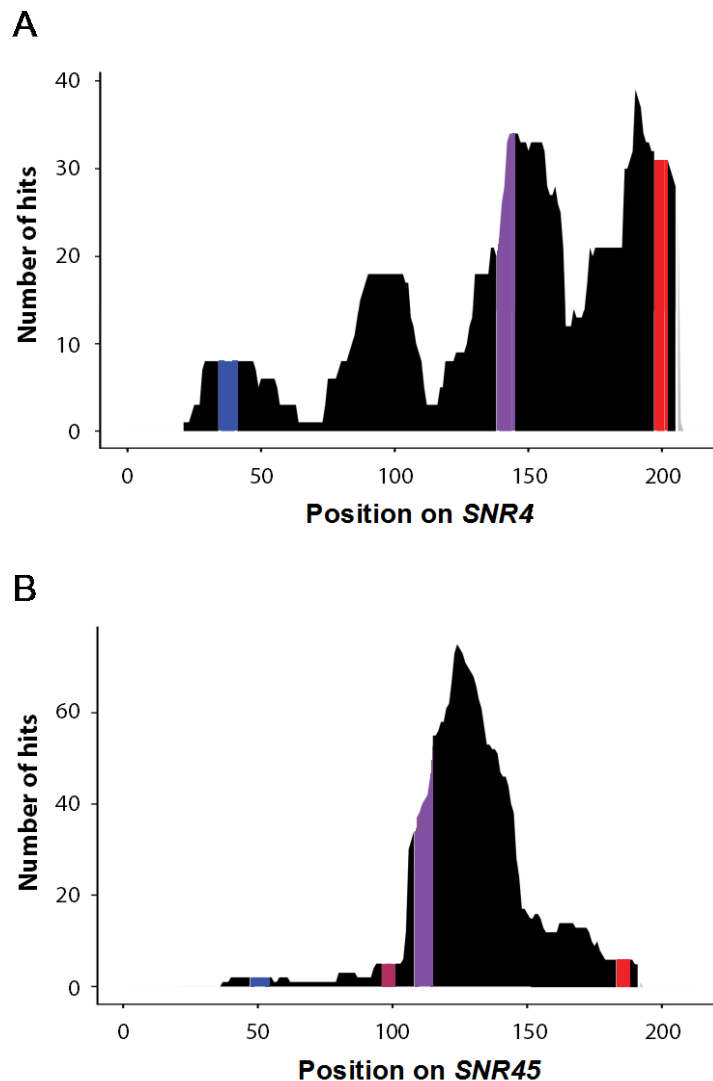
Figure 4.3.1 Pile-up of snoRNA-mRNA hybrid hits on snoRNA. A) The snR4 sequence obtained from snR4-mRNA hybrids mapped onto *SNR4*. B) The snR45 sequence obtained from snR45-mRNA hybrids mapped onto *SNR45*. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Hybrid data obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Figure 4.3.1 showed the snoRNA binding sites for all mRNA hybrids found in CLASH. However, comparing the CLASH data to the RNA sequencing data shows

the binding sites for mRNAs that change in differential expression. This may result in more precise identification of binding sites. The genes were grouped based on where they hybridised to the snoRNA, then the base-paired nucleotides were highlighted on the snoRNA sequence (green bar, Figure 4.3.2). The pile-up of all CLASH hybrids is also shown across the snoRNA sequence for comparison. As shown in Figure 4.3.1A, mRNAs did not bind to one distinct site on *SNR4*, but rather at a number of sites across the snoRNA (Figure 4.3.2A). These coincided with the sites displayed in Figure 4.3.1A, with Figure 4.3.2A showing the exact nucleotides involved. A number of hybrids base-paired directly with the D box and 3' end indicating that they are unlikely to be valid targets, as the 3' end of the snoRNA must be free to base-pair to the 5' end for structural stability. Given the short length of the 3' hybrid sequences, this is unlikely to be due to trailing ends of valid cross-linked targets, as discussed above. The mRNAs within groups did not show an overall trend in differential expression, meaning that individual binding sites did not correlate with a specific change in mRNA expression. However, it is noteworthy that all ribosomal protein genes (RPGs) within all groups on *SNR4* showed a decrease in expression level. No RPGs showed differential expression in *snr45Δ. ALD6* was the only mRNA from this dataset to bind at the 5' end of the snoRNA, but was included due to its high fold change. However, this was base-paired directly to the C box, which may also be an artefact, as the 5' end must be free to interact with the 3' end for structural stability.
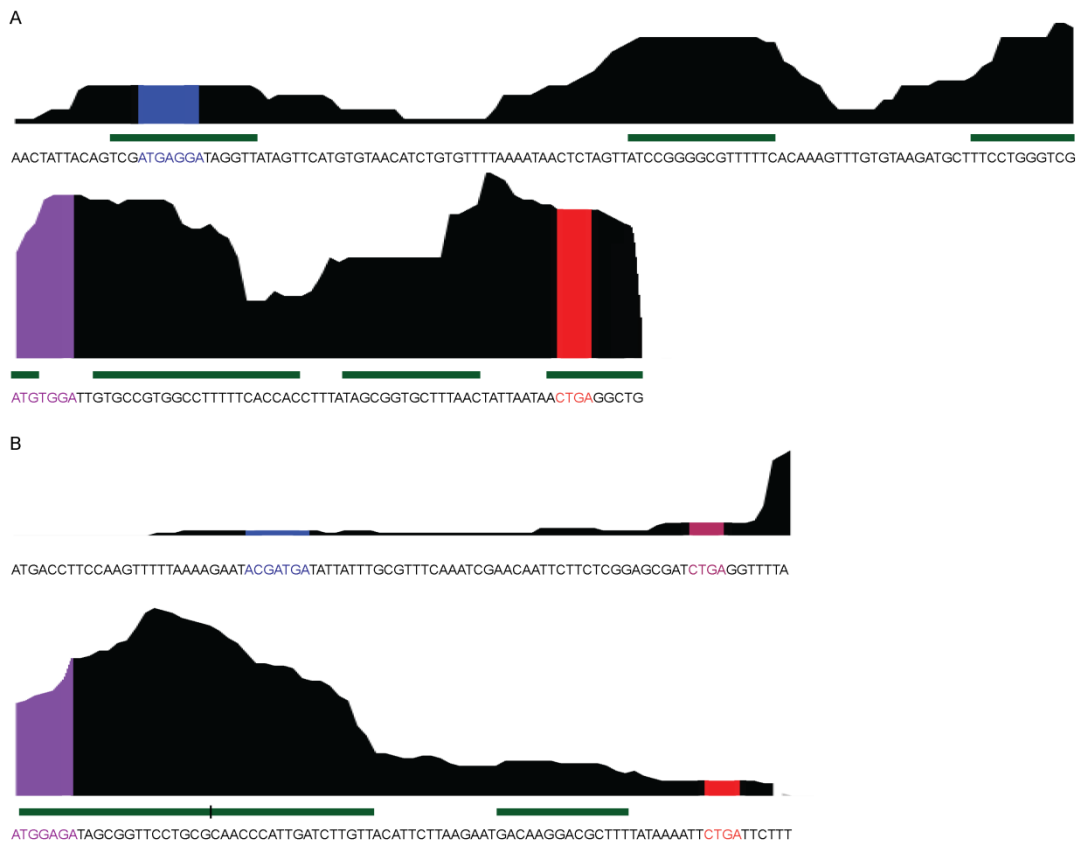
Figure 4.3.2 Distribution of differentially expressed hybrids across snoRNA sequence. A) snR4-mRNA hybrids from Figure 4.3.1A aligned to the *SNR4* sequence. B) snR45-rRNA hybrids from Figure 4.3.1B aligned to the *SNR45* sequence. Green lines indicate which nucleotides are directly base-paired with the differentially expressed hybrids. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Hybrid data obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

The nucleotides involved in base-pairing to mRNAs from *snr45Δ* sequencing data coincided with the hybrid peak shown in Figure 4.3.1B (Figure 4.3.2B). However, only five mRNA CLASH targets were identified in the RNA sequencing data. The binding site in Figure 4.3.2B was split into two sites (denoted by a black vertical line) – the first half bound *ASN1* and *ARG4*, and the second half bound *TPS2* and *COX1*. *HRP1* was bound upstream of the D box. Notably, *ASN1* and *ARG4* both decreased in expression level upon *SNR45* deletion, to a similar magnitude. However, *TPS2* and *COX1* both increased in expression level, but to differing magnitudes. It is possible that the two sites, while adjacent, have differing effects upon target

mRNAs. This could only be investigated with more CLASH hits identified in the sequencing data. Furthermore, both *ASN1* and *ARG4* were base-paired to the C' box. If snR45 is structurally similar to canonical snoRNAs, the region around the C' and D' boxes would be base-paired for structural stability. However, some snoRNAs lack C' and D' boxes. It is possible that snR4 and snR45 do not need the C'-D' box interaction for stability, thus allowing these sites to interact with other RNAs. Table 4.2.1 shows that there were approximately 10-fold fewer mRNAs with differential expression in *snr45Δ* strains than in *snr4Δ* strains. This explains why so few targets were able to be mapped onto the snoRNA, and supports the theory that snR45 binding to mRNAs may not affect expression level.

4.4: Validation of non-canonical snoRNA targets

Six genes were both CLASH targets and showed significant differential expression in the RiboMinus-treated sequencing data (chapter 4.2, Figure 4.2.1). These interactions were experimentally tested. *snr4Δ* and WT strains were grown and harvested in both standard and ice conditions, as previously described. The RNA was isolated and reverse-transcribed using random decamer primers, and quantified by qPCR. For each sample, three or more biological replicates were analysed, with three technical replicates each. The $2^{-\Delta\Delta Ct}$ (**Livak and Schmittgen, 2001**) for each experiment was determined by calculating the difference between the C_t (threshold cycle) of the test gene and the housekeeping gene (*SCR1*) in the deletion strain, then subtracting the value of the difference between the C_t of the test gene and the housekeeping gene in the WT strain, to give the $\Delta\Delta C_t$ value. The exponential of $-\Delta\Delta C_t$ gives the relative expression of the test gene in the deletion strain. The average of $2^{-\Delta\Delta Ct}$ for replicates of each gene was taken to determine fold change. Figure 4.4.1 shows the relative fold change for each gene compared to WT along the y axis, for both standard and ice conditions. Error bars indicate standard deviation of the replicates, and the red dashed line denotes a fold change of 1, showing the WT normalised expression of each gene. A non-target gene, *ALG9,* is also shown for reference. A one-sample t-test was performed on the $2^{-\Delta\Delta Ct}$ replicates to determine statistical significance. Fold changes and p values for each gene are shown in Table 4.4.1.

PDC1 showed a similar qPCR fold change in both standard and ice conditions as was reported in the RiboMinus-treated RNA sequencing (Table 4.4.1). Furthermore, the 34% reduction in *PDC1* expression in the ice condition was confirmed as statistically significant. *TDH3, CCW12* and *FBA1* all showed the same direction of change as was observed in RNA sequencing, but with a lower magnitude. Fold changes of *CCW12* in both conditions were statistically significant, as was *TDH3* in the ice condition. *CCW12* and *FBA1* showed a greater fold change in ice condition than in standard conditions, though the changes in *FBA1* were not statistically significant in either condition. *RPL28* showed little change in both standard and ice conditions, in contrast to the results of RNA sequencing. This may be due to differences that can be observed for results obtained through high throughput screens compared to small scale experiments. While *RPS20* showed limited fold change under standard conditions, there was a 22% reduction in expression in the ice condition which was statistically significant.
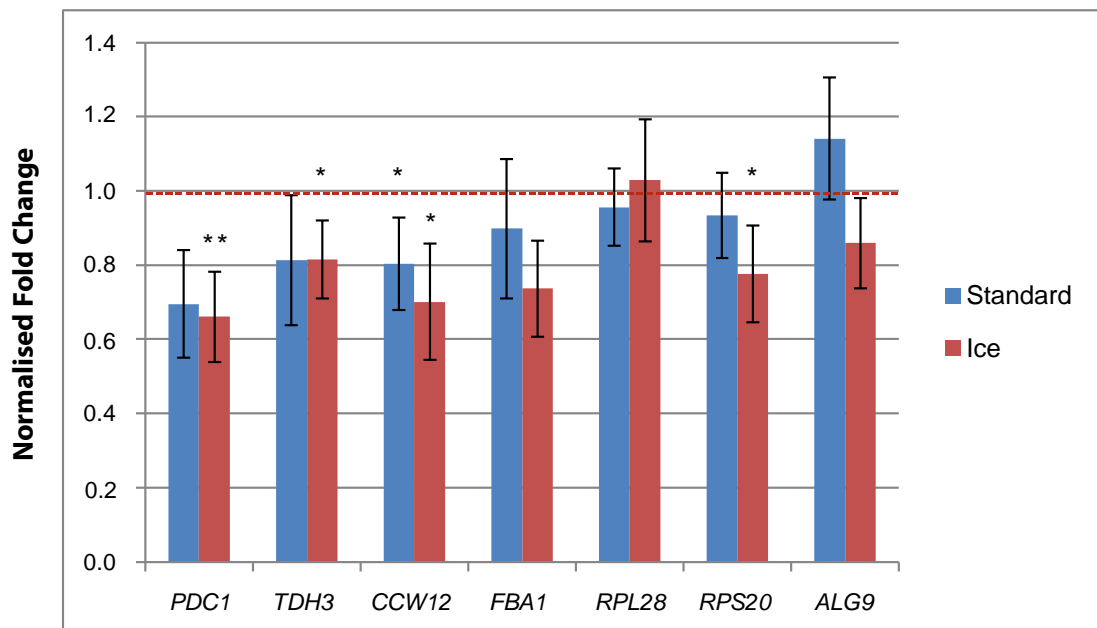
Figure 4.4.1 RT-qPCRs showing fold change of snR4 CLASH targets following growth in standard or ice conditions. *ALG9* **is not a target, and shown for comparison. All samples were normalised to** *SCR1,* **then to WT gene expression level. Red dotted line denotes relative WT expression level. RT performed using random decamer primers on total RNA, and based on RiboMinus-treated sequencing data. One asterisk denotes p<0.05, two asterisks denote p<0.01.**

Gene name	FC: Sequencing	FC: qPCR - Standard	p value: Standard	FC: qPCR - Ice	p value: Ice
PDC1	0.672	0.69	0.068	0.66	0.003
TDH3	0.555	0.81	0.12	0.82	0.017
CCW12	0.592	0.80	0.05	0.70	0.013
FBA1	0.568	0.90	0.29	0.74	0.072
RPL28	0.557	0.96	0.46	1.0	0.75
RPS20	0.575	0.93	0.27	0.78	0.041
ALG9	1.29	1.1	0.28	0.86	0.10

Table 4.4.1 Comparison between RiboMinus-treated sequencing values and RT-qPCR values. The average FC of the three biological replicates from RiboMinus-treated RNA sequencing is listed, followed by the FC and p value obtained by three or more replicates of RT-qPCR following growth in standard conditions, followed by FC and p value obtained by three or more replicates of RT-qPCR following growth in ice conditions. qPCR FC and p values shown to two significant figures.

Taking into account the positions of the hybrids on *SNR4, PDC1* and *CCW12* form hybrids directly with, and surrounding, the D box. This indicates that they are unlikely to be valid targets of snR4, given that the D box is critical for the structure of the snoRNA, and thus cannot be base-paired to a target. Moreover, *PDC1, TDH3* and *CCW12* are frequently found in other experiments, and may be non-specific hits. *ALG9* showed differential expression under both standard and ice conditions, indicating that it may be affected downstream by targets of snR4. *RPS20* therefore appears to be the only high confidence target.

To further check the validity of these targets, the RiboMinus-treated RNA sequencing data was compared to the poly(A)+ selected sequencing data. CLASH targets that showed differential expression in the same direction in both datasets were more likely to be valid. Table 4.4.2 shows the fold changes for each target in the RiboMinus-treated sequencing data and the poly(A)+ selected sequencing data. The observed changes of *PDC1, TDH3* and *CCW12* in the poly(A)+ selected sequencing data demonstrate that these genes are not valid targets of snR4, as they are not subject to consistent change. Neither *FBA1* nor *RPL28* showed significant differential expression in the poly(A)+ selected sequencing data.

However, *RPS20* showed comparable changes in both sets of data, further supporting its potential as a direct target.

Gene	FC: RiboMinus-treated RNA Sequencing	FC: poly(A)+ selected RNA Sequencing
PDC1	0.672	1.25
TDH3	0.555	1.09
CCW12	0.592	1.20
FBA1	0.568	-
RPL28	0.557	-
RPS20	0.575	0.874

Table 4.4.2 Comparison between CLASH targets in RiboMinus-treated sequencing versus poly(A)+ selected sequencing. Column 2 lists the average FC of the three biological replicates from RiboMinus-treated RNA sequencing, while column 3 lists the average FC of the three biological replicates from poly(A)+ selected RNA sequencing.

As only one target was potentially valid from the overlap, the search criteria were widened. 30 CLASH targets showed significant differential expression in *snr4Δ* standard condition in poly(A)+ selected sequencing, and five CLASH targets showed significant differential expression in *snr45Δ* standard condition, compared to WT. Therefore, more genes were able to be investigated. To determine which snR4 CLASH targets were most likely to be valid, the list of 30 differentially expressed CLASH targets from poly(A)+ selected sequencing data was compared to the RiboMinus data, both significant ($p < 0.05$) and not significant ($p > 0.05$). Any gene that appeared in both the list of 30 and the RiboMinus-treated sequencing data that also showed a fold change in the same direction was chosen. Non-coding RNAs and RNAs with a paralogue were excluded: ncRNAs because they should not appear in poly(A)+ selected datasets, and paralogues because they prove more difficult to study by qPCR. This left four genes: *ALD6, RPS20, TMA7* and *SES1. SES1* however had a predicted fold change of only 0.9, so was excluded. *RCK1* is not a CLASH target, but had the highest fold change from poly(A)+ selected RNA sequencing, so was also included. As RiboMinus-treated sequencing was not performed for *snr45Δ,* a cut-off of ±20% FC was used. snR45 had only two target

genes with a fold change greater than ±20%, after excluding ncRNAs: *ASN1* and *ARG4*.

snr4Δ, snr45Δ and WT strains were grown and harvested in standard conditions as previously described, and the RNA was isolated. This was reverse-transcribed using random decamer primers, and the genes were analysed by qPCR. However, this did not give consistent results. Therefore, the RNA was reverse-transcribed using oligo (dT) primers, to replicate the poly(A)+ selection process used in the sequencing. As before, samples were analysed in three biological replicates, which each had three technical replicates. Given that *SCR1* has no poly(A) tail, this could not be used for normalisation. *ACT1* cannot be used as it is an snR4 CLASH target, and is unreliable for normalisation (**Teste et al., 2009**). Therefore, a subunit of the TFIID complex, *TAF10,* was used, as it is a reliable reference gene. In addition, *Schizosaccharomyces pombe* RNA was added at a concentration 10-fold lower than the isolated RNA for normalisation purposes, and *ACT1* was quantified from this. Each technical replicate for both the gene of interest and *TAF10* was normalised to the corresponding replicate of *S. pombe ACT1,* to account for experimental variance between samples*.* The gene of interest was then normalised to *TAF10,* and the average for each biological replicate taken. These were then normalised to the average of each WT biological replicate, and a two-tailed homoscedastic t-test applied (two samples, equal variance).

Figure 4.4.2A shows the relative fold changes for the three snR4 CLASH targets and *RCK1* (a non-CLASH target) along the y axis*,* compared to WT*.* The red dashed line indicates a fold change of 1, showing the WT normalised expression for each gene. In standard conditions upon *SNR4* deletion, *ALD6* had a fold change of 1.54. This matched the direction of the fold change of 2.04 shown in poly(A)+ selected RNA sequencing. *RCK1* showed a fold change of 2.4, which was similar to the 2.83 fold change observed in RNA sequencing. Both *ALD6* and *RCK1* had statistically significant differential expression upon *SNR4* deletion, with p values of 0.019 and 0.002, respectively. *RPS20* showed a fold change of 0.83 by qPCR, which is comparable to its fold change of 0.874 from RNA sequencing. However, its p value was 0.091, which is not significant. *TMA7* showed a fold change of 0.860 in RNA sequencing, but had a fold change of 0.96 in qPCRs, with a p value of 0.75.

A
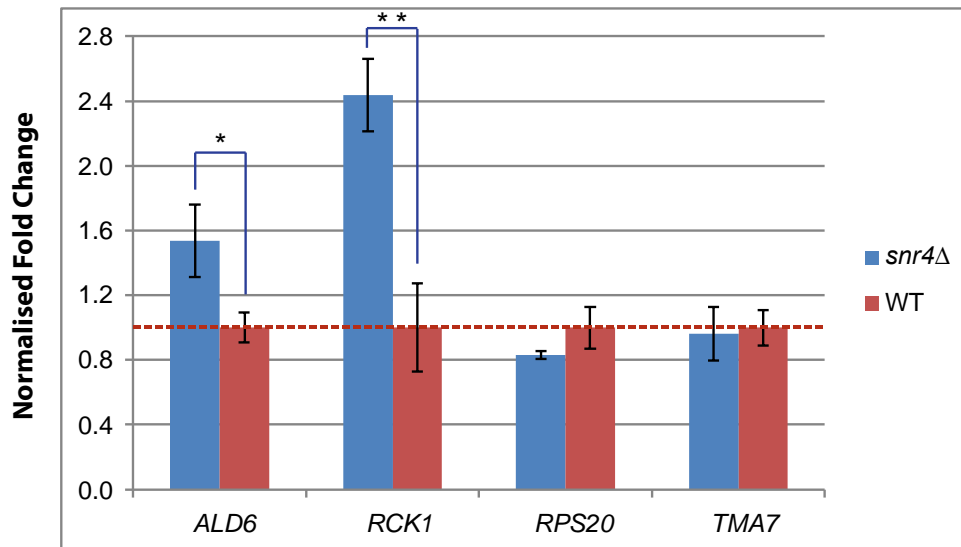
B
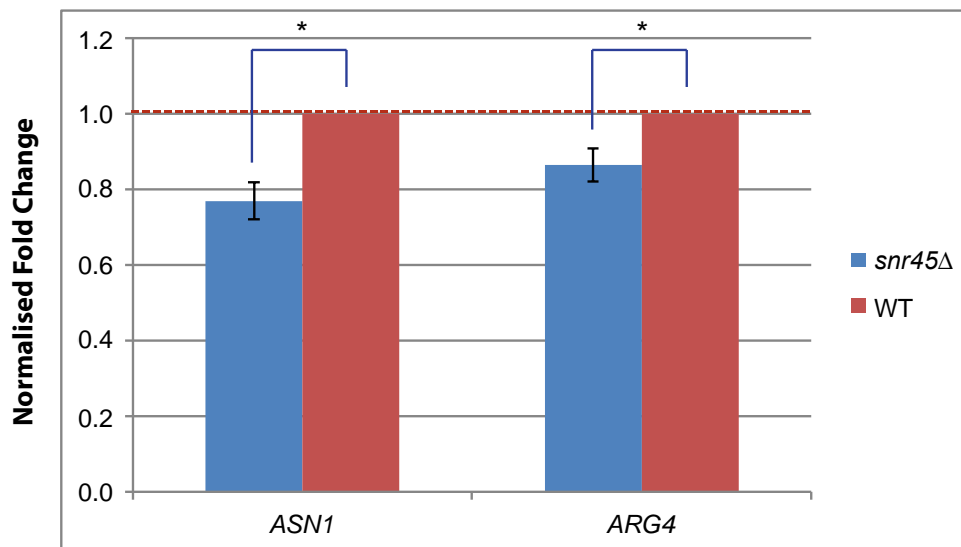
Figure 4.4.2 RT-qPCRs showing fold change of CLASH targets following growth in standard conditions. A) FC of snR4 targets in the _snr4Δ_ strain compared to WT strain. All genes apart from _RCK1_ are CLASH targets. B) FC of snR45 CLASH targets in the _snr45Δ_ strain. All samples were normalised to _S. pombe ACT1_, then to _TAF10_, then to WT gene expression level. Red dotted line denotes relative WT expression level. RT was performed using oligo (dT) primers. One asterisk denotes p<0.05, two asterisks denote p<0.01.

As the RNA extracted for one sample in the _snr45Δ_ analysis was below the threshold needed to continue with qPCR validation, two further replicates were

performed. The $2^{-\Delta\Delta Ct}$ for each replicate was calculated, using *TAF10* as the reference gene, and the average taken to determine average fold change. *ASN1* showed a fold change of 0.77, giving a consistent drop in expression of 23%. This was similar to its fold change of 0.747 shown by RNA sequencing. A one-sample t-test showed this was statistically significant, with a p value of 0.015. *ARG4* showed a fold change of 0.86, with a consistent drop in expression of 14%. This also matched the direction shown by RNA sequencing, which gave a fold change of 0.759. A one-sample t-test showed this was also statistically significant, with a p value of 0.034.

As the targets with highest differential expression that were also direct CLASH targets were now validated, *ALD6* and *ASN1* were chosen for further analysis. Tagging strains with GFP allows for visualisation of protein expression levels. This could validate the interactions observed with both *ALD6* and *ASN1* at the protein level. Strains were created tagging both proteins with a GFP tag, with kanamycin resistance marker for selection purposes, in the respective *snr4Δ* and *snr45Δ* strains. This was also performed in the WT strain. *ACO1* and *RNR2* were chosen as normalisation proteins and also tagged, as these showed a fold change close to 1 in RNA sequencing and had similar levels of expression to *ALD6* and *ASN1*, respectively. However, due to time constrains, further experiments were not possible.

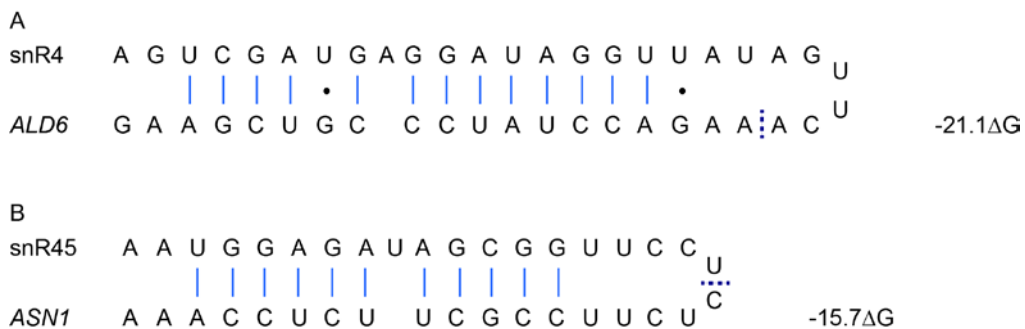**Figure 4.4.3 snoRNA-mRNA hybrids from CLASH showing target interactions. A) snR4-*ALD6* hybrid. B) snR45-*ASN1* hybrid. Watson-Crick base-pairs denoted by a blue straight line, G-U base-pairings denoted by a dot. Dashed purple line denotes the boundary between snoRNA and mRNA regions of the hybrid. Hybrid drawn from full chimeric sequence from CLASH, and base-pairing predicted by the ViennaD folding programme.**

To further investigate this interaction, the hybrids were identified from CLASH data, (Figure 4.4.3). Base-pairing indicates the nucleotides involved in the interaction. Once a robust effect is observed, these sequences can be mutated to determine whether the effect is the same as snoRNA deletion, and then a compensatory mutation made to rescue the interaction. The mutations made and the effects observed may identify a binding motif, which can be used as a basis for other potential interactions of the same nature.

4.5: Discussion

The purposes of the investigations described in this chapter were to identify mRNA targets of the two yeast orphan snoRNAs, and to uncover the effect of snoRNA binding on these targets. It was unclear what the functional consequences of snoRNA-mRNA interactions would be, but it seemed plausible that it would alter protein-mRNA interactions, resulting in either stabilisation or destabilisation that could be detected by RNA sequencing.

The results of RiboMinus-treated RNA sequencing revealed that deletion of *SNR4* caused the differential expression of numerous RNAs, a high proportion of which were mRNAs. The majority of RNAs only showed significant differential expression in the ice condition, with a global decrease in expression. This indicated that snR4 may have a protective effect on mRNAs during cold shock. However, performing poly(A)+ selection on sequencing libraries revealed that more poly(A)-tailed RNAs increased in expression than decreased. Upon further analysis, a cut-off of ±0.5 $\log_2$(FC) of WT showed that more RNAs increased in expression in standard conditions, whereas more RNAs decreased in expression in ice conditions, in *snr4Δ* strains. Fewer RNAs showed differential expression upon *SNR45* deletion than *SNR4* deletion, but relative abundance of RNAs in *snr45Δ* ice condition was generally much higher than in standard conditions. Analysis showed that the majority of genes identified in multiple conditions showed a change in expression level in the same direction. This shows that the snoRNA is affecting the same RNAs in the same way, in different conditions. snR4 still appears to have a protective function over RNAs in the ice condition, as was initially observed, and this function is also seen in standard conditions for the same RNAs. However, no global direction was observed for snR45 targets.

Neither RiboMinus depletion nor poly(A)+ selection were as efficient as expected from the descriptions by the manufacturers. A high percentage of the RiboMinus-treated RNA identified in the sequencing was rRNA, and a number of non-coding RNAs without poly(A) tails were identified in the poly(A)+ selected sequencing. Valid targets thus may not have been observed due to presence of other RNA species. Therefore, more repeats of RNA sequencing must be completed in order to highlight more reliable targets, remove false positives, and observe more consistent global effects. Repeats of sequencing would also be beneficial to obtain higher numbers of snR45 targets. Similar numbers of direct targets for snR4 and snR45 were uncovered by CLASH. They also have similar WT cellular expression. It was therefore puzzling that 7.4-fold more RNAs showed differential expression upon *SNR4* deletion than *SNR45* deletion in standard conditions, and 10-fold more in ice conditions. This led to 6-fold higher detection of CLASH targets that change in RNA sequencing for snR4. It is possible that snR4 affects expression of more RNAs than snR45, but the reason for this is unclear.

The RNAs with the greatest changes in the RNA sequencing data were not CLASH targets, and the majority of CLASH targets showed minimal fold changes. This led to the theory that snR4 and snR45 may not predominantly function to affect mRNA expression level. It is possible that snoRNA binding affects at least some mRNAs during translation, which would not be observed by analysing mRNA fold changes. It is also likely that the CLASH analysis did not saturate all snoRNA-target interactions. A large number of experimental replicates were performed to collect as many reproducible CLASH targets as possible, but given the low yield of hybrids compared to single hits, the target list may still be incomplete. Furthermore, weaker or more transient interactions may not be obtained by this method, for example *RCK1*, or excluded due to poor reproducibility. It would be advantageous to perform a different method of obtaining hybrids (see chapter 1.5), although this would not necessarily provide better snoRNA-mRNA interaction data. Alternatively, non-bound mRNAs may be indirectly affected by alterations in functionally related mRNAs that are direct snoRNA targets.

A further possibility is that the snoRNAs observed binding to mRNAs in CLASH data affect neither their expression nor translation, but their modification. As discussed in chapter 1.3, snoRNAs have been shown to play a role in post-transcriptional mRNA modifications, either by ribose modification or blocking base editing. Canonical

snoRNAs guide the methylation of rRNA transcripts, thus it would be possible that they target mRNA transcripts in a similar manner. However, mRNA hybrids mapped to *SNR45* did not show a peak of binding at the expected location of the modification guide sequence, directly adjacent to the D or D' boxes. Neither snoRNA showed a binding profile similar to canonical snoRNAs. This reduces the likelihood that these snoRNAs guide Nop1p in a methyltransferase reaction. This does not preclude their ability to guide a different modification. They may be bound by other proteins to carry out an alternative mRNA modification. Alternatively, they may function by binding to a site which consequentially blocks modification by another complex.

mRNA binding sites over the snoRNA differed between *SNR4* and *SNR45*. While *SNR4* was bound at multiple sites across the gene, *SNR45* only contained one peak. No specific site in snR4 was clearly associated with one direction of change. Notably, several RPGs had decreased differential expression in *snr4Δ*. This suggests that snR4 has a protective effect on RPG mRNAs, which would aid efficient translation. Further work will be necessary in order to uncover an explanation for the numerous snR4-mRNA binding sites across the gene, and to identify a higher number of differentially expressed snR45 targets, to observe if there is correlation between change in abundance and binding site.

Comparison of RiboMinus-treated sequencing with poly(A)+ selected sequencing identified four genes for snR4 and two genes for snR45 that were likely to be valid targets. *RPS20* is a protein component of the 40S subunit of the ribosome, binding the head region near the A site and helix 34 (**Walker *et. al.,* 2013**). This is adjacent to the mRNA entry channel, and may facilitate translation. *RPS20* abundance decreased in both conditions upon *SNR4* deletion, as measured by RiboMinus sequencing, but its value was statistically significant only in ice conditions. This was supported by qPCR analysis (Figure 4.4.1). *RPS20* also showed a decrease in abundance in the poly(A)+ selected sequencing under standard conditions. This was also supported by qPCR analysis, but was not shown to be statistically significant, so the validity of this interaction is more questionable. Further work would be needed to validate the interaction, and determine whether this has an effect on translation efficiency. *RCK1* was not a CLASH target, but showed the highest fold change throughout all datasets. *RCK1* is a MAPK-activated protein kinase involved in the oxidative stress response (**Dahlkvist and Sunnerhagen, 1994; Bilsland *et. al.,* 2004**). *rck1* null mutants show sensitivity to oxidative stress

agents. Deletion of *SNR4* resulted in increased expression of *RCK1*, which may mean lack of snR4 in cells triggers the oxidative stress pathway.

ALD6 showed a consistent significant increase in fold change upon *SNR4* deletion. *ALD6* is the major cytosolic aldehyde dehydrogenase (**Meaden et. al., 1997**). It is required for the conversion of acetaldehyde into acetate during metabolism. Acetate can be converted into acetyl-CoA, which can be used for many processes such as cholesterol synthesis, fatty acid synthesis, histone acetylation and RNA acetylation. Upon *SNR4* deletion, expression of *ALD6* increased. This would potentially lead to higher production of acetate. *ALD4* is the mitochondrial aldehyde dehydrogenase, and also showed increased expression upon both *SNR4* and *SNR45* deletion, with a higher magnitude in *snr4Δ*. *ADH5* was also identified in the sequencing data upon both *SNR4* and *SNR45* deletion, but showed decreased expression. *ADH5* is an alcohol dehydrogenase, which converts acetaldehyde into ethanol. These results indicate that in *snr4Δ* or *snr45Δ*, conversion of acetaldehyde into ethanol is reduced and conversion of acetaldehyde to acetate is increased. snR4 and snR45 may thus regulate levels of acetate in the cell via regulation of the expression of these genes.

Both *ASN1* and *ARG4* showed consistent significant decrease in abundance upon *SNR45* deletion, and both mRNAs were bound to the same nucleotides of the snoRNA. Further to this, simultaneous deletion of both *ASN1* and *SNR45* resulted in a growth defect (Table 4.2.2B). *ASN1* is an asparagine synthetase catalysing the synthesis of asparagine from aspartate. (**Jones, 1978**). *ARG4* is an argininosuccinate lyase, catalysing the final step in the arginine biosynthesis pathway (**Beacham et. al., 1984**). GO term analysis of the overlap between *snr4Δ* and *snr45Δ* sequencing targets showed enrichment for small molecule biosynthetic processes, including alpha-amino acid biosynthetic processes and arginine metabolic processes. Analysis of targets from the DMA screen and from RNA sequencing revealed other enzymes in the arginine metabolic pathway: *ARG1, ARG3, ARG7, CPA2, CAR2* and *ORT1*. Double mutants with *ARG4, CPA2* and *CAR2* showed little growth defect, but the mRNAs showed a greater fold change. Upon *SNR4* or *SNR45* deletion, *ARG1, ARG3, ARG7, CPA2* and *ORT1* expression levels decreased. Figure 4.5.1 shows their positions in the arginine biosynthesis pathway as the final six enzymes of the process. *CAR2* is an ornithine transaminase, catalysing the second step of arginine degradation (**Middlehoven, 1964**). *CAR2* mRNA expression was increased in *snr4Δ* and *snr45Δ*. These results

indicate that both snR4 and snR45 may regulate arginine biosynthesis by enhancing the expression of arginine biosynthesis enzymes, and reducing expression of arginine degrading enzymes. One model would be that snR4 and snR45 affect expression of an upstream transcription factor - such as *GCN4,* an activator for amino acid biosynthetic genes - given the high number of amino acid biosynthesis enzymes identified in the sequencing data. However, *GCN4* mRNA levels were not clearly altered in *snr4Δ* or *snr45Δ* strains. Further investigation would be needed to discover the involvement of snR4 and snR45 in the coordination of these processes.
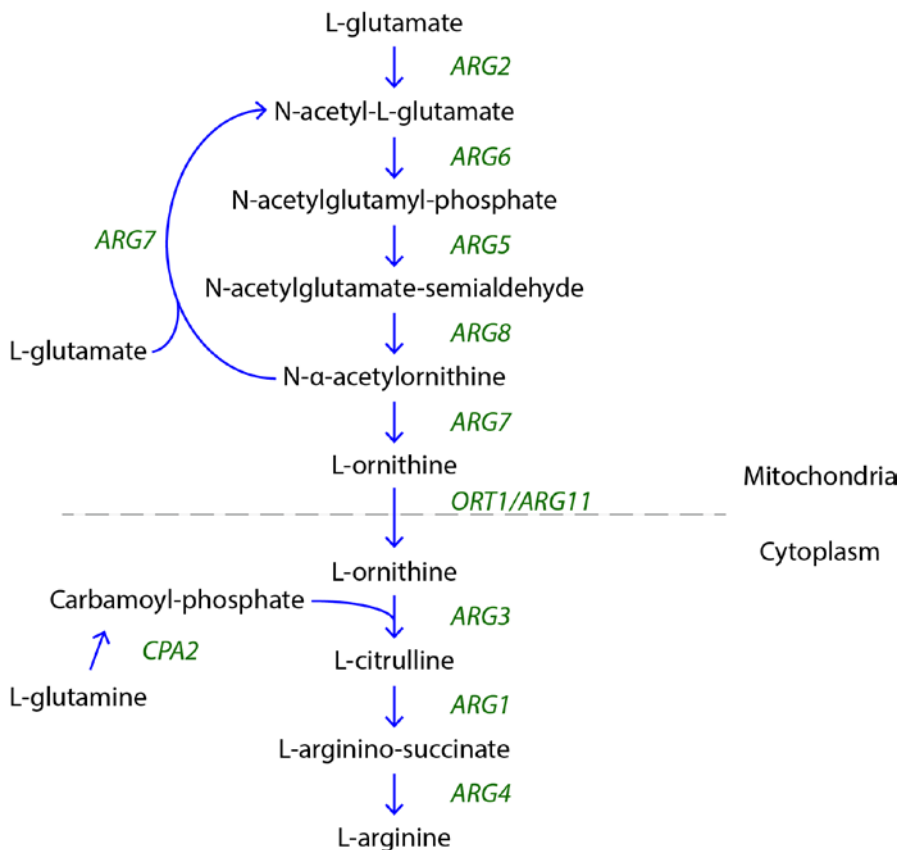
Figure 4.5.1 Arginine biosynthesis pathway. L-glutamate is converted into L-arginine by a series of modifications. Substrates are listed in black, enzymes listed in green. Dashed line represents the mitochondrial membrane to the cytoplasm. Figure adapted and simplified from Crabeel *et. al.* (1997).

The overlap between the RNA sequencing data and CLASH data discussed in this chapter can be compared to the overlap between the RNA sequencing and SGA data (Table 4.2.2). Table 4.5.1 summarises the top targets identified within these screens. The top genes identified from the overlap between RNA sequencing and CLASH, tested by RT-qPCR, are listed in 'CLASH Overlap'. Mutants that showed

the strongest SGA scores combined with the largest changes in mRNA abundance upon snoRNA deletion are listed in 'SGA Overlap', under either TSA or DMA. Genes or alleles listed in blue indicate a decreased mRNA expression level, whereas genes or alleles listed in orange indicate an increased mRNA expression level upon snoRNA deletion. *RCK1* was not a CLASH target, but showed the highest fold change of any mRNA. *nop1-2* showed limited mRNA fold change upon snoRNA deletion, but the strongest score in all SGA TSA screens. Notably, the only gene that was identified in RNA sequencing data, CLASH data and the SGA screen was *ASN1.*

snR4Δ			*snR45Δ*		
CLASH Overlap	**SGA Overlap**		**CLASH Overlap**	**SGA Overlap**	
	TSA	**DMA**		**TSA**	**DMA**
TMA7	*orc6-ph*	*DSE1*	ARG4	*ded1-f144c*	ASN1
RPS20	*nop2-3*	*KNS1*	ASN1	*ded1-95*	
ALD6	*rsp5-3*				
	ded1-95				
RCK1	*nop1-2*				

Table 4.5.1 Summary of top targets identified within RNA sequencing, CLASH and the SGA screen. Genes or alleles listed in blue indicate a decreased mRNA expression level, whereas genes or alleles listed in orange indicate an increased mRNA expression level, upon snoRNA deletion. Dashed line separates genes or alleles with a strong score in one screen but limited overlap in others.

This study has revealed novel potential targets of orphan snoRNAs, which were previously unknown. Work is underway to develop an assay that can systematically detect mRNA 2'-*O*-methyl modifications. This would potentially clarify whether snR4 and snR45 possess the ability to guide mRNA methylation, or have a function linked to translation.

5 Novel interactions between snoRNAs and Ribosomal RNA

5.1: Introduction

The canonical targets of modification guide box C/D snoRNAs are rRNA and snRNA (chapter 1.2). A number of conserved features are expected in these snoRNAs, which are necessary for their biogenesis and function. The box C and box D elements base-pair and, together with a 3' terminal stem, are essential for snoRNA stability (**Watkins *et. al.,* 2000**). Related, but generally less well conserved, are the sequences that form the C' and D' boxes. The modification guide sequences are generally very highly conserved and are located immediately 5' to box D, box D' or both. As discussed in chapter 1.2, *in silico* predictions of the targets for methylation guided by box C/D snoRNAs can be made, based on certain criteria. However, the orphan snoRNAs snR4 and snR45 do not meet these criteria, and had unknown targets and functions. Chapter 4 discussed the interaction of orphan snoRNAs with mRNA. Orphan snoRNAs were thus not predicted to share the canonical box C/D snoRNA function of guiding methylation of rRNA or snRNA. Contrary to this, both snR4 and snR45 were found in multiple hybrids with rRNA in preliminary CLASH data (**Kudla *et. al.,* 2011,** and unpublished work, T. Dudnakova).

While this project was ongoing, HPLC (high performance liquid chromatography) was performed on yeast rRNA to discover if any acetylation marks were present (**Sharma *et. al.,* 2015**). This identified two acetylated cytosines in yeast 18S rRNA, at positions C1280 and C1773, which are conserved to humans at positions C1337 and C1842, respectively. Acetylation is not an essential rRNA modification, as cells are fully viable without it. Kre33p was first identified as the yeast homologue to a bacterial tRNA acetyl-transferase (TmcA, **Ikeuchi *et. al.,* 2008, Ito *et. al.,* 2014**), but was found associated with 40S pre-ribosomes and is essential during ribosome biogenesis, for which acetylation activity is not required (**Grandi *et. al.,* 2002**). Mutations in the acetyl-coA binding site of Kre33p abolished 18S acetylation (**Sharma *et. al.,* 2015**). Similarly, depletion of NAT10p showed loss of acetylation and NAT10p was demonstrated to be the human homologue of Kre33p.

In a collaboration between the Entien and the Granneman labs, CRAC was performed on Kre33p to identify the 18S rRNA binding sites, and uncover other RNA targets of the protein. Within this dataset, hits were found in snR4 and snR45, indicating that Kre33p binds to snR4 and snR45 (**Sharma *et. al.,* 2017**). The Entien group also identified the human box C/D snoRNA U13 as co-migrating with NAT10p in sucrose gradient fractions corresponding to 40S precursors (**Sharma *et. al.,* 2015**). HPLC carried out in cells depleted of U13 resulted in 55.4% reduction in acetylation levels – corresponding to loss of one of the two acetylated residues. Phylogenetic analysis and comparison of nucleotide sequences showed that U13 is the human homologue of snR45 (**Sharma *et. al.,* 2017**). Furthermore, deletion of either *SNR4* or *SNR45* resulted in a 50% reduction in acetylation levels. It was demonstrated that deletion of *SNR4* led to complete loss of ac^{4}C1280, whereas deletion of *SNR45* led to complete loss of ac^{4}C1773. Therefore, snR4 and snR45 guide acetylation of C1280 and C1773, respectively, by Kre33p and U13 guides acetylation of C1842 by NAT10p.

Secondary structure models were generated based on *in vivo* DMS RNA structure probing data, in which secondary structures protect RNA from DMS methylation but accessible sequences are methylated (**Sharma *et. al.,* 2017**). This identified two guide sequences in each snoRNA (GS). Mutation of the 5' ends of snR4 and snR45 (GS1) resulted in abolition of acetylation at C1280 and C1773, respectively. Mutation of GS2 (downstream of the C' box) conferred an 88% and 82% reduction of acetylation in *snr4* and *snr45* mutant strains, respectively. The Entien group also performed mutations of other regions of snR4 and snR45, which similarly reduced or abolished acetylation. This indicated that the secondary structure maintaining the architecture of the snoRNA is important to facilitate acetylation, not just the guide sequences.

Finally, Kre33p contains an N-terminal DEAD-box like helicase module (TmcA homologue, PDB 2ZPA, **Chimnaronk *et. al.,* 2009, Sharma *et. al.,* 2015**). A K289A mutation in the helicase domain of Kre33p reduced 18S acetylation by 90%, indicating that helicase activity is important for acetylation (**Sharma *et. al.,* 2017**). The helicase mutant significantly altered association of certain snoRNAs with pre-ribosomes compared to WT, in sucrose gradient fractionation. Therefore, it was proposed that Kre33p actively facilitates binding of snR4 and snR45 to their rRNA targets. kre33-K289A also affected the association of snR55 with pre-ribosomes;

however, the sedimentation profiles of snR40 indicated that Kre33p was important for its release instead of facilitating binding. It was further found that the kre33-K289A mutation resulted in significantly reduced 2'-O-methylation of snR40 and snR55 targets in close proximity to the snR4 binding site. The hypothesis was thus that Kre33p triggers release of snR40 before facilitating association of snR4 and/or snR55.

This chapter examines the rRNA interactions of snR4 and snR45, further investigates their function as acetylation guide snoRNAs through association with Kre33p, and searches for a human snR4 homologue.

5.2: Orphan snoRNA sequence conservation

Given the small size of the yeast genome, genes that have no function are unlikely to be retained. If both snR4 and snR45 are present in the genomes of distantly related organisms, they are unlikely to be redundant or completely non-essential, and conservation would strongly support a significant function for these snoRNAs.

To assess their conservation, the *SNR4* and *SNR45* sequences were analysed using the NCBI BLAST tool, excluding *Saccharomyces* and optimising for 'somewhat dissimilar sequences' (blastn). A region of ~200nt was taken around each of the most distantly related homologous fungal sequences and analysed using the MultAlin online alignment tool (**Corpet, 1988**). Nine sequences were analysed for *SNR4* (Figure 5.2.1A) and 11 analysed for *SNR45* (Figure 5.2.1B). Red colouring indicates conservation of ≥90% between the analysed sequences, while blue colouring indicates conservation of ≥50%. Black colouring is a conservation score of <50%. The sequences are ordered by similarity, with the most similar sequences to *S. cerevisiae* at the top and least similar at the bottom. The final line in each row gives the consensus sequence of all the sequences based on conservation score; no base is written if there is <50% conservation at that nucleotide.

Figure 5.2.1 Sequence conservation of *SNR4* and *SNR45* among fungi. A) *SNR4* fungal sequence conservation. Nine homologous sequences from distantly related fungi identified using BLAST were analysed using MultAlin online alignment tool. The C box is from nucleotides 15-21, C' box from nucleotides 123-129, and D box from nucleotides 184-187. **B)** *SNR45* fungal sequence conservation. 11 homologous sequences from distantly related fungi identified using BLAST were analysed using MultAlin. The C box is from nucleotides 28-34, D' box from nucleotides 97-100, C' box from 108-114, and D box from nucleotides 193-196. **C)** *SNR4* and *SNR45* alignment in MultAlin. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. Black lines denote the box motifs, and conserved regions of unknown function.

Figure 5.2.1A shows that *SNR4* has a highly conserved region at the 5' end of the snoRNA, which continues until downstream of the C box (nucleotides 15-21). The C' box, at nucleotides 123-129, is also highly conserved. As noted in chapter 4.3, snR4 lacks a clear D' box consensus sequence, and this is supported by the lack of any highly conserved element in the expected location between the C and C' boxes. Reducing the number of sequences inputted to MultAlin to the most similar fungal homologues increased the number of highly conserved nucleotides (≥90%), as expected. However, a clear D' box was still not identified within the conserved nucleotides, based on the D' box consensus for the majority of box C/D snoRNAs (**van Nues *et. al.,* 2011**). The D box, at nucleotides 184-187, and the remaining 3' end are also highly conserved. Furthermore, the 3' end nucleotides are complementary to the 5' end nucleotides, and would be predicted to base-pair *in vivo* to form the termini stem. Notably, nucleotides 146-155 showed a region of high conservation of unknown function.

Figure 5.2.1B shows a region of high conservation located at the 5' end of the *SNR45* gene. The C box, at nucleotides 28-34, is highly conserved. The D' box, at nucleotides 97-100, is also well conserved, with high conservation extending through the C' box (nucleotides 108-114) until nucleotide 130. A highly conserved region of unknown function is located at nucleotides 60-69, with another region spanning nucleotides 140-151, which occupies a similar location to a conserved region within *SNR4*. Finally, the 3' end is very well conserved, including the D box (nucleotides 193-196). The nucleotides downstream of the D box are complementary to the nucleotides directly upstream of the C box, and are predicted to base-pair in a 3' terminal stem.

SNR45 is overall more highly conserved than *SNR4* and was identified in more fungal species. However, given the similarities in positioning of the unpredicted regions of high conservation, and the similarities in results of the SGA screen (Tables 3.3.1 and 3.3.2), it was important to check the sequence conservation between the two orphan snoRNAs. Figure 5.2.1C shows the two snoRNAs aligned to each other using the same alignment software. The two clearly show little sequence similarity, indicating that the two yeast orphan snoRNAs are not redundant. This is consistent with the evolutionary retention of homologues of both snoRNAs, and with Figure 3.2.5 showing that the cell remains viable with double depletion of both snoRNA genes.
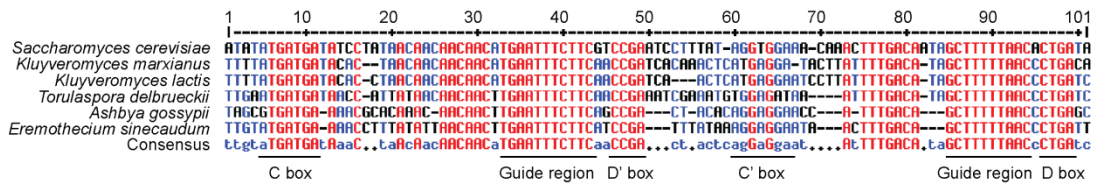
Figure 5.2.2 Sequence conservation of *SNR47* among fungi. The five most distantly related fungal homologue sequences identified using BLAST were analysed using MultAlin. The C box is from nucleotides 5-11, D' box from nucleotides 46-49, C' box from nucleotides 60-66, and D box from nucleotides 96-99. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. Black lines denote the box motifs and guide regions.

For comparison, Figure 5.2.2 shows the sequence conservation for *SNR47*. snR47 is another box C/D snoRNA, but with a canonical methylation guide function. BLAST was performed, and a block of ~100nt taken around the regions of sequences from the most distantly related fungal homologues. These were submitted to MultAlin, again with red nucleotides indicating a score of ≥90%, and blue nucleotides indicating a score of ≥50%. Conservation is very high over the boxes, shown at nucleotides 5-11 (C box), 46-49 (D' box), 60-66 (C' box) and 96-99 (D box). In contrast to *SNR4* and *SNR45*, regions of high conservation are found directly upstream of the D' and D boxes, and represent the canonical targeting sequences of box C/D snoRNAs (as discussed in chapter 5.1). The nucleotides indicated as T41 and T91 in the gene sequences base-pair with A2220 in 25S rRNA and A619 in 18S rRNA, respectively, to guide methylation of these nucleotides. Comparing Figure 5.2.2 to Figure 5.2.1 shows the clear difference in the pattern of conservation between the canonical and orphan snoRNAs. In particular, *SNR4* and *SNR45* lack highly conserved sequences immediately upstream of the D and D' boxes indicating that these two snoRNAs are unlikely to have methylation guide functions.

5.3: CLASH analysis of orphan snoRNA-rRNA interactions

snR4 and snR45 were identified in hybrids with both mRNAs and rRNAs (preliminary data and chapter 4). To further understand the variety of targets for these snoRNAs, CLASH data was analysed by hybrid type (experiments performed

by T. Dudnakova). Reproducible hybrids with a predicted Gibb's free energy (folding energy) of ≤-12ΔG were counted by the number of hybrids with RNAs from distinct genes, to demonstrate the variety of targets within each RNA class. Figure 5.3.1A shows a pie chart representing the percentages of snR4 target genes identified within different RNA classes. The percentages were calculated by tallying the number of genes from each RNA class that formed hybrids with snR4, and comparing this to the total number of genes that formed hybrids with snR4. 71% of the total snR4 targets consisted of distinct mRNA genes. 20% were from distinct snoRNAs (including self-interactions), while 4% comprised rRNA and 4% were from tRNA hybrids. 1% of gene targets were from other RNA classes. To further analyse the distribution of hybrids, the total number of snR4 hybrids identified for each RNA class was tallied. This was calculated by tallying the number of hybrids for each gene from each RNA class that formed with snR4. 959 hybrids were identified between snR4 and rRNAs (Figure 5.3.1B). This was the highest number of snR4 hybrids identified for all RNA classes. 215 internal hybrids were identified within snR4 (intramolecular hybrids). Contrasting to this, snR4 was identified in only 60 hybrids with other snoRNAs (intermolecular hybrids). 70 mRNA hybrids were identified in total, while only five hybrids were identified between snR4 and tRNAs, and only two with other RNAs.

These analyses were also performed on snR45. Figure 5.3.1C shows the pie chart representing the percentages of snR45 target genes identified within different RNA classes. snR45 targeted a higher number of mRNAs, with this class comprising 73% of total snR45 targets. 12% consisted of hybrids with distinct snoRNAs, while 3% were with rRNAs and 3% were with XUTs. 9% of gene targets were from other RNA classes. A fewer number of hybrids were identified between snR45 and rRNA genes than were identified for snR4-rRNA (259, Figure 5.3.1D). 70 mRNA hybrids were identified for snR45, which was the same number as snR4-mRNA hybrids. However, snR45 formed fewer snoRNA hybrids than snR4 both inter-molecularly and intra-molecularly (13 and 9, respectively). Three hybrids were identified between snR45 and antisense RNAs, two with XUTs, and six hybrids with other RNA classes.
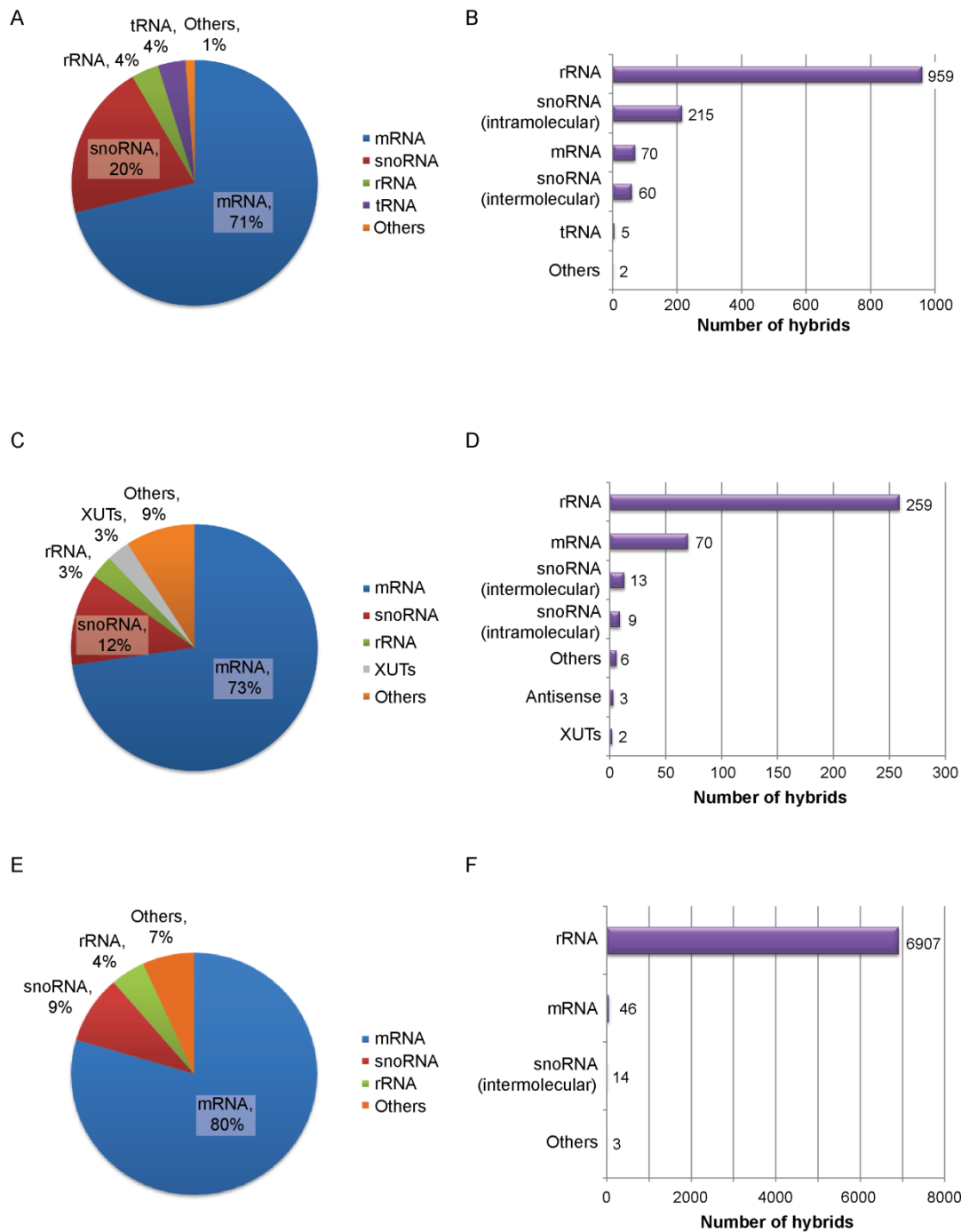
Figure 5.3.1 Numbers of CLASH hybrid hits in different RNA classes for snR4 and snR45. A) Pie chart showing the percentages of snR4 target genes identified within each RNA class. B) The total number of snR4 hybrid hits identified within each RNA class. C) As in 'A' but for snR45. D) As in 'B' but for snR45. E) As in 'A' but for snR55. F) As in 'B' but for snR55.

To demonstrate how this data compares to canonical box C/D snoRNA targets, this analysis was performed for snR55, which guides the methylation of U1269 in 18S rRNA. Figure 5.3.1E shows the pie chart representing the percentages of snR55 target genes identified within different RNA classes. Compared to snR4 and snR45, snR55 targeted a higher number of mRNA genes, comprising 80% of total distinct gene targets. Less snoRNA genes were targeted, comprising 9% of total targets, while a similar percentage of rRNA genes were targets (as there are a limited number of rRNA genes). 7% of gene targets were from other RNA classes, with only one gene identified for each class. Analysing the number of hybrids identified within each RNA class demonstrates the large contrast between the orphan snoRNAs and the canonical snoRNA (Figure 5.3.1F). snR4 formed 959 hybrids with rRNA genes, while snR45 formed 259 hybrids with rRNA genes (Figures 5.3.1B and 5.3.1D, respectively). However, snR55 formed 6,907 hybrids with rRNA genes, showing a difference of 7-fold compared to snR4, and 27-fold compared to snR45. However, the total number of mRNA hybrids identified with snR55 was lower than both snR4 and snR45, at 46 hybrids. The number of intermolecular snoRNA hybrids identified with snR55 was similar to that of snR45, but much lower than snR4, at 14 hybrids. Furthermore, no intramolecular hybrids were identified for snR55. Finally, only three hybrids were found with other classes of RNA. The ratios between the number of hybrids identified with rRNA genes and the number of hybrids identified with other RNA classes strongly highlights the contrast between the orphan snoRNAs and the canonical snoRNA, in both diversity of targets and function.

The majority of snR4 and snR45 hybrids were with rRNA. In order to visualise which nucleotides of the snoRNAs were involved in rRNA base-pairing, a pile-up was created of all the nucleotides found in hybrids with rRNA (Figure 5.3.2). As described in chapter 4.3, the snR4 and snR45 sequences from each hybrid were mapped onto the full *SNR4* or *SNR45* sequence, respectively, to visualise which regions of the snoRNA were most frequently involved in rRNA binding. The x axis denotes nucleotide position over the gene, and the y axis represents the number of independent CLASH hybrids identified for each nucleotide. The different colours indicate positions of the C, D', C' and D boxes, and no D' box is indicated for snR4.

Figure 5.3.2A shows the pile-up profile for *SNR4*. The first peak was between +75-110, close to where the D' box would be expected to be positioned. There was a large peak centred on +150, which included the C' box, then a steep dip in the

number of hybrids, before peaking again before and including the D box. This profile was extremely similar to the mRNA hybrid profile (Figure 4.3.1A), except that the ratio of the C' peak in Figure 5.3.2A was much higher than in the mRNA profile. Additionally, the number of hybrids found for the snR4-rRNA interaction was approximately 4-fold greater than the number found for snR4-mRNA hybrids.

The rRNA binding profile for *SNR45* (Figure 5.3.2B) was also very similar to the mRNA binding profile in Figure 4.3.1B. There was a large peak in the number of hybrids directly after the C' box, between +115 and +150. However, the rRNA binding profile also showed a peak of slightly smaller height from +150-180, which was not present in the mRNA binding profile. There was also a much smaller peak directly downstream of the C box present in the rRNA profile, that was absent from the mRNA profile. However, the number of hybrids found for snR45-mRNA and snR45-rRNA was comparable for each peak, on the scale. The similarity between interactions of snR4 with mRNA and rRNA was notable, as were the additional rRNA binding sites present in the snR45 interactions. Both findings suggest that these snoRNAs function via base-pairing with target mRNAs and rRNA, although the consequences of such interactions were unclear.
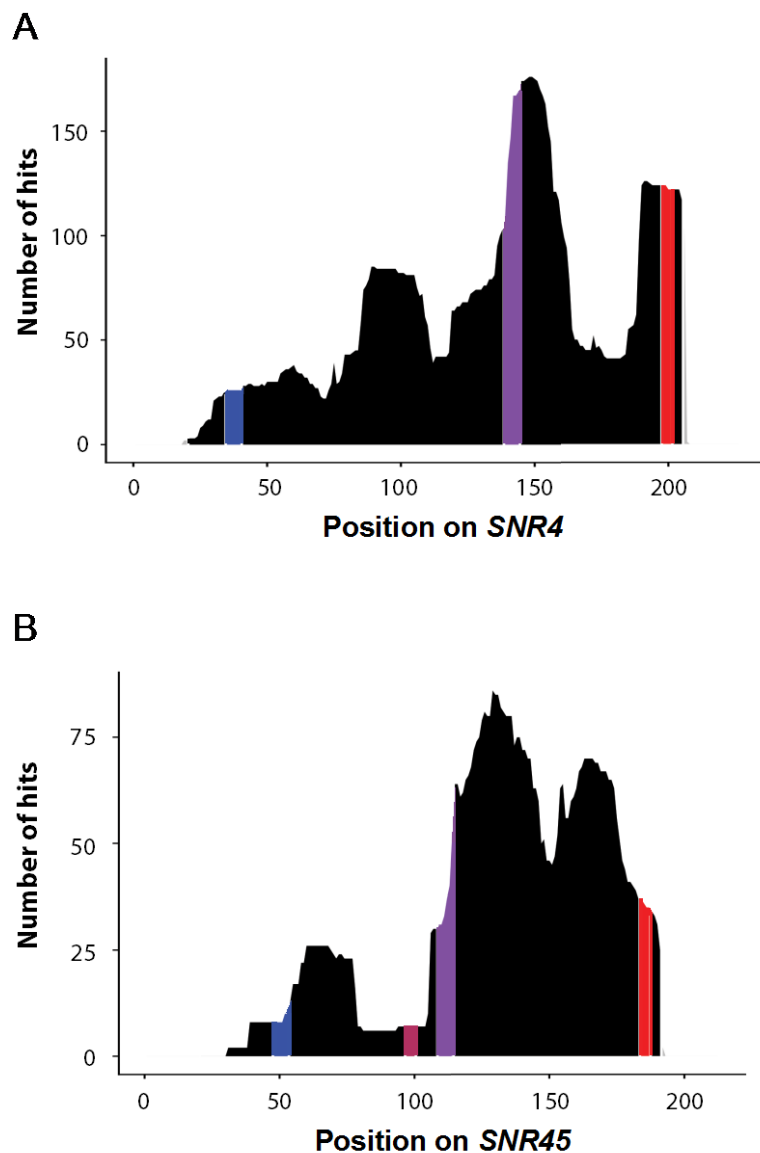
Figure 5.3.2 Pile-up of snoRNA-rRNA hybrid hits on snoRNAs. A) The snR4 sequence obtained from snR4-rRNA hybrids mapped onto *SNR4*. B) The snR45 sequence obtained from snR45-rRNA hybrids mapped onto *SNR45*. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Since snR4 and snR45 are conserved, but dispensable during normal growth, it was considered that they would be more important during stress conditions. A commonly used stress condition for budding yeast is transfer from glucose-containing medium

to ethanol plus glycerol, to mimic the naturally occurring "diauxuic shift", as performed in chapter 3.2. Thus, HTP-tagged strains were grown in a preculture medium containing 2% w/v glucose and then shifted to medium containing 2% v/v ethanol plus 2% v/v glycerol as carbon source for 30min prior to crosslinking with a 5min incubation on ice prior to crosslinking (unpublished work, T. Dudnakova). CLASH was carried out as described in chapter 2.10, and the snR4 and snR45 sequences from each hybrid were mapped onto the full *SNR4* and *SNR45* sequences, respectively.

Figure 5.3.3A shows snR4-rRNA hybrids over *SNR4.* The peak between +75 and +110 was still present, but at an altered ratio compared to Figure 5.3.2A. Following growth in glucose medium, the peak around +100 was approximately 67% of the peak at the D box, whereas following growth in ethanol-glycerol medium, the peak around +100 had approximately 80% of the number of hybrids at the D box. The peak over the D box was much sharper, and contained the highest number of hybrids. Furthermore, the peak at the C' box was absent following growth in ethanol-glycerol medium, showing that this interaction site is not essential under stress conditions.

Comparison of Figure 5.3.3B with Figure 5.3.2B showed a change in the ratio of the two peaks between the C' and D box. In Figure 5.3.2B, the peak from +115-150 was approximately 1.2-fold the size of the peak between +150 and +180, whereas Figure 5.3.3B showed no "peak" between +115-150, and the number of hybrids was approximately 67% that of the latter peak. Binding between +150 and +180 of snR45 was better preserved under poor carbon nutrition, suggesting a role in maintaining ribosomal synthesis or function.

The number of snR4-rRNA hybrids recovered following growth in ethanol-glycerol medium was 10-fold lower than following growth in glucose medium, and the number of snR45 hybrids was 4-fold lower. This may reflect the lower rate of ribosome synthesis predicted for the reduced growth rate observed on the poorer carbon sources (chapter 3.2). Therefore, there would be substantially less pre-rRNA available for interaction, resulting in fewer hybrids.

A

B
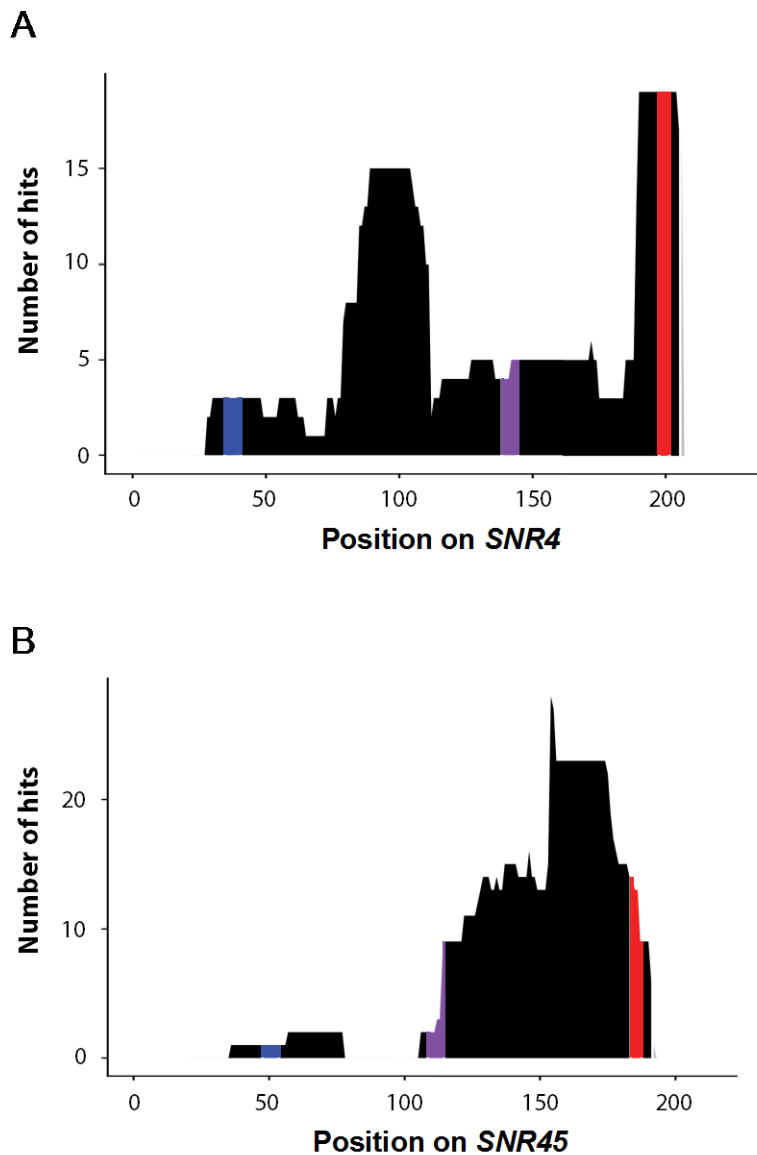
Figure 5.3.3 Pile-up of snoRNA-rRNA hybrid hits on snoRNAs following growth in ethanol-glycerol medium. A) The snR4 sequence obtained from snR4-rRNA hybrids following growth in ethanol-glycerol medium, mapped onto *SNR4*. B) The snR45 sequence obtained from snR45-rRNA hybrids following growth in ethanol-glycerol medium, mapped onto *SNR45.* X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Figure 5.3.4 places these hybrid peaks onto the snoRNA sequence conservation seen in Figure 5.2.1. The peak around +100 in snR4 in Figures 5.3.2A and 5.3.3A corresponded to very little sequence conservation, which makes the significance of its preservation under poor carbon source unclear (Figure 5.3.4A). If this binding site is biologically important during stress response, it would be expected that the sequence would be well conserved between fungi, but this is not the case. The peak around the D box was also poorly conserved. However, the end of the peak around the C' box involved hybrids with the unpredicted highly conserved region. It is therefore interesting that this peak disappears during growth on a poor carbon source. Similarly, for snR45, the peak upstream of the D box showed little sequence conservation, making the significance of its preservation under poor carbon source unclear (Figure 5.3.4B). However, the peak between +115 and +150, visible in Figures 5.3.2B and 5.3.3B, included the unpredicted region of high conservation at the 3' end. This may indicate that there is a biological function for rRNA binding at this site. This peak was also reduced during growth in a poor carbon source.
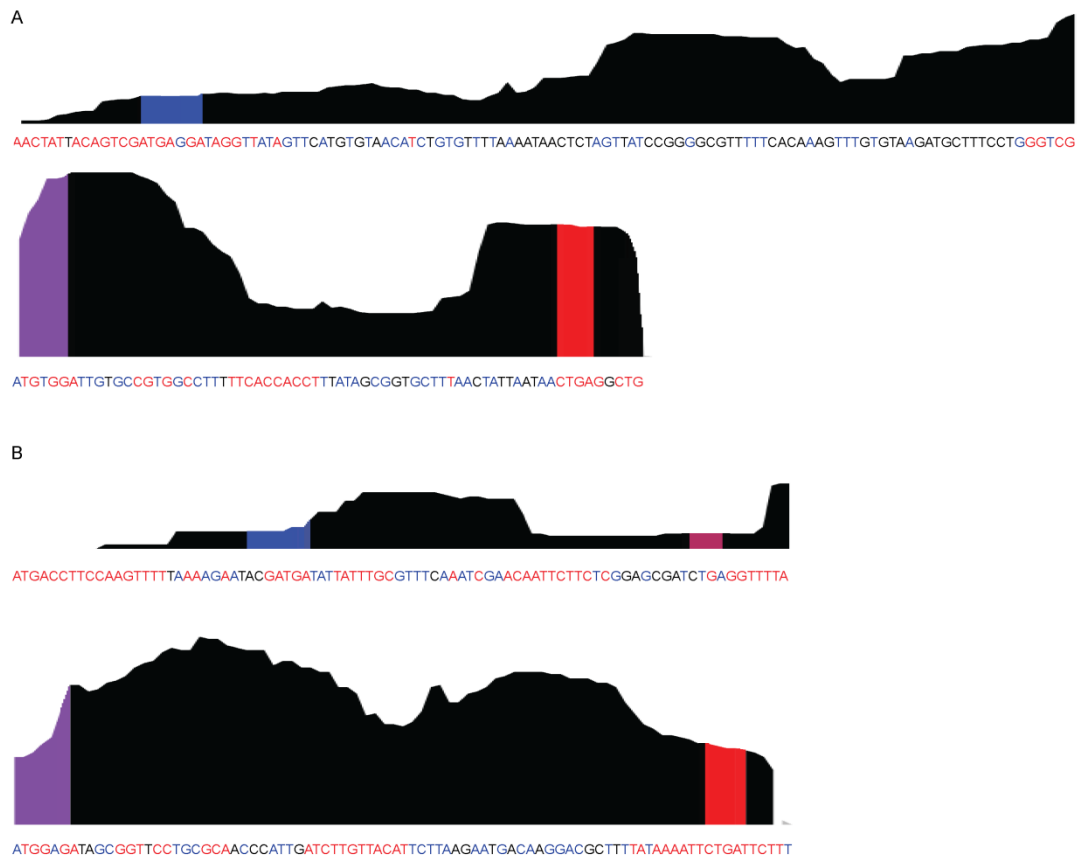
Figure 5.3.4 Distribution of hybrids across conserved snoRNA sequences. A) snR4-rRNA hybrids from Figure 5.3.2A aligned to the *SNR4* sequence conservation from Figure 5.2.1A. B) snR45-rRNA hybrids from Figure 5.3.2B aligned to the *SNR45* sequence conservation from Figure 5.2.1B. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Red letters indicate conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

The CLASH hybrid data were also mapped across the rDNA gene, which is termed '*RDN37'* in yeast. The pile-ups are shown in Figure 5.3.5. The x axis denotes the nucleotide position over the *RDN37* gene, and the y axis shows the number of independent CLASH hybrids identified for each position. Each subsection of the gene is represented by a different colour – as indicated in the key – and grey lines highlight known methylation sites within the rRNA.
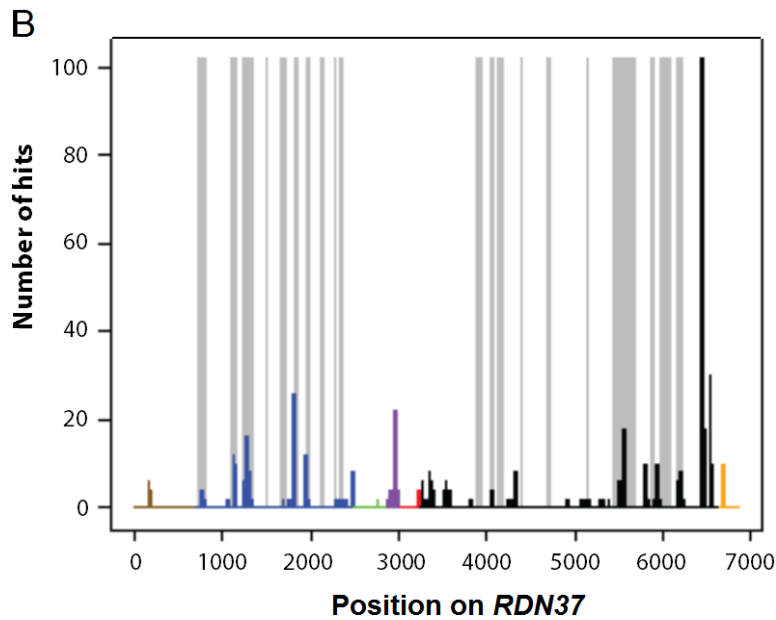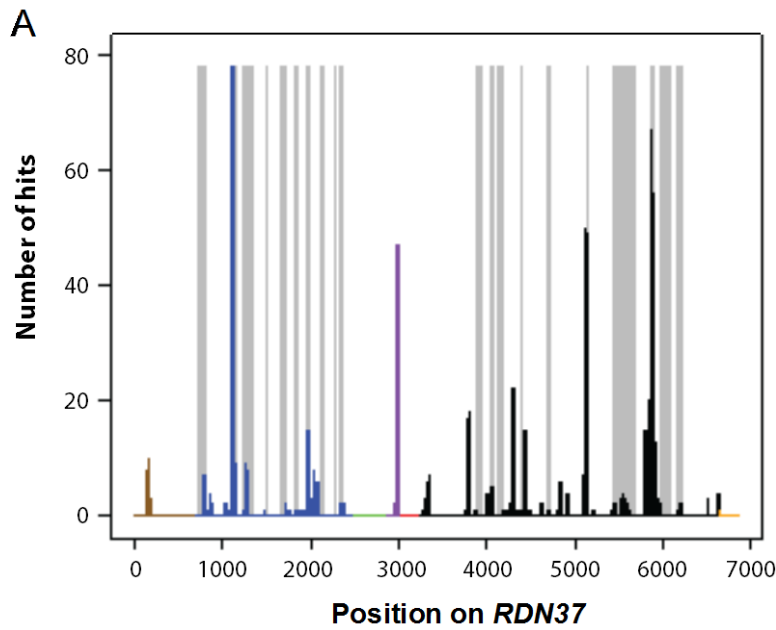
Figure 5.3.5 Pile-up of snoRNA-rRNA hits on rRNA. A) The rRNA sequence obtained from snR4-rRNA hybrids mapped onto *RDN37*. B) As in 'A' but for snR45. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Brown: 5' ETS, blue: 18S rRNA, green: ITS1, purple: 5.8S rRNA, red: ITS2, black: 25S rRNA, orange: 3' ETS. Grey bars indicate known methylation sites. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Hybrids between snR4 and rRNA are mapped in Figure 5.3.5A. There was a large peak in the number of hybrids around position +1097-1129, which corresponds to 397-429nt of 18S rRNA. There was another large peak at +2970-2986, which is within 5.8S rRNA, and two large peaks within 25S rRNA, at positions +5109-5136 and +5890-5926 corresponding to 1858-1885nt and 2639-2675nt, respectively. snR45 hybrids are shown in Figure 5.3.5B, and bind in small numbers to a few sites across the gene, with one large peak at +6429-6458, corresponding to 3178-3207nt of 25S rRNA. None of these peaks overlapped with known methylation sites, though a number were in close proximity.

The hybrids obtained in the datasets for strains grown in ethanol-glycerol medium were then mapped onto *RDN37* (Figure 5.3.6). As expected, the overall number of hybrids within each peak was fewer than when grown in glucose medium, and the distribution of hybrids across the gene was more restricted following growth in ethanol-glycerol medium. Similar numbers of hybrids were recovered for snR45-rRNA and for snR4-rRNA in ethanol-glycerol medium, whereas in glucose medium the majority of snR4-rRNA peaks contained more hybrids than snR45-rRNA peaks. Comparing Figure 5.3.6A with Figure 5.3.5A showed that the peak observed around +1100 following growth in glucose medium was reduced by 25-fold following growth in ethanol-glycerol medium. Similarly, the peaks observed in 25S rRNA were also reduced, with the peak at +5100 reduced 10-fold, and the peak around +5900 reduced ~35-fold. However, the peak at +2970 within 5.8S rRNA was still present, with hybrid numbers only reduced ~3-fold, meaning this was the tallest peak. Figure 5.3.6B shows that very few hybrids were recovered between snR45 and *RDN37* following growth in ethanol-glycerol medium. However, the large peak observed at +6450 within 25S was relatively well-preserved, with 2.5-fold fewer hybrids. The peak at +1900 was proportionally higher following growth with a poor carbon source, relative to growth in glucose medium.
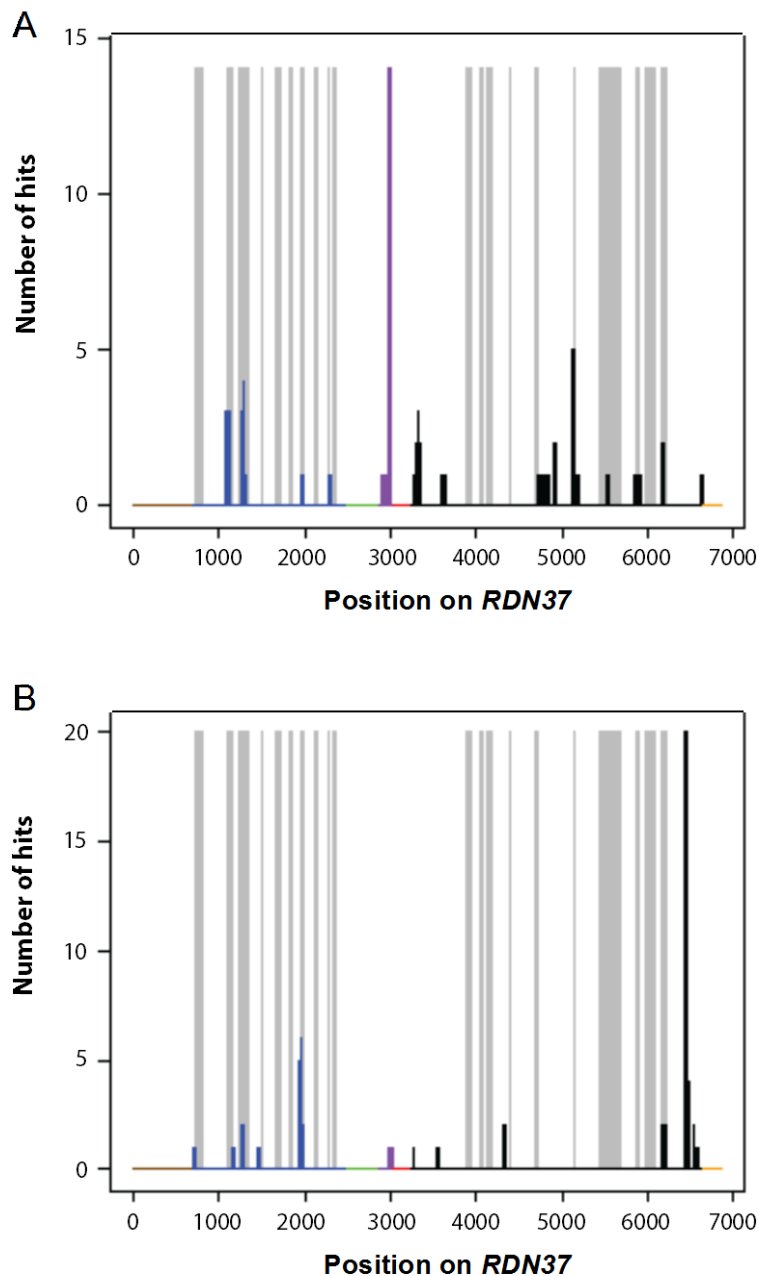
Figure 5.3.6 Pile-up of snoRNA-rRNA hybrid hits on rRNA following growth in ethanol-glycerol medium. A) The rRNA sequence obtained from snR4-rRNA hybrids following growth in ethanol-glycerol medium, mapped onto *RDN37*. B) As in 'A' but for snR45. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Brown: 5' ETS, blue: 18S rRNA, green: ITS1, purple: 5.8S rRNA, red: ITS2, black: 25S rRNA, orange: 3' ETS. Grey bars indicate known methylation sites. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

The CLASH data was then analysed to identify which peaks of snoRNA binding corresponded to which peaks of rRNA binding (Figure 5.3.7). Isolating all snR4-rRNA hybrids between +0 and +80 of *SNR4* identified the sole rRNA binding site as +5890-5926 (2639-2675nt of 25S rRNA). This was the highest 25S peak in Figure 5.3.5A. This rRNA peak was identified in a very low number of hits following growth in ethanol-glycerol medium, corresponding to the low number of hybrids identified between +0 and +80 of *SNR4* following growth in ethanol-glycerol medium. Isolating hybrids between +70 and +120 of *SNR4* revealed two main rRNA binding sites: +2970-2986 (109-125nt of 5.8S) and +3772-3807 (521-556nt of 25S). These comprised the majority of hybrids at this region of *SNR4*. Two hybrids each were also formed with site +4603-4619 (1352-1368nt of 25S) and site +6494-6516 (3243-3265nt of 25S).

Analysis of the peak at the C' box from +120-180 of *SNR4* revealed the majority of hybrids were formed by +140-160. A large number of these hybrids were bound to the peak around +1100 of *RDN37* (400nt of 18S). The majority of hybrids were also found at site +5850-5884 (2599-2633nt of 25S). Further hybrids were found at sites +1269-1290 (569-590nt of 18S), +3987-4014 (736-763nt of 25S), and two hybrids with the 5' of *RDN5-6* (1-23nt). In contrast, hybrids showing base-pairing with different nucleotides to those in the above hybrids were found between +140-178 of *SNR4*. These hybrids formed with +1953-1986 of *RDN37* (1253-1286nt of 18S). Finally, isolating the D box peak from +165-210 showed the vast majority of hybrids formed with the 5' of *RDN5-6* (20-44nt). A small number of hybrids were also found at +5109-5136 (1858-1885nt of 25S), with single or two hybrids found at sites +4033, +4231, +4413 and +6495 of *RDN37* (782nt, 980nt, 1162nt and 3244nt of 25S, respectively). Analysis of these peaks following growth in ethanol-glycerol medium showed only two binding sites of *SNR4:* +75-108 and +165-206. The former site only formed two hybrids, both with the large peak at 5.8S. The latter site formed many hybrids, all with the same site of *RND5-6.*
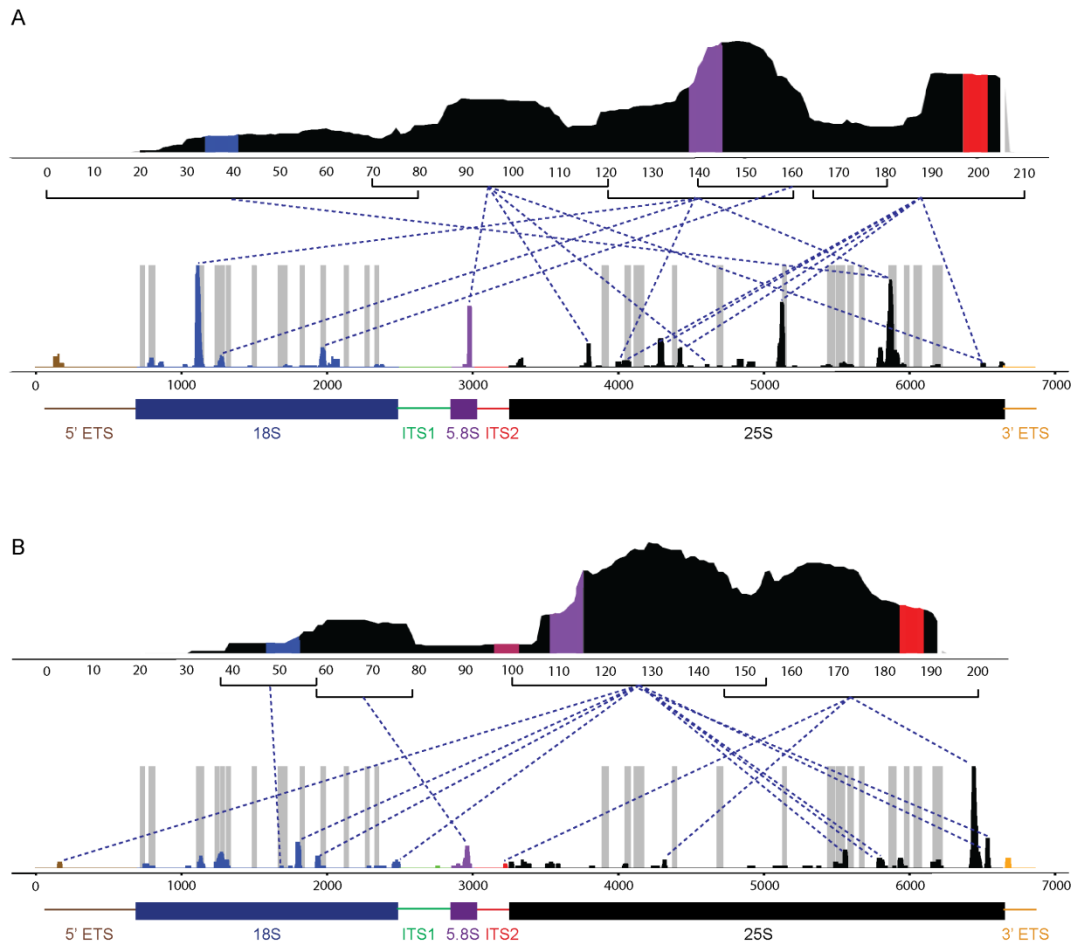
Figure 5.3.7 RNA-RNA interaction map. A) Pile-up of the snR4 sequences of hybrids from Figure 5.3.2A segmented by binding peaks, mapped to the pile-up of the *RDN37* sequences of hybrids from Figure 5.3.5A. B) Pile-up of the snR45 sequences of hybrids from Figure 5.3.2B segmented by binding peaks, mapped to the pile-up of the *RDN37* sequences of hybrids from Figure 5.3.5B. Dashed blue lines denote the interaction sites. Blue indicates the C box, mauve indicates the D' box (unknown for *SNR4*), purple indicates C' box and red indicates D box. Grey bars indicate known methylation sites. A cartoon of *RDN37* is shown below the pile-up. Hybrid data was obtained from aggregated Nop1p CLASH experiments, performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

The same analysis was applied to groups of binding sites of *SNR45*. Segregating all hybrids between +0 and +100 of *SNR45* showed that +39-58 bound to +1703-1721 of *RDN37* (1003-1021nt of 18S rRNA). There was also a hybrid between +58-78 that bound to +2952-2986 (91-125nt of 5.8S rRNA), which corresponded to the

peaks in Figure 5.3.2B and Figure 5.3.5B. This was hybridised upstream of the 5.8S site bound by snR4. Separating hybrids bound by +100-155 of *SNR45* identified the majority of hybrids at +6523-6553 (3272-3302nt of 25S). A high number of hybrids were also found with peak +1788-1815 (1088-1115nt of 18S). Also within 18S rRNA were binding sites at +1956-1986 (1256-1286nt of 18S) and +2452-2496 (1752-1796nt of 18S). Notably, the former was exactly the same sequence of 18S rRNA as bound by snR4. Further hybrids were formed in 25S with sites +5481-5502 (2230-2251nt), +5781-5802 (2530-2551nt), +5812-5845 (2561-2594nt of 25S), and +6664-6692 (3413-3441nt). There was also a binding site within the 5'ETS, at +152-174.

Finally, analysing hybrids bound throughout the peak at +145-200 identified the majority of hybrids bound at +6429-6458 (3178-3207nt of 25S). This corresponds to the highest peak in Figure 5.3.5B. A few hybrids bound at +4307-4326 (1056-2075nt of 25S) and one at +3216-3231 (197-212nt of ITS2). Analysing these sites following growth in ethanol-glycerol medium revealed that the majority of hybrids were formed between +145-200 of *SNR45,* mostly to the peak at +6429-6458 (as supported by Figure 5.3.6B). The peak between +4307-4326 was also still present in these conditions. Few hybrids were still present following growth in ethanol-glycerol medium at the *SNR45* peak from +100-150. These were at +1244-1284 (544-584nt of 18S), +1791-1811 (1091-1111nt of 18S), +2458-2479 (1758-1779nt of 18S), and +6529-6544 (3278-3293nt of 25S).

5.4: snR4 and snR45 guide rRNA acetylation

CRAC was performed on Kre33p, identifying snR4 and snR45 hits within the dataset (**Sharma *et. al.,* 2017**). This indicated that Kre33p binds to snR4 and snR45. However, the CRAC approach is not optimised to provide hybrid data, and therefore there was limited evidence as to whether snR4 and snR45 guided Kre33p to the 18S acetylation marks. To investigate this further, the Tollervey lab attempted CLASH on a Kre33-HTP strain and on a NAT10-HTP strain (the human homologue of Kre33p) (unpublished work, T. Dudnakova). However, neither protein cross-linked well, and the experiment provided only limited data.

It appeared possible that CLASH data on the Nop proteins might contain hybrids between snR4 and/or snR45, and the acetylated residues. Therefore, nucleotides

flanking each acetylated residue were identified within hybrid files. Figure 5.4.1A shows one example of the 17 hybrids found between snR4 and the region of 18S rRNA surrounding C1280. Base-pairing occurred upstream of the ac⁴C1280, with a predicted Gibb's free energy of -13.8ΔG. It is notable that the nucleotides involved in base-pairing upstream of ac⁴C1280 corresponded exactly to the unpredicted region of high conservation identified in Figure 5.2.1A. As hypothesised, it indeed appears that this region is well conserved because it has an important biological function. In a number of the hybrids, there were an additional five Watson-Crick base-pairs predicted surrounding the acetylated cytosine, with a predicted Gibb's free energy of -22.7ΔG for the hybrid. However, as the residue would need to be free to accept the chemical group, it is unlikely that these residues are base-paired i*n vivo* at the time of modification.

Figure 5.4.1 snoRNA hybrids at 18S-C1280 acetylation site. A) snR4-18S rRNA hybrid from CLASH. B) snR45-18S rRNA hybrid from CLASH. Acetylated cytosine is coloured in red, and denoted by an arrow and labelling. Watson-Crick base-pairs are denoted by a blue straight line, G-U base-pairings denoted by a dot. Dashed purple line denotes the boundary between snoRNA and rRNA regions of the hybrid. Hybrid drawn from full chimeric sequence from Nop1p CLASH, and base-pairing indicated by the ViennaD folding programme.

Interestingly, snR45 was also identified in two hybrids at this same site (Figure 5.4.1B). Base-pairing occurred over the same residues on the ribosomal RNA as for snR4 binding, although the snR45-rRNA base-pairing was not as consecutive as the snR4-rRNA base-pairing. The hybrid had a low predicted Gibb's free energy of -21.7ΔG, but only when multiple bulges were permitted within the base-pairing. The

region of snR45 involved in base-pairing is not well conserved. Furthermore, given the low number of hybrids recovered this seems less likely to have significant cellular function.

Figure 5.4.2A shows the snR45 hybrid with ac[4]C1773. This interaction occurred upstream of the site of modification and did not include base-pairing with the acetylated residue. The predicted Gibb's free energy of the hybrid was -22.5ΔG, but again only when multiple bulges were present within the base-paired regions. The representation in Figure 5.4.2A had a predicted Gibb's free energy of -10.4ΔG, with a shortened version of the hybrid having -12ΔG.
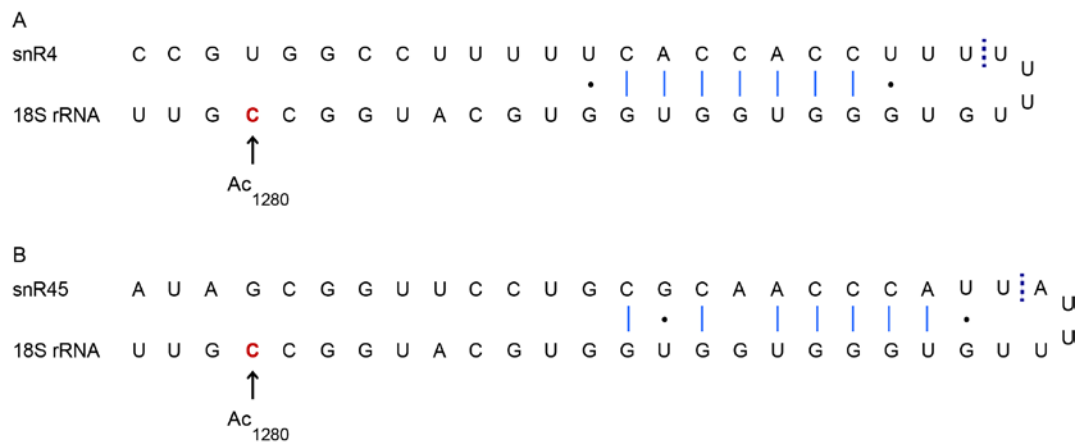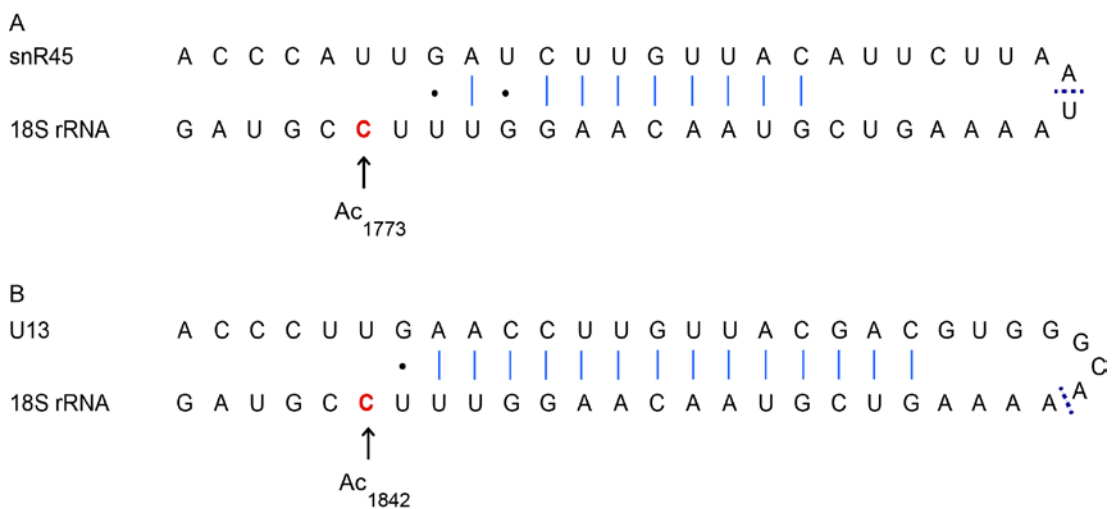

Figure 5.4.2 snoRNA hybrids at 18S-C1773 acetylation site. A) snR45-18S rRNA hybrid from yeast CLASH. B) U13-18S rRNA hybrid from human CLASH. Acetylated cytosine is coloured in red, and denoted by an arrow and labelling. Watson-Crick base-pairs denoted by a blue straight line, G-U base-pairings denoted by a dot. Dashed purple line denotes the boundary between snoRNA and rRNA regions of the hybrid. Hybrid drawn from full chimeric sequence from Nop1p/Fibrillarin CLASH, and base-pairing indicated by the ViennaD folding programme.

As Nop protein CLASH data contained hybrids between snR4, snR45 and the acetylated residues, it appeared possible that human CLASH data might contain hybrids between U13 and the corresponding human acetylated residues. Searching the corresponding human acetylation sites (C1337 and C1842) within Fibrillarin CLASH data identified 431 hybrids between U13 and 18S rRNA around ac[4]C1842 (≤-12ΔG). The representation of one of these hybrids in Figure 5.4.2B had a

predicted Gibb's free energy of -22.3ΔG. It showed substantial base-pairing, again not including the acetylated residue. The bases involved in this hybrid are very similar to the bases in the snR45-C1773 hybrid (Figure 5.4.2A), indicating that this interaction is well conserved.
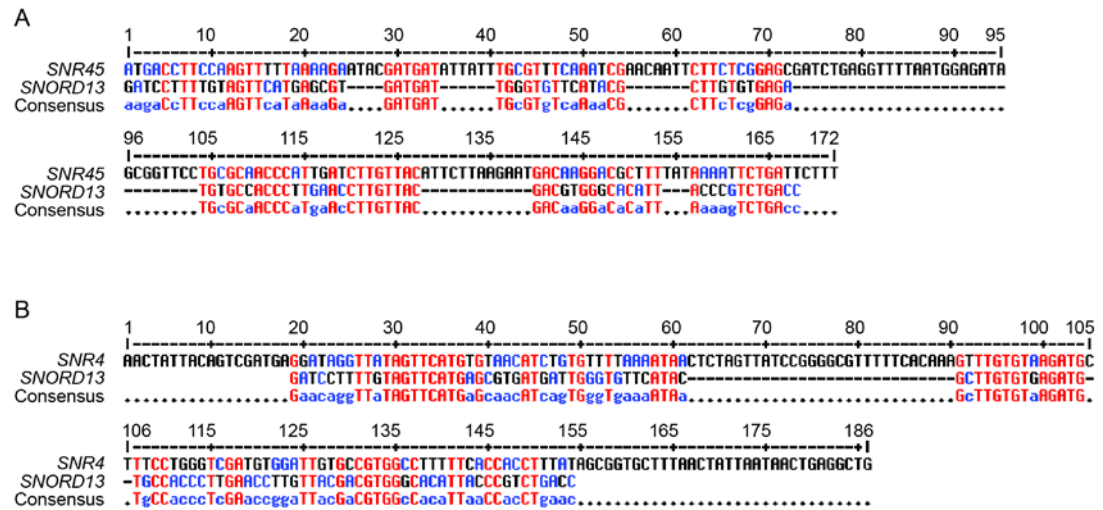
Figure 5.4.3 Sequence conservation of *SNORD13* with yeast orphan snoRNAs. A) *SNR45* and *SNORD13* sequences were aligned using MultAlin. B) *SNR4* alignment with *SNORD13*. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath.

Comparison of the *SNR45* sequence to *SNORD13* (U13 gene) using the MultAlin software showed good conservation throughout the snoRNAs, with very high conservation at the interaction site (nucleotides 120-127, Figure 5.4.3A). This corresponds to the region of high conservation identified in Figure 5.2.1B, supporting the conclusion that the region is well conserved because it has the biological function of guiding Kre33p to its target RNA. Notably, *SNORD13* also has regions well conserved with *SNR4* but, crucially, not including the interaction site. Furthermore, snR4 was not found in hybrids with the region around C1773 in yeast Nop1p CLASH, and U13 showed limited base-pairing to the region around C1337 in human Fibrillarin CLASH.

5.5: Searching for a human homologue of snR4

Both snR4 and snR45 have a biological function (**Sharma *et. al.,* 2017**), which is conserved to humans for snR45. It was therefore predicted that the function of snR4 would also be conserved to humans. To test this hypothesis, *SNR4* was analysed using BLAST against all higher organisms, optimising for 'somewhat dissimilar sequences' (blastn). However, this returned few results with plausible conservation. Therefore, BLAST was performed against individual model organisms to determine whether the *SNR4* sequence was conserved in some organisms but not others. A block of ~200nt was taken around regions of results that contained homology with the conserved rRNA binding site, and analysed using MultAlin. Figure 5.5.1 shows the alignments.

Only one BLAST result was obtained for the rRNA binding site in *Trypanosoma brucei*. Figure 5.5.1A shows *SNR4* aligned with a region of chromosome 11 of *Trypanosoma brucei gambiense*. The rRNA binding site of *SNR4* was highly conserved in this region (nucleotides 157-165), including a number of nucleotides upstream. However, there was no discernible D box downstream of this site in the region taken, indicating this well conserved sequence may not be part of a snoRNA. A number of BLAST results with good conservation of the rRNA binding site were seen in *Drosophila melanogaster*. Figure 5.5.1B shows alignment with a region of chromosome X, and Figure 5.5.1C shows alignment with a region of chromosome 3R. Once again, the rRNA binding site and surrounding regions were very well conserved, but there was no discernible D box downstream of this in the sequences analysed. Blastn was also performed for *Xenopus laevis*, however the only hit for the rRNA binding region was in the coding sequence of ribosomal protein S6 kinase II, and conservation over the rRNA binding site was poor. Interestingly, when BLAST was performed against *Homo sapiens, Danio rerio, Arabidopsis thaliana* and *Caenorhabditis elegans,* there were no hits over the rRNA binding site.
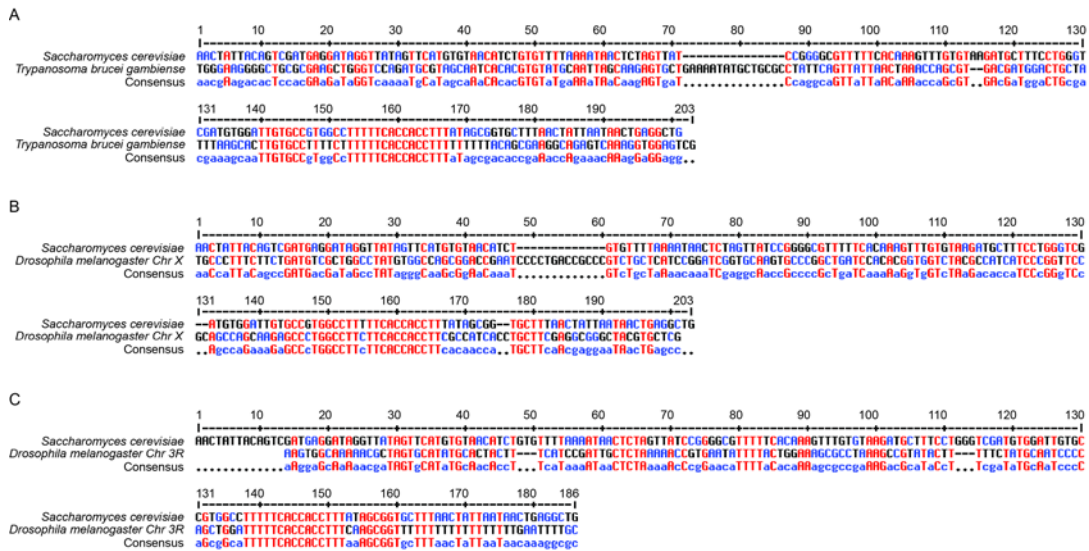
Figure 5.5.1 Individual *SNR4* alignments with higher eukaryotes. A) *SNR4* aligned with 769,685-769,885 of chromosome 11 of *Trypanosoma brucei gambiense*. B) *SNR4* aligned with 15,081,970-15,082,170 of chromosome X of *Drosophila melanogaster*. C) *SNR4* aligned with 23,032,755-23,032,935 of chromosome 3R of *Drosophila melanogaster*. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath.

While some U13-C1842 hybrids were recovered using NAT10p CLASH data, no hybrids could be found around C1337. However, hybrids were discovered between snR4, snR45 and the acetylation sites by analysing Nop1p CLASH data, and hybrids between U13 and the acetylation sites were discovered using Fibrillarin CLASH data. Therefore it was plausible that a human snR4 homologue could be uncovered by analysing Fibrillarin CLASH data around the 18S-C1337 site. Table 5.5.1 lists all the human snoRNAs that formed hybrids within a 70bp region around C1337 of pre-47S rRNA. The table also lists any known human methylation target sites for each snoRNA, the yeast homologue if known, and the corresponding target site for the yeast homologue, obtained from the snOPY database (**Yoshihama *et. al.,* 2013**). The majority of the snoRNAs in the table were known to guide methylation at one or more sites, and 69% had a known yeast homologue. However, as a number of these snoRNAs target multiple sites on pre-47S rRNA, and others lack a clear yeast homologue, it was possible that they still may be the functional snR4 homologue.

Human	Human Target Site	Yeast	Yeast Target Site
SNORD1B	28S:G4362	snR38	25S:G2815
SNORD3D	-	snR17	-
ACEA-U3	-	snR17	-
SNORD13	18S:C1842	snR45	18S:C1773
SNORD14	18S:C462	U14	18S:C414
SNORD24	28S:C2338,28S:C2352	U24	25S:C1437,25S:A1449,25S:G1450
SNORD25	18S:G1490	snR56	18S:G1428
SNORD32A	18S:G1328,28S:A1511	snR59	25S:A807
SNORD33	18S:U1326	snR55	18S:U1269
SNORD51	28S:A1511	snR59	25S:A807
SNORD65	18S:U627	snR77	18S:U578
SNORD70	18S:A512	-	-
SNORD83A	-	-	-
SNORD100	18S:G436	-	-
SNORD117	-	-	-
SNORD124	-	-	-

Table 5.5.1 Human snoRNAs identified in hybrids around 18S-C1337, and their yeast homologues. Column 1 lists the human box C/D snoRNA in hybrids within 90bp around 18S-C1337. Column 2 indicates the target nucleotide of each SNORD, if known, and the corresponding rRNA subunit. Column 3 states the yeast homologue of each SNORD, if known. Column 4 indicates the target nucleotide of the yeast homologue, if known, and the corresponding rRNA subunit.

Initially, all snoRNA sequences from the hybrids were aligned using MultAlin. Comparison immediately showed that SNORD14 and SNORD83A bind too far upstream of the acetylation site and did not include any of the nucleotides from the snR4 binding site. It is therefore unlikely that either of these are the snR4 homologue. *SNORD13* was also an unlikely candidate, not only because it showed poor conservation with the *SNR4* binding site (Figure 5.4.3), but also because the predicted base-pairing of the hybrid was poor (>-12ΔG). Each snoRNA was then considered for homology based on how well conserved the snoRNA was to *SNR4*, particularly to the rRNA-binding site, and how similar the hybrid sequence and predicted base-pairing were to the snR4-C1280 hybrids. Given that *SNORD13* only showed high conservation to *SNR45* over the rRNA binding site, conservation over

the *SNR4* rRNA binding site was considered much more important than overall conservation. Furthermore, snR45 and U13 hybrids showed very similar positioning with respect to the acetylated residue, and showed similar base-pairing, with U13 forming a longer, more stable hybrid. It was therefore presumed that the snR4 human homologue would also show similar positioning and hybridisation, possibly with a longer hybrid for the human RNAs.

The SNORD1B hybrid was dissimilar to the snR4 hybrid and showed very poor conservation. SNORD3D base-paired downstream of the acetylation site, which was incomparable to the snR4 binding site. Similarly, ACEA-U3 showed non-consecutive base-pairing, and showed poor conservation to *SNR4*. Both SNORD3D and ACEA-U3 are variants of the U3 gene. Given that U3 is essential for site-specific cleavage of pre-rRNA, it appeared unlikely that it also acts as the human snR4 homologue. *SNORD24* showed good conservation with the snR4-rRNA binding site, and these nucleotides were involved in base-pairing in the hybrid. However, SNORD24 targets two different sites for methylation – using both guide regions - and its yeast homologue targets three sites, all within 25S rRNA. Furthermore, the hybrid formed fewer consecutive base-pairs than snR4 with the acetylation site of rRNA, meaning the interaction would be too weak to act as the snR4 homologue. *SNORD25* showed good conservation with the snR4-rRNA binding site; however, the site itself contained extra nucleotides which were not conserved. Furthermore, the hybrid included a high number of predicted G-U base-pairs, which is very dissimilar to the snR4-rRNA hybrid and also much weaker. *SNORD32A* showed very good conservation with the snR4-rRNA binding site (Figure 5.5.2A). However, SNORD32A targets G1328 for methylation, which is nine nucleotides upstream of the acetylated cytosine. Thus, the snoRNA target sequences would be expected to be very similar, and hence well conserved, as it is targeting the same region. Surprisingly, the sequence showing high *SNR4* conservation was not the same sequence present in the C1337 hybrid (Figure 5.5.2B). The conserved sequence is used to target A1511 in 25S rRNA (Figure 5.5.2C), for which snR59 is the yeast homologue. Notably, snR4 does not form hybrids at the respective yeast site, A807, despite having a highly conserved targeting sequence. Therefore, SNORD32A cannot be the human snR4 homologue.
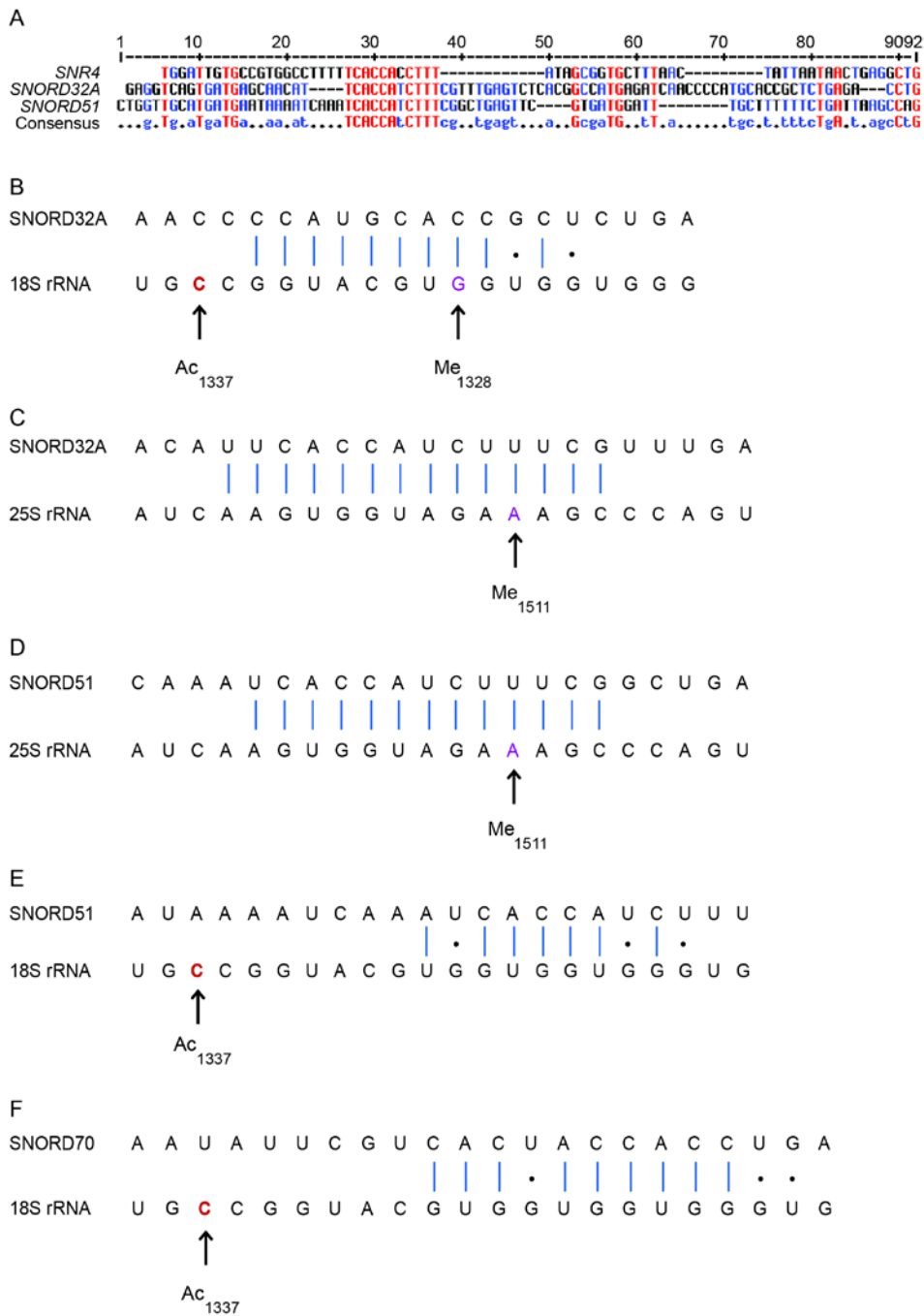
Figure 5.5.2 *SNR4* alignment with human snoRNAs. A) *SNR4* aligned with *SNORD32A* and *SNORD51*. **Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. B) SNORD32A hybrid with 18S rRNA with its target, G1328. C) SNORD32A hybrid with 25S rRNA with its target, A1511. D) SNORD51 hybrid with 25S rRNA with its target, A1511. E) SNORD51 hybrid with 18S rRNA, upstream of C1337. F) SNORD70 hybrid with 18S rRNA, upstream of C1337. Acetylated cytosine coloured in red, and denoted by an**

(Figure 5.5.2 cont.) arrow and labelling. Methylated residue is coloured in purple, and denoted by an arrow and labelling. Watson-Crick base-pairs are denoted by a blue straight line, G-U base-pairings denoted by a dot. Hybrid drawn from human CLASH, with base-pairing indicated by the ViennaD folding programme.

SNORD33 guides methylation at C1326, 11 nucleotides upstream of C1337, and hence showed strong hybridisation with the region around the acetylated cytosine, but its conservation with *SNR4* was otherwise poor. *SNORD51* showed high conservation over the snR4-rRNA binding site, similar to *SNORD32A* (Figure 5.2.2A). Similarity between targeting sequences of SNORD32A and SNORD51 was expected, given that they target the same nucleotide: A1511 in 25S rRNA (Figures 5.5.2C and D). However, the SNORD51-C1337 hybrid also showed very similar positioning and base-pairing with respect to the snR4-C1280 hybrid (Figure 5.5.2E), which SNORD32A did not. It is possible that SNORD51 may act as a guide to both its canonical site and also this acetylation site, but given that the targeting sequence is directly upstream of the D box (which is not the case for snR4), it is more likely that it forms a hybrid due to high conservation, rather than because of actual function. *SNORD70* showed some conservation with the snR4-rRNA binding site, and its hybrid showed high similarity in position and base-pairing to snR4 (Figure 5.5.2F). However, the predicted base-pairing included three nucleotides of the D box, which is not expected *in vivo,* and this sequence is the targeting sequence for methylation and so is unlikely to act as a targeting sequence for acetylation also. *SNORD100* and *SNORD124* showed poor conservation with *SNR4* as well as limited base-pairing with the acetylation site.

Elimination of these snoRNAs left SNORD65 and SNORD117. Neither *SNORD65* nor *SNORD117* showed strong conservation with the snR4-rRNA binding site (Figures 5.5.3A and 5.5.4A). However, both showed very similar positioning with respect to the acetylated cytosine, and showed similar base-pairing. snR4 binds upstream of C1280, downstream of the snR4 C' box (Figures 5.5.3B and 5.5.4B). SNORD65 is known to guide the methylation of 18S-U627; however this targeting sequence is directly upstream of the D box. Notably, the sequence in SNORD65 hybrids in Figure 5.5.3C is upstream of C1337 and directly downstream of the C box, which is very similar to the snR4 hybrid. SNORD65 could therefore have dual-function using different binding sites to guide a different modification.

A

```
                1        10        20        30        40 44
          |--------+---------+---------+---------+---------+---|
SNR4      CGTGGCCTTTTTCACCACCTTTATAGCGGTGCTTTAACTATTAA
SNORD65   AAATGATGAAATCACC--CAAAATAGCTG-GAATTACCGGCAGA
Consensus aaagGacgaaaTCACC..CaaaATAGCgG.GaaTTAaCgacaaA
```

B

snR4   C G U G G C C U U U U U C A C C A C C U U U A U

18S rRNA U G **C** C G G U A C G U G G U G G U G G G U G U U

Ac$_{1280}$

C

SNORD65   A A A U G A U G A A A U C A C C C A A A A

18S rRNA U G **C** C G G U A C G U G G U G G U G G G U G C C

Ac$_{1337}$

Figure 5.5.3 SNORD65 similarity to snR4-C1280 binding. A) *SNR4* **conserved region aligned with** *SNORD65* **region involved in hybrids***.* **Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. B) snR4 hybrid with 18S-C1280, as seen in Figure 5.4.1A. C) SNOR65 hybrid with 18S rRNA, upstream of C1337. Acetylated cytosine is coloured in red, and denoted by an arrow and labelling. Watson-Crick base-pairs are denoted by a blue straight line, G-U base-pairings denoted by a dot. Hybrid drawn from Fibrillarin CLASH, with base-pairing indicated by the ViennaD folding programme.**

SNORD117 has no known function and is thus an orphan snoRNA, like snR4. The hybrid formed at C1337 was very similar to the snR4-C1280 hybrid (Figure 5.5.4C, compared to Figure 5.5.4B). snR4 also showed additional predicted base-pairing directly around the acetylated residue, which can also be predicted for SNORD117, so showing further similarity. However, the SNORD117 sequence present in hybrids was directly upstream of the D box, which shows more similarity to canonical box C/D snoRNA binding. The acetylation binding site of snR4 is not directly upstream of the D box, but partway between the C' box and D box, and this is also the case for snR45 and U13. Therefore, SNORD117 did not show similar positioning of the

binding site within the snoRNA. Positioning is likely very important, as Kre33p binding presumably does not occlude the Nop1p binding site over the D box.
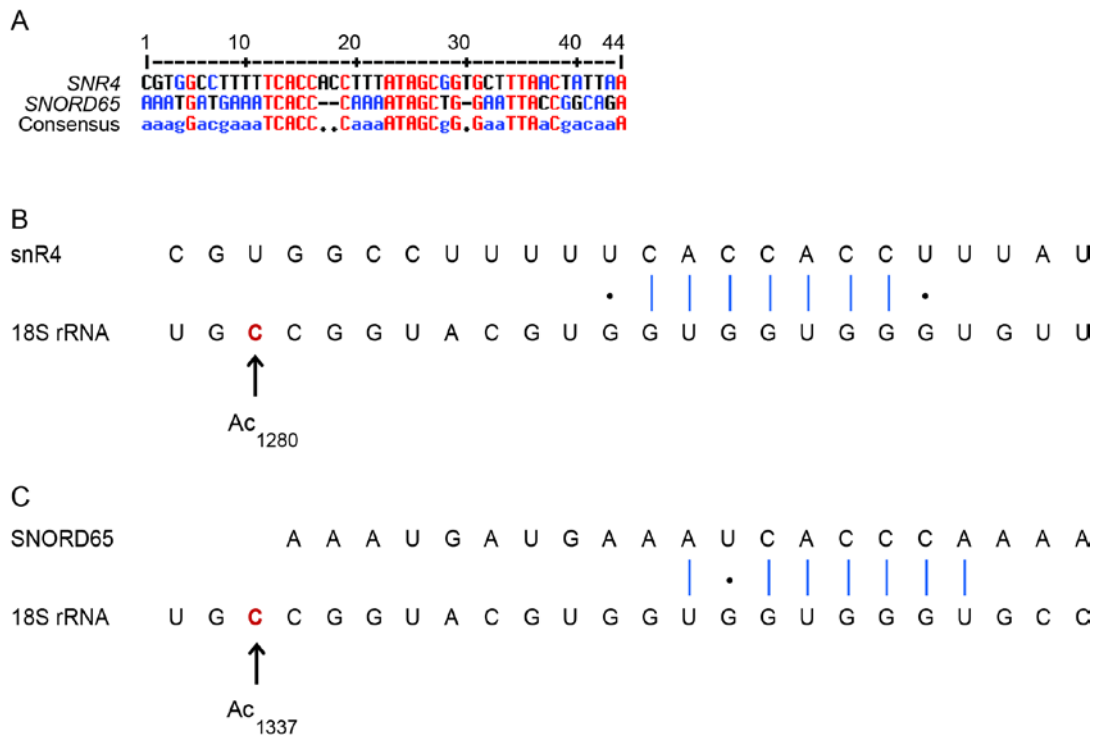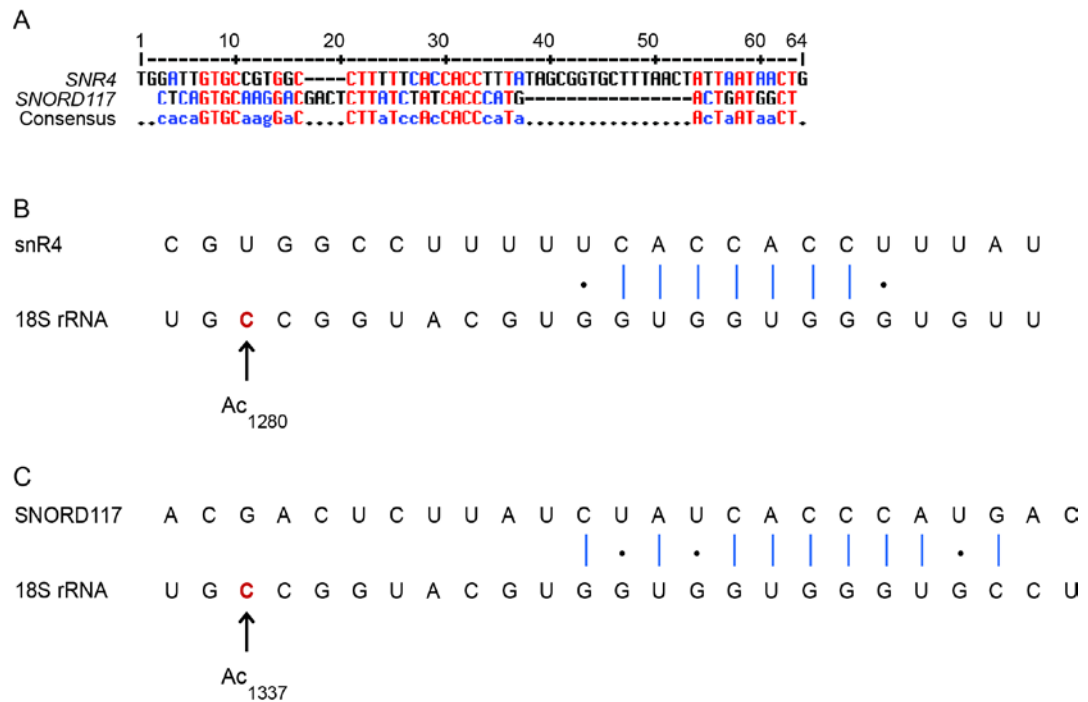

Figure 5.5.4 SNORD117 similarity to snR4-C1280 binding. A) *SNR4* **conserved region aligned with** *SNORD117* **region involved in hybrids. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. B) snR4 hybrid with 18S-C1280, as seen in Figure 5.4.1A. C) SNOR117 hybrid with 18S rRNA, upstream of C1337. Acetylated cytosine is coloured in red, and denoted by an arrow and labelling. Watson-Crick base-pairs are denoted by a blue straight line, G-U base-pairings denoted by a dot. Hybrid drawn from Fibrillarin CLASH, with base-pairing indicated by the ViennaD folding programme.**

SNORD65 and SNORD117 have the potential to be the human snR4 homologue. However, evidence for this remains limited. Further experiments will be needed to determine whether either of these is indeed the homologue of snR4. The human genome is much larger than the yeast genome, and much of it is still to be identified. It is therefore possible that the snR4 homologue has not yet been discovered or classified as a snoRNA. Based on this premise, a Hyb database was created from the human NAT10p CLASH data. BLAST was then run on this database, using the snR4-C1280 binding site, with the sequence '*TGGCCTTTTTCACCACCT*'. This

returned three hits, two of which were the same. The latter two sequences matched almost identically to 80nt of *SNR4*. However, when this sequence was inputted to Ensembl BLAT, no results were returned. The fasta sequence used for BLAST with the hyb database was then used for Ensembl BLASTN, which returned two results. The first was a region on chromosome 15, matching the entire sequence, and the second a region on chromosome 9, matching all but one nucleotide. Running BLASTN on the same sequence without the 5' 'T' also gave two more results: a region on chromosome X and a region on chromosome Y, both matching all but one nucleotide. It should be noted that Ensembl BLAST gives different results depending on the exact nucleotides included, and thus may produce more hits if the inputted sequence was adjusted upstream or downstream.

Approximately 200nt were then taken around each hit, and the entire region compared to *SNR4* using MultAlin. Figure 5.5.5 shows the results of these alignments. All four results showed very high conservation over the rRNA binding site. This is expected as this was the sequence used to find these hits. Conservation across the rest of the sequence appeared in patches, but no substantial regions of conservation were identified. The region of chromosome X was from 5,128,278-5,128,426 on the '-' strand, and the region of chromosome Y was from 16,497,279-16,497,449 on the '+' strand (Figures 5.5.5A and B). The regions of chromosomes X and Y were very similar, as BLAT of one showed 83% identity to the other. The region of chromosome 9 was from 36,807,148-36,807,400 on the '+' strand, and the region on chromosome 15 was from 46,322,824-46,323,024 on the '-' strand (Figures 5.5.5C and D).

In order to test the validity of these regions, each region was searched for a potential D box. Neither the region at chromosome X nor Y contained the 'CTGA' motif downstream of the potential binding site, only upstream. However, both regions on chromosomes 9 and 15 contained a 'CTGA' sequence downstream of the potential binding site. Analysis of the region of chromosome 9 revealed two 'CTGA' sequences upstream of the conserved region, which could possibly be D' boxes. Figure 5.5.5C shows that the first of these sequences aligned well with 'CTGT' at nucleotides 61-64. The second potential D' box aligned well with 'CTGG' at nucleotides 132-135, just upstream of the *SNR4* C' box at nucleotides 146-152. Furthermore, there were seven highly conserved nucleotides at the 5' end of the

alignment followed by further patches of conservation. However, the 'CTGA' motif downstream of the potential binding site did not align with the *SNR4* D box.
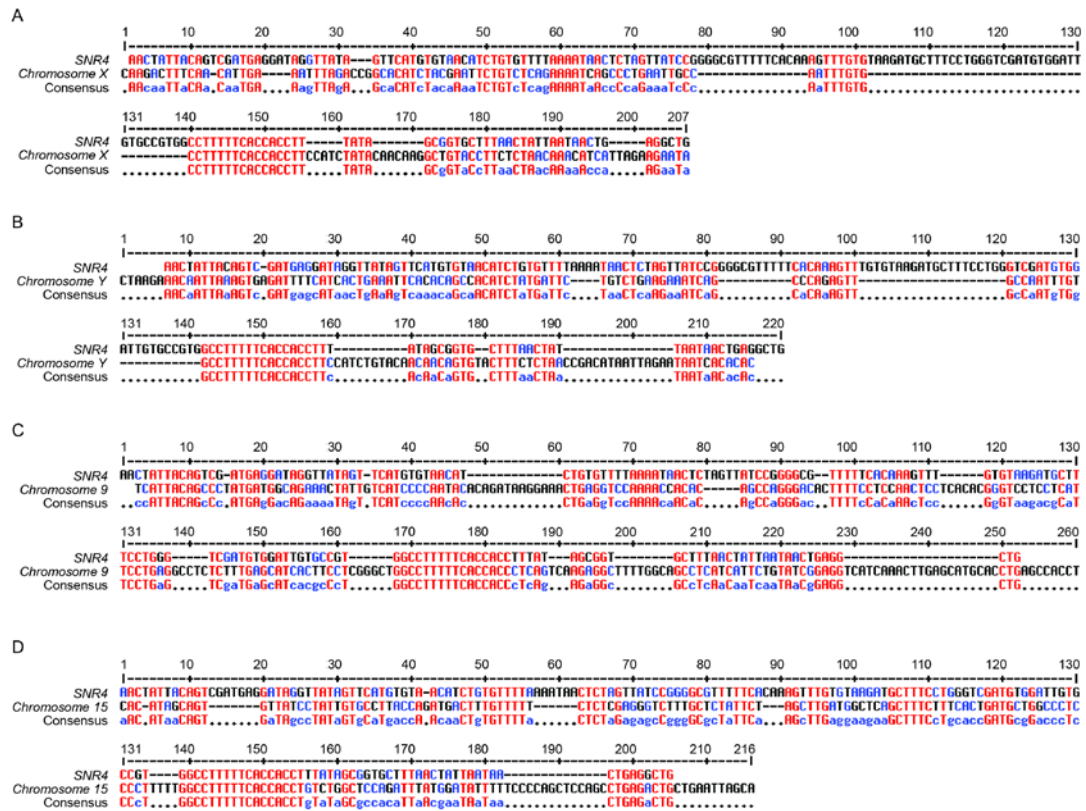
Figure 5.5.5 *SNR4* **alignment with unassigned human target regions. A)** *SNR4* **aligned with 5,128,278-5,128,426 of chromosome X. B)** *SNR4* **aligned with 16,497,279-16,497,449 of chromosome Y. C)** *SNR4* **aligned with 36,807,148-36,807,400 of chromosome 9. D)** *SNR4* **aligned with 46,322,824-46,323,024 of chromosome 15. Red indicates conservation of ≥90%, blue indicates ≥50% conservation, and black is <50% conserved. The consensus sequence is denoted underneath. All regions taken from the hg38 assembly of** *Homo sapiens.*

The region on chromosome 15 included two D box consensus sequences downstream of the conserved site (Figure 5.5.5D). One aligned with the *SNR4* D box, and the other was immediately 3' of the end of the *SNR4* sequence. If this region is in fact a box C/D snoRNA, it would be uncertain which the true D box was; however, the nucleotides after the first D box sequence aligned well with *SNR4.* There was another D box sequence upstream of the conserved region on chromosome 15, which could potentially be a D' box. This is shown at nucleotides

116-119. However, the last nucleotide of this potential D' box aligned with the first nucleotide of the *SNR4* C' box, and so is either unlikely to be the correct alignment, or unlikely to be the site of the *SNR4* D' box. Given the variety in C and C' boxes, these are harder to identify, and current analysis of both sequence and conservation gave no indication of whether a C or C' box was present in the regions on chromosomes 9 or 15.

Further analysis would be required to determine whether either of these regions contained an unidentified box C/D snoRNA. However, the analysis demonstrated in chapter 5.5 has narrowed a large number of potential candidates to four possible human snR4 homologues: SNORD65, SNORD117, the region described on chromosome 9 and the region described on chromosome 15. Each SNORD (and potential SNORD) will have to be deleted from the genome, and the effect on ac^{4}C1337 observed by HPLC, as initially done by Sharma *et. al.* (**2015 and 2017**). Only then can it be determined if any of these four are the true human snR4 homologue.

5.6: Kre33p binding analysis

CLASH using Kre33-HTP did not produce many hybrids (unpublished work, T. Dudnakova). However, single RNA hits found in the experiment can still reveal the position of protein binding on crosslinked RNAs. Thus, pile-ups were created to indicate Kre33p-RNA binding sites.

The majority of Kre33p hits on *SNR4* were centred over the C box (Figure 5.6.1A). There was minimal binding across the rest of *SNR4* (<10 his per nucleotide), with the peak number at approximately 55 hits, near the 5' end of the snoRNA. This was unexpected as the region of conservation - which is also the region found in hybrids with the acetylation site - is much further downstream (~+170). It might have been anticipated that Kre33p would bind at this conserved region, so that it was optimally positioned to transfer the acetyl group to the correct nucleotide. It is possible, however, that the structure of snR4 brings this 5' region close to the conserved region, meaning that Kre33p is still in position to transfer the acetyl group to the correct nucleotide, but without binding directly to the guide sequence. It should be noted that the canonical box C/D proteins would protect the regions of bound RNA

from RNases, during the CLASH protocol, which would lead to the recovery of longer RNA fragments than the Kre33p protein-RNA contact site. Therefore, Kre33p may not bind to the C/D boxes, even if the recovered fragments include this site.

A
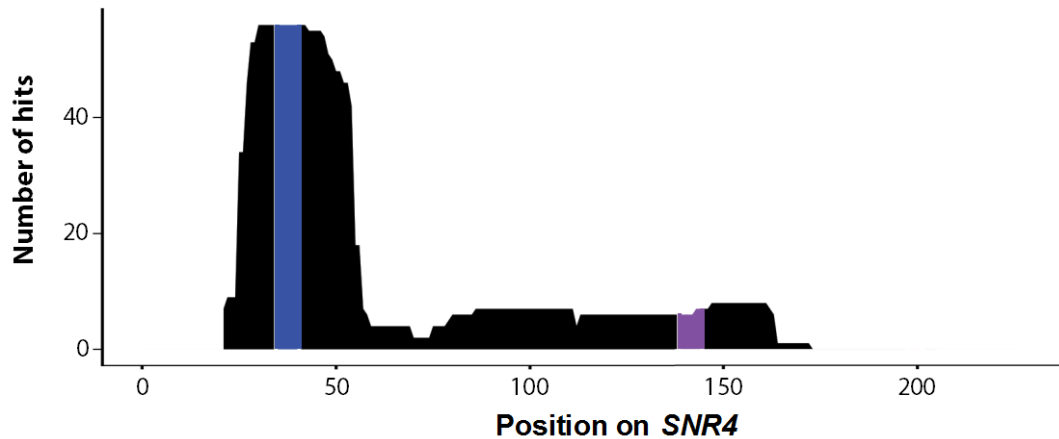
B
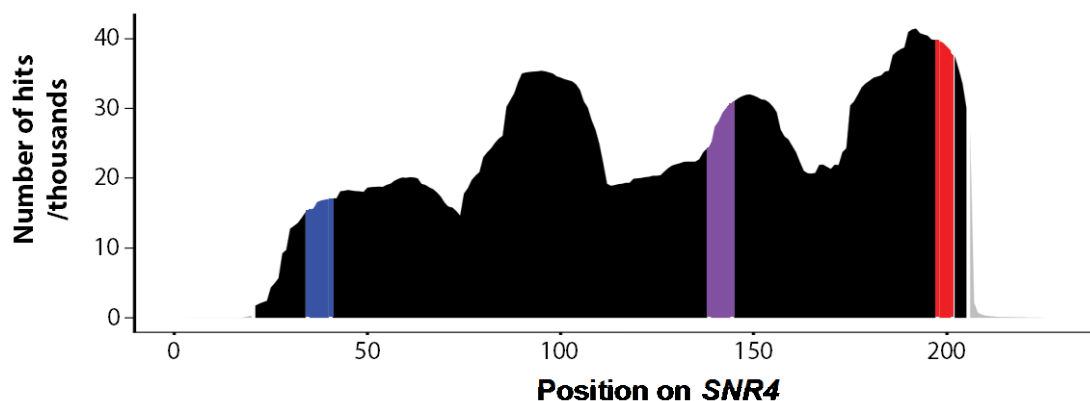
Figure 5.6.1 Single hit CLASH of Kre33p and Nop1p binding over *SNR4*. A) Kre33p binding sites. B) Nop1p binding sites. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Blue indicates the C box, mauve indicates the D' box, purple indicates C' box and red indicates D box. CLASH performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Nop1p CRAC and CLASH were optimised to yield good results, and thus a number of experiments have been performed using Nop1p. The Nop1p data for single hits over *SNR4* was aggregated onto one pile-up (Figure 5.6.1B). This image shows a

similar binding profile to the Nop1p CLASH data in Figure 5.3.2A, but with different peak ratios. The scale exceeds 40,000 hits, which is approximately 200-fold greater than the scale of snR4-rRNA hybrid hits. The highest peak in hybrid hits was over the C' box, whereas the single hit peaks over the predicted D' box and D box were higher than the C' box peak. This is expected, as Nop1p is well known to bind to the D' and D boxes (**van Nues *et. al.,* 2011**). Furthermore, the D box peak was much broader and the C box peak had a higher proportion of hits than in Figure 5.3.2A. Comparison of Kre33p hits over *SNR4* and Nop1p hits over *SNR4* revealed that the highest peak of Kre33p binding was the lowest peak of Nop1p binding and vice versa, as there was no Kre33p binding over the D box. This may indicate that Kre33p and Nop1p occupy distinct sites on snR4, possibly to avoid steric hindrance whilst still carrying out their functions. However, the linker ligation steps in the CLASH method did not include a de-capping step, which may have resulted in the small peak of Nop1p binding to the 5' region of *SNR4.*

Single hit CLASH data was also used to identify Kre33p binding sites on ribosomal RNA. Kre33p bound directly around C1280 in 18S rRNA (Figure 5.6.2A). Within the 100bp range around this site, there were few other hits, with the highest being 3-fold smaller than the peak around C1280. The centre of the peak lay upstream of the acetylated cytosine. This is expected, as the hybrid forms with the nucleotides upstream of C1280 (Figure 5.4.1A). The peak also overlapped a known methylation site. Similarly, Nop1p binds at C1280, with the highest number of hits upstream of the acetylated cytosine (Figure 5.6.2B). Given that both Nop1p and Kre33p are bound to snR4, it is expected that both proteins would have a peak at this site. However, many other snoRNAs also target this region of *RDN37* for methylation; hence Nop1p is bound at a number of sites within this 100bp range, resulting in multiple peaks.
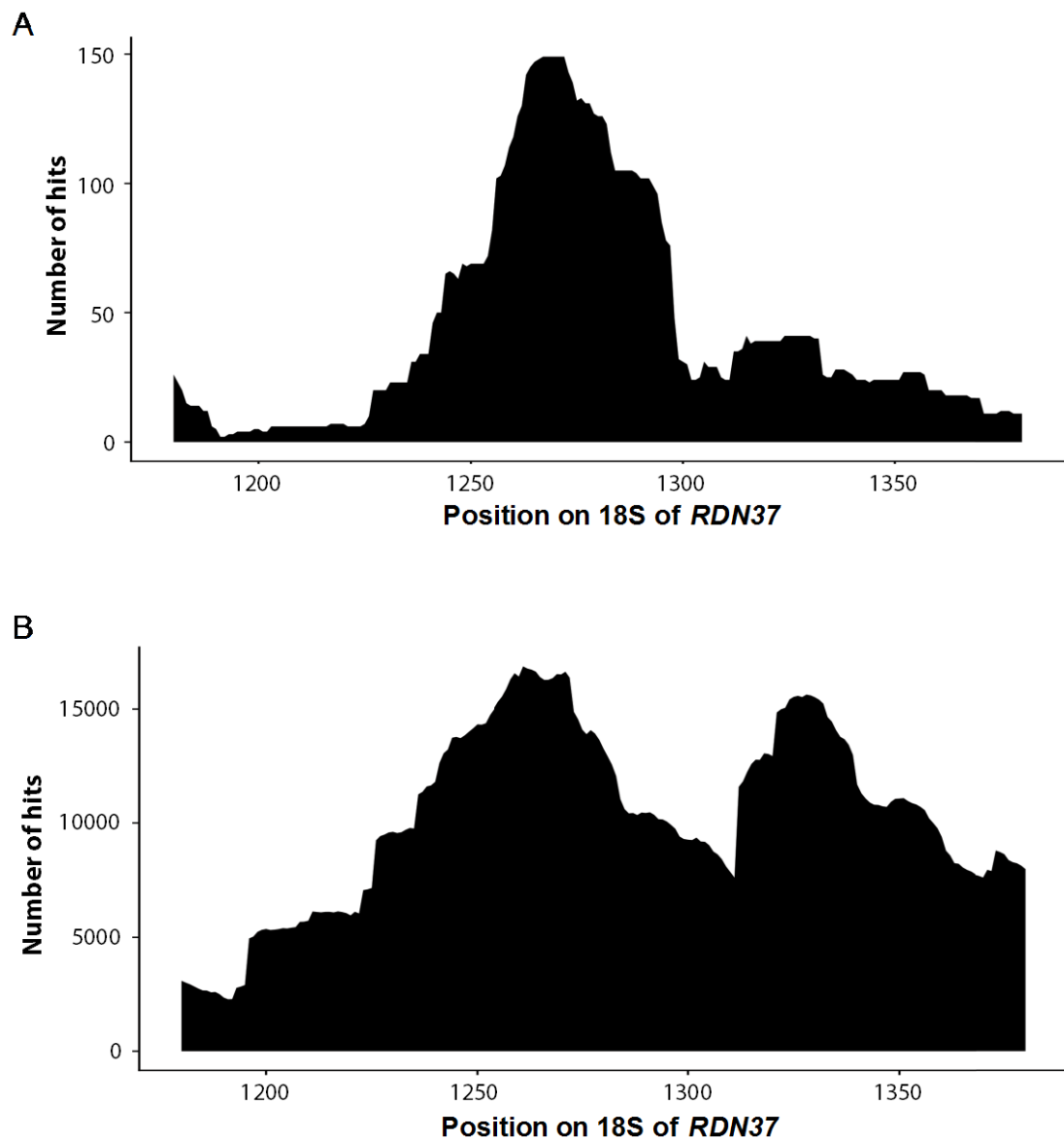
Figure 5.6.2 Single hit CLASH of Kre33p and Nop1p binding 100bp around 18S-C1280 site of 18S rRNA on _RDN37_. A) Kre33p binding sites. B) Nop1p binding sites. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. CLASH performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Similar to _SNR4_, the majority of Kre33p hits on _SNR45_ were centred over the C box (Figure 5.6.3A). However, there was also a peak 2-fold smaller over the D box. Furthermore, the peak present in _SNR4_ had approximately 55 hits, whereas the peaks in _SNR45_ had 4 and 2 hits, respectively. Such disparity between the Kre33p hits indicates that either coverage over RNAs is limited, due to the poor quality of the CRAC/CLASH, or that Kre33p may bind to snR45 less efficiently than to snR4.

Kre33p does not bind to the conserved region of *SNR45*, which lies upstream of +150 in Figure 5.6.3. As with *SNR4*, it is possible that the structure of snR45 brings the 5' end close to the conserved region, so that Kre33p can still transfer the acetyl group to the correct nucleotide without binding directly to the guide sequence. This then highlights the question of why Kre33p might bind upstream of the D box. It is possible that the limited single hit data does not represent the true binding sites of Kre33p on *SNR45*. Alternatively, the single hit data may falsely represent binding over *SNR4*, and did not cover sites upstream of the D box. Assuming the structure of canonical box C/D snoRNAs, the C box and D box would be brought close together with the adjacent sequences base-paired to form the termini stem. Therefore Kre33p may straddle the 5' and 3' end of snR45 (and possibly snR4) in order to function correctly.

The aggregated pile-up of Nop1p single hits over *SNR45* also showed a similar profile to snR45-rRNA CLASH hits in Figure 5.3.2B (Figure 5.6.3B). Again, the scale was much larger for the single hits, with the tallest peak consisting of ~70,000 hits, whereas the hybrid data had a peak of 80 hits. The ratio between peaks was very similar for both single hits and hybrid hits. The peak over the C' box contained the highest number of hits in both cases, though the peak upstream of the D box was slightly smaller for single hits. There was no peak upstream of the D' box, which differed from the binding profile for hybrid hits. The peak upstream of the D box is expected, as Nop1p is well-known to bind to the D box; however, it was much smaller proportional to the other peaks than was observed for *SNR4*. The absence of a peak upstream of the D' box was reciprocal to the Kre33p binding profile, as was observed for *SNR4*. This supports the proposal that Kre33p and Nop1p occupy distinct sites on the snoRNAs. This is not the case over the D box, where both Nop1p and Kre33p are bound. However, given the limited data available for Kre33p binding to snR45, a true picture of the interactions and precise binding sites cannot be built.
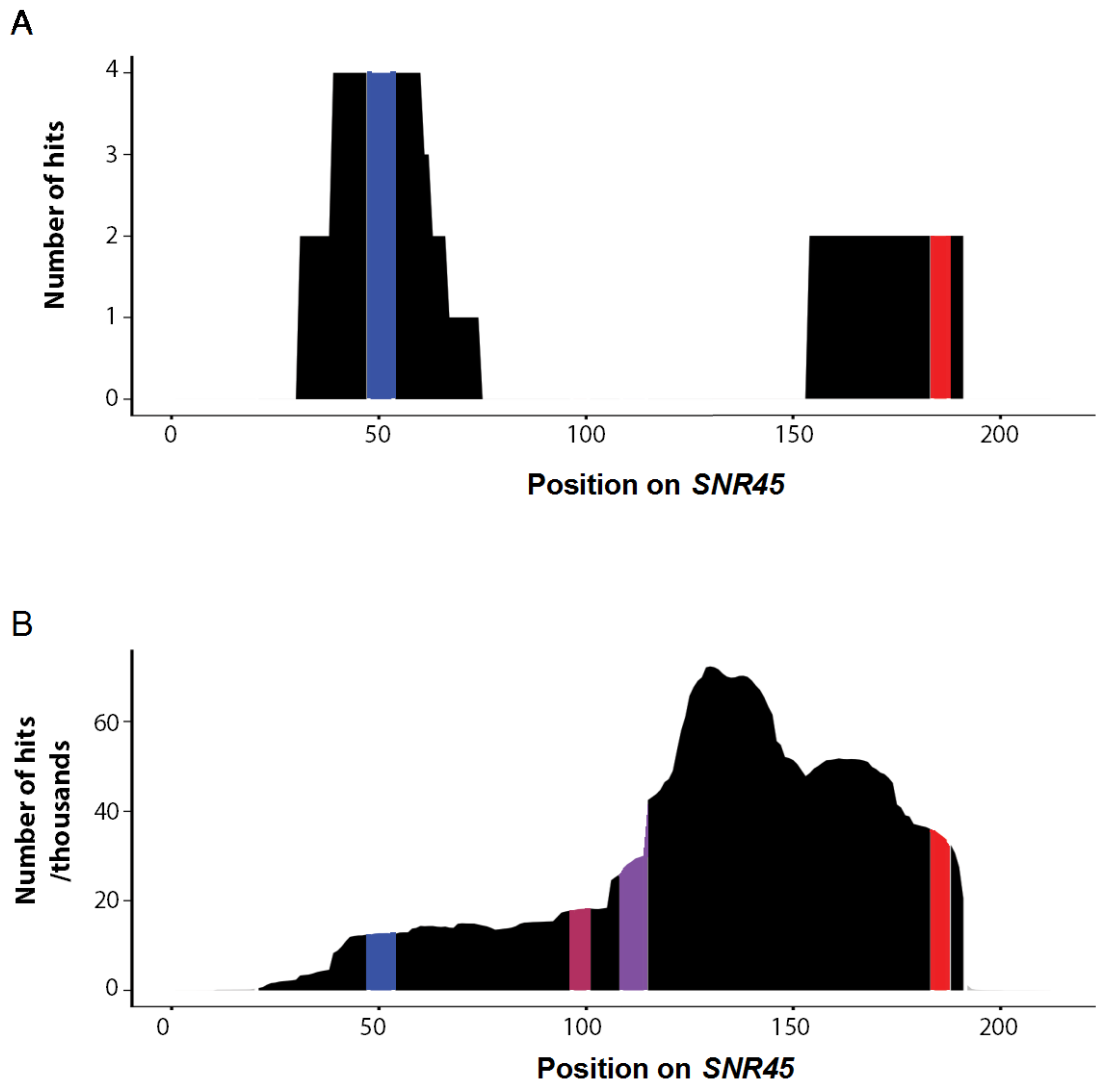
Figure 5.6.3 Single hit CLASH of Kre33p and Nop1p binding over *SNR45*. A) Kre33p binding sites. B) Nop1p binding sites. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Blue indicates the C box, mauve indicates the D' box, purple indicates C' box and red indicates D box. CLASH performed by T. Dudnakova. Plots produced by H. Dunn-Davies.

Single hit data was then analysed for binding at 18S-C1773. Kre33p binds directly around C1773 in 18S rRNA (Figure 5.6.4A). However, within the 100bp range around this site there was a larger peak upstream of C1773, approximately 2-fold bigger. The functionality of binding at this region is unclear. The peak at C1280 contained 100-150 hits, whereas the peak at C1773 contained approximately 30

hits. Again, this corroborates the earlier stated proportions between the snR4-C1280 interaction and the snR45-C1773 interaction. The centre of the peak lies upstream of the acetylated cytosine. This is predicted, as the hybrid forms with the nucleotides upstream of C1773 similar to snR4-C1280 (Figure 5.4.2A). Similarly, Nop1p binds at C1773, with approximately 5700 hits around this nucleotide (Figure 5.6.4B). Given that both Nop1p and Kre33p are bound to snR45, it is expected that both proteins would have a peak at this site. However, this is the only Nop1p peak within 100bp of C1773, with all other nucleotides having >2-fold fewer hits. This is very dissimilar to Nop1p binding around C1280 (Figure 5.6.2B). This correlates with the lack of methylation sites within 100bp of C1773. Furthermore, comparison of Figure 5.6.4 with Figure 5.6.2 reveals that high binding of Nop1p at either side of the acetylated residue occurs alongside low binding of Kre33p, and vice versa.

There are only two known acetylated nucleotides in *RDN37* – 18S-C1280 and 18S-C1773 (**Sharma *et. al.,* 2015**). Kre33p is the acetyltransferase catalysing this reaction. It is also required for export of 40S subunits, for which acetylation activity is not required (**Grandi *et. al.,* 2002; Sharma *et. al.,* 2017**). Figure 5.6.4A shows Kre33p binding to sites surrounding the acetylated nucleotides. This poses the questions of where else on *RDN37* does Kre33p bind to, and to what extent are these interactions functional. Therefore, all single hits from the Kre33p CLASH data were mapped onto *RDN37*, as shown in Figure 5.6.5. The x axis denotes the position along *RDN37*, and the y axis denotes how many CLASH single hits were identified for each nucleotide. There was a large peak at +1803 (1103nt of 18S rRNA) and a number of other peaks throughout 18S rRNA. This peak did not overlap with known methylation sites, but is in close proximity to the snR41-guided methylation site at 18S-1126. Notably, the 18S-1773 rRNA acetylation site (+2473 of *RDN37*), had a comparatively small peak. As earlier discussed, the snR45-C1773 interaction had a much smaller number of hybrids and individually much fewer hits than the snR4-C1280 interaction. There were also five peaks in 25S rRNA that contained >100 hits. There are no acetylated residues in 25S rRNA as analysed by HPLC, and Kre33p is not required for 25S rRNA synthesis, meaning these peaks were not expected. It is possible that Kre33p has an additional function that would explain its binding to other sites in *RDN37* that do not require acetyl-group transference or SSU export.
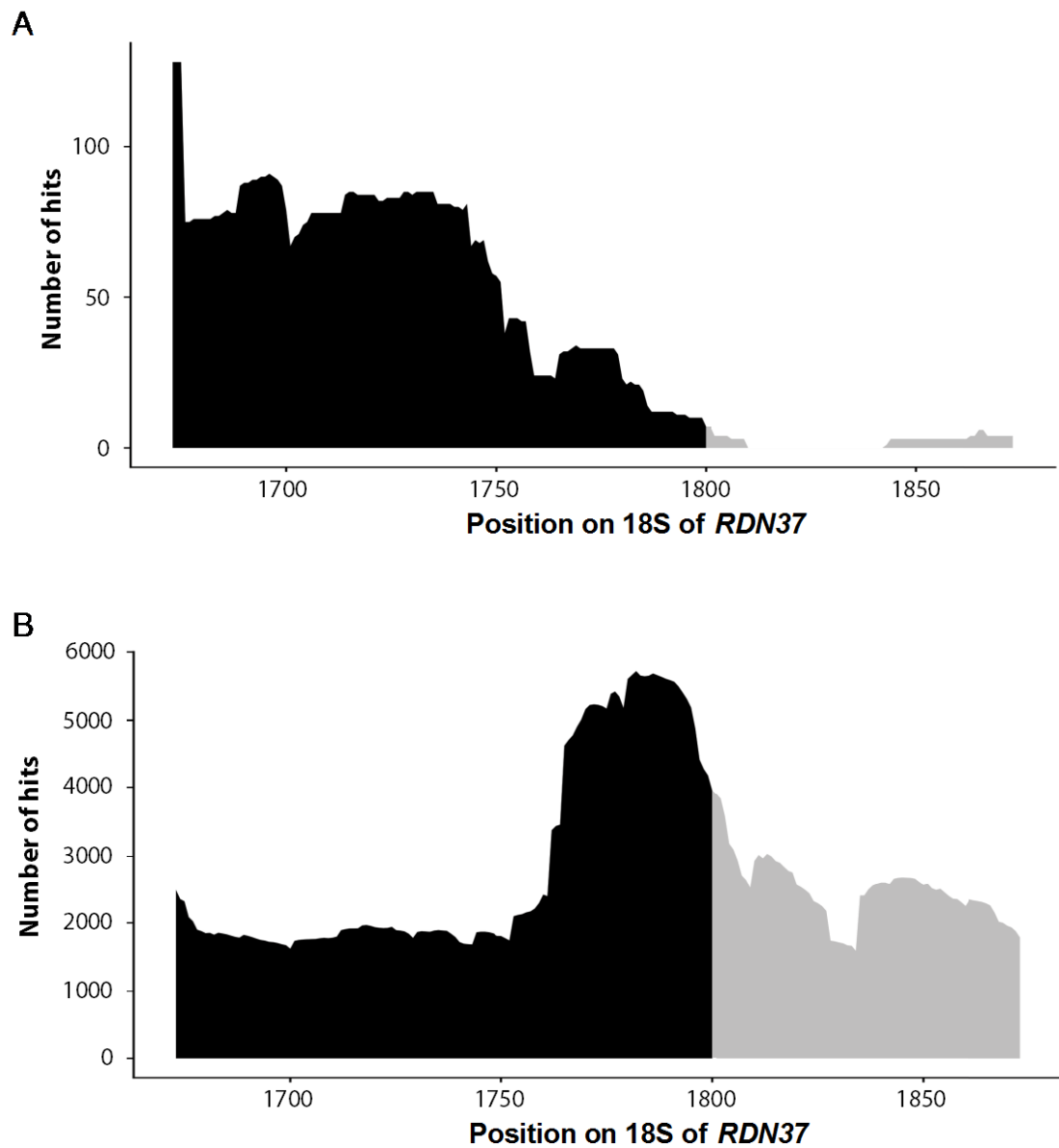
Figure 5.6.4 Single hit CLASH of Kre33p and Nop1p binding 100bp around 18S-C1773 site of *RDN37*. A) Kre33p binding sites. B) Nop1p binding sites. X axis denotes the position on the gene. Y axis indicates the number of CLASH hybrid hits identified for each nucleotide. Black: 18S rRNA; grey: ITS1. CLASH performed by T. Dudnakova. Plots produced by H. Dunn-Davies.
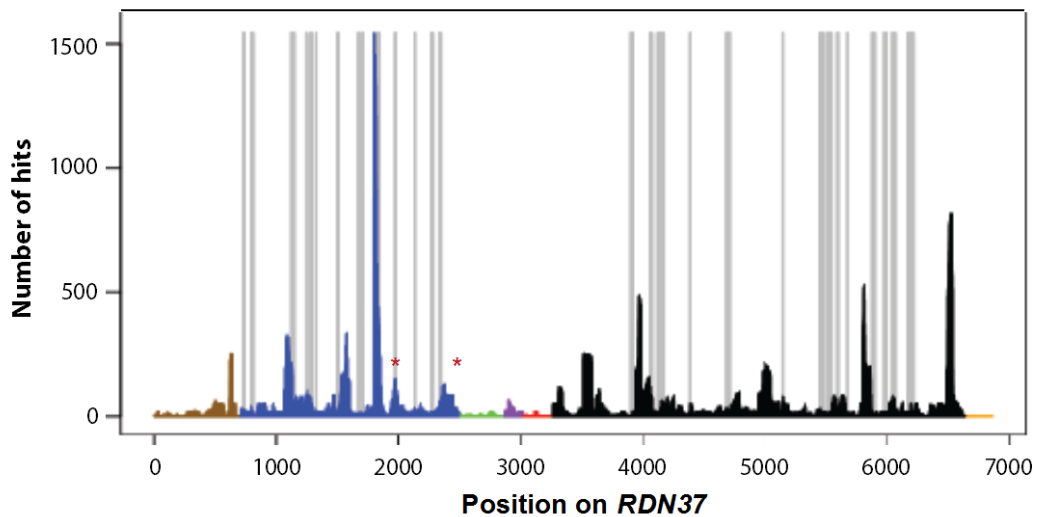
Figure 5.6.5 Single hit CLASH of Kre33p binding over *RDN37*. X axis denotes the position on the gene. Y axis indicates the number of CLASH single hits identified for each nucleotide. Brown: 5' ETS, blue: 18S rRNA, green: ITS1, purple: 5.8S rRNA, red: ITS2, black: 25S rRNA, orange: 3' ETS. Grey bars indicate known methylation sites. Red asterisks indicate the two acetylation sites, at +1980 and +2473. CLASH performed by T. Dudnakova. Plot produced by H. Dunn-Davies.

Kre33p binds to sites on *RDN37* that are not known to contain acetylated residues. Furthermore, Kre33p has been shown to interact with other snoRNAs, in addition to snR4 and snR45 (**Sharma *et. al.*, 2017**). CLASH analysis did not identify snoRNA-mRNA hybrids (including with snR4 and snR45), nor snoRNA-tRNA hybrids. One snoRNA-snoRNA hybrid was discovered, between snR64 and snR128. Selecting for snoRNA-rRNA hybrids, four interactions were found. One was the snR4-C1280 interaction (Figure 5.6.6A). Another was between snR9 and +4759-4795 of *RDN37* (1508-1544nt of 25S rRNA). snR9 is a box H/ACA snoRNA that guides the pseudouridylation of U2340 of 25S rRNA. The site bound by Kre33p was between the two box motifs, and was very distinct from the guide sequence (Figure 5.6.6B). Eight of the hybrids were with snR30: four with site +5791-5822 (2540-2567nt of 25S, Figure 5.6.6C), and four with site +1760-1780 (1060-1080nt of 18S, Figure 5.6.6D). snR30 is also a box H/ACA snoRNA, functioning in 18S synthesis (chapter 1.1), with no known modification target. Both hybrid sites were bound by a sequence part-way between the two box motifs. Four hybrids were found between snR40 and site +1254-1272 (554-572nt of 18S, Figure 5.6.6.E). snR40 is a box C/D snoRNA,

and targets G1271 of 18S and U898 of 25S rRNA. It has recently been discovered to also target G562 of 18S rRNA, which is the site identified in Figure 5.6.6E (**Yang et. al., 2015**). The snR40 sequence identified in the hybrid is the 5' region of the snoRNA up to the D' box (the 3' C of the snR40 sequence is the first nucleotide of the D' box). It hybridises using the same base-pairs as used to target the 18S-G1271 site. Future analyses may reveal the functional relevance of Kre33p binding to these RNAs.

Figure 5.6.6 Other targets of Kre33p from CLASH. A) snR4 hybrid with 18S-C1280. B) snR9 hybrid with 25S rRNA. C) snR30 hybrid with 25S rRNA. D) snR30 hybrid with 18S rRNA. E) snR40 hybrid with 18S rRNA. Acetylated cytosine is coloured in red, and denoted by an arrow and labelling. Watson-Crick base-pairs are denoted by a blue straight line, G-U base-pairings denoted by a dot. Dashed purple line denotes the boundary between snoRNA and rRNA regions of the hybrid. Hybrids drawn from Kre33p CLASH, with base-pairing indicated by the ViennaD folding programme.

5.7: Discussion

The aim of this chapter was to characterise the interactions of snR4 and snR45 with rRNA. Examining the conservation between distantly related fungi revealed that both *SNR4* and *SNR45* are well conserved box C/D snoRNAs. However, both showed a pattern of conservation that was different from that of canonical box C/D snoRNAs - particularly in sequences flanking the box motifs - and contained a region of unpredicted conservation. During this work, snR4 and snR45 were shown to guide the acetylation by Kre33p of residues C1280 and C1773 in yeast 18S rRNA, respectively (**Sharma *et. al.,* 2017**). The unpredicted conserved regions were subsequently identified as the guide region for rRNA acetylation. The conservation of these sequences therefore reflects an important biological function for these snoRNAs. Identifying this function for the two snoRNAs removes their 'orphan' status, and suggests a novel potential function for the other orphan snoRNAs in humans.

The binding profiles of snR4 and snR45 to rRNA and mRNA were similar. However, snR45 had a second rRNA binding site upstream of its D box. During carbon shift, the binding site downstream of the C' box formed significantly fewer hybrids, as did the same region of snR4. Incidentally, these two peaks include the acetylation guide sequences. Acetylation is not an essential rRNA modification (cells are fully viable without it, **Sharma *et. al.,* 2015**); therefore, yeast may limit the interaction guiding this modification in order to preserve more energy following nutrient deprivation. Alternatively, reduced acetylation may fine tune the activity of the ribosome under these conditions. The peak over the potential D' box of snR4 was bound to a region of 5.8S rRNA. The peak over the D box of snR45 was bound to the largest peak in 25S rRNA. Both of these interactions were preserved under carbon shift. The snR4 site in 5.8S rRNA is part of the stem that makes expansion segment ES3, while the snR45 site in 25S rRNA is part of the short stem in ES39. As all other interactions were diminished following growth in poor carbon source, these two interactions may be functionally significant. It is possible that the snoRNAs bind to these sites to stabilise/facilitate the rRNA folding/interactions, which may make ribosome biogenesis more efficient or consume less energy under nutrient deprivation.

The Entien lab discovered that U13 is the human homologue of snR45. However, no human homologue has been discovered for snR4. This chapter aimed to identify

possible human homologues and confine these to a reasonable number to experimentally validate. BLAST revealed *SNR4* conservation with regions of *Trypanosoma brucei gambiense* and *Drosophila melanogaster.* However, there were no closely located, predicted D boxes. Analysis on the human CLASH data around C1337 showed SNORD65 and SNORD117 had the potential to be the snR4 homologue. However, both had their caveats. SNORD117 (U83) has also been shown to direct site-specific methylation of an artificial substrate, consistent with the positioning of the guide sequence with respect to the D box (**Jády and Kiss, 2000**). No such experiment has been performed with snR4, but the positioning of its guide sequence makes this function unlikely. Further analysis revealed that a region of chromosome 9 and a region of chromosome 15 showed high conservation to the snR4 guide region, and both had the potential to be an unidentified box C/D snoRNA. These regions must be further investigated as to whether they are in fact unidentified box C/D snoRNAs, which could be assessed by northern hybridisation. In order to functionally validate these four potential homologues, each would be deleted from the genome. Phenotyping should reveal small to no change in growth rate compared to WT. As deletion of *SNR4, SNR45* and *SNORD13* confer no clear growth phenotypes, it would be difficult to determine whether deletion of the potential homologue showed no phenotype because it was indeed a homologue, or because its deletion was redundant. HPLC could then be performed to determine if this abolished acetylation of C1337. Only if this stipulation is met, could the snoRNA be definitively identified as the human snR4 homologue.

Single hit data showed Kre33p was bound to the 5' ends of both snR4 and snR45, and also the 3' end of snR45. The Entien group observed Kre33p binding to the 5' and 3' ends of snR4, and the 3' end of snR45 (**Sharma *et. al.,* 2017**). This would suggest that Kre33p is bound to both the 5' and 3' ends of both snR4 and snR45. Given that the two snoRNAs guide the same function, it is logical that they are both bound in the same way by the same protein, and that neither experiment had sufficient data to clearly show this. Mutation of the 5' ends of snR4 and snR45 (GS1) resulted in abolition of acetylation at C1280 and C1773, respectively (**Sharma *et. al.,* 2017**). The high level of Kre33p binding at this region of the snoRNA supports this result that GS1 is important for guiding transfer of the acetyl group. It is unclear why there were no hybrids surrounding the conserved guide region downstream of the C' box (GS2) (Figures 5.6.1A and 5.6.3A), especially considering that mutation of this sequence conferred an 88% and 82% reduction of acetylation in *snr4* and

snr45 mutant strains, respectively (**Sharma *et. al.,* 2017**). The Entien group performed mutations of other regions of snR4 and snR45 which also reduced or abolished acetylation. This indicates that the whole structure/sequence of the snoRNA is important to facilitate acetylation, not just the guide sequences. However, they identified the D' box of snR45 as 'CGAA' at 53-56nt, whereas our bioinformatic analysis identifies 'CUGA' at 76-79nt, which is the canonical D/D' box sequence. Variation can be found in D' boxes; however, analysis of *S. cerevisiae* D' box sequence conservation indicates a 'CUGA' motif is much more likely than 'CGAA' (**van Nues *et. al.,* 2011**). An *snr45* mutant in their study contained a mutated D' box (assuming the true D' box is the 'CUGA') with mutations in the nucleotides between the D' and C' boxes. This would not only affect base-pairing within the snoRNA and potentially alter the structure, but also potentially affect the binding of Nop56p and Nop1p (**van Nues *et. al.,* 2011**). Such changes to the snoRNP would be very likely to affect its function and stability. Furthermore, another *snr45* mutant altered the highly conserved region immediately upstream of the D' box (Figure 5.2.1B). This resulted in 0% acetylation of C1773. Whether this is due to alteration of the structure of the snoRNA or its function, it would be valuable to understand why this sequence is also highly conserved.

Kre33p had a wide distribution of binding sites across *RDN37* (Figure 5.6.5). The main peaks were found in 18S and 25S rRNA, but peaks were also found in the 5' ETS, and a small number in 5.8S rRNA. Interestingly, the peaks at C1280 and C1773 comprised a very small fraction of total hits. Given the importance of Kre33p binding at these sites, it raises the question of why such a small ratio of binding is observed here, compared to other sites which currently have no known significance. The Entien group found only two acetylated residues across pre-rRNA, both in 18S rRNA. This suggests that Kre33p has an additional function or additional rRNA targets, in addition to the 18S acetylation sites. The helicase mutant, kre33-K289A, reduced 18S acetylation by 90% and altered snoRNA association with pre-ribosomes compared (**Sharma *et. al.,* 2017**). snR40 and snR55 guide methylation of target in close proximity to the snR4 binding site. Based on snoRNA sedimentation profiles, it was proposed that Kre33p actively facilitates binding of snR4, snR45 and snR55 to their rRNA targets. However, the sedimentation profile of snR40 indicated that Kre33p is important for its release instead of facilitating its binding. Kre33p-snR40 interactions were also found in our CLASH data (Figure 5.6.6E). Due to the reduction of 2'-*O*-methylation observed in the kre33-K289A mutant, the hypothesis

was thus that Kre33p triggers release of snR40 before facilitating association of snR4 and/or snR55. In Figure 5.6.5, there was a large peak of Kre33p binding present at 18S-1103. This is 23nt upstream of the snR41-guided methylation site at 18S-1126. It is therefore possible that Kre33p binds strongly at this site in order to facilitate snR41 binding to its target sequence. Furthermore, the Kre33p binding observed upstream of the 18S-C1773 site may similarly facilitate efficient snR45 binding at the correct site. Nop1p also showed high distribution of binding across *RDN37,* in hybrids with both snR4 and snR45 (Figure 5.3.5). None of these peaks overlapped with known methylation sites, though a number were in close proximity. The two proteins may therefore interact at more than just the acetylation site to contribute to efficient ribosome processing and modification. Further work will be needed to fully explore snoRNA binding and release facilitated by Kre33p, and the function of additional interactions of snR4 and snR45 with rRNA.

The question of why ribosomes need acetylation is still unanswered. Lack of single modifications has little effect on ribosome synthesis or function; however, lack of multiple modifications in specific regions of the ribosome has severe effects on growth rate and translational fidelity (as discussed in chapter 1.1). The majority of 2'-*O*-methylations are transferred co-transcriptionally in early biogenesis (**Kos and Tollervey, 2010**). However, the two acetylations are thought to be a late event during ribosome biogenesis, due to NAT10p and U13 association with late forms of SSU precursors. Yeast lacking either or both acetylations are viable (**Sharma et. al., 2015**). It is suggested that C1280 acetylation might play a role in translation accuracy, due to its proximity to the mRNA entry channel, whereas C1773 is very close to the decoding site, where the mRNA and tRNAs interact (**Sharma et. al., 2015**). However, the exact function of this modification is unclear, with the current hypothesis being that it enforces conformational constraints on the RNA. Given the positioning of all modifications within the core of the ribosome rather than the periphery, the ongoing hypothesis is that rRNA modifications contribute to rRNA folding for efficient ribosome function.

Finally, it is puzzling that there were a higher number of hybrid hits with snR4 than with snR45, including at the respective acetylation sites. This is also true of the Kre33p single hits with both the snoRNAs and the respective acetylation sites. Both snR4 and snR45 have similar expression levels. snR45 is more highly conserved with distantly related fungi (Figure 5.2.1) and has a known human homologue

(Figures 5.4.2 and 5.4.3). The acetylation it guides is also only a few nucleotides upstream from the two most highly conserved modified nucleotides, A1781 and A1782 of 18S rRNA. It would therefore be logical for this to be the more abundant interaction, due to its conservation and the importance of the target site. Given the number of CLASH experiments performed using Nop1p, the reason for the fraction of coverage over snR45 compared to snR4 is not due to lack of data. Therefore, the snR45-C1773 interaction may occur much less frequently than snR4-C1280. As there are no methylation sites within 100bp of C1773, the snoRNA may have little to no competition with other snoRNAs in that area. The interaction would not need to be maintained once the acetyl group had been transferred, and thus would be captured less frequently. C1280 is surrounded by a number of other modified sites, however. It is possible that snR4 is important for the timely modification of the other sites (assisted by its Kre33p interaction) and thus is found in more hybrids, because it is bound to the substrate for longer.

This chapter has explored the novel function of snR4 and snR45 interaction with ribosomal RNA. It has revealed a multitude of previously unknown interactions, and analysed the details of the recently discovered function of guiding acetylation. Further work must now be done to better understand the mechanistic significance of these interactions, and ascertain whether these two snoRNAs play additional roles in ribosome biogenesis.

6 FINAL DISCUSSION

At the outset of this project, three main deficiencies were identified in the current knowledge within the field: the phenotype of deletion of orphan snoRNAs had not been extensively investigated; a global effect of snoRNA knockdown on target mRNAs in yeast had not been characterised; and a general mechanism by which orphan snoRNA binding alters target RNA pathways had not been determined. The data presented in this work addressed each of these points, in the process utilising three high-throughput screens: an SGA screen, RNA-sequencing and CLASH.

Investigation of the growth of *snr4Δ* and *snr45Δ* strains under many different growth and stress conditions identified no clear phenotypes (chapter 3.2). Furthermore, analysis of the double mutant also showed no evident growth phenotype, indicating that the two snoRNAs are not functionally redundant. However, during the course of this work it became clear that snR4 and snR45 function as the guides for rRNA acetylation of sites C1280 and C1773, respectively, in 18S rRNA. It had previously been shown that, for the majority of ribosomal modifications, removal of individual modifications has no evident effect on ribosome biogenesis or cell viability (**Parker et. al., 1988; Liang et. al., 2007; Liang et. al., 2009**). Consistent with this, the majority of snoRNAs are non-essential, and Sharma *et. al.* demonstrated that yeast lacking either or both acetylations are still viable (**Sharma et. al., 2015**). It is therefore logical that neither snR4 nor snR45 is essential. However, these acetylations are conserved between yeast and humans, and must optimise ribosome function in some way; if not, the cell would have no need to conserve them. It is possible that removal of the acetylations in combination with other ribosomal modifications would identify a phenotype and correlated function, as was shown for the loss of modifications in the PTC, the inter-subunit bridge and the decoding centre (**King et. al., 2003; Liang et. al., 2007; Liang et. al., 2009**). As the two acetyl groups are on residues in the mRNA entry channel and decoding site, respectively, it is likely that their presence assists in optimising translation.

Analysis by SGA screen highlighted the interaction of both snoRNAs with genes in microtubule pathways (chapter 3.3). It was hypothesised that snR4 and snR45 may coordinate spindle pole body duplication and M-phase checkpoints through their

interactions with Mps1p. It is also possible that they may coordinate cytokinesis and the exiting of mitosis, based on their interactions with *mob1* and *mob2* mutants. Notably, the human acetyltransferase, NAT10p, has been implicated in cytokinesis. During interphase, NAT10p is localised to the nucleolus – where it functions as an rRNA acetyl-transferase – but re-localises to the midbody (a dense structure containing microtubules, functioning in cytokinesis and mitotic completion) during telophase (**Shen *et. al.,* 2009**). Mutations of NAT10p have been shown to cause defects in nucleolar assembly, and affect the time taken to complete cytokinesis. It is also postulated to be the protein catalysing acetylation of α-tubulin, possibly improving microtubule stability. Furthermore, NAT10p has been linked to both cancer and laminopathies (defects in the nuclear-architecture proteins Lamin A and C). Hallmarks of these diseases are defects in chromosome organisation, misshapen nuclei and poor fitness. An inhibitor of NAT10p (or depletion of NAT10p) was found to rescue nuclear shape defects in Lamin A/C depleted cells and Hutchinson-Gilford progeria sydrome (HGPS) cells. This was due to reorganisation of the aberrantly organised microtubules mediated by NAT10p (**Larrieu *et. al.,* 2014**). Thus, interaction between NAT10p and microtubules has clearly been demonstrated. Whether the snoRNA-microtubule interactions observed in the SGA screen are linked to snoRNA association with the NAT10p homologue - Kre33p - remains to be determined.

An interaction was also observed between both snoRNAs and the *nop1-2* mutant, shown by severely impaired growth of both double mutants. Mapping of snR4 and snR45 CLASH hybrids onto rRNA showed that both snoRNAs bind to many sites within rRNA (Figure 5.3.5). CLASH hits recovered with Kre33p also showed multiple binding sites on rRNA (Figure 5.6.5). As discussed in chapter 5.7, snR4 and snR45 may have additional functions that are independent of guiding the acetylation of the two rRNA residues, such as facilitating the timely folding of the pre-rRNA and/or association with r-proteins or assembly factors. As both Nop1p and Kre33p are bound to snR4 and snR45, it is possible that they interact with each other to facilitate cooperative binding to the snoRNA. It is thus plausible that the regions affected by mutations within the *nop1-2* strain mediate Kre33p interaction or alternative snoRNA-rRNA function, which in turn could result in the processing defects observed by Tollervey *et. al.* (**1993**). rRNA analysis of the *nop1* mutant strains created in chapter 3.5 will elucidate the involvement of snR4 and snR45 in the synthetic lethal phenotype.

Further genetic interactions were identified with genes involved in actin pathways, cell cycle pathways, splicing pathways and translation pathways for both *snr4Δ* and *snr45Δ* in the SGA TSA screen. Targets of both snoRNAs showed a high degree of overlap, linking the functions of the two snoRNAs. Chapter 3.6 discussed the possibility of snR4 and snR45 facilitating the polymerisation of actin filaments and thus regulating cell polarisation, based on their interaction with *act1-2.* It was also hypothesised that snR4 and snR45 may negatively regulate the branching of actin filaments, based on their interactions with *arp2* and *arp3* mutants. Overlap was also observed for both snoRNAs with genes identified in the SGA DMA screen. Genes identified in the DMA screens involved in cell cycle pathways, ribosome biogenesis pathways and microtubule pathways link to those interactions observed in the TSA screen. However, a large proportion of genes identified in the SGA screen were not identified in more than one replicate. Additionally, a number of pathways showed unrelated functions, and a number of artefactual genes were identified in the DMA screen. This indicates that a large proportion of targets in the screen were unreliable. Furthermore, none of the interactions tested in the lab could be experimentally validated. Further experiments will be needed to test what proportion of the targets identified by the SGA screen is reliable.

CLASH analyses of snoRNA-RNA interactions supported the recent reports described in chapter 1.3, demonstrating that snoRNAs bind to mRNAs (chapter 4). Deletion of both *SNR4* and *SNR45* affected the expression levels of multiple mRNAs in both standard and stress conditions. The majority of RNAs in *snr4Δ* stress conditions decreased in expression level, in both RiboMinus treated samples and poly(A)+ selected samples. CLASH targets were identified within the dataset of mRNAs that showed differential expression. The observed number of CLASH targets that changed in expression level upon snoRNA deletion was statistically significant compared to the expected number, thus showing a correlation (Table 4.2.1). Analysis of mRNA binding sites over the snoRNA genes demonstrated that the orphan snoRNAs were unlikely to guide methylation of these targets (Figure 4.3.1). No direction of change was associated with any particular binding site over *SNR4,* though there was not enough data to assess any association of this kind for *SNR45.* Furthermore, mRNA targets were identified binding over snoRNA boxes, which are required for stability. While these fragments may be the ends of sequences cross-linked next to the boxes, it draws into question how many of the observed interactions are valid.

Six CLASH targets with differential expression in the RiboMinus treated sequencing were tested for validity by RT-qPCR; however, *RPS20* was the only high confidence target. Further experiments showed the fold change in *RPS20* as statistically insignificant, questioning its validity as an snR4 target. The expression levels of two mRNA CLASH targets in the poly(A)+ selected sequencing dataset (*ALD6* and *RCK1)* were reproducibly increased upon *SNR4* deletion, while the expression levels of two other mRNA CLASH targets (*ASN1* and *ARG4)* were reproducibly decreased upon *SNR45* deletion (chapter 4.4). These changes in abundance indicate that snoRNA binding has a role in mRNA stability and/or degradation, as proposed in chapter 1.3. Furthermore, the fold changes observed are comparable to those previously reported upon SNORD83B depletion (**Sharma *et. al.,* 2016**). Supporting these results, genes in related pathways had altered expression levels upon snoRNA deletion. These results suggested that snR4 and snR45 may regulate levels of acetate in the cell via regulation of the expression of genes related to ALD6. Further evidence suggested that snR4 and snR45 may regulate arginine biosynthesis by enhancing the expression of arginine biosynthesis enzymes, and reducing expression of arginine degrading enzymes (chapter 4.5). *ASN1* was the only CLASH target with altered expression level that was identified in the SGA screen, demonstrating the limited overlap between these three screens. However, only a small fraction of CLASH targets were identified among RNAs that showed altered expression, suggesting that altered RNA stability may not be the main consequence of snoRNA-mRNA interactions. Similarly, many RNAs with altered abundance in the sequencing data were not identified as CLASH targets, indicating that either these are indirect effects of snoRNA deletion or that CLASH did not saturate all snoRNA-mRNA interactions. How the apparent effects of snR4 and snR45 on mRNA stability are mediated must now be ascertained, including which other factors are implicated.

Chapter 4.5 discussed the possibility of snoRNA-mRNA interactions functioning to guide modifications of mRNA instead of affecting stability. Given that snR4 and snR45 guide Kre33p to acetylate rRNA, it is possible that they also function to guide acetylation of mRNAs. CLASH analysis provided a list of mRNA targets for both snoRNAs. Any mRNAs bound to the conserved acetylation guide sequence would be potential candidates for acetylation. There would be too little material to assess acetylation by HPLC and acetylation sites cannot currently be identified by sequence-based analysis, so a novel assay would need to be developed to

determine whether mRNAs are acetylated. No snoRNA-mRNA hybrids were identified by Kre33-HTP CLASH, including with snR4 and snR45. However, as discussed in chapter 5.4, the Kre33-HTP strain cross-linked very poorly, thus the experiment provided limited data. An optimised CLASH protocol could determine whether Kre33p has the potential to guide mRNA acetylation through snR4 or snR45. Such a modification could allow post-transcriptional regulation of gene expression, as has been demonstrated for other mRNA modifications (**reviewed in Hoernes *et. al.,* 2016**).

Both snoRNAs were highly conserved amongst fungi, showing a different pattern of conservation to canonical snoRNAs (chapter 5.2). The unpredicted region of high conservation was later identified as the acetylation guide region for Kre33p. While snR4, snR45 and snR55 (a canonical snoRNA) all formed the highest number of hybrids with rRNA, snR55 formed 7-fold more hybrids with rRNA than snR4, and 27-fold more than snR45. Analysis of rRNA binding sites over the snoRNA genes further demonstrated that the orphan snoRNAs were unlikely to guide methylation of rRNA targets (Figure 5.3.2). The binding profiles for snoRNA-rRNA interactions were very similar to snoRNA-mRNA interactions, though an additional peak was identified for snR45-rRNA interactions. Analysing snoRNA-rRNA interactions showed that most binding sites on the snoRNA corresponded to numerous sites on *RDN37,* demonstrating that there was no observable link between high peaks of snoRNA and rRNA binding. The ubiquity of snR4 and snR45 binding to rRNA, as well as the similarity in binding profiles between snoRNA-mRNA and snoRNA-rRNA interactions, casts doubts upon the validity of these interactions. It is possible that a proportion of observed hybrids are not true interactions. This is further supported by the identification of rRNA targets over snoRNA boxes (which are required for stability). These fragments may be regions protected by the canonical box C/D proteins rather than binding sites; however; these arguments indicate that snoRNA-mRNA and snoRNA-rRNA interactions cannot be presumed to be valid unless reproducibly identified in other screens or experiments, or found specifically linked under different conditions. Certain snoRNA-rRNA interactions identified under standard conditions were reduced upon carbon shift, while others were preserved, demonstrating a specific link under stress conditions. Among the nutritionally-impacted interactions were the acetylation binding site on the snoRNA and the acetylated sequence in the rRNA. Further work will be needed to measure the functional significance of these observations.

The work presented in this study has addressed each of the three aims set out in chapter 1.6, providing interesting new insights into the mechanisms of orphan snoRNA interactions. Together with other recent work in the field, these data provide an improved comprehension of the targets and functions of these snoRNAs. As knowledge of these interactions is developed, studies on human diseases can be further informed and enhanced, allowing for better treatment of patients in the future.

REFERENCES

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399–5410.

Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.-L., Tapsin, S., Chan, Y.-S., Tan, C.-P., Sim, A.Y.L., et al. (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. Mol. Cell *62*, 603–617.

Bachellerie, J.-P., Michot, B., Nicoloso, M., Balakin, A., Ni, J., and Fournier, M.J. (1995). Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. Trends in Biochemical Sciences *20*, 261–264.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. Cell *86*, 823–834.

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010). Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Meth. Enzymol. *470*, 145–179.

Beacham, I.R., Schweitzer, B.W., Warrick, H.M., and Carbon, J. (1984). The nucleotide sequence of the yeast ARG4 gene. Gene *29*, 271–279.

Ben-Shem, A., Loubresse, N.G. de, Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. Science *334*, 1524–1529.

Bernstein, K.A., Gallagher, J.E.G., Mitchell, B.M., Granneman, S., and Baserga, S.J. (2004). The Small-Subunit Processome Is a Ribosome Assembly Intermediate. Eukaryot Cell *3*, 1619–1626.

Bilsland, E., Molin, C., Swaminathan, S., Ramne, A., and Sunnerhagen, P. (2004). Rck1 and Rck2 MAPKAP kinases and the HOG pathway are required for oxidative stress resistance. Molecular Microbiology *53*, 1743–1756.

Bohmann, K., Ferreira, J., Santama, N., Weis, K., and Lamond, A.I. (1995). Molecular analysis of the coiled body. J. Cell Sci. Suppl. *19*, 107–113.

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8*, 574–585.

Bortolin-Cavaillé, M.-L., and Cavaillé, J. (2012). The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Res. *40*, 6800–6807.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Res *39*, 675–686.

Briggs, M.W., Burkard, K.T.D., and Butler, J.S. (1998). Rrp6p, the Yeast Homologue of the Human PM-Scl 100-kDa Autoantigen, Is Essential for Efficient 5.8 S rRNA 3′ End Formation. J. Biol. Chem. *273*, 13255–13263.

Broach, J.R. (2012). Nutritional Control of Growth and Development in Yeast. Genetics *192*, 73–105.

Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol *10*, 1623–1630.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature *387*, 303–308.

Bussiere, C., Hashem, Y., Arora, S., Frank, J., and Johnson, A.W. (2012). Integrity of the P-site is probed during maturation of the 60S ribosomal subunit. J Cell Biol *197*, 747–759.

Cahill, N.M., Friend, K., Speckmann, W., Li, Z.-H., Terns, R.M., Terns, M.P., and Steitz, J.A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. EMBO J *21*, 3816–3828.

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. U.S.A. *97*, 14311–14316.

Chamberlain, J.R., Lee, Y., Lane, W.S., and Engelke, D.R. (1998). Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. Genes Dev *12*, 1678–1690.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism11Edited by J. Karn. Journal of Molecular Biology *284*, 975–988.

Chang, D.D., and Clayton, D.A. (1987). A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. EMBO J *6*, 409–417.

Chang, L.-S., Lin, S.-Y., Lieu, A.-S., and Wu, T.-L. (2002). Differential expression of human 5S snoRNA genes. Biochemical and Biophysical Research Communications *299*, 196–200.

Charette, M., and Gray, M.W. (2000). Pseudouridine in RNA: What, Where, How, and Why. IUBMB Life *49*, 341–351.

Chimnaronk, S., Suzuki, T., Manita, T., Ikeuchi, Y., Yao, M., Suzuki, T., and Tanaka, I. (2009). RNA helicase module in an acetyltransferase that modifies a specific tRNA anticodon. EMBO J *28*, 1362–1373.

Chu, S., Archer, R.H., Zengel, J.M., and Lindahl, L. (1994). The RNA of RNase MRP is required for normal processing of ribosomal RNA. Proc Natl Acad Sci U S A *91*, 659–663.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res *16*, 10881–10890.

Crabeel, M., Abadjieva, A., Hilven, P., Desimpelaere, J., and Soetens, O. (1997). Characterization of the Saccharomyces cerevisiae ARG7 gene encoding ornithine acetyltransferase, an enzyme also endowed with acetylglutamate synthase activity. Eur. J. Biochem. *250*, 232–241.

Dahlkvist, A., and Sunnerhagen, P. (1994). Two novel deduced serine/threonine protein kinases from Saccharomyces cerevisiae. Gene *139*, 27–33.

Darzacq, X., Jády, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. EMBO J *21*, 2746–2756.

Doe, C.M., Relkovic, D., Garfield, A.S., Dalley, J.W., Theobald, D.E.H., Humby, T., Wilkinson, L.S., and Isles, A.R. (2009). Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. Hum Mol Genet *18*, 2140–2148.

Dong, X.-Y., Rodriguez, C., Guo, P., Sun, X., Talbot, J.T., Zhou, W., Petros, J., Li, Q., Vessella, R.L., Kibel, A.S., et al. (2008). SnoRNA U50 is a candidate tumor suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. Hum Mol Genet *17*, 1031–1042.

Dong, X.-Y., Guo, P., Boyd, J., Sun, X., Li, Q., Zhou, W., and Dong, J.-T. (2009). Implication of snoRNA U50 in human breast cancer. J Genet Genomics *36*, 447–454.

Dragon, F., Gallagher, J.E.G., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., et al. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature *417*, 967–970.

Dunbar, D.A., Wormsley, S., Agentis, T.M., and Baserga, S.J. (1997). Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. Mol Cell Biol *17*, 5803–5812.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Molecular Cell *32*, 519–528.

Faber, A.W., Van Dijk, M., Raué, H.A., and Vos, J.C. (2002). Ngl2p is a Ccr4p-like RNA nuclease essential for the final step in 3'-end processing of 5.8S rRNA in Saccharomyces cerevisiae. RNA *8*, 1095–1101.

Falaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., Nevo, Y., Eyras, E., Sperling, R., and Stamm, S. (2016). Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. Proc

Natl Acad Sci U S A *113*, E1625–E1634.

Fatica, A., Oeffinger, M., Dlakić, M., and Tollervey, D. (2003). Nob1p Is Required for Cleavage of the 3′ End of 18S rRNA. Mol Cell Biol *23*, 1798–1807.

Ferreira, H.J., Heyn, H., Moutinho, C., and Esteller, M. (2012). CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer. RNA Biol *9*, 881–890.

Ferreira-Cerca, S., Pöll, G., Gleizes, P.-E., Tschochner, H., and Milkereit, P. (2005). Roles of Eukaryotic Ribosomal Proteins in Maturation and Transport of Pre-18S rRNA and Ribosome Function. Molecular Cell *20*, 263–275.

Ferreira-Cerca, S., Pöll, G., Kühn, H., Neueder, A., Jakob, S., Tschochner, H., and Milkereit, P. (2007). Analysis of the In Vivo Assembly Pathway of Eukaryotic 40S Ribosomal Proteins. Molecular Cell *28*, 446–457.

Fox, M.J., Gao, H., Smith-Kinnaman, W.R., Liu, Y., and Mosley, A.L. (2015). The Exosome Component Rrp6 Is Required for RNA Polymerase II Termination at Specific Targets of the Nrd1-Nab3 Pathway. PLoS Genet *11*.

Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., and Bozzoni, I. (2002). Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2′-O-Methylation of Target RNA In Vitro. Mol Cell Biol *22*, 6663–6668.

Gallagher, J.E.G., Dunbar, D.A., Granneman, S., Mitchell, B.M., Osheim, Y., Beyer, A.L., and Baserga, S.J. (2004). RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. Genes Dev *18*, 2506–2517.

Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997a). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev. *11*, 941–956.

Ganot, P., Bortolin, M.-L., and Kiss, T. (1997b). Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. Cell *89*, 799–809.

García-Gómez, J.J., Fernández-Pevida, A., Lebaron, S., Rosado, I.V., Tollervey, D., Kressler, D., and de la Cruz, J. (2014). Final Pre-40S Maturation Depends on the Functional Integrity of the 60S Subunit Ribosomal Protein L3. PLoS Genet *10*.

Gartmann, M., Blau, M., Armache, J.-P., Mielke, T., Topf, M., and Beckmann, R. (2010). Mechanism of eIF6-mediated Inhibition of Ribosomal Subunit Joining. J Biol Chem *285*, 14848–14851.

Gee, H.E., Buffa, F.M., Camps, C., Ramachandran, A., Leek, R., Taylor, M., Patil, M., Sheldon, H., Betts, G., Homer, J., et al. (2011). The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. Br J Cancer *104*, 1168–1177.

Geerlings, T.H., Vos, J.C., and Raué, H.A. (2000). The final step in the formation of

25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. RNA *6*, 1698–1703.

Grandi, P., Rybin, V., Baßler, J., Petfalski, E., Strauß, D., Marzioch, M., Schäfer, T., Kuster, B., Tschochner, H., Tollervey, D., et al. (2002). 90S Pre-Ribosomes Include the 35S Pre-rRNA, the U3 snoRNP, and 40S Subunit Processing Factors but Predominantly Lack 60S Synthesis Factors. Molecular Cell *10*, 105–115.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. U.S.A. *106*, 9613–9618.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res *34*, 721–733.

Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y., and Caizergues-Ferrer, M. (2001). Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. Nucleic Acids Res *29*, 2733–2746.

Henry, Y., Wood, H., Morrissey, J.P., Petfalski, E., Kearsey, S., and Tollervey, D. (1994). The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. EMBO J *13*, 2452–2463.

Hernandez-Verdun, D. (1991). The nucleolus today. Journal of Cell Science *99*, 465–471.

Hoernes, T.P., Hüttenhofer, A., and Erlacher, M.D. (2016). mRNA modifications: Dynamic regulators of gene expression? RNA Biol *13*, 760–765.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J *19*, 1357–1365.

Hughes, J.M., and Ares, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. EMBO J *10*, 4231–4239.

Ikeuchi, Y., Kitahara, K., and Suzuki, T. (2008). The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. EMBO J *27*, 2194–2203.

Ito, S., Akamatsu, Y., Noma, A., Kimura, S., Miyauchi, K., Ikeuchi, Y., Suzuki, T., and Suzuki, T. (2014). A Single Acetylation of 18 S rRNA Is Essential for Biogenesis of the Small Ribosomal Subunit in Saccharomyces cerevisiae. J Biol Chem *289*, 26201–26212.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R., Meskauskas, A., Musalgaonkar, S.,

Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R., et al. (2011). rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. Mol Cell *44*, 660–666.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). THE MECHANISM OF EUKARYOTIC TRANSLATION INITIATION AND PRINCIPLES OF ITS REGULATION. Nat Rev Mol Cell Biol *11*, 113–127.

Jády, B.E., and Kiss, T. (2000). Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2′-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. Nucleic Acids Res *28*, 1348–1354.

Jády, B.E., and Kiss, T. (2001). A small nucleolar guide RNA functions both in 2′-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. EMBO J *20*, 541–551.

Jansen, R., Tollervey, D., and Hurt, E.C. (1993). A U3 snoRNP protein with homology to splicing factor PRP4 and G beta domains is required for ribosomal RNA processing. EMBO J *12*, 2549–2558.

Jones, G.E. (1978). L-Asparagine auxotrophs of Saccharomyces cerevisiae: genetic and phenotypic characterization. J. Bacteriol. *134*, 200–207.

Kempers-Veenstra, A.E., Oliemans, J., Offenberg, H., Dekker, A.F., Piper, P.W., Planta, R.J., and Klootwijk, J. (1986). 3'-End formation of transcripts from the yeast rRNA operon. EMBO J *5*, 2703–2710.

King, T.H., Liu, B., McCully, R.R., and Fournier, M.J. (2003). Ribosome Structure and Activity Are Altered in Cells Lacking snoRNPs that Form Pseudouridines in the Peptidyl Transferase Center. Molecular Cell *11*, 425–435.

Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science *311*, 230–232.

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Hum Mol Genet *19*, 1153–1164.

Kiss, T., Fayet-Lebaron, E., and Jády, B.E. (2010). Box H/ACA small ribonucleoproteins. Mol. Cell *37*, 597–606.

Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. Cell *85*, 1077–1088.

Kiss-László, Z., Henry, Y., and Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. EMBO J *17*, 797–807.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol *17*, 909–915.

Koš, M., and Tollervey, D. (2010). Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. Mol Cell *37*, 809–820.

Kressler, D., Hurt, E., and Baβler, J. (2010). Driving ribosome assembly. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1803*, 673–683.

Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc. Natl. Acad. Sci. U.S.A. *108*, 10010–10015.

Kufel, J., Dichtl, B., and Tollervey, D. (1999). Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. RNA *5*, 909–917.

Lafontaine, D.L., and Tollervey, D. (1999). Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. RNA *5*, 455–467.

Lafontaine, D.L., and Tollervey, D. (2000). Synthesis and assembly of the box C+D small nucleolar RNPs. Mol. Cell. Biol. *20*, 2650–2659.

Lafontaine, D.L.J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. *12*, 527–537.

Larrieu, D., Britton, S., Demir, M., Rodriguez, R., and Jackson, S.P. (2014). Chemical inhibition of NAT10 corrects defects of laminopathic cells. Science *344*, 527–532.

Lebaron, S., Schneider, C., van Nues, R.W., Swiatkowska, A., Walsh, D., Böttcher, B., Granneman, S., Watkins, N.J., and Tollervey, D. (2012). Proof reading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. Nat Struct Mol Biol *19*, 744–753.

Lee, S.J., and Baserga, S.J. (1999). Imp3p and Imp4p, Two Specific Components of the U3 Small Nucleolar Ribonucleoprotein That Are Essential for Pre-18S rRNA Processing. Mol Cell Biol *19*, 5441–5452.

Li, H.D., Zagorski, J., and Fournier, M.J. (1990). Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in Saccharomyces cerevisiae. Mol Cell Biol *10*, 1145–1152.

Li, Z., Vizeacoumar, F.J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F.S., Min, R., Vandersluis, B., Bellay, J., Devit, M., et al. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat. Biotechnol. *29*, 361–367.

Liang, X., Liu, Q., and Fournier, M.J. (2007). rRNA Modifications in an Intersubunit Bridge of the Ribosome Strongly Affect Both Ribosome Biogenesis and Activity. Molecular Cell *28*, 965–977.

Liang, X., Liu, Q., and Fournier, M.J. (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. RNA *15*, 1716–1728.

Liao, J., Yu, L., Mei, Y., Guarnera, M., Shen, J., Li, R., Liu, Z., and Jiang, F. (2010). Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol Cancer *9*, 198.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Liu, B., and Fournier, M.J. (2004). Interference probing of rRNA with snoRNPs: A novel approach for functional mapping of RNA in vivo. RNA *10*, 1130–1141.

Liu, B., Liang, X., Piekna-Przybylska, D., Liu, Q., and Fournier, M.J. (2008). Mis-targeted methylation in rRNA can severely impair ribosome synthesis and activity. RNA Biol *5*, 249–254.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Longtine, M.S., McKenzie, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279.

Luca, F.C., and Winey, M. (1998). MOB1, an Essential Yeast Gene Required for Completion of Mitosis and Maintenance of Ploidy. Mol. Biol. Cell *9*, 29–46.

Luca, F.C., Mody, M., Kurischko, C., Roof, D.M., Giddings, T.H., and Winey, M. (2001). Saccharomyces cerevisiae Mob1p Is Required for Cytokinesis and Mitotic Exit. Mol Cell Biol *21*, 6972–6983.

Lygerou, Z., Allmang, C., Tollervey, D., and Séraphin, B. (1996). Accurate Processing of a Eukaryotic Precursor Ribosomal RNA by Ribonuclease MRP in Vitro. Science *272*, 268–270.

Mannoor, K., Liao, J., and Jiang, F. (2012). Small nucleolar RNAs in cancer. Biochim Biophys Acta *1826*.

Massenet, S., Bertrand, E., and Verheggen, C. (2016). Assembly and trafficking of box C/D and H/ACA snoRNPs. RNA Biol *14*, 680–692.

Meaden, P.G., Dickinson, F.M., Mifsud, A., Tessier, W., Westwater, J., Bussey, H., and Midgley, M. (1997). The ALD6 gene of Saccharomyces cerevisiae encodes a cytosolic, Mg2+-activated acetaldehyde dehydrogenase. Yeast *13*, 1319–1327.

Middelhoven, W.J. (1964). The pathway of arginine breakdown in Saccharomyces cerevisiae. Biochimica et Biophysica Acta (BBA) - General Subjects *93*, 650–652.

Mitchell, P., Petfalski, E., and Tollervey, D. (1996). The 3' end of yeast 5.8S rRNA is

generated by an exonuclease processing mechanism. Genes Dev. *10*, 502–513.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell *91*, 457–466.

Morrissey, J.P., and Tollervey, D. (1993). Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol Cell Biol *13*, 2469–2477.

Natarajan, K., Meyer, M.R., Jackson, B.M., Slade, D., Roberts, C., Hinnebusch, A.G., and Marton, M.J. (2001). Transcriptional Profiling Shows that Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast. Mol Cell Biol *21*, 4347–4368.

Nicholls, R.D., Saitoh, S., and Horsthemke, B. (1998). Imprinting in Prader-Willi and Angelman syndromes. Trends Genet. *14*, 194–200.

Niewmierzycka, A., and Clarke, S. (1999). S-Adenosylmethionine-dependent Methylation in Saccharomyces cerevisiae IDENTIFICATION OF A NOVEL PROTEIN ARGININE METHYLTRANSFERASE. J. Biol. Chem. *274*, 814–824.

Nizami, Z., Deryusheva, S., and Gall, J.G. (2010). The Cajal Body and Histone Locus Body. Cold Spring Harb Perspect Biol *2*.

van Nues, R.W., Granneman, S., Kudla, G., Sloan, K.E., Chicken, M., Tollervey, D., and Watkins, N.J. (2011). Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. EMBO J *30*, 2420–2430.

Oeffinger, M., Zenklusen, D., Ferguson, A., Wei, K.E., El Hage, A., Tollervey, D., Chait, B.T., Singer, R.H., and Rout, M.P. (2009). Rrp17p Is a Eukaryotic Exonuclease Required for 5′ End Processing of Pre-60S Ribosomal RNA. Mol Cell *36*, 768–781.

Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. FEBS Letters *514*, 17–25.

Ohmayer, U., Gamalinda, M., Sauert, M., Ossowski, J., Pöll, G., Linnemann, J., Hierlmeier, T., Perez-Fernandez, J., Kumcuoglu, B., Leger-Silvestre, I., et al. (2013). Studies on the assembly characteristics of large subunit ribosomal proteins in S. cerevisae. PLoS ONE *8*, e68412.

Ono, M., Yamada, K., Avolio, F., Scott, M.S., van Koningsbruggen, S., Barton, G.J., and Lamond, A.I. (2010). Analysis of Human Small Nucleolar RNAs (snoRNA) and the Development of snoRNA Modulator of Gene Expression Vectors. Mol Biol Cell *21*, 1569–1584.

Ono, M., Scott, M.S., Yamada, K., Avolio, F., Barton, G.J., and Lamond, A.I. (2011). Identification of human miRNA precursors that resemble box C/D snoRNAs. Nucleic Acids Res *39*, 3879–3891.

Paranjape, T., Slack, F.J., and Weidhaas, J.B. (2009). MicroRNAs: tools for cancer diagnostics. Gut *58*, 1546–1554.

Parker, R., Simmons, T., Shuster, E.O., Siliciano, P.G., and Guthrie, C. (1988). Genetic analysis of small nuclear RNAs in Saccharomyces cerevisiae: viable sextuple mutant. Mol Cell Biol *8*, 3150–3159.

Petfalski, E., Dandekar, T., Henry, Y., and Tollervey, D. (1998). Processing of the Precursors to Small Nucleolar RNAs and rRNAs Requires Common Components. Mol Cell Biol *18*, 1181–1189.

Pöll, G., Braun, T., Jakovljevic, J., Neueder, A., Jakob, S., Woolford, J.L., Tschochner, H., and Milkereit, P. (2009). rRNA Maturation in Yeast Cells Depleted of Large Ribosomal Subunit Proteins. PLoS One *4*.

Ramani, V., Qiu, R., and Shendure, J. (2015). High-throughput determination of RNA structure by proximity ligation. Nat Biotechnol *33*, 980–984.

Richard, P., Darzacq, X., Bertrand, E., Jády, B.E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. EMBO J *22*, 4283–4293.

Rife, J.P., and Moore, P.B. (1998). The structure of a methylated tetraloop in 16S ribosomal RNA. Structure *6*, 747–756.

Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat. Genet. *40*, 719–721.

Schimmang, T., Tollervey, D., Kern, H., Frank, R., and Hurt, E.C. (1989). A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. EMBO J *8*, 4015–4024.

Schmitt, M.E., and Clayton, D.A. (1993). Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in Saccharomyces cerevisiae. Mol Cell Biol *13*, 7935–7941.

Scott, M.S., Avolio, F., Ono, M., Lamond, A.I., and Barton, G.J. (2009). Human miRNA Precursors with Box H/ACA snoRNA Features. PLoS Comput Biol *5*.

Scott, M.S., Ono, M., Yamada, K., Endo, A., Barton, G.J., and Lamond, A.I. (2012). Human box C/D snoRNA processing conservation across multiple cell types. Nucleic Acids Res. *40*, 3676–3688.

Sharma, S., and Lafontaine, D.L.J. (2015). 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. Trends in Biochemical Sciences *40*, 560–575.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. Mol. Cell *62*, 618–626.

Sharma, S., Langhendries, J.-L., Watzinger, P., Kötter, P., Entian, K.-D., and Lafontaine, D.L.J. (2015). Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. Nucleic Acids Res *43*, 2242–2258.

Sharma, S., Yang, J., van Nues, R., Watzinger, P., Kötter, P., Lafontaine, D.L.J., Granneman, S., and Entian, K.-D. (2017). Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. PLoS Genet *13*.

Shen, Q., Zheng, X., McNutt, M.A., Guang, L., Sun, Y., Wang, J., Gong, Y., Hou, L., and Zhang, B. (2009). NAT10, a nucleolar protein, localizes to the midbody and regulates cytokinesis and acetylation of microtubules. Experimental Cell Research *315*, 1653–1667.

Sillevis Smitt, W.W., Vlak, J.M., Molenaar, I., and Rozijn, T.H. (1973). Nucleolar function of the dense crescent in the yeast nucleus: A biochemical and ultrastructural study. Experimental Cell Research *80*, 313–321.

Steitz, T.A. (2008). A structural understanding of the dynamic ribosome machine. Nature Reviews Molecular Cell Biology *9*, 242–253.

Strunk, B.S., Loucks, C.R., Su, M., Vashisth, H., Cheng, S., Schilling, J., Brooks, C.L., Karbstein, K., and Skiniotis, G. (2011). Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. Science *333*, 1449–1453.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature *519*, 491–494.

Szewczak, L.B.W., DeGregorio, S.J., Strobel, S.A., and Steitz, J.A. (2002). Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. Chem. Biol. *9*, 1095–1107.

Teste, M.-A., Duquenne, M., François, J.M., and Parrou, J.-L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. BMC Molecular Biology *10*, 99.

Thomson, E., and Tollervey, D. (2010). The Final Step in 5.8S rRNA Processing Is Cytoplasmic in Saccharomyces cerevisiae. Mol. Cell. Biol. *30*, 976–984.

Tollervey, D. (1987). A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. EMBO J *6*, 4169–4175.

Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. Cell *72*, 443–457.

Travis, A.J., Moody, J., Helwak, A., Tollervey, D., and Kudla, G. (2014). Hyb: A bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. Methods *65*, 263–273.

Turowski, T.W., and Tollervey, D. (2015). Cotranscriptional events in eukaryotic ribosome synthesis. WIREs RNA *6*, 129–139.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science *302*, 1212–1215.

Valleron, W., Ysebaert, L., Berquet, L., Fataccioli, V., Quelen, C., Martin, A., Parrens, M., Lamant, L., Leval, L. de, Gisselbrecht, C., et al. (2012). Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. Blood *120*, 3997–4005.

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. Annu. Rev. Genet. *33*, 261–311.

Venema, J., Vos, H.R., Faber, A.W., van Venrooij, W.J., and Raué, H.A. (2000). Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association. RNA *6*, 1660–1671.

Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaillé, J., and Huttenhofer, A. (2005). ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. J Cell Biol *169*, 745–753.

Volkmann, N., Amann, K.J., Stoilova-McPhie, S., Egile, C., Winter, D.C., Hazelwood, L., Heuser, J.E., Li, R., Pollard, T.D., and Hanein, D. (2001). Structure of Arp2/3 Complex in Its Activated State and in Actin Filament Branch Junctions. Science *293*, 2456–2459.

Walker, S.E., Zhou, F., Mitchell, S.F., Larson, V.S., Valasek, L., Hinnebusch, A.G., and Lorsch, J.R. (2013). Yeast eIF4B binds to the head of the 40S ribosomal subunit and promotes mRNA recruitment through its N-terminal and internal repeat domains. RNA *19*, 191–207.

Wang, C., and Meier, U.T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. EMBO J. *23*, 1857–1867.

Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.-H. (2000). Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 Å resolution. EMBO J *19*, 317–323.

Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Lührmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. RNA *4*, 1549–1568.

Watkins, N.J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000). A Common Core RNP Structure Shared between the Small Nucleoar Box C/D RNPs and the Spliceosomal U4 snRNP. Cell *103*, 457–466.

Watkins, N.J., Dickmanns, A., and Lührmann, R. (2002). Conserved Stem II of the Box C/D Motif Is Essential for Nucleolar Localization and Is Required, Along with the 15.5K Protein, for the Hierarchical Assembly of the Box C/D snoRNP. Mol Cell Biol *22*, 8342–8352.

Weiss, E.L., Kurischko, C., Zhang, C., Shokat, K., Drubin, D.G., and Luca, F.C. (2002). The Saccharomyces cerevisiae Mob2p–Cbk1p kinase complex promotes polarized growth and acts with the mitotic exit network to facilitate daughter cell–specific localization of Ace2p transcription factor. J Cell Biol *158*, 885–900.

Wiederkehr, T., Prétôt, R.F., and Minvielle-Sebastia, L. (1998). Synthetic lethal interactions with conditional poly(A) polymerase alleles identify LCP5, a gene involved in 18S rRNA maturation. RNA *4*, 1357–1372.

Wilson, D.N., and Doudna Cate, J.H. (2012). The Structure and Function of the Eukaryotic Ribosome. Cold Spring Harb Perspect Biol *4*.

Winey, M., and Weiss (1996). The Saccharomyces cerevisiae spindle pole body duplication gene MPS1 is part of a mitotic checkpoint. J Cell Biol *132*, 111–123.

Woolford, J.L., and Baserga, S.J. (2013). Ribosome Biogenesis in the Yeast Saccharomyces cerevisiae. Genetics *195*, 643–681.

Xu, G., Yang, F., Ding, C.-L., Zhao, L.-J., Ren, H., Zhao, P., Wang, W., and Qi, Z.-T. (2014). Small nucleolar RNA 113–1 suppresses tumorigenesis in hepatocellular carcinoma. Mol Cancer *13*.

Yoshihama, M., Nakao, A., and Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. BMC Res Notes *6*, 426.

Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L., and Carbon, J. (1999). Point Mutations in Yeast CBF5 Can Abolish In Vivo Pseudouridylation of rRNA. Mol Cell Biol *19*, 7461–7472.

Zhang, B., Pan, X., Cobb, G.P., and Anderson, T.A. (2007). microRNAs as oncogenes and tumor suppressors. Developmental Biology *302*, 1–12.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of Polycomb-associated RNAs by RIP-seq. Mol Cell *40*, 939–953.