UNIVERSITY OF TURKU

RNA-Decoder:

Development of a novel DNA-sequencing-based method to identify proteins binding to a single specific RNA transcript

Master's thesis

University of Turku
Department of Life Technologies
Molecular Biosciences, Biochemistry

May/2024

Beda Anttila

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

UNIVERSITY OF TURKU

Master's thesis

Subject: Molecular Biosciences, Biochemistry
Author: Beda Anttila
Title: RNA-Decoder: Development of a novel DNA-sequencing-based method to identify proteins binding to a single specific RNA transcript
Supervisor(s): PhD. Matti Turtola, PhD. Desire Garcia Pichardo and Prof. Sebastian Marquardt
Number of pages: 58 pages + 1 appendix
Date: 21.05.2024

Long non-coding RNAs (lncRNAs) constitute a diverse class of RNA-molecules that serve as important regulators of various cellular processes such as metabolism, despite not encoding proteins. Typically, lncRNAs do not act alone within cells; instead, they interact with RNA-binding proteins to form functional ribonucleoprotein complexes (RNPs). RNPs can alter gene expression through epigenetic mechanisms by modelling chromatin structure or affecting transcription or translation.

The aim of this study is to develop a new method called RNA-Decoder which focuses on identifying the RNA-binding proteins within RNPs that interact with protein coding RNA transcripts (mRNA) and lncRNA transcripts originating from a divergent transcription locus called GCG1-SUT098 in *S. cerevisiae*. RNA-Decoder utilizes Epi-Decoder yeast library which consists of thousands of yeast strains, each containing two unique DNA-barcodes at the locus of interest and a unique protein with a tandem immunopurification tag (TAP-tag). In RNA-Decoder method, RNPs formed in the *S. cerevisiae* cells were isolated by TAP-tag immunoprecipitation, and gathered RNA transcripts from isolated RNPs are transcribed into cDNA. cDNA was amplified by PCR to be able to sequence the barcoded regions. Sequencing results enable comparison between mRNA and lncRNA interacting RNA-binding proteins.

The preliminary results of RNA-Decoder indicate its potentiality, and functionality. Western blot -analysis indicates successful immunoprecipitation of RNPs. The correct sizes PCR products from barcoded region of cDNAs indicate successful cDNA synthesis and barcode amplification. However, PCR reaction targeting GCG1 barcode region would require more optimization due primer-dimer formation during the reaction. To validate RNA-Decoder method and to detect the differences between mRNA-protein and lncRNA-protein complexes the amplified barcode regions needs to be sequenced in the future.

Key words: long non-coding RNA, lncRNA, divergent transcription, ribonucleoprotein complex, RNP, immunoprecipitation, TAP-tag, DNA barcode, Saccharomyces cerevisiae

Table of contents

Abbreviations 3

1 Introduction 4

1.1 Long non-coding RNAs 5

1.2 Genomic context of long non-coding RNAs 6

1.3 Functional mechanisms of long non-coding RNAs 8

1.3.1 Regulation of chromatin structure 9

1.3.2 Regulation of transcriptional activity 10

1.3.3 Post-transcriptional gene regulation 11

1.4 Long non-coding RNAs in fungi 11

1.4.1 Meiosis and sporulation 13

1.4.2 Senescence and telomere maintenance 16

1.4.3 Nutrient metabolism 16

1.4.4 Stress response 21

1.4.5 Circadian rhythms 22

2 Aims of the study 24

3 Materials and methods 25

3.1 Epi-Decoder yeast library 25

3.2 Yeast cultivation 25

3.3 Cross-linking RNA-protein complexes 25

3.4 Cell lysis 25

3.5 DNA sonication and sample pre-clearing 26

3.6 Immunoprecipitation 26

3.7 Immunoprecipitation analysis 27

3.8 Cross-link reversal and RNA sample purification 28

3.9 cDNA synthesis 28

3.10 qPCR 28

3.11 Barcode amplification 29

3.12 Barcode library assembly 29

3.13 Artificial Intelligence 29

4 Results 30

4.1 Small-scale RNA-Decoder experiment 30

4.1.1 Epi-Decoder library yeast strain selection for method development 30

4.1.2 Selection of cross-linker for RNA-protein complexes 32

4.1.3 Dynabead A and Dynabead G comparison 35

4.1.4 Optimizing PCR reactions for barcode amplification 36

4.2 Big-scale RNA-Decoder experiment 39

4.2.1 Chromatin shearing 39

4.2.2 Immunoprecipitation 40

4.2.3 Barcode amplification and concentrations of assembled libraries 41

5 Discussion 45

5.1 Reflection of cross-linking and immunoprecipitation results 45

5.2 Reflection of barcode amplification results 47

6 Conclusion 49

References 51

Appendices 59

Appendix 1: Illumina primers for Barcode amplification 59

Abbreviations

bp = basepair

CUT = cryptic unstable transcript

IgG = Immunoglobulin G

IP = immunoprecipitated

lncRNA = long non-coding RNA

mRNA = messenger RNA

miRNA = microRNA

NAT = natural antisense transcripts

Pol II = RNA polymerase II

RBP = ribonucleobinding protein

RIP = RNA immunoprecipitation

RNP = ribonucleoprotein

rRNA = ribosomal RNA

RT = room temperature

SUT = stable unstable transcript

TAP tag = tandem affinity purification tag

TERRA = telomeric repeat-containing RNA

tRNA = transfer RNA

1 Introduction

During past decade perception of significance of pervasive transcription and importance of non protein coding RNA transcripts has become more and more evident. Pervasive transcription has been noted to have valuable role in various gene regulating processes instead of being just a transcriptional background noise. (Lybecker et al. 2014.) Previous studies have affirmed that, for example in the case of yeast, approximately 80 - 85 % of the genome is transcribed, around 70% of transcriptome being protein coding RNAs and almost 25 % of transcriptome non-coding RNAs (David et al. 2006; Juan Li et al. 2021; Parker et al. 2018). Long non-coding RNAs, exceeding 200 nucleotides in length, are distinguished by their inability to encode proteins. Despite this, they play significant regulator roles in many biological processes in cells (Quinn and Chang 2016; Yamashita et al. 2016).

In recent years modern biotechnological tools like high-throughput sequencing have multiplied the number of discovered lncRNAs. Nevertheless, the knowledge of the molecular functions of lncRNAs still remains limited. The most well-known roles of lncRNAs are their functional regulation roles in cellular processes. They are involved for example in meiosis control, nutrient starvation response, telomere maintenance, stress response and nutrient biosynthesis. (Li et al. 2021; Till et al. 2018; Yamashita et al. 2016.) In humans certain lncRNAs or disorders in their expression have also been connected in some neurodegenerative conditions including Alzheimer's disease, Parkinson’s disease and Huntington’s disease and even autism spectrum disorders (Johnson et al. 2010; Magistri et al. 2015; Ni et al. 2017; Ziats and Rennert 2013). Some lncRNAs have been linked to different cancers and tumor metastasis (Jiang et al. 2019).

lncRNAs may be utilized as biomarkers, therapeutic targets, and biotechnical tools in the future. For example PCAT-1 can be detected from the urine of prostate cancer patients, so it might have a function as a diagnostic marker of prostate cancer. (Prensner et al. 2011). A filamentous fungus, called *Trichoderma reesei* is used as a producer strain to produce cellulases in bioindustry and lncRNAs have been observed to enhance its expression of cellulaces and hemicellulases (Till et al. 2018).

Even though research of lncRNAs and knowledge regarding their functions and significance has increased remarkably in 2010s, our understanding of lncRNAs is still limited. Previous studies have focused on especially the roles of lncRNAs in mammals, but the research of fungal lncRNAs has remained aside in spite of them being linked to several specific regulatory roles in biological processes also in fungi. (Davati and Ghorbani 2023; Li et al.

2021.) More research is necessarily to identify the number of lncRNAs comprehensively and understand their significance and roles in biological functions.

The aim of this study is to reveal differences between lncRNA and messenger RNA (mRNA) transcripts by comparing mRNA-protein and lncRNA-protein complexes assembled in *Saccharomyces cerevisiae* cells. To perform the comparison, a new method called RNA-Decoder was developed during this study. RNA-Decoder can be utilized to identify proteins that bind to a specific RNA transcript. Identification of different proteins binding to lncRNA, and mRNA transcripts could reveal the unknown secrets of lncRNAs and help to understand their roles in cells better.

1.1 Long non-coding RNAs

Overall, RNA molecules can be classified into two main categories: non-protein-coding RNAs and protein-coding RNAs i.e. protein encoding mRNAs. Non-protein coding RNA class include structural RNAs, for example transfer RNA (tRNA) and ribosomal RNA (rRNA) molecules which participate in protein synthesis, as well as lncRNAs and small non-coding RNAs which act as regulators of gene expression in various ways. (Jing Li and Liu 2019.) For some RNA molecules, placing the RNA in non-coding or protein-coding category might be challenging (Ma et al. 2013). Dinger et al. (2008) have even suggested that perhaps RNA molecules might be molecular multi-taskers instead of being only strictly protein-coding or non-coding molecules.

The word that best describes the class of lncRNAs is heterogeneity – lncRNAs establish functionally a remarkably diverse class. All ribonucleic acid molecules whose length is longer than 200 nt and do not encode proteins are considered as lncRNAs. (Li et al. 2021; Ma et al. 2013.) Another typical shared feature for lncRNAs is their tendency to form highly organized secondary structures (Mercer and Mattick 2013).

Small non-coding RNAs and lncRNAs share similarities in despite differences in their mechanisms of action. Usually, small non-coding RNAs work as negative regulators of cellular processes, interacting with mRNA transcripts and silencing gene expression whereas gene regulatory mechanisms of lncRNAs are more complicated and subtle. (Palazzo and Lee 2015.) For example, small interfering RNAs (siRNAs) and micro-RNAs (miRNAs) bind to their complementary sequences in mRNA transcripts inhibiting their translation into proteins either directly or indirectly (Lam et al. 2015). On the other hand, gene regulatory mechanisms of lncRNAs are more complicated and subtle. lncRNAs can enhance or inhibit gene expression by interacting with various other macromolecules in the cell for example

chromatin remodelling complexes, transcription factors and DNA (Ma et al. 2013; Palazzo and Lee 2015). Therefore, they both are important administrators of cellular processes such as cell cycle regulation, developmental processes, metabolism, and stress responses (Huang et al. 2013; Juan Li et al. 2021; Saliminejad et al. 2019).

Biogenesis of lncRNA transcripts is similar to biogenesis of mRNA transcripts because both are transcribed by RNA polymerase II and processed in a similar manner(Fernandes et al. 2019; Mercer and Mattick 2013; Quinn and Chang 2016). Common RNA processing steps for lncRNAs include 5'-capping, polyadenylation and splicing (Mishra and Kanduri 2019). In some rare cases lncRNAs are even perceived to translated into functional micropeptides(Juan Li et al. 2021.).

In mammalian cell, lncRNAs consist of only 0.03 % to 0.2 % of the total RNAs while mRNAs constitute 3 % to 7 % (Palazzo and Lee 2015S). The expression levels of lncRNAs are typically lower than expression levels of mRNAs in cells and the half-lives of lncRNAs tend to be short due to their susceptibility to degradation which makes studying of them difficult (Yamashita et al. 2016). A short half-life allows quick changes of lncRNA levels in cells enabling quick response to altered environmental conditions and subtle modifications in gene expression. (Yamashita et al. 2016.)

The number of different lncRNAs in humans is unclear but it has been estimated to be from 5,400 to 53,000. In fungi, thousands of lncRNAs have been discovered. (Atkinson et al. 2018; Davati and Ghorbani 2023; Palazzo and Lee 2015.) However, not all of discovered lncRNAs have a significant function and part of them are still supposed to be just a transcriptional background noise. This is supported by the observations that i) only a small fraction of lncRNAs are transcribed on the level that is high enough to have some kind of function into cells, ii) RNA polymerase II can sometimes start transcription on any available DNA, iii) some areas of strictly packed heterochromatins are actively transcribed in low levels and iv) some lncRNAs can encode functional micropeptides, meaning that they might really be misannotated untranslated regions of mRNA transcripts. The rapid degradation of lncRNAs can be seen as an evidence of part of them being just transcriptional background noise: rapid degradation acts as a quality control degrading unintentionally transcribed unnecessary transcripts. (Palazzo and Lee 2015.)

1.2 Genomic context of long non-coding RNAs

lncRNAs can be divided in cis or trans acting lncRNAs derived from do they act in their nearby genomic location or not (Quinn and Chang 2016). While many lncRNAs function in

trans, regulating genes in remote location, there is a large group of cis acting lncRNAs that regulate the genes nearby their genomic location. In these cases, knowledge about the genomic locus of a lncRNA can offer a valuable information about its functionality. (Kung et al. 2013.)

Based on their genomic location lncRNAs can be divided into intergenic lncRNAs and intronic lncRNAs (Figure 1A-B). Intronic lncRNAs are transcribed only from introns of protein coding genes and locate either in the sense or antisense direction relative to the nearby genes (Fernandes et al. 2019). Intergenic lncRNAs are located between protein coding genes and can locate in sense strand of DNA or in antisense strand of DNA (Figure 1A). Intergenic lncRNAs can be divided further into stand-alone lncRNAs and pseudogenes. Stand-alone lncRNAs remind mRNAs more than other lncRNAs because they are usually spliced and polyadenylated and their average length is approximately 1 kb. Pseudogenes are "fossil genes" which have lost their protein coding features during evolution. Pseudogenes might have lost their potential due different point mutations, retrotransposons or they might be products of tandem gene duplication. Majority of pseudogenes are not expressed anymore, but a few of them are still transcribed and some of them have been noted to regulate gene expression. (Kung et al. 2013.)

Antisense-sense transcription pairs are common, because 70 % of sense transcripts have been detected to have an antisense pair. In addition, most of the eukaryotic promoters are bidirectional, initiating transcription in both directions so that mRNA is transcribed in the sense direction and lncRNA in the antisense direction (Quinn and Chang 2016). In such "divergent promoters" (Katayama et al. 2005) Sense-antisense transcript partners do not overlap each other at all. In other cases, sense-antisense pairs can overlap complete or partly. It is common, that the antisense transcript overlaps the promoter or the terminator area of the sense transcript (Figure 1C). This kind of overlapping can cause consequences the expression of sense transcript for example blocking transcription factor binding sites. (Kung et al. 2013.)

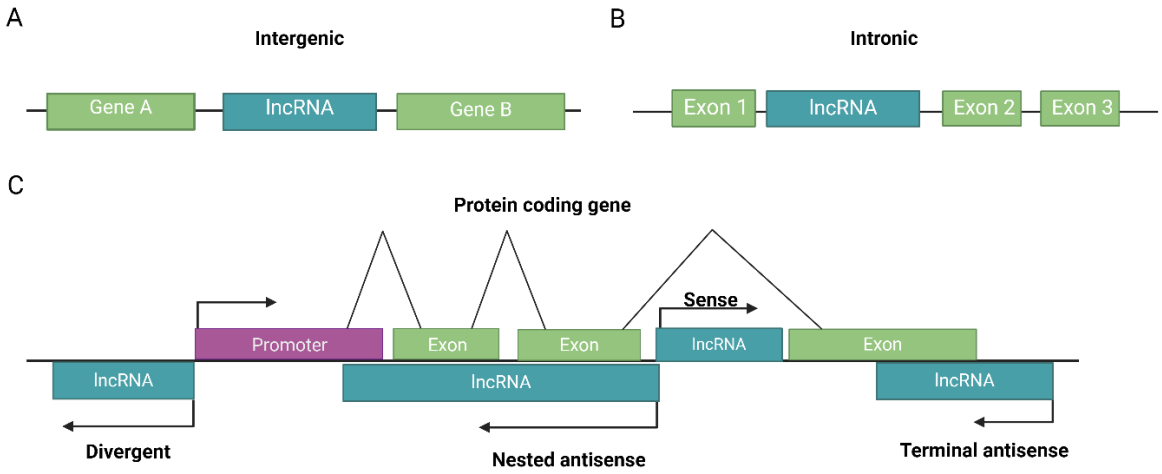

Figure 1. Different models of biogenesis of lncRNAs. (A-B) Intergenic lncRNAs are located between protein-coding genes and intronic lncRNAs are transcribed from intron regions of genes. (C) Direction of transcription of lncRNAs can be sense or antisense related to its nearby genes and the overlap of lncRNAs and genes can be partly (terminal antisense on the picture), completely (nested antisense on the picture) or there might be no overlapping at all (divergent on the picture). (Created with BioRender.com)

1.3 Functional mechanisms of long non-coding RNAs

The action mechanisms of lncRNAs vary a lot and they can influence gene regulation at multiple levels: including levels of epigenetic, transcription and post-transcriptional regulation (Kung et al. 2013). The lncRNAs can be localized in the nucleus, the nucleolus or in the cytoplasm. Depending on their localization and interaction partners (DNA, RNA, or proteins) lncRNAs can control gene expression in different mechanisms, for example by modifying the chromatin structure, recruiting, or decoying transcription factors away from their target genes, interrupting transcription machinery or altering mRNA stability (Figure 2). (Statello et al. 2021.)

Some lncRNAs mediate their functions by interacting directly with other RNA molecules or DNA. However, the most common described action mechanism of lncRNAs is to interact with RNA-binding proteins (RBPs) and to form functional ribonucleoprotein complexes (RNPs). (Kung et al. 2013.) Proteins binding to lncRNAs participate in transcriptional regulation and are known to bind to DNA can also bind on lncRNAs. Usually, these proteins have different binding sites to interact with DNA and lncRNAs and this simultaneous binding is needed to form a functional RNP complex. (Long et al. 2017.)

The easiest way to view the different action mechanisms of lncRNA is to distinguish their roles as regulators of chromatin structure, transcriptional activity, and post-transcriptional activity gene expression.

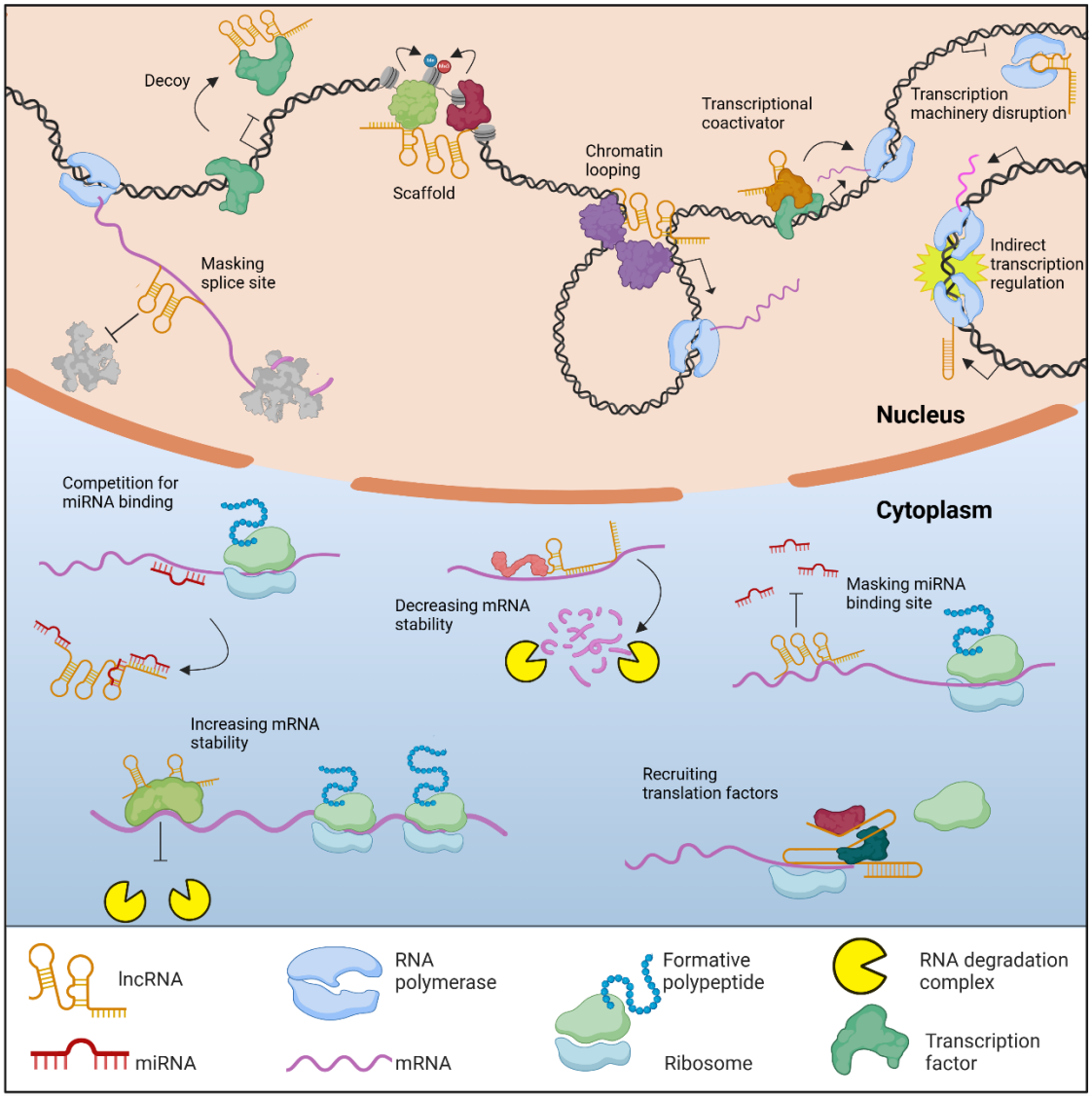

Figure 2. LncRNAs regulate gene expression through various mechanisms. Adapted from Kung et al. 2013. (Created with BioRender.com)

1.3.1 Regulation of chromatin structure

At chromatin stage actions of lncRNAs cause epigenetic modifications, typically via chromatin modifying which causes either activation or silencing of the gene. Additionally, lncRNAs can recruit chromatin modifying complexes on the DNA or act as scaffolds. (Kung et al. 2013.)

Acting as a scaffold lncRNA can interact with multiple molecule partners at the same time and bring these partners together, causing changes in gene expression. For example, *HOTAIR* lncRNA can attach on PRC2 and LSD1 proteins at the same time. Both PRC2 and LSD1 are capable to repress genes by different mechanisms and with *HOTAIR* lncRNA they form the *HOTAIR*/PRC2/LSD1 complex which can inhibit expression of several genes by various

mechanisms. (Wang and Chang 2011.) Recruiter lncRNAs can recruit chromatin modelling complexes to the promoter areas of genes to activate or inhibit gene expression. Chromatin modifying complexes can acetylate or methylate histones affecting chromatin density and accessibility to transcription factors. For example, lncRNA called *HOTTIP* can recruit MLL trimethylase complex to promoter area of HOXA genes to promote histone trimethylation and activating gene expression. (Statello et al. 2021.)

Some lncRNAs have also a possibility to interact directly with DNA. These interactions create triple helix or R-loop structures on chromatin affecting chromatin accessibility and allowing, for example transcription activation complexes to bind to the promoter area. (Kung et al. 2013.) For example, *TARID* lncRNA can form R-loop structure on the promoter area of *TCF21* gene. Other transcription activator proteins can recognize this structure and activate expression of *TCF21.* (Statello et al. 2021.)

1.3.2 Regulation of transcriptional activity

At transcription stage lncRNAs can regulate gene expression by recruiting transcription factors, interrupting transcription, acting as coregulators with transcription factors, indirectly mechanisms, or interacting directly with RNA polymerase II (Pol II). (Kung et al. 2013; Long et al. 2017.) Some lncRNAs have an ability to recruit transcription factors to the promoter regions of genes, thereby activating gene expression in this way. Additionally, certain lncRNAs can decoy transcription factors away from the promoter areas of genes and supress gene expression by isolating these transcription factors. (Long et al. 2017.) For example lncRNA *PANDA* decoys transcription factors from the promoter area of the pro-apoptotic genes, preventing gene expression and blocking apoptosis (Hung et al. 2011).

Coregulator acting lncRNAs regulate gene expression together with other molecules. For example, lncRNA *RMST* interacts with SOX2 transcription factor during neurogenesis. Together *RMST* and SOX2 can bind to the promoter areas of a wide range of other neurogenesis involved genes. The interaction between *RMST* and SOX2 determines the fate of the nerve cell. (Ng et al. 2013.)

In addition to regulating transcription through transcription factors, lncRNAs can bind directly to Pol II and disturb its activity. For example in humans, Alu lncRNAs can bind directly on Pol II, preventing formation of preinitiation complex and expression of heat-shock proteins. (Mariner et al. 2008.)

Indirect transcription regulation presents an intriguing regulation mechanism, where the transcription event itself is more significant than the nascent lncRNA transcript. In this case,

the transcription of lncRNA can enhance or supress the expression of nearby genes. (Long et al. 2017.) A principle of indirect transcription regulation can be observed in the transcription of lncRNA *SRG1* in *S. cerevisiae.* The transcription of *SRG1* interrupts binding of Pol II to the promoter area of SER3 gene preventing the expression. (Martens et al. 2005.)

1.3.3 Post-transcriptional gene regulation

At translation stage, lncRNAs can control mRNA processing, interrupt translation machinery by binding ribosomes, influence translational factors, alter a stability of mRNAs, or compete with miRNAs (Kung et al. 2013). Several lncRNAs controlling mRNA processing influences alternative splicing in one way or another. For example lncRNA named *MALAT1* controls alternative splicing by re-localizing splicing factors to right transcription sites where splicing happens (Tripathi et al. 2010). Instead *Gomafu/MIAT* lncRNA can isolate SPF1 (splicing factor 1) preventing formation of spliceosome complex and disturb splicing of several mRNAs (Kung et al. 2013). Natural antisense transcripts (NATs) have an ability to affect alternative splicing by masking the splice sites on mRNA and protecting mRNA from spliceosome complex (Krystal et al. 1990).

In cytoplasm, lncRNAs can destabilize or stabilize mRNAs by recruiting deadenylation, decapping and other factors or interacting with miRNAs (Kung et al. 2013). For example, NATs from iNOS region (inducible nitric oxide synthase) have been detected to interact with its sense mRNA and recruit stabilization factors to stabilize sense mRNA (Matsui et al. 2008).

LncRNAs can compete against miRNAs or bind directly on either mRNAs or miRNAs to cover miRNA binding sites (Kung et al. 2013). *BACEAS* is an antisense transcript of BACE gene, and it can increase stability of BACE mRNA. *BACEAS* and specific miRNA compete for binding on the same binding site of BACE mRNA. Binding of *BACEAS* on the BACE mRNA obscures binding site for miRNA preventing miRNA-mediated degradation of BACE mRNA. (Faghihi et al. 2010.) Some lncRNAs have miRNA-binding sites on their 3'-untranslated regions. These lncRNAs do not compete with miRNAs but they can act as "sponges", bind miRNAs on themselves and isolate them from their target mRNAs. (Salmena et al. 2011.)

1.4 Long non-coding RNAs in fungi

To date, thousands of lncRNAs have been discovered in fungi. Previous research has discovered 472 lncRNAs in *Aspergillus flavus* and over 7,000 lncRNAs in *Schizosaccharomyces pombe* and in *Candida albicans.* (Atkinson et al. 2018; Davati and Ghorbani 2023; Hovhannisyan and Gabaldón 2021.) In *S. cerevisiae* 25 % of the

transcriptome has been identified as non coding RNA and classified into Stable Unannotated Transcripts (SUTs) and Cryptic Unstable Transcripts (CUTs) (Xu et al. 2009). Despite the high number of discovered lncRNAs, only a few of these lncRNAs in fungi have been functionally characterized (Table 1). Furthermore, the experimental characterization extends to only eight different fungali species: *S. cerevisiae, S. pombe, T. reesei, Cochiliobolus heterosphus, Neurospora crassa, Ustilago maydis, Cryptococcus neoformans* and *Fusarium oxysporum.* (Davati and Ghorbani 2023; Juan Li et al. 2021.)

Table 1. Characterized lncRNAs in fungi (Davati and Ghorbani 2023; Juan Li et al. 2021).

Organism	Characterized lncRNAs
S. cerevisiae	GAL10-lncRNA, GAL4 lncRNA, ncASP3, PHO84 antisense transcripts, REM2, IRT1, REM3, SRG1, pHO-lncRNA, PWR1, ICR1, AS-PHO5, Antisense lncRNA of CDC28, Antisense lncRNA of Ty1, TERRA, TLC1, ADF1
S. pombe	prt1, nc-tgp1, prt2, mlonRNAs, SPNCRNA.1164, meiRNA-S and L, TER1
T. reesei	HAX1
C. heterosphus	Antisense of transcription factor CMR1
N. crassa	Qrf
U. maydis	Antisense to gene um02151
C. neoformans	RZE1
F. oxysporum	Fo-carP and Ff-carP60

Majority of lncRNA research has been conducted in *S. cerevisiae* and *S. pombe* as a model organism because both are well-studied model organisms, and there exists molecular biology tools for their genomic manipulation. Both belong into same division Ascomycota, but they are not very near kin. Unlike *S. pombe, S. cerevisiae* has lost its RNA interference mediated degradation pathway during evolution (Drinnenberg et al. 2009). This ability makes *S. cerevisiae* a unique organism for studying functions of lncRNAs because effects of lncRNAs might be hard to detect in the presence of the RNA interference machinery(Wery et al. 2016).

Fungal lncRNAs have been functionally linked to various cellular network systems, for example control of meiosis, mating, sporulation, metabolism, stress conditions, circadian rhythms, cell-cell adhesion, cell aging and telomere replication. An interesting detail is that the majority of characterized lncRNAs in fungi regulate the expression of their nearby genes in *cis*. (Juan Li et al. 2021; Till et al. 2018.) Next, a further description of some characterized lncRNAs is given to explain their importance and functional mechanisms comprehensively.

1.4.1 Meiosis and sporulation

A few different lncRNAs are known to have an impact on meiosis or sporulation in yeasts. LncRNAs *IRT1, RME2* and *RME3* participate in preventing meiosis in *S. cerevisiae.* They act in a cis, inhibiting their nearby genes which are needed for cell differentiation and sporulation. The expression of those genes can cause cell to enter meiosis in nutrient starvation conditions. (Hongay et al. 2006; van Werven et al. 2012.)

A key regulator, of meiosis *IME1*(inducer of meiosis) is expressed during nitrogen starvation and the presence of Ime1 protein allows diploid cells to enter meiosis (van Werven et al. 2012). However, when cells are grown in nutrient rich conditions in a haploid state, a lncRNA *IRT1* inhibits the expression of *IME1*. *IRT1* is transcribed from a locus upstream of *IME1* gene and is itself regulated by the transcription factor Rme1.

IRT1 prevents initiation of transcription of *IME1* by attracting histone methyltransferase and histone deacetylase Set2 and Set3 enzymes. Set2 and Set3 modify chromatin structure of the promoter area of *IME1* by adding methyl groups and removing acetyl groups of histones. This prevents transcription factor *Pog1* to bind on the promoter area of *IME1,* thereby inhibiting its expression (Figure 3A). (Gelfand et al. 2011.) In sporulating MAT a/α cells the a1–α2 repressor complex inhibits the expression of the transcription factor Rme1 which shuts off *IRT1* transcription, thereby enabling the expression of *IME1* (Figure 3B) (van Werven et al. 2012).

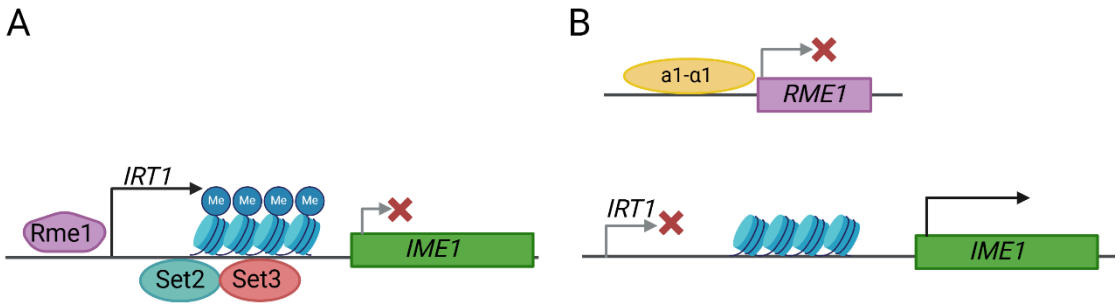

Figure 3. LncRNA mediated gene silencing determines the cell's fate. (A) In haploid *S. cerevisiae* cells *IME1* is silenced by lncRNA *IRT1* mediated mechanism. Transcription factor *Rme1* strengthens transcription of *IRT1*. *IRT1* recruits Set2 and Set3 enzymes to remodel chromatin and the formation of facultative heterochromatin on the promoter area of *IME1* causes silencing of the gene and these cells cannot undergo sexual differentiation. (B) In diploid *MAT a/α S. cerevisiae* cells expression of *RME1* is blocked and transcription of *IRT1* is prevented permitting expression of *IME1*. The expression of *IME1* allows the cell to entry to meiosis. (Created with BioRender.com)

Besides of *IME1*, also *IME4* and *ZIP2* genes are required for the control of meiosis in S. *cerevisiae*. *IME4* codes for a methyltransferase and it activates *IME1* as a response for nutritional conditions. *ZIP2* helps pairing of homologues chromosomes during meiosis. (Chua and Roeder 1998; Shah and Clancy 1992.) LncRNA *RME2* (regulation of meiosis 2) inhibits expression of *IME4* and lncRNA *RME3* (regulation of meiosis 3) inhibits expression of *ZIP2*. Both *RME2* and *RME3* act in cis, are transcribed from the antisense strand of DNA and they locate downstream of their regulated genes, so that transcription directions of lncRNAs and their regulated genes overlap (Figure 4.). (Gelfand et al. 2011; Till et al. 2018.)

RME2 is constitutively transcribed in haploid *S. cerevisiae* cells instead of *IME4* sense transcript because the promoter of *RME2* is stronger meaning that the transcription of IME4 is prevented (Figure 4A). In diploid *S. cerevisiae* cells the a1–α2 repressor complex binds to the promoter area of the *RME2* preventing the transcription of *RME2* and allowing transcription of weaker *IME4* promoter, so cells can undergo meiosis (Figure 4B) (Chen and Neiman 2011; Hongay et al. 2006). The specific mechanism how *RME2* and *RME3* act is not entirely understood yet but previous research indicates that lncRNAs block transcription elongation, not initiation, because there is no evidence that they would prevent binding of transcription factors to the promoter areas of *IME4* and *ZIP2* (Gelfand et al. 2011).

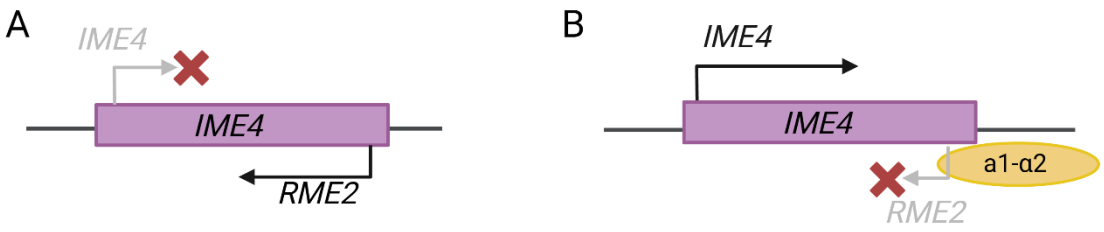

Figure 4. Transcriptional balance between *IME4* and lncRNA *RME2* controls the cell's entry to meiosis. (A) *RME2* is transcribed from the antisense strand of *IME4* gene. In haploid *S. cerevisiae* cells *RME2* is transcriptionally active, which inhibits the transcription of *IME4.* (B) In diploid *S. cerevisiae* cells, the a1–α2 repressor complex prevents transcription of *RME2*, which activates transcription of *IME4* and allows the cell to enter meiosis (Created with BioRender.com)

In *S. pombe* meiosis is also regulated by lncRNA but the regulation system differs from the regulation system in *S. cerevisiae.* In *S. pombe* lncRNA called meiRNA regulates progression of meiosis and in addition to that it also helps chromosome pairing during prophase in meiosis. (Till et al. 2018.)

During meiosis, two isoforms of meiRNA (meiRNA-S and meiRNA-L) are transcribed and accumulate on their genomic locus *sme2*. There they interact with RNA-binding Mei2 protein

forming meiRNA-Mei2 dot complex (Figure 5A). Mei2 dot complex binds to Mmi1, a RNA degradation and heterochromatin formation factor, inhibiting its action and allowing a stable expression of meiosis progression genes which are normally degraded by Mmi1. (Till et al. 2018; Yamashita et al. 2016.) Mei2 dot complex also binds and isolates Mmi1 RNA transcripts preventing expression of new Mmi1 proteins (Figure 5B) (Mukherjee et al. 2018). In mitotic cells Mei2 protein is not expressed and meiRNA is degraded by Mmi1 directed RNA degradation. Activation of Mmi1 also provokes formation of heterochromatin. (Till et al. 2018.)

In *S. pombe,* meiotic prophase chromosomes bind together in the spindle pole body structure in the middle of the cell. Transcription and accumulation of meiRNAs near to the *sme2* locus helps pairing of homologous chromosomes during early prophase. Transcription needs to happen from both *sme2* loci in homologous chromosomes to cause the pairing. This might mean that meiRNA transcripts from different chromosomes can recognize and interact with each other. The exact reaction mechanism is not clear yet and there could be also more factors beside to meiRNAs to associate the pairing. (Juan Li et al. 2021.)

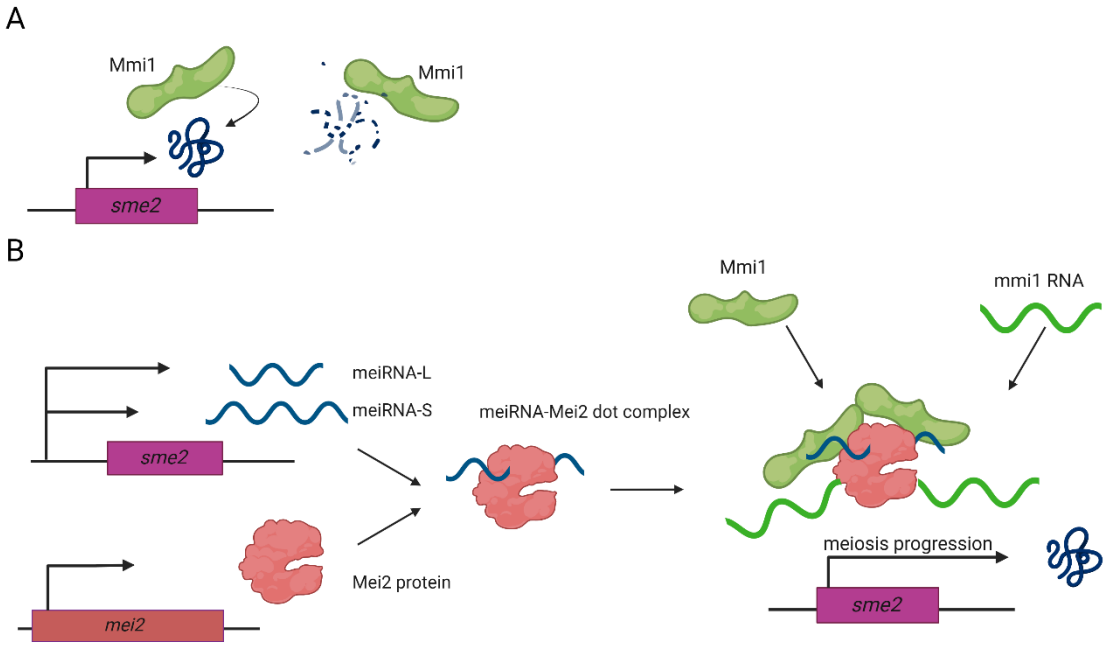

Figure 5. lncRNA-mediated regulation of meiosis in *S. pombe.* (A) In mitotic cells gene products which can cause meiosis are degraded by Mmi1 (B) In early meiosis, meiRNA-L and meiRNA-S are transcribed, and they form meiRNA-Mei2 dot complex by interacting with Mei2 protein. meiRNA-Mei2 dot complex prevents action of a meiotic inhibitor Mmi1 by binding both Mmi1 proteins and RNA transcripts allowing expression of *sme2* and inducing meiosis. (Created with BioRender.com)

1.4.2 Senescence and telomere maintenance

LncRNAs are known to be involved in senescence by regulating telomerase activity and telomere synthesis in cells. Senescence is a cell's response to aging related damage, for example telomere damage or DNA damage, but it has also significant roles during development. The cell cycle of senescent cells is arrested: they do not undergo meiosis or multiply, but they are not apoptotic either. (Huang et al. 2022.) In *S. cerevisiae* mother cells enter senescence after budding and ultimately dies(Fehrmann et al. 2013).

Telomeric repeat containing RNAs, also known as TERRAs (telomeric repeat-containing RNA) participate homeostasis of chromosomes in higher eukaryotes, but they have also been found in *S. cerevisiae* and *S. pombe*. (Luke et al. 2008; Moravec etal. 2016.)

TERRAs have been perceived to act as a scaffold by interacting with telomeres and attracting telomerase in cells where telomeres have already shortened. (Luke et al. 2008; Moravec et al. 2016.) To form a scaffold structure, TERRAs start to build up and form a nuclear focus which acts as a scaffold. The function of the scaffold structure is to help to recruit chromatin remodelling factors to the end of chromosome provoking lengthening of chromosome. Shortened telomeres and aggregation of TERRAs activate telomerase leading to telomere elongation (Cusanelli et al. 2013).

Besides *TERRAs*, other lncRNAs can participate to telomere maintenance. In *S. cerevisiae* lncRNA *TLC1*, participates formation of telomerase enzyme and serves a template of telomeric repeats. (Zeinoun et al. 2023.) The scaffold structure of *TERRAs* aggregate and bind *TLC1* into *TERRA* focus to elongate telomeres (Cusanelli and Chartrand 2015). S. *pombe* has a corresponding lncRNA to *TLC1*, called *TER1* (Leonardi et al. 2008).

1.4.3 Nutrient metabolism

Nutrition starvation conditions cause activation of different lncRNAs in fungi. As previously mentioned, nitrogen and glucose starvation cause *S. cerevisiae* cells to enter meiosis and start sporulation (Herskowitz 1988). Additionally, lncRNAs control other genes associate with metabolism and these lncRNAs have been found in various fungi(Till et al. 2018).

Metabolic stress-induced lncRNAs (mlonRNAs) respond to glucose starvation conditions in *S. pombe*. (Juan Li et al. 2021.) Those lncRNAs are expressed from the upstream area of the *fbp1* gene, which encodes essential gluconeogenesis enzyme fructose-1,6-bisphospatase1. In the absence of glucose a transcription factor *atf1* attaches on the upstream promoter area 1 of *fbp1*. Binding of *atf1* remodels chromatin structure accessible for Pol II initiating mlonRNAs

transcription. The passage of Pol II causes remodelling on the upstream promoter area 2 of *fbp1*. Transcription factor Rst1 binds on the upstream promoter area 2 and *fbp1* is expressed. (Figure 6). (Hirota et al. 2008; Juan Li et al. 2021.)

Figure 6. *mlonRNAs* induce the expression of gluconeogenesis enzyme. Transcription factor *atf1* binds on upstream promoter area (UAS) of *fbp1* causing transcription of lncRNA called *mlonRNA*. Transcription of *mlonRNA* causes chromatin rearrangement enabling binding of another transcription factor Rst1 on upstream promoter area, causing expression of *fbp1*. (Created with BioRender.com)

In S. pombe phosphate starvation is regulated by three different lncRNAs prt1, prt2 and nc-tpg1. They regulate three genes, *pho1* (a cell-surface acid phosphatase), *pho84* (an inorganic phosphate transporter) and *tgp1* (a glycerophosphate transporter) which are important for phosphate metabolism and repressed in phosphate rich conditions. *Prt1*, *prt2* and *nc-tpg1* locate upstream of their target genes and act in *cis* by inhibiting the expression of their target genes. (Till et al. 2018.)

Nc-tgp1 is an unstable transcript and it is expressed in phosphate-rich conditions. The expression of *nc-tgp1* increases density of chromatin on the promoter area of *tgp1,* preventing the binding of a transcription factor Pho7. As a result, *tgp1* transcription is repressed. In phosphate starvation conditions the expression of *nc-tgp1* is inhibited causing decreased chromatin density and allowing Pho7 to bind on the promoter area, leading stable expression of *tgp1* (Figure 7A). (Ard et al. 2014.)

Similar to *nc-tgp1*, transcription of *prt1* and *prt2* lncRNAs inhibits the expression of *pho1* and *pho84* genes in phosphate-rich medium but unlike *nc-tgp1,* they do not cause changes to chromatin density. Transcription of *prt1* is initiated on the upstream of *pho1* and, elongate across its promoter area. Elongation disturbs binding of the transcription factor Pho7 to the *pho1* promoter area, preventing its transcription. (Figure 7B) The mechanism of *prt1*-mediated inhibition is unclear but might involve supercoiled DNA structure that builds up of on the promoter area of *pho1* as a result of *prt1* transcription which could prevent transcription factor binding. Alternatively, transcriptionally engaged Pol II might physically block binding of Pho7 to the promoter area of *pho1* during elongation of prt1 transcription might physically block binding of Pho7. (Chatterjee et al. 2016.)

Prt2 controls gene silencing by disturbing transcription of *pho84,* but its exact mechanism remains somewhat unclear. The entire *prt2-pho84* locus is transcribed in phosphate rich conditions. This might imply that transcription of lncRNA *prt2* across the *pho84* gene disturbs productive transcription of *pho84* (Figure 7B). (Garg et al. 2018.)

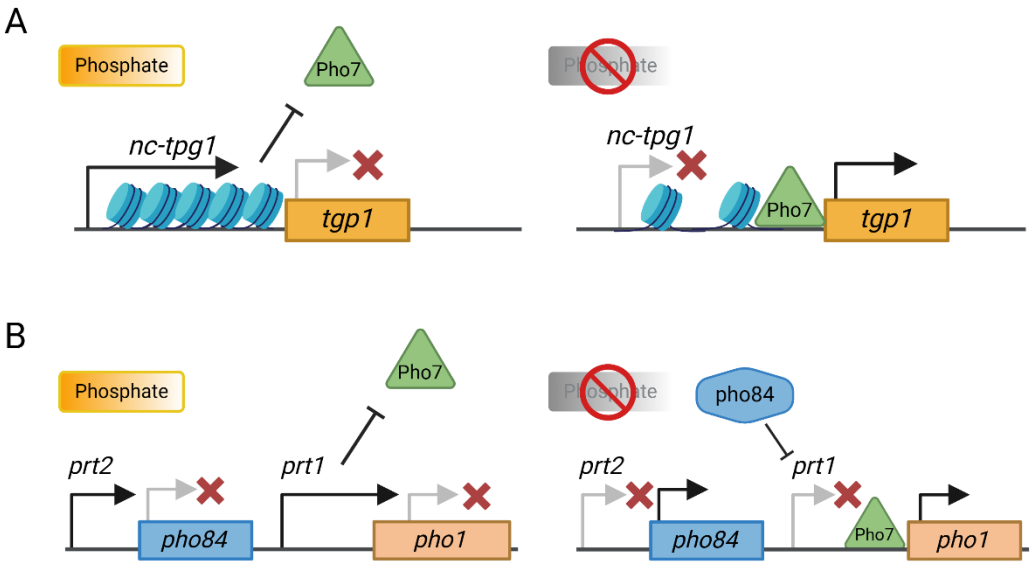

Figure 7. Availability of extracellular inorganic phosphate regulates expression of phosphate metabolism genes in *S. pombe.* (A) *Left:* In phosphate rich conditions expression of lncRNA *nc-tpg1* causes increased chromatin density and inhibition of *tgp1*gene. *Right:* Absence of phosphate represses expression of *nc-tpg1* decreasing density of chromatin structure. As a result, Pho7 can bind to the promoter *tgp1* gene and induce its expression. (B) *Left:* In the presence of phosphate, the expression of lncRNAs *prt1* and *prt2* inhibit the expression of *pho1* and *pho84*, respectively. *prt1* disturbs the binding of Pho7 but the action mechanisms of *prt2* is unknown. *Right:* In the absence of phosphate, *prt1* and *prt2* are repressed allowing expression of *pho1* and *pho84*. (Created with BioRender.com)

In *S. cerevisiae* the expression of *PHO84* and antisense transcript of *PHO84* are anticorrelated to each other. Accumulation of antisense *PHO84* transcripts causes reduction of sense *PHO84* transcripts and vice versa. (Castelnuovo et al. 2013.) Normally, *PHO84* antisense transcripts are quickly degraded by the exosome component Rrp6. In aged cells the antisense transcript becomes stabilised whereas the sense transcription of *PHO84* is decreased. (Camblong et al. 2007.) Loss of Rrp6 promotes expression of antisense transcripts of *PHO84* (Castelnuovo et al. 2013). Similar to *IRT1* function in the control of meiosis, *PHO84* antisense transcripts cause histone rearrangement. *PHO84* antisense transcripts recruit Hda1, a histone deacetylase causing deacetylation on the *PHO84* promoter resulting *PHO84* gene repression. (Camblong et al. 2007.)

GAL1, *GAL7* and *GAL10* genes are associated with galactose metabolism in *S. cerevisiae*. These three genes form a cluster in genome and depending on the availability of different carbon sources, these genes can be repressed or expressed, enabling *S. cerevisiae* to utilize

different carbon sources and thrive in different environments. In the absence of galactose and the presence of glucose, genes are not transcribed whereas in the presence of galactose genes are expressed. (Juan Li et al. 2021.)

GAL10-ncRNA is transcribed in *S. cerevisiae* cells once in every 50 minutes and in the absence of galactose, its transcription represses the expression of *GAL1* and *GAL10* genes. *GAL10*-ncRNA is an antisense transcript which initiates near the 3' end of the *GAL10* gene and terminates at the *GAL1* gene locus. Transcription of *GAL10*-ncRNA is induced by Reb1 transcription activator, which binds on the coding region of *GAL10*. Transcription of *GAL10*-ncRNA promotes activity of methyltransferase Set1 and indirectly histone deacetylase Rdp3S leading to nucleosome deacetylation on the GAL cluster area and gene silencing (Figure 8A) (Houseley et al. 2008.)

The decapping enzyme Dcp2 seems to control stability and lifecycle of *GAL10*-ncRNAs in cells. Glucose starvation causes phosphorylation and activation of Dcp2 leading to decapping of *GAL10*-ncRNAs by Xrn1 and Rat1. Consequently, in the presence of galactose (and when glucose is absent) *GAL10*-ncRNAs are degraded, changing the chromatin state at GAL cluster and leading to the expression of *GAL1* and *GAL10* genes (Figure 8B) (Yamashita et al. 2016.)

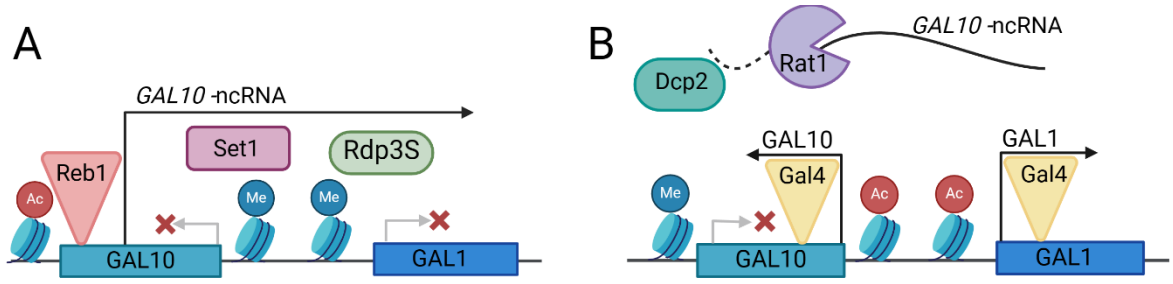

Figure 8. *GAL10*-ncRNA is an important regulator of galactose metabolism genes. (A) In the absence of galactose, Reb1 transcription activator induces the transcription of *GAL10*-ncRNA. Transcription of *GAL10*-ncRNA recruits Set1 methyltransferase and Rdp3S histone deacetylase to remodel chromatin structure, which leads to the repression of *GAL1* and *GAL10* genes. (B) *GAL10*-ncRNAs are degraded quickly by decapping enzyme Dcp2 and the 5'-3' exonuclease Rat1. The presence of glucose activates Dcp2 enzyme leading to the degradation of *GAL10*-ncRNA and expression of *GAL10* and *GAL1* genes. Gal4 is the main activator of transcription of *GAL10* and *GAL1* genes. (Created with BioRender.com)

S. cerevisiae lncRNA *SRG1* regulates serine metabolism by controlling the expression of the phosphoglycerate dehydrogenase gene *SER3*. Under serine rich conditions, *SRG1* represses *SER3* in *cis* via transcription interference. (Till et al. 2018). Transcription of *SRG1* is activated by Cha4 transcription activator, which and the SAGA and Swi/Snf coactivator complexes (Martens et al. 2005). *SRG1* locates upstream of SER3, overlapping its promoter area. During the transcription of *SRG1* RNA Pol II disassembles the nucleosomes on the way disturbing

binding of transcription factors on the *SER3* promoter area. Nucleosomes are reassembled by Spt6 and Spt16 enzymes later (Figure 9A). (Hainer et al. 2011.) In the absence of serine Cha4 transcription activator does not recruit SAGA and Swi/Snf complexes on the promoter area of *SRG1* repressing its expression and allowing expression of *SER3* (Figure 9B) (Martens et al. 2005.)

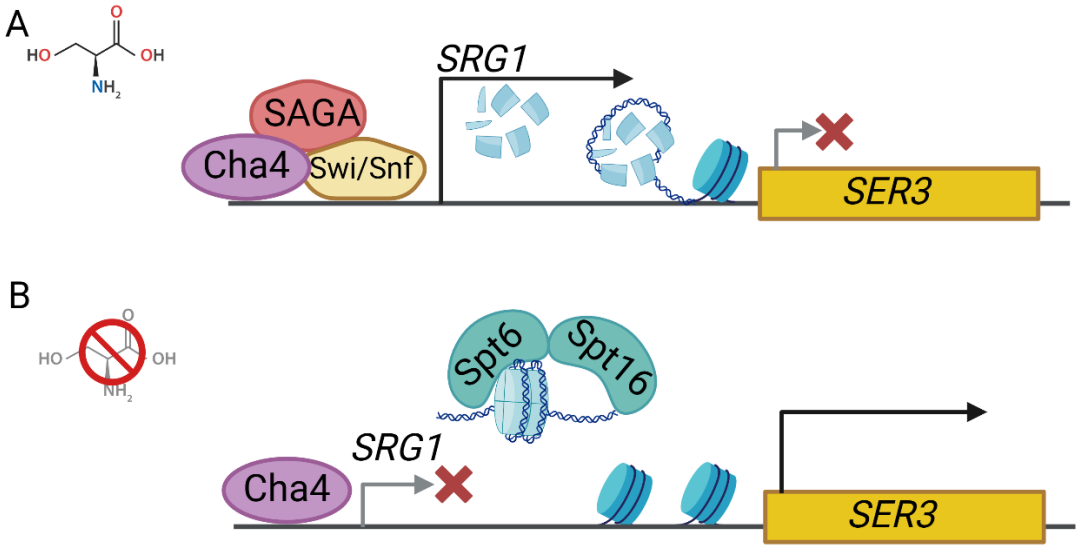

Figure 9. The expression of *SER3* is regulated by the lncRNA *SRG1*. (A) In the presence of serine expression of *SER3* is repressed. Cha4 recruits other transcription activators on the promoter area of *SRG1*. Transcription of *SRG1* causes disassembly of nucleosomes on the promoter of *SER3* repressing its expression. (B) Spt6 and Spt16 reassemble nucleosomes later. In the absence of serine expression of *SRG1* is repressed leading to the expression of *SER3*. (Created with BioRender.com)

HAX1 lncRNA is involved in the regulation of cellulose metabolism in filamentous fungi *Trichoderma reesei.* Unlike many other characterized lncRNAs in fungi, *HAX1* acts in *trans* controlling expression of genes distant from its own locus. Furthermore, *HAX1* has a stimulatory effect for the expression of its target genes. *T. reesei* produces different kinds of cellulases and hemicellulases, and therefore it is widely utilized as a producer strain in industry. (Till et al. 2018.)

Transcription of cellulases and hemicellulases is controlled by transcription factor Xyr1 which binds to two different binding sites, XBS and XRE, on the promoter areas of the target gene genes. Binding of Xyr1 to the XRE site found on the promoter area of *XYR1* gene itself suppresses transcription. This forms a negative feedback loop that controls the cellular amount of Xyr1 transcription factor. XBS sites are located on the promotor areas of cellulase genes *cbh1*, *cbh2* and *xyn1*. Binding of Xyr1 to XBS activates gene expression. Thus, reduced Xyr1 amount in the cell also decreases the expression of cellulases. *HAX1* controls cellulose

metabolism by interacting with Xyr1 transcription factor and disturbing this negative feedback loop of cellulose metabolism. *HAX1* transcripts can attach to Xyr1 proteins preventing their binding to the XRE site of the *XYR1* promoter , leading to increased Xyr1 level and increased expression of cellulase genes (Figure 10). (Till et al. 2020.)

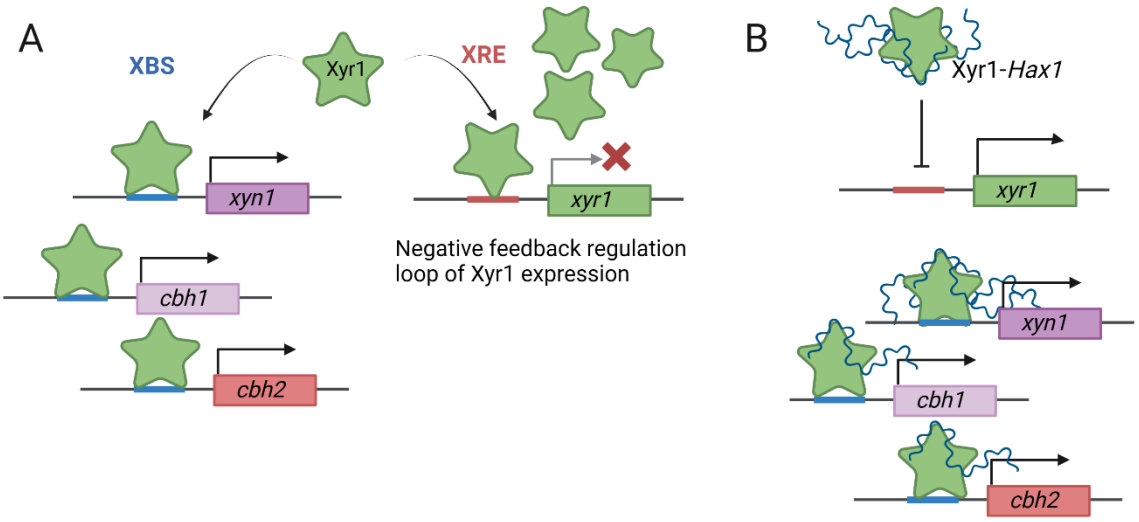

Figure 10. *HAX1* lncRNA controls the expression of cellulase genes. (A) Xyr1 transcriptional activator can bind on the promoter area of *xyn1, cbh1* and *cbh2* genes and activate expression of these cellulases. If Xyr1 is saturated in the cell, it can also bind to its own promoter area causing negative feedback loop and repressing its own expression. (B) *HAX1* lncRNA can prevent Xyr1 to bind its own promoter area, preventing negative feedback regulation and increasing expression of cellulases. (Created with BioRender.com)

An interesting aspect of *HAX1* is that different strains of *T. reesei* express various forms of *HAX1* lncRNA, differing in length from each other. All three different forms activate expression of cellulases, but the longest isoform affects overexpression of cellulases and increases the cellulase activity the most whereas the shortest isoform increases the cellulase activity only marginally. (Till, Mach, et al. 2018). The longer versions of *HAX1* contain more suitable binding sites for Xyr1, sequestering it more efficiently away from the promoter area of the *XYR1* gene (Till et al. 2020).

1.4.4 Stress response

In *S. cerevisiae* the antisense transcript of the *CDC28* and in *S. pombe SPNCRNA.1164* are activated during stress to aid cells to adapt to changed environmental conditions.

In osmotic stress conditions, in *S. cerevisiae* stress-activated protein kinase Hog1 induces transcription of hundreds of lncRNAs. One of those lncRNAs is antisense transcript of *CDC28* which acts in a *cis,* regulating expression of *CDC28* gene. *CDC28* encodes a cell

cycle controlling protein, cyclin-dependent kinase 1 and its expression is increased during osmostress allowing a rapid re-enter to cell cycle progression after stress.

Under osmotic stress, Hog1 attaches on the 3'-end of the CDC28 gene causing antisense transcript of *CDC28* transcription. This transcription remodels chromatin and causes formation of gene looping structure between 5'- and 3'-end of the *CDC28*. Gene looping causes Hog1 translocation to 5'-end of *CDC28* gene, chromatin remodelling by RSC complex, leading increased expression of *CDC28* gene (Fig 11). (Nadal-Ribelles et al. 2014.)

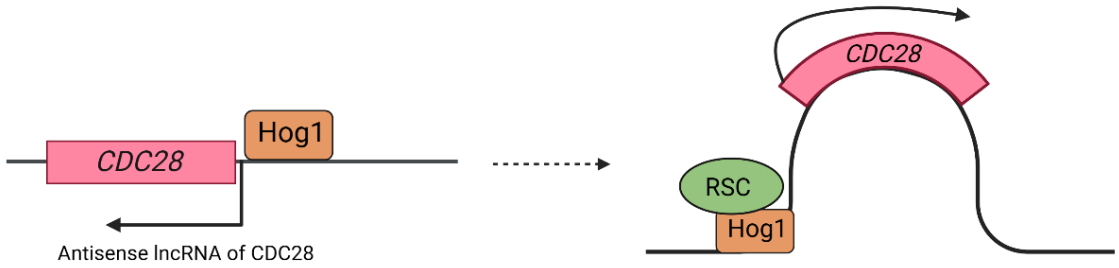

Figure 11. The expression of *CDC28* is activated by its antisense counterpart. Under osmotic stress Hog1 activates the transcription of *CDC28* antisense transcript. Antisense lncRNA causes formation of gene looping structure between 5'- and 3'-end of the CDC28 gene. Consequently, Hog1 translocates increasing expression of *CDC28* sense transcript. (Created with BioRender.com)

In *S. pombe* lncRNA *SPNCRNA.1164* transcription initiates as a response to oxidative stress conditions. The exact reaction mechanism of *SPNCRNA.1164* is still unknown but it has been detected to act in *trans* to reduce a cell's Atf1 levels. Atf1 is a promoter of stress activated MAP kinase pathway of yeast. (Leong et al. 2014.)

1.4.5 Circadian rhythms

Numerous biological processes are coordinated by circadian rhythms for example biochemical pathways and behaviour of an organism. Even in lower eukaryotes like yeasts and filamentous fungi things like metabolism and gene expression are partially regulated by a circadian clock. Studies in *N. crassa* has revealed that the expression of *frq* gene and lncRNA *qrf* (an antisense transcript of *frq*) are required for keeping a circadian rhythm synchronized to a time of a day. (Cemel et al. 2023.)

The detailed molecular mechanism of how *qrf* participates regulation of circadian rhythm is still uncertain but *qrf* controls *frq* levels in cells. Previous research has revealed that transcription of lncRNA *qrf* inhibits expression and controls levels of *frq* via negative feedback. In addition, expression of *frq* suppress transcription of *qrf*. Together *frq* and *qrf* create a negative feedback loop to regulate timing of a circadian clock. (Xue et al. 2014.) A

more recent study suggested that *qrf* could have bidirectional effect the *frq* levels in cells because *qfr* have been detected to also promote the expression of *frq* (N. Li et al. 2015).

2 Aims of the study

The aim of the study is to reveal differences between RNA-protein complexes in coding RNA transcripts and lncRNA transcripts and to understand which of these proteins mediate regulatory functions within the lncRNA RNA-protein complexes.

To achieve this goal, this study focuses on the development of a new method called RNA-Decoder to identify proteins associating with specific RNA transcripts. RNA-Decoder method is based on DNA barcoding, RNA immunoprecipitation (RIP) and high-throughput sequencing (Figure 12A-C).

RNA-Decoder method harnesses *S. cerevisiae* library, called Epi-Decoder library, which consists of thousands of yeast strains, each strain containing two unique DNA-barcodes (18 random, nucleotides) at the SUT098-GCG1 locus, one of them at the mRNA coding GCG1 region and the other is at the non-coding region SUT098, and a unique protein with a TAP-tag (Tandem immunopurification tag). Barcodes and TAP-tagged proteins make it possible to compare which proteins bind to protein coding mRNA transcripts and which to lncRNA transcripts.

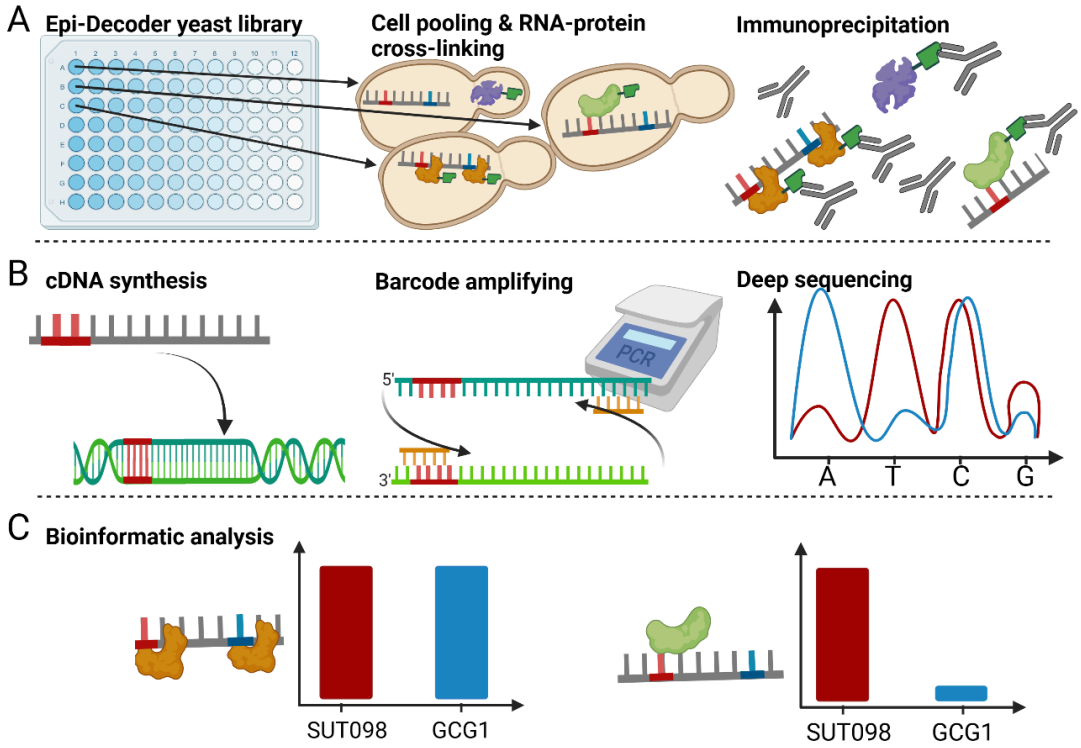

Figure 12. Design of RNA-Decoder method. (A) Yeast cells from Epi-Decoder library are pooled, proteins and RNA transcripts are cross-linked, and RNA-protein complexes are immunoprecipitated. (B) Then cross-links are removed, surroundings of barcodes from RNA transcripts are translated into cDNA, amplified by PCR, and sequenced. (C) Bioinformatic analyses are used to study amounts of different RNA transcript binding proteins. (Created with BioRender.com)

3 Materials and methods

3.1 Epi-Decoder yeast library

Epi-Decoder yeast library had been generated by crossbreeding the custom-made *S. cerevisiae MAT*α barcoder library and the commercial *S. cerevisiae MATa* TAP-tag library by utilizing SGA methodology and the RoToR machine. Epi-Decoder library contains 4 247 different yeast strains each containing two unique DNA-barcodes at the SUT098-GCG1 locus and a unique protein with a TAP-tag.

3.2 Yeast cultivation

The frozen 1 ml Epi-Decoder yeast library pellet was thawed on ice, spinned down to remove glycerol, washed with Milli-Q water and resuspended in 50 ml of YPD media (10g/L yeast extract; 20 g/L peptone; 20 g/L D-Glucose). The culture was grown overnight and diluted back to an OD600 = 0.15 in 500 ml of YPD, grown to a mid-log phase, an OD600 = 0.7 – 0.8 and divided into three 150 ml technical replicates.

3.3 Cross-linking RNA-protein complexes

RNP complexes were cross-linked by formaldehyde or UV-light. Cross-linking by formaldehyde was made by adding 37 % formaldehyde into the yeast cultures to a final concentration of ~ 1 % and shaken slowly on a platform for 15 minutes at RT. Cross-linking was stopped by adding 2 M glycine to a final concentration of ~ 2 % and incubating five minutes at RT (room temperature). UV-light induced cross-links were made by Stratagene Stratalinker UV 2400 Crosslinker (Marshall Scientific, USA). The yeast cultures were pelleted at 3 000 g, for three minutes, at RT, washed twice with ice-cold TBS (140 mM NaCl; 10 mM Tris-Cl) and resuspended in 15 ml of TBS. The cells were spread onto Ø 15 cm petri dish and exposed to 1,200 mJ/cm^2 (3 X 400 mJ/cm^2) with 2 minutes breaks and gentle mixing on ice.

3.4 Cell lysis

The cells were centrifuged at 2 500 g, for five minutes, at +4 ˚C, washed twice with ice-cold TBS and resuspended in 0.75 ml of FA lysis buffer (1 mM EDTA pH 8.0; 50 mM HEPES-KOH pH 7.5; 140 mM NaCl; 0.1 % (w/v) Sodium deoxycholate; 1 % (v/v) Triton X-100; 0,1 % (w/v) SDS; Pierce Protease Inhibitor Tablets, EDTA-free (Thermo Scientific)). The cells were centrifuged at 15 000 g, for 30 seconds, at +4 ˚C and resuspended in 1.125 ml of ice-cold FA lysis buffer, containing 40 U/ml RNasin (Thermo Scientific) to protect RNAs.

The cells were lysed by adding ø 0.5 mm Zirconia/Silica beads (Fisher Scientific) and vortexing in a multivortex at full speed, at +4 ˚C, for two twenty-minute cycles followed by five-minute rest on ice. The beads were separated from the samples, by creating a puncture hole on the bottom of the 1.5 ml microcentrifuge tube with a syringe needle, the tube was placed into 15 ml centrifuge tube and centrifuged at 2 000 g, for 2 minutes, at +4 ˚C.

3.5 DNA sonication and sample pre-clearing

The yeast samples were sonicated in a water bath using Q700 sonicator with Cup Horn (Qsonica) to shear the DNA. The cells were sonicated at an amplitude of 100 V, delivered in 20-second ON/OFF pulses, sonication lasting for 10 minutes. The efficiency of the sonication was verified by taking 10 µl of each sonicated sample, adding 10 µl of Milli-Q water and 1 µl of 5 M NaCl. The samples were incubated for 10 minutes at +95 ˚C, 1 µl of RNase A/T (Thermo Scientific) and 1 µl of DNase (Thermo Scientific), were added and the samples were incubated for 30 minutes at +37 ˚C. 0.2 µl of Proteinase K (Thermo Scientific) was added and samples were incubated 60 minutes at 56 ˚C and ran in 1% agarose TAE gel (120 V, 30 minutes).

The sonicated samples were centrifuged two-times at 21 130 g, for 10 minutes, at +4 ˚C and a bead volume of 75 µl of Sepharose CL-4B (Sigma Aldrich) was added to pre-clear the extract. The samples were incubated with end-over-end rotation for 1 h, at +4 ˚C and centrifuged at 1 000g, for 2 minutes, at +4 ˚C to remove the Sepharose beads. The samples were adjusted to 25 mM $MgCl_2$ and 5 mM $CaCl_2$, optimizing the environment for DNase I. 3 µl of RNasin (40 U/µl, Thermo Scientific) was added into the samples to protect the RNAs and 6 µl of DNase I (Invitrogen) was added to remove the DNA. The samples were incubated for 20 minutes, at +37 ˚C and reaction was stopped by adding 0,5 M EDTA to a final concentration of 20 mM. The samples were centrifuged 21 130 g, for 10 minutes, at +4 ˚C and the extracts were transferred into 1,5 DNA loBind tubes.

3.6 Immunoprecipitation

The input RNA samples were prepared by diluting 50 µl of the pre-cleared sample into 100 µl of RIP elution buffer (10 mM EDTA; 1 % SDS; 100 mM Tris-Cl, pH 8.0).

Dynabead™ Protein G (Invitrogen) or Dynabead™ Protein A (Invitrogen) were coated with Rabbit IgG antibodies (Invitrogen) by diluting 100 µl of Dynabeads to 200 µl of PBS (0.137 M NaCl; 0.0027 M KCl; 0.01 M Na_2HPO_4; 0.0018 M KH_2HPO_4), containing 1 mg/ml of Rabbit IgG antibodies and incubated for 4 hours at +4 ˚C. For the negative, non-antibody control, Dynabeads were incubated in 200 µl of PBS for 4 hours, at +4 ˚C. After incubating,

Dynabeads were resuspended into 50 µl of PBS, containing 0,02 % Tween-20 to prevent aggregation. Resuspended Dynabeads were added into 500 µl of the pre-cleared sample and incubated with end-over-end rotation for overnight, at +4 ˚C, to couple TAP-tags of RNA-proteins complexes with IgG antibodies.

After incubation, Dynabeads were collected by using a magnet and supernatant was removed. Dynabeads underwent a series of washes by using 1 ml of ice-cold FA lysis buffer, FA500 buffer (1 mM EDTA, pH 8.0; 50 mM HEPES-KOH, pH 7.5; 500 mM NaCl; 0.1 % (w/v) sodium deoxycholate, 1 % (v/v) Triton X-100), Li-Cl wash buffer (1 mM EDTA, pH 8.0; 250 mM LiCl; 0.5 % v/v Nonidet P-40; 0.1 % w/v sodium deoxycholate; 10 mM Tris-Cl, pH8.0) and TE buffer (100 mM Tris-Cl, pH 8.0; 10 mM EDTA, pH 8.0). For each wash, Dynabeads were collected with a magnet, the supernatant was removed, beads were resuspended into next washing buffer and incubated with end-over-end rotation, for five minutes, at +4 ˚C.

Following the washes, RNA-protein complexes were eluted from Dynabeads by resuspending Dynabeads into 75 µl of RIP elution buffer, containing 40 U/ml of RNasin and incubating in a Thermomixer at 1 400 rpm, for 10 minutes, at +37 ˚C. The eluate was transferred into a new 1,5 ml DNA loBind tube, the elution was repeated and both eluates were pooled together.

3.7 Immunoprecipitation analysis

The 20 µl samples for protein analysis were collected after DNA sonication (insoluble sample), after diluting input sample to RIP buffer (input sample), after immunoprecipitation from the final washing buffer (unbound sample) and after immunoprecipitation from RIP buffer containing eluted RNA-protein complexes (IP). The samples were diluted in 1:9:2:10 ratio to ensure that differences observed among the samples on the membrane were not caused by varying initial concentrations.

The samples were diluted 1:1 to SDS loading buffer (0.25 % bromophenol blue, 0.5 M DTT, 50 % glycerol, 10 % SDS, 0.25 M Tris-Cl, pH6.8), denatured for five minutes, at +95 ˚C and loaded into Criterion™ TGX Stain-Free™ Precast gel 4-15 % (Bio-Rad). The gel was run for 45 minutes at 180 V. Proteins were transferred into Trans-Blot® Turbo™ Mini PVDF Transfer Pack (Bio-Rad) by using Trans-Blot Turbo Transfer System (Bio-Rad, USA) at 2.5 A, at 25 V, for 10 minutes. The PVDF membrane was blocked for 2 h on a shaker at RT in 5% low-fat milk powder in PBS and incubated in 3% low-fat milk powder in PBS containing 10 µg/ml IgG Rabbit antibody (Invitrogen), on a shaker at + 4 ˚C for overnight. The membrane was rinsed two times with PBS and incubated in 1.4 % low-fat milk powder in PBS containing 1:10 000 Polyclonal Swine Anti-Rabbit Immunoglobulins/HRP (Dako) for 1

h, at RT. The membrane was washed 2 x 10 minutes in PBS and 10 minutes in PBS-Tween at RT on a shaker. Images were generated by using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) and ChemiDoc MP imaging System (Bio-Rad).

3.8 Cross-link reversal and RNA sample purification

5 M NaCl was added into each sample (input, immunoprecipitated (IP) and non-antibody control) to a final concentration of ~ 200 mM to adjust the conditions suitable for RNA-protein cross-link reversal. 20 µg of Proteinase K (Thermo Scientific) was added into each sample to reversal the cross-links, and the samples were incubated for 1 hour, at +42 ˚C to remove cross-linked polypeptides and for 1 hour at +65 ˚C to unravel the RNA-protein cross-links.

Following the cross-link reversal, nuclease-free water (Promega) was added up to 400 µl to the samples. The RNA was purified and collected by using miRNeasy Mini Kit (Qiagen), following the kit's protocol. The only exception was on a last step, when the RNA was diluted in 30 µl of nuclease-free water.

Remaining DNA fragments were removed from the samples by using Turbo DNA-*free*™ -kit (Invitrogen), following kit's protocol.

3.9 cDNA synthesis

Input samples and immunoprecipitated samples were translated into cDNA by using SuperScript IV Reverse Transcriptase -kit (Thermo Scientific), following kit's protocol. Used primers were 5'-CTATGCCACCCGTGTTTCTCC for SUT098 and 5'-TGGGTTTGCAGGTGTACC for GCG1 cDNA synthesis.

3.10 qPCR

qPCR was carried out by using C1000Touch Thermal Cycler (Bio-Rad), using 10 µl qPCR reactions containing GoTaq qPCR Master Mix (Promega), 3 µl of template cDNA, 500 nM forward primer and 500 nM reverse primer. Used primers were 5'-TACTTTGCCAGAAACGACGC and 5'-TTACAAGCCTCCGTCTCATTA for GCG1 samples and 5'-CAATTTCTGGCAAACGACGC and 5'-AGTACTTGACTGCTTCGTCTTC for SUT098 samples. The initial denaturation was 2 minutes at 95 ˚C followed by qPCR-reaction cycle (15 seconds at 95 ˚C, 1 minute at 60 ˚C and 15 seconds at 72 ˚C) repeated 39 times and 5 minutes at 72 ˚C. Melting curve analysis

was performed in the range of 65°C to 95°C, 0.5°C per 5 seconds increments. The data was analysed by using delta-delta Ct ($2^{-\Delta\Delta Ct}$) method.

3.11 Barcode amplification

Barcode regions were amplified by PCR reactions. The PCR conditions (extension and annealing times and temperatures, number of the cycles, primer concentration, magnesium concentration, amount of used cDNA, used PCR buffer and DNA polymerase enzyme) were optimized for GCG1 IP, GCG1 input, SUT098 IP and SUT098 input samples meaning that parameters of reactions changed during the study. The final PCR reactions were 25 µl reactions containing 10 X Taq buffer with $(NH_4)2SO_4$ (Thermo Scientific), 1.5 mM MgCl2, 0.2 mM dNTPs, 250 nM forward and reverse primers and 6.5 U Taq DNA polymerase (Thermo Scientific). The volume of template cDNA was 4 µl for input PCR reactions and 6 µl for IP PCR reactions. Primers used in PCR reaction contained P7 and P5 adapters for Illumina sequencing and they are listed in appendix 1. The initial denaturation was 2 minutes at 95 ˚C followed by PCR-reaction cycle (30 seconds at 95 ˚C, 15 seconds at annealing temperature and 15 seconds at 72 ˚C) and 5 minutes at 72 ˚C. The annealing temperature for GCG1 samples was 58.1 ˚C and 55.8 ˚C for SUT098 samples. The PCR -reaction cycle was repeated 29 – 35 times for IP samples and 25 times for input samples. The success of the PCR was perceived in the 2 % agarose gel, ran at 120 V for 60 minutes or on the 6% native polyacrylamide gel (native-PAGE), ran at 300 V, for 65 minutes. The gel was photographed by ChemiDoc MP imaging System (Bio-Rad) and analysed by ImageJ.

3.12 Barcode library assembly

The PCR reactions from barcode amplification from technical replicates were pooled together to assemble GCG1 IP, GCG1 input, SUT098 IP and SUT098 input libraries. The libraries were run in the 2 % agarose gel at 120 V, for 60 minutes and extracted and purified from gel by using GeneJET Gel Extraction Kit (Thermo Scientific) following kit's protocol but at the end the samples were eluted into nuclease-free water (Promega) instead of elution buffer. The DNA concentrations of the libraries were measured by Qubit 4.

3.13 Artificial Intelligence

An AI language model, ChatGPT, developed by OpenAI, was utilized as a proofreading tool for fixing grammatical errors and suggesting improvements for clarity throughout this thesis.

4 Results

The results section is divided into two parts: the first section, “Small-scale RNA-Decoder experiment” contains the results from the developing process of RNA-Decoder method, while the second section, “Big-scale Epi-Decoder method” contains the results from RNA-Decoder experiment harnessing the whole Epi-Decoder yeast library.

4.1 Small-scale RNA-Decoder experiment

4.1.1 Epi-Decoder library yeast strain selection for method development

In total, twelve strains from Epi-Decoder yeast library were cultivated, lysed and TAP-tagged fusion proteins were detected by using Western blot analysis. The purpose of these cultivations and Western blot analysis was to verify that the yeast cells still expressed the correct TAP-tagged fusion proteins and to select a couple of yeast strains from Epi-Decoder library for development of RNA-Decoder method.

The selection of twelve yeast strains was based on *Saccharomyces* genome database, focusing on known interactions between wild type versions of expressed TAP-tagged fusion proteins and RNA-transcripts. Among these twelve strains, four expressed TAP-tagged fusion proteins which wild type versions participate cell metabolism (Adh2, Pyk2, Gly1 and Tor1), and these proteins were not expected to attach to RNA transcripts in the cells. Another four yeast strains expressed fusion proteins which wild type versions were known to interact RNA transcripts in nucleus (Nab2, Nab3, Mpp6 and Sto1), and the remaining strains expressed fusion proteins which wild type versions are capable to bind to RNA transcripts in nucleus and in cytoplasm (Hek2, Pab1, Ski2 and Tif1). The sizes of the TAP-tagged fusion proteins are detailed in the table 2. The sizes of the TAP-tagged fusion proteins were determined based on the protein information from *Saccharomyces* Genome Database and the known size of TAP-tag, which is 21 kDa (Gloeckner et al. 2007).

Table 2. The sizes of the TAP-tagged fusion proteins.

Name of the TAP-tagged fusion protein	Size of the TAP-tagged fusion protein (kDA)
Adh2	58
Gly1	64
Hek2	63
Mpp6	42
Nab2	79
Nab3	111
Pab1	85
Pyk2	76
Ski2	167
Sto1	121
Tif1	66
Tor1	49

The table 2 and the figure 13 together illustrate, that almost all of the twelve selected yeast strains expressed the correct sized TAP-tagged fusion proteins. The fusion proteins Pab1 and Pyk1 were not detected in any of the repeated Western blot analysis or replicate Epi-Decoder yeast cultivations, while Gly1 and Mpp6 were detected in the repeated Western blot analyses (figure does not show). Based on the results from Western blot analysis, four yeast strains - Adh2, Hek2, Nab2 and Nab3 - were selected for the development of RNA-Decoder method.

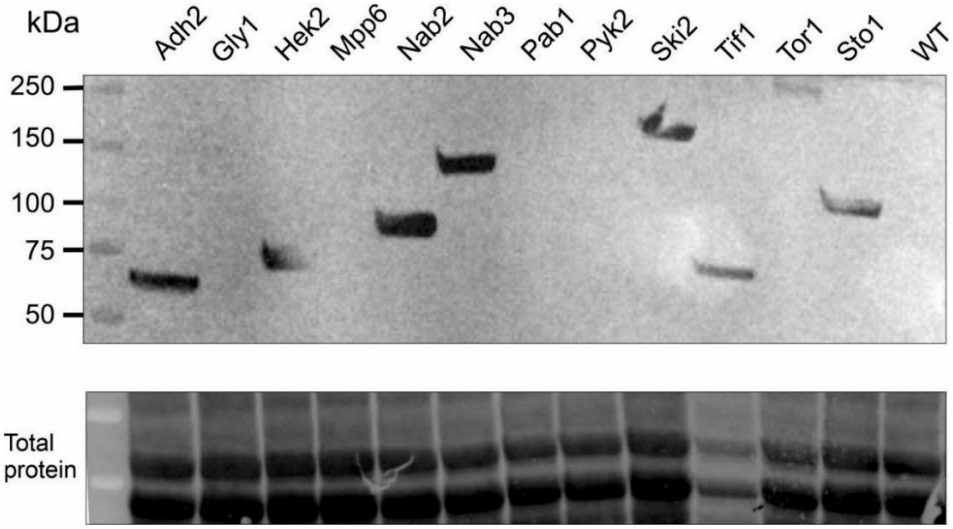

Figure 13. Western blot analysis of twelve tested Epi-Decoder strains alongside wild type *S. cerevisiae* (WT) yeast. WT yeast serves as a negative control of immunoprecipitation of TAP-tagged fusion proteins. The protein band on Ski2 lane is malformed due the rough handling of membrane.

4.1.2 Selection of cross-linker for RNA-protein complexes

RNP complexes can be studied in native conditions or denaturing conditions to stabilize RNP complexes. In native conditions, RNP complexes are immunoprecipitated and purified under physiological conditions of the cells. While this method offers the possibility to purify RNP complexes as they appear in the cells, it also has its disadvantages. Without using any kind of denaturing agent, it is possible that RNP complexes rearrangement during cell lysis or immunoprecipitation steps. (McHugh et al. 2014.)

In denaturing conditions, RNP complexes are cross-linked in the cells before purification by using a cross-linker. Common cross-linkers are chemical cross-linkers, like formaldehyde or UV-light. The general advantage of using a cross-linker is their stabilization effect to RNP complexes preventing rearrangement of RNP complexes in later steps. UV-light and formaldehyde cross-link RNP complexes in distinct ways, both having their own benefits and weaknesses. The benefit of using UV-light is cross-linking RNAs and proteins only at their interaction points preventing formation of incorrect cross-links. The disadvantage of UV-light induced cross-links is that they are irreversible, meaning that protein-RNA cross-links may inhibit later PCR related reactions and prolonged UV-light exposure can damage the sample. The advantage of formaldehyde created cross-links is that they are reversible allowing utilization of different types of PCR reactions to detect the RNAs. On the other hand, formaldehyde can create unintentional cross-links for example protein-protein crosslinks making identification of RNA-protein complexes from RNA-protein-protein complexes challenging. (Majumder and Palanisamy 2021.)

The necessity of using a cross-linker before immunoprecipitation was investigated by comparing the success of the immunoprecipitation under native conditions and denaturing conditions. In denaturing conditions, RNP complexes were cross-linked using either UV-light or formaldehyde. To perform the comparison, RNP complexes were pulled down by targeting the TAP-tags of the RNA associated proteins. Next, previously created cross-links were reversed, and RNA transcripts were reverse transcribed into cDNA by using specific primers that bind on the SUT098 and GCG1 areas. The cDNA transcripts were amplified by qPCR using the specific primers to the barcode areas of SUT098 and GCG1 (Figure 14). Cycle threshold values were analysed using the $2^{-\Delta\Delta Ct}$ method (Figure 15), to compare native conditions and different cross-linkers (Figure 15).

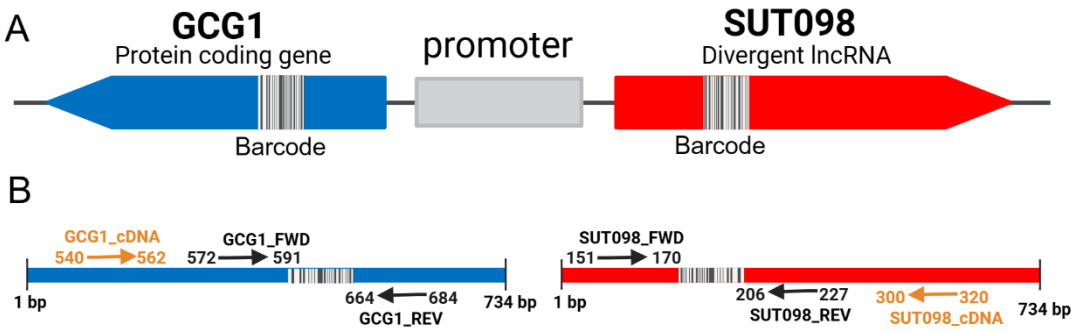

Figure 14. The locus of interest. (A) The divergent transcription locus GCG1-SUT098 containing the DNA barcodes. The direction of the arrow represents the direction of the transcription. (B) A schematic representation of the primers used to convert the immunoprecipitated RNA transcripts into cDNAs (orange) and amplify the barcode regions from cDNAs (black). The representation is not in a scale. (Created with BioRender.com)

The Figure 15 shows, that the smallest amount of the RNA was recovered from the cells under native conditions. Without cross-linking the RNP complexes, the amount of GCG1 cDNA was so low that it was not detected at all during the qPCR run. The amount of SUT098 cDNA was a slightly higher (Figure 15A). Using UV-light or formaldehyde as a cross-linker resulted in higher RNA recovery (Figure 15B - C).

In general, the RNA recovery of UV-light cross-linked RNP complexes was higher from RNP complexes interacting with GCG1 transcripts than those interacting with SUT098 (Figure 15B). The result was the opposite than expected because the previous research has detected higher levels of divergent transcription than protein coding transcription from GCG1-SUT098 loci (Gowthaman et al. 2021). It was assumed, that higher levels of SUT098 transcription than GCG1 transcription would result in a higher amount of SUT098 associated RNP complexes than GCG1 linked RNP complexes leading to higher SUT098 cDNA recovery than GCG1 cDNA. The only exception was the Adh2 sample where GCG1 cDNA was not detected during the qPCR run and the recovery of SUT098 cDNA was low (Figure 15B). This was the expected result because Adh2 should not interact with RNA transcripts in cells.

The best RNA recovery was obtained from the formaldehyde cross-linked samples. Formaldehyde cross-linked proteins strictly to both GCG1 and SUT098 RNA transcripts resulting in the successful recovery of RNA from all the samples (Figure 15C). In most cases, formaldehyde cross-linking resulted in higher $2^{-\Delta\Delta Ct}$ values and better recovery of RNA transcripts compared UV-light crosslinking or immunoprecipitation performed in native conditions. For example, $2^{-\Delta\Delta Ct}$ value of formaldehyde cross-linked SUT098 Adh2 and Hek2 samples were 75.3 times and 67.4 times higher than in UV-cross-linked samples. The only

exceptions were GCG1 Nab2 and Nab3 samples for which UV-cross-linked samples exhibited higher $2^{-\Delta\Delta Ct}$ values than the corresponding formaldehyde cross-linked samples (Figure 15 B – C).

The error bars implicate that there was high variation in the $2^{-\Delta\Delta Ct}$ values among formaldehyde treated replicates (Figure 15C). Formaldehyde may have been too effective cross-linker, creating cross-links between RNA transcripts and proteins that do not normally interact with each other.

Formaldehyde was selected as a cross-linker for RNA-Decoder experiment due to its best recovery of RNA compared to native conditions and UV-light cross-linking.

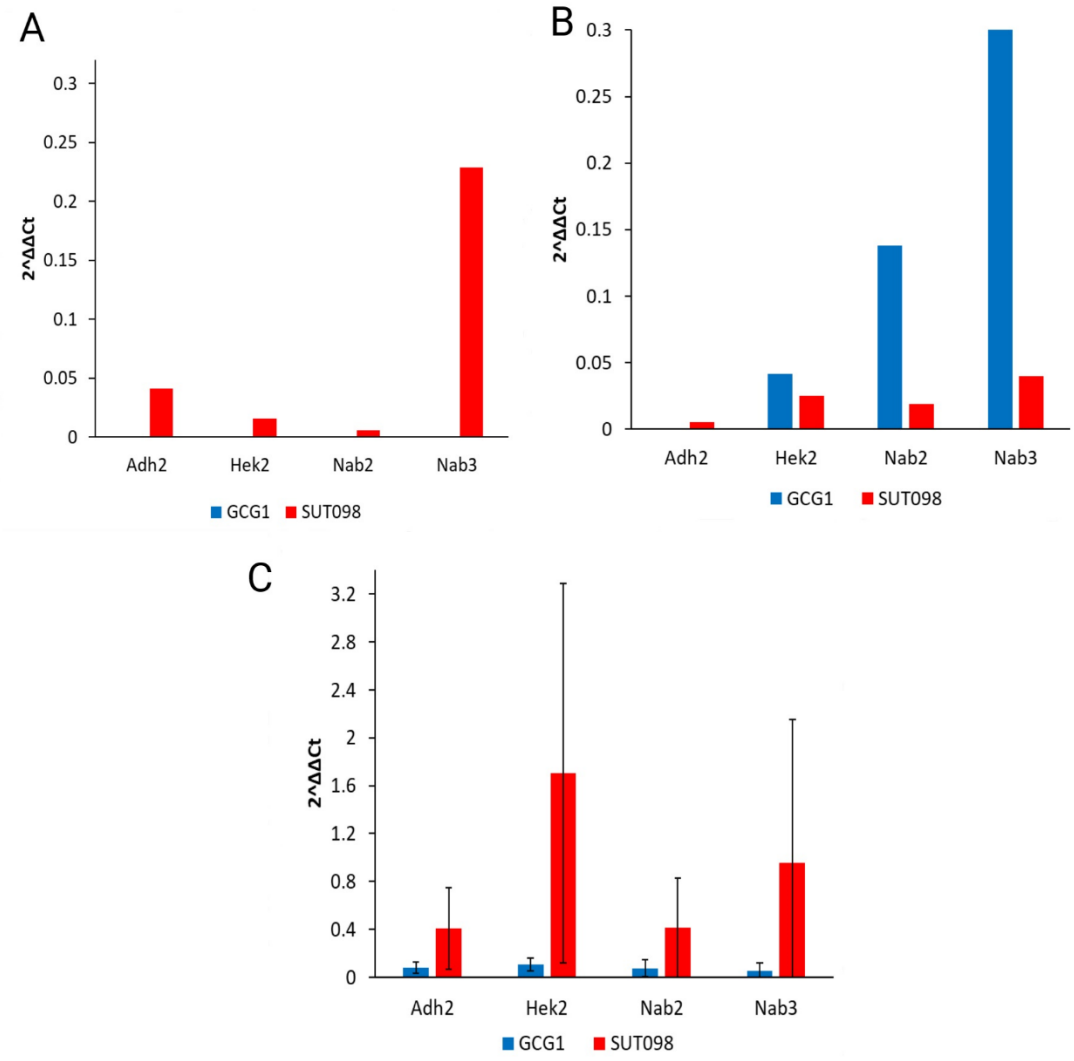

Figure 15. Comparison of RNA recovery from different cross-linking conditions. RNA was recovered from native conditions (A) and denaturing conditions where RNP complexes were cross-linked by UV-light (B) or formaldehyde (C). Data was normalized against wild type *S. cerevisiae.* The error bars represent standard deviation among the replicates. The Y-axel is formed to start from zero because the amount of RNA cannot have negative values. It is important to note the different maximum value of the Y-axel on the figure15 C.

4.1.3 Dynabead A and Dynabead G comparison

The cross-linked RNP complexes were immunoprecipitated by using IgG coated Dynabeads. During immunoprecipitation TAP-tags of the RNP complexes bound to IgG antibodies. TAP tag contains two different domains: a calmodulin binding peptide, which is coupled to the protein of interest and Protein A, which has an affinity for IgG antibodies.

The Dynabeads compared in this study have either protein A or protein G covalently coupled on the surface of the beads. During the method development, immunoprecipitation was performed by using both Dynabead As and Dynabed Gs to compare their affinity against TAP-tagged RNP complexes and their effect of on the immunoprecipitation efficiency. In both cases, more RNA from SUT098 and GCG1 samples was recovered by using IgG coated Dynabead Gs (Figure 16A-B). However, there was a significant variation in the distribution among the replicates of Dynabead A.

Dynabead Gs were selected to use in immunoprecipitation for RNA-Decoder experiment due their better RNA recovery results compared to Dynabead As.

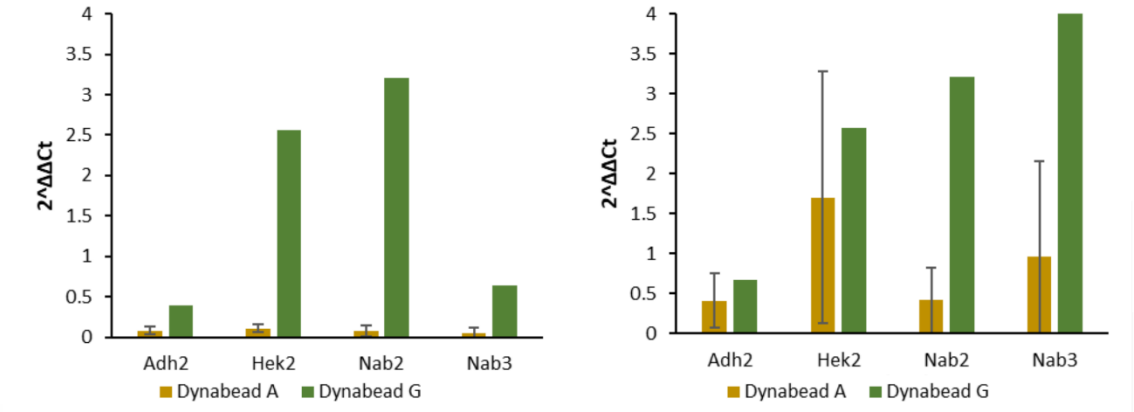

Figure 16. Comparison of recovered RNA using Dynabead As and Dynabead Gs. Recovered RNA from GCG1 (A) and SUT098 samples (B). RNP complexes were formaldehyde cross-linked, immunoprecipitated by using Dynabead As or Dynabead Gs, transcribed into cDNA, and amplified by qPCR. The error bars represent standard deviation among the replicates. The Y-axel is formed to start from zero because the amount of RNA cannot have negative values.

A one explanation for the higher RNA recovery using Dynabead Gs might be the different affinity of IgGs antibodies for protein A and protein G. Since both TAP-tags and Dynabead As contain protein A there might be more available binding sites for TAP-tags on the surface of IgG coated Dynabead G's than the surface of IgG coated Dynabead A's. This is because the protein G of Dynabeads and protein A domain of TAP-tag do not necessarily compete for the same binding spots on IgG antibodies.

4.1.4 Optimizing PCR reactions for barcode amplification

The immunoprecipitated GCG1 and SUT098 RNA transcripts needed to convert to cDNA and the barcoded regions of the cDNA needed to amplify by specific Illumina primers via PCR (Figure 17). Barcode amplification facilitated assembly of barcode libraries from PCR products, sequencing of barcodes, and analyse the proteins associated to RNP complexes.

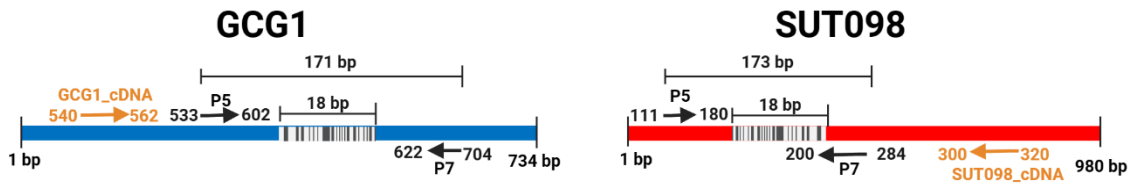

Figure 17. A schematic representation of the primers used in both small-scale and big-scale experiments. GCG1_cDNA and SUT098_cDNA primers (orange) are used to convert the immunoprecipitated RNA transcripts into cDNAs, while Illumina primers (black) are used to amplify the barcode regions from cDNAs. The size of correctly amplified GCG1 barcode region is 171 basepairs (bp) and SUT098 barcode region is 173 bp, while the barcodes themselves are only 18 bp. The representation is not in a scale. (Created with BioRender.com)

The conditions of the PCR reactions needed optimization due to several factors affecting the success of the barcode amplification. In this experiment, the annealing temperature, number of the cycles, primer concentration, magnesium concentration, amount of cDNA used, and DNA polymerase enzyme and PCR buffer were optimized for both input and IP (immunoprecipitated) samples for both regions GCG1 and SUT098. The number of cycles was kept as low as possible to minimize amplification bias and nucleotide miscorporation of DNA polymerase that could lead skewed results. GCG1 primers caused difficulties during the amplification of the barcode area at GCG1 region because they formed unintentional primer-dimers.

The PCR optimization for SUT098 IP and input samples was successful. As shown in the figure 18 SUT098 primers did not form unintentional artifacts during the PCR reaction. The optimal annealing temperature was determined by gradient PCR (Figure 18A). Additionally, the required volumes of cDNA from IP and input samples to obtain clear, visible bands in agarose gel were determined experimentally (Figure 18B-C). However, the faint bands on the negative control lane in the figure 18C (cDNA synthesis reaction without reverse transcriptase enzyme), suggest that the DNAse treatment of the input samples was not efficient enough.

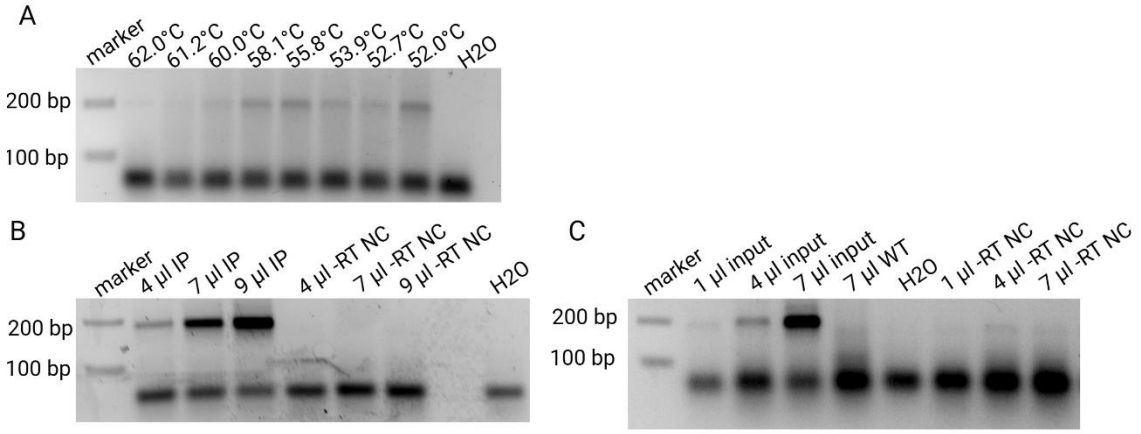

Figure 18. PCR reactions optimization for SUT098 IP and input samples. (A) The optimal annealing temperature was determined by gradient PCR, and it was found to be 55.8 °C. (B - C) The required volume to achieve distinct bands for gel extraction were determined to be 7 µl for IP samples and 4 µl for input samples. The faint bands on the negative control lanes of the input samples (-RT NC) could imply incompetent DNAse treatment.

The optimization of PCR reaction for GCG1 IP and input samples proved to be challenging due to primer-dimer formation during the PCR reaction. As can be seen from the figure 19, primer-dimers were almost same size as the desired, correct size PCR product, making the detection and extracting of the right PCR product from agarose gel very challenging.

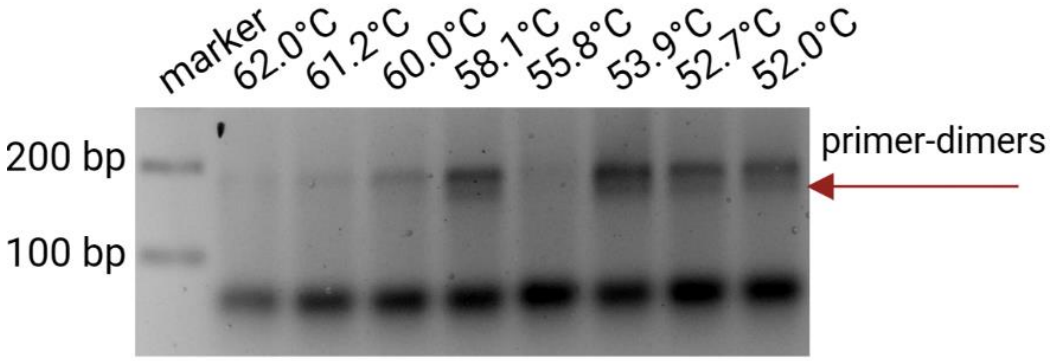

Figure 19. Gradient PCR for GCG1 samples. The optimal annealing temperature for GCG1 region was determined by gradient PCR. Primer-dimers formed during the run can be detected as blurry bands located just below the PCR product bands. Primer-dimers were formed nearly all temperatures. The optimal annealing temperature was found to be 58.1 °C. At this temperature, the band of the PCR product is strong and the most distinct from the primer dimer band.

Various PCR conditions, such as different PCR buffers, DNA polymerases, extension and annealing temperatures, primer concentrations, Mg^{2+} -concentrations and adding DMSO to the PCR reaction, were tested to eliminate primer-dimers. However, it proved to be an impossible task to find the conditions to eliminate them permanently. The results varied among PCR runs – sometimes primer-dimers formed during the run and sometimes they did

not appear at all even though the conditions (including pipetting order, sample handling and PCR program) were kept exactly the same (Figure 20A-B).

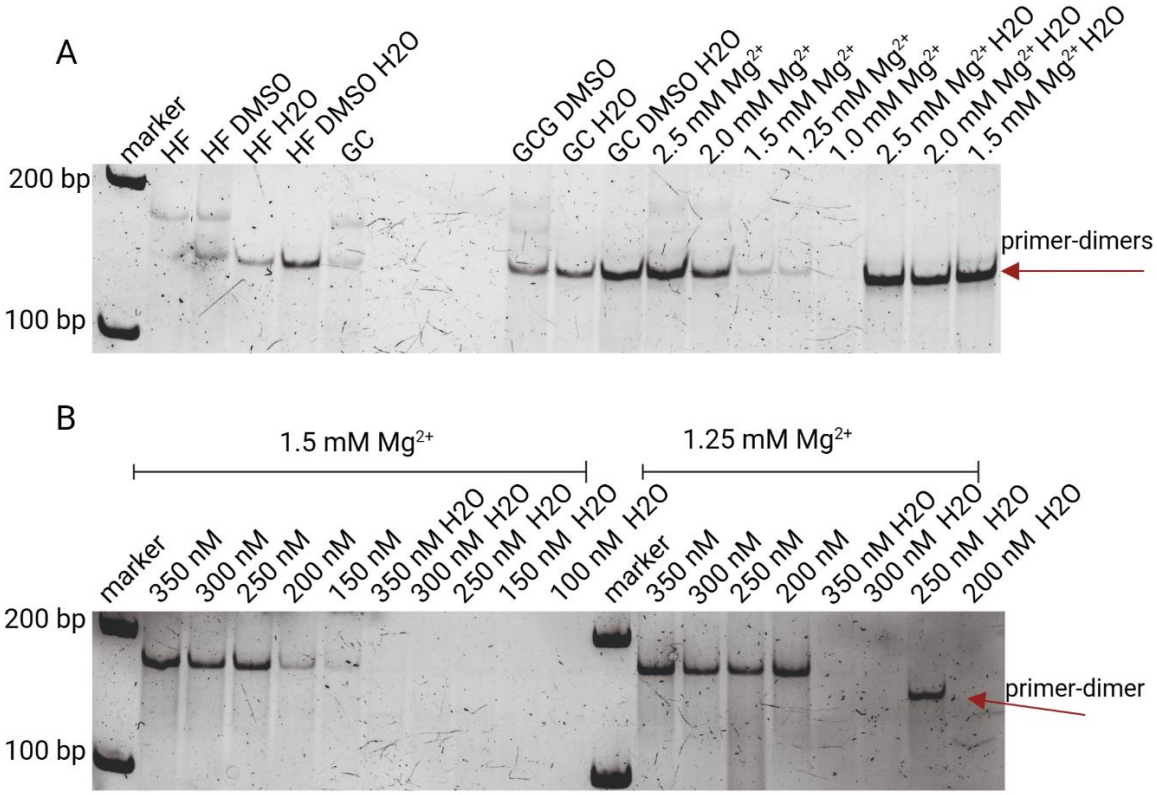

Figure 20. Optimizing the PCR conditions for GCG1 samples. (A) Phusion High-Fidelity buffer (HF) (New England Biolabs), Phusion GC buffer (GC) (New England Biolabs) and Taq buffer with different magnesium concentrations were compared in presence and absence of DMSO. Primer-dimers were formed in each condition, even in negative control samples (H2O). The only exception was PCR reaction in Taq buffer with 1.0 mM Mg^{2+} but in that condition concentration of magnesium was too low for the correct PCR product to be detected. (B) The effect of primer concentration (100 nM – 350 nM) on the formation of primer-dimers. The most important factors to avoid primer-dimers during PCR reaction were usage of Taq-polymerase and Taq buffer with 1.25 – 1.5 mM magnesium concentration. However, the complete elimination of primer-dimer formation was impossible. Primer-dimers formed spontaneously in Taq buffer with 250 nM Mg^{2+} PCR reaction but not in any other PCR reactions.

The best possible explanation for formation of primer-dimers is that the affinity of GCG1 primers to each other is higher than their affinity to the GCG1 barcode region. Sometimes GCG1 primers attach each other spontaneously right after the first denaturation step of PCR. If the primers attach each other spontaneously during the beginning of the reaction, the formed artifact interacts with more primers after each PCR cycle, creating a primer-dimer artifact which can be detected on the gel electrophoresis after PCR.

4.2 Big-scale RNA-Decoder experiment

4.2.1 Chromatin shearing

The chromatin from lysed yeast cells was needed to be sheared before immunoprecipitation to enhance the effectiveness of immunoprecipitation. Chromatin fragmentation increases the efficiency of RNA immunoprecipitation because non-fragmented nucleic acids might disturb the accessibility of antibodies to target the specific RNPs during the immunoprecipitation (Gilbert and Svejstrup 2006).

The shearing of the chromatin was performed by sonication and the efficiency of sonication was perceived by digesting proteins, DNA and RNAs from sonicated samples and running the samples on agarose gel. Sonication sheared the nucleic acids for smaller fragments, they can be observed as long, faint, smear bands on the gel (Figure 21). The sizes of the DNA fragments varied between 100 bp and over 1 000 bp.

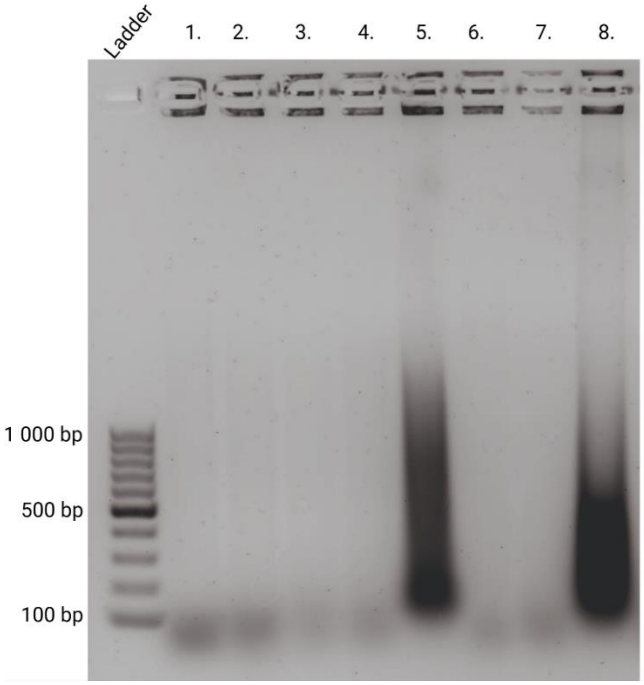

Figure 21. Fragmented chromatin. The lanes 1 – 3 represent the technical replicates from the same Epi-Decoder yeast library while the lane 4. represents wild type *S. cerevisiae* yeast used as a negative control strain for immunoprecipitation. The samples on the lanes 1 – 4 were treated with Proteinase K and Rnase A/T to remove proteins and RNA. The faint long bands on the lanes 1 – 4 indicate successful DNA fragmentation. The lanes 5. – 8. were control samples for Proteinase K and Rnase A/T treatments. The sample on the lane 5. did not receive any proteinase K, Dnase or Rnase A/T treatment meaning that the long, blurry, dark band contains mostly proteins and RNA. The efficiency of sonication is impossible to detect on the lane 5. due protein and RNA interference. The samples on the lanes 6. And 7. received Proteinase K, Dnase I and Rnase A/T treatments – these lanes are empty proven that long, faint bands on the lanes 1–4 are fragmented DNA. The sample on the lane 8. was treated with proteinase K indicating that both Proteinase K and Rnase A/T treatments are needed to observe the success of the sonication.

4.2.2 Immunoprecipitation

The success of the immunoprecipitation of TAP-tagged RNP complexes was confirmed by Western blot analysis. The Western blot targeted TAP-tags of fusion proteins using Anti-Rabbit IgGs. Four sample types were collected and analysed: insoluble, input, unbound and immunoprecipitated (IP).

The insoluble sample, collected after chromatin shearing, represents all insoluble proteins in the lysis buffer from Epi-Decoder library. The input sample represents the entire protein content of the cell lysate including both specific TAP-tagged RNP complexes and all other nonspecific proteins from the yeast cells. The unbound sample was collected after immunoprecipitation, from the final washing buffer of Dynabeads and it contains the proteins which were not specifically captured by immunoprecipitation. The IP sample was taken after immunoprecipitation, from the pool containing eluted RNP complexes, representing all TAP-tagged proteins which were pulled-down by immunoprecipitation.

As shown in the figure 22, the lane representing the IP sample exhibits clearer and more visible protein bands compared to the control lanes representing the insoluble and input samples before immunoprecipitation. This means that immunoprecipitation captured majority of TAP-tagged fusion proteins. However, there are still some not captured TAP-tagged proteins in the unbound sample, indicating that not all of the TAP-tagged fusion proteins were captured during the immunoprecipitation. The negative control where immunoprecipitation was performed without the Anti-Rabbit IgGs (IP -AB), shows an empty lane. This indicates that only TAP-tagged proteins have been pulled down, confirming that the immunoprecipitation has targeted only the RNP complexes of interest.

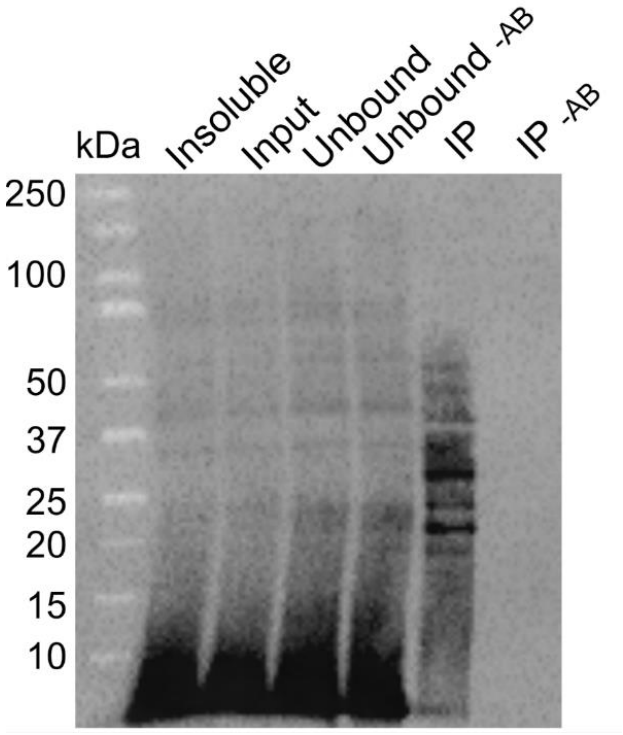

Figure 22. The success of immunoprecipitation was confirmed by Western Blot analysis. The lane of the immunoprecipitated sample (IP) contains more distinct protein bands than the insoluble, input, and unbound samples. The unbound samples contain the proteins which were not captured by antibody-coated Dynabeads during immunoprecipitation. By comparing the negative controls (IP -AB and unbound -AB), can be perceive that no RNP complexes were pulled-down without antibody coated Dynabeads indicating that RNPs and other proteins remained in the discarded cell lysate.

4.2.3 Barcode amplification and concentrations of assembled libraries

The barcode regions for SUT098 input and IP, as well as GCG1 input and IP samples were amplified by PCR and ran in the native-PAGE or agarose gel to perceive the success of the barcode amplification and to extract the barcodes for library assembly and subsequent sequencing.

The amplification of SUT098 input and IP samples was adequate, consistent with the previous optimization of PCR reaction. No primer-primers or other artifacts detected on the gel, indicating successful amplification of both input and IP samples (Figure 23). However, the bands of the IP samples were faint and almost invisible. This suggests that either the immunoprecipitation process should be modified for more efficient or barcode amplification would require more cycles to achieve fainter bands.

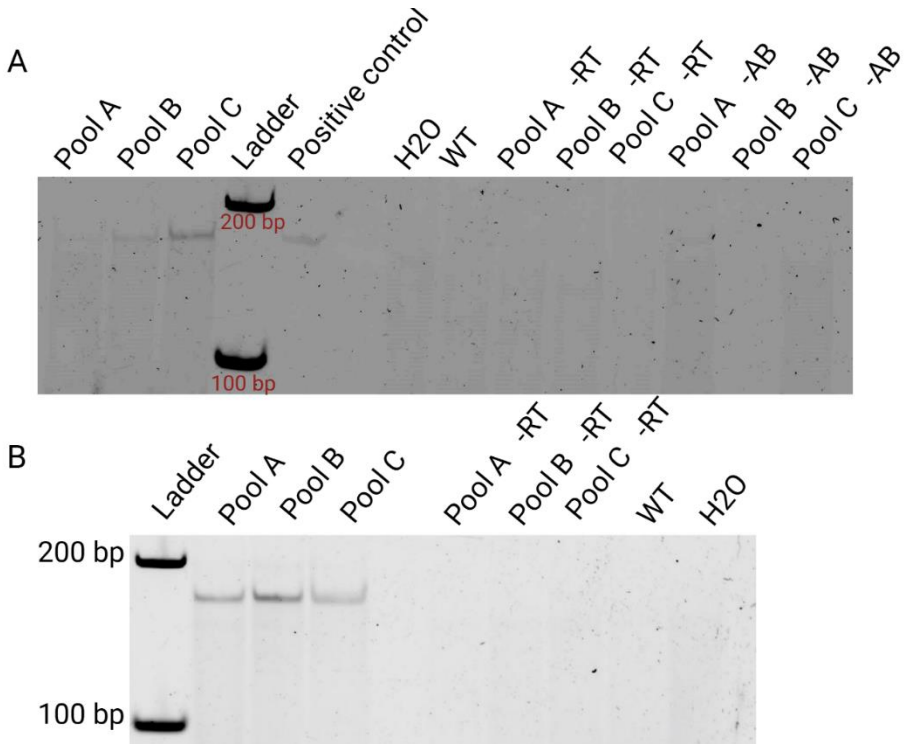

Figure 23. Amplification of SUT098 barcode regions from IP and input samples. (A) The amplified SUT098 IP barcode regions are only barely noticeable on the gel indicating that more optimization for PCR would be needed. There are no visible bands on the reverse transcription control (-RT) or antibody control (-AB) lanes indicating that DNAse treatment has been successful and Dynabeads have targeted only TAP-tagged RNPs. (B) The amplification of SUT098 input samples has been successful, the right size bands can be detected from all technical replicates (pool A, B and C) and there are any undesired bands on the control lanes.

The amplification of GCG1 barcode regions turned out to be significantly more challenging compared to the amplification of SUT098 samples, consistent with prior PCR reaction optimization experiments. Primer-dimers formed during the amplification of GCG1 IP barcodes, can be detected on the gel. Also, the bands of the amplified GCG1 IP samples are smear, faint and nearly invisible, indicating potential issues with immunoprecipitation (Figure 24).

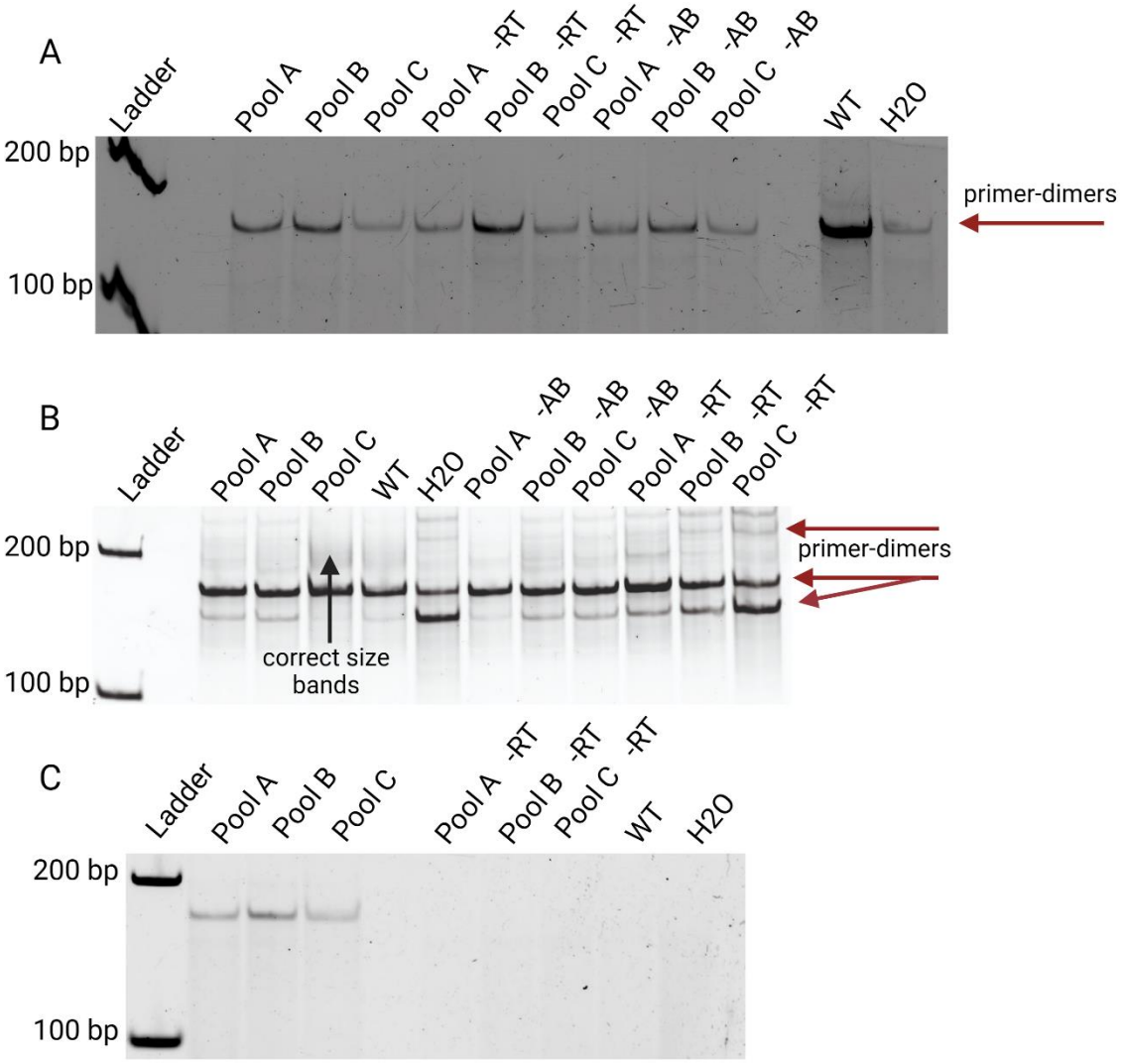

Figure 24. The amplification results of GCG1 barcode regions from IP and input samples. The amplification of GCG1 barcode regions from IP samples turned out to be challenging. (A) Using a PCR reaction with 29 cycles, no correct size bands were observed on the sample lanes (Pool A, B and C) but primer-dimers can be detected on all lanes. (B) Increasing the number of cycles to 35 resulted in the detection of faint smear bands of the correct size on the sample lanes (Pool A, B and C), but also led to a higher occurrence of incorrect artifacts. Notably, faint, correct size bands were also observed on the reverse transcriptase control (-RT) and immunoprecipitation control (-AB) lanes. The presence of these bands complicates the determination of the correct size bands on the sample lanes – do they represent amplified barcode regions or just artifacts. (C) The amplification of GCG1 input samples was successful. The correct size bands can be observed on the sample lanes (Pool A, B and C) and no undesired bands observed on the control lanes.

The amplified barcode regions of SUT098 input, SUT098 IP, GCG1 input and GCG1 IP samples were extracted from the agarose gel and the technical replicates for each sample (Pool A, B and C) were combined and pooled together for sequencing. The concentrations of the pooled samples were determined by Qubit 4 fluorescence-based assay to evaluate the success of the barcode amplification and extraction before sequencing. The concentrations of the extracted samples are presented in the table 3.

Table 3. Sample concentrations according to Qubit 4.

Library name	Original sample concentration (ng/µl)
SUT098 IP	0.362
SUT098 input	13.8
GCG1 IP	7.02
GCG1 input	1.19
GCG1 -RT	1.21
GCG1 -AB	2.54

Overall, the concentrations of barcoded libraries are notably low, except for SUT098 input barcode library concentration. One explanation for low concentrations could be the short length of the desired, correct amplified barcode regions. The correct size amplified barcode region for SUT098 is 173 bp and 171 bp for GCG1. So, despite the successful amplification, concentrations might seem lower than used to.

The higher concentration of GCG1 IP library compared to concentration of SUT098 IP library indicates potential issues during the extraction of either amplified GCG1 IP barcodes or amplified SUT098 barcodes from agarose gel. Visually, SUT098 IP bands appeared stronger and more distinguishable compared to GCG1 IP bands in agarose gel. One possible explanation is that GCG1 IP library contains some artifacts in addition to amplified barcodes, because the extraction of the GCG1 IP band was challenging due the artifact bands being so close to the correct size GCG1 IP band. Another explanation could be the issues during the extraction of the SUT098 IP band from agarose gel, because the concentration of SUT098 IP library is so minimal.

The bands from the negative control samples (GCG1 -RT and GCG1 -AB) were also extracted and purified for sequencing. Subsequent sequencing of the negative control samples will help ascertain whether the undesired bands are primer-dimers or PCR artifacts or, if there was a potential issue with the immunoprecipitation or the DNAse treatment was inefficient.

5 Discussion

The five-month time period was not enough to finish the method development as planned. The next step in the development of the method would have been the sequencing of GCG1 and SUT098 barcode libraries and bioinformatic analyses of the sequencing data to make a proof of concept. The aim of bioinformatic analyses would be to compare RBPs associated with coding mRNA transcripts and with lncRNA transcripts.

Since the assembled barcode libraries were not sequenced by the end of this study, the proof of concept for RNA-Decoder method is still missing. Additionally, this type of method is not previously used to study lncRNA-protein interactions, it is difficult to compare the results of this study with previous research. However, Korhout et al. (2018) and Breugel et al (2023) have used the similar type of method to detect the chromatin associated proteins. Korhout et al (2018) were able to identify even 469 chromatin binder proteins in yeast. In their research, they used two barcode regions (one at the upstream area of the gene and one at the downstream area of the gene) and were able to determine 18 proteins which were specific to bind only the upstream area of the gene and 273 the downstream area specific binder proteins. (Korthout et al. 2018.) Whereas Breugel et al. (2023) determined the different factors participating the regulation of tRNA gene expression. Their key finding was Fpt1, a newly identified protein which promotes the removal of RNA polymerase III from chromatin (Breugel et al. 2023).

Although Breugel et al. (2023) and Korhout et al. (2018) utilized this kind of method to identify chromatin binding proteins not RNA transcript binding proteins, their results refer to that with a little further optimization also RNA-Decoder method could work. In this study, the results of Western blot analysis of immunoprecipitation and PCR amplification of SUT098 barcodes look very promising, indicating the RNA-Decoder method's potential effectiveness. Still, some steps of the method as amplification of GCG1 barcode region would require more optimisation to maximize method's efficiency. Therefore, the critical reflections from cross-linking, immunoprecipitation and PCR optimisation have been gathered to this discussion to further develop this method in the future.

5.1 Reflection of cross-linking and immunoprecipitation results

The approach of this study was pragmatic, meaning that during the development of RNA-Decoder method the number of repetitions remained limited. For example, during the development process, RNA-protein cross-links were created either by UV-light or formaldehyde. The recovered RNA from UV-light cross-linked and formaldehyde cross-

linked samples were measured by qPCR but because formaldehyde cross-linking seemed more promising, immunoprecipitation and qPCR measurements for RNA recovery were repeated only for formaldehyde samples. Later, formaldehyde was selected as a cross-linker for big-scale RNA-Decoder experiment despite the high variation among formaldehyde replicates. The decision was made without repeated measurements with UV-light cross-linked samples, leaving uncertainty whether UV-light would have been more prominent cross-linker.

Another factor advocating for the use of UV-light is that, in case of formaldehyde treated samples cDNA was obtained even from Adh2 samples. This suggests that formaldehyde may have been overly effective cross-linker, potentially creating cross-links between RNA transcripts and Adh2 proteins, although Adh2 is involved in amino acid metabolism of the yeast, and it is not expected to interact with RNA transcripts.

The efficiency of the cross-linkers could be tested by repeating qPCR measurements for UV-light cross-linked samples or performing PCR reaction with Illumina primers for UV-cross-linked samples and comparing intensity of bands obtained from PCR reactions between UV-light and formaldehyde cross-linked samples.

The issue of limited repetitions also considers results from Dynabead A versus Dynabead G comparison experiment. While immunoprecipitation using Dynabead As was repeated three times, immunoprecipitation with Dynabead Gs was performed only once during the method development. Significantly higher yield of the recovered cDNA with Dynabead Gs justified their selection for big-scale experiment despite the lack of repetitions.

Overall, the success of the immunoprecipitation seemed to vary among different experiments. During the method development phase, immunoprecipitation appeared to yield a higher amount of RNPs compared to the immunoprecipitation step performed in big-scale RNA-Decoder experiment. One possible explanation for this result is, that the optimization of immunoprecipitation was carried out in different laboratory where big-scale RNA Decoder experiment was conducted, and there were some variations in equipment between the laboratories.

In this study, targeted RNPs were eluted simply by washing the Dynabeads with washing buffers and finally shaking and heating them in elution buffer. However, TAP-tags have a TEV protease cleavage site, enabling that after washing the Dynabeads it would be possible to add TEV protease to release the RNPs bound on the Dynabeads to get cleaner and better yield (Rigaut et al. 1999). This is something worth considering for further method development.

5.2 Reflection of barcode amplification results

At the first, the entirety of PCR optimisation for both SUT098 and GCG1 barcode regions proved to be a challenging task due to the need to balance multiple goals. On the other hand, the one goal was to keep the cycle number low as possible to decrease formation of PCR artifacts and minimize other errors in the barcode areas during the amplification such as nucleotide miscorporation and the amplification bias. Amplification bias means unequal amplification of barcodes during PCR leading inaccurate representation of the original barcode composition in the amplified product. However, it was also necessarily to keep the cycle number a high enough to get a distinct, visible band in the gel for the extraction for barcode sequencing.

The sequencing of the barcode libraries assembled in this study is necessary to determine whether the cycle numbers used were appropriate. Too low cycle number would result in insufficient number of barcodes for sequencing and too high cycle number would result in amplification bias.

The effectiveness of immunoprecipitation has a direct impact to its subsequent steps like barcode amplification. The higher yields of RNP complexes obtained from immunoprecipitation directly improve the success of barcode amplification during PCR. Improving the conditions of immunoprecipitation reaction would enable the use of reduced number of PCR cycles during barcode amplification, preventing amplification bias, primer-dimer formation, and nucleotide miscorporation, ensuring more accurate results. Therefore, focusing on to improve the PCR reaction for SUT098 barcodes is related to improvising the conditions and yield of immunoprecipitation.

The major challenge on the PCR optimization was the consistent formation of GCG1 primer-dimers across almost all PCR reactions, despite attempts to modify PCR reaction conditions. Enhancing reaction yield and eliminating the PCR artifacts would require designing for the new primers targeting on the GCG1 barcode area. However, the designing of the new primers for GCG1 area would also be challenging. The genomes of Epi-Decoder yeast strains differ only a slightly from the genome of wild type *S. cerevisiae* yeast, complicating the design new primers that selectively target Epi-Decoder yeast genomes without binding to the wild type yeast genome used as a negative control.

The distinguish between primer-dimer bands and the correct size PCR product was more distinct on the native-PAGE gel than on the agarose gel. Therefore, in the future, a potential

solution could be the extraction of the amplified barcodes from native-PAGE gels instead of agarose gels.

6 Conclusion

The results obtained from immunoprecipitation and barcode amplification steps from RNA-Decoder method were promising, suggesting that the method could serve as a valuable tool to identify novel RNP complexes in the future. The next step in the method development process would be the sequencing of the assembled barcode libraries to establish a proof of concept, evaluate the efficiency of RNA-Decoder method and pinpoint the parts of the method that would require further optimization. It is already evident, that at least PCR reaction for GCG1 barcode region needs either the design of new set of primers or more stringent optimization of reaction conditions to ensure the method's efficiency.

RNA-Decoder method has two significant limitations. The first limitation is the requirement of integrated DNA barcodes at the locus of the interest. The size of the barcodes is only a few tens of nucleotides, so usually the barcodes do not disturb a transcription of the locus but this needs to be verified for every locus of interest. The DNA barcodes are generally easy to integrate by using CRISPR/Cas9 technique. Due the requirement of the integrated DNA barcodes, RNA-Decoder method is not suitable for a general discovery tool for all potential, yet undiscovered lncRNA-protein complexes in the cell. Other methods such as mass spectrophotometry may provide better applicability for such studies in the future. The second limitation of RNA-Decoder method is the requirement of tagged proteins library. It may limit the utility of the method in cases where such libraries are not available. However, tagged protein libraries of other organisms are already used in research (Cho et al. 2022; Dewari et al. 2018; Schwinn et al. 2020). The accessibility of tagged protein libraries will probably increase in the future with the improvement of techniques like CRISPR-Cas9.

The strength of RNA-Decoder method is its possibility compare RNP complexes formed from different loci of interest to each other or analyse formed RNP complexes from one locus of interest in different conditions.

In the future, after the optimization and proof of concept, RNA-Decoder method could be utilized to study how protein composition bound to specific RNA transcripts changes when the yeast cells are cultivated in different environments or kept in different conditions. For instance, the method could be used to identify possible lncRNA-protein complexes which are active only in specific conditions such as during the nutrient depletion. Moreover, harnessing the method to mammalian cells, plant cells or human cells could provide valuable information for the roles of lncRNA-protein complexes in various cellular processes including cell growth, differentiation, metabolism and in the onset of diseases. The wider utilization of

RNA-Decoder method could lead the to a deeper understanding of lncRNA-protein interactions and their functional significance across different organisms.

References

Ard, R., Tong, P. & Allshire, R. C. (2014) Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nat Commun* **5**:5576.

Atkinson, S. R., Marguerat, S., Bitton, D. A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., … Bähler, J. (2018) Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* **24**:1195–1213.

Breugel, M. E. van, Kruijsbergen, I. van, Mittal, C., Lieftink, C., Brouwer, I., Brand, T. van den, … Leeuwen, F. van (2023) Locus-specific proteome decoding reveals Fpt1 as a chromatin-associated negative regulator of RNA polymerase III assembly. *Molecular Cell* **83**:4205-4221.e9.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. (2007) Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in *S. cerevisiae*. *Cell* **131**:706–717.

Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F. & Zenklusen, D. (2013) Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* **20**:851–858.

Cemel, I. A., Diernfellner, A. C. R. & Brunner, M. (2023) Antisense Transcription of the Neurospora Frequency Gene Is Rhythmically Regulated by CSP-1 Repressor but Dispensable for Clock Function. *J Biol Rhythms* **38**:259–268.

Chatterjee, D., Sanchez, A. M., Goldgur, Y., Shuman, S. & Schwer, B. (2016) Transcription of lncRNA prt, clustered prt RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of pho1 gene expression under phosphate-replete conditions in fission yeast. *RNA* **22**:1011–1025.

Chen, H.-M. & Neiman, A. M. (2011) A conserved regulatory role for antisense RNA in meiotic gene expression in yeast. *Current Opinion in Microbiology* **14**:655–659.

Chua, P. R. & Roeder, G. S. (1998) Zip2, a Meiosis-Specific Protein Required for the Initiation of Chromosome Synapsis. *Cell* **93**:349–359.

Cho, N. H., Cheveralls, K. C., Brunner, A.-D., Kim, K., Michaelis, A. C., Raghavan, P., … Leonetti, M. D. (2022) OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**:eabi6983.

Cusanelli, E. & Chartrand, P. (2015) Telomeric repeat-containing RNA TERRA: A noncoding RNA connecting telomere biology to genome integrity. *Frontiers in Genetics* 6.

Cusanelli, E., Romero, C. A. P. & Chartrand, P. (2013) Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell* **51**:780–791.

Davati, N. & Ghorbani, A. (2023) Discovery of long non-coding RNAs in Aspergillus flavus response to water activity, CO2 concentration, and temperature changes. *Sci Rep* **13**:10330.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., … Steinmetz, L. M. (2006) A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences* **103**:5320–5325.

Dewari, P. S., Southgate, B., Mccarten, K., Monogarov, G., O'Duibhir, E., Quinn, N., … Pollard, S. M. (2018) An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. *eLife* **7**:e35069.

Drinnenberg, I. A., Weinberg, D. E., Xie, K. T., Mower, J. P., Wolfe, K. H., Fink, G. R. & Bartel, D. P. (2009) RNAi in Budding Yeast. *Science* **326**:544–550.

Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., … Wahlestedt, C. (2010) Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology* **11**:R56.

Fehrmann, S., Paoletti, C., Goulev, Y., Ungureanu, A., Aguilaniu, H. & Charvin, G. (2013) Aging yeast cells undergo a sharp entry into senescence unrelated to the loss of mitochondrial membrane potential. *Cell Rep* **5**:1589–1599.

Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M. & Muxel, S. M. (2019) Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA* **5**:17.

Garg, A., Sanchez, A. M., Shuman, S. & Schwer, B. (2018) A long noncoding (lnc)RNA governs expression of the phosphate transporter Pho84 in fission yeast and has cascading effects on the flanking prt lncRNA and pho1 genes. *J Biol Chem* **293**:4456–4467.

Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., … Vershon, A. K. (2011) Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast ▿ . *Mol Cell Biol* **31**:1701–1709.

Gilbert, C. & Svejstrup, J. Q. (2006) RNA immunoprecipitation for determining RNA-protein associations in vivo. *Curr Protoc Mol Biol* Chapter 27:Unit 27.4.

Gloeckner, C. J., Boldt, K., Schumacher, A., Roepman, R. & Ueffing, M. (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *PROTEOMICS* **7**:4228–4234.

Gowthaman, U., Ivanov, M., Schwarz, I., Patel, H. P., Müller, N. A., García-Pichardo, D., … Marquardt, S. (2021) The Hda1 histone deacetylase limits divergent non-coding transcription and restricts transcription initiation frequency. *The EMBO Journal* **40**:e108903.

Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M. & Martens, J. A. (2011) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* **25**:29–40.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C. S., Shibata, T. & Ohta, K. (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**:130–134.

Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. (2006) Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. *Cell* **127**:735–745.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D. & Vogelauer, M. (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell* **32**:685–695.

Hovhannisyan, H. & Gabaldón, T. (2021) The long non-coding RNA landscape of Candida yeast pathogens. *Nat Commun* **12**:7317.

Huang, W., Hickson, L. J., Eirin, A., Kirkland, J. L. & Lerman, L. O. (2022) Cellular senescence: The good, the bad and the unknown. *Nat Rev Nephrol* **18**:611–627.

Huang, Y., Zhang, J. L., Yu, X. L., Xu, T. S., Wang, Z. B. & Cheng, X. C. (2013) Molecular functions of small regulatory noncoding RNA. *Biochemistry Moscow* **78**:221–230.
Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., … Chang, H. Y. (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* **43**:621–629.

Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y. & Zhuo, W. (2019) Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J Cancer Res* **9**:1354–1366.

Johnson, R., Richter, N., Jauch, R., Gaughwin, P. M., Zuccato, C., Cattaneo, E. & Stanton, L. W. (2010) Human accelerated region 1 noncoding RNA is repressed by REST in Huntington's disease. *Physiol Genomics* **41**:269–274.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., … Wahlestedt, C. (2005) Antisense Transcription in the Mammalian Transcriptome. *Science* **309**:1564–1566.

Korthout, T., Poramba-Liyanage, D. W., van Kruijsbergen, I., Verzijlbergen, K. F., van Gemert, F. P. A., van Welsem, T. & van Leeuwen, F. (2018) Decoding the chromatin proteome of a single genomic locus by DNA sequencing. *PLoS Biol* **16**:e2005542.

Krystal, G. W., Armstrong, B. C. & Battey, J. F. (1990) N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. *Mol Cell Biol* **10**:4180–4191.

Kung, J. T. Y., Colognori, D. & Lee, J. T. (2013) Long Noncoding RNAs: Past, Present, and Future. *Genetics* **193**:651–669.

Lam, J. K. W., Chow, M. Y. T., Zhang, Y. & Leung, S. W. S. (2015) siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy - Nucleic Acids* **4**:e252.

Leonardi, J., Box, J. A., Bunch, J. T. & Baumann, P. (2008) TER1, the RNA subunit of fission yeast telomerase. *Nat Struct Mol Biol* **15**:26–33.

Leong, H. S., Dawson, K., Wirth, C., Li, Y., Connolly, Y., Smith, D. L., … Miller, C. J. (2014) A global non-coding RNA system modulates fission yeast protein levels in response to stress. *Nat Commun* **5**:3947.

Li, Jing & Liu, C. (2019) Coding or Noncoding, the Converging Concepts of RNAs. *Front Genet* **10**:496.

Li, Juan, Liu, X., Yin, Z., Hu, Z. & Zhang, K.-Q. (2021) An Overview on Identification and Regulatory Mechanisms of Long Non-coding RNAs in Fungi. *Frontiers in Microbiology* 12.

Li, N., Joska, T. M., Ruesch, C. E., Coster, S. J. & Belden, W. J. (2015) The frequency natural antisense transcript first promotes, then represses, frequency gene expression via facultative heterochromatin. *Proceedings of the National Academy of Sciences* **112**:4357–4362.

Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. (2017) How do lncRNAs regulate transcription? *Science Advances* **3**:eaao2110.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z. & Lingner, J. (2008) The Rat1p 5′ to 3′ Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. *Molecular Cell* **32**:465–477.

Lybecker, M., Bilusic, I. & Raghavan, R. (2014) Pervasive transcription: Detecting functional RNAs in bacteria. *Transcription* **5**:e944039.

Ma, L., Bajic, V. B. & Zhang, Z. (2013) On the classification of long non-coding RNAs. *RNA Biol* **10**:924–933.

Magistri, M., Velmeshev, D., Makhmutova, M. & Faghihi, M. A. (2015) Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid-β Homeostasis, and Deregulated Expression of Long Noncoding RNAs. *J Alzheimers Dis* **48**:647–665.

Majumder, M. & Palanisamy, V. (2021) Compendium of Methods to Uncover RNA-Protein Interactions In Vivo. *MPs* **4**:22.

Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F. & Goodrich, J. A. (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* **29**:499–509.

Martens, J. A., Wu, P.-Y. J. & Winston, F. (2005) Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes Dev* **19**:2695–2704.

Matsui, K., Nishizawa, M., Ozaki, T., Kimura, T., Hashimoto, I., Yamada, M., … Okumura, T. (2008) Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. *Hepatology* **47**:686–697.

McHugh, C. A., Russell, P. & Guttman, M. (2014) Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol* **15**:203.

Mercer, T. R. & Mattick, J. S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* **20**:300–307.

Mishra, K. & Kanduri, C. (2019) Understanding Long Noncoding RNA and Chromatin Interactions: What We Know So Far. *Noncoding RNA* **5**:54.

Moravec, M., Wischnewski, H., Bah, A., Hu, Y., Liu, N., Lafranchi, L., … Azzalin, C. M. (2016) TERRA promotes telomerase-mediated telomere elongation in Schizosaccharomyces pombe. *EMBO Rep* **17**:999–1012.

Mukherjee, K., Futcher, B. & Leatherwood, J. (2018) Mmi1 and rep2 mRNAs are novel RNA targets of the Mei2 RNA-binding protein during early meiosis in Schizosaccharomyces pombe. *Open Biology* **8**:180110.

Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L. M., de Nadal, E. & Posas, F. (2014) Control of Cdc28 CDK1 by a Stress-Induced lncRNA. *Molecular Cell* **53**:549–561.

Ng, S.-Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. (2013) The Long Noncoding RNA *RMST* Interacts with SOX2 to Regulate Neurogenesis. *Molecular Cell* **51**:349–359.

Ni, Y., Huang, H., Chen, Y., Cao, M., Zhou, H. & Zhang, Y. (2017) Investigation of Long Non-coding RNA Expression Profiles in the Substantia Nigra of Parkinson's Disease. *Cell Mol Neurobiol* **37**:329–338.

Palazzo, A. F. & Lee, E. S. (2015) Non-coding RNA: What is functional and what is junk? *Frontiers in Genetics* 6.

Parker, S., Fraczek, M. G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., … O'Keefe, R. T. (2018) Large-scale profiling of noncoding RNA function in yeast. *PLoS Genet* **14**:e1007253.

Prensner, J. R., Iyer, M. K., Balbin, O. A., Dhanasekaran, S. M., Cao, Q., Brenner, J. C., … Chinnaiyan, A. M. (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**:742–749.

Quinn, J. J. & Chang, H. Y. (2016) Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**:47–62.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. & Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**:1030–1032.

Saccharomyces Genome Database (SGD). <https://www.yeastgenome.org/>

Saliminejad, K., Khorram Khorshid, H. R., Soleymani Fard, S. & Ghaffari, S. H. (2019) An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *Journal of Cellular Physiology* **234**:5451–5465.

Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. (2011) A ceRNA hypothesis: The Rosetta stone of a hidden RNA language? *Cell* **146**:353–358.

Schwinn, M. K., Steffen, L. S., Zimmerman, K., Wood, K. V. & Machleidt, T. (2020) A Simple and Scalable Strategy for Analysis of Endogenous Protein Dynamics. *Sci Rep* **10**:8953.

Shah, J. C. & Clancy, M. J. (1992) IME4, a Gene That Mediates MAT and Nutritional Control of Meiosis in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **12**:1078–1086.

Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**:96–118.

Till, P., Derntl, C., Kiesenhofer, D. P., Mach, R. L., Yaver, D. & Mach-Aigner, A. R. (2020) Regulation of gene expression by the action of a fungal lncRNA on a transactivator. *RNA Biology* **17**:47–61.

Till, P., Mach, R. L. & Mach-Aigner, A. R. (2018) A current view on long noncoding RNAs in yeast and filamentous fungi. *Appl Microbiol Biotechnol* **102**:7319–7331.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., … Prasanth, K. V. (2010) The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol Cell* **39**:925–938.

van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., … Amon, A. (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**:1170–1181.

Wang, K. C. & Chang, H. Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**:904–914.

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.-S., Gautheret, D. & Morillon, A. (2016) Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol Cell* **61**:379–392.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., … Steinmetz, L. M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**:1033–1037.

Xue, Z., Ye, Q., Anson, S. R., Yang, J., Xiao, G., Kowbel, D., … Liu, Y. (2014) Transcriptional interference by antisense RNA is required for circadian clock function. *Nature* **514**:650–653.

Yamashita, A., Shichino, Y. & Yamamoto, M. (2016) The long non-coding RNA world in yeasts. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**:147–154.

Zeinoun, B., Teixeira, M. T. & Barascu, A. (2023) TERRA and Telomere Maintenance in the Yeast Saccharomyces cerevisiae. *Genes* **14**:618.

Ziats, M. N. & Rennert, O. M. (2013) Aberrant Expression of Long Noncoding RNAs in Autistic Brain. *J Mol Neurosci* **49**:589–593.

Appendices

Appendix 1: Illumina primers for Barcode amplification

Name of the primer	Sequence from 5' to 3'
P7_SUT098_IP	CAAGCAGAAGACGGCATACGAGATagctgtgtGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACTGCTTCGTCTTcAGAcgA
P7_GCG1_IP	CAAGCAGAAGACGGCATACGAGATgtcatgtgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCACGGTTTTGCTCGgaGa
P7_SUT098_input	CAAGCAGAAGACGGCATACGAGATacatgtcgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACTGCTTCGTCTTcAGAcgA
P7_GCG1_input	CAAGCAGAAGACGGCATACGAGATctctgtacGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCACGGTTTTGCTCGgaGa
P7_GCG1_-reverse transcriptase control	CAAGCAGAAGACGGCATACGAGATtcgtgtgaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCACGGTTTTGCTCGgaGa
P7_GCG1_-antibody control	CAAGCAGAAGACGGCATACGAGATtgactgtcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACTGCTTCGTCTTcAGAcgA
P5_POOL_A_technical replicate	AATGATACGGCGACCACCGAGATCTACACtacagcagACACTCTTTCCCTACACGACGCTCTTCCGATCT
P5_POOL_B_technical replicate	AATGATACGGCGACCACCGAGATCTACACtgagtgtcACACTCTTTCCCTACACGACGCTCTTCCGATCT
P5_POOL_C_technical replicate	AATGATACGGCGACCACCGAGATCTACACactcgaagACACTCTTTCCCTACACGACGCTCTTCCGATCT