THE UNIVERSITY
of EDINBURGH

Investigating the role of RNAi and epigenetic modifications in the basidiomycetes yeast *Cryptococcus deneoformans*

Charlotte Scoynes

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

February 2022

Contents

Table of Figures

Table of Tables

Abbreviations

5-FOA	5-Fluoroorotic Acid
5mC	5-Methyl Cytosine
BLAST	Basic Local Alignment Search Tool
ChIP	Chromatin Immunoprecipitation
ClrC	Clr4 containing Complex
CsA	Cyclosporin A
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded RNA
FTH	Flag tag, Tev cleavage site and Histone x 6 tag
HGS	Homologous Gene Silencing
LTR	Long Terminal Repeat
miRNA	Micro RNA
MSUD	Meiotic Silencing of Unpaired DNA
Mya	Million years ago
NCRIP	Non-Canonical RNAi Pathway
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
piRNA	Piwi-interacting RNA

PTGS	Post-Transcriptional Gene Silencing
PRC2	Polycomb Repressive Complex 2
PRSC	P-body-associated RNA Silencing Complex
RdRP	RNA dependent RNA Polymerase
RF	Retrotransposon Fragment
RISC	RNA Induced Silencing Complex
RITS	RNA Induced Transcriptional Silencing
RNAi	RNA interference
SCANR	Spliceosome-Coupled And Nuclear RNAi complex
siRNA	Short interfering RNA
SIS	Sex-Induced Silencing
sRNA	Short RNA
SNP	Single Nucleotide Polymorphism
ssRNA	Single stranded RNA
TE	Transposable Element
TGS	Transcriptional Gene Silencing
WT	Wild Type

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Elizabeth Bayne, for giving me the opportunity to work on this project and develop as a scientist. I would also like to thank Dr Rob Van Nues for teaching me everything about working with *Cryptococcus*, for the endless ideas and for encouraging me to critically discuss my work.

I would like to extend my thanks to the BBSRC EastBio DTP for funding this project, and to both the University of Edinburgh and the Institute of Cell Biology for providing me with this opportunity. I would also like to thank all of my colleagues for the support, both within the lab and the BTO, along with those in my EastBio cohort who brightened up every training and symposium.

I am especially grateful to my friends in Edinburgh for their continued interest in my studies, but most importantly always being there - without those many hours rowing, hiking, climbing or at band together I would not have been able get to where I am today.

Finally, I would like to thank my family for their unconditional love and support, despite not understanding what I've been working on these past few years, and as we always knew – nothing makes me more productive than the last minute.

Declaration

I declare that the work presented in this thesis was conducted solely by myself, and that any contributions by other people have been clearly indicated and referenced. This work has not been submitted for any other degree or professional qualification.

Signed,

Charlotte Scoynes

Lay Summary

Many species of fungi can infect humans and cause disease. Among these are the species which give rise to Cryptococcosis, an infection affecting the lungs, causing pneumonia, and the brain causing meningitis. The species of *Cryptococcus* determines who is susceptible to cryptococcosis, with some species only infecting immunocompromised people whilst other species can infect anyone. What differentiates the different species of *Cryptococcus* is the genetic material within each cell. In *Cryptococcus* species, the DNA is split over 14 chromosomes. Majority of the DNA within each chromosome is made up of genes which encode proteins which are essential for cellular function. Each chromosome also has one region which does not contain any genes, but is important for cell division, named the centromere.

In *Cryptococcus*, the centromeres are also home to mobile genetic elements which if left unchecked can jump in and out of random places within the DNA, causing damage. Specific parts of the DNA including mobile elements are commonly silenced by epigenetic methods, which means that the DNA itself is not changed but reversible changes are made at other levels beyond the genetic information. One common epigenetic silencing pathway involves packaging the DNA really tightly around proteins to create heterochromatin. This makes it really hard to access the genes within the tightly packaged DNA, therefore silencing them. Another mechanism employs a small piece of RNA, a form of genetic material similar in structure to DNA, which is complementary to the DNA sequence. This RNA can be used to target protein complexes to specific regions of the DNA to initiate silencing by interfering with the pathway from gene to functional protein. This mechanism is termed RNA interference (RNAi). The aim of this thesis is to see if there is a connection between these two epigenetic silencing pathways, RNAi and heterochromatin, within the pathogenic fungi *Cryptococcus deneoformans,* as the

presence of these two pathways correlates with differences in the infectivity of *Cryptococcus* species. Here I have looked to see if RNAi initiates silencing by packaging the DNA up into heterochromatin. I have also looked at whether RNAi is required to suppress the movement of mobile genetic elements, by monitoring insertion of these elements into certain genes that cause drug resistance. These mobile elements are usually found within the centromeres, the regions which don't have any genes, and which are usually tightly packaged up into heterochromatin. By looking at the movement of these mobile elements I explore the links between RNAi and heterochromatin. I also look further into the proteins required for the RNAi pathway to function, as well as the sites within the genome that the RNAi pathway is silencing.

Abstract

The basidiomycete yeasts from the *Cryptococcus neoformans/gattii* species complex are major fungal pathogens and are particularly prevalent in the developing world. The species within the complex are rapidly evolving, with several species loosing genes encoding proteins required for silencing by the RNA interference (RNAi) pathway which correlate with an increase virulence. The species C. *deneoformans,* however, has retained all five of the core RNAi components (Rdp1, Ago1, Ago2, Dcr1, Dcr2), and has been shown to have a functional RNAi pathway involved in the silencing of transposable elements (TEs). Centromeric TEs have also been shown to coincide with DNA methylation in C. *deneoformans,* and also with H3K9 methylation in neighbouring species C. *neoformans*. Here I look at the relationship between these three potential mechanisms of silencing in C. *deneoformans*, RNAi, DNA methylation and H3K9 methylation, focussing on how RNAi interacts with both methylation marks in TE regulation.

Identification of the H3K9 methyltransferase Clr4 and H3K9me2-ChIP confirmed the presence of H3K9 methylation at the centromeres in C. *deneoformans*. Analysis of strains with deletions of core RNAi components revealed wild-type levels of centromeric H3K9 methylation, confirming that RNAi is not required for maintenance of this heterochromatin mark. Analysis of transcript levels at RNAi target sites showed no difference between wild-type and RNAi deficient strains. This suggests that RNAi silences targets through a post-transcriptional gene silencing (PTGS) method out with RNA degradation.

To investigate the role of RNAi in suppressing transposon activity, mutation rate assays were carried out by screening for spontaneous 5-FOA resistance that results from

disruption of the *URA3* or *URA5* genes. Strains lacking both H3K9 methylation and DNA methylation (*clr4Δdnmt5Δ*) had the highest drug resistance rates. PCR screening determined if 5-FOA resistance was due to transposon insertion into *URA3* or *URA5*, and both T1 and T2 DNA transposon insertions were identified. The rate of inserts identified within *rdp1Δ* and *clr4Δdnmt5Δ* strains was significantly higher than in WT, showing increased transposon mobility in both strains. Analysis of DNA transposable element expression showed large variance between replicate cultures but suggested that T3 may be regulated by an RNAi-independent mechanism, unlike T1 and T2 where suppression appears dependent on Rdp1. Analysis of retrotransposon copy numbers showed no significant increase in any strains tested when compared to WT. Overall this shows a potential role for H3K9 and/or DNA methylation in controlling transposon mobility alongside RNAi.

Finally, analysis was carried out into the roles of both Argonaute proteins within C. *deneoformans,* as Ago2 is frequently lost within the species complex, and is not present within neighbouring species C. *neoformans*. Mass spectrometry of tagged proteins showed that each Ago binds to a different subset of proteins, suggesting a different role for each protein within the RNAi pathway. Deletion of Gwo1, the main Ago1 interactor, increases the interaction of Ago1 with Ago2.

The work undertaken here contributes to the further understanding of the interaction between RNAi and the DNA and H3K9 methylation silencing pathway in C. *deneoformans* and shed lights on the different roles of the two Argonaute proteins in this species.

Chapter 1 Introduction

1.1 Pathogenic fungi

During the evolution of many fungal species, pathogenicity has arisen through the need to adapt to different environments. Pathogenic fungi are distributed throughout the kingdom, suggesting that many transitions from nutritional mode to pathogenic, and the reverse, have occurred (James et al., 2006). Fungi have been shown to be pathogenic to plants, mammals, and insects, with most species only infecting across one kingdom or phylum. Most mammalian and plant fungi belong to the Ascomycota or Basidiomycota phyla (Taylor, 2015). These include the common human pathogen *Candida albicans,* the fungus responsible for Dutch Elm Disease, *Ophiostoma ulmi*, and the opportunistic human pathogen *Cryptococcus,* the focus of this thesis. Fungal pathogens are often underestimated even though fungal diseases cause over 1.7 million deaths per year, are responsible for major agricultural losses through plant and crop damage and are also threatening to make species of bats and reptiles extinct (Fausto et al., 2019; Fisher et al., 2012; Kainz et al., 2020). Treatment for fungal pathogens is currently limited and no vaccines are available for any fungal pathogen as of yet (Reedy et al., 2007). Changes to the environment, such as global warming and climate change, are also increasing the incidence of fungal pathogenicity, with pressures to adapt to the increasing environmental temperature allowing species to breach the mammalian thermal barrier (Nnadi & Carter, 2021). With the potential for pathogenic fungi to cause a severe global threat, the study of these pathogenic species is of growing importance.

1.2 *Cryptococcus* Overview

1.2.1 *Cryptococcus neoformans/gattii* species complex

In 1894/95, there were two independent reports of a *Saccharomyces*-like organism, one found acting as a pathogen in a leg wound by Busse and one discovered in the non-pathogenic environment of fermenting fruit juice by Sanfelice. This organism was renamed *Cryptococcus neoformans* in 1901 due to its inability to produce ascospores like *Saccharomyces* yeasts (Barnett, 2010), and was later classified in the Basidiomycota phylum (Kwon-Chung, 1975). However, over the past century many different names have been coined when referring to this organism, partly due to using different names for different sexual states and the many species variants identified, which are hard to distinguish between within the *Cryptococcus neoformans/gattii* species complex (Barnett, 2010). Eventually, two different species were distinguished, C. *neoformans*, including serotypes A, D and AD and C. *gattii* including serotypes B and C (Kwon Chung et al., 1978; Kwon-Chung et al., 2002). Within C. *neoformans* were the variants var. grubii and var. neoformans, initially separated by serotypes (A and D respectively) but then more accurately defined by genotypes (Kwon-Chung & Varma, 2006). The C. *neoformans/gattii* species complex was more recently split into seven species, based on the seven different genotypes present, along with four hybrid species. This splits C. *neoformans* into two distinct species: C. *neoformans* which is representative of the species variant grubii, and C. *deneoformans* which is representative of the species variant neoformans (Hagen et al., 2015). This new classification is going to be used throughout, however, the majority of the past and current studies have used the previous two species naming system, so this will be converted to the new system accordingly when referenced to enable distinction between the species used. The new species classification for the C. *neoformans/gattii* species complex can be seen in table 1.1, including the genotypes

Table 1.1 *Cryptococcus neoformans/gattii* species complex

Current species	Genotype	Previous species name	Important strains
Cryptococcus neoformans	VNI, VNII	*Cryptococcus neoformans* var. *grubbii*	H99
Cryptococcus deneoformans	VNIV	*Cryptococcus neoformans* var. *neoformans*	JEC21
Cryptococcus gattii	VGI	*Cryptococcus gattii*	WM276
Cryptococcus deuterogattii	VGII	*Cryptococcus gattii*	R265
Cryptococcus bacillisporus	VGIII	*Cryptococcus gattii*	
Cryptococcus tetragattii	VGIV	*Cryptococcus gattii*	
Cryptococcus decagattii	VGIV/VGIIIc	*Cryptococcus gattii*	
Hybrid species	`		
Cryptococcus neoformans x Cryptococcus deneoformans hybrid	VNIII	*Cryptococcus neoformans* intervariety hybrid	
Cryptococcus deneoformans x Cryptococcus gattii hybrid	-	*Cryptococcus neoformans* var. *neoformans x Cryptococcus gattii AFLP/VGI* hybrid	
Cryptococcus neoformans x Cryptococcus gattii hybrid	-	*Cryptococcus neoformans* var. *grubbii x Cryptococcus gattii* ALP4/VGI hybrid	
Cryptococcus neoformans x Cryptococcus deuterogattii hybrid	-	*Cryptococcus neoformans* var. *grubbii x Cryptococcus gattii* AFLP6/VGII hybrid	

and previous names under the two species complex. A phylogenetic tree shows the evolutionary divergence of the seven species, including the estimated time of divergence of major branches (Figure 1.1). Here, I will focus on C. *deneoformans* strain JEC21 as a model organism. This strain was generated from mating of one pathogenic and one non-pathogenic strain, NIH12 and NIH433 respectively, and then crossing their F_1 progeny, strains B-3501 and B-3502, to yield the F_2 progeny. A subsequent series of back crossing ten times of the F_2 strain JEC20, initially with another F_2 progeny and then with a daughter strain, resulted in JEC21 (Heitman et al., 1999; Kwon-Chung et al., 1992). Comparison of the karyotypes shows that JEC21 is more closely related to the environmental parental strain than the pathogenic strain (Heitman et al., 1999; Wickes et al., 1994).

1.2.2 *Cryptococcus* Lifecycle

Cryptococcus species are environmental yeasts found all over the world, with C. *deneoformans* and C. *neoformans* species favouring cooler climates, widespread across Europe and Northern America, and C. *gattii* and related species favouring more tropical climates. They also differ in their preferential hosts, with both C. *deneoformans* and C. *neoformans* being found in pigeon guano and soil (Bennett et al., 1977), whereas C. *gattii* is associated with eucalyptus trees (Ellis & Pfeiffer, 1990). One thing in common between all species, however, is the abundance of the alpha mating type within the natural population, estimated to be >95% for all species, though this is lower in some micro populations (Halliday et al., 1999; Kwon-Chung & Bennett, 1978; Yah et al., 2002; Zhao et al., 2019).

Sexual mating within the C. *neoformans/gattii* species complex is bipolar, where there is one mating-type locus (*MAT*) within the genome indicating the mating type – either **a**

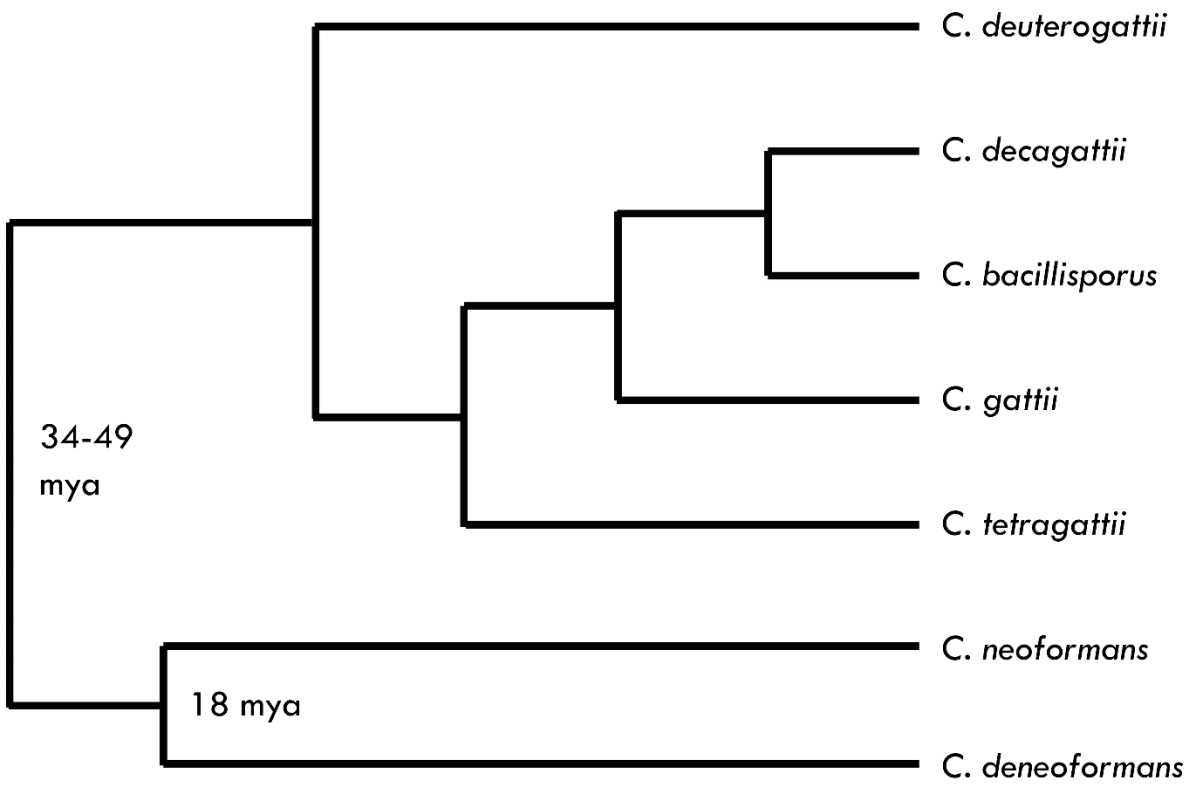

Figure 1.1 Phylogenetic tree of the *Cryptococcus neoformans/gattii* species complex
Phylogenetic tree showing the evolutionary relationship between the seven species of the *Cryptococcus neoformans/gattii* species complex. The estimated time of divergence is shown as million years ago (mya).

or α. This differs from many other basidiomycetes which have evolved to become tetrapolar, a mating type system unique to this phylum, where there are two unlinked *MAT* loci on different chromosomes which during mating segregate independently producing four different progenies, increasing the chance of outcrossing (Heitman, 2015). In *Cryptococcus,* fusion of a *MAT*α cell with a *MAT***a** cell forms a dikaryon with clamp-like connections between the dikaryotic hyphal cells. Nuclear fusion occurs in the basidium, followed by meiosis producing four initial spores. Meiotic progeny are mitotically replicated and form spores which extend from the basidium in four long basidiospore chains with random distribution, before germinating to produce haploid cells once more (Figure 1.2A) (Kwon-Chung, 1975, 1976, 1980). These species can also replicate asexually (Lin et al., 2005; Wickes et al., 1996), potentially providing an evolutionary survival advantage due to the lack of *MAT***a** cells in the natural environment or due to a gradual evolution of the species into becoming asexual (Hull & Heitman, 2002). Asexual mating is carried out through monokaryotic fruiting of the MATα cells, where two haploid cells undergo diploidisation by fusion of the two nuclei, and then meiosis creating basidiospores with genetic variation similar to sexual reproduction. However, a major difference in this pathway compared with sexual mating is the presence of monokaryotic hyphae without clamp connections fusing, likely because the clamp connections ensure fusion of two different hyphal cells and in asexual mating all cells are of the same mating type. (Figure 1.2B) (Fu et al., 2015; Wickes et al., 1996).

1.2.3 Pathogenicity

The *Cryptococcus* genus is one of the main human fungal pathogens alongside *Candida* and *Aspergillus*. Although *Cryptococcus* is primarily an environmental yeast, several species including C. *neoformans* and C. *gattii* can cause disease and are often referred

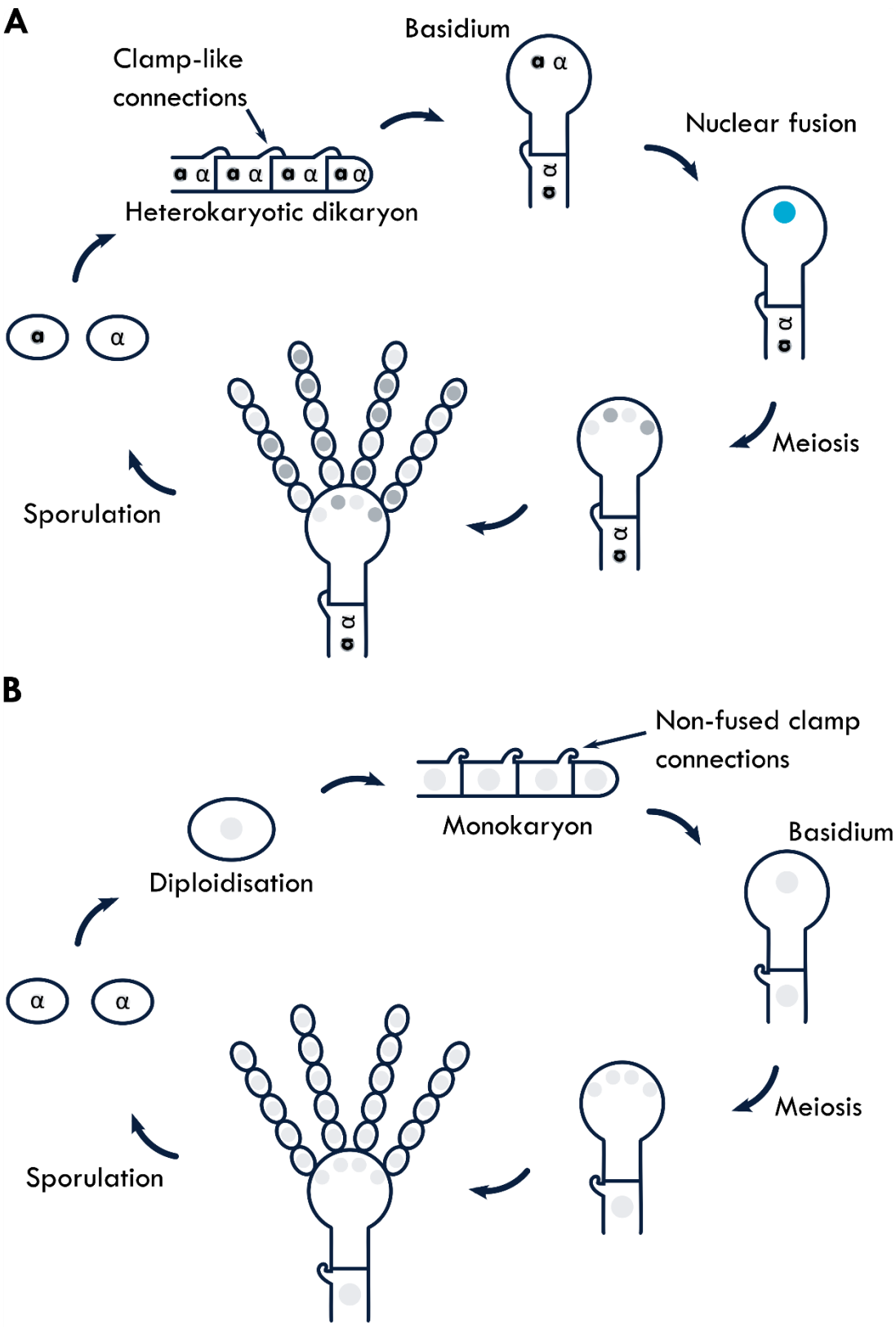

Figure 1.2 Sexual cycles of *Cryptococcus*

(A) Diagram showing the sexual mating life cycle when a *MAT***a** and a ***MAT***α cell fuse. **(B)** Diagram showing the asexual mating life cycle when two ***MAT***α cells fuse through the process of monokaryotic fruiting.

to as having accidental pathogenicity as infection of a host is not required for completion of their life cycle (Casadevall & Pirofski, 2007). Out of the pathogenic *Cryptococcus* species, C. *neoformans* is the most prevalent, infecting immunocompromised individuals and being a major cause of death within Sub-Saharan Africa, Asia, and South America (Watkins et al., 2017). Susceptible patients include those with HIV/AIDS and organ transplant recipients (Singh et al., 2008). C. *deneoformans* is also pathogenic, though it is much less prevalent, with its highest rates of infection found within Europe (Cogliati et al., 2016). Contrary to these, C. *gattii* infects immunocompetent individuals, though at a much lower rate than other *Cryptococcus* species, however it has been responsible for outbreaks of cryptococcosis over the last couple of decades in Vancouver and the Pacific Northwest region of the US (Byrnes et al., 2009; Hoang et al., 2004). These outbreaks also saw the identification of a new variant of the C. *deuterogattii* species, R265, which is thought to have significant genomic differences aiding the increased pathogenicity of this strain (Blake Billmyre et al., 2014).

Infection with *Cryptococcus* occurs through inhalation of spores into the lungs where, in patients with compromised immune systems, they evade recognition and germinate to establish infection (Maziarz & Perfect, 2016). This requires the fungi to be able to adapt to tolerate, survive and thrive in a new environment with new stresses including higher temperatures, different nutrient sources, different pH, and oxidative stress. However, the species does have two major virulence factors which aid this survival, including its polysaccharide capsule and pigment melanin (Lin & Heitman, 2006). Mutations in either of these pathways results in decreased virulence. The polysaccharide capsule, mostly composed of glucuronoxylomannan and glucuronoxylomannogalactan polysaccharides, protects the yeast against desiccation and phagocytic predators in

the environment but within a host it aids virulence through several methods (Casadevall et al., 2019). The capsule protects the yeast, preventing phagocytosis and offers protection against other stresses such as dehydration and free radicals. The capsule polysaccharides are also secreted and have been shown to alter the host immune response resulting in immune unresponsiveness (Decote-Ricardo et al., 2019). The pigment melanin also helps protect the yeast against stress factors including free radicals and heat, as well as decreasing the efficiency of antifungal drugs (Lin & Heitman, 2006; Y. Wang et al., 1995). The formation of titan cells within the lungs, cells up to 100 microns in diameter compared with the average cell at 5-7 microns in diameter, also helps with evasion of the immune host due to their large size preventing phagocytosis (Okagaki et al., 2010; Zaragoza et al., 2010). Once infections are established, *Cryptococcus* is then able to disseminate from the lungs to other tissues, primarily the brain resulting in cryptococcal meningitis and without treatment, death.

1.3 *C. deneoformans* genome

1.3.1 Genome structure and evolution

The genomes of two different strains of C. *deneoformans,* JEC21 and B-3501A, were sequenced in 2005 to aid with understanding the genomic basis for its pathogenic behaviour (Loftus et al., 2005). The ~20 Mb genome of C. *deneoformans* spans 14 chromosomes and is predicted to encode ~6500 proteins. The genome is unusually intron-rich compared with other fungi, with an average of 6.3 exons and 5.3 introns per gene, and a majority of protein coding genes having at least one intron. This differs hugely from the genomes of other model fungi, with *Saccharomyces cerevisiae* having only 5% of genes containing an intron, and nearly all of these with just one intron (Spingola et al., 1999), and *Schizosaccharomyces pombe* having 40% of genes containing introns, with the majority of these having just 1 or 2 introns (Wood et al.,

2002). C. *deneoformans* has a functional alternative splicing pathway allowing for the removal of the introns, with the species favouring introns of around 52 bp, with an insertion or deletion bias against introns which are shorter or longer than this optimal size (Hughes et al., 2008).

Unlike the genome of *S. cerevisiae*, there is no evidence of a whole-genome duplication event in C. *deneoformans* (Kellis et al., 2004), however, in JEC21 a ~60 kb segment has been duplicated when compared against the genome of the F_1 parental strain B-3501A. This is the result of an unstable telomere fusion between two chromosomes which when broken produced a translocation duplicating 22 genes from chromosome 12 onto the end of chromosome 8 (Fraser et al., 2005; Loftus et al., 2005). The genome of the JEC21 strain also contains an Identity Island, a ~40 kb region containing 14 genes on chromosome 5 which is nearly identical to the C. *neoformans* H99 genome, in contrast to the 85-90% sequence identity of the rest of the genome. This is likely to have occurred through a genetic exchange event within a hybrid species of C. *neoformans* and C. *deneoformans* allowing for C. *neoformans* to retain the original copy of these genes whilst C. *deneoformans* also received this region of the C. *neoformans* genome, in addition to its own homologous region. However, parts of the Identity Island were evolutionarily maintained in the majority of C. *deneoformans* strains tested, both clinical and environmental, with 5 out of the 14 genes conserved in all strains containing the Island, whilst the original homologs of these 14 genes within the C. *deneoformans* genome were lost after the transfer of the Identity Island, suggesting that the 5 conserved genes are either essential or provide a selective advantage, and that the C. *neoformans* homologs are preferred. A smaller ~8 kb fragment containing 4 genes on chromosome 13 is thought to have also been exchanged from C. *neoformans* to C. *deneoformans* through a similar older exchange event. Within the JEC21 strain there is

also a partial copy of the Identity Island at the other end of chromosome 5, likely occurring through a telomere fusion event similar to the larger ~60 kb fragment duplication, with duplication of 4 of the 14 genes from the Identity Island (Kavanaugh et al., 2006).

The genomes of C. *deneoformans* JEC21 and C. *neoformans* H99 are mostly syntenic, however, detailed analysis comparing the two genomes of JEC21 and H99 showed four other large translocations between the species as well as the previously mentioned translocation resulting in the chromosome arm duplication of chromosome 12 onto chromosome 8 (Figure 1.3) (Sun & Xu, 2009). These four other translocations have caused a large rearrangement of chromosomes 3 and 11. Comparison of synteny also shows nine simple inversions and 18 complex rearrangements which include areas featuring both small inversion and translocations. Unlike the majority of the simple rearrangements, these complex rearrangements are mostly (16 out of 18) found within the centromeres or *MAT* locus, with both of these locations known to contain chromosomal rearrangements and size variations between the species (Yadav et al., 2018), and are rich in transposable elements which may aid these rearrangements (Loftus et al., 2005). When these two genomes were compared against C. *deuterogattii* R265 and C. *gattii* WM276 to discern the ancestral genome arrangement, it was found that some of the inversion events such as the centromeric chromosome 1 inversion, alongside the large chromosome 3 and 11 translocation, are unique to C. *neoformans* (Janbon et al., 2014; Sun & Xu, 2009). This highlights that although there is genomic diversity between C. *deneoformans* and C. *neoformans*, C. *deneoformans* is more representative of the ancestral species and only events such as the formation of the Identity Islands and chromosomal duplication between regions of chromosomes 8 and 12 are unique to the species. With the rapid genome evolution that has taken place, it is likely that other

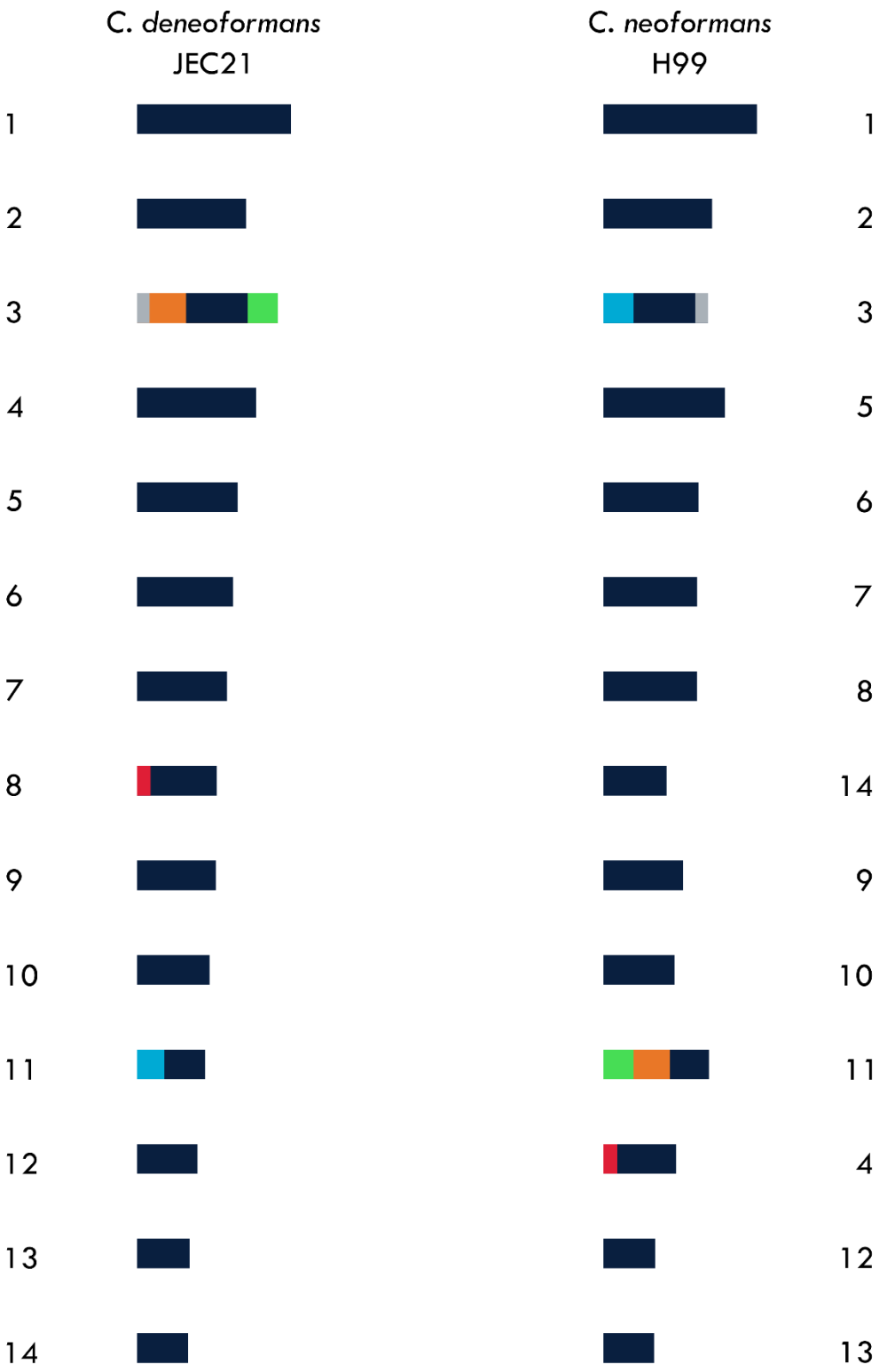

Figure 1.3 Large translocations have occurred between the genomes of *C. deneoformans* and *C. neoformans*

Comparison of the 14 chromosomes of *C. deneoformans* strain JEC21 with neighouring species *C. neoformans* strain H99. Homologous regions between chromosomes are shown in navy. Large translocations between the species are indicated by coloured sections, showing the positions of the homologous region within each strain.

genes as well as the ones mentioned here may have also been duplicated, resulting in paralogous copies within C. *neoformans* and C. *deneoformans* that are not present in closely related species, as seen when comparing the number of protein coding genes in conserved gene clusters between C. *deneoformans* JEC21, C. *neoformans* H99 and C. *gattii* WM276 (Janbon et al., 2014). This can also be seen when comparing the overall genome sizes, as the genome of C. *deuterogattii* is 1.3 Mb smaller than the genome of C. *neoformans,* with two-thirds of this difference occurring within the protein coding regions of the genome and one-third due to differences in centromere size (Yadav et al., 2018).

1.3.2 Centromeres and Telomeres

The locations of the centromeres in C. *deneoformans* were confirmed through the presence of CENP-A, a centromeric variant of histone H3, and CENP-C, a protein involved in chromosome segregation and kinetochore formation, both of which are known to localise to centromeres (Yadav et al., 2018). This identified one locus on each of the 14 chromosomes which corresponded with an open reading frame (ORF)-free and poorly transcribed region and a cluster of transposons (Loftus et al., 2005). The centromeres identified all vary in the size of the CENP-A associated region, unlike many fungi which have consistently sized CENP-A regions across all chromosomes (Roy & Sanyal, 2011), suggesting varying sizes of centromeres across the genome which are sequence independent. However, the centromeres in C. *deneoformans* are all on average larger (62 kb) than those in the closely related species C. *neoformans* and C. *deuterogattii* with their average centromeric length at 44 kb and 14 kb respectively. This difference in centromere size, and importantly genetic variation amongst closely related species, has been referred to as a centromere paradox, where the centromeres rapidly evolve and diverge at a greater speed than the remainder of the genome

(Henikoff et al., 2001). In the *Cryptococcus* genus, the centromeres have undergone expansion and contraction due to transposon mobility. The centromeres in both C. *deneoformans* and C. *neoformans* are rich in fully functional transposons which are usually silenced, however, the centromere of C. *deuterogattii* only contains partial transposons which are unable to function and therefore do not require silencing (Loftus et al., 2005; Yadav et al., 2018). It is thought that at some point in the past, the transposons in C. *deuterogattii* have mobilised and begun inserting into random sequences. Here the centromere would often be favoured, due to the lack of essential genes in this area, meaning that over time the centromeres would grow with transposons. However, as transposons get closer to each other they can excise parts of neighbouring transposons leaving unfunctional and incomplete transposons behind, causing the shrinking of its centromeres again (Yadav et al., 2018). As the centromeres in C. *deneoformans* are large and contain fully functional transposons, a lack of transposon silencing could trigger further expansion and then shrinking of its centromeres.

The gene density throughout the C. *neoformans* and C. *deneoformans* genomes remains consistent with the exception of the centromeres, indicating that the subtelomeric regions are not observable by a poor gene density such as seen in *S. cerevisiae* (Brown et al., 2010; Winzeler et al., 2003). Instead, the subtelomeric regions in C. *neoformans* and C. *deneoformans* were identified through the comparison of the synteny between the two strains. 2-36 kb regions at the end of each chromosome were shown to be less syntenic than the remainder of the genome and unusually enriched in hexose transporters. This signals the subtelomeric region due to the known potential for telomeric regions to undergo rapid evolution and accumulate genes involved in niche adaptation, like the hexose transporter which may contribute to virulence in the formation of the C. *neoformans* protective sugar capsule (Chow et al., 2012). The

telomere itself is formed of repeated sequences, which in C. *neoformans* are tandem repeats of the octanucleotide AGGGGGTT (Edman, 1992). This telomeric sequence is also seen embedded in between coding regions on the chromosomes which underwent the previously mentioned telomere fusion event in C. *deneoformans* JEC21, as well as being added to the ends of the newly made chromosomes 8 and 12 (Fraser et al., 2005). Telomeric repeats have been shown to increase linearised plasmid transformation efficiency when added to the ends of the introduced DNA. De *novo* telomeres also have been shown to form at the sites of induced double-strand breaks in the centromeres, allowing smaller and a greater number of chromosomes to form, with a similar natural formation of extra chromosomes occurring in hybrid species where there is less genomic stability (Brown et al., 2010; Priest, Coelho, et al., 2021; Yadav et al., 2020).

1.3.3 Transposons

Transposons are DNA sequences that have the ability to move around the genome. Originally discovered in Maize, transposable elements (TEs) are present in almost all life forms (McClintock, 1950, 1953). Sequence analysis of C. *deneoformans* identified that almost 5% of the genome is made up of TEs. These are mostly found within the centromere on each chromosome, but can also be seen within the telomeres, near to rDNA repeats and at the mating type *MAT* locus (Loftus et al., 2005). C. *deneoformans* contains several distinct families of TE representing both DNA transposons and retrotransposons.

Retrotransposon is a term first coined after the discovery that *S. cerevisiae* Ty elements transpose through an RNA intermediate in a 'copy-and-paste' mechanism, requiring a reverse transcriptase in a similar way that retroviruses require one for replication

(Boeke et al., 1985). The two main types of retrotransposons are LTR retrotransposons, characterised by long terminal repeats (LTRs) flanking the retrotransposon, and non-LTR retrotransposons (Finnegan, 1989). LTR retrotransposons encode two main elements, a *GAG* gene encoding a nucleocapsid protein, and the *POL* gene encoding a protease, a reverse transcriptase, RNaseH and an integrase (Wicker et al., 2007). In *C. deneoformans* the LTR retrotransposons can be split into three groups – Ty3/*gypsy* superfamily elements, Ty1/*copia* superfamily elements and solo LTRs which haven't been able to be attributed to either superfamily. Ty3/*gypsy* and Ty1/*copia* differ through the coding order of the four *POL* gene proteins (Figure 1.4A). 10 retrotransposon families were identified within *C. deneoformans* JEC21 with flanking LTRs (tcn1-10), although only tcn1-6 are present as complete, full-length retrotransposons. Of these, tcn-6 is the only Ty1/*copia* element, with the others belonging to the Ty3/*gypsy* superfamily. Further phylogenetic analysis has shown that Tcn1 is of the Tf1/sushi subgroup, a type of chromovirus retrotransposon which contains a predicted chromodomain at the C-terminal of the integrase protein (Goodwin & Poulter, 2001; Marín & Lloréns, 2000). These six tcn1-6 retrotransposons are unique to the pathogenic *Cryptococcus* species, as they are absent in the closely related non-pathogenic species *C. amylolentus* (Yadav et al., 2018). Another 10 retrotransposon fragment (RF) families, RF1-10, were also identified in the *C. deneoformans* genome, where internal parts of a retrotransposon from either Ty3/*gypsy* or Ty1/*copia* superfamilies are present without a corresponding LTR. A further 5 families of solo LTRs, LTR11-15, are found where no internal retrotransposon region has been identified (Goodwin & Poulter, 2001). These elements are all found within the centromeres, with all centromeres containing at least one copy of Tcn5 and Tcn6 (Loftus et al., 2005).

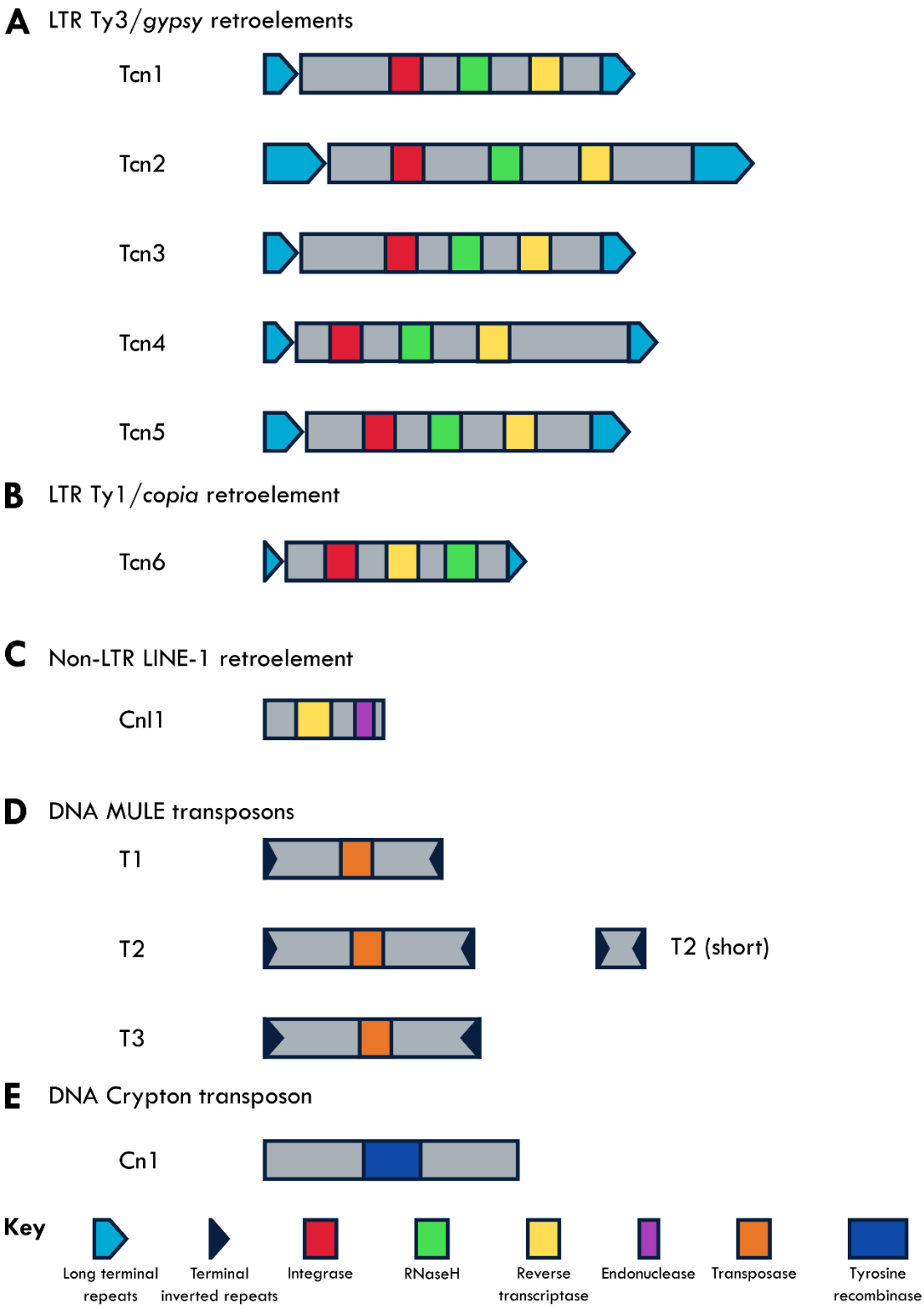

Figure 1.4 Transposon families within C. *deneoformans*
Schematics showing a selection of transposable elements within C. *deneoformans*: **(A)** LTR Ty3/*gypsy* superfamily retroelements, **(B)** LTR Ty1/*copia* superfamily retroelement, **(C)** Non-LTR LINE-1 superfamily retroelement, **(D)** DNA MULE superfamily transposons and **(E)** DNA Crypton superfamily transposon. Coloured blocks represent key features and genes which distinguish different TE superfamilies.

The most abundant non-LTR retrotransposon is Cnl1, which is unique in its location being primarily found in the subtelomeres (Figure 1.4C). The only copies outside of the telomeres are partial copies, thought to be remnants from the chromosome end fusion event and where the gene translocation event occurred in the formation of the Identity Island (Fraser et al., 2005; Kavanaugh et al., 2006).

DNA transposons are the other major class of transposons, replicating through a 'cut-and-paste' mechanism, and split into two subclasses depending on the number of DNA strands cut during transposition, with most encoding just a transposase (Finnegan, 1989; Wicker et al., 2007). Families of DNA transposons present in C. *deneoformans* include *Crypton, EnSpm, Harbinger, Mariner/Tc1* and MULE which are all subclass I, with C. *deneoformans* lacking any of the full-length subclass II *Helitron* transposons which are present in C. *neoformans* (Castanera et al., 2016). MULE TEs include T1, T2 and T3 which are present in complete and partial versions, primarily found throughout the coding regions of the genome (Figure 1.4D). T2 exists in both a long and short form, with only the long version encoding a putative gene, but with both containing an identical 5' end. Neither T2 or T3 transposons are found within the C. *neoformans* genome suggesting that they have been lost during the divergence of C. *neoformans* and C. *deneoformans* (Janbon et al., 2010). *Crypton* transposons encode a tyrosine recombinase instead of the typical DDE-type transposase of the other subclass I DNA transposons (Figure 1.4E). This element is unique to pathogenic fungi, named after the identification in C. *deneoformans,* with *Cn1-4* being identified within the genome (Goodwin et al., 2003).

1.4 Epigenetics

During the very early studies of genetics, it was understood that any phenotypic change was the result of a genotypic change. The idea that there is a whole other level of regulation between these two, the epigenotype, was first proposed by Waddington in 1942. More commonly known as epigenetics, the name comes from the Greek etymology of epi meaning "over, outside or around", referring to changes to the genetic material that are irrespective of the DNA sequence. The field of epigenetics initially was linked to developmental biology, with the epigenetic landscape serving as a metaphor for changes that are occurring to the genome of cells during differentiation (Waddington, 2012). However, with further understanding of genetics at a molecular level the field of epigenetics has grown, and the term is now used to describe a phenomenon where one genotype can produce two or more phenotypes. The nature of whether this should be an inheritable trait or include transient modifications that are long-term is still up for debate (Bird, 2007). The role of the environment in epigenetic traits is still being understood (Feil & Fraga, 2012).

The most iconic epigenetic modifications are those referred to as covalent modifications (Tollefsbol, 2017). These include DNA methylation, typically cytosine methylation, and histone modifications, such as methylation, acetylation, and ubiquitination at specific amino acid residues. These modifications remodel the chromatin, creating areas of euchromatin and heterochromatin which respectively are lightly packed regions which are accessible for transcription or tightly packed regions making transcription harder or impossible. Other epigenetic regulation mechanisms include non-coding RNAs such as *Xist* involved in X-chromosome inactivation, misfolded proteins associated with prion disease as observed in *S. cerevisiae* and miRNA-mediated gene regulation (Agrawal et al., 2003; Brockdorff et al., 1991; Liebman & Chernoff, 2012). Examples of

epigenetics range from harmless characteristics such as the tortoiseshell colouring on cats to detrimental changes responsible for certain cancer formations and mental illnesses (Lyon, 1961; Nestler et al., 2016; Sharma et al., 2009).

1.5 Covalent chromatin modifications in *Cryptococcus*

DNA is packaged into the nucleus with the help of histones, forming a dynamic chromatin structure. The DNA winds itself around these octamer proteins, creating the 'beads on a string' structure (Felsenfeld & Groudine, 2003). The formation of chromatin allows DNA to be highly compacted, allowing it to fit into the nucleus, but it also regulates the accessibility of DNA in processes such as replication and transcription. Depending on how condensed the chromatin is, will depend on whether it is referred to as euchromatin (less condensed) or heterochromatin (more condensed). Chromatin writers are enzymes which are able to modify the histones within chromatin through post-translational modifications, which in turn can alter the state of the chromatin. This is typically through altering specific amino acids in the N-terminal tail of histones by covalently adding groups such acetyl or methyl groups. These modified histones are then recognised by chromatin readers, protein factors which have a role controlling DNA transcription (Gillette & Hill, 2015). Different chromatin writers catalyse the addition of different modifications, with specific modifications causing characteristic changes to the chromatin, such as the formation of heterochromatin or euchromatin.

1.5.1 Histone methylation

Both H3K27me3 and H3K9me3 repressive histone methylation marks have been identified within C. *neoformans* H99 (Dumesic et al., 2015). H3K27me3 is deposited by the widely conserved Polycomb Repressive Complex 2, PRC2, with the subunit Ezh2 having the methyltransferase activity. Within C. *neoformans*, ChIP-seq data has showed

that the H3K27me3 mark is located in subtelomeric regions with an average domain size of 41kb. This differs in localisation to the repressive H3K9me3 mark, which is primarily located in the centromeres, but is also found in small (13kb average) domains at subtelomeres (Dumesic et al., 2015). The laying of the H3K9me3 mark is reliant on the histone methyltransferase Clr4, a variant of the SUV39H/Su(var)3-9 methyltransferases in *Drosophila Melanogaster* (Ivanova et al., 1998; Tschiersch et al., 1994). The distributions of the two repressive marks have been shown to be linked: loss of the PRC2 subunit Ccc1, a chromodomain containing protein, causes a reduction in the size of the subtelomeric H3K27me3 domain to an average 14kb, as well as the accumulation of centromeric H3K27me3 (Dumesic et al., 2015). This altered distribution is similar to the normal H3K9me3 pattern seen, and deletion of both Clr4 and Ccc1 causes the ectopic redistribution of H3k27me3 to be lost, suggesting that the redistribution of the H3K27me3 is dependent on H3K9me3. This link in location between the two marks has been put down to the nonspecific histone tail methyl-lysine residue binding pocket in the PRC2 subunit Eed1, which in the absence of Ccc1 for specific targeting, can bind the H3K9me3 instead of the H3k27me3 mark and change the distribution of the mark through this nonspecific maintenance(Dumesic et al., 2015).

Heterochromatin serves a structural function at the telomeres and centromeres which mainly consist of inactive, non-coding DNA, although are known sites for transposon localisation. However, heterochromatin is also used to repress genes. Removing these repressive marks allows the genes within the previously heterochromatic regions to be expressed, with deletion of Ezh2 and loss of H3K27me3 causing many transcripts to increase at least 3-fold, and deletion of Clr4 and loss of H3K9me3 causing many centromeric transcripts to increase approximately 6-fold (Dumesic et al., 2015).

1.5.2 DNA methylation

Although some fungi lack DNA methylation, including the popular model organisms *S. cerevisiae* and *S. pombe*, the presence of DNA methylation is seen throughout many Basidiomycetes including the majority of the *Cryptococcus* species. DNA methylation is mediated by Dnmt family proteins, and there are six subfamilies of Dnmt proteins currently identified within eukaryotes with majority of species encoding at least two Dnmt homologs (Ponger & Li, 2005). In mammals and plants, DNA methylation is carried out *de novo* by Dnmt3 and maintained by Dnmt1, or by homologs of either proteins (Law & Jacobsen, 2010). However, *C. neoformans* and *C. deneoformans* encodes only a sole DNA methyltransferase, Dnmt5, which is also present in several *Aspergillus* species. Dnmt5 has an N-terminal chromodomain followed by a RING finger domain and a C-terminal SNF2-type ATPase related domain. C5 cytosine methylation (5mC) is found primarily at the centromeres (Yadav et al., 2018), with high levels of CG methylation coinciding with TEs within the centromeres. Under wild-type (WT) conditions, little to no RNA is produced at these regions due to the presence of silencing marks (Huff & Zilberman, 2014).

With Dnmt5 being the only DNA methyltransferase present in the *C. neoformans* and *C. deneoformans* genomes, the presence of CG methylation is dependent on it (Catania et al., 2020; Huff & Zilberman, 2014). However, once this methylation is lost, these species have no ability to re-establish this, making Dnmt5 a maintenance enzyme unable to carry out *de novo* methylation. The original establishment of this 5mC has been suggested to be due to DnmtX, a predicted DNA methyltransferase present in related species, such as in several *Kwoniella* species (Catania et al., 2020). DnmtX has been shown to have *de novo* methylation properties when introduced into *C. neoformans*, suggesting that historically this protein would have been present in the *C. neoformans*

genome, however, would have been lost 150-50 mya when this species last shared a common ancestor with *Kwoniella*. This shows that the maintenance of 5mC has survived for a very long time, which is surprising when passaging experiments to study the 5mC maintenance showed that over 120 generations there were 20 times more loss events than gaining events, resulting in only 99% of 5mC sites being maintained over the 120 generations (Catania et al., 2020). This makes the equilibrium of gaining and loosing 5mC very unbalanced and suggests that methylation could only be retained for ~130 years – nowhere near as long as it has been predicted to be maintained. Therefore, DNA methylation must be maintained by selection for it to have remained in C. *neoformans* for so long.

1.5.3 Relationship between H3K9 methylation and DNA methylation

Histone H3K9 methylation and DNA methylation are closely linked in species where both are present, with the histone methylation acting as a temporary silencing method before the more permanent DNA methylation takes over the silencing (Cedar & Bergman, 2009). The interaction between the histone and DNA methylation, and their respective methyltransferases Clr4 and Dnmt5, is starting to be understood in *Cryptococcus*. The areas of genomic H3K9 methylation and 5mC are known to overlap almost fully, with the majority present at the centromeres but with small domains in sub-telomeric regions also present (Catania et al., 2020; Huff & Zilberman, 2014). The chromodomain of Dnmt5 has been shown to bind H3K9me, and when Clr4 is deleted, and H3K9 methylation is lost, the levels of 5mC also decrease as Dnmt5 cannot be recruited to the H3K9 domains (Catania et al., 2020). Recruitment of Dnmt5 also occurs through interactions with the HP1 ortholog Swi6, providing two parallel methods of recruitment to H3K9me regions. Uhf1 also works with Clr4 to promote 5mC, by binding hemimethylated DNA to then promote symmetrical methylation of the DNA through

recruitment of Dnmt5. Whilst Dnmt5 recruitment and WT levels of 5mC require Clr4 and H3K9 methylation, it has also been shown that deletion of Dnmt5 causes an altered distribution of H3K9 methylation when compared with WT (Catania et al., 2020). In the absence of Dnmt5, the H3K9 methylation levels increase in the subtelomeric regions and decrease at the centromeres. This suggests that the recruitment of Clr4 to the H3K9me regions and WT levels of methylation relies on either the DNA methylation, Dnmt5, or another factor which is altered when Dnmt5 is deleted. This interdependent relationship suggests a reliance on both Clr4 and Dnmt5 to be present for maintaining normal histone and DNA methylation levels.

1.6 RNA interference

1.6.1 RNAi pathway

RNA interference (RNAi) was discovered in *Caenorhabditis elegans,* where it was seen that double-stranded RNA (dsRNA) was able to manipulate gene expression (Fire et al., 1998). This phenomenon explained the mechanism behind the previously identified post-transcriptional gene silencing (PTGS) occurring in plants (Napoli et al., 1990), as well as quelling in the fungus *Neurospora crassa* (Cogoni et al., 1996; Romano & Macino, 1992). RNAi has since been identified in many eukarya, including humans, plants, and fungi (Wilson & Doudna, 2013).

The RNAi pathway comprises of an Argonaute-like protein interacting with a 20-30 nucleotide (nt) single-stranded (ss)RNA, which acts as a sequence-specific guide, guiding Argonaute and its interactors to the target site to initiate silencing (Agrawal et al., 2003). Three types of small RNA molecules can induce RNAi: short interfering (si)RNA, micro (mi)RNA and piwi-interacting (pi)RNA, with the pathway varying slightly in the protein interactors and mechanism depending on the type of RNA used.

Both siRNA and miRNA are produced from dsRNA precursors (Martienssen & Moazed, 2015). In the case of siRNAs, these can be exogenously produced, such as from viral genomes, or endogenously produced through hairpin formations of RNA molecules. miRNAs are also formed through hairpin structures, although these differ from siRNA hairpins due to the presence of mismatches within the hairpin while siRNA hairpins form perfect base pairing (Wilson & Doudna, 2013). In plants, worms, and fungi, endogenously produced dsRNA can rely on the presence of an RNA-dependent RNA Polymerase (RdRP) (Czech & Hannon, 2011). The dsRNA molecules are then processed and cleaved by Dicer enzymes to produce the 21-24nt dsRNA duplexes. Although both siRNA and miRNA are processed differently, the pathways converge when the dsRNA duplexes bind Argonaute proteins (Martienssen & Moazed, 2015; Wilson & Doudna, 2013). Here the RNA duplexes unwind forming a ssRNA guide, which together with Argonaute and other protein interactors form the RNA Induced Silencing Complex, RISC. RISC recognises target mRNAs, guided by the sequence-specificity of the bound siRNA or miRNA, and initiates silencing of the target site through PTGS (Figure 1.5A). This differs in *S. pombe* where Argonaute and its bound RNA duplex forms the RNA Induced Transcriptional Silencing complex, where the siRNA targets the complex to regions of the chromosome for silencing via chromatin modifications in a transcriptional gene silencing (TGS) mechanism (Verdel et al., 2004).

The other class of sRNAs, piRNAs, are produced in a Dicer-independent manner. These are generated from piRNA clusters and can then be amplified through a ping-pong cycle. piRNAs bind PIWI proteins, such as piwi-containing Argonaute proteins and other proteins also containing the piwi domain, and can direct silencing through either PTGS or TGS. piRNAs are often found in the germline of metazoans and have been shown to silence transposable elements in metazoans (Huang et al., 2021).

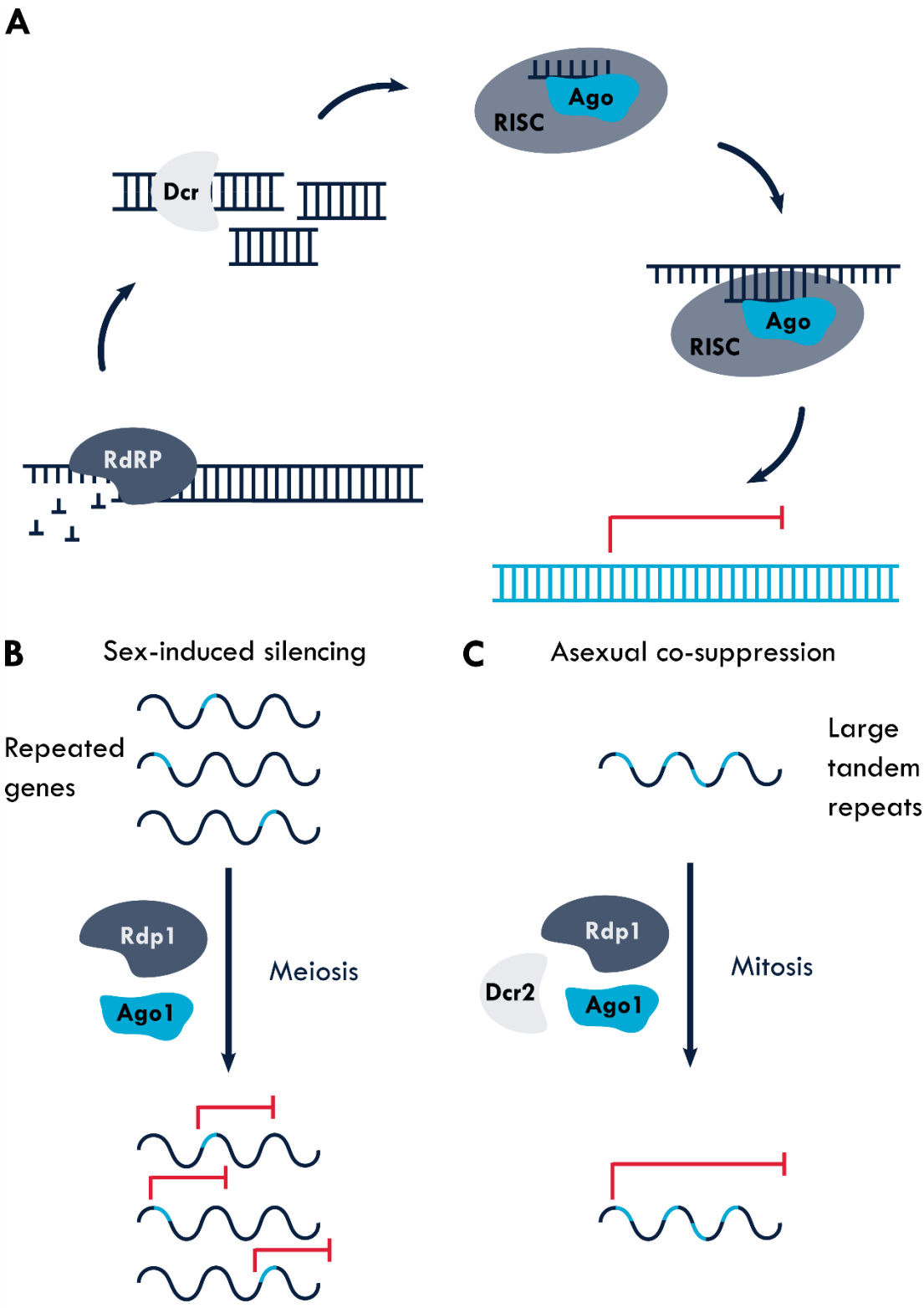

Figure 1.5 RNA intereference mechanisms
Schematics showing **(A)** the canonical RNAi mechanism of silencing DNA, and representations of the transgene-induced **(B)** sex-induced silencing and **(C)** asexual cosuppression present within C. *deneoformans,* with essential proteins for the silencing pathway labelled.

The canonical RNAi pathway relies on the presence of RdRP, Dicer and Argonaute proteins, or homologous versions, for silencing. However non-canonical pathways have been shown to exist in fungi and plants, which rely on the presence of only RdRP out of the core RNAi machinery (Cuerda-Gil & Slotkin, 2016; Trieu et al., 2015). These pathways allow species without the genes encoding Dicer and/or Argonaute proteins to still have a functioning RNAi pathway or have parallel pathways which are independent from these core RNAi factors. This is seen in *Mucor circinelloides* where a Dicer-independent non-canonical RNAi pathway exists alongside the canonical RNAi pathway (Trieu et al., 2015).

1.6.2 RNAi and genome defence

One of the major roles of RNAi across eukarya is as a mechanism of protecting the genome by silencing unwanted or potentially harmful DNA. The idea of homology dependent gene silencing (HDGS) and co-suppression were formed before the identification of RNAi; however, RNAi has since been shown to be the mechanism involved in these processes (Matzke et al., 1989, 2002; Napoli et al., 1990; Romano & Macino, 1992). Here, introduction of transgenes enabled silencing of endogenous genes with sufficient homology, and this has been shown to also apply to regulation of transposable elements which repeat throughout the genome (Nolan et al., 2005). The best-characterised example of RNAi acting in genome defence is quelling (Romano & Macino, 1992). Initially identified in *Neurospora crassa,* quelling involves the core RNAi machinery, including Dicer, Argonaute and RdRP-like proteins and acts as a mechanism for silencing repetitive loci (Catalanotto et al., 2000; Cogoni & Macino, 1999a, 1999b). Aberrant RNA is produced from these repeated regions, which can then be processed into the siRNAs, silencing through the RISC complex, here made up of Argonaute-like Qde-2 and Qip (Maiti et al., 2007). The siRNA targets the RISC complex

to mRNA, and silencing occurs through degradation of the mRNA transcripts. Whilst quelling occurs during vegetative growth, a similar process occurs during meiosis. Here Meiotic Silencing by Unpaired DNA (MSUD) occurs through a parallel RNAi-dependent mechanism of processing aberrant RNA into siRNA and PTGS via degradation, though it utilises different homologs of the core RNAi machinery to quelling (Aramayo & Metzenberg, 1996; Shiu & Raju, 2001). MSUD silences both the unpaired DNA, but also homologous copies of the unpaired DNA, even if it is paired during meiosis.

As well as protecting the genome from transposable elements, RNAi has been shown to have a role in genome defence against exogenous genomic material including viruses. During infection siRNAs are derived from the viral genetic material, allowing the RNAi pathway to silence the virus, thus aiding antiviral immunity (Ding & Voinnet, 2007).

1.6.3 RNAi and heterochromatin

As well as PTGS seen in quelling, RNAi can silence through chromatin modification, such as the formation of heterochromatin. This has been well studied in the yeast *S. pombe* and plant *Arabidopsis thaliana* (Martienssen & Moazed, 2015). In *S. pombe,* the centromeres are flanked by 'innermost' (*imr)* and 'outermost' (*otr)* repeats, with the *otr* containing tandem alternating copies of dg and dh repeats (Pidoux & Allshire, 2005). The pericentromeres are rich in H3K9 methylation, silencing the region via heterochromatin formation as DNA methylation is absent within the species. RNAi is required for the formation of this heterochromatin, with the RITS complex, formed of Ago1, Chp1 and Tas3, recruiting the sole H3 lysine 9 methyltransferase Clr4 (Verdel et al., 2004). Clr4 binds H3K9 methylation marks via its chromodomain and is required for the spreading of heterochromatin by propagating the H3K9 methylation mark through its catalytically active SET domain (Ivanova et al., 1998). Clr4 forms a complex,

the Clr4-containing complex (ClrC), including proteins Raf1, Raf2, Rik1 and Cul4. H3K9 methylation by the ClrC recruits Chp2, which in turn recruits the SHREC complex which has histone deacetylase activity to reinforce heterochromatin formation (Hong et al., 2005; Horn et al., 2005; Jia et al., 2005; Motamedi et al., 2008; Pidoux & Allshire, 2005; Thon et al., 2005). Swi6 is also recruited by ClrC-deposited H3K9 methylation, and maintenance of H3K9 methylation relies on Swi6 recruitment and binding to the H3K9 methylation mark, allowing the mark to be epigenetically maintained during miosis and mitosis (Grewal & Klar, 1996; I. M. Hall et al., 2002; Zhang et al., 2008). The RNAi pathway and histone methylation pathway are intrinsically linked, with deletion of components from either the RNAi pathway or the Clr4 complex CLRC affecting H3K9 methylation (Volpe et al., 2002). This means that it is unclear which pathway drives the establishment of heterochromatin and if other factors are involved.

1.6.4 Loss of RNAi in fungi

Although RNAi is seen throughout all four eukaryotic kingdoms, it is not found in every species. The ability to lose RNAi shows that in at least some organisms RNAi is not an essential process, with many fungi lacking RNAi components and therefore the silencing process. This includes the well-studied baker's yeast *S. cerevisiae* which lacks RdRP, Argonaute and Dicer proteins (Aravind et al., 2000). Other species within the *Saccharomyces* clade, including several *Candida* species, of which many are opportunistic pathogens, also lack several or all RNAi core components. This suggests that RNAi may have been lost in a common ancestor to the complex (Nakayashiki et al., 2006). However, this loss of RNAi is seen throughout the Fungal kingdom in other unrelated species suggesting that many independent loss events have occurred throughout evolution, resulting in evolutionarily distant species sharing this polymorphism. Independent gene expansion events have also been seen across many

unrelated species, where the number of core machinery components present within the genome dramatically increases (Nakayashiki et al., 2006). Although species which have lost RNAi have often lost all RNAi components, resulting in a polymorphism of functionally linked genes (Feretzaki et al., 2016), some species have lost only a subset of the core machinery. In some cases, species have evolved and adapted to allow canonical RNAi to still occur without the full complement of proteins, such as in *S. castellii* where the loss of RdRP has not resulted in the loss of the RNAi pathway (Drinnenberg et al., 2009). Other species which have lost RNAi have maintained some core components due to their role in other essential functions. This is seen in C. *albicans* where RNAi is unable to be triggered by a dsRNA hairpin despite having conserved Argonaute and Dicer-like proteins, but the Dicer protein has been shown to be maintained for its essential role as the sole RNase III enzyme, required for RNA maturation (Bernstein et al., 2012; Staab et al., 2011). Non-canonical RNAi pathways also exist, which can occur in species alongside the canonical RNAi pathway or individually. Non-canonical RNAi pathways can utilise some of the core RNAi components, but importantly silences through a different mechanism. An example of a non-canonical RNAi pathway (NCRIP) is the RdRP-dependent degradation pathway in *M. circinelloides,* which interacts with the canonical RNAi pathway within the species (Calo et al., 2017; Trieu et al., 2015). The ability for species to utilise a non-canonical method or find other ways to counteract the loss of RNAi is thought to relate to differences between losing the entire RNAi gene network in a short space of time and losing the machinery gradually over time (Nicolás et al., 2013).

Within the *Cryptococcus neoformans/gattii* species complex, whilst C. *deneoformans* and C. *neoformans* both have RNAi, C. *deuterogattii* has lost all RNAi components, with genes encoding Rdp1 and Ago1 being deleted and Dcr1 present but truncated and non-

functional. This very clearly illustrates the gene network polymorphism, with all components being selectively lost in one species but present in all closely related species. Moreover, analysis of other genes within the polymorphism highlights additional factors that might be part of the same functional network out with the core RNAi machinery (Feretzaki et al., 2016). Something similar can also be seen in *Ustillago maydis* which has lost RNAi while other closely related species *U. hordei* and *U. bromivora* still have functional pathways (Laurie et al., 2008). Both C. *gattii* and *U. maydis* share other similarities in comparison with their close relatives, such as their shorter centromere length, loss of DNA methylation (Yadav et al., 2018), and lack of transposable elements (Kämper et al., 2006; Laurie et al., 2008). It is unknown whether these features are a result of losing RNAi, or if this is a consequence of other evolutionary events.

1.7 RNAi in *C. deneoformans*

1.7.1 RNAi machinery in *C. deneoformans*

RNA interference was first identified in C. *deneoformans* as a method of artificial gene silencing as early experiments using *Cryptococcus* as a model organism were hindered by the inability to alter its genome due to its low recombination rate (Liu et al., 2002). RNAi was proposed as an alternative method, allowing the study of mutant phenotypes recreated through gene silencing. Although this method showed variation in the efficiency of silencing different genes, it proved that C. *deneoformans* has a functional RNAi pathway and therefore must include the RNAi machinery within its genome. Database searches and phylogenetic analysis against *Neurospora* RNAi components identified two Argonaute-like proteins (Ago1 and Ago2), two Dicer-like proteins (Dcr1 and Dcr2) and one RNA dependent RNase Polymerase-like protein (Rdp1) within the C. *deneoformans* genome (Janbon et al., 2010). Here it was discovered that both Dicer-

like proteins lacked the DEAD/H box helicase domain responsible for unwinding dsRNA, suggesting diversification of the Dicer protein from other fungi (Nakayashiki et al., 2006). Although unusual, the lack of a helicase is also seen in one of the three *Tetrahymena* Dicer-like proteins which is still functional (Mochizuki & Gorovsky, 2005) and in both of the slime mould *Dictyostellium* Dicer-like proteins (Martens et al., 2002). Further analysis showed that the closely related species C. *neoformans* also has RNAi machinery, with two Dicer-like proteins which also lack the DEAD/H box domains, however this species variant only has one Argonaute-like protein unlike the two present in the C. *deneoformans* genome (Nakayashiki & Nguyen, 2008). The Argonaute and Dicer-like genes present in C. *deneoformans* were named according to their sequence similarity with the C. *neoformans* genes for consistency. This variation in the core RNAi machinery encoded within the genomes of the two species is reflected throughout the whole C. *neoformans/gattii* species complex, with two other species also having the full complement of RNAi components that C. *deneoformans* has, and two other species having lost Ago2 from their genome, similar to C. *neoformans* (X. Wang et al., 2010). Out of the seven species within the complex, only C. *deuterogattii* does not have a functional RNAi pathway with Dcr2 being the only full-length core RNAi component encoded within its genome (Figure 1.6).

As there are paralogous genes present for some of the core RNAi machinery, the requirement of each RNAi component in C. *deneoformans* was tested using an artificial RNAi system, whereby expression of an RNA hairpin generates siRNAs against the *ADE2* gene, resulting in a visible silencing phenotype in WT cells. Analysis of individual deletion strains showed that whilst Rdp1, Ago1 and Dcr2 are required for RNAi to function, with Rdp1 deletion showing the greatest phenotype, deletion of either Ago2 or Dcr1 has little effect on the RNAi pathway. However, deleting both Ago1 and Ago2

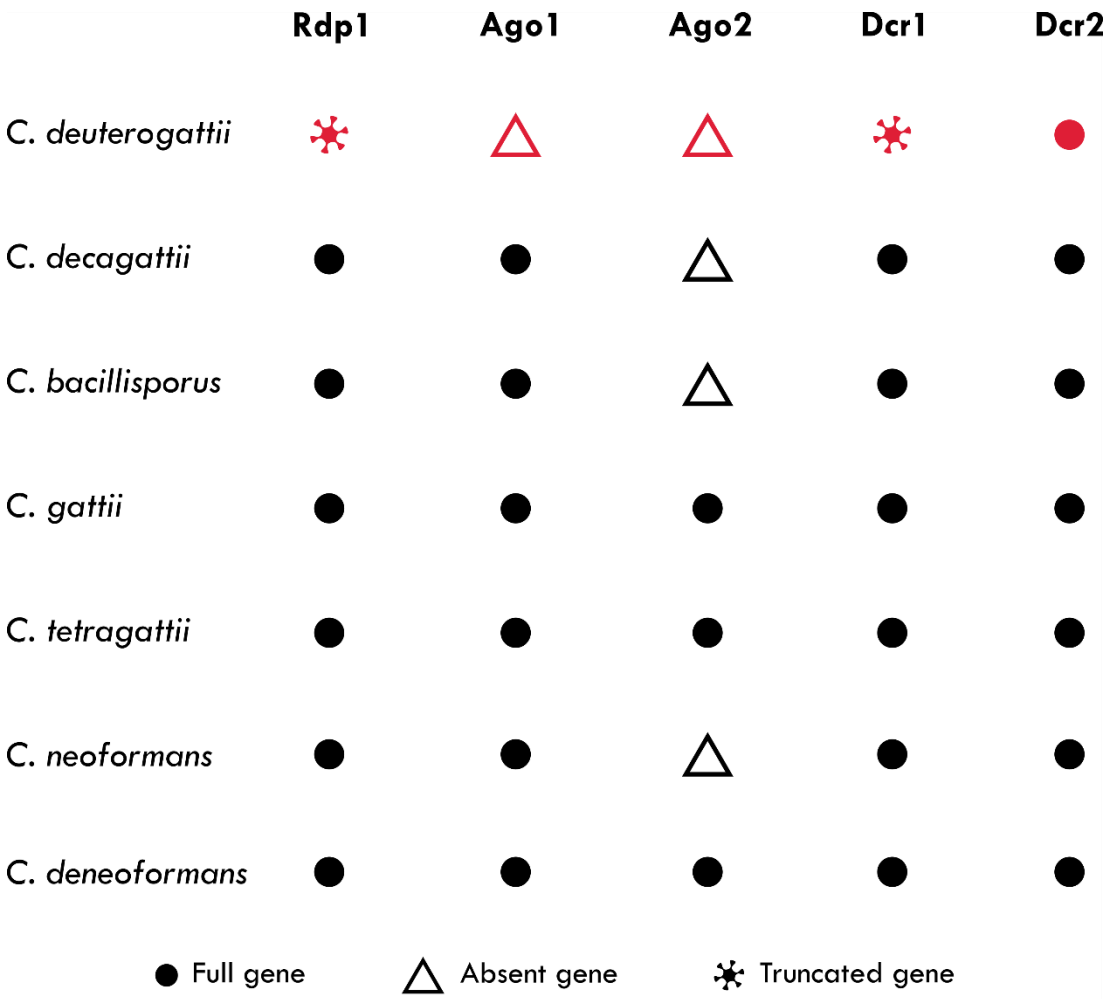

Figure 1.6 The core RNAi machinery varies between species in the C. *neoformans*/*gattii* complex

Schematic showing which of the five core RNAi machinery components (Rdp1, Ago1, Ago2, Dcr1, Dcr2) are encoded within the genome of each species in the C. *neoformans*/*gattii* species complex. Red symbols indicates a total loss of RNAi within the species. Figure is adapted from Feretzake et al., 2016.

showed a greater deficiency in RNAi than either single deletion, suggesting that although Ago2 is not essential for RNAi, it must have at least a minor role within the pathway (Janbon et al., 2010). Further experiments on both Dicer proteins in C. *deneoformans* have shown they also both have a role in RNAi as neither Dcr1 or Dcr2 deletion strains were able to support silencing of a *URA5* reporter by two identified miRNAs, miR1 and miR2, in a transgene silencing assay (Jiang et al., 2012). The presence of both Argonaute proteins within C. *deneoformans,* along with their individual roles, is of interest to further understand why both genes are conserved in this species and not in some closely related species.

1.7.2 Sex-induced silencing and asexual co-suppression

Two different transgene-induced types of RNA interference have been identified in *Cryptococcus*. The first, Sex-Induced Silencing (SIS), was discovered in C. *neoformans* when the mating type locus genes *SXI1alpha* and *SXI2a* were deleted and reintroduced by plasmid at the *URA5* locus. After crossing 50% of the F_1 progeny were unexpectedly uracil auxotrophic despite having the intact *URA5* gene present, indicating silencing of the *URA5* gene. This transgene induced silencing occurred when multiple copies of the plasmid were present and at significantly higher rates during meiosis than mitosis, and was dependent on Rdp1 and Ago1, identifying it as an RNAi pathway (Figure 1.5B) (X. Wang et al., 2010). RNA-Pol-II-ChIP and qPCR showed similar *URA5* transcript levels in both Ura^+ and Ura^- strains, suggesting a post transcriptional gene silencing method, with the conditions required to trigger silencing differing between genes as *SXI2a* was never silenced even when there were similar copy numbers present to that of *URA5*. This shows similar variations in the efficiency of silencing different genes by RNAi as previously seen in other systems (Liu et al., 2002). Further analysis of small (s)RNAs in C. *neoformans* led to the identification of siRNAs against transposable

elements, which show increased transcript accumulation during mating in Rdp1 deletion mutants, suggesting that mating induced RNAi focuses on silencing repetitive elements which is why multiple insertions of the transgene were able to induce SIS (X. Wang et al., 2010). As fungi can undergo unisexual reproduction, which is particularly common in C. *deneoformans,* the transgene induced SIS was also tested during α-α asexual mating in C. *deneoformans*. This proved that SIS also functions in this species, and that this type of RNAi silencing can also occur during unisexual mating, where it is again Rdp1 dependent and more prominent during meiosis compared with vegetative growth. Here SIS was also shown to work with another transgene not related to mating, showing that the pathway can be induced by transgenes independent of a role in mating (X. Wang et al., 2013).

The second type of RNAi recognised in C. *neoformans* is co-suppression, which follows a similar principle to quelling in *N. crassa* along with having similarities to SIS (Figure 1.5C). This form of transgene induced RNAi was identified when multiple copies of a *cpa1*:*ADE2* disruption allele, comprising the full-length *CPA1* gene with *ADE2* insertion, were inserted into the genome, and 25% of strains were found to be CsA resistant, something that usually only occurs when both *CPA1* and *CPA2*, which share 85% nucleotide homology, are disrupted. This unexpected silencing of both *CPA1* and *CPA2* occurred during vegetative growth and was dependent on the core RNAi machinery: Rdp1, Ago1 and Dcr2 (X. Wang et al., 2012). The main difference between co-suppression and SIS is the occurrence of transgene-induced silencing during asexual growth unlike the silencing identified previously during sexual growth (X. Wang et al., 2010). Co-suppression is also partially dependent on the copy number, favouring loci where the transgenes are organised into large tandem repeats, with the levels of siRNA produced reflecting the efficiency of silencing in each strain. This form of silencing once

again showed a gene dependent difference in silencing efficiency, with *ADE2* being less sensitive to silencing than *CPA1* and *CPA2* (X. Wang et al., 2012).

1.7.3 SCANR complex

The Spliceosome-Coupled And Nuclear RNAi Complex, SCANR, has been uniquely identified in *Cryptococcus*. Discovered in C. *neoformans*, this complex represents a potential link between targets for the RNAi pathway and the splicing of introns (Dumesic et al., 2013). Composed of four proteins, the SCANR complex includes the essential RNAi proteins Rdp1 and Ago1 along with Gwc1, a protein containing five GW/WG dipeptide motifs often found in Argonaute binding proteins (El-Shami et al., 2007), and Qip1, an exonuclease whose homolog in *N. crassa* is known to bind Argonaute proteins and have a role in degrading passenger strands of siRNA duplexes (Maiti et al., 2007). All four members of the SCANR complex are either absent or severely truncated within the C. *gattii* strain R265 which does not have a functional RNAi pathway, confirming their suggested link to the RNAi polymorphism (Feretzaki et al., 2016). The SCANR complex is thought to localise within the nucleus, associating and competing with some spliceosomal complexes. RNAi requires the formation of siRNA against target sites, and as cryptococcal siRNAs have been shown to map to both introns and exons, this suggests that precursor mRNA could be required for siRNA production. When DNA is transcribed, the precursor mRNA is targeted towards the spliceosome for processing to produce mRNA ready for translation. The splicing process requires sequence specific splice sites in the RNA along with splicing factors and associated enzymes for efficient spliceosome function, with other factors such as intron length also affecting splicing efficiency (Warnecke et al., 2008). Inefficient splicing can cause pre-mRNA to stall at the spliceosome, and it is proposed that whilst stalled, the pre-mRNA can instead be targeted by SCANR for siRNA production. This is thought to be how at least a subset of

siRNAs accumulate, as their levels have been shown to be dependent on both entry of precursor RNAs to the spliceosome pathway and the activity of lariat debranching enzymes during splicing (Dumesic et al., 2013). It has also been suggested that this may be how siRNAs against transposons accumulate, as introns within transposons are less likely to splice at similar efficiencies as introns within host genes, and introns are likely insertion sites for transposons due to minimal host gene disruption.

1.7.4 Other RNAi factors

Other proteins have been identified as having a link to siRNA biogenesis and function alongside the core RNA-dependent polymerase, Argonaute and Dicer RNAi proteins. As mentioned above, comparison of the C. *deneoformans* JEC21 and B-3501A, C *neoformans* H99 and C. *gattii* WM276 genomes against the C. *deuterogattii* R265 genome identified an RNAi gene network polymorphism (Feretzaki et al., 2016). In the C. *deuterogattii* genome, 14 genes were found to be either completely absent or partially deleted with greater than 50% of the gene missing. These 14 genes were all suspected as being involved within the RNAi pathway, based on the roles of the known genes included within the 14, specifically the genes encoding the canonical RNAi components Rdp1, Ago1 and Dcr1, along with the SCANR components Gwc1 and Qip1 (Dumesic et al., 2013; Feretzaki et al., 2016).

Gwo1 was one of the factors also identified within the network polymorphism. Gwo1 was previously identified as an Ago1 interactor and has been shown to interact with Ago1 independently of the SCANR complex, with the Ago1-Gwo1 complex named the P-body-associated RNA Silencing Complex (PRSC) due to its localisation within the cytoplasm at the P-bodies (Dumesic et al., 2013). Deletion of Gwo1, a protein similar to the SCANR Gwc1 component due to the presence of GW/WG dipeptide motifs

often found in Argonaute binding proteins, resulted in increased RNAi target transcript levels. However, siRNA levels remained the same suggesting that Gwo1 has a role in RNAi out with siRNA biogenesis. In the proteomic analysis that identified Gwo1 as an Ago1 interactor independent of the SCANR complex, five other proteins were also identified, however none of these five proteins (Skp1, Aga1, Aga2, Bre1, and Aga3) showed any role in RNAi (Dumesic et al., 2013).

Other proteins within the network polymorphism include two proteins involved in unisexual mating, Znf3 and Cpr2 (Feretzaki et al., 2016). Znf3 is a zinc finger protein involved in the regulation of unisexual reproduction. Deletion of *ZNF3* and subsequent transcript analysis during unisexual reproduction showed increases in transcript levels of genes encoding putative transposases, endonucleases, and RNA-dependent DNA polymerases, all of which are typically encoded by transposons, along with an increase in T1 and T3 TE expression. This transcript profile is similar to that of a *RDP1* deletion strain, suggesting that Znf3 functions along with Rdp1 in RNAi-mediated regulation of transposons alongside its essential role in unisexual reproduction (Feretzaki & Heitman, 2013). Further studies into Znf3 showed that it is not required for silencing during mitosis, and that its role in RNAi is only during SIS (Feretzaki et al., 2016).

The remaining genes identified within the network polymorphism are *FCR1, OXR1, CDP1, MEH1* and transcription factors *FZC47* and *FZC28,* the latter of which was shown to be involved in SIS but not silencing during mitosis. The other genes haven't been investigated in respect to their roles in RNAi, but they are likely to have an involvement within the pathway due to their inclusion within the network polymorphism (Feretzaki et al., 2016).

Another assay to identify RNAi factors used a *ura5::HAR1* background strain, where lack of RNAi silencing allows the harbinger transposon to cut and paste elsewhere in the genome leaving the *URA5* gene intact. They then introduced a *NAT*-resistance cassette bound by T-DNA from *A. tumefaciens* which randomly inserted into the genome. Strains that are both *NAT* resistant and uracil prototrophic were analysed to identify the gene in which the T-DNA *NAT*-resistance cassette was inserted into. Five novel proteins were identified, Rde1-5, whose deletion results in partial loss of siRNAs, reminiscent of the loss of siRNAs when canonical RNAi machinery is lost (Burke et al., 2019). A further protein, Prp43, was identified which associated with Rde1, with Prp43 mutants also exhibiting a loss in siRNAs. Prp43 is a DEXD-box helicase involved in ribosome biogenesis and pre-mRNA splicing by disassembling stalled spliceosomes. Mutants of Prp43 also increased the amount of Ago1 pulled down with Rde1, and Gwo1, which is absent in wild-type Prp43 pull downs, was also detected. This again links stalled spliceosomes with RNAi and suggests a potential relocalisation of Ago1 when there are changes to spliceosome stalling. (Burke et al., 2019).

1.8 Role of RNAi in *C. deneoformans*

1.8.1 sRNA targets

Although the two different types of RNAi mechanisms identified in C. *deneoformans*, SIS and asexual co-suppression, have both been induced through transgene introduction, it is assumed that similar pathways function throughout the normal life cycle. To gain insight into the role of RNAi within a cell without transgene introduction, studies have carried out sRNA sequencing to identify potential target sites of the endogenous RNAi pathway. Within C. *neoformans,* majority of the sRNAs identified map to centromeric retrotransposons, with the production of these sRNAs largely dependent on the presence of the core RNAi machinery (Burke et al., 2019; Dumesic et al., 2013; X. Wang et al.,

2010). However, sRNAs have also been identified in the RNAi-deficient species C. *deuterogattii* (Ferrareze et al., 2017). These also map to the centromeric retrotransposon sequences, even though the retrotransposons are mostly truncated when compared with the retrotransposons in C. *deneoformans* and C. *neoformans*. This suggests that a sRNA silencing mechanism may be possible within C. *deuterogattii,* acting through a non-canonical RNAi mechanism, with the sRNAs biogenesis and function via other unidentified factors.

Screening for miRNA within C. *deneoformans* also identified miR1 and miR2 as being involved in RNAi silencing. Both miRNAs have multiple copies present throughout the genome, with most loci encoding a transposable element or pseudogene. These repetitive sequences are thought to initiate silencing of their loci via RNAi as silencing is abolished in RNAi deficient strains (Jiang et al., 2012).

1.8.2 Transposons and RNAi

One of the major roles of RNAi is in genome preservation via the control of transposable elements. As transposons are usually present in several copies throughout the genome, a primary role for RNAi in targeting such repetitive elements within C. *deneoformans* would be consistent with the observed silencing of multi-copy transgenes (X. Wang et al., 2010, 2012). In support of this, Rdp1 deletion strains also show increases in DNA transposon mobility (Janbon 2010) and deletion of other proteins involved in RNAi such as Qip1 and Znf3 result in an increase in retrotransposon transcript abundance (Feretzaki 2016).

Transposition assays have been undertaken on several strains of C*ryptococcus deneoformans* both *in vitro* and within host infection environments (Gusa et al., 2020). These analyse transposition by assaying insertions into a reporter gene whose disruption

confers drug resistance, and subsequent screening identifies the type of elements, frequency of transposition events occurring and position of insertion sites. Transposition rates in C. *deneoformans* are higher during the infection of mice than culture growth *in vitro,* although transposition rates *in vitro* are temperature dependent, with rates higher at 37°C than at 30°C. However, this level is still lower than the rate within a mouse model. Deletion of Rdp1 does not increase the rate of transposon insertions at 30°C to the rate seen at 37°C, suggesting that the increase in transposon mobility at 37°C is not because of a lack of RNAi function due to protein denaturing, but rather due to some other mechanism (Gusa et al., 2020). These assays only screen one or two genes whose disruption confers drug resistance, for instance selecting for 5-FOA resistance resulting from disruption of *URA3* and *URA5*. However, biases have been seen in the transposon insertions in *URA3* and *URA5*, both in the type of TE as well as the position and orientation of the insertion, suggesting either a sequence or folding preference. Differences have also been found between the C. *deneoformans* wild-type strains showing that there are vast differences in the mutation rate, with some becoming hypermutators (Priest, Yadav, et al., 2021). By looking at transposon expression and mobility, it is possible to look at repeated sequences within the genome without having to introduce a transgene, therefore looking at endogenous processes and not effects artificially induced due to the foreign material. Utilising these assays can help understand the role of RNAi within the cell in a more natural environment.

1.8.3 H3K9 methylation, DNA methylation and RNAi

A link between RNAi and DNA methylation has been seen in plants, with sRNA directing the locations of DNA methylation (Erdmann & Picard, 2020). Dnmt5 is truncated within C. *deuterogattii*, potentially linking DNA methylation in with the RNAi polymorphism within this species (Feretzaki et al., 2016). The truncated form is non-functioning causing

a loss of 5mC (Yadav et al., 2018). With majority of the 5mC selectively maintained over TEs in both C. *deneoformans* and C. *neoformans*, the lack of full-length TEs within C. *deuterogattii* could explain the loss of Dnmt5 as methylation would no longer be required to silence them (Catania et al., 2020). However, without knowing the order of loss and truncation of *DNMT5* and the 14 RNAi polymorphism genes within C. *deuterogattii,* it is not possible to confirm whether the loss of 5mC contributed to the loss of RNAi, or if it is a side-effect of the loss. Studying both RNAi and DNA methylation in a species containing the two functional pathways, such as C. *deneoformans,* will help explore if Dnmt5 might be part of the RNAi polymorphism.

Also, with a link between RNAi and histone methylation determined in *S. pombe,* the possibility of a similar relationship in C. *neoformans* has been briefly explored. H3K9 methylation levels at Tcn1 elements were shown to be comparable between WT and *rdp1* mutant cells, suggesting that Rdp1 and RNAi do not control H3K9 methylation levels at these elements. However, in WT and *rdp1*Δ cells, a three-fold decrease in H3K9me2 was seen during mating, along with increases in Tcn1 transcription, suggesting that silencing of retrotransposons occurs via histone methylation independent of RNAi (X. Wang et al., 2010). This is consistent with a six-fold increase in centromeric transcript levels when Clr4, and subsequentially H3K9 methylation, has been lost (Dumesic et al., 2015). However, this contradicts the increased expression levels of Tcn1, Tcn3 and Tcn4 seen during mating of *rdp1, ago1* and *dcr1/2* mutant strains, which suggests that RNAi silences retrotransposons during mating. This increase in expression level occurring during mating corresponds with a time when RNAi is more active, as shown through SIS (X. Wang et al., 2010). Taking all of this data into consideration, it could be that there is a cross-talk between the two pathways, however, further investigation into the link

between RNAi and heterochromatin is required before any firm conclusion can be drawn.

1.9 Aims of study

The yeast C. *deneoformans* is growing in popularity as a model organism, partly from a medical perspective to help understand how C. *deneoformans* infection manifests and further treatment possibilities, but also due to its evolutionary relationship with other species within the C. *neoformans/gattii* species complex. Differences in genome size, centromere size, TE content and pathogenicity of species within the complex seem to correlate with whether the species have a functional RNAi pathway or not. These relationships with RNAi are of interested, as it is unknown whether they are caused by the lack of RNAi or RNAi is lost as a result of these differences. C. *deneoformans,* as opposed to C. *neoformans,* is of particular interest as a model in which to study RNAi, as it has the full complement of RNAi machinery present, and also in which to study the transposon landscape due to the increased levels and types of TEs present within the genome.

In this study, I aimed to investigate the role of RNAi within C. *deneoformans,* focussing on endogenous genome regulation. Initially I aimed to elucidate the mechanisms of silencing via RNAi, exploring whether the loss of RNAi is associated with the loss of H3K9 methylation and/or increased accumulation of target transcripts. I also aimed to further explore the link between RNAi and DNA methylation at the centromeres as well as exploring the role of RNAi in silencing transposons through transposition assays. Finally, I aimed to explore the functional differences between Ago1 and Ago2, to try to understand the evolutionary reasons why both proteins have been maintained within the C. *deneoformans* genome. Through mass spectrometry I identified different binding

partners for each Argonaute protein and looked further at the interaction between Ago1 and Gwo1.

Chapter 2 Materials and methods

2.1 *C. deneoformans* culture and media

2.1.1 Strains

The strain used throughout is C. *deneoformans* JEC21, with all genetic manipulations originating from this strain. The full list of strains used can be found in Table 2.1.

2.1.2 Growth media

C. *deneoformans* was usually cultured in nutrient rich YPSUC agar for both WT and genetically modified strains. YPSUC + Hygromycin (HYG) or CAS-SUC-AWU were used for strains with a *HYG*^{R} or *URA5* insert respectively when a selection pressure was needed to be maintained. The replacement of sucrose in the media to either galactose or glucose was used to alter gene expression when the *pGAL7* promoter was present, and glucose-based medium was use as the rich media for the mutation rate assay. CAS-GLU-AW with 5-Fluroorotic acid (5-FOA) was used to identify cells which were auxotrophic for uracil. The composition of each media is listed below:

YPSUC medium: 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) sucrose. For YPSUC agar, 2% (w/v) agar was added. For YPSUC + HYG, 0.1 mg/mL Hygromycin B was added. For YPGAL or YPD medium sucrose was replaced with equivalent amounts of galactose or glucose respectively.

CAS-SUC-AWU medium: 2% (w/v) cas amino acids, 2% (w/v) sucrose, 0.8% (w/v) yeast nitrogen base (YNB). For CAS-SUC-AWU agar, 2% (w/v) agar was added. For CAS-GAL-AWU or CAS-GLU-AWU medium sucrose was replaced with equivalent amounts of galactose or glucose respectively.

Table 2.1 *C. deneoformans* strain list

Strain Number	Strain Name	Genotype
3084	Jec21	*Cryptococcus deneoformans* JEC21, haploid and mating type alpha, serotype D (From Robin May Lab, Birmingham)
3166	WSA1226	JEC21 *URA5*Δ::hisG
3434	*dcr1*Δ*dcr2*Δ	JEC21, CNC03670-CNC03680Δ::*NAT*^{R}
3639	FTH-Ago2	WSA1226 FTH:CNJ00610
3695	*dcr1*Δ*dcr2*Δ	WSA1226 CNC03670Δ-CNC03680Δ::*NAT*^{R}
4000	FTH-Ago1	WSA1226 FTH:CNJ00490
4044	*dcr1*Δ*dcr2*Δ FTH-Ago2	WSA1226 FTH:CNJ00610, CNC03670Δ-CNC03680Δ::*HYG*^{R}
4730	CNH0900Δ	WSA1226 CNH00900Δ::*HYG*^{R}
4731	*clr4*Δ	WSA1226 CNH00720Δ::*URA5*
4920	*clr4*Δ*dnmt5*Δ	WSA1226 CNH00720Δ::*URA5* CNK01170Δ::*HYG*^{R}
5022	*dnmt5*Δ	WSA1226 CNK01170Δ::*URA5*
6018	*rdp1*Δ	WSA1226 CNH01230Δ::*URA5*
6019	*rdp1*Δ*dnmt5*Δ	WSA1226 CNK01170Δ::*HYG*^{R} CNG01230Δ::*URA5*
6020	*rdp1*Δ*clr4*Δ	WSA1226 CNH00720Δ::*URA5* CNH01230Δ::*HYG*^{R}
6568	*dcr1*Δ*dcr2*Δ	JEC21, CNC03670-CNC03680Δ::*HYG*^{R}
7883	*pGAL7*-FTH-Clr4	WSA1226 *HYGR*:*URA5*:pGAL7:FTH:CNH00720
7884	*RI-CLR4*	WSA1226 CNH00720Δ::*URA5*::CNH00720
7885	*RI-CLR4*	WSA1226 CNH00720Δ::*URA5*::CNH00720
7886	FTH-Ago1 *gwo1*Δ	WSA1226 FTH:CNJ00490, CNA05020Δ::*URA5*

5-FOA CAS-GLU-AW agar: 2% (w/v) cas amino acids, 2% (w/v) sucrose, 0.8% (w/v) yeast nitrogen base (YNB), 2% (w/v) agar, 200 mg/L uracil, 0.25 g/L 5-FOA.

2.1.3 Cell culture

C. *deneoformans* was usually grown on agar medium at 30°C for 2 days. Liquid cultures were usually grown in 100 mL liquid media at 140 rpm, 30°C overnight using inoculum taken from a fresh plate streak. Cell cultures would usually be knocked back and allowed to regrow and harvested during exponential growth phase as determined by the OD_{600} (exponential growth was determined at OD_{600} = 1.2). Cell concentration was determined using a haemocytometer only for growth curves.

2.1.4 Mutation rate assay

C. *deneoformans* cultures were grown in 50 mL liquid rich YPD media at 140 rpm, 30°C for 3 days using inoculum taken from a fresh plate streak grown on CAS-GLU-AW to check that all strains have a functioning *URA* pathway. After 3 days of growth, the OD_{600} was determined and 1 mL of culture at OD4, pelleted and resuspended in 100 µL YPD liquid media, was plated onto 2 CAS-GLU-AWU 5'-FOA plates and incubated at 30°C for 7 days. Colonies were counted and marked after days 2, 4 and 7, and up to 24 colonies across all days were restreaked onto CAS-GLU-AW 5'-FOA plates and incubated at 30°C for 7 days, or until grown, for genomic extraction. Viable cell counts were determined by plating out 100 µL of OD0.001 culture onto YPD plates and incubating at 30°C for 2 days before colonies on one quarter of the plate were counted. This was repeated for 10 independent cultures for each strain (Figure 2.1).

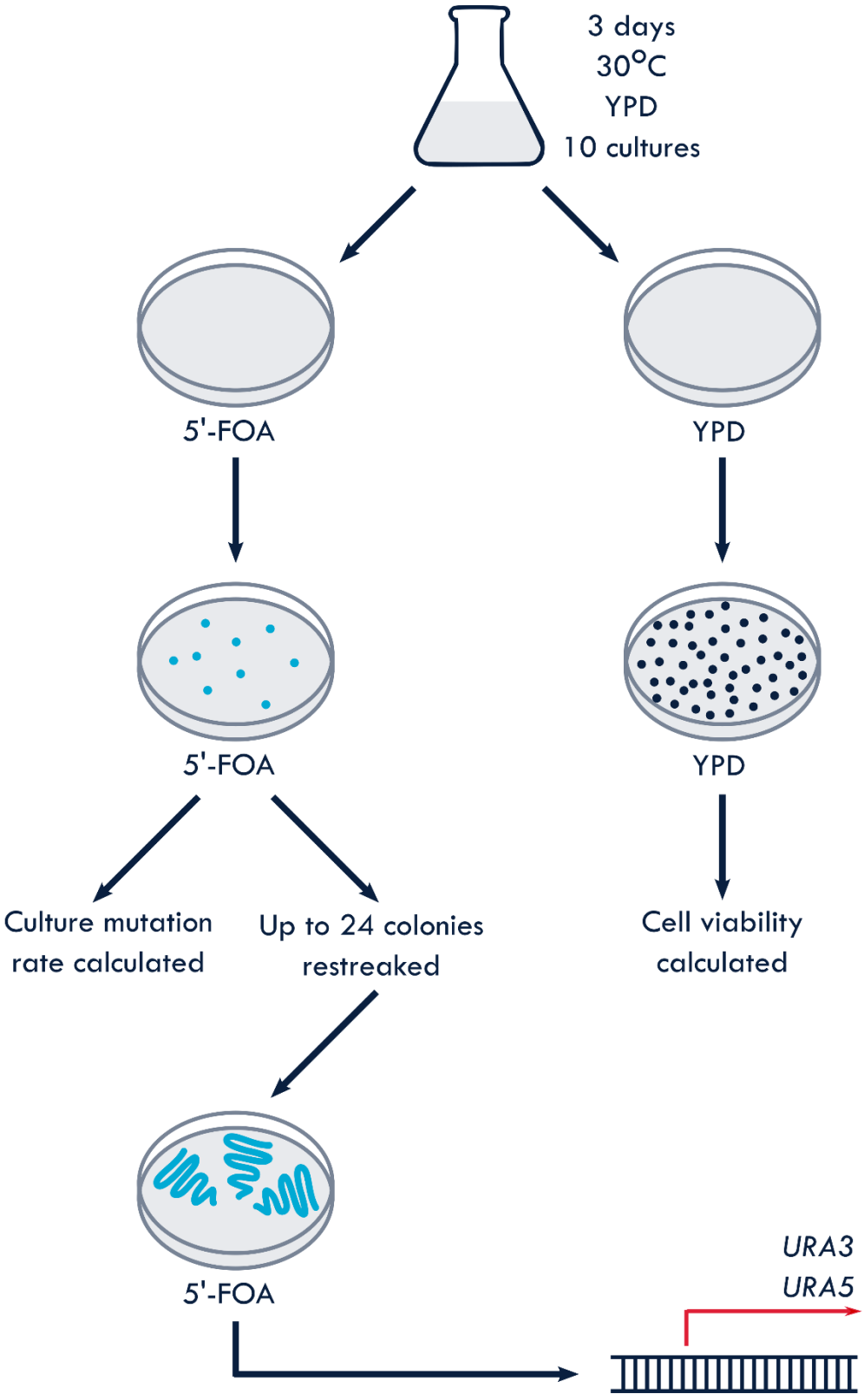

Figure 2.1 Mutation rate assay protocol
Schematic representation of the mutation rate assay, indicating culture length and medium, repeat and screening sample sizes and genes screened to determine transposition rates, as well as calculations determined.

2.2 *C. deneoformans* molecular genetics

2.2.1 CRISPR method for tagging and deletions

C. *deneoformans,* and in particular JEC21, has a low homologous recombination rate making genome editing hard (Liu et al., 2002) so to increase the efficiency of successfully editing the genome a split marker suicide CRISPR-Cas9 method was used. This combined the suicide CRISPR method allowing for the removal of gDNA and C*as9* endonuclease after editing to prevent restoration of the edited section (Y. Wang et al., 2016), and a split marker system which reduces the number of transformants that need to be screened for the correct integration (Kim et al., 2009). DNA fragments were inserted into either RH6 or KA7 plasmid base, an ampicillin resistant plasmid containing the C*as9* gene and sgRNA for CRISPR. gDNAs were manually designed, with both plasmids in the split marker system having a different DNA target for either end of the insertion site. ~1 kb regions homologous to either side of modified site were used to ensure recombination into the correct site. Tagged proteins were initially placed under a *pGAL7* promoter with two marker genes inserted upstream of the gene. These markers were removed through homologous recombination with repair plasmids combining up to four different gDNAs cleaving the markers at multiple sites to encourage their complete removal and placing the gene under its endogenous promoter (Figure 2.2). Deletion strains were created using the same CRISPR method, inserting one marker gene into the gene of interest (Figure 2.3A). Re-introduction strains involved the removal of the marker gene, with gDNAs created for each end of the marker inserted (Figure 2.3B). Plasmids were linearised for electroporation, ensuring that an overlap of ~1 kb occurred within the marker gene for recombination, and that this one recombination event would give a functional linear CRISPR construct with gDNA, insert and C*as9*.

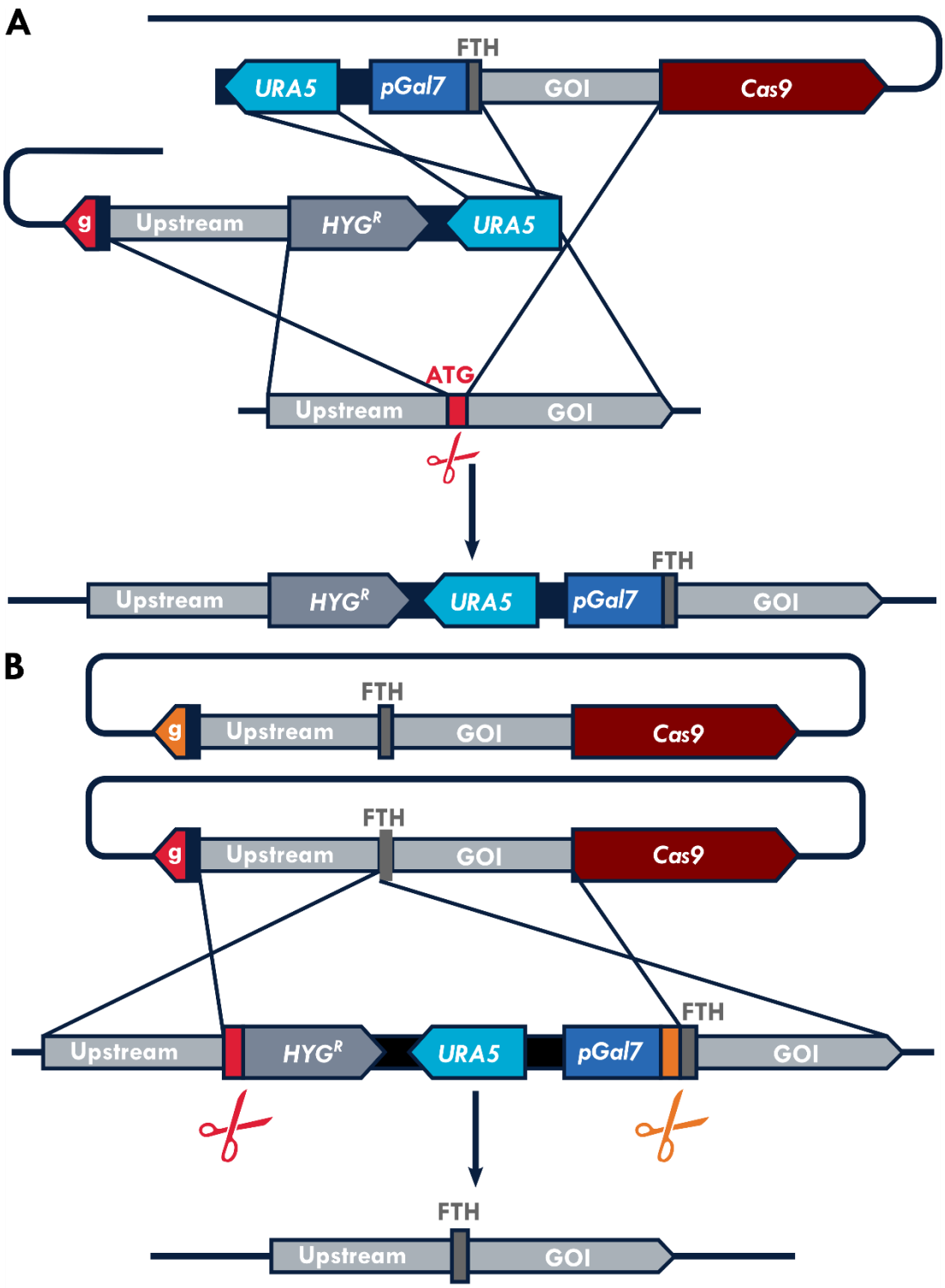

Figure 2.2 Split marker suicide CRISPR method for tagging genes of interest
Schematic showing **(A)** the initial tagging step inserting marker genes *URA5* and *Hyg*^{R}, *pGAL7* promoter and Flag tag, Tev cleavage site and His x6 tag (FTH) and **(B)** showing the removal of the marker genes *URA5* and *HYG*^{R} and *pGAL7* promoter, leaving the FTH tag and the gene of interest (GOI) under control of the endogenous promoter. Scissors show cutting sites of the guide sequence (g). Lines show regions where recombination occurs.

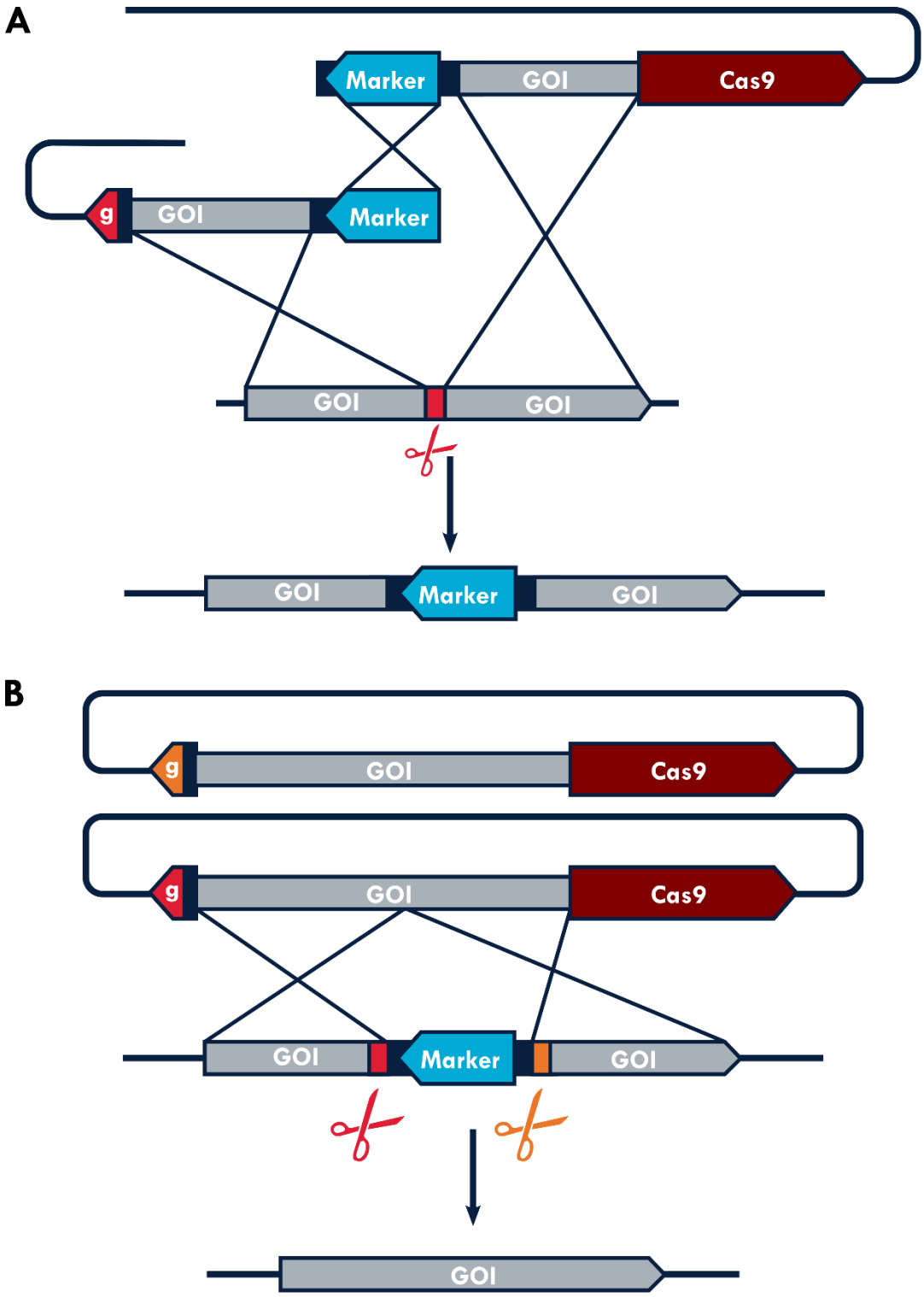

Figure 2.3 Split marker suicide CRISPR method for disrupting genes of interest and re-introducing the gene

Schematic showing **(A)** the deletion disruption of a gene of interest by inserting a marker gene into the middle of the gene of interest and **(B)** removing the marker gene disruption and re-introducing the gene of interest. Scissors show cutting sites of the guide sequence (g). Lines show regions where recombination occurs.

2.2.2 Electroporation

A single colony of C. *neoformans* was cultured in 50 mL at 140 rpm, 30°C overnight. Cultures were diluted to OD 0.1 in 150 mL medium and grown until OD 0.3. Cells were pelleted by centrifugation at 3,000 rpm, 4°C for 5 minutes. Pellets were washed twice in 40 mL cold water, centrifuging at 3,000 rpm, 4°C for 5 minutes each time before resuspending in 45 mL cold electroporation buffer (EB) (10 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$, 270 mM Sucrose (9.24 g/100 mL) with 200 µL 1 M DTT and incubated on ice for 15 minutes. Cells were centrifuged at 3,200 rpm, 4°C for 5 minutes, pellets resuspended in 45 mL cold EB and centrifuged at 3,000 rpm, 4°C for 5 minutes before resuspending in 200 µL cold EB. 40 µL cells were added to 5 µL of linearised plasmid dissolved in EB and incubated on ice for 1 minute. Cell and plasmid suspension were then transferred to ice cold 2mm cuvettes for electroporation using Gene Pulser Xcell electroporation systems (Bio-Rad) at the following settings: V = 1400 v, C = 25 µF, R = 600 Ω, Cuv = 2 mm with the pulse length recorded in ms for each sample. Cells were immediately mixed with 1 mL warm YPSUC and incubated at 30°C for 10 minutes. For transformations with plasmids to remove a URA5 marker gene, cells were incubated at 30°C for a further hour, before being plated directly onto 5-FOA media for selection, and incubated at 30°C. For all other transformations, cells were plated onto YPSUC and incubated at 30°C overnight before replica plating onto selective plates. For all transformations, single colonies were screened by re-streaking on selective plates and confirmed through genomic DNA isolation and sequencing.

2.3 Molecular cloning

2.3.1 Ligation of guide DNA (gDNA)

Annealed gDNA oligonucleotides were ligated into either RH6 or KA7 plasmids linearised with BspQI using the following reaction: 1 µL 50-fold diluted annealed gDNA (2 µM), 100 ng digested plasmid, 0.5 µL 10x T4 ligase buffer, 0.5 µL T4 ligase (NEB) and water to 5 µL. Ligation reactions were incubated at 4°C overnight before either storing at -20°C or transformation into *E. coli*.

2.3.2 Plasmid construction using Gene Assembly

Plasmids for the tagging or deletion of genes were generated from either RH6 or KA7 plasmids linearised with MluI and SpeI and PCR amplified marker and gene of interest DNA fragments with 10-20bp overlapping ends, extracted from agarose gels. These fragments were joined together using NEBuilder HiFi DNA Assembly master mix, following manufacturer's instructions, using 50 ng vector with a 1:2 ratio of vector to inserts.

2.3.3 Generation of competent *E. coli*

A single colony of *E. coli* was cultured in 5 mL LB shaking at 37°C overnight. The culture was diluted 1:200 into 100 mL warm LB with 20 mM $MgSO_4$ and shaken at 37°C until OD 0.48. Cells were incubated on ice for 10 minutes before centrifugation at 3,500 rpm, 4°C for 15 minutes. Pellets were resuspended in 40 mL of TFB1 buffer (30 mM KAc, 100 mM $RuCl_2$, 10 mM $Cacl_2$, 50 mM $MnCl_2$, 15 % glycerol and adjusted to pH 5.8 with HAc) per 100 mL culture and incubated on ice for 5 minutes. Cells were centrifuged at 3,000 rpm, 4°C for 10 minutes and pellets resuspended in 4 mL of TFB2 (10 mM MOPS, 10 mM $RuCl_2$, 75 mM $CaCl_2$, 15% glycerol and adjusted to pH 6.5

with KOH) per 100 mL culture and incubated on ice for 15 minutes. 50-200 µL was aliquoted into pre-chilled tubes and stored at -80°C.

2.3.4 Bacterial transformation

50 µL of competent *E. coli* was thawed on ice and incubated with 5 µL ligation mix, 2.5 µL of assembly mix or 0.5 µL of miniprep sample on ice for 30 minutes. Cells were heat-shocked at 42°C for 35 s and incubated on ice for 5 minutes. 1 mL of LB was added, and cells grown at 37°C, 200 rpm for 1 hour before plating onto LB agar with 100 µg/mL ampicillin.

2.3.5 Plasmid miniprep

Single bacterial colonies were grown in 8 mL of LB with 100 µg/mL carbenicillin shaking at 37°C overnight. Cells were pelleted and plasmids were extracted using QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions, with samples eluted in 70 µL warm water. DNA was quantified by NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.3.6 Plasmid linearisation

5 µg of each plasmid for electroporation were linearised by incubation with the appropriate restriction enzymes in 10x cutsmart buffer (NEB) at 37°C overnight. Digested DNA was extracted using Phenol:Chloroform:Isoamyl alcohol 25:24:1 and the aqueous layer collected and combined for pairs of split marker plasmids to be used for the same electroporation reaction. DNA was precipitated with 1/10 volume 3 M NaAc and 3x volume 100% EtOH. DNA pellets were washed in 70% EtOH and air dried before dissolving in 5 µL electroporation buffer.

2.4 DNA protocols

2.4.1 Genomic DNA isolation

Clumps of cells grown on a suitable plate were used for genomic DNA isolation. Cells were resuspended in 500 µL extraction buffer (550 mM Tris-HCl, pH 7.5, 20 mM EDTA, 1% SDS) on ice, disrupted by vortexing for 10 minutes, and incubated at 70°C for 10 minutes. 200 µL 5 M KOAc and 200 µL 5 M NaCl was added, mixed with inversion and incubated on ice for 20 minutes before centrifuged at 13,000 rpm for 20 minutes. 500 µL Phenol:Chloroform:Isoamyl alcohol 25:24:1 was added to the supernatant, mixed by inversion and centrifuged at 13,000 rpm for 10 minutes. 500 µL chloroform was added to the aqueous phase, mixed by inversion and centrifuged at 13,000 rpm for 10 minutes. The aqueous phase was collected, and DNA precipitated with 500 µL isopropanol, mixed with inversion, incubated at room temperature for 20 minutes and centrifuged at 13,000 rpm for 10 minutes. DNA pellets were washed with 100 µL 75% EtOH, centrifuged at 13,000 rpm for 10 minutes and pellets were air dried before being resuspended in 100 µL water.

2.4.2 Annealing oligos

gDNA oligonucleotide pairs were annealed using the following reaction: 5 µL 100 µM forward gDNA oligo, 5 µL 100 µM reverse gDNA oligo, 5 µL 10x NEBuffer 2.1 (NEB) and 35 µL water. Samples were heated to 95°C for 5 minutes and cooled to room temperature over 2 hours before diluting 50-fold.

2.4.3 PCR

DNA fragments for molecular cloning were generated using Q5 High-Fidelity DNA Polymerase (NEB) following manufacturer's instructions, using genomic DNA or plasmids as templates. Primers for PCR reactions are in table 2.2. PCR reactions were performed

using the following thermocycling programme, with the Tm calculated for each primer pair using the NEB Tm Calculator (tmcalculator.neb.com):

Denature: 98°C for 30 s

Denature: 98°C for 15 s

Anneal: Tm for 25 s

Elongation: 72°C for 1 min/kb → 30 cycles

Final elongation: 72°C for 5 minutes

2.4.4 Fusion PCR

PCR amplified gene tags were fused to corresponding PCR amplified genes using Q5 High-Fidelity DNA Polymerase (NEB) following manufacturer's instructions, with initial reactions set up without PCR primers and the following thermocycling programme used with the Tm of the overlap and primers calculated using the NEB Tm calculator:

98°C for 30 s

98°C for 15 s

Tm overlap for 25 s

72°C for 1 min/kb → 5 cycles

PCR primers were then added to the annealed fragment reaction and the following thermocycling programme used:

98°C for 15 s

Tm for 25 s

72°C for 1 min/kb → 25 cycles

72°C for 5 minutes

2.4.5 PCR purification

PCR samples were purified using QIAquick PCR Purification Kit (Qiagen) following manufacturer's instructions, with samples eluted in 30 µL warm water. Samples were quantified using NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.4.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse the presence and size of DNA fragments after PCR amplification of cloning fragments or restriction digestion of plasmids before extraction. 1% agarose gels were prepared in 1x TBE buffer with ethidium bromide to visualise DNA, with all samples loaded with the appropriate amount of 6x loading buffer (NEB). Gels were run in 1x TBE buffer at 80 V. For PCR cloning fragments, all 50 µL of the amplification reaction was loaded across two wells, DNA visualised under UV light and bands of interest excised. For plasmid restriction digests for gDNA insertion, 1 µg DNA was loaded across two wells, DNA visualised under UV light and bands of interest excised. For plasmid screening using restriction digestion, 500 µg DNA was loaded, and visualised using the GelDoc XR system (Bio-Rad).

2.4.7 Gel extraction

DNA fragments were extracted from excised gel using Monarch DNA Gel Extraction Kit (NEB) following manufacturer's instructions, with samples eluted in 15 µL warm water. Samples were quantified using NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.4.8 qPCR

ChIP samples and reverse transcribed RNA were analysed by qPCR performed in 96 well plates with LightCycler 480 SYBR Green master mix (Roche), using the LightCycler

96 System (Roche), using the following reaction: 5 µL 2x LightCycler 480 SYBR Green master mix, 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 1 µL water and 3 µL template DNA or cDNA. DNA from ChIP was diluted 80-fold for inputs and 8-fold for samples and cDNA from reverse transcribed RNA extractions was diluted 5-fold for both positive and negative samples, with all reactions set up in triplicate. Primers for qPCR reactions are in table 2.2, and standard curves were used to determine the efficiency of each primer pair. qPCR reactions were performed using the following programme:

95°C for 2 minutes

95°C for 20 s

55°C for 20 s

72°C for 20 s → 45 cycles

ChIP-qPCR enrichment levels were quantified as IP over input and normalised against *ACT1* expression. RT-qPCR expression levels were multiplied to as per the dilution and normalised against the *ACT1* expression. Gene copy numbers were determined by normalising against the *ACT1* copy number.

2.4.9 Sanger sequencing

Sanger sequencing reactions were generated using BigDye Terminator v3.1 Sequencing kit (Thermo Fisher), using the following reaction: 2 µL 5x BigDye sequencing buffer, 2 µL BigDye Terminator v3.1 reaction mix, 0.6 µL 100 µM primer, 400ng DNA sample and water to make 10 µL. Sequencing reactions were performed using the following thermocycling programme, before being sent to the Edinburgh Genomics sequencing service for analysis:

95°C for 5 minutes

95°C for 30 s

50°C for 20 s

60°C for 4 minutes → 25 cycles

60°C for 1 minute

For Sanger sequencing reactions performed by GeneWiz, 5 µL 10-50 ng/µL PCR extracted DNA and 5 µL 5 µM primer were combined and sent to their sequencing service for analysis.

2.5 RNA protocols

2.5.1 RNA isolation

Total RNA was extracted from 1.5 mL cell culture at mid-log phase with MasterPure Yeast RNA Purification Kit (Epicentre) following manufacturer's instructions. RNA was quantified with NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.5.2 Reverse transcription

DNase treatment reactions were set up using the following: 1 µg total RNA, 1 µL 10x Turbo DNAse buffer, 1 µL Turbo DNAse (Ambion) and water to 10 µL total. Samples were incubated at 37°C for 1 hour. Reverse transcription was performed using the Superscript III Kit (Invitrogen). 2 µL 10 mM dNTPs, 2 µL 100 ng/µL random hexamers and 12 µL water was added to each DNase treated sample and incubated at 65°C for 5 minutes before incubated on ice for 5 minutes. 8 µL 5x Superscript III buffer, 2 µL 0.1 M DTT and 2 µL water were added and each reaction was mixed and split into two 19 µL aliquots. 1 µL Superscript III Reverse Transcriptase enzyme was added to one aliquot to synthesise cDNA (positive sample), with the other aliquot acting as the

control (negative sample). All aliquots underwent the following thermocycling programme:

Annealing 25°C for 5 minutes

Synthesis 50°C for 60 minutes

Termination 70°C for 15 minutes.

2.6 Protein protocols

2.6.1 Protein immunoprecipitation

Protein G Bead Protocol – Clumps of cells grown on a suitable plate were resuspended in 120 µL TN150 (50 mM Tris, pH7.6, 150 µM NaCl, 0.1% NP-40, 5mM β-mercaptoethanol) with proteinase inhibitor (EDTA free, complete, mini). 200 µL Zirconia beads were added, and cells lysed by bead beating twice for 2 minutes. Whilst on ice, 900 µL TN150 was added and samples centrifuged at 14,500 rpm, 4°C for 5 minutes. Supernatant was transferred to a fresh tube and centrifuged at 14,500 rpm, 4°C for 20 minutes and the supernatant collected. 40 µL protein G beads per sample were added to two tubes (one for blocking, one for antibody) and equilibrated by washing in 900 µL TN150 three times. One tube of protein G beads resuspended in 1 mL TN150 buffer and equally split over each protein sample, and incubated on a rotating wheel at 4°C for 20 minutes. 20 µL input samples were taken, the remainder of supernatant transferred, and the other tube of protein G beads was resuspended in 1 mL TN150 and equally split over each protein sample. 1 µL M2 FLAG antibody was added and samples incubated on a rotating wheel at 4°C overnight. Protein G beads were washed three times in 900 µL TN300 buffer (100 mM Tris, pH 7.6, 300 µM NaCl, 0.2% NP-40, 10 mM β-mercaptoethanol). Protein G beads were transferred to a fresh tube in 1 mL of cold PBS and all supernatant removed. 30 µL PLB was added to protein G beans

and 20 µL PLB to inputs before incubation at 95°C for 5 minutes in order to unfold the proteins and add a negative charge ready for analysis via SDS-PAGE and Western blot.

Magnetic Bead Protocol – Either cell cultures grown to OD 1.2-2 and washed in ice cold water and pelleted, or clumps of cells grown on a suitable plate were resuspended in 200 µL Lysis buffer (50 mM Tris, pH 7.8, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% NP-40 with anti-peptidase inhibitor and anti-phosphatase inhibitor). 400 µL Zirkonia beads were added and cells lysed by bead beating for 2 minutes twice, incubating on ice for 5 minutes between rounds. 400 µL lysis buffer with triton X-100 (0.05%) was added and samples centrifuged at 14,000 rpm, 4°C for 5 minutes. Supernatant was transferred and centrifuged twice at 14,000 rpm, 4°C for 20 minutes. 7 µL M2 anti-FLAG magnetic beads per sample were individually equilibrated by washing in 500 µL lysis buffer three times. 10 µL input samples were taken and the remainder of lysate added to the magnetic beads and incubated on a rotating wheel at 4°C for 25 minutes. 10 µL flow-through samples were taken before supernatant discarded and beads were washed three times in 1 mL wash buffer (lysis buffer without the protease and phosphatase inhibitors) by inversion. Proteins were eluted twice by incubation with 50 µL 100 ng/mL 3x FLAG peptide in TBS on a rotating wheel at 4°C for 30 minutes. 200 µL acetone was added, mixed and incubated at -20°C overnight. Samples were centrifuged at 14,000 rpm, 4°C for 20 minutes and pellets dissolved in 10 µL TBS.

2.6.2 ChIP

Cell cultures at OD 1 were fixed in 1% PFA at room temperature for 15 minutes. Fixing was inhibited by incubating with 0.125 M glycine at room temperature for 5 minutes. Cells were washed twice in cold water, pelleting at 3000 rpm, 4°C for 5 minutes and

stored at -80°C. Cell pellets were thawed on ice and resuspended in 500 µL ChIP lysis buffer (50 mM HEPES-KOH pH7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% NaDOC) with 0.1% SDS, proteinase inhibitor (10 µL/mL) and 1 mM PMSF. 500 µl zirconia beads were added and cells lysed with six rounds of 90 s bead beating, incubating on ice for 1 minute every two rounds. Cell lysates were collected by piercing two holes in the bottom of each tube with a 25G needle and placing in another tube for centrifugation at 1,000 rpm, 4°C for 1 minute. Cell lysates were centrifuged at 8,000 rpm, 4°C for 10 minutes, supernatant discarded and pellets resuspended in 350 µL ChIP lysis buffer with 0.1% SDS. Chromatin was sheared using a Bioruptor sonicator with six rounds of 10 minutes with 30 s ON and 30 s OFF on high setting before centrifuging at 14,000 rpm, 4°C for 20 minutes. 30 µL protein G beads per sample were added to two tubes (one for blocking, one for antibody) and equilibrated by washing in 1 mL lysis buffer three times and resuspended in half the beads volume of lysis buffer. 30 µL equilibrated protein G beads were added to each sample and incubated on a rotating wheel at 4°C for 1 hour. Supernatant was collated for samples of the same cell type, made up to 1 mL per IP with lysis buffer, and 10 µL input collected. The remainder of the collated samples were split for each IP (~1 mL) and 30 µL equilibrated protein G beads was added. Either 2 µL H3 antibody (Rabbit polyclonal, Abcam ab1791), 1 µL H3K9me antibody (Mouse monoclonal, (Nakagawachi et al., 2003)) or no antibody for control samples was added and samples incubated on a rotating wheel at 4°C overnight. Protein G beads were sequentially washed in 1 mL lysis buffer, 1 mL high salt lysis buffer (50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% NaDOC), 1 mL wash buffer (10 mM Tris-Cl pH 8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% NaDOC, 1 mM EDTA) and 1 mL TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA), each incubating on a rotating wheel at room temperature for 10 minutes. To reverse the cross-linking and isolate the DNA for qPCR

analysis, 100 µL Chelex-100 resin (Bio-Rad) in water was added to both IP and input samples, and incubated at 100°C for 12 minutes. 2.5 µL 10 mg/mL proteinase K was added to each sample and incubated with shaking at 1,000 rpm at 55°C for 30 minutes, followed by a further 10 minute incubation at 100°C to inactivate proteinase K. 50 µL of the supernatant was collected for each sample for qPCR analysis.

2.6.3 Mass spectrometry

Cell cultures were grown to OD 1.2-2, washed in ice cold water three times, pelleted and stored at -80°C. Cells were resuspended in 200 µL Lysis buffer (50 mM Tris pH 7.8, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% NP-40 with anti-peptidase inhibitor and anti-phosphatase inhibitor) and 5 µL Superasin added. 400 µL Zirkonia beads were added and cells lysed by bead beating for two minutes twice, incubating on ice for five minutes between rounds. 400 µL lysis buffer with triton X-100 (0.05%) was added and samples centrifuged at 14,000 rpm, 4°C for 5 minutes. Supernatant was transferred and centrifuged at 14,000 rpm, 4°C for 20 minutes twice before some samples were passed through a 0.45 µm cellulose acetate filter. 7 µL M2 anti-FLAG magnetic beads per sample were equilibrated by washing three times in 500 µL of lysis buffer. 10 µL input samples were taken, and the remainder of the lysate was incubated with the magnetic beads on a rotating wheel at 4°C for 25 minutes. 10 µL flow through samples were taken and the remainder of the supernatant removed. Beads were washed in 1 mL wash buffer (Lysis buffer without the protease and phosphatase inhibitors) by inversion three times. Proteins were eluted in 50 µL of 0.1 Rapigest in 50 mM Tris:HCl pH 8 twice, incubated at 50 °C for 10 minutes (100 µL sample total). 1M DTT was added to a final concentration of 25 mM and samples incubated at 300 rpm, 95°C for 7 minutes. Samples were cooled to room temperature and urea added to a final concentration of 8 M. Samples were added to a Vivacon 500 spin column 30k (Sartorius) and

centrifuged at 12,000 rpm for 15 minutes. 100 µL 0.05 M IAA in 8M urea in 0.1 M Tris-HCl pH 8.2 was added to each column, shaken at 600 rpm at room temperature for 1 minute and incubated in the dark at room temperature for 20 minutes before centrifuged at 12,000 rpm for 10 minutes. 100 µL 8 M urea in 0.1 M Tris-HCl pH 8.2 was added to each column and centrifuged at 12,000 rpm for 15 minutes. 100 µL 0.05 M ABC was added to each column and centrifuged at 12,000 rpm for 15 minutes. 100 µL 0.05 M ABC and 150 ng trypsin in 0.1% TFA was added to each column and incubated at 37°C overnight. Peptides were eluted by centrifugation at 12,000 rpm for 15 minutes, with remaining peptides eluted in a further 100 µL 0.05 M ABC and centrifuged at 12,000 rpm for 10 minutes. 20 µL 10% TFA was added to stop trypsin digestion. C18 StageTips (Rappsilber et al., 2007) were prepared with 3 layers and equilibrated sequentially with 50 µL MeOH, 50 µL 80% ACN in 0.1% TFA and 50 µL 0.1% TFA. Samples were loaded on the StageTip and washed with 70 µL 0.1% TFA. StageTips with bound peptides were then handed to Dr Christos Spanos for mass spectrometry.

2.7 Data analysis

2.7.1 Genome and protein analysis

The reference genome assembly used is *C. deneoformans* JEC21 ASM9104v1, accessed via EnsemblFungi, using the updated version from April 2018. Both NCBI Basic Local Alignment Search Tool (BLAST) (Altschup et al., 1990) and Conserved Domain Database (CDD) (Lu et al., 2020) were used to identify hypothetical proteins. Identification of GW/WG binding motifs in proteins was through combio.pl/agos/submit web tool.

2.7.2 Fluctuation analysis

Fluctuation analysis was carried out using the FALCOR interface to determine drug resistance rates for whole cultures (B. M. Hall et al., 2009). MSS maximum likelihood method was used, with the 10 independent cultures of each strain grouped together. The upper and lower differences were displayed on graphs as 95% confidence intervals.

2.7.3 Transposon insertion rates

Total and individual transposon insertion rates were calculated by multiplying the overall mutation rate for the strain by the proportion of unique insertions over colonies screened. The 95% confidence intervals were calculated as per Moore et al., 2018. Vassarstats.com was used to determine the 95% confidence limits of a proportion. The following equation was then used to determine the 95% confidence intervals:

$$\delta Q = Q \times \sqrt{\left[(\frac{\delta \mathrm{A}}{\mathrm{A}})^2 + \left(\frac{\delta B}{AB}\right)^2\right]}$$

Q = Rate of a specific event

A = Rate of 5-FOA resistance

B = Proportion of a specific event

2.7.4 T-test

Statistical significance was determined for all ChIP-qPCR and RT-qPCR data by the students t-test, calculated via the mean, standard error of the mean and sample size. This method was chosen due to different repeat sizes between strains. Statistical significance is represented by * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$ in figures.

2.7.5 ANOVA

Statistical significance was determined for all TE qPCR data by ANOVA, where all repeat sizes were consistent. One-way ANOVA was used for all assays looking at differences just between strains. Two-way ANOVA with repeats was used for assays looking at differences between strains and another variable, such as loci or days. HSD Tukey post hoc tests were used to determine individual significant differences, and represented by * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$ in figures.

2.7.6 Mass spectrometry analysis

Mass spectrometry data was normalised based on the intensity of peptides to calculate LFQ values (Rgen Cox et al., 2014). Interpretation of LFQ values was carried out in Perseus. Filtering of the data sets removed contaminants and proteins with less than 1 peptide. Zero values were replaced with normalisation values. T-tests were used to calculate p-values. Significance levels were set at false discovery rate (FDR) = 0.5 and s0 = 1 (Goss Tusher et al., 200 C.E.).

2.7.7 Genoppi

Genoppi, open-source software designed to analyse protein interaction datasets, was used for analysis of sRNA loci of Ago1-IP and Ago2-IP (Pintacuda et al., 2021). Fold change was calculated for each repeat, and Genoppi was used to calculate the p-value for each locus based on the 3 independent datasets used.

Table 2.2 DNA oligonucleotides

Primer name	Primer sequence (5'-3')
Molecular cloning	
C1: CLR4 gDNA fwd	TTGATGGGCAGCATAGTAGTCGC
C2: CLR4 gDNA rev	AACGCGACTACTATGCTGCCCAT
C3: FTH - N-term CLR4 fwd	CATCACCATCACCATGCT CTGCCCATACAAAAAGACGTC
C4: N-term CLR4 - FTH rev	GACGTCTTTTTGTATGGGCAG AGCATGGTGATGGTGATG
C5: Pactin - C-term CLR4 rev	GTGAGTCCGCTCTCTAGA ACTA GGATCC TCATAAAGTTTCGTCAAATGGTATC
C9: FTH - CLR4 5' UTR rev	CGTCGTCATCCTTGTAGTC CATAGTAGTCGCAGGTTTTTTCTG
C10: UpCLR4 - HYG fwd	CAGAAAAAACCTGCGACTACT GTCAAATGCGATGCGTG
C11: HYG - UpCLR4 rev	CACGCATCGCATTTGAC AGTAGTCGCAGGTTTTTTCTG
C12: Pu6 - UpCLR4 fwd	CTTTGTTTTTAGTTCTAATGCAA CGCGT CATCAGTTATTTTTAAGGTGATGTTGGTG
C13: URA5 rev	GGACGAATCGTCATCATCGAC
R48: URA5-split-R	GAGGTAAGAACATCGTCGATGATG
R164:F-FLAG	ATGGACTACAAGGATGACGACGAT
R165:R-HIS	AGC ATGGTGATGGTGATGGTGAGCAG
C14: Check UpCLR4 - HYG fwd	GAAGAGTGCAGAAATGTGTG
C15: UpCLR4 - HYG gDNA fwd	TTGTCGCATTTGACAGTAGTCGC
C16: Check CLR4 - FTH rev	CTCCCTGTAGAAGAAATAGGAG
C17: UpCLR4 - HYG gDNA rev	AACGCGACTACTGTCAAATGCGA
C18: CNH00910 fwd	GAACTTACGCAGCAGAAGAG
C19: DwnCLR4 rev	CTAATTCGCAAAGACACGCAG
R21: HYG check-Nrev	CGCTGTCGAACTTTTCGATC
C35: Pu6 - UpCLR4 fwd	CTTTGTTTTTAGTTCTAATGCAA CGCGT AGACGTGCATTGTATTAGAGG
C36: CLR4 - PHis3 fwd	CAATTAATTCCTGATCCCTTATCTTTCTT GTCAAATGCGATGCGTG

Primer name	Primer sequence (5'-3')
C37: PHis3 - CLR4 rev	CACGCATCGCATTTGAC AAGAAAGATAAGGGATCAGGAATTAATTG
C38: THis3 - CLR4 fwd	GTCTACAAATACTTGTAATCGCGTATAG TGTTTACCCTTGTTATAATTTCGGG
C39: CLR4 - THis3 rev	CCCGAAATTATAACAAGGGTAAACA CTATACGCGATTACAAGTATTTGTAGAC
C40: Pactin - DwnCLR4 rev	GTGAGTCCGCTCTCTAGA ACTA GGATCC GCTGCTCTCAGAGCTC
C47: CLR4 deletion gDNA fwd	TTGAGGGTAAACAAAGAAAGATA
C48: CLR4 deletion gDNA rev	AACTATCTTTCTTTGTTTACCCT
C41: Pu6 - CNH00720 fwd	CTTTGTTTTTAGTTCTAATGCAA CGCGT ATGTCTATAGCAAATTCGTCTTCTTC
C42: CNH00720 - URA5 fwd	CACCAGCTAGAGCCC TCTACTAGAGTTCAGAGTTTGATTG
C43: URA5 - CNH00720 rev	CAATCAAACTCTGAACTCTAGTAGAGGGCTCT AGCTGGTG
C44: URA5 - CNH00720 fwd	GATCGTCTTCAATACCATCC AGAACTGATTAGGGAGGC
C45: CNH00720 - URA5 rev	GCCTCCCTAATCAGTTCT GGATGGTATTGAAGACGATC
C46: Pactin - CNH00720 rev	GTGAGTCCGCTCTCTAGA ACTA GGATCC TTACCATCGACAGATATCCTTCAC
C49: CNH00720 deletion gDNA 1 fwd	TTGGTTACACAGGAGAACTGATT
C50: CNH00720 deletion gDNA 1 rev	AACAATCAGTTCTCCTGTGTAAC
C51: CNH00720 deletion gDNA 2 fwd	TTGGATAACGATGGGGCTCTAGC
C52: CNH00720 deletion gDNA 2 rev	AACGCTAGAGCCCCATCGTTATC
C85: Check CNH00730	CTGAGCCGGCATTAGAATAG
C53: CNH00720 gDNA fwd	TTGCACTGCGCTCTAAAAATGCC
C54: CNH00720 gDNA rev	AACGGCATTTTTAGAGCGCAGTG
C55: FTH - N-term CNH00720 fwd	CATCACCATCACCATGCT CCAGGCGCCAGCTC

Primer name	Primer sequence (5'-3')
C56: N-term CNH00720 - FTH rev	GAGCTGGCGCCTGG AGCATGGTGATGGTGATG
C57: Pactin - C-term CNH00720 rev	GTGAGTCCGCTCTCTAGA ACTA GGATCC TGAAGCAGCCGGTACAG
C58: UpCNH00720 - HYG fwd	CTCTCACTGCGCTCTAAAA GTCAAATGCGATGCGTG
C59: HYG - UpCNH00720 rev	CACGCATCGCATTTGAC TTTTAGAGCGCAGTGAGAG
C60: Pu6 - UpCNH00720 fwd	CTTTGTTTTTAGTTCTAATGCAA CGCGT CTTTCTCCTCGCCTTTCCAG
C61: Check CNH00720 - FTH rev	TAGGTGAGCTTGGAACGTAC
C62: Check CNH00713	CTTGACTTGTCGGCATGTAAC
C63: Check DwnCNH00720	GAAGATGTAGCGGGTGAAATG
C64: UpCNH00720 - HYG gDNA fwd	TTGTAAAAGTCAAATGCGATGCG
C65: UpCNH00720 - HYG gDNA rev	AACCGCATCGCATTTGACTTTTA
C66: FTH - UpCNH00720 rev	CGTCGTCATCCTTGTAGTCCAT TTTTAGAGCGCAGTGAGAG
C102: CNH00720 mid fwd	CAACCAGCAGCGTTTTATACTC
C103: CNH00720 mid rev	TGTCTTAGCAGCTGAGGTG
C104: CNH00720 del repair gDNA fwd	TTGAATACCATCCAGAACTGATT
C105: CNH00720 del repair gDNA rev	AACAATCAGTTCTGGATGGTATT
C106: CNH00720 del repair gDNA 2 fwd	TTGCTCTAGTAGAGGGCTCTAGC
C107: CNH00720 del repair gDNA 2 rev	AACGCTAGAGCCCTCTACTAGAG
C108: CNH00720 start check rev	CTACATCTTCTCCAGCTGGTC
C109: CNH00720 end check fwd	GTTGGTCCCTTGCTTACGAG

Primer name	Primer sequence (5'-3')
C110: CNH00720 middle check fwd	TGCTCCACTTGCTCTTCGAC
C111: CNH00720 middle check rev	ACGTCACAATTTCATCACCATC
C112: CNH00720 middle check fwd	AAGGTCCATCACAATCGCAC
C113: GWO1 gDNA 1 fwd	TTGGCAATGTTACTTTGAACGGA
C114: GWO1 gDNA 1 rev	AACTCCGTTCAAAGTAACATTGC
C115: GWO1 gDNA 2 fwd	TTGTGAGCACCATACTCCCAAGT
C116: GWO1 gDNA 2 rev	AACACTTGGGAGTATGGTGCTCA
C117: PU6 - UpGWO1 fwd	CTTTGTTTTTAGTTCTAATGCAACGCGT AAGAAGAAGAAGAGTTACGAACG
C118: GWO1 - URA5 fwd	CTCCGGTTACTGGCAATGTTAC TCTACTAGAGTTCAGAGTTTGATTG
C119: URA5 - GWO1 rev	CAATCAAACTCTGAACTCTAGTAGA GTAACATTGCCAGTAACCGGAG
C120: URA5 - GWO1 fwd	GATCGTCTTCAATACCATCC ACTCCCAAGTCGGTCG
C121: GWO1 - URA5 rev	CGACCGACTTGGGAGT GGATGGTATTGAAGACGATC
C122: Pactin - GWO1 rev	GTGAGTCCGCTCTCTAGAACTAGGATCC TTAATAACGAGAAATCTCTTGTAGACTAG
C123: Check CNA05015 fwd	GTTGGTAGCATATTTGCGTAGAAC
C124: Check CNA05030 rev	CTAAACCGCAGTACCTCTAC
C125: Check GWO1 mid fwd	CTGCTCGTACATCGTGGAAG
C126: Check GWO1 mid rev	GAACGGACCAGCCAGTG
C149: GWO1 mid fwd	TCTGATGTGGGTGCTGATG
C150:GWO1 mid rev	GAGATGTGTCGCTGCACAG
C151: URA3 inside fwd 2	TGCTTCGTACCTTGCTCTG
C152: URA3 inside rev 2	GTAGTATCACCATGCTTATTGC
qPCR	
C67: Chr2 centromere A fwd	TCTCGCCAAGAAACTGCTGCCA
C68: Chr2 centromere A rev	TGCAAGGGCTCTGCCAGCTT

Primer name	Primer sequence (5'-3')
C73: Ch1 centromere A fwd	TCCGACTCCTCGTAACCGACA
C74: Ch1 centromere A rev	TCCCAGCTGATGGACGGATT
C81: Ch3 centromere B fwd	GGTAAGGCGCCGAGTGTTTG
C82: Ch3 centromere B rev	TGCGAAATTTCCCCACCATC
C86: Ch1 sRNA A fwd	CAATGCGCGGCCTACATCGC
C87: Ch1 sRNA A rev	TGCTGACTGGGGAGCCAAACC
C92: Ch6 sRNA B fwd	AGCGCGTCGGCGATTAGACA
C93: Ch6 sRNA B rev	AGCCAGGGGTCAACGAGAGACT
C96: Ch8 sRNA B fwd	ACGTCCTCAATTTGGGCCCCG
C97: Ch8 sRNA B rev	GGCCTTCGATCTCTGTGCCGT
C100: Ch12 sRNA B fwd	CCCCTGTCGGGCAAAGCCAT
C101: Ch12 sRNA B rev	TCCTTCCCGATGACCTTGAGCCT
C153: T1-1 fwd	TAACTTTGCCAGCGACAGGA
C154: T1-1 rev	CTGTCTGACATCGCATCGGT
C157: T2short-1 fwd	AATTCACTGACCCCCACCGA
C158: T2short-1 rev	GGCTCAGGGAATGACGACAG
C163: T2long-2 fwd	CTGTAAAAGGGGGTTCGCCT
C164: T2long-2 rev	CGCAAACTGCATGTCAACCA
C183: Tcn1-2 fwd	TCCAGACGAAGAACAGACAGC
C184: Tcn1-2 rev	GCTCTTGTTCCTGGCTACCTT
C199: Tcn5-2 fwd	TGGCAGGTAAGCTGGAAGTG
C200: Tcn5-2 rev	AGGTCTTCTCAAACGCGAGG
C203: Tcn6-2 fwd	CACTCCCCCTTGATCCAACC
C204: Tcn6-2 rev	CTTGCCCATGTCCCATCAGT
R37: act1-exon4 1 F	GCACCCCGTTCTCCTCACTG
R38: act1-exon4 1 R	CGTAGAAGGCAGGGGCATTG
Mutation rate assay screening	
C141: URA3 fwd	GATCCCGCTGATTAGTGGAGTCG
C142: URA3 rev	GTATCACCATGCTTATTGCGTATTC
C143: URA3 inside fwd	CGACCACGCTCAAGACGTAC

Primer name	Primer sequence (5'-3')
C144: URA3 inside rev	TACTCAGCCTCTCCTCGTAC
C145: URA6 fwd	TACAACGCTGGTTGGCATTA
C146: URA6 rev	TGCACCAATTTGTTACGATC
C147: URA5 fwd	TTGAATACTCATCACGGCAC
C148: URA5 rev	TAAGACCTCTGAACACCGTAC
R623: URA3fwd3	CCATCATCTCCATCTGCACCT
R621: URA3rev2	CATGACGTTGCACGCAAGTAG
R47: Ura5.F.split	GCTTCTTTTCGGCAACTTTACC
R48: Ura5-split-R	GAGGTAAGAACATCGTCGATGATG

Chapter 3 Investigating the interaction between RNAi and H3K9 methylation

3.1 Introduction

RNA interference has been identified within C. *deneoformans* through the transgene silencing phenomena of SIS and asexual co-suppression, as well as through the identification of genes encoding major components of the RNAi machinery (X. Wang et al., 2010, 2012, 2013). However, the mechanism of silencing the target loci is still unknown. Argonaute proteins bind the siRNA which target the protein and its interactors to the target site, but once here it is unknown how the actual process of silencing occurs.

The RNAi pathway in *S. pombe* mediates silencing through the introduction of H3K9 methylation and subsequent heterochromatin formation at target loci. Once the Argonaute-containing RITS complex is at the target site, directed there by the siRNA, the Clr4 complex, CLRC, is then recruited. Clr4 is a histone methyltransferase which can lay down the epigenetically repressive mark H3K9 methylation (Nakayama et al., 2001). This allows for heterochromatin formation to occur, and prevents the target site from being transcribed, therefore successfully silencing the region. Previous transcriptome and sRNA analyses of *rdp1*Δ strains in C. *deneoformans* and C. *neoformans* have shown that RNAi is linked to silencing of transposons, both retrotransposons and DNA transposons (Janbon et al., 2010; X. Wang et al., 2010).

The interaction between RNAi and H3K9 methylation has been briefly explored in C. *neoformans*, with H3K9 methylation levels at the centromeric retrotransposon Tcn1 found to be comparable in WT and Rdp1 deletion strains (X. Wang et al., 2010). However, in both WT and *rdp1*Δ strains a three-fold decrease in H3K9 methylation was seen during mating, along with corresponding increases in RNA Pol-II loading indicating transcription of Tcn1. This would suggest that retrotransposons expression is regulated by H3K9 methylation, at least during vegetative growth, and not by the RNAi pathway.

However, during mating, although RNA Pol-II loading at Tcn1 elements increases in both WT and *rdp1Δ* strains, mRNA transcript levels increase only in RNAi deficient strains. This suggests that RNAi may silence retrotransposon expression during mating, potentially via SIS, and that the silencing may occur via a mechanism which controls transcript levels but not their production (X. Wang et al., 2010). Even during vegetative growth, siRNAs have been shown to map to transposons, with 22% of small RNA sequencing reads mapping to the centromeres which are rich in retrotransposons (Dumesic et al., 2013). This shows that even during vegetative growth there is still some RNAi activity occurring at the centromeres.

Establishment and maintenance of heterochromatin often requires different processes and factors. In *S. pombe,* maintenance of H3K9 methylation relies on Swi6 recruitment and binding to H3K9 methylation, and methylation marks can be epigenetically maintained during miosis and mitosis (Grewal & Klar, 1996; I. M. Hall et al., 2002). Clr4 is also required for the spreading of heterochromatin, with it binding to H3K9 methylation marks through its chromo domain and then propagating the mark through its catalytically active SET domain. However, as deletion of components from either the RNAi pathway or the Clr4 complex CLRC affect H3K9 methylation, it is unclear which pathway drives the establishment of heterochromatin and if other factors are involved (Volpe et al., 2002). This is also the case in species where H3K9 methylation is generally not related to RNAi, such as *D. melanogaster*. Here the epigenome state is of importance, with Su(var)3-3 required to remove the H3K4 methylation mark before Su(var)3-9 can lay the H3K9 methylation mark (Rudolph et al., 2007).

The establishment and maintenance of DNA methylation has been studied in C. *neoformans.* Dnmt5 has been shown to be the active DNA methyltransferase able to form 5mC, however once DNA methylation is lost it cannot be re-stablished (Catania et

al., 2020). This is due to the loss of a *de novo* DNA methyltranserase, and only after the introduction of DnmtX, a *de novo* DNA methyltransferase, is methylation then able to be re-established. With DnmtX predicted to have been lost ~150-50 mya, Dnmt5 has managed to maintain the 5mC for a remarkably long time, suggesting either selection or preference for the methylated state (Catania et al., 2020). With maintenance of 5mC being so favourable, loss of DnmtX is unsurprising as the many genomic rearrangements between species of the C. *neoformans/gattii* species complex suggest that unrequired genes are lost.

To aide in exploring the endogenous RNAi pathway, sRNA-seq data generated previously in the lab is used to determine RNAi target loci. sRNA-seq data is available for both Ago1-IP and Ago2-IP to distinguish between Ago1-unique sRNAs and Ago2-unique sRNAs. These respective IPs were also carried out in strains with a deletion background to determine if sRNAs rely on the presence of other components in the RNAi, H3K9 methylation or DNA methylation pathways. Total sRNAs-seq data is also available for WT control, but also for other deletion backgrounds.

The aim of this chapter is to explore some of the potential modes of silencing via RNAi. To ensure that the native RNAi pathway is explored, I will focus on regions within the genome which are thought to be targets of RNAi, instead of establishing a faux RNAi response due to transgene introduction as seen in the discovery of SIS and asexual co-suppression. These experiments will help understand the role of RNAi within the cell, and further explore the relationship between histone methylation and RNAi.

3.2 RNAi does not silence through the introduction of H3K9 methylation

C. *deneoformans* has been shown to silence transposons via RNAi (Janbon et al., 2010; X. Wang et al., 2010). A majority of C. *deneoformans* transposons (~80%) are retrotransposons, of which Tcn1-6 all reside within the centromeric regions of the genome (Castanera et al., 2016; Goodwin & Poulter, 2001; Yadav et al., 2018). This coincides with the regions of H3K9 methylation in C. *neoformans,* although the presence or location of H3K9 methylation or location has not yet been confirmed in C. *deneoformans* (Dumesic et al., 2015). Given the apparent link between RNAi and retrotransposon silencing, the correlation in location of the retrotransposons and H3K9 methylation led me to hypothesise that RNAi was silencing retrotransposons via H3K9 methylation and the formation of heterochromatin.

Although it has been suggested that RNAi does not silence target loci through H3K9 methylation, previous research has only looked at Tcn1, just one of several retrotransposons present throughout the centromeres (X. Wang et al., 2010). I decided to explore this link further, looking at other centromeric regions and also sRNA target sites out with the centromere to see if these two coinciding repressive pathways work together.

3.2.1 Identification of the H3K9 histone methyltransferase Clr4

Su(var)3-9 was identified in *Drosophila melanogaster* as a suppressor of position-effect variegation with potential connections to heterochromatin (Tschiersch et al., 1994). A homologous protein within *S. pombe*, Clr4, was identified soon after and both have since be confirmed as H3K9 methyltransferases (Ivanova et al., 1998; Nakayama et al., 2001; Schotta et al., 2002). To identify the Clr4 homolog within C. *deneoformans,* BLAST

searches with both the *S. pombe* Clr4 and *Drosophila melanogaster* Su(var)3-9 were run against the *C. deneoformans* JEC21 genome (Altschup et al., 1990). This identified two potential candidates, CNH00720 and CNH00900; both labelled as putative proteins (Figure 3.1A). CNH00720 is an 1815 amino acid (aa) protein, with a higher homology to both the *S. pombe* and *D. melanogaster* Clr4 proteins than CNH00900. However, CNH00900 is a similar size to both Clr4 homologs at 340 aa in comparison to 490 aa and 635 aa for *S. pombe* Clr4 and *D. melanogaster* Su(var)3-9 respectively. Conserved domain searches of both CNH00720 and CNH00900 showed the presence of a C-terminal SET domain, the fundamental domain for Clr4 proteins that catalyses the methylation of lysine residues on histones (Lu et al., 2020). However, both putative proteins lack the N-terminal chromodomain also present in *S. pombe* Clr4 and *D. melanogaster* Su(var)3-9, which enables the proteins to bind methylated lysine residues and thus associate with heterochromatic regions for the maintenance and spreading of heterochromatin. As the absence of a chromodomain could be compensated for via interaction with another chromodomain-containing protein, both were still seen as possible candidates due to the presence of the SET domain, essential for their role as a histone methyltransferase.

The conserved domain searches also reveal other domains within the compared proteins. Su(var)3-9 shares part homology with the eukaryotic translation initiation factor 2, gamma subunit (eIF2γ). However, this is due to the overlapping of two antisense genes – the C-terminal of Su(var)3-9 where this homology lies and the first exon of eIF2γ protein. Therefore, although this domain is recognised in Su(var)3-9, it is unlikely to be contributing to its role as a histone methyltransferase. *S. pombe* Clr4 shares homology with endonuclease 4-like protein in between its chromo and SET domains and CNH00720 is predicted to have region of homology with the tegument protein UL36

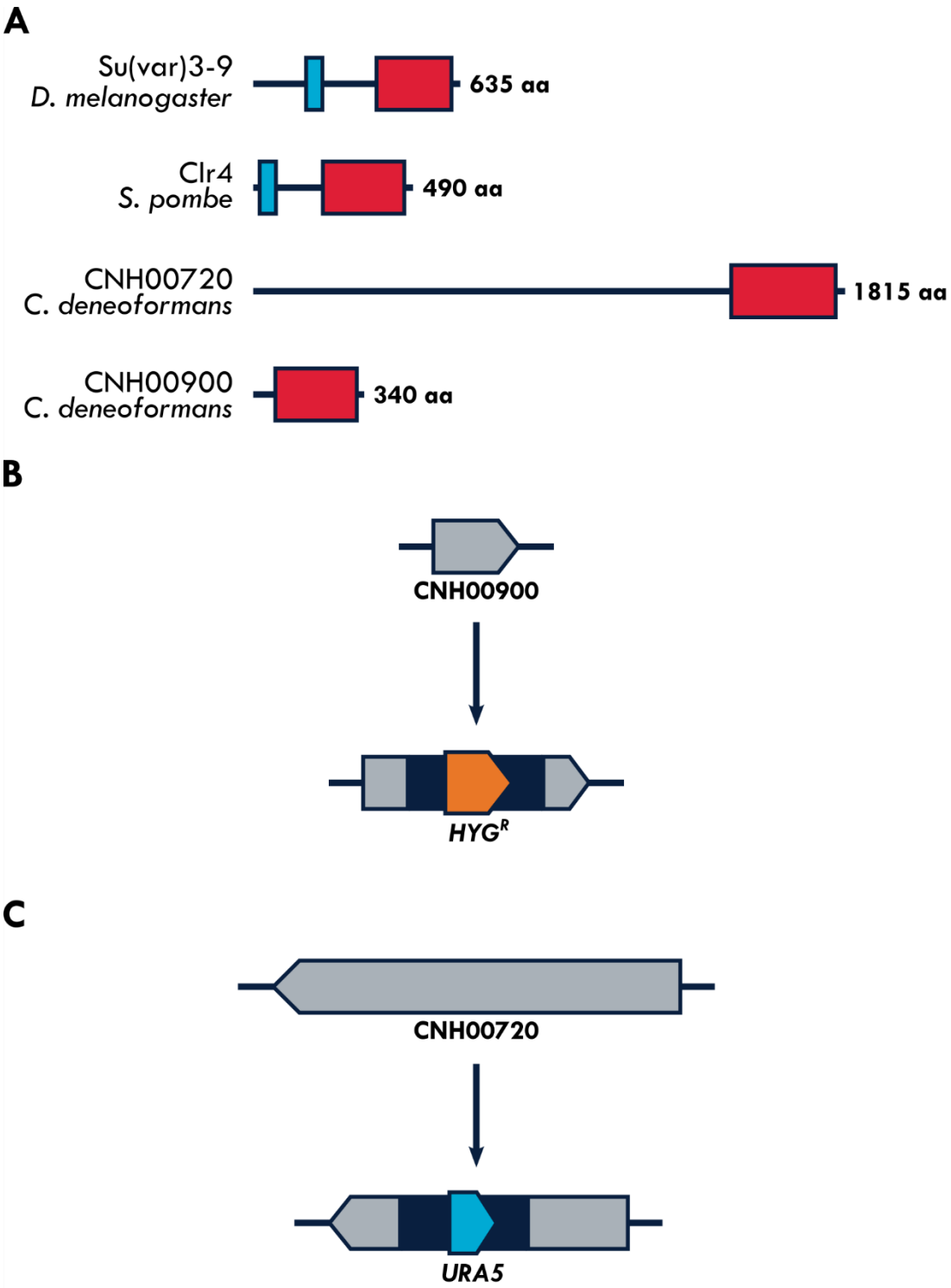

Figure 3.1 Identification of potential Clr4 homologs
(A) Diagram showing structure of potential of Clr4 homologs in C. *deneoformans* against *D. melanogaster* Su(var)3-9 and *S. pombe* Clr4. Domains are as followed: Red = SET, Blue = Chromo **(B)** Schematic of CNH00900 disruption with *HYG*^{R} insertion. **(C)** Schematic of CNH00720 disruption deletion with *URA5* insertion. For **(B)** and **(C)** arrowheads indicate direction of gene insertion relative to chromosome.

within the C-terminal region, although both of these have low probability and with no confirmed function of either regions. The other possible candidate, CNH00900, only contains the SET domain which covers almost the entire length of the protein. Interestingly, the other candidate gene CNH00720 was found to overlap the final three exons of the gene with the final exon of the antisense gene CNH00730.

With both candidates having different merits as the possible ortholog of Clr4, CNH00720 having higher homology but CNH00900 having a similar length, I set about making deletion strains for each candidate. These were made using the combined split-marker suicide CRISPR method (Kim et al., 2009; Y. Wang et al., 2016), whereby the gene of interest was disrupted via insertion of either a hygromycin resistance (*HYG^R*) or *URA5* marker gene. With the different lengths of candidate genes, the disruption of CNH00900, being shorter, resulted in a straight disruption insertion in the middle of the gene, whilst CNH00720 being considerably longer meant that a deletion disruption was formed, with ~1200 bp of the GOI either side of the marker and the remaining ~3700 bp removed (Figure 3.1B and C).

To determine which candidate is the functional Clr4 homolog, H3K9me2 ChIP-qPCR was carried out to assess H3K9 methylation levels at several centromeric regions in wild-type and mutant cells. qPCR primers used correspond to centromeric regions on chromosomes 1, 2 and 3 and vary in their distance into the centromeres allowing the methylation levels to be assessed across different loci and chromosomes in case of variation. The site chosen on chromosome 2, primer pair B, corresponds to a region of mapped sRNA reads as identified through the work of other lab members (Figure 3.2A). These sRNA reads are unique to the Ago2-IP compared against Ago1-IP, and is still detected in Ago2-IP when either Dcr1, Dcr2 or Clr4 proteins are deleted. Deletion of Dnmt5 causes a slight reduction in the number of reads detected as well as their

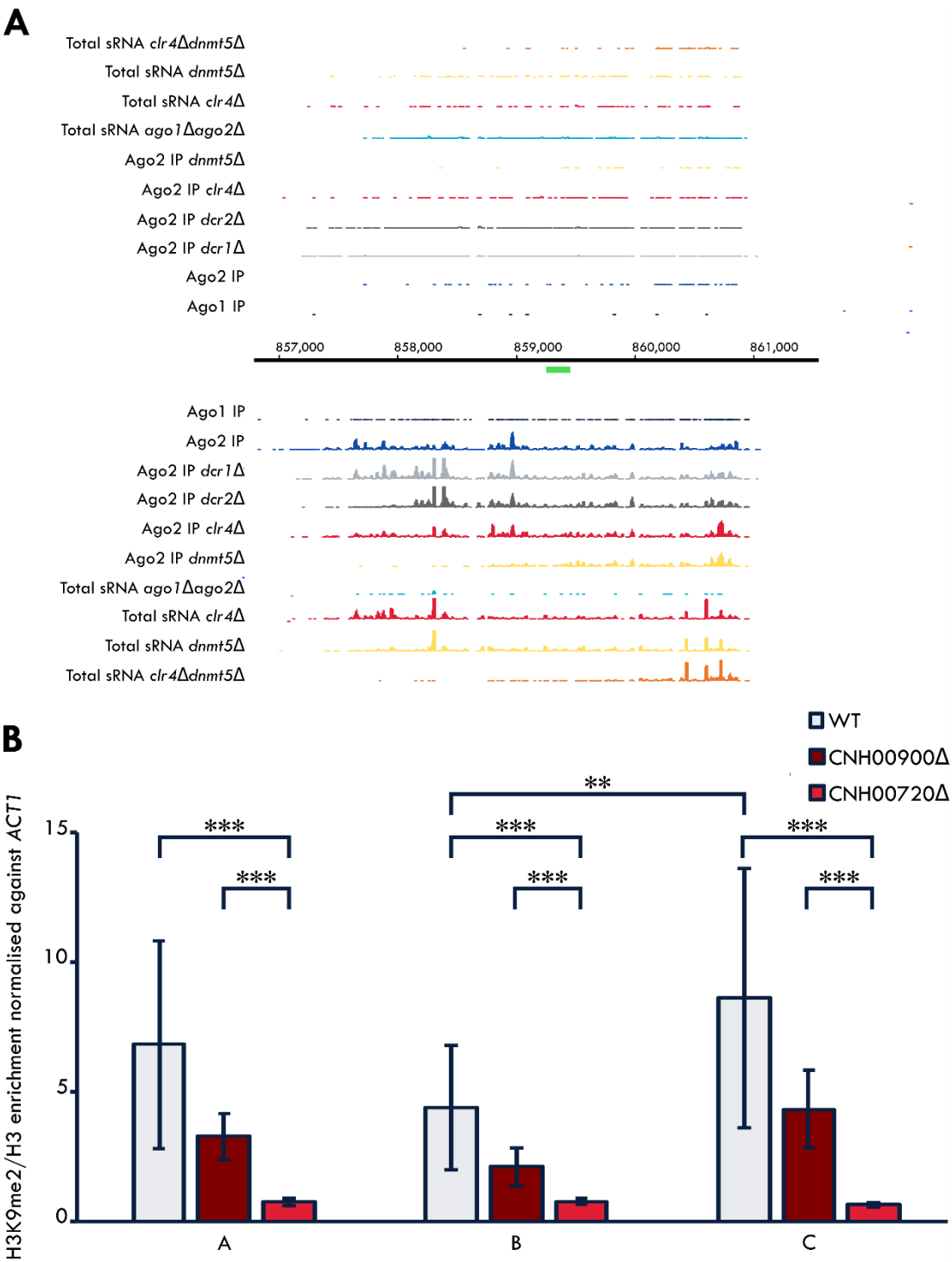

Figure 3.2 CNH00720 is the H3K9 methyltransferase

(A) Plot showing the intensity of small RNA reads mapping to the region on centromere where qPCR primer pair B (green bar) are located, in the indicated strains. **(B)** Graph showing H3K9me2 levels at three different centromeric sites in WT and deletion strains CNH00900Δ and CNH00720Δ, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. n=14 for WT, n=3 for CNH00900Δ and n=8 for CNH00720Δ. WT is WSA1226 strain.

distribution over the area, which is further altered in the total sRNA reads of a double Clr4 and Dnmt5 deletion strain. This pattern of sRNA reads is uncommon within the centromeres, where sRNA reads are only usually seen in Dnmt5 single or Clr4 Dnmt5 double deletion strains. The presence of Ago2 specific reads suggests that this region may differ in silencing regulation from majority of the centromeres. Conserved domain finder analysis of this region identified a reverse transcriptase, which when compared against the previously identified locations of retrotransposons in the centromeres of C. *deneoformans*, matches a region corresponding with Tcn6 (Yadav et al., 2018). Primer pairs A and C correspond to predicted Tcn1 and Tcn3 regions respectively, though without the presence of sRNA reads in these IPs. H3K9me2-ChIP-qPCR in WT cells showed the presence of H3K9me2 at all centromeric regions tested, with levels up to 8-fold higher than at the non-centromeric actin *ACT1* negative control locus at region C. Comparison of the three regions shows no statistical significance in the difference in the levels between regions A and C, however statistical significance is seen between regions B and C ($p=0.0084$) although the difference between A and B is not quite significant ($p=0.0606$). When comparing WT with both potential Clr4 candidate deletion strains, CNH00900Δ strains have levels of H3K9 methylation similar to WT at all centromeric regions tested, whereas H3K9me2 was lost in the CNH00720Δ strain at these regions, with significantly less H3K9me2 at all regions when compared with both the WT and CNH00900Δ strains (Figure 3.2B). This indicates that CNH00720 encodes the functional Clr4 H3K9 methyltransferase within C. *deneoformans,* and all further mentions of Clr4 within this species will be referring to this gene. This also confirms the presence of H3K9 methylation at the centromeres in C. *deneoformans,* which up until now has only been confirmed in the related species C. *neoformans*.

3.2.2 RNAi deficient strains retain H3K9 methylation

To determine if H3K9 methylation is linked to the RNAi pathway, I used two RNAi deficient strains made previously in the lab to compare methylation levels. Although RNAi in C. *deneoformans* has been shown to primarily involve Ago1 and Dcr2, as there are two version of both of these components used strains where both homologs of the genes were deleted to ensure full loss of RNAi (Janbon et al., 2010). In the case of the double *dcr1dcr2*Δ deletion strain, the two genes are disrupted through the insertion of just one marker gene due to them being neighbouring genes, however in the *ago1*Δ*ago2*Δ strain the two genes are disrupted with two with separate markers due to them being spatially separated, although still on the same chromosome. If RNAi silences target loci through Clr4 recruitment and subsequent H3K9 methylation, it was predicted that deleting some of the core RNAi machinery, resulting in a lack of RNAi, would also stop Clr4 recruitment and histone methylation.

H3K9me2-ChIP-qPCR was carried out with both RNAi deficient strains to assess levels of H3K9me2 at the previously tested centromeric regions. The levels of H3K9me2 in *dcr1dcr2*Δ cells are comparable to WT at all three regions analysed. Surprisingly, *ago1*Δ*ago2*Δ strains show significantly higher levels of H3K9me2 when compared with WT (Figure 3.3A). However, there is no significant difference in methylation levels between *dcr1dcr2*Δ and *ago1*Δ*ago2*Δ strains, likely due to the large variation between samples tested, so further repeats (n>10 for all strains) would be required before a full conclusion should be made. However, both RNAi deficient strains have significantly higher H3K9me2 levels than those seen in the c*lr4*Δ strains, showing that H3K9 methylation remains intact even when RNAi is lost.

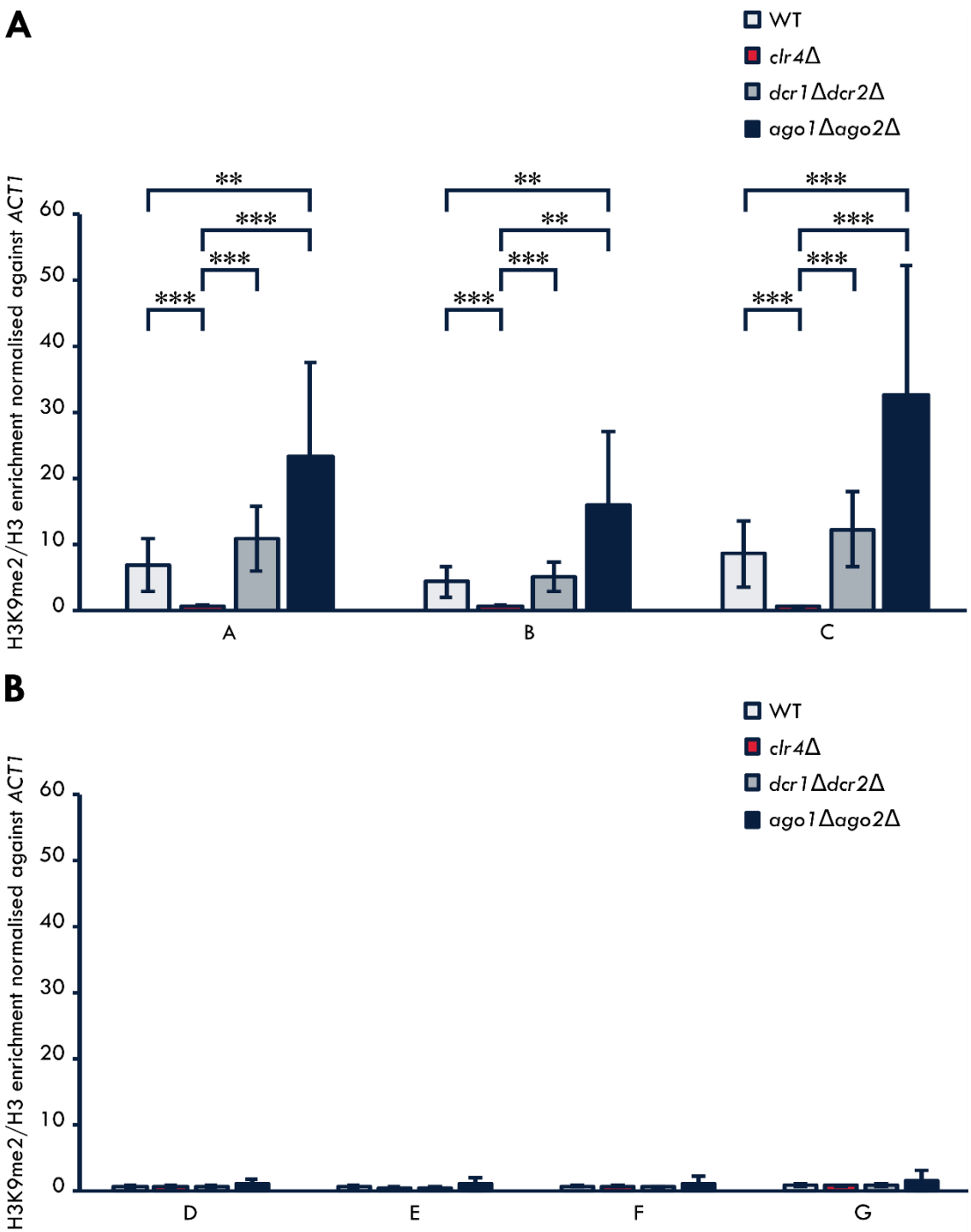

Figure 3.3 RNAi deficient strains have H3K9 methylation

(A) Graph showing H3K9me2 levels at three different centromeric sites in WT and deletion strains *clr4*Δ, *dcr1*Δ*dcr2*Δ and *ago1*Δ*ago2*Δ, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. WT is WSA1226 strain. **(B)** Graph showing H3K9me2 levels at four non-centromeric small RNA target sites in WT and deletion strains *clr4*Δ, *dcr1*Δ*dcr2*Δ and *ago1*Δ*ago2*Δ, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. WT is WSA1226 strain. For both graphs, n=14 for WT, n=8 for *clr4*Δ, n=4 for *dcr1*Δ*dcr2*Δ and n=3 for *ago1*Δ*ago2*Δ.

In C. *neoformans* H3K9 methylation has been shown to be primarily at the centromeres, with small domains also occurring at sub-telomeres. However, sRNA sequencing data from our lab shows that Ago1 and Ago2 associated sRNAs mainly coinciding with regions out with the centromeres. As these are predicted RNAi target sites, I wanted to check the H3K9 methylation levels at these short loci, in case there are small islands of H3K9 methylation throughout the coding regions of each chromosome, and if so if they are linked to RNAi. Four regions with sRNAs were identified (D-G), chosen for having high levels of corresponding sRNAs that were associated with Ago1 or Ago2 in immunoprecipitation experiments (Figures 3.4 and 3.5).

Region D is in the gene CNA02205, a hypothetical protein with no predicted conserved domains, with a BLAST search only showing homology to other hypothetical proteins within various *Cryptococcus* species. The sRNAs are predominantly antisense to the gene and their accumulation requires the presence of both Argonaute proteins as they are seen in both Ago1 and Ago2 IP, although to a greater extent in Ago2. Detection of these sRNAs in Ago1-IPs is dependent on the presence of both Dicer proteins, whereas detection in Ago2-IPs is primarily reliant on the presence of Dcr2. Region E lies immediately downstream of the gene CNF00360, a hypothetical protein with homology to a Atg22 autophagy-related transporter protein. Here there are sRNAs that are present in Ago2-IPs and not Ago1-IPs, and can be produced by either Dcr1 or Dcr2. sRNA reads corresponding to region F cover the gene CNH03540, which contains conserved domains corresponding to the retrotransposon Gag protein, and integrase core domain. Due to the close proximity of the Gag and Int domains, and the absence of reverse transcriptase and protease domains, it is likely that this is an incomplete retrotransposon. Here the sRNA reads from the Ago1-IP cover the C-terminal region where the integrase core domain resides, and the sRNA reads from the Ago2-IP cover

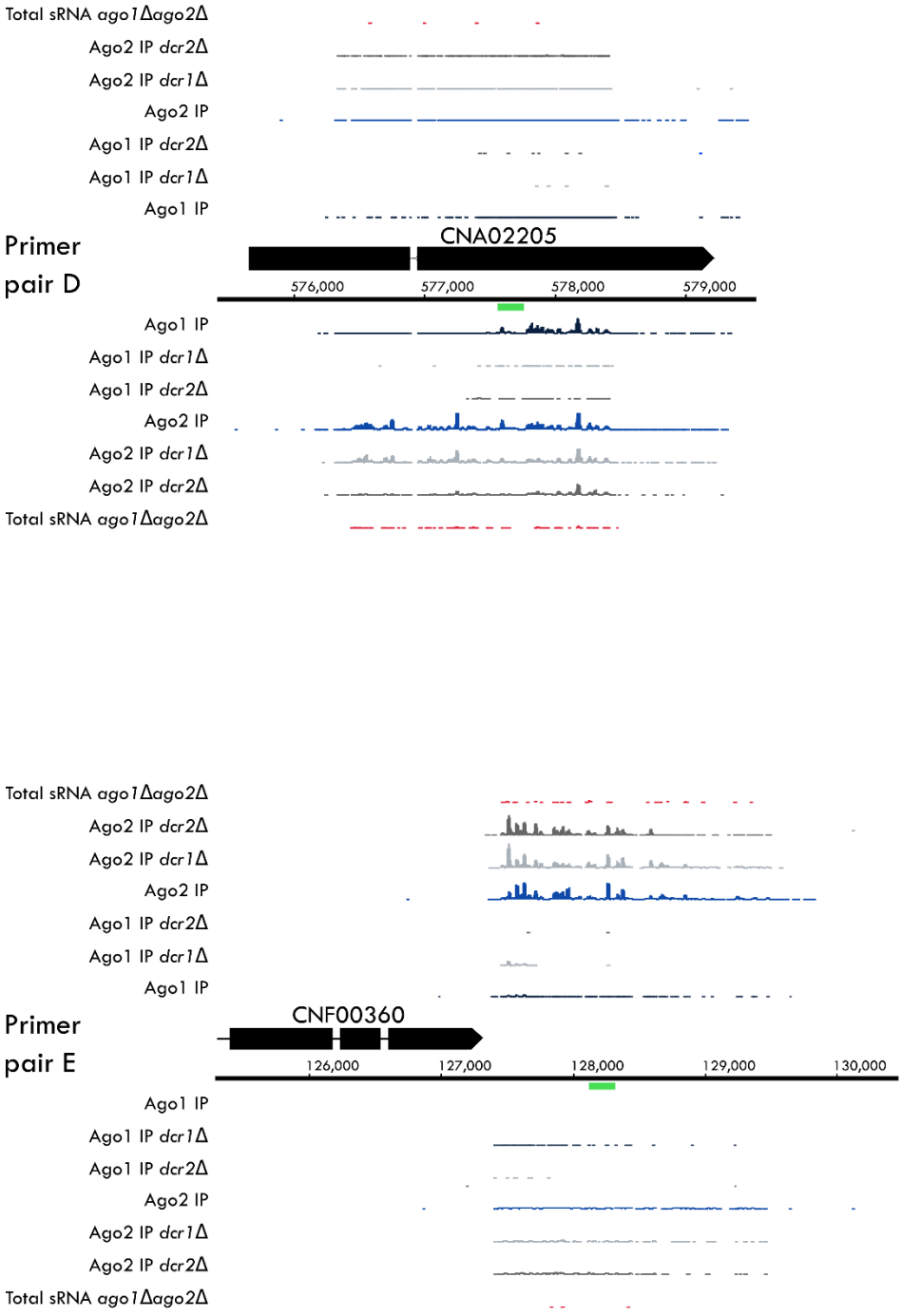

Figure 3.4 Features of non-centromeric sRNA target regions D and E
Plots showing the intensity of small RNA reads mapping to non-centromeric sRNA target regions D and E, in the indicated strains. Positions of the qPCR primers are indicated by the green band.

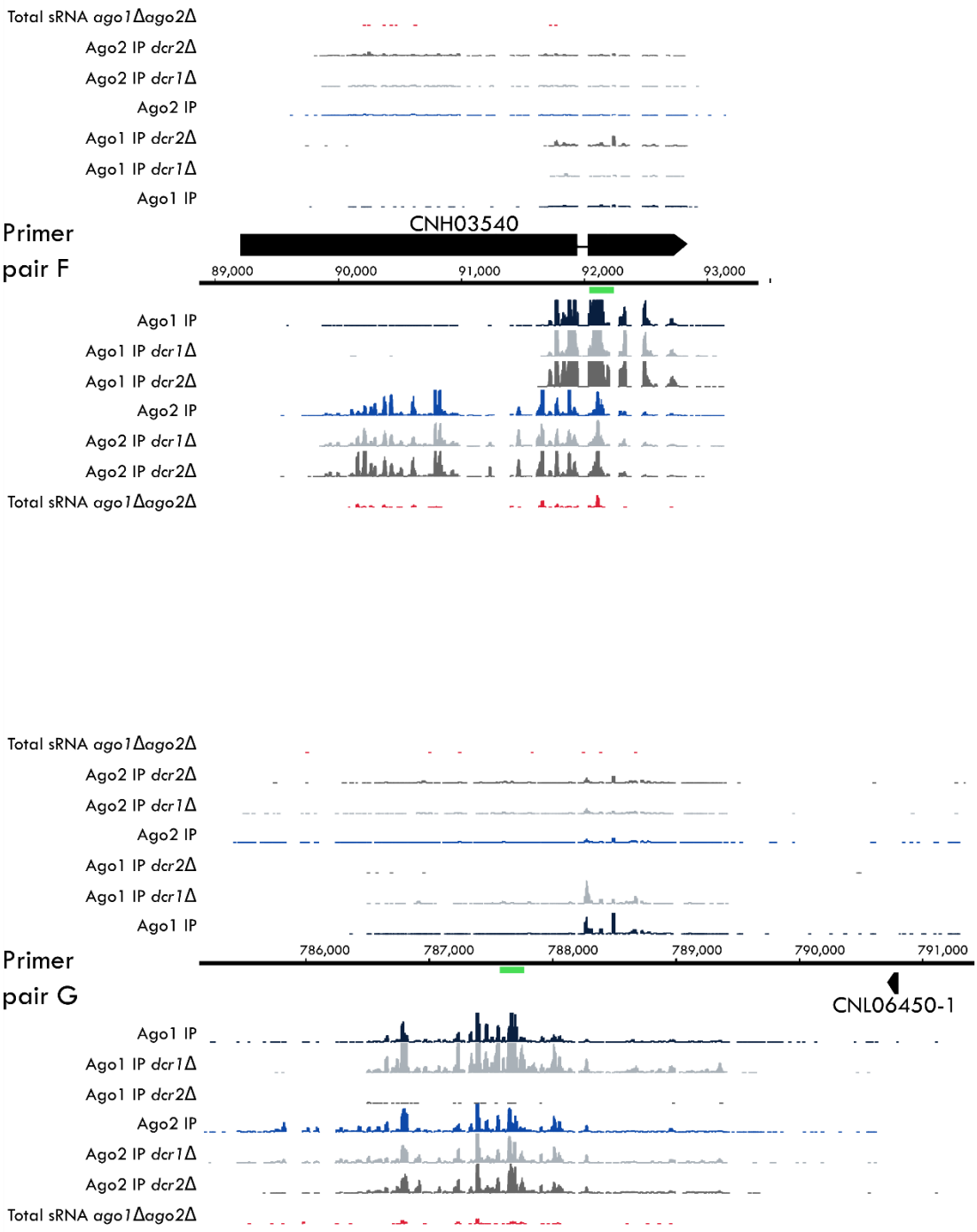

Figure 3.5 Features of non-centromeric sRNA target regions F and G
Plots showing the intensity of small RNA reads mapping to non-centromeric sRNA target regions F and G, in the indicated strains. Positions of the qPCR primers are indicated by the green band.

the N-terminal region where the Gag protein domain resides. sRNAs in bit regions are unaffected by deletion of either Dcr1 or Dr2, and the primer pair used lie in region of Ago1 associated sRNAs. The final region G lies between genes, with the closest gene being CNL06450-1 which produces a tRNA. sRNA reads from this region are present in both Ago1 and Ago2 IPs, although only the sRNAs associated with Ago1 are dependent on the presence of Dcr2. Comparing all these regions against total sRNAs of other related strains (not shown), shows that total sRNA of *dcr1dcr2*Δ strains show no reads present for these four regions, similar to *ago1Δago2*Δ total sRNA. sRNAs are largely unaffected by the loss of DNA methylation and H3K9 methylation, with the exception of sRNAs from region G which are absent in a *clr4*Δ*dnmt5*Δ strain. Comparing these total sRNAs confirms that these reads are RNAi specific, as require the presence of at least one Argonaute and one Dicer protein to be present, and only one site relies on the presence of H3K9 or DNA methylation for sRNA production.

ChIP-qPCR analysis of H3K9me2 levels at these four non-centromeric sRNA target regions showed no enrichment of H3K9me2 in WT relative to *clr4*Δ, with both RNAi deficient strains showing comparable background levels of H3K9me2 (Figure 3.3B). This shows that these sRNA target regions are not silenced by H3K9 methylation, even in region G where the sRNA accumulation relies on the presence of Clr4 and Dnmt5. From this data I conclude that in C. *deneoformans* RNAi silences through a H3K9 methylation independent pathway, at both centromeric and non-centromeric target sites.

3.3 H3K9 methylation can be re-established in strains which have lost H3K9 methylation

Although I did not find a link between RNAi and H3K9 methylation in C. *deneoformans*, I was still interested in the role which this epigenetic modification plays within the cell. Establishment and maintenance of histone methylation hasn't been looked at in C. *deneoformans,* although the establishment and maintenance of DNA methylation has been studied in C. *neoformans* (Catania et al., 2020). Based on this study, I hypothesised that H3K9 methylation would also be equally selected or favoured with Clr4 maintaining the mark over generations, and that any establishment factors would have been lost from the genome.

To determine if H3K9 methylation could be re-established after having been lost, I initially carried out H3K9me2 ChIP on strains where the *CLR4* gene is under the promoter of the *GAL7* gene. These were created in the process of tagging Clr4 with an N-terminal FTH tag, where the intermediate step involves the insertion of the FTH tag, *pGAL7* and selection markers *URA5* and HYG^R upstream of the start codon via CRISPR (Figure 3.6A). *pGAL7* promoters are active when cells are grown on galactose medium, allowing expression of the gene of interest. However, growth on glucose medium results in repression of the gene of interest. For this experiment, I grew the *pGAL7-FTH-CLR4* strains for 1 week on GAL medium and collected samples for ChIP. I then transferred these cultures to GLU medium for 1 week and collected samples for ChIP, before transferring them back to GAL medium for 1 week and collecting the final ChIP samples. This allowed for direct comparison between the different media without having to factor in any differences in methylation levels between the independent biological replicates, if there was any. H3K9me2 ChIP-qPCR showed that H3K9 methylation levels at the

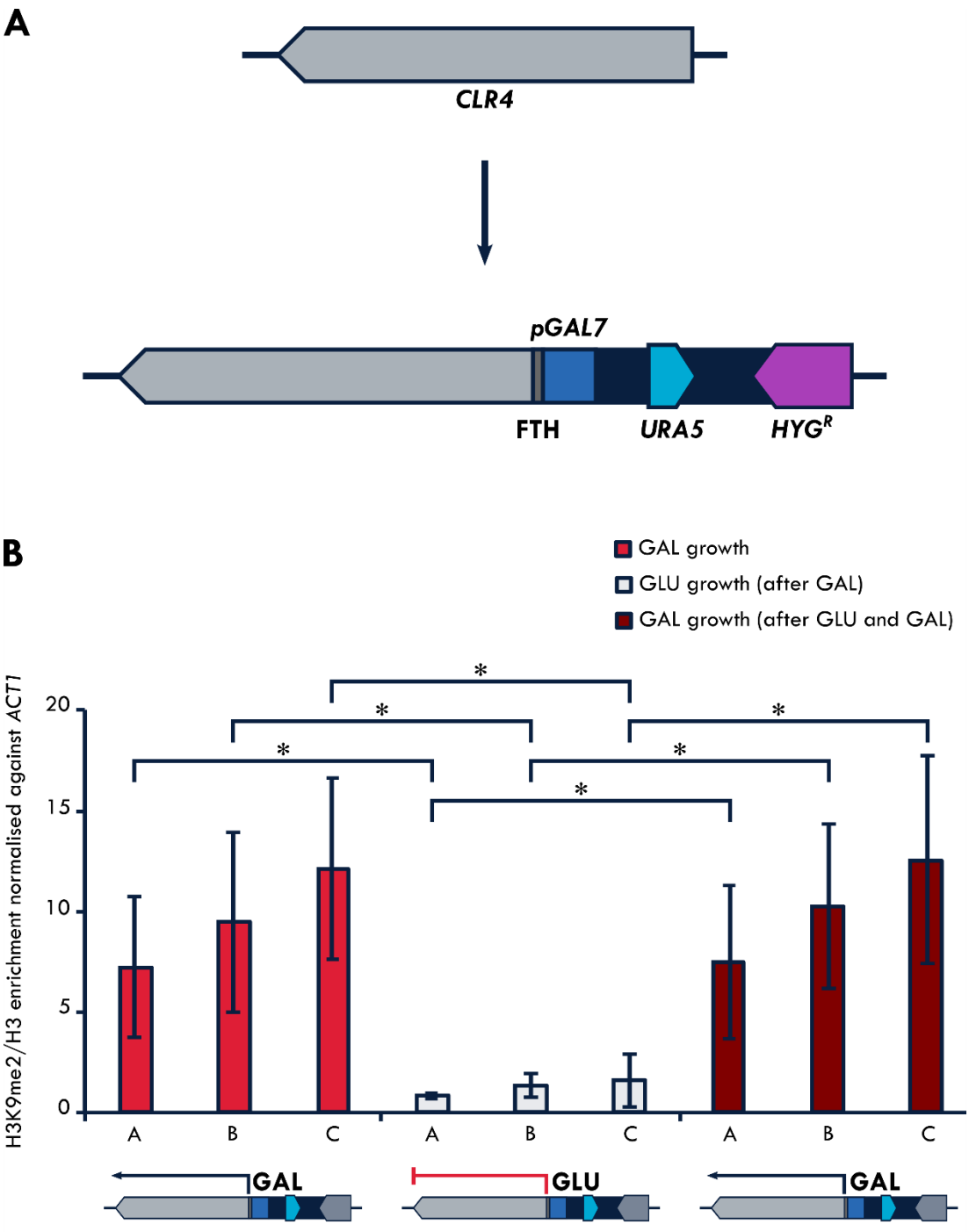

Figure 3.6 H3K9me2 is re-established after Clr4 is switched off and on again
(A) Schematic of Clr4 being placed under a *GAL7* promoter with the addition of a FTH (Flag, Tev Cleavage site and 6xHis) tag and two seletion marker genes, *URA5* and *HYG*^{R}. **(B)** Graph showing H3K9me2 levels at three centromeric sites in strains in which Clr4 is expressed under a GAL7 promoter, normalised against *ACT1*. Samples for ChIP were collected after 1 week growth on GAL, then a further 1 week growth on GLU and 1 week growth on GAL in order to switch on, off and back on Clr4 transcription. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. n=3 for all.

three centromeric regions significantly decreased during growth on GLU, showing repression of Clr4 expression (Figure 3.6B). The levels of methylation then significantly increased again after growth on GAL, to levels comparable with the initial GAL growth.

As it is possible that growth on GLU does not fully supress Clr4 expression, and that residual amounts of H3K9 methylation remain which Clr4 can then amplify to WT levels again, I confirmed this result using deletion and re-introduction. Taking the deletion disruption strain of Clr4, I removed the *URA5* marker and re-inserted the full-length gene using CRISPR (Figure 3.7A). Sequencing of the full length of the gene showed successful re-insertion of Clr4 in two independent strains which were comparable and the data for each was combined. H3K9me2 ChIP-qPCR showed that the re-introduction strain, RI-C*LR4*, has comparable levels of methylation to WT, and significantly higher methylation levels than in c*lr4*Δ (Figure 3.7B). This confirms that H3K9 methylation can be re-established after its loss.

3.4 RNAi target sites do not show an increase in transcript levels

After determining that RNAi does not silence target loci through H3K9 methylation, I started to look at other possible mechanisms of silencing. Most RNAi pathways silence their targets via a post-transcriptional gene silencing (PTGS) mechanism. Whereas histone methylation prevents transcription, in PTGS transcription occurs but the transcripts are unable to complete the pathway to functional proteins. The most common mechanism is through degradation of transcripts, as seen during quelling in *N. crassa* (Catalanotto et al., 2002). To assess whether a similar mechanism may operate in C. *deneoformans*, I decided to look at changes in transcript levels in RNAi deficient strains. I hypothesised that target transcript levels would increase in RNAi deficient strains

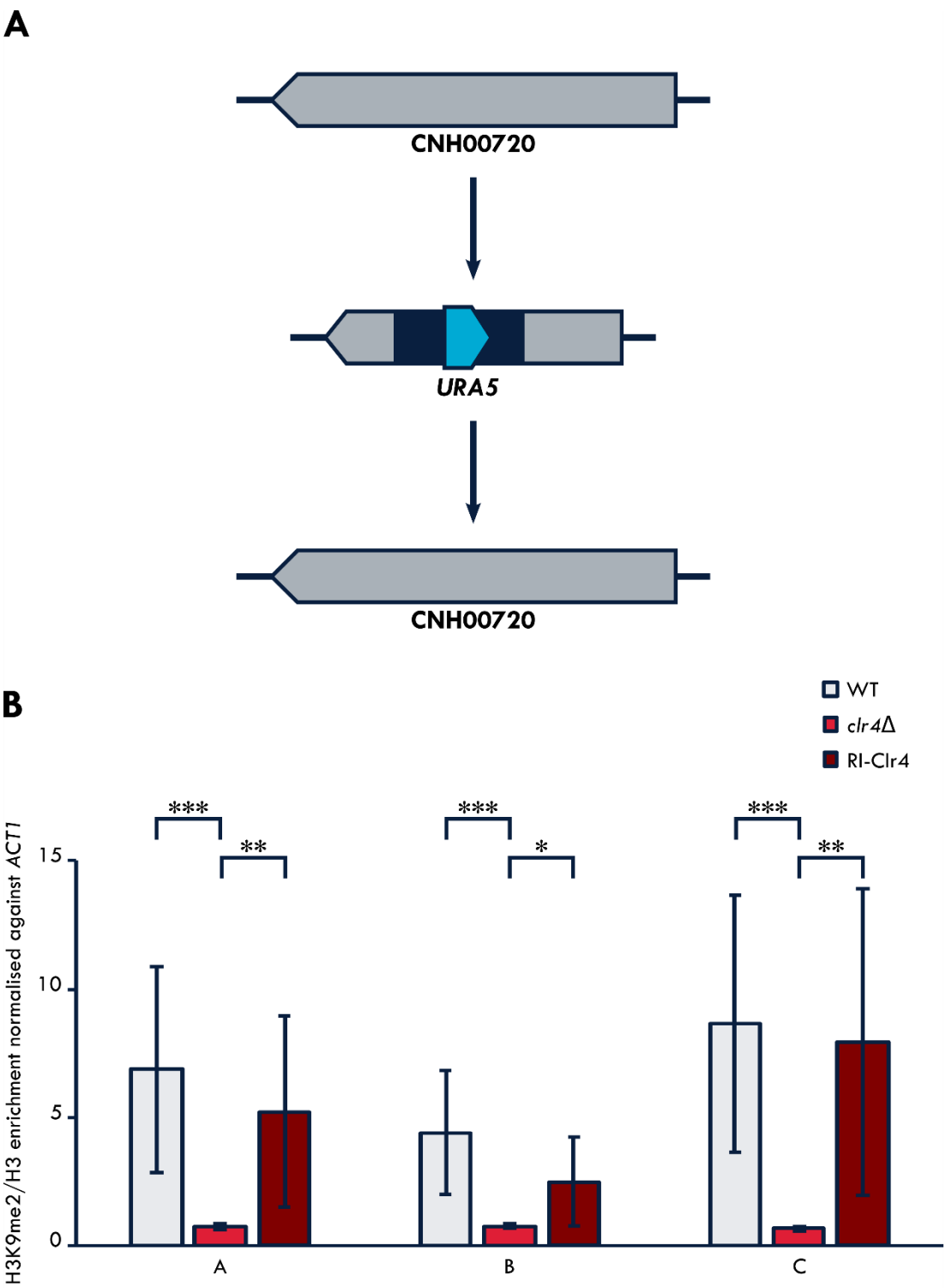

Figure 3.7 H3K9me2 is re-established after Clr4 is deleted and re-introduced
(A) Schematic of Clr4 disruption deletion with *URA5* insertion, followed by removal of the URA5 and replacement with full-length Clr4 **(B)** Graph showing H3K9me2 levels at three centromeric sites in WT, *clr4*Δ and RI-Clr4 strains, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. WT is WSA1226 strain. n=14 for WT, n=8 for *clr4*Δ and n=4 for RI-Clr4 from two independent strains.

compared with WT, as when a functional RNAi pathway is present the mRNA transcripts would be targeted for degradation. Without the presence of RNAi, these transcripts would no longer be degraded, causing increases in their levels.

Four strains were selected for analysis: WT, two RNAi deficient strains (*dcr1*Δ*dc2*Δ and *rdp1*Δ, expected to give similar results), and an H3K9 methylation and DNA methylation deficient strain (*clr4*Δ*dnmt5*Δ), since accumulation of siRNAs from some loci was found to be reduced in this background. RNA extraction and RT-qPCR was performed to analyse levels of transcripts from five genomic regions associated with abundant sRNA reads, as identified earlier in the chapter – one centromeric (B) and 4 non-centromeric (D-G). Comparison of the transcript levels in WT and mutant strains showed no significant difference between all four of the strains tested at each site tested (Figure 3.8). There was also no significant difference in transcript levels between different loci in WT cells. However, in some genetic backgrounds significant differences are seen in the levels of transcripts between different loci – locus E is significantly different to B, F and G in *dcr1dcr2*Δ, and locus E is significantly different to F and G in *clr4*Δ*dnmt5*Δ. This suggests variation in the transcript levels between the regions, though significant differences are only seen in samples with a small standard deviation, so further repeats (n=10) would be required to determine if they are seen in WT or if these differences are unique to these deletion strains. The standard deviation for *rdp1*Δ is very large, suggesting large differences in the transcript levels between independent cultures. Increasing the sample size (n>3) would help confirm the result.

The data presented here shows no significant increase in transcript levels in RNAi deficient strains. That transcript levels are not altered during the loss of RNAi suggests

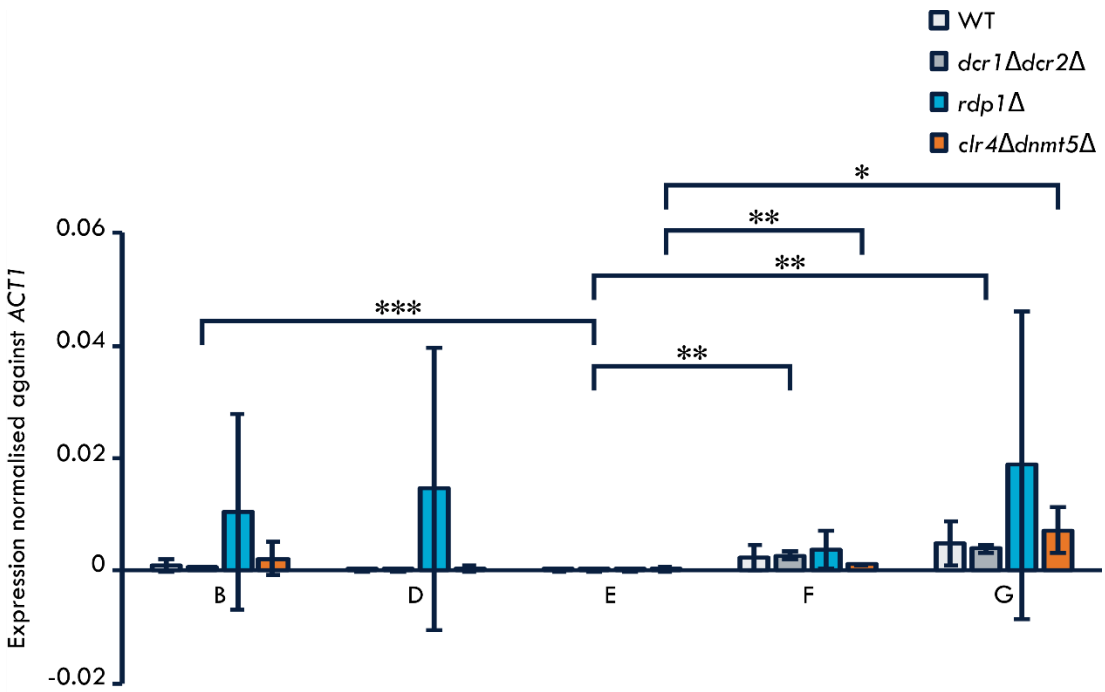

Figure 3.8 Transcript levels in RNAi deficient strains are comparable to WT
(**A**) Graph showing expression levels at five sRNA target sites in WT, *dcr1Δdcr2Δ*, *rdp1Δ* and *clr4Δdnmt5Δ* strains, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05, ** = p≤0.01, *** = p≤0.001. WT is WSA1226 strain. n=3 for all.

that an alternate PTGS mechanism other than RNA degradation is the mode of silencing at these regions.

3.5 Discussion

This chapter aimed to further understand the silencing mechanism of RNAi within C. *deneoformans*. Although the method of silencing has not been elucidated, several options have been ruled out.

Initial work focused on the role of Clr4 within C. *deneoformans*. Identification of gene CNH00720, which contains a SET domain but not a chromo domain, as the functional H3K9 methyltransferase in C. *deneoformans* showed that the presence of a chromo domain is not required for functionality. This differs from *S. pombe,* where the function of Clr4 is dependent on both the chromo and SET domains (Ivanova et al., 1998). The Clr4 homolog in C. *deneoformans* is a bigger protein, approximately 4x the length of the *S. pombe* homolog and 3x the length of the *D. melanogaster* homolog. Without the chromodomain, another means of localising Clr4 to H3K9 methylated regions is required, and it is possible that this extended region not present in the chromodomain-containing homologs is required for localisation. Initially, I had planned to carry out Clr4 immunoprecipitation and mass spectrometry to identify interacting partners, however after tagging *CLR4* with FTH at the N-terminus I was unable to detect the protein via Western blot, even after optimisation of the protocol. Without being able to visualise the protein, it would be unlikely that enough Clr4 was precipitated for mass spectrometry. Tagging of the other candidate SET domain protein CNH00900 was successful, however after confirmation that it was not the Clr4 homolog, no further experiments were carried out with it. The presence of the SET domain in this protein

suggests that it functions as a lysine methyltransferase at another residue position. As both of the candidate genes are closely positioned on chromosome 8, a chromosome which underwent genomic rearrangement with chromosome 12 in the creation of the JEC21 strain resulting in the duplication of some genes (Loftus et al., 2005; Sun & Xu, 2009), I was interested to see if CNH00720 and CNH00900 were a result of this duplication. However, homologs of both CNH00900 and CNH00720 are present in the parental strain B-3015, as well as C. *neoformans* strains, suggesting that if they are a result of a duplication event it would have occurred before these two species diverged.

H3K9 methylation was confirmed at the centromeres, although full ChIP-seq would be required to determine if the genome-wide methylation pattern is similar to that seen in C. *neoformans* with subtelomeric methylation as well (Dumesic et al., 2015). The presence of H3K9 methylation in RNAi deficient strains shows that there is no direct link between the RNAi pathway and maintenance of H3K9 methylation, and that recruitment of Clr4 to heterochromatic regions is unlikely to be through direct interaction with RNAi components such as either Argonaute protein. The relationship between Clr4 and the RNAi pathway has more recently been further explored through analysis of sRNA populations in C. *neoformans*. As expected, loss of Ago1 causes a loss of sRNAs, however the sRNA population increases with the loss of Clr4, with those lost and gained under the two respective conditions largely mutually exclusive (Burke et al., 2019). This suggests that Clr4 and RNAi silence different regions, but once H3K9 methylation is lost, the RNAi pathway is able to act as a backup mechanism. This would form a non-direct link between RNAi and H3K9 methylation, and could also explain why both are present at the centromeric regions, along with DNA methylation which although I haven't explored has also been shown to be present at the centromeres (Catania et al., 2020; Yadav et al., 2018). It would be interesting to explore the relationship between RNAi,

Clr4 and Dnmt5 and their silencing pathways in more detail at the centromeres to discern if there are patterns between the sequences silenced in each pathway or if one is more prominent at specific conditions, such as during mating when RNAi is thought to be prevalent and H3K9methyltation has been shown to decrease. For this exploration, ideally strains would be constructed containing deletions of RNAi components, Clr4 and Dnmt5, which is problematic as we currently only have two functional markers which produce good results for genomic manipulation, and increased rounds of CRISPR increases the risk of genome instability, so other approaches may be required.

From here, I chose to look in more depth at the role of Clr4 within the cell in respect to establishment and maintenance of H3K9 methylation. With the knowledge that Dnmt5 is only a maintenance protein, with the establishment DNA methyltransferase being lost ~150-50 mya, I was intrigued to see if *C. deneoformans* still has the ability to establish H3K9 methylation (Catania et al., 2020). The data presented here suggests that Clr4 is the sole H3K9 methyltransferase within the genome, although other SET containing proteins are encoded for methylation of other histone residues and proteins. The finding that H3K9 methylation returns to WT levels after both repression and subsequent activation, and deletion and re-introduction of *CLR4* shows that *C. deneoformans* still has the ability to establish H3K9 methylation. It is unlikely that this result is due to remnants of H3K9 methylation having remained to allow propagation via maintenance mechanisms, due to the length of time the *clr4*Δ strain was passaged before re-introduction. This suggests that the proteins involved in H3K9 methylation establishment are maintained in the genome of *C. deneoformans*. To determine if there is a link between RNAi or DNA methylation and establishment of H3K9 methylation, further investigation carrying out the deletion and re-introduction of Clr4 in a *dnmt5*Δ or *rdp1*Δ

background could help determine if the localisation of Clr4 and establishment of H3K9 methylation relies on either pathway.

Finally, after ruling out H3K9 methylation as mode of silencing for the RNAi pathway, I decided to look at the transcript levels at regions known to be associated with sRNAs. As these regions are predicted to be silenced, loss of RNAi would be expected to cause an increase in expression levels, however no difference was found between WT, RNAi deficient strains and double H3K9 and DNA methylase deficient strains. This suggests that transcript levels remain constant, and that silencing occurs via a PTGS mechanism or that silencing is maintained by an RNAi-independent mechanism. RNA degradation is unlikely to be the mechanism as RNA levels would also likely increase in RNAi deficient strains in this case. The overall expression levels of sRNA target regions were generally low when compared with actin, even those corresponding to a hypothetical protein-coding sequence, which could suggest that only small amounts of transcripts are produced in RNAi functional and non-functional cells, making it harder to determine a difference. Analysis of Pol-II loading would help confirm if the level of transcription is related to the level of transcripts seen. Amongst the different sRNA target regions tested, there does not seem to be a correlation between the level of sRNAs seen in different strains and the transcript levels; even with the *clr4*Δ*dnmt5*Δ deletion strains where two regions show reduced sRNA accumulation compared to WT, but no increase in transcript accumulation. As RNAi silencing doesn't seem to regulate the transcription of target regions or RNA stability, the mode of silencing could occur further along the pathway from gene to protein. This could be through regulation of translation, or transport of the transcripts from the nucleus to the cytoplasm. Other things to bear in mind are that many sRNA reads don't cover predicted proteins, as seen in region G, or appear immediately up- or down-stream of a coding gene, as seen in region E.

Therefore, although these areas are where sRNAs are produced, the targeted area for silencing may be the neighbouring region such as through regulation of 5'UTR or 3'UTRs, so it might be worth analysing transcript levels for neighbouring regions instead of just the sRNA hits. If the non-coding sRNA associated regions are the targets for silencing, then other downstream PTGS mechanisms of silencing might need to be addressed, as the regions aren't predicted to code for genes or proteins and therefore translation is unlikely to occur. Other possibilities for further experiments could include looking at conditions where RNAi is thought to be more important, allowing greater differences to be seen in transcript levels, if there are any. This would include mating, as expression levels of Tcn1 have been shown to increase during mating (X. Wang et al., 2010), or stress conditions such as different temperatures, although initial findings from my own experiments didn't provide any consistent or clear trends.

Whilst the actual method of silencing is still not understood in C. *deneoformans,* this chapter has confirmed some of the findings from studies in the closely related C. *neoformans* species, and has suggested that silencing occurs through a PTGS mechanism independent of RNA stability. A direct link between the silencing pathways of Clr4 and RNAi has not been shown despite the lab sRNA sequencing dataset showing the presence of some sRNAs which are Clr4 and Dnmt5-dependent. This suggests that a link may be present at some target loci, and could provide the focus for future work.

Chapter 4 Investigating the role of silencing transposable element activity

4.1 Introduction

RNAi is often used as a mechanism of silencing transposable elements within genomes, acting as a genome defence method. When TEs are active, they can mobilise and insert into the genome at new loci. Silencing TEs prevents this mobilisation, thus preventing any deleterious insertions occurring and protecting the genome. RNAi within C. *deneoformans* has been shown to have a role in silencing both DNA transposons and retrotransposons within the genome. sRNAs have been shown to be produced against Tcn LTR-retroelements, and the expression levels of these elements have been shown to increase during mating when SIS is inactive (X. Wang et al., 2010). It has also been shown that transcription of some DNA transposons increases in RNAi-deficient strains (Janbon et al., 2010), and increased mobilisation of a DNA transposon has also been utilised as a readout in assays to identify new RNAi factors (Burke et al., 2019).

The aim of this chapter is to look at the efficiency of transposon silencing within strains deficient in RNAi, H3K9 methylation and/or DNA methylation. Again, the native RNAi pathway is being explored through the identification of transposition events of TEs from their endogenous loci within the genome into target genes which confer drug resistance. Utilising this transposition assay will ensure that I am identifying the TE silencing methods at their endogenous loci. By comparing the rates of transposition between different deletion strains, I aim to determine if there are any differences between silencing of different transposons, particularly between DNA transposons and retrotransposons. This is of particular interest as although RNAi has been linked to regulation of both types of TEs, a functional role for the chromatin marks H3K9 methylation and DNA methylation in TE regulation has not been shown, even though they are present over the centromeres where retrotransposons reside. This transposition assay will help determine the specific

TEs which are silenced by RNAi, as well as exploring the relationship between TEs and DNA and H3K9 methylation.

4.2 Strains lacking both H3K9 methylation and DNA methylation have elevated mutation rates

To look at the native role of RNAi in silencing of transposons, I chose an assay where I could measure transposon mobilisation from the original loci into one of two genes, *URA3* and *URA5,* whose disruption I could select for on media containing 5-FOA. From here the drug resistance rate can be calculated, defined as the number of mutations per cell per generation. I hypothesised that RNAi-deficient strains would have the highest mutation rates. This is because RNAi has been shown to silence TEs, and a loss of silencing would increase transposon transcription and subsequent mobilisation into either of the two target genes. This would result in a higher mutation rate, as in WT cells the only mutations occurring would be Mendelian mutations such as single nucleotide polymorphisms (SNPs).

Although the previous chapter showed that there is no link between RNAi and H3K9 methylation, I chose to carry out this transposition assay on strains deficient in H3K9 methylation and DNA methylation alongside RNAi-deficient strains, as well as combinations of these three. This is because although RNAi has been shown to silence both DNA transposons and retrotransposons, I hypothesised that H3K9 methylation and/or DNA methylation may contribute to the silencing of retrotransposons within the centromeres in an RNAi-independent manner. Eight strains were used in total: JEC21, *dcr1Δdcr2Δ*, *rdp1Δ*, *clr4Δ*, *dnmt5Δ*, *clr4Δdnmt5Δ*, *rdp1Δclr4Δ* and *rdp1Δdnmt5Δ*. As mentioned previously, only two markers are available for genomic manipulation so a strain deficient in RNAi, H3K9 methylation and DNA methylation could not be

constructed. JEC21 was used as the WT control in this assay, as it contains both *URA5* and *URA3* genes, whereas the WSA1226 strain used previously is lacking *URA5* so already 5-FOA resistant. WSA1226 is often used as the starting point for genomic manipulation, as out of the two markers available the *URA5* gene is the most frequently used. This means that the majority of the strains (all apart from JEC21 and *dcr1*Δ*dcr2*Δ) are in the WSA1226 background, with *URA5* inserted to disrupt one of the genes and not in its endogenous position. The only deletion strain with JEC21 as a background is *dcr1*Δ*dcr2*Δ, where the two genes are disrupted through a single HYG^R gene.

To measure the mutation rates of each strain, 10 independent cultures were grown for 3 days in rich medium which would allow for random mutations to occur. Each culture was then plated onto 5-FOA medium, allowing for growth only of cells which have a non-functional *URA3* or *URA5* gene. Appropriate dilutions of cells were also plated onto rich medium allowing for the cell viability to be determined. The occasional culture which did not have any viable cells after 3 days of growth were repeated. The number of 5-FOA resistant cells per mL of culture, along with the number of viable cells per mL were determined, and fluctuation analysis using the Ma-Sandri-Sarkar Maximum Likelihood Estimator (MSS-MLE) was carried out (Sarkar et al., 1992). Fluctuation analysis is based on the model of expansion of mutation clones by Luria and Delbruck (Luria & Delbruck, 1943), and the MSS-MLE method of calculating the mutation rate was chosen due to its accuracy by calculating the mutation rate taking into account the entire data set and not just the median as calculated via the Lea-Coulson method (Lea & Coulson, 1949).

Surprisingly, the fluctuation analysis shows that the *clr4*Δ*dnmt5*Δ strain has the highest mutation rate, approximately five-times greater than WT, whilst the *dcr1*Δ*dcr2*Δ strain

has a comparable mutation rate to WT (Figure 4.1A). However, *rdp1*Δ strains have a mutation rate approximately three-times higher than WT, suggesting that some RNAi-deficient strains do show an increase in mutation rates when compared with WT. Combinatory strains of RNAi deficiency and either H3K9 methylation deficiency or DNA methylation deficiency do not have an increased mutation rate beyond either individual deletion strain mutation rate, suggesting that the two deletions affect the same pathway. However, the mutation rates of neither single *clr4*Δ or *dnmt5*Δ strain is as high as the double deletion *clr4*Δ*dnmt5*Δ strain, suggesting a combinatory effect with the two deletions disrupting parallel pathways.

To determine if the increase in mutation rate was due to an increase in growth rate, growth curves were calculated for all strains. This is important to determine as the mutation rates are determined from a set length of culture growth and not a specific number of cell generations. Therefore, a higher growth rate would naturally expect to have a higher mutation rate due to increased cell replication providing increased opportunity for error and evolution. Cultures were grown for 4 days in rich medium, as per the mutation rate cultures, and the OD_{600} was measured regularly. The growth curves show that culture growth starts at ~12hrs after inoculation and that the culture grows exponentially up to ~48hrs where the growth rate starts to fluctuate and plateau (Figure 4.1B). The exponential growth rate of all deletion strains is lower than WT, with *rdp1*Δ*clr4*Δ having the slowest rate and *dcr1*Δ*dcr2*Δ being closest to WT. However, the main differences between the growth curves are the level at which each culture plateaus, showing that this occurs after ~48hrs of growth, no matter the density of the culture. This suggests that the cultures stop growing due to a factor out with overcrowding of the culture. From 48hrs, the WT cultures continues to slowly grow up to 96hrs, showing the greatest overall growth, whilst the other strains fluctuate. However,

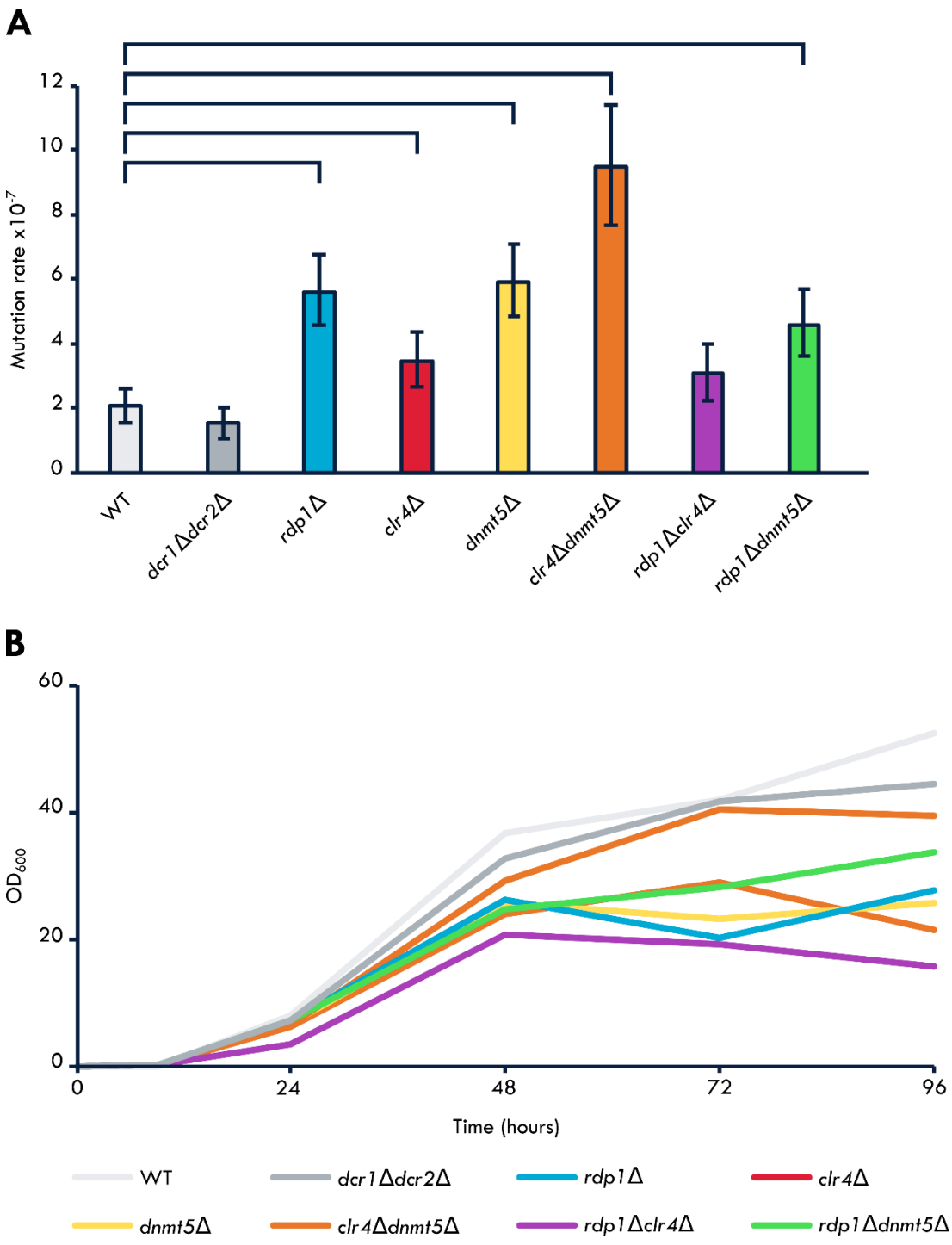

Figure 4.1 *clr4Δdnmt5Δ* strain has the highest mutation rate
(A) Graph showing the mutation rates of WT, *dcr1Δdcr2Δ*, *rdp1Δ*, *clr4Δ*, *dnmt5Δ*, *clr4Δdnmt5Δ*, *rdp1Δclr4Δ* and *rdp1Δdnmt5Δ* on 5-FOA calculated via MSS-MLE fluctuation analysis from 10 parallel cultures. Error bars show 95% confidence intervals; rates are statistically significant if error bars don't overlap shown through brackets. WT is JEC21 strain. Mutation rate is defined as the number of mutations per cell per division. **(B)** Graph showing growth curves for the eight strains used in the transposition assay, grown in rich medium for 92 hours. n=1 for all growth curves.

as no culture has a growth rate quicker than WT, it suggests that the changes in mutation rate are not due to increased passaging which would result in increased chances of mutations occurring and increased duplication of any mutation that did occur.

This data unexpectedly shows that a lack of DNA and H3K9 methylation causes the greatest increase in mutation rates, however some RNAi deficient strains did also have an increased mutation rate when compared with WT. Increases in mutation rates were shown to not be linked to an increase in growth rate, suggesting a mechanism involved in preserving the genome has been lost.

4.3 *rdp1Δ* and *clr4Δdnmt5Δ* strains have higher rates of transposable element insertions than WT

Whilst the drug resistance rates identify the strains with the highest number of mutations occurring within the culture, 5-FOA resistance can be caused by any mutation within either *URA3* or *URA5*. These could be SNPs, frame shifts, or larger insertions or deletions which disrupt the gene. Therefore, an increase in the drug resistance rate doesn't necessarily mean higher rate of transposon mobilisation. Taking this into account, I hypothesised that the increase in the 5-FOA resistance rate in RNAi deficient strains would be due to an increased rate of TE insertions within *URA3* and *URA5*, whereas the increase in 5-FOA resistance rate in H3K9 methylation and DNA methylation deficient strains would be due to other causes such as SNPs and not TE insertion.

To analyse the type of mutation resulting in 5-FOA resistance, up to 24 colonies from the 5-FOA plates from each independent culture were restreaked onto 5-FOA plates and genomic DNA was extracted. PCR amplification of both *URA3* and *URA5* genes was used to determine if the drug resistance was due to a large insertion or deletion,

followed by sequencing of a selection of samples with each size of insertion to determine the TE present. Sometimes colonies did not grow after restreaking, suggesting that the drug resistance in some cases was temporary and reversible, with *rdp1*Δ strains having the lowest percentage of colonies re-growing (188 out of 240, 78.3%). Reversible drug resistance is likely due to a large insertion which is subsequentially excised, or through a silencing mechanism targeting the genes of interest. When selecting colonies to be restreaked for genomic extraction and screening, colonies chosen were taken from those that appeared on 5-FOA plates up to 2, 4 and 7 days after plating in even proportions. This is because I was unsure if there might be a bias in the origin of mutations in colonies appearing early or late. Visible phenotypes were also noted in a number of colonies and their subsequent restreaks. While C. *deneoformans* usually grows to form perfectly circular, off-white-coloured colonies, several colonies were noted to form irregular circle shapes (245 out of 1525). These were usually more yellow in colour too, and observations under the microscope showed flocculation of the cells. Other visible phenotypes included those which formed much larger colonies than usual (67 out of 1525), and those which had a shiny appearance (51 out of 1525), though no differences could be observed under the microscope in either of these cases.

PCR amplification of both genes identified colonies with large insertions and deletions. Wild-type *URA3* gives a ~1.2 kb band and wild-type *URA5* gives a ~1 kb band and any band larger or smaller than these was counted as an insertion or deletion (Figure 4.2A). Screening showed that the majority of colonies, across all the cultures from all the strains, did not contain an insertion or deletion (1457 out of 1624). However, all strains had at least one culture with an insertion or deletion, with WT having the lowest

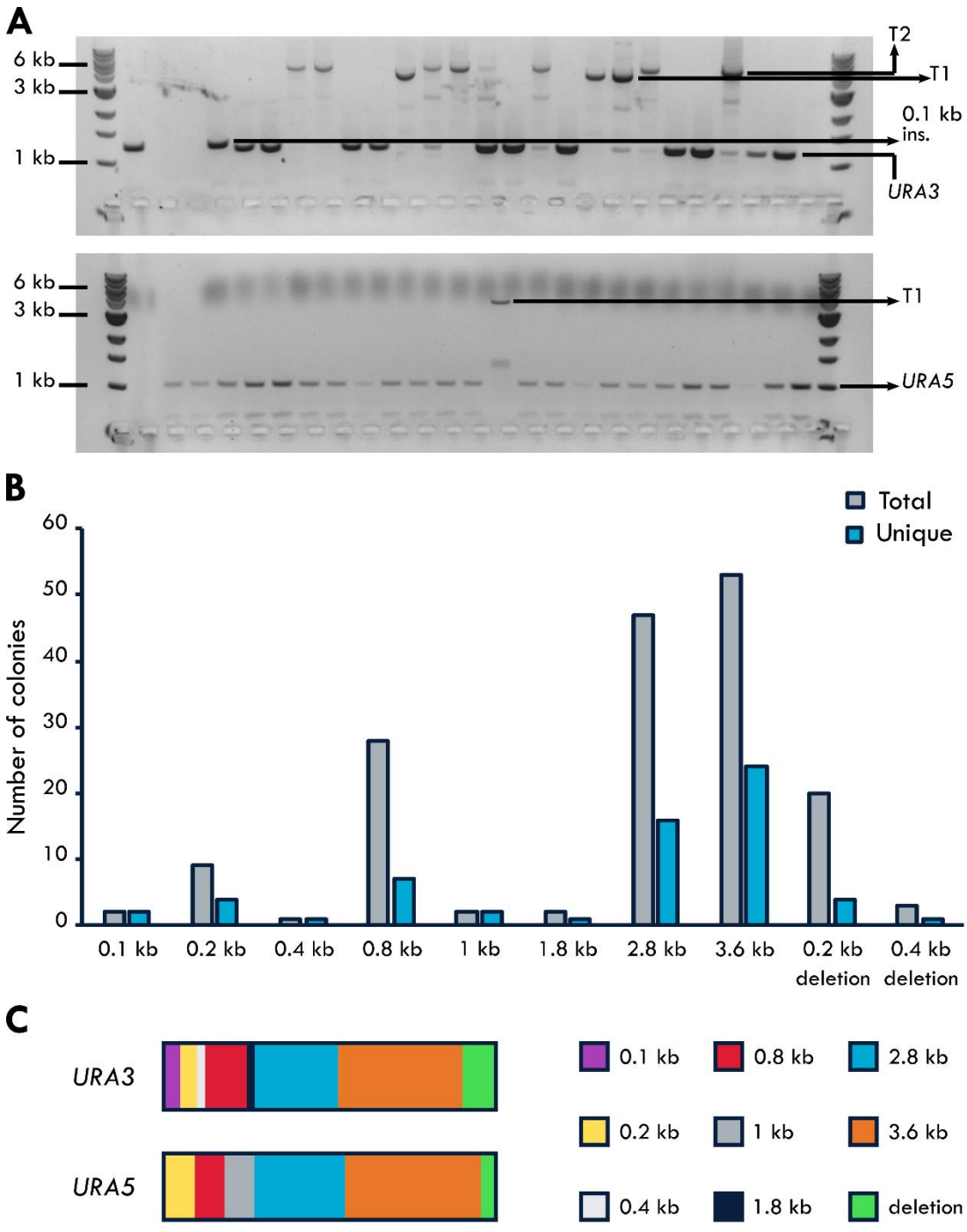

Figure 4.2 A proportion of 5-FOA resistance is due to insertions of transposable elements

(**A**) Gel electrophoresis of PCR amplified *URA3* and *URA5* genes for the 24 colonies screened from one culture (*dcr1*Δ*dcr2*Δ), showing both large and small insertions. Colonies with the same insertion within the same gene are counted as only one unique event. (**B**) Graph showing the combined total number of colonies with each size of inertion or deletion from all 8 strains tested, along with the number of unique insertions for each size of insertion or deletion. (**C**) Schematic showing the distribution of unique insertions and deletions between both genes screened (*URA3* and *URA5*).

number of insertions and deletions detected out of the total colonies screened (6 out of 200) whilst *clr4*Δ*dnmt5*Δ had the highest (51 out of 234) (Table 4.1).

Eight different sized insertions were identified, ranging from 0.1 kb to 4 kb, whereas only two deletions were identified, 0.2 kb and 0.4 kb. Occasionally multiple bands would be present in samples containing an insert, often including the wild-type product size along with several different insertion sizes. In this case, the most intense band was considered to be the product of interest. The reason behind this was assumed to be mixed populations, potentially through some of the cells having lost the TE which had inserted into the gene. On other occasions no product was amplified for one of the genes, when the amplification of the other gene produced a band with no insertion. As the presence of one band confirms that genomic DNA was successfully extracted, the absence of a PCR product is likely to be the result of either an insertion over one of the primer sites, an insertion which is too large to be amplified by PCR or human error. In all these cases, it was assumed that there was no insert within either gene.

Within each culture, insertions of the same size would often be seen within several colonies. This is likely due to cells with insertions dividing within the culture, and therefore the earlier that the mutation occurs, the more cells that are likely to contain the same insertion. As I was more interested in the number of unique insertions and the different TEs, and not how quickly the mutation arises, in quantifying insertions I chose to only look at unique insertions. Therefore, out of the (up to) 24 colonies screened from one culture, those with the same size insertion within the same gene were assumed to be the result of just one transposition event and counted just once, no matter how many times I identified colonies containing the insertion (Figure 4.2B). Duplication events would also occur in 5-FOA resistant colonies without TE insertions, though without sequencing all *URA3* and *URA5* genes it would be impossible to determine how many non-TE insertion

Table 4.1 Transposon assay insertions per culture

Strain		1	2	3	4	5	6	7	8	9	10	Total
WT	Insertions and deletions	0	0	0	0	0	0	0	2	0	4	6
	Colonies restreaked	24	24	24	24	24	9	24	20	21	22	216
	Colonies screened	24	24	24	24	19	9	24	16	17	19	200 (92.6%)
	Unique insertions and deletions	0	0	0	0	0	0	0	1	0	1	2
dcr1Δdcr2Δ	Insertions and deletions	0	0	8	2	2	13	0	0	0	0	25
	Colonies restreaked	3	24	24	24	24	24	17	13	8	8	169
	Colonies screened	2	24	22	24	24	24	15	13	7	8	163 (96.4%)
	Unique insertions and deletions	0	0	3	1	1	4	0	0	0	0	9
rdp1Δ	Insertions and deletions	4	3	4	7	0	0	1	0	0	2	21
	Colonies restreaked	24	24	24	24	24	24	24	24	24	24	240
	Colonies screened	24	24	23	24	24	23	15	13	5	13	188 (78.3%)
	Unique insertions and deletions	3	1	2	1	0	0	1	0	0	1	9

Strain		1	2	3	4	5	6	7	8	9	10	Total
clr4Δ	Insertions and deletions	0	0	0	6	0	1	0	0	0	1	8
	Colonies restreaked	10	24	24	24	16	24	24	24	24	24	218
	Colonies screened	10	18	17	22	13	23	23	23	22	22	193 (88.5%)
	Unique insertions and deletions	0	0	0	2	0	1	0	0	0	1	4
dnmt5Δ	Insertions and deletions	9	0	0	0	1	0	0	0	0	0	10
	Colonies restreaked	24	24	24	24	24	24	24	24	24	24	240
	Colonies screened	24	24	24	24	24	23	24	24	24	24	239 (99.6%)
	Unique insertions and deletions	1	0	0	0	1	0	0	0	0	0	2
clr4Δdnmt5Δ	Insertions and deletions	1	7	3	4	14	0	13	0	4	5	51
	Colonies restreaked	24	24	24	24	24	24	24	18	24	24	234
	Colonies screened	24	24	24	24	24	24	24	18	24	24	234 (100%)
	Unique insertions and deletions	1	3	1	2	1	0	2	0	2	3	15

Strain		1	2	3	4	5	6	7	8	9	10	Total
rdp1Δclr4Δ	Insertions and deletions	0	3	4	2	2	0	9	0	1	5	26
	Colonies restreaked	24	24	22	22	24	15	24	16	8	24	203
	Colonies screened	23	24	22	15	16	14	24	16	8	24	186 (91.6%)
	Unique insertions and deletions	0	1	2	2	1	0	2	0	1	2	11
rdp1Δdnmt5Δ	Insertions and deletions	8	0	1	0	0	2	4	3	1	1	20
	Colonies restreaked	24	24	24	24	5	24	24	24	24	24	221
	Colonies screened	24	24	24	24	5	24	24	24	24	24	221 (100%)
	Unique insertions and deletions	3	0	1	0	0	1	2	1	1	1	10

mutations occurred. As a result, when calculating the ratio of unique insertions out of colonies screened, I did not adjust the total number to remove those containing non-unique insertions. This does mean that the ratios calculated are likely to be lower than the true ratios if only unique TE insertion and non-TE insertion events were counted, but by using these lower estimates it ensures that any differences seen between strains are likely to be real. Whilst mutations in either *URA3* and *URA5* equally result in 5-FOA resistance, a bias was seen towards insertions and deletions occurring in the *URA3* gene, with twice as many unique insertions and deletions identified (22 in *URA5* compared with 44 in URA3). Comparing the proportion of different TEs and deletions between the two genes showed that some elements were only identified within one gene, though these were usually elements which had a low frequency of occurrence. For those with a higher frequency of occurrence, a similar proportion of unique insertions was seen within each gene (Figure 4.2C). However, individual strain biases did occur, with neither *rdp1*Δ or *dnmt5*Δ strains having any *URA5* insertions (Table 4.2).

The overall TE insertion rate was calculated for each strain, taking into account the overall mutation rates and the proportion of unique insertions into both *URA3* and *URA5* from all colonies screened. Both *clr4*Δ*dnmt5*Δ and *rdp1*Δ have a significantly higher transposon insertion rate than WT, with *clr4*Δ*dnmt5*Δ having the highest rate (Figure 4.3A). The proportion of colonies screened with a unique insertion was also calculated. This showed that there is no significant difference between the number of unique insertions and deletions identified in any of the eight strains, although *drc1*Δ*dcr2*Δ, *rdp1*Δ, *clr4*Δ*dnmt5*Δ, *rdp1*Δ*clr4*Δ and *rdp1*Δ*dnmt5*Δ strains have a similar proportion whilst WT, *clr4*Δ and *dnmt5*Δ strains have a similar, lower proportion (Figure 4.3B). Together this analysis shows that there is no significant difference in the overall TE

Table 4.2 Unique insertions in *URA3* and *URA5* per strain

Strain	Mutation rate x10^{-7}	Gene	T1 2.8kb	T2 short 0.8kb	T2 long 3.6kb	0.1kb	0.2kb	0.4kb	1kb	1.8kb	Deletion	Total unique	Colonies screened
WT	2.04	URA3									1	1	200
		URA5	1									1	
dcr1Δdcr2Δ	1.53	URA3	3		1	1						5	163
		URA5	3		1							4	
rdp1Δ	5.60	URA3	2		4	1	1			1		9	188
		URA5										0	
clr4Δ	3.45	URA3			2						1	3	193
		URA5			1							1	
dnmt5Δ	5.92	URA3	1	1								2	239
		URA5										0	
clr4Δdnmt5Δ	9.54	URA3		4	3						1	8	234
		URA5	1	2	2				2			7	
rdp1Δclr4Δ	3.05	URA3	2		3		1					6	186
		URA5	1		3		1					5	
rdp1Δdnmt5Δ	4.59	URA3	2		2			1			1	6	221
		URA5			2		1				1	4	

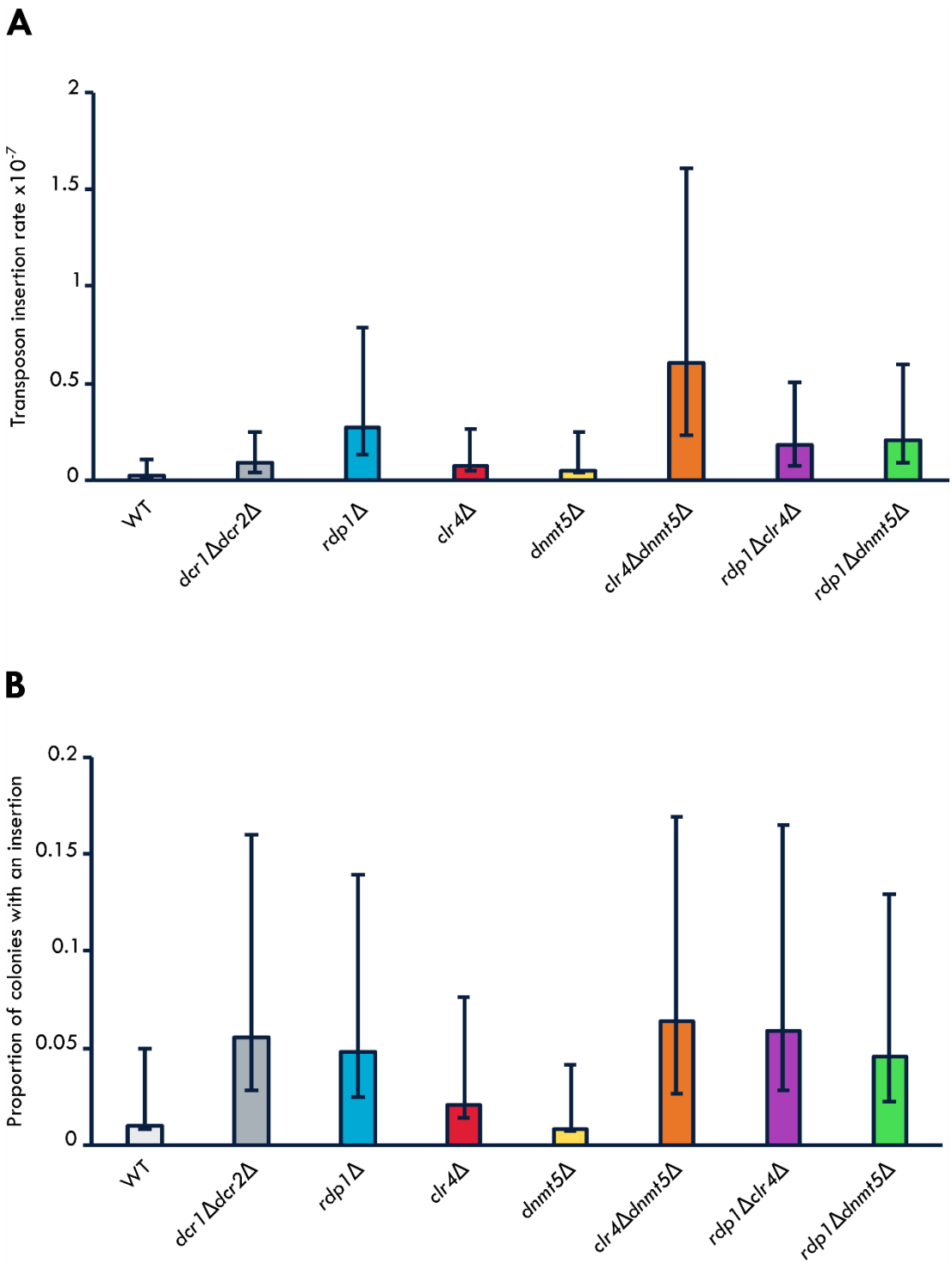

Figure 4.3 Both *rdp1Δ* and *clr4Δdnmt5Δ* strains have signifiantly higher transposon insertion rates than WT

(**A**) Graph showing the unique transposon insertion rate for each strain. Error bars show 95% confidence intervals. WT is JEC21 strain. (**B**) Graph showing the proportion of colonies with unique insertions out of the total number of colonies screened. Error bars show 95% confidence intervals of a proportion. WT is JEC21 strain.

insertion rates between RNAi deficient strains and H3K9 and DNA methylation deficient strains, although both do show an increase compared to WT, and that TE insertions account for a similar proportion of the drug resistance in these strains.

Sanger sequencing was carried out to determine the origin of each different insertion size identified. This proved problematic as *URA3* could only be successfully sequenced from one end, and sequencing reads would not cover the whole of the gene. This meant that only two transposable elements could be identified: T1 giving a 2.8kb insert and T2 giving an ~3.5 kb insert whilst the short version of T2 gives a 0.8kb insert. The 1.8 kb insert matches the size of the non-LTR retrotransposon Cnl1, although further attempts at sequencing are needed to confirm this. The smaller insertions (≤0.4 kb) are too small for full length transposable elements, though they could represent truncated versions which have been able to mobilise or solo LTRs.

Transposon mobilisation rates were determined for each individual element, or insertion size, identified to see if different strains had different rates of insertion of different elements (Figure 4.4). No significant differences were seen between strains for elements which were identified in multiple strains, likely due to the overall low number of unique insertions identified compared with the high number of colonies screened. However, different elements were identified in different strains. T1 was seen to mobilise in all strains apart from *clr4*Δ, and potentially more in RNAi-deficient strains. T2 was shown to mobilise in all strains apart from WT and *dnmt5*Δ, potentially at higher rates in *clr4*Δ*dnmt5*Δ strains. However, the short version of T2 was only identified in *clr4*Δ*dnmt5*Δ and *dnmt5*Δ strains suggesting that its mobilisation is regulated by DNA methylation. If the 1.8 kb insert is representative of Cnl1, it would be the only

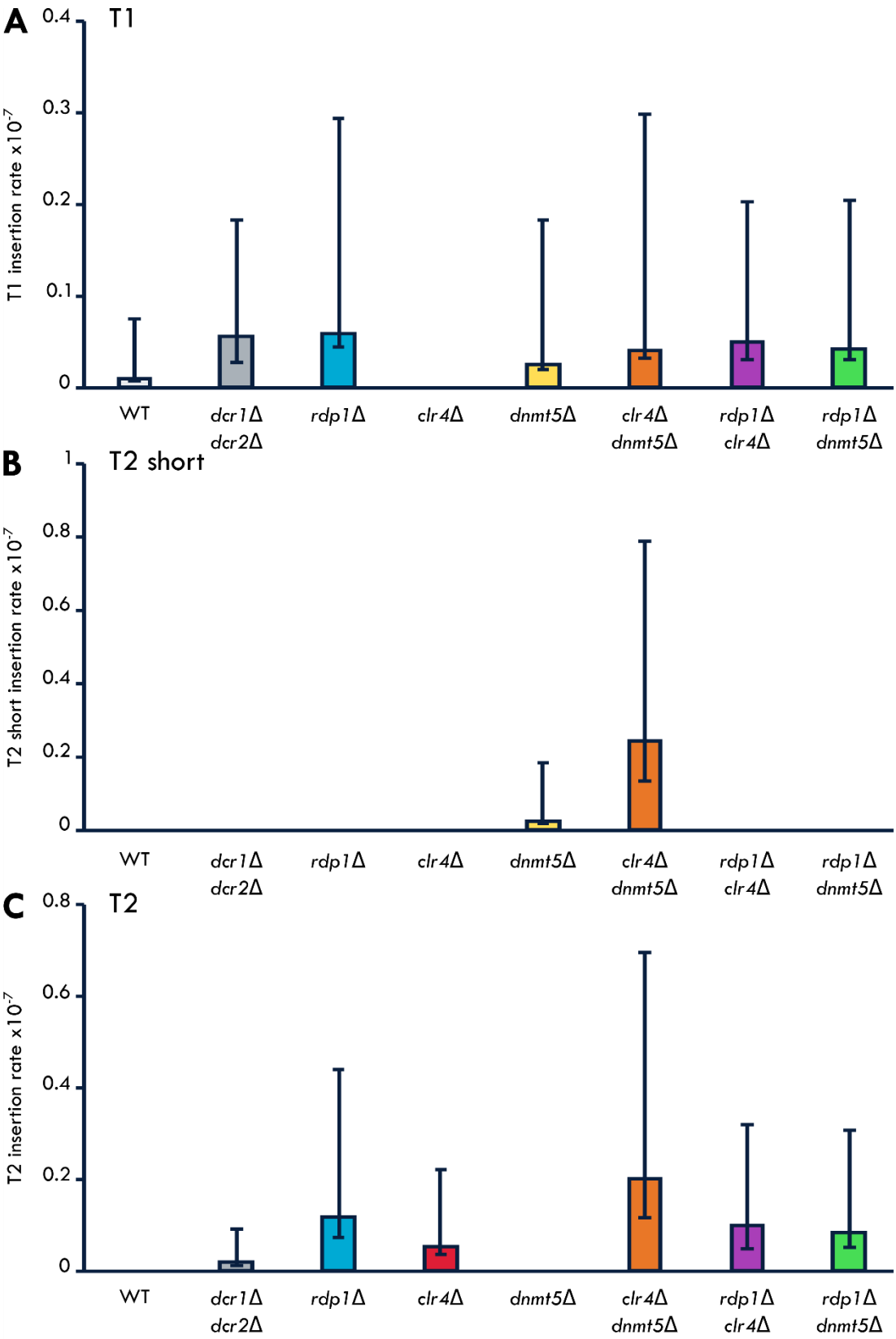

Figure 4.4 Transposon insertion rates of all transposons identified in the transposition assay

Graphs showing the individual mutation rates for (A) T1, (B) T2 short and (C) T2. Error bars represent 95% confidence intervals. WT is JEC21 strain.

retroelement identified, though this was only identified in the *rdp1*Δ strain at a very low rate with only one unique insertion.

Overall, this data shows that TE mobilisation occurs in all strains, and that this accounts for a small proportion of the overall 5-FOA resistance rates in both RNAi deficient and H3K9 and DNA methylation deficient strains. Initial comparison of the different TEs suggests that T2 short is likely to be silenced through H3K9 and DNA methylation whereas Cnl1 may be silenced by the RNAi pathway, however further screening of more colonies and/or cultures is required to determine significant differences between the identified TE insertion rates.

4.4 LTR-retrotransposons do not have a higher copy number in RNAi deficient strains than WT

Although the transposition assay was successful in identifying transposons which had mobilised into either *URA3* or *URA5* genes, no LTR-retrotransposons were identified. LTR-retrotransposons are localised to the centromeres and have also been reported to be silenced by RNAi, as RNAi deficient strains show increased transcription (Loftus et al., 2005; X. Wang et al., 2010). LTR-retrotransposons are mostly larger than the DNA transposons identified, ranging from ~3.5 kb – 6 kb in length. If an LTR-retrotransposon was to mobilise and insert into either *URA3* or *URA5* genes, the PCR product would be ~1.2kb longer than the length of the LTR-retrotransposon itself, possibly making it too long for standard PCR amplification. This could be one explanation as to why some colonies showed a wild-type sized band for one of the genes of good intensity, confirming genomic DNA has been extracted, but no band in the PCR amplification of the other gene. Other explanations are possible, such as an insertion of a TE over the PCR primer binding site, and of course human error, though the pattern and frequency

of samples producing no PCR product suggest that something else is the cause of this. While WT, *dcr1*Δ*dcr2*Δ and *clr4*Δ*dnmt5*Δ strains had less than 2.5% of colonies showing this pattern of failed amplification for one of the two genes, the percentage was higher in all other strains, with the strains combining both RNAi deficiency and either H3K9 methylation deficiency or DNA methylation deficiency having the highest percentages (15% and 19% respectively). I therefore hypothesised that LTR retroelements were active in some strains, however I was unable to detect them during the transposition assay due to their length. I also hypothesised that LTR retroelements would be more active in strains combining RNAi and either H3K9 methylation or DNA methylation deficiencies, as suggested by the proportion of empty inserts identified.

As an alternative method to determine if LTR-retroelements are active in some strains, I carried out qPCR analysis on genomic extractions from cultures grown in rich media to determine element copy number. As LTR-retroelements are 'copy-and-paste' elements, the number of elements will increase within the genome if the element is active. Therefore, more copies will be present throughout the genome, giving a higher starting point for the qPCR, resulting in higher levels after amplification. Analysis was carried out for three LTR-retroelements: Tcn1, Tcn5 and Tcn6. Tcn6 was chosen as the only family from the Ty1/*copia* superfamily, and Tcn6, alongside Tcn5 from the Ty3/*gypsy* superfamily, are the most abundant LTR-retroelements with copies in each of the 14 centromeres. The other element chosen was Tcn1, also of the Ty3/*gypsy* superfamily, as it is the only element with an identified chromodomain (Goodwin & Poulter, 2001). Retroelements often have chromodomains which directs them to insert in regions of heterochromatin, where insertions are less likely to result in deleterious effects (Gao et al., 2008). Due to the presence of the chromodomain, Tcn1 would have been unlikely to insert into either *URA* genes in the transposition assay.

Cultures used for this copy number analysis were independent from those used in the transposition assay, though grown under the same conditions. Samples for genomic extraction were collected after day 1 and day 3 of growth to see if any changes occurred over the growth period. qPCR levels were normalised to *ACT1,* of which there is just a single copy within the genome, and the number of copies should remain constant over the 3-day growth period (Figure 4.5). Data were analysed via two-way ANOVA with repeats for each of the LTR-retrotransposon tested. This showed that there is a positive interaction between the strain and the day of culture for Tcn1 ($F(5, 24)=3.20$, $p=0.0236$) (Figure 4.5A). This means that for Tcn1 the differences in copy number are due to both the day and the strain used whereas for Tcn5 and Tcn6 any significant differences in copy number are due to one or the other. A Tukey post hoc test showed significant differences between the copy number of Tcn1 at day 3 in JEC21 and the copy number at day 1 in both *rdp1*Δ and *dcr1Δdcr2*Δ ($p\leq0.05$). A significant difference was also seen in the copy number between the total day 1 and day 3 levels ($F(1, 24)=9.99$, $p=0.00422$), showing that the copy number is likely to increase over the course of each culture. Whilst Tcn5 shows no statistical interaction, significant differences were seen in the copy numbers between strains ($F(5,24)=13.8$, $p=0.00000205$), with Tukey post hoc test showing significant differences between the overall copy numbers in *rdp1*Δ and WT, *clr4Δdnmt5*Δ, *rdp1*Δ*clr4*Δ and *rdp1Δdnmt5*Δ strains ($p\leq0.05$) (Figure 4.5B). This shows that regardless of any change in copy number during the 3-day culture growth, *rdp1*Δ has a lower copy number of Tcn5 than WT and strains with H3K9 and/or DNA methylation deficiency. Tcn6 also shows no significant interaction, though a significant difference was seen in the copy number between the total day 1 and day3 levels ($F(1, 24)=4.35$, $p=0.0479$), showing that the copy numbers are likely to increase over the course of the 3-day culture (Figure 4.5C).

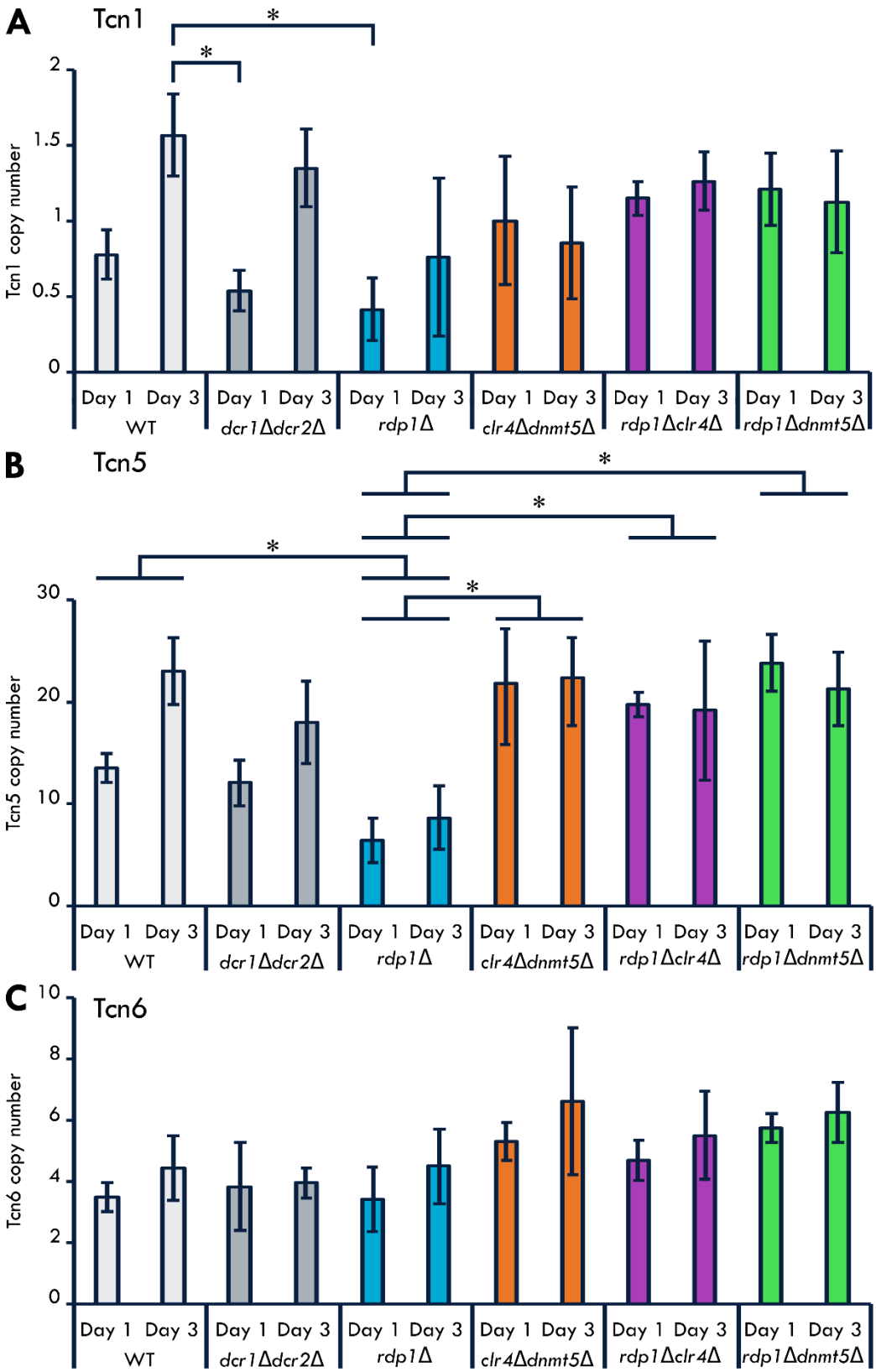

Figure 4.5 *rdp1Δ* has lower copy numbers of some LTR-retrotransposons
Graphs showing the copy number of **(A)** Tcn1, **(B)** Tcn5 and **(C)** Tcn6, normalised against *ACT1*. Error bars represent the standard deviation, and statistical significance is shown by * = p≤0.05. WT is JEC21 strain.

Overall, this shows that, depending on the LTR-retroelement, the copy numbers can increase from day 1 to day 3 of culture, showing mobilisation of the elements during the culture. This suggests that these elements can mobilise and could insert into *URA3* or *URA5,* so may be responsible for the absence of a PCR product in some colonies in the transposition assay. However, no significant increases were seen in the RNAi, H3K9 methylation or DNA methylation strains, nor was there a lack of mobilisation in the strains which had a low percentage of absent PCR bands. Therefore, without determining the reason for the lack of bands in all the strains, I cannot determine if there is likely to be increased mobilisation of LTR retroelements into *URA3* or *URA5* in *rdp1*Δ*clr4*Δ and *rdp1*Δ*dnmt5*Δ strains compared with the other strains. There is also no evidence of an overall increase in mobilisation in any strain. A significant reduction in the copy number of Tcn5 was seen in *rdp1*Δ cells, suggesting a reduction in full length Tcn5 copies throughout the genome.

4.5 Transcription of DNA transposons increases in some independent cultures of strains lacking RNAi

The transposition assay only screens for transposons which have inserted into either the *URA3* or *URA5* genes – a very small percentage of the overall C. *deneoformans* genome. It is therefore likely that transposons are active and mobilising in cultures where inserts were not identified through PCR screening. It is also likely, based on previous studies carrying out similar transposition assays in C. *deneoformans,* that transposons have a sequence preference for insertion, thus creating biases towards insertion into some genes over others (Gusa et al., 2020). Therefore, it is possible that other transposons beyond those identified through the PCR screening are also mobilised in some strains and were just not identified due to having a bias towards inserting into

other genes than *URA3* or *URA5*. To look at DNA transposon activity transcript levels have to be analysed. This is because DNA transposons are 'cut-and-paste' transposable elements, so the number of copies present within the genome will remain constant. However, if the elements are actively mobilising, the transposes in which they encode will be expressed to allow the element to excise and reinsert elsewhere in the genome. Therefore, the transcript levels should be indicative of the activity of the transposons. Therefore, I hypothesised that across whole cultures the expression of DNA transposons would increase in RNAi deficient strains when compared against WT and H3K9 and DNA methylase deficient strains.

To analyse the culture-wide expression of DNA transposons, I looked at the transcript levels of the three MULE TE families present in C. *deneoformans:* T1, T2 and T3. Separate cultures from those grown as part of the transposition assay were used, so that there was no bias towards those with insertions of MULE TEs already identified within the screened genes and those without. RNA was extracted after three days of growth in rich medium, as per the transposition assay, and RT-qPCR performed to analyse the transcript levels of each transposon family. The transcript levels were normalised to *ACT1,* which should be consistently expressed in all strains. As there are differences in the number of copies of each transposon within the genome, the overall levels of transcription cannot easily be compared between elements. The RT-qPCR data shows no significant differences in the transcript levels of all MULE TEs tested between any of the deletion strains and WT, as analysed by one-way ANOVA (Figure 4.6). The lack of significant differences is due to the large variance in TE transcripts between the three independent cultures measured for each strain. This suggests that events within individual cultures, which cannot be controlled for, alter the rate of DNA transposon expression. Comparing just the average transcription levels for each strain shows that

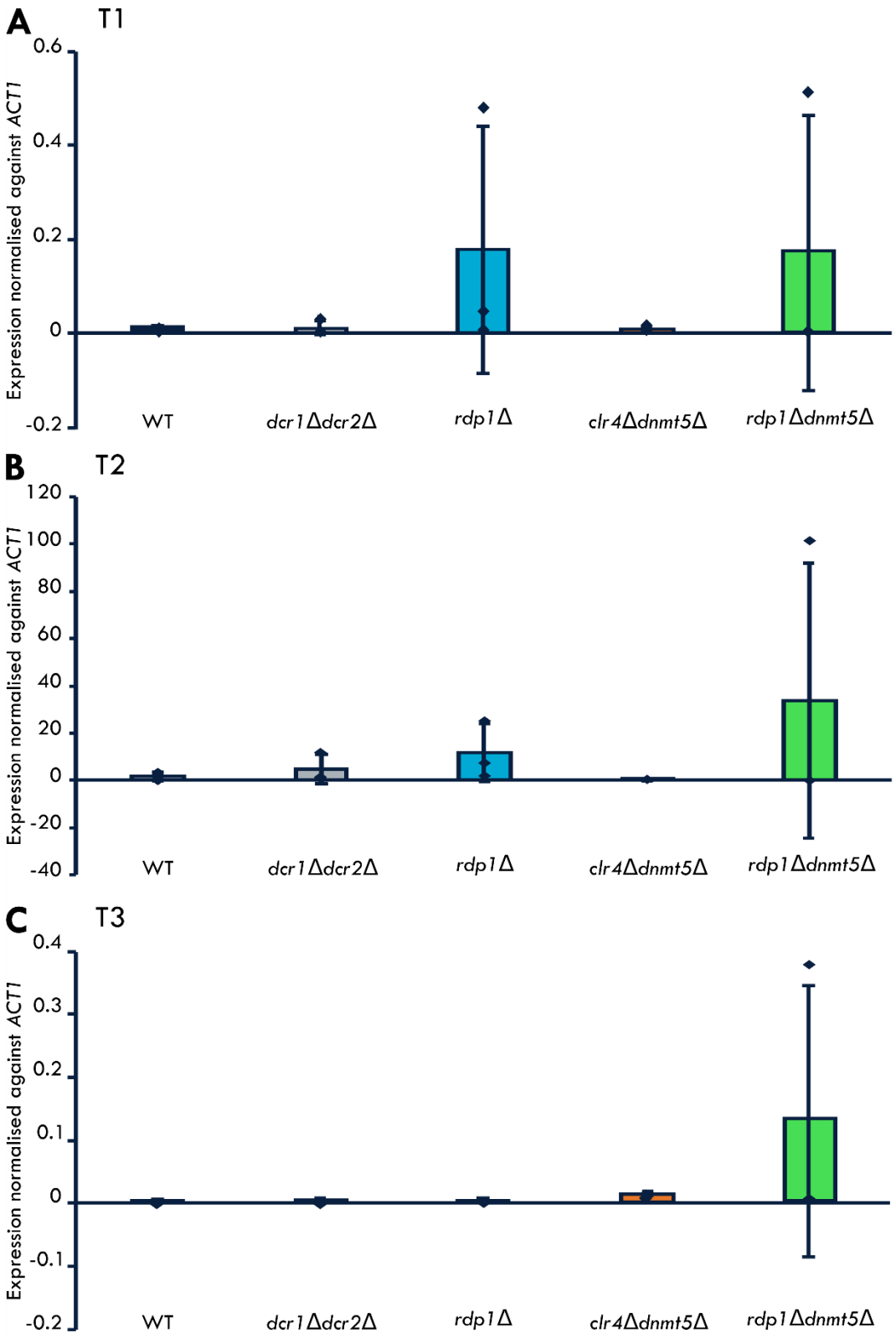

Figure 4.6 Transcript levels of DNA transposons

Graphs showing the transcript levels of **(A)** T1, **(B)** T2 and **(C)** T3, normalised against *ACT1*. Error bars represent the standard deviation, and no statistical significance has been determined. WT is WSA1226 strain. n=3 independent cultures for each strain, plotted on each graph as individual data points.

for both T1 and T2 the RNAi deficient strains show higher TE transcript accumulation than WT and methylation deficient strains, with *rdp1Δ* and *rdp1Δdnmt5Δ* strains showing greater increases than *dcr1Δdcr2Δ* strains (Figure 4.6A and B). For T3 however, the *rdp1Δdnmt5Δ* strain has the greatest increase in transcription, with *clr4Δdnmt5Δ* showing a small increase too (Figure 4.6C). These observations suggest that RNAi may play a dominant role in regulation of T1 and T2, which RNAi and DNA methylation may function redundantly in regulation of T3. This does also show that just because one MULE family is actively transcribed, the other families are not necessarily also actively transcribed, as seen in the lack of increase of T3 transcription in the *rdp1Δ* strain whilst at least one of the three cultures had a moderate increase in both T1 and T2.

Overall, these results suggest that RNAi may regulate the transcription of T1 and T2 elements, but that DNA methylation and potentially H3K9 methylation may contribute to regulating the transcription of T3 elements, potentially alongside RNAi. However, further repeats (n=10) would be required with more independent cultures to determine if any significant conclusions can be drawn, though this initial analysis shows promising potential differences in the regulation of different DNA transposons which are worth exploring more in the future.

4.6 Discussion

This chapter aimed to identify transposable elements which are silenced by RNAi by determining if the lack of RNAi results in increased transposition. It also aimed to determine if either H3K9 methylation or DNA methylation had a parallel role in silencing retrotransposons alongside the previously identified RNAi silencing.

Initial mutation rates for each strain was determined based on rates of acquisition of 5-FOA resistance, based on the experimental approach applied previously in a different strain of C. *deneoformans* (Gusa et al., 2020). Occasionally, after 3 days of growth cultures would have no viable cells, and repeat cultures were set up. The lack of viable cells was not due to a lack of growth as cultures had comparable OD_{600} to other cultures, nor due to infection. Therefore, it can be assumed that after a period of growth, the cells started to become non-viable and die. One possibility could be that to many mutations were occurring, therefore resulting in deleterious effects. It is likely, therefore, that the cultures with the highest mutation rates, and potentially the highest transposon mobilisation rates, were not screened as they could not survive the culture period. This suggests that only those with lower drug resistance rates would have been screened, with the top extremities removed from the assay. This may alter the overall trend, and analysis of the mutation rates over a shorter culture period may be worthwhile. This would also determine if the drug-resistance rate increased due to stressors within the culture, such as overcrowding.

The observation that *clr4*Δ*dntm5*Δ strains have the highest mutation rates of the strains tested, was unexpected. This suggested that H3K9 methylation and DNA methylation may have a role in silencing TEs. Previous studies have shown differences in the mutation rates in C. *deneoformans* between 30°C and 37°C, and confirmed that the increased mutation rate at 37°C was not solely due to a loss of RNAi at the higher temperature (Gusa et al., 2020). Based on the mutation rates calculated here, it could be that a loss of both H3K9 methylation and DNA methylation at 37°C is the cause of the increase in drug resistance at 37°C. This, paired with the knowledge that *clr4*Δ*dnmt5*Δ has a temperature sensitivity at 37°C, is worth exploring in the future to determine if there is a link between the temperature sensitivity and transposon insertion rate.

When cultures were plated onto 5-FOA after growth, colonies appeared over the course of a week, showing differences in growth rate. Also, colonies appearing between days 4 and 7 on 5-FOA media tended to be smaller in size than those which appeared by day 4. Different mutation types have been shown to have a difference in the growth of colonies in similar assays carried out in *S. cerevisiae* due to the varying degrees of 5-FOA resistance generated by the mutation (Radchenko et al., 2018). Large colonies indicate strong 5-FOA resistance and could be indicative of large deletions and nonsense mutations. Insertions within the gene are likely to result in average sized colonies, as the gene expression might not be totally reduced, though the likelihood of functional *URA3* or *URA5* expression depends on the size of the insertion with those such as TE insertions unlikely to have functional expression. The smallest colonies are likely to be those with indirect resistance or only downregulation of expression. Therefore, the small colonies appearing between days 4 and 7 were likely to have indirect resistance, meaning that the type of mutation will not be identifiable by screening just *URA3* and *URA5*. This general trend was observed during screening, as these colonies were less likely to have insertions than those appearing before day 4. Therefore, in future it may be worthwhile only analysing colonies appearing on 5-FOA before day 4. Whilst this may exclude some colonies of interest, the TE insertion rates would potentially be higher which may help determine further significant differences between strains. It would also be worth comparing the size of colonies on 5-FOA to those on rich medium for the full 7 days. Here only growth up to day 2 was analysed on rich medium to calculate culture viability, and no difference in colony size at this point was identified. Continuing the analysis over 7 days on rich medium would confirm if the size of colony was related to the 5-FOA mutation or if the slower, smaller growth was due to 'sick' cells, less able to grow generally and not related to the presence of 5-FOA.

Flocculation was also seen in a number of colonies and their subsequent restreaks. Whilst the cause of flocculation has been studied in *S. cerevisiae,* it has only briefly been explored in *Cryptococcus*. Homologs of the genes involved in the flocculation process in *S. cerevisiae* have not been identified in *Cryptococcus* species (Li et al., 2006). However, flocculation does still occur, and flocculating strains usually have reduced virulence due to increased phagocytosis by macrophages. The increase in flocculation is thought to be due to changes in the polysaccharide capsule altering the adherence of the cells. Studies on the pyrimidine biosynthesis pathway in C. *neoformans* have shown that an intact pyrimidine biosynthesis pathway is required for wild-type capsule formation (de Gontijo et al., 2014). This suggests that the observation of flocculation might be a result of mutations within the pyrimidine pathway, and might explain why this phenotype shows a higher proportion of TE insertions within either *URA3* or *URA5* than colonies without flocculation.

Screening of colonies with 5-FOA resistance showed a large proportion without any insertion into either the *URA3* or *URA5* gene. Whilst these could be due to the previously mentioned SNPs and frame shifts, another possibility is indirect mutations. For 5-FOA resistance, mutations could also occur in genes involved in uptake and transport of 5-FOA into the cell along with other genes encoding proteins regulating components of the *URA* pathway (though obviously mutations in any gene which is essential, such as those involved in protein synthesis, would be deleterious). Whilst *URA3* and *URA5* are those most likely to confer 5-FOA resistance, it is possible that other genes within the *URA* pathway may also cause resistance. The altered polysaccharide capsule formation which has previously been shown to cause flocculation was most often seen in mutations in *URA4,* which functions upstream of both *URA3* and *URA5* in the pyrimidine biosynthesis pathway (de Gontijo et al., 2014). *URA6* is the only *URA* gene which

functions downstream of *URA3* and *URA5*. Initial screening of a selection of colonies which had no insert in *URA3* or *URA5* showed no insertion into that *URA6* either. As further screening of *URA6* would greatly increase the screening, and mutations in *URA6* haven't been shown to induce 5-FOA resistance this was not continued. Other possible causes of 5-FOA resistance which wouldn't be identifiable via PCR would be small insertions or deletions. Previous studies have seen 3-5 bp deletions in genes after a transposon has excised (Gusa et al., 2020). Therefore some 'empty' PCR products may be the result of an earlier TE mobilisation event, however, this was not detected in any of the strains chosen for sequencing, although the entire gene was not often sequenced due to only being able to sequence from one end of *URA3*.

Upon screening 5-FOA resistant colonies, it also became obvious that in some cases 5-FOA resistance could be reversed, as some selected colonies were unable to grow, or grew much slower, upon restreaking. This is reminiscent of epimutations identified in *M. circinelloides* (Calo et al., 2014). Epimutations refer to the silencing of a gene via an epigenetic mechanism. In *M. circinelloides,* epimutations have been shown to be responsible for some cases of FK506 resistance, where resistant cells did not contain a mutation within the *fkbA* gene (Calo 2014). These mutants were also able to revert easily within a few generations, similar to some of the mutants I observed. The epimutations in *M. circinelloides* have been shown to be due to an RNAi response, based on the presence of sRNAs complementary to the *fkbA* gene which confers the FK506 resistance and the absence of epimutations in RNAi deficient strains. Taking this into account, an RNAi-induced epimutation may not account for the unstable resistance seen here, as the RNAi-deficient *rdp1Δ* strain has the lowest percentage of colonies showing stable resistance upon restreaking (78.3%). However, it is possible that an epimutation may be occurring via another epigenetic mechanism. This could be through localised

DNA methylation silencing the gene, as for all of the strains lacking DNA methylation (*dnmt5Δ*, *clr4Δdnmt5Δ* and *rdp1Δdnmt5Δ*) almost every colony was able to grow upon restreaking (99.6%, 100% and 100% respectively for each strain). This could also explain why such a large proportion of colonies were shown to have no insertion, if some of the mutations conferring 5-FOA resistance were not actually altering the gene. Possible ways to screen for epimutations would be to regularly restreak colonies on rich, non-selective media and then regularly replica-plate onto 5-FOA media to see if the resistance is still present. Initial tests were carried out using this method, and one colony was shown to have reverted back to become 5-FOA susceptible after 4 generations. However, due to the number colonies being screened across all eight strains, this method was too time consuming to continue. Other potential methods of determining the presence of epimutations could be to analyse the 5mC levels of *URA3* and *URA5* in a manageable number of colonies from WT and *dnmt5Δ* strains, along with any cultures showing evidence of unstable resistance, to determine if any DNA methylation is present, before confirming if it is forming epimutations.

PCR screening identified a number of insertions within both the *URA3* and *URA5* genes, and for quantification purposes only the unique insertions were considered to ensure that I was only looking at insertion events and not cell replication. Comparison of the proportion of unique TE insertions amongst the colonies screened showed comparable proportions in *clr4Δdnmt5Δ* and *rdp1Δ* strains which were significantly higher than WT. The high proportions were expected in *rdp1Δ* cells due to the previous data linking RNAi to transposon silencing, but no link has previously been shown between heterochromatin formation and silencing of the retrotransposons (Dumesic et al., 2015; Loftus et al., 2005; Yadav et al., 2018). However, H3K9 methylation and DNA methylation occur at the centromeres, and that this is where retrotransposons are

located, therefore it might be expected. However, surprisingly, sequencing results confirmed that the inserts identified within *URA3* and *URA5* in *clr4Δdnmt5Δ* cells showed the presence of only DNA TEs, which largely reside in non-centromeric regions of the genome. Neither H3K9 methylation or DNA methylation have been shown to be linked to DNA transposons, nor have high levels of heterochromatin been identified out with the centromeres or subtelomeric regions. However, this data suggests that there is a link between H3K9 methylation, DNA methylation and silencing of DNA transposons. The combinatory effect also seen in the *clr4Δdnmt5Δ* strain, where the double deletion strain shows a higher mutation rate than either single deletion strains, also implies that H3K9 methylation and DNA methylation function to silence these elements in parallel pathways. H3K9 methylation and DNA methylation are inter-dependent, such that wild-type localisation of both marks relies on the presence of each other (Catania et al., 2020). This fits with this finding, as assuming that both H3K9 methylation and DNA methylation are present at these regions, deleting one or other of the pathway could cause a slight redistribution of the remaining silencing marks allowing some expression to occur, and removal of both would further increase the expression. This suggests a novel role for H3K9 methylation and DNA methylation is silencing DNA transposons.

Difficulties in sequencing prevented confirmation of the identity of all the different inserted products. However, the most frequently occurring elements were identified as T1, and T2. Only one unidentified insertion of 1.8 kb is of the predicted size of a transposon – in this case the non-LTR Cnl1 which has been shown to mobilise in hypermutator strains of C. *neoformans* (Priest, Yadav, et al., 2021). The other unidentified insertions were all smaller in size (≤1 kb), and could be solo LTRs as identified in other strains of C. *deneoformans* (Gusa et al., 2020), or truncated versions of larger previously identified transposons. Consistent with this, sequencing of the 1 kb

insertion showed homology to the full-length T2 transposon. The detection of the short version of T2 mobilisation was unexpected as whilst the full-length T2 transposon contains a putative transposase gene, the T2 short transposon does not encode any predicted genes, and therefore should not have the ability to mobilise by itself. As T2 short was seen inserted into the *URA* genes, the transposase activity must have been from another DNA transposon. The most likely source is full length T2, however as the mobilisation of T2-short insertions was almost exclusively seen in *clr4*Δ*dnmt5*Δ cells it could be from a H3K9/DNA methylation regulated transposon. Based on the DNA transposon expression data, the most likely candidate however is T3 as it is the only transposon which showed an increase in expression in this strain, although this is increase is not significant. The reason why T3 insertions were not identified within the transposition assay could be a bias towards insertion in sites other that *URA3* and *URA5* or a different mechanism of regulation independent of RNAi, DNA methylation and H3K9 methylation.

The types of transposons identified through the PCR screening differed from those seen in a previous study, where T1, T3 and Tcn12 were reported to have mobilised into *URA3* or *URA5* genes (Gusa et al., 2020). However, the previous study used a different strain of C. *deneoformans* which could explain the differences in TEs identified. For example, that mobilisation of Tcn12 was not detected here is not surprising as this element has not been identified in C. *deneoformans* JEC21, although solo LTR12 sequences do exist. However, it is also possible that the genome assembly over the centromeres is incomplete, and that Tcn12 does exist within C. *deneoformans* JEC21 but just does not appear in the reference genome.

Comparison of the transcript levels of the DNA transposons against the insertions identified during the transposition assay shows differences between the findings.

Transcript levels appear to increase in RNAi deficient strains, but no corresponding increase in mobilisation is seen in these strains compared to others. The combined data for T2 is more contradictory, with H3K9 and DNA methylation deficient strains producing the highest number of T2 insertions, a combination of both the full length and short versions. However, no increase was seen in the transcript levels for T2 in these strains. This does however confirm that not all actively transcribed TEs are identified within the transposition assay, as T3 was not seen to mobilise into *URA3* or *URA5* in any strain but does show increases in transcript levels of some strains. However, to draw any significant conclusions from the comparison of the data, the same number of repeats should be carried out for the transcript analysis as the transposon assay due to the large variance seen between individual cultures.

A difference between the *rdp1*Δ and *dcr1*Δ*dcr2*Δ strains is consistently seen throughout the transposition, copy number and expression data. This corresponds with similar previous findings that Rdp1 deletion strains have a stronger effect in activating silenced loci than double dicer deletion strains (Janbon et al., 2010). Here, this difference can be seen in the transposition assay, where the *rdp1*Δ strain has a higher drug resistance rate and a higher rate of transposon insertion than the *dcr1*Δ*dcr2*Δ strain, and in the transcript analysis of DNA transposons where the *rdp1*Δ strain has higher average transcript levels than the *dcr1*Δ*dcr2*Δ strain. However, analysis of the copy number of LTR-retrotransposons shows that the *rdp1*Δ strain has a significantly lower Tcn5 copy number than WT whereas the *dcr1*Δ*dcr2*Δ strain has a copy number comparable with WT. This appears contradictory to the model that RNAi silences retrotransposons, as a lack of silencing would be expected to increase the copy number of each element. However, this data shows a decrease in the copy number in strains lacking Rdp1. A

possible explanation for this could be that so many transposition events have occurred within this strain that the number of full-length copies, or at least the number where the region amplified remaining intact, is lower due to high levels of TE mobilisation and possible recombination. This could occur in a similar way to the predicted loss of full-length retrotransposons within C. *deuterogattii,* where retrotransposons are thought to have inserted into each other producing many retrotransposon fragments but no full-length copies (Yadav et al., 2018). Overall, though, neither RNAi deficient strain showed a consistent difference in transcript levels of DNA transposons or copy number of retrotransposons. Repeating these experiments with cells undergoing meiosis may reveal effects of RNAi not evident during vegetative growth, as seen previously when determining the role of SIS in C. *neoformans* (X. Wang et al., 2010).

In conclusion, this chapter shows that both strains lacking Rdp1 and strains lacking Clr4 and Dnmt5 have increased transposon insertion rates, suggesting a role for both RNAi and heterochromatin formation in the silencing of DNA transposons. Generating a *rdp1*Δ*clr4*Δ*dnmt5*Δ triple mutant strain would be ideal to repeat these experiments, and would help determine if these pathways act individually or if there is a relationship between them both. The potential for Dnmt5 controlled epimutations is a novel concept, but is an exciting possibility within C. *deneoformans* which can serve as the basis for future experiments.

Chapter 5 Investigating the role and interactors of Ago1 and Ago2 within the RNAi pathway

5.1 Introduction

Argonaute proteins play an essential role in canonical RNAi pathways with their ability to bind sRNA, and are present in both the RISC and RITS effector complexes. Two Argonaute proteins are encoded within the C. *deneoformans* genome, whereas closely related species including C. *neoformans* only have one. The fact that both Argonaute genes are conserved within the genome suggests that they are both functional, otherwise evolutionary loss would likely have occurred as seen in other C*ryptococcus* species. This makes C. *deneoformans* an interesting model to study as it is likely that the RNAi pathway functions differently than in other closely related species due to this extra component.

Ago1 has been shown to be the main Argonaute protein required for RNAi-mediated silencing out of the two Argonaute proteins within C. *deneoformans*. This was shown through the expression of an RNA hairpin against the *ADE2* gene, which induces RNAi-mediated silencing of *ADE2* giving an observable phenotype. Although deleting Ago2 had little effect on silencing, deleting both Ago1 and Ago2 gave a stronger defect than deleting Ago1 along, showing that Ago2 is not completely redundant in RNAi function. The two Argonaute proteins share a similar relationship to that of the two Dicer proteins, where Dcr2 is the main Dicer protein required for RNAi but Dcr1 still has a role within the pathway (Janbon et al., 2010).

Although both Argonaute proteins appear to have roles within the RNAi pathway, it is unclear what these are and how they differ from each other. Given that both have been conserved through evolution, it can be assumed that they each have distinct roles, otherwise one protein would be able to carry out the role making the other redundant. Therefore, are some regions silenced preferentially by Ago1 and other by Ago2, or

do the two proteins initiate different silencing mechanisms once the region is targeted? With the majority of studies so far looking at RNAi in the C. *neoformans* H99 strain, it can be hard to translate the findings to C. *deneoformans* where both Ago1 and Ago2 proteins are present. In C. *neoformans,* Ago1 has been shown to form the SCANR complex, involved in recognising stalled spliceosomes which act as a trigger for producing sRNAs (Dumesic et al., 2013). The SCANR complex was shown to localise within the nucleus, which separates the role of Ago1 in the SCANR complex from its other role as part of the PRSC complex which localises to p-bodies within the cytoplasm. Assuming that these complexes also form in C. *deneoformans,* it is hard to predict how Ago2 will fit into this model, although an obvious possibility would be that one Ago protein forms the SCANR complex and the other forms the PRSC complex.

The aim of this chapter is to explore the differences between the two Ago proteins in C. *deneoformans*, both in terms of protein interactors and also sRNA targets. Combining both of these analyses will help distinguish differences in function between the two proteins, and potentially indicate potential roles for both Argonaute proteins within the RNAi pathway.

5.2 Ago1 and Ago2 have different interaction partners

In C. *neoformans* the interacting partners of Ago1 have been identified through mass spectrometry, with Ago1 interacting with Rdp1, Gwc1 and Qip1 as part of the SCANR complex, and with Gwo1 as part of the PRSC complex (Dumesic et al., 2013). I hypothesised that one of the two Argonaute proteins in C. *deneoformans* would form the SCANR complex and one the PRSC complex. This would separate the functions of each Ago protein, as well as the localisation which could explain why both genes are maintained within the genome.

To determine the interacting partners of both Argonaute proteins in C. *deneoformans,* I used strains expressing FTH-tagged Ago1 or Ago2 available in the lab. Immunoprecipitation was carried out using anti-flag beads to bind to the Flag-tag section of the FTH tag, followed by mass spectrometry to identify protein interactors. Initial samples showed high background levels of non-specific protein binding relative to the tagged protein, however this was addressed by adapting the number and length of washes of the beads, changing the salt concentration of the wash buffer and by adding in an extra filtration step. This modified protocol was then used in two repeat experiments where FTH-Ago1, FTH-Ago2 and WT control strains were analysed in parallel. Data analysis was carried out using Max Label-Free Quantification (MaxLFQ) which normalises the proteins in each repeat based on the intensity of the peptides identified (Rgen Cox et al., 2014). The data was filtered to remove contaminants and those proteins with less than one peptide, and the repeats for each IP were averaged. Values for proteins which were missing in one sample compared with the others were replaced with values from the normal distribution so as to keep the standard deviation as small as possible. Volcano plots were generated by plotting the Log_2(Difference between FTH-Ago1 LFQ values and FTH-Ago2 LFQ values) against Log_{10}(p-value) to determine the differences in proteins identified with each Ago IP (Figure 5.1). Statistical significance was determined at a False Discovery Rate (FDR) of 0.5 and s0 (minimal fold change) of 1, as shown on the plot by the curved lines (Goss Tusher et al., 200 C.E.).

Comparison of the protein interactors with a differential enrichment between Ago1 and Ago2 IPs showed only 3 proteins significantly increased in Ago1 IPs, whereas 32 proteins were shown to significantly increase in Ago2-IPs. In both cases this included the tagged Ago1 or Ago2 for each respective IP. The other two proteins enriched in Ago1

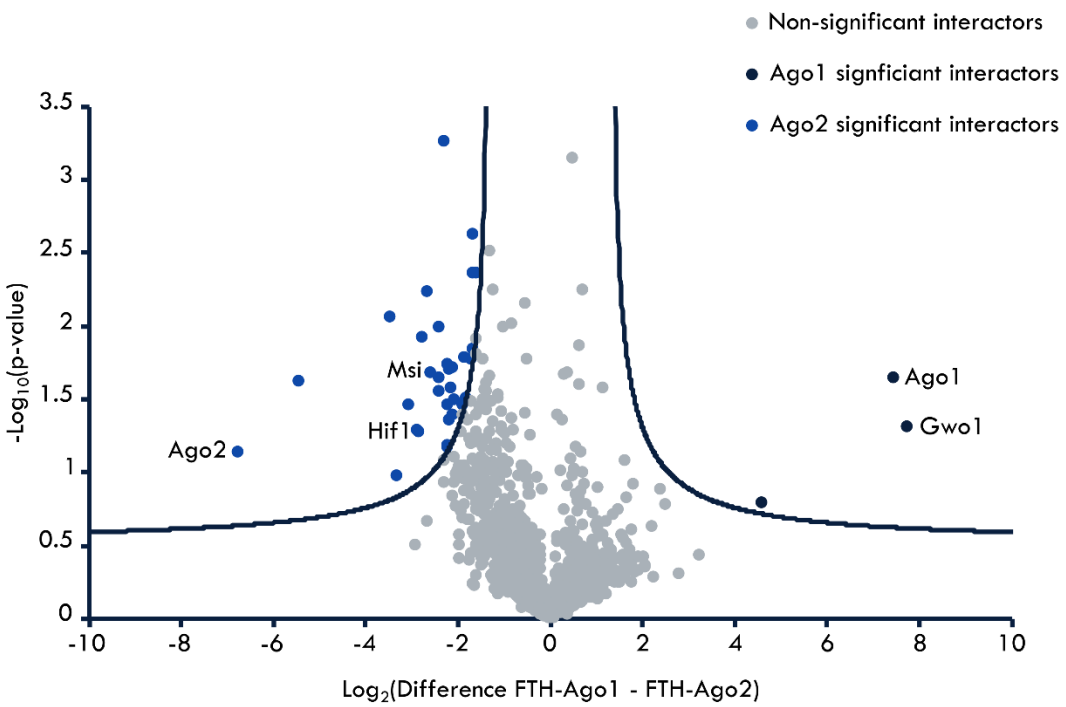

Figure 5.1 Differential interaction of Ago1 and Ago2

Volcano plot showing the Log_2 difference in FTH-Ago1 - FTH-Ago2 normalised LFQ values against the $-Log_{10}$ p-value. Statistical significance is drawn at FDR=0.5, s0=1. n=2 repeats.

IPs include Gwo1, a previously identified interactor of Ago1 forming the PRSC complex in C. *neoformans* (Dumesic et al., 2013). This suggests that Ago1 and Gwo1 form the PRSC complex in C. *deneoformans* as well, and therefore that Ago1 may localise within the p-bodies. The other protein is 6 Phosphofructose kinase (Pfk), which is unlikely to have a functional role in RNAi (Table 5.1). The proteins significantly enriched in Ago2-IPs can be split into seven groups. The largest group contains 14 metabolism related proteins, which are unlikely to be of interest in having a role with Ago2 in RNAi. The same can be said for the group of structural proteins, as these are highly abundant within the cell. The other groups include proteins involved in protein synthesis and degradation, proteins associated with the cell capsule, membrane proteins, a chromatin modulator and one hypothetical protein with no predicted domains which is only conserved in *Cryptococcus* species (Table 5.2). Of these, the proteins which are most of interest include Hif1 which is a Hat1 interacting factor. In *S. cerevisiae,* Hif1 resides in the nucleus and acts as a molecular chaperone for the nuclear type-B HAT complex for acetylation of free histone H4 (NuB4) (Poveda et al., 2004). Another protein of potential interest is the RNA binding protein Musashi, which in *Xenopus* has been shown to have a role in regulating the activation of translation of mRNA (Cragle et al., 2019). This is consistent with the identification of several other proteins involved in protein synthesis including ribosomal proteins, and could suggest a role of Ago2 in silencing RNAi targets by regulating translation of mRNA.

Overall, this shows that Ago1 and Ago2 interact with different binding partners, which could lead to different roles for each protein within the RNAi pathway. Ago1 was shown to interact with Gwo1, a previously identified RNAi factor and Ago1 interactor in C. *neoformans,* suggesting that the presence of the PRSC complex and its role in RNAi is conserved within C. *deneoformans*.

Table 5.1 Proteins with a significant differential increase in FTH-Ago1 IP-MS compared with FTH-Ago2 IP-MS

Gene ID	Gene name	Putative function and domains	Known RNAi?
CNJ00490	**Ago1**	**Argonaute 1 with PAZ and PIWI domains**	**Yes**
CNA05020	Gwo1	GW/WG dipeptide and zinc-finger containing protein with DNA pol-III subunits gamma and tau and tegument homology	Yes
CNJ01080	Pfk	6-Phosphofructose Kinase	No

Table 5.2 Proteins with a significant differential increase in FTH-Ago2 IP-MS compared with FTH-Ago1 IP-MS

Gene ID	Gene name	Putative function and domains	Known RNAi?
CNJ00610	**Ago2**	**Argonaute 2 with PAZ and PIWI domains**	**Yes**
CNC03980	Kynu	Kynureninase	No
CNC05850	Rfk	Riboflavin kinase	No
CNB01250	Sdr1	Short chain dehydrogenase	No
CNK01760	Ogf	Hypothetical protein with opioid growth factor receptor domain and homology	No
CNB05610	Hif1	Expressed protein with SHNi-TPR domain and homology to HAT-1 interacting factor 1	No
CND01080	Ctr	Copper uptake transporter	No
CNC02410	Smo	C-4 methyl sterol oxidase	No
CNC05040	Erp	Endoplasmic reticulum protein	No
CNN01940	Msi	Hypothetical protein with two RNA recognition motifs and homology to mushashi	No
CND02230	Sdr2	Hypothetical protein with SDR oxidoreductase domain	No
CND01340	Gst	Glutathione transferase	No
CNN00800	Sec	Hypothetical protein with Sec61 domain transmembrane helix	No
CNC06260	Znd	Zinc binding dehydrogenase	No

Gene ID	Gene name	Putative function and domains	Known RNAi?
CNA05830	Cap1	Capsular associated protein with mannosyltransferase domain	No
CNA02550	Idi	Isopentyl-diphosphate delta isomerase	No
CNA06560	UreG	Urease accessory protein	No
CND04400	Rpl17	Hypothetical protein with homology to ribosomal protein L17	No
CNF01800	Cda	Chitin deacetylase	No
CNK00110	Fbl	SAM methyltransferase with homology to fibrillarin	No
CND02720	Arp1	Hypothetical protein with homology to ARP2/3 complex 20 kDa	No
CNB03530	Twf	Protein tyrosine kinase with twinfilin domain	No
CND03100	Nef	Adenyl nucleotide exchange factor	No
CNA02900	Hyp	Hypothetical protein with no domains, *Cryptococcus* specific protein	No
CNC02805	Rpl31	Hypothetical protein with homology to ribosomal protein L31	No
CNJ01090	Xdh	Xylitol dehydrogenase	No
CNA07130	Ssd	Succinate semialdehyde dehydrogenase	No
CNA07000	Cap2	Capsular associated protein with mannosyltransferase domain	No

Gene ID	Gene name	Putative function and domains	Known RNAi?
CNJ01950	Arp2	Arp2/3 complex 21 kDa subunit	No
CNF03970	Bis	Biotin synthase	No
CNE03880	Ube	Hypothetical protein with homology to ubiquitin conjugating enzyme E2	No

5.3 Deletion of Gwo1 increases Ago1-Ago2 interaction

Since Gwo1 was identified as the main interactor of Ago1 in C. *deneoformans,* and it has previously been shown to be involved in RNAi, I decided to explore this interaction further. Previous studies in C. *neoformans* have shown that deletion of *GWO1* decreases sRNA levels, although not to the same extent as deletion of other RNAi components such as Ago1 and Rdp1 (Dumesic et al., 2013). In C. *neoformans,* Ago1 was shown to interact with Gwo1 to form the PRSC complex which is independent from the other Ago1 protein complex, SCANR (Dumesic et al., 2013). However, no other proteins were identified within the PRSC complex. I hypothesised that deleting Gwo1 would alter the binding partners of Ago1. With Gwo1 apparently the main Ago1 interactor, deleting Gwo1 could destabilise other interactions, which would allow for identification of proteins either in a complex with both Ago1 and Gwo1, or which rely on the presence of Gwo1 for their interaction with Ago1. This exclusion mass spectrometry could help identify which other Ago1 interactors are also involved in the RNAi pathway.

Gwo1 deletion was carried out through the insertion of *URA5* into the *GWO1* gene via CRISPR-Cas9. Gwo1 deletions were made in FTH-Ago1 strains, and upon confirmation of the disruption, the FTH-Ago1 IP mass spectrometry was repeated using the deletion strain. Two repeats were carried out, with wild-type FTH-Ago1 being run in parallel as the control. The same analysis was carried out as above, and volcano plots were generated plotting the Log_2(Difference in FTH-Ago1 LFQ values and FTH-Ago1.*gwo1*Δ LFQ values) against Log_{10}(p-value) (Figure 5.2). Again, statistical significance was indicated by the curve representing FDR=0.5 and s0=1. The mass spectrometry confirmed the successful disruption of *GWO1* as Gwo1 was absent in the IP from the deletion strain.

Initially I looked at the proteins which were reduced in the FTH-Ago1.*gwo1*Δ IP. This surprisingly included Ago1, which suggests that Gwo1 may act to stabilise this protein and therefore the loss of stability may result in the loss of other Ago1 interactors. This is seen for the other identified Ago1-specific interactor, Pfk, which was also significantly reduced in the FTH-Ago1.*gwo1*Δ IP. The other two proteins with a significant reduction are involved in metabolism and not considered to be of interest. However one other protein was found to have a large (but not significant) difference between FTH-Ago1 IP and FTH-Ago1.*gwo1*Δ IP - this is a Dhx RNA helicase which is uniquely identified in FTH-Ago1 IP, but not FTH-Ago2 IP, and lost upon the deletion of Gwo1. Although this difference is not significant, further repeats ($n \geq 3$) of this mass spectrometry analysis would decrease the variance, potentially bringing it into significant range.

Only one protein was shown to increase interaction with Ago1 upon deletion of Gwo1, and surprisingly this was Ago2. In strains containing Gwo1, Ago1 and Ago2 show only weak interaction, with only the tagged protein in each IP being significantly increased. The substantial increase in Ago2 enrichment in Ago1 IPs upon deletion of Gwo1 could suggest that Ago2 interacts with Ago1 in a Gwo1-independent manner, and the absence of Gwo1 results in an increase of this interaction as less Ago1 is sequestered into the PRSC complex. Another alternative reflects the possible loss in stability of Ago1. If this loss in stability results in a loss of function, the increased Ago2 interaction could act as a backup by providing a functioning Argonaute protein to the Ago1 pathway and complexes. As no other Ago2 interactors are shown to have a significant increase upon deletion of Gwo1, it may suggest that Ago2 joins the Ago1 complexes rather than Ago1 joining the Ago2 complexes.

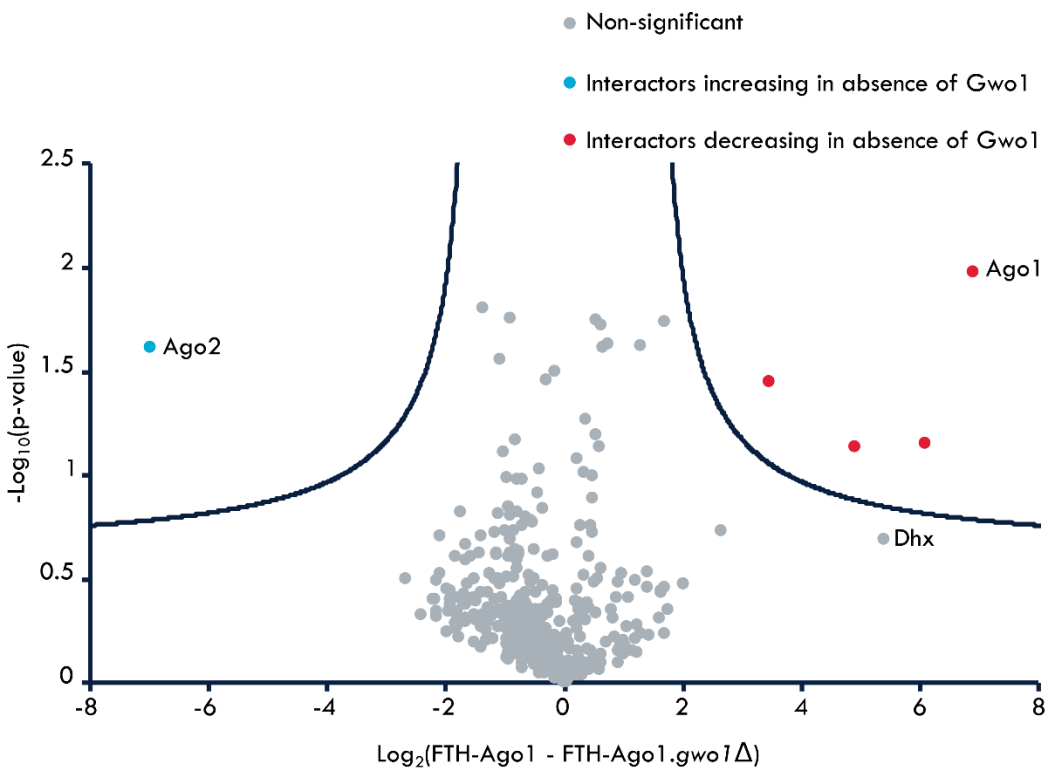

Figure 5.2 Differential association of interactors with Ago1 upon deletion of Gwo1

Volcano plot showing the Log_2 difference in FTH-Ago1 and FTH-Ago1.*gwo1*Δ normalised LFQ values against the $-Log_{10}$ p-value. Statistical significance is drawn at FDR=0.5, s0=1. n=2 repeats.

Overall, this data shows that upon deletion of Gwo1, there are some changes in Ago1 binding partners. It also suggests that Gwo1 may be required for Ago1 stability, and further investigation via Western blot should be conducted to determine if this is true. The inferred loss in stability may account for the loss of some Ago1 interactors, with Dhx1 being identified as a protein of interest potentially linked to RNAi due to the loss of Ago1 association upon Gwo1 deletion. An increase in Ago2 association with Ago1 is also observed in the absence of Gwo1, suggesting that the interaction of two Argonaute proteins is Gwo1-independent.

5.4 More sRNA loci specifically associate with Ago2 than Ago1

As the mass spectrometry analysis identified differences in the protein interactors for each Argonaute protein, this suggests that each may have a different role. Argonaute proteins bind sRNAs, which in turn target Ago and its interactors to specific loci for silencing. With potentially different roles for each Argonaute protein, I hypothesised that there may be differences in the sRNAs associating with each Ago which could determine whether loci are silenced via Ago1 or Ago2 dependent methods.

For this analysis, I used the Ago1-IP and Ago2-IP sRNA sequencing dataset available in our lab, which I previously used in chapter 3 to identify RNAi target loci. The raw read counts for each sRNA generating locus were corrected for the length of each target locus, generating a count of reads per bp for each locus. For each locus, the fold change in read count between Ago1-IP and Ago2-IP was calculated, for three independent experiments. The p value was calculated for each locus, and a volcano plot was created by plotting the average Log_2(Fold-change) against the -Log_{10}(p-value) (Figure 5.3). Cut offs were applied at a p-value of ≤ 0.05, and a fold change

≥100. Using these thresholds, there were six sRNA loci which showed strong and significant enrichment in Ao1-IPs, and 71 that showed enrichment in Ago2-IPs. This shows that there are differences in sRNAs associating with Ago1 and Ago2, and that Ago2-specific sRNA loci are more common than Ago1-specific sRNA loci. This is also represented in the whole data set where 174 sRNA loci were more enriched in Ago1-IPs whilst 248 sRNA loci were more enriched in Ao2-IPs.

Genome database searches confirmed the genomic locations of each of the identified sRNA loci, along with any coinciding genes or predicted transposable elements. Out of the six Ago1 enriched sRNA loci, three were over predicted protein coding genes, whilst one is centromeric and two are in the rDNA repeats on chromosome 2. However, one of the important things to note is that the genome annotation of *C. deneoformans* is not complete, and most coding regions have only been predicted. This means that sometimes protein coding genes are predicted over regions which correspond to transposable elements, and many TEs are unannotated. This is the case for one of the predicted genes here, where further analysis revealed that it corresponds to the non-LTR retroelement Cnl1. Similarly, both of the loci in the rDNA repeats region correspond to the transposon T1. The centromeric sRNA locus does not appear to map to a retrotransposon, although the mapping of the centromeric retroelements may not be complete.

Looking at the locations of the 71 Ago2-enriched sRNA loci showed that the majority were either centromeric (29/71) or over non-coding regions (27/71). BLAST analysis on a sample of these non-coding regions revealed that some contained unannotated transposable elements including crypton Cn1 and the non-LTR retroelement Cnl1. Other non-coding regions where no TE homology was found often appeared either immediately upstream or downstream of a neighbouring coding region. Those that were centromeric were most likely to correspond to predicted Tcn3 elements, even though

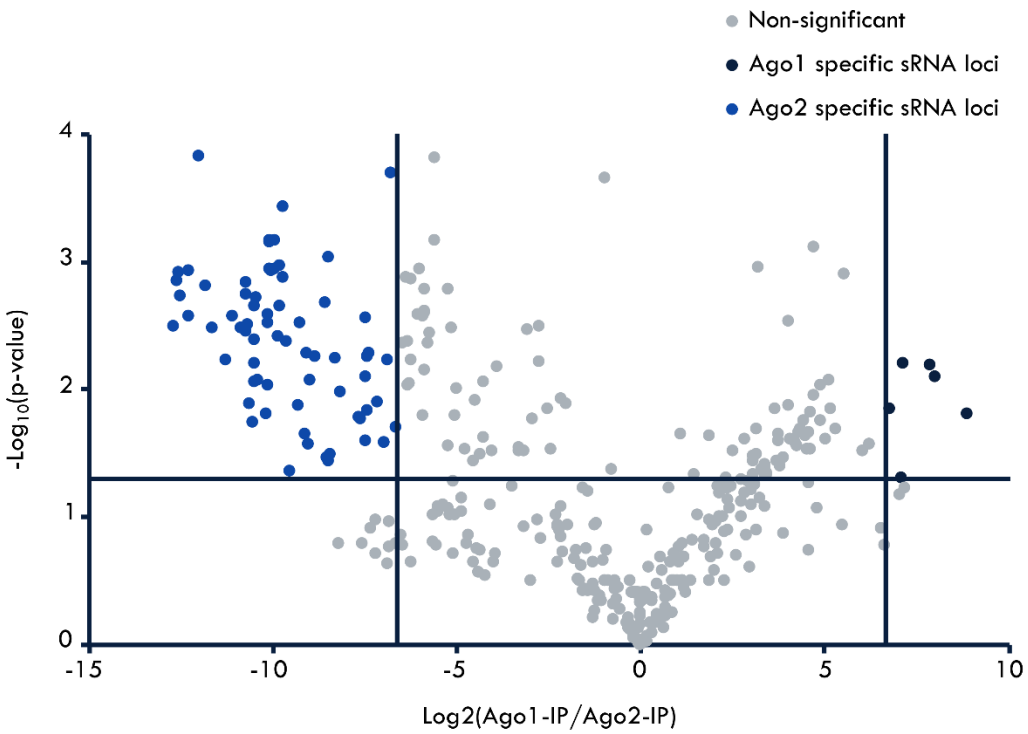

Figure 5.3 Differential enrichment of sRNAs in Ago1-IP and Ago2-IP
Volcano plot showing the Log_2 fold change in Ago1-IP sRNAs over Ago2-IP sRNAs against the $-Log_{10}$ p-value. Statistical significance is drawn at p=0.05 and fold-change=100. n=3 repeats.

Tcn5 and Tcn6 are the most prevalent within the centromeres. Of the other sRNA loci identified, four were telomeric and eleven corresponded to coding regions. Telomeric sRNAs are likely to coincide with Cnl1 elements. Several of the sRNA loci in coding regions also mapped to an annotated Harbinger TE as well as the protein-coding gene – these were counted as coding, though it is unclear whether the predicted protein is related to the function of Harbinger or not.

Overall, this comparison shows that there is some specificity in sRNA binding by Ago proteins, with more Ago2-specific loci than Ago1-specific loci. In both cases, sRNAs map to different structural regions of the genome, though with a bias towards areas with predicted transposable elements. Although this does not identify any reason why some sRNAs are Ago1-associated and some are Ago2-associated, it does suggest that Ago2 has an important role in the RNAi pathway at least at some loci.

5.5 Discussion

This chapter aimed to try to distinguish different roles for Ago1 and Ago2 within the RNAi pathway. With both proteins present within the C. *deneoformans* genome, and neither redundant, the aim was to understand how the role of Argonaute within the species is split over the two proteins. Understanding the individual roles of both Ago1 and Ago2 within the RNAi pathway will aid further understanding of the RNAi pathway itself, and help understand why both proteins have been evolutionarily conserved in C. *deneoformans* when Ago2 has been so frequently lost in related species.

Initial work focused on identifying protein interactors of both Ago1 and Ago2 to see if the two proteins either work together, binding the same interactors or if they have a different subset of protein partners which could help determine the function of each Argonaute. Mass spectrometry analysis identified interactors of both Ago1 and Ago2.

However, determining those that were significantly enriched in one IP over the other required the use of higher than optimal FDR and s0 values. These were used over the standard values due to only two repeats being carried out, and these repeats being run independently which could introduce additional variation due to potential differences before and after mass spectrometry. By repeating mass spectrometry experiments using 3 repeats, carried out simultaneously, the standard deviation would be reduced, thus the significance level would be able to be lowered whilst still detecting the same differentially expressed proteins.

Of the proteins identified as interacting specifically with Ago2, many had metabolic functions and are unlikely to be of interest. Such proteins would often be removed during the analysis; however this was not done here as a lot of the C. *deneoformans* gene annotations are still hypothetical, and as such screening them for function would be time consuming. This was therefore carried out after the enrichment analysis allowing the focus to be on those with high fold changes. Proteins were considered to be unrelated to RNAi if they were metabolic-related proteins or structural proteins as all of the above are thought to give non-specific hits due to their relatively high abundance within the cell. Due to the non-specific binding of Ago1 to the anti-FLAG beads, I was unable to discount all proteins which were detected within the control IP from WT cells, and therefore it is possible that proteins of interest with a small fold-change are missed in this method. However, having removed potentially non-specific abundant proteins, I identified several potential specific interaction partners for both Ago1 and Ago2.

Ago1-IP showed strong enrichment for the previously identified Gwo1 protein, a known interactor of Ago1 in C. *neoformans* (Dumesic et al., 2013). Neither Ago1 or Ago2 showed a significant interaction with any of the three components of the SCANR complex – Rdp1, Qip1 and Gwc1 (Dumesic et al., 2013). This could suggest that the

complex is not prevalent in C. *deneoformans,* or that the complex is destabilised during the IP protocol. Gwo1, along with the members of the SCANR complex, are all part of the 14 gene RNAi network polymorphism identified within the C. *neoformans/gattii* species complex (Feretzaki et al., 2016). As Gwo1 was identified here as an Ago1 interactor, I checked whether any of the other proteins in the polymorphism were interacting with Ago1 or Ago2. None of the others were present in the mass spectrometry data apart from Qip1, Rdp1 and Gwc1, which as previously mentioned, were not detected consistently or at high frequency in either Ago IP.

Of the Ago2-specific interactors, one of the most interesting identified interactors was Hif1. Hif1 is a component of the Nu4B complex which acetylates free histone H4 before it is integrated into chromatin (Poveda et al., 2004). This complex has been shown to be nuclear in *S. cerevisiae,* suggesting that at least a subset of Ago2 is nuclear, contrasting with Ago1 where at least a subset is likely to be localised to p-bodies (Dumesic et al., 2013; Poveda et al., 2004). The identification of a protein involved in a chromatin modification complex is of interest, as it links the role of RNAi back to chromatin modifications, even if it isn't directly through H3K9 methylation as previously predicted. In *S. cerevisiae*, the other proteins which make up the Hif1-containing Nu4B complex, Hat1 and Hat2, have been shown to be involved in telomeric silencing. Nu4B complex acetylates lysine residue 12 on histone H4, and it is this acetylation that has been shown to mediate telomeric silencing (Kelly et al., 2000). It has also been shown that loss of Hif1 in *S. cerevisiae* causes an over expression of histone H3 (Dannah et al., 2018). If the role of Hif1 is also altered when Ago2 is deleted, this could explain the significant increase in H3K9me2 identified in chapter 3 in *ago1Δago2Δ* strains, as it could increase the number of H3 histones available to be methylated. Overall, this interaction with Hif1 suggests a possible role for Ago2 in telomeric silencing, and

indicates that this role in telomeric silencing may be through chromatin formation and transcriptional regulation instead of the PTGS predicted at centromeres and other sRNA target sites.

Another interesting Ago2-interactor was Msi. Msi is a Musashi family protein with two RNA binding motifs. Two Musashi factors in *Xenopus* have been shown to be involved in both repression and activation of mRNA translation through interaction with other factors at the poly(A) tail of mRNA (Cragle et al., 2019). This protein could therefore have a role in PTGS by Ago2, by regulating the translation of target mRNAs. However, it is also possible that this protein is controlling the levels of Argonaute proteins themselves. Studies in *D. melanogaster* have shown that the expression of Piwil1, a PIWI domain containing protein involved in piRNA-mediated regulation, is translationally regulated by Murashi-2 (Sutherland et al., 2018). Both Argonaute proteins contain a PIWI domain, and therefore it is possible that a similar regulation of Argonaute is occurring here.

Deletion of Gwo1 provided unexpected results through the apparent destabilisation of Ago1 and the increase in Ago1 interaction with Ago2. The only other significant specific interactor of Ago1 was Pfk, which also decreased in enrichment, mirroring the decrease in Ago1 upon deletion of Gwo1. This suggests that Ago1 may be unable to form its normal complexes without the presence of Gwo1 to stabilise it. This is surprising, as in C. *neoformans* Ago1 was previously found within the SCANR complex, which is a Gwo1-independent complex (Dumesic et al., 2013). Therefore, either the stability of Ago1 is different between C. *deneoformans* and C. *neoformans,* or the components of the SCANR complex are able to stabilise the protein. This could be the case, as another GW/WG containing protein other than Gwo1 is present in the SCANR complex - Gwc1. Whether the stability of Ago2 depends on its interactions with other proteins is

yet to be confirmed. However, the increase in Ago2 interaction with Ago1 upon deletion of Gwo1, suggests that Ago2 may be able to compensate for the loss in function of Ago1. An alternative possibility could be that a loss of Gwo1 means that more Ago1 is available to interact with Ago2 due to the redistribution of Ago1 from the disbanded PRSC complex.

The other protein identified as potentially of interest from the Gwo1 deletion analysis is the DHX helicase. Although not significant, the large decrease in enrichment of this protein in Ago1-IPs in the absence of Gwo1, along with the role of Dhx helicases in RNA related processes, makes it a protein worth investigating further. Dhx helicases have a D-E-X-H catalytic motif in the helicase domain, and this protein also contains two dsRNA binding motifs and N-terminal tegument homology. BLAST analysis of this protein shows that the full-length protein is only evolutionarily conserved in *Cryptococcus* species, although the helicase domain and dsRNA binding motifs are conserved in P-loop containing nucleoside triphosphate hydrolase (NTPase) proteins, which include a diverse group of proteins including DHX helicases (Pathak et al., 2014). Without a well-studied homologous protein, it is hard to predict the function of Dhx1. However, DHX helicases have NTPase activity which isn't specific to ATP, and various DHX helicases are involved in translation initiation and regulation of G4-containing mRNAs (Shen & Pelletier, 2020). Overall, this suggests that Ago1, via interaction with this RNA helicase could play a role in translational control of sRNA targets. It is also possible that this helicase assists Dicer activity, as neither Dcr1 or Dcr2 contain the evolutionarily conserved DEAD-box helicase domain alongside their other domains. Whereas other species such as the slime mould *Dictyostellium* have been shown to have functional helicase domains incorporated in two of the Rdp homologs (Martens et al., 2002), in C. *deneoformans* the role of the helicase could be through an interacting partner.

Analysis of sRNA populations using the sRNA datasets available in the lab showed some specificity in binding by Ago1 or Ago2, consistent with distinct function. Surprisingly, more sRNA loci showed preferential association with Ago2 than with Ago1. This contradicts the prior knowledge that Ago1 is essential for RNAi, and that Ago2, although is not redundant in function, only has a minor role in the pathway. However, as this knowledge is based on analysis of RNAi at one locus in an artificial system, it is possible that it does not represent the role of Ago1 and Ago2 in endogenous silencing across the whole genome. Therefore, further exploration of the role of Ago1 and Ago2 is required, looking at the silencing ability of each Argonaute protein at RNAi target sites across the whole genome, to determine if this dependency for Ago1 silencing is irrespective of the distribution in favour for Ago2-associated loci. Mapping of the sRNA loci preferentially associated with Ago1 or Ago2 showed a large proportion of both mapping to predicted transposable elements, which fits with the knowledge that RNAi silences TEs. However, other regions including coding regions and non-coding regions were identified, suggesting that RNAi also functions to silence other targets alongside TEs.

Overall, this data suggests that RNAi mechanisms differ between C. *deneoformans* and C. *neoformans* due to the presence of Ago2. Whilst Gwo1 was identified as Ago1 interacting, the SCANR complex was not, suggesting that this may be a complex which was derived to compensate for the loss of Ago2. Ago2 may silence different loci through different mechanisms, with a potential for chromatin-modulation silencing occurring at the telomeres alongside PTGS RNAi silencing via translation regulation. The increase in interaction between Ago1 and Ago2 when Gwo1 is deleted suggests that Gwo1 may have a role in separating the two RNAi pathways. The differences in sRNA abundance in Ago1-IPs versus Ago2-IPs also suggests a specific role for Ago2 within

the RNAi pathway. Further exploration into the role of the identified proteins Hif1, Msi and Dhx1 and the stabilisation Ago1 should be considered to help elucidate further the different roles for each Argonaute protein within C. *deneoformans,* and determine if these factors newly identified factors are required for a functional RNAi pathway.

Chapter 6 Discussion

6.1 RNAi, H3K9 methylation and DNA methylation – is there a link?

Whilst other species in the C. *neoformans/gattii* complex have lost components of epigenetic silencing pathways, C. *deneoformans* has preserved the core machinery for the RNAi pathway, including Ago2 which is most frequently lost within the complex (Janbon et al., 2010). It has also preserved the methyltransferases for H3K9 methylation and DNA methylation. RNAi in C. *deneoformans* has been shown to function during meiosis in the form of SIS, and during vegetative growth in the form of asexual co-suppression (X. Wang et al., 2010, 2012). These both rely on repetitive sequences to induce the silencing pathways, and have been shown to have a role in protecting the genome against transposable elements (Janbon et al., 2010; X. Wang et al., 2010). This includes retrotransposons, which are primarily found within the centromeres (Loftus et al., 2005). The centromeres are also known to be the location of DNA 5mC methylation, which coincides with the regions of H3K9 methylation in C. *neoformans* (Dumesic et al., 2015; Yadav et al., 2018). One of the main aims of this thesis was to explore potential interactions of these three epigenetic silencing mechanisms, particularly focussing on RNAi and H3K9 methylation due to the known link in *S. pombe* (Ivanova et al., 1998; Volpe et al., 2002).

H3K9 methylation and DNA methylation have been shown to be linked in C. *neoformans,* where both H3K9me2-ChIP and bisulphate sequencing has been carried out in deletion strains (Catania et al., 2020). Dnmt5, the sole DNA methyltransferase, only has maintenance activity and cannot re-establish *de novo* 5mC after it has been lost. DNA methylation is reduced in strains lacking the H3K9 methyltransferase Clr4, and also Uhrf-1, with a combinatory effect showing a greater reduction in DNA methylation. It

was suggested that Dnmt5 is recruited by H3K9 methylation, as well as by Swi6, which also interacts with H3K9 methylation, and that Uhrf-1 recruits Dnmt5 to hemi-methylated DNA. However, even with deletion of both Clr4 and Uhrf-1, a small proportion of DNA methylation is still present, and this is still localised over the centromere. This suggests that another method of recruitment for Dnmt5 to the centromeres exists. In a similar way that loss of Clr4 alters DNA methylation, the loss of Dnmt5 causes a redistribution of H3K9 methylation, with centromeric methylation reduced and subtelomeric methylation increased (Catania et al., 2020). Reintroducing Dnmt5 does not restore DNA methylation, and nor does it restore the WT distribution of H3K9 methylation. This suggests that Clr4 is recruited, at least in part, to the centromeres by 5mC, either directly or indirectly. However, similar to 5mC, although it is redistributed in the absence of Dnmt5, the H3K9 methylation is still localised over the centromeres and sub-telomeres and does not spread across the whole chromosome, suggesting that another mode of targeted recruitment exists alongside potentially 5mC. The method of Clr4 recruitment is currently unknown, as whereas Clr4 homologs in *S. pombe* and *D. melanogaster* have a chromodomain, which can direct themselves to heterochromatic regions (Ivanova et al., 1998; Tschiersch et al., 1994), this is missing in the C. *deneoformans* Clr4 homolog. Although an interacting protein could provide this required function to target Clr4 to heterochromatin, it has yet to be identified.

The link between RNAi and H3K9 methylation has previously been established in *S. pombe,* where RNAi is required for heterochromatin formation (Volpe et al., 2002). The link had also previously been explored briefly in C. *neoformans,* and it was concluded that there was no link between the two pathways, however only one locus, Tcn1, was analysed (X. Wang et al., 2010). The H3K9me2-ChIP-qPCR that I carried out confirmed similar findings in C. *deneoformans,* and showed that centromeric H3K9 methylation

does not rely on RNAi as WT levels were seen in *dcr1Δdcr2Δ* double deletion strains. I also showed that sRNA targets in non-centromeric regions are not H3K9 methylated. Interestingly, the levels of H3K9 methylation at Tcn1 have been shown to reduce during meiosis, which corresponds with an increase in expression of Tcn elements in RNAi deficient strains, which could suggest different modes of regulation for different stages of the life cycle (X. Wang et al., 2010).

However, having concluded that there was no link between H3K9 methylation and RNAi, the transposition assay suggests that heterochromatin may contribute to silencing similar loci to RNAi. Although the *clr4Δ* strain doesn't show a significant difference in transposon insertion rates, *clr4Δdnmt5Δ* strains do show increased transposon insertion rates compared to WT, similar to *rdp1Δ* strains. This suggests that both RNAi and H3K9 and DNA methylation, potentially working together, have a role in silencing transposons. This is particularly linked to DNA transposons, as they were the only type of transposon identified through the assay. DNA transposons localise in non-centromeric regions and have not been shown to coincide with H3K9 or DNA methylation. This surprising increase in mobilisation was not however reflected in the transcript levels of the MULE DNA transposons or the copy number of the Tcn retrotransposons. However, assuming that the empty PCR products most frequently seen in *rdp1Δclr4Δ* and *rdp1Δdnmt5Δ* strains are retrotransposons, which are too large to amplify, suggests that a loss of RNAi and either H3K9 methylation or DNA methylation has greater effects than just the loss of one of these pathways or the loss of both methylation marks alone. Overall, this suggests that both H3K9 and DNA methylation as well as RNAi control DNA transposon mobilisation, and potentially retrotransposon mobilisation. This control of retrotransposons is likely to be through two parallel pathways, due to the combinatory effects seen.

The interaction of Ago2 with Hif1, a chaperone of the nuclear type-B Hat1 complex for acetylation of free histone H4, suggests that there is a link between an RNAi pathway and chromatin modifications. Although the acetylation of H4 is linked with telomeric silencing, there are small domains of sub-telomeric H3K9 and DNA methylation in C. *deneoformans*. This chromatin regulation of telomeric regions could also have a link to transposon silencing, due to the presence of the non-LTR retrotransposon Cnl1 residing in the telomeres. It is possible that altering the epigenome can initiate H3K9 methylation at this region, providing a link between the two pathways.

6.2 Epimutations and Non-canonical RNAi

The transposition assay also revealed some cases of reversible drug resistance, suggesting possible occurrence of epimutations. It was hypothesised that these may occur via DNA methylation, based on the absence of reversible drug resistance in all strains lacking Dnmt5. This would mean that all strains with a functional Dnmt5 have a proportion of the drug-resistance colonies resistant due to epimutations. As these would be absent in *clr4*Δ*dnmt5*Δ strains due to the lack of Dnmt5, it would suggest that a lower rate of drug resistance should have been seen, though this was not the case. If mutations cannot occur by one means, it would suggest that fewer mutations would be expected overall, or at least that the proportions would differ as one mechanism has been removed. However, this was not the case. Similar proportions of colonies with inserts were found in *clr4*Δ*dnmt5*Δ and *rdp1*Δ strains. Therefore, if Dnmt5-dependent epimutations are occurring, something else is also happening to increase the drug-resistance rate in *clr4*Δ*dnmt5*Δ strains, such as increased occurrence of another type of mutation. This type of mutation could also explain the high drug resistance but low TE transposition rate in *dnmt5*Δ strains.

Epimutations are a relatively new field, but an RNAi linked epimutation pathway has been shown to exist in *M. circinelloides* (Calo et al., 2014). Whilst the potential epimutation identified here looks to be Dnmt5-linked and not RNAi-linked, what is interesting in *M. circinelloides* is the relationship between RNAi-dependent epimutations and a non-canonical RNA induced pathway (NCRIP) mechanism which acts in parallel and has been shown to regulate the canonical RNAi pathway responsible for epimutations (Calo et al., 2017; Trieu et al., 2015). The NCRIP has also been shown to have a role in silencing retrotransposons within the species (Pérez-Arques et al., 2020). One of the main points to take from the research on RNAi pathways in *M. circinelloides* is that RNAi regulation is often very complex, and here there are two inter-linked pathways with one regulating the other.

The example from *M. circinelloides* highlights the possibility of more than one RNAi pathway acting within C. *deneoformans*. It has been well observed, in both this study and previous studies, that *rdp1*Δ strains usually have a greater defect in silencing transgenes and TEs than *dcr1*Δ*dcr2*Δ or *ago1*Δ*ago2*Δ strains, when RNAi is expected to be equally lost in all strains (Janbon et al., 2010; X. Wang et al., 2010). A reason for this could be that Rdp1 acts in two RNAi pathways whilst the other components only belong to one. It has been briefly suggested that a non-canonical RNAi pathway may exist in C. *deuterogattii,* which only has a functional *DCR2* gene out of the core RNAi machinery, possibly explaining why this gene alone has been evolutionarily conserved whilst the other have not (Ferrareze et al., 2017). This could also explain the presence of sRNAs in the species which could act as regulators for retrotransposons, even if these are truncated versions. It is therefore possible that a form of NCRIP does exist within C. *deneoformans*. It is likely that this would be a Dicer independent pathway, due to consistent differences seen between the *rdp1*Δ and *dcr1*Δ*dcr2*Δ strains in the transposon

assay, though the involvement of Argonaute proteins is unknown as Ago deletion strains were not included in the transposition assays. If a NCRIP pathway was to exist acting at the centromeres, it could be that this pathway interacts with heterochromatin formation. This could be a direct interaction, as H3K9me2 levels were tested only in Dicer and Argonaute deletion strains, and not an *rdp1*Δ strain. However, it could also be an indirect effect, with each pathway contributing to the recruitment of the others. There being some interaction would be consistent with the co-localisation of the silencing pathways, and explain why the absence of both can initiate transposon movement. The known and proposed interactions are shown in Figure 6.1.

6.3 Future directions

Regardless of whether there is a NCRIP pathway or not, a link between the three silencing pathways is still likely, especially at the centromeric regions. Both Dnmt5 and Clr4 have other unidentified recruiters to the centromeres based on the targeted localisation even when the other has been lost, and it is possible that this could involve RNAi directly or indirectly. One of the unknowns with Clr4 is how it localises to regions of heterochromatin without a chromodomain. Within the 14 gene RNAi network polymorphism, one chromodomain containing protein was identified – Cdp1 (Feretzaki et al., 2016). As it is part of the polymorphism it is thought to have a role in RNAi, and the chromodomain function and H3 binding site could relate to a function in recruiting Clr4 to heterochromatin. It is also possible that the chromodomain functions to recruit RNAi machinery to heterochromatin regions, but either way it could form the link between the RNAi and heterochromatin pathways. Studying the H3K9 methylation levels in cells lacking Cdp1, along with the efficiency of RNAi silencing via transposition assay, would be worth investigating.

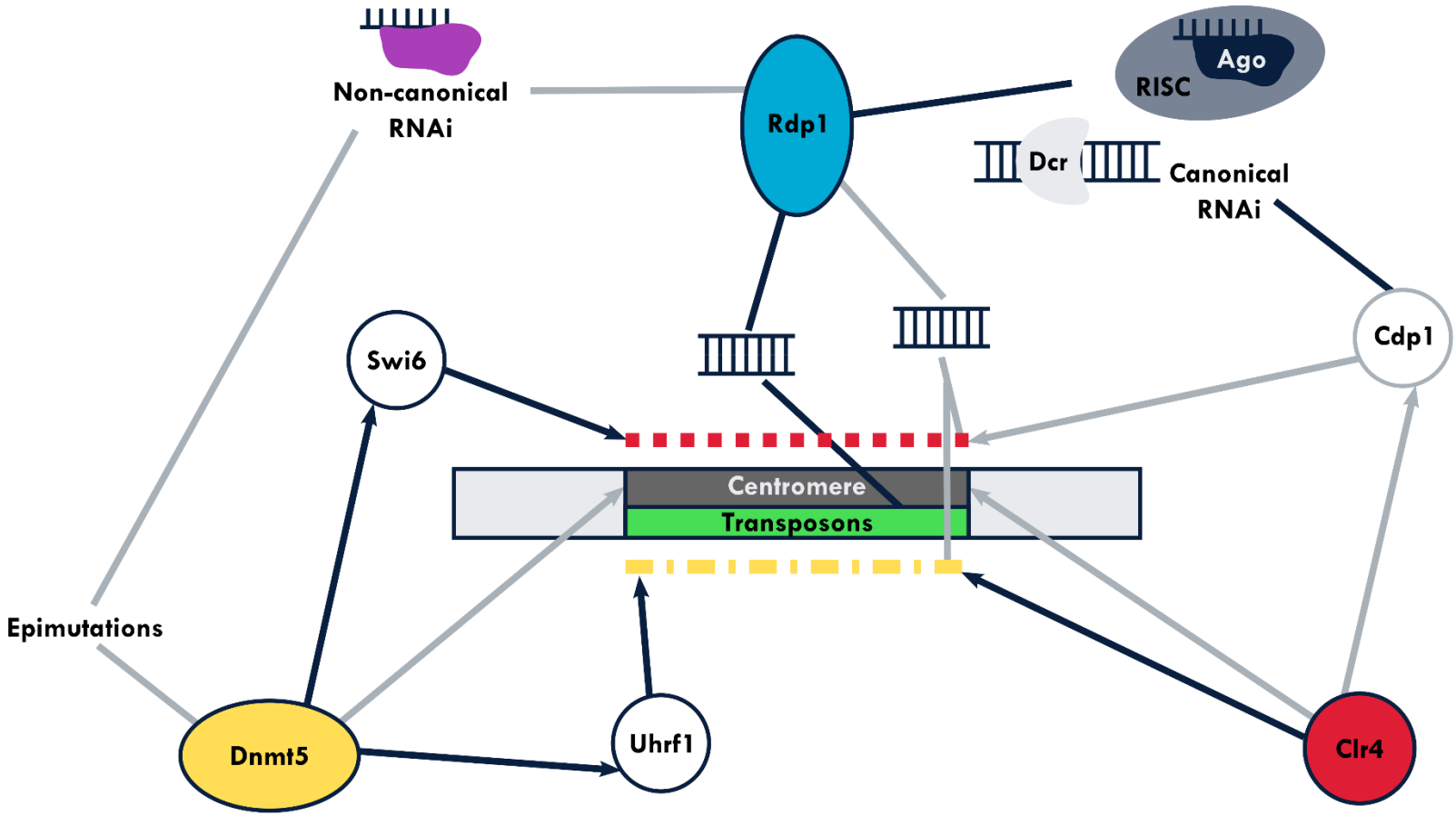

Figure 6.1 Proposed connections between the RNAi, H3K9 methylation and DNA methylation pathways in *C. deneoformans*
Schematic showing the proposed connections between the three silencing mechanisms occurring at the centromeres to silence transposable elements. Lines connected in navy are representative of published interactions in C. *deneoformans,* with arrowheads showing direction of recruitment. Lines connected in grey are representative of proposed novel interactions and mechanisms in C. *deneoformans,* with arrowheads showing direction of proposed recruitment. Red dashed line represents H3K9 methylation. Yellow dashed line represents DNA methylation.

It is also unknown how the siRNAs are produced against the centromeric regions. In plants, the RdDM relies on siRNA production through the transcription of heterochromatin regions (Cuerda-Gil & Slotkin, 2016). If siRNA production in C. *deneoformans* relies on the presence of heterochromatin, this could establish a link between the two pathways, without methylation being a mechanism of silencing. This would fit with the findings that some sRNA reads identified are Clr4 and Dnmt5 dependent, even if this is only one of several methods of producing sRNAs. It is then possible that a feedback loop is created, with components of the RNAi pathway in turn ensuring that both DNA Dnmt5 and Clr4 are recruited, with RNAi potentially acting as the missing recruitment mechanism mentioned earlier. To determine if this is occurring, a number of loci where the sRNAs are dependent on Clr4 and Dnmt5 will need to be identified. Comparison of both the H3K9 methylation and DNA methylation levels of these loci against loci where sRNAs are independent of Clr4 and Dnmt5 could determine if differences are seen in heterochromatin of target sites. Also, immunoprecipitation of either Clr4 or Dnmt5 followed by mass spectrometry would identify interacting partners and could help determine recruitment mechanisms.

Exploring the link between RNAi and telomeric silencing will help determine if the telomeres are regulated by a transcriptional gene silencing mechanism. Repeating the transposition assay using strains with components of the Nu4B complex such as Hif1 deleted could confirm if the complex has a role in silencing the telomeric regions, as in these cases Cnl1 should be able to mobilise. The link between silencing telomeric Cnl1 and an Ago2-specific RNAi pathway could also be confirmed by carrying the transposition assays out in *ago2*Δ and *ago1*Δ strains. Also, investigating the presence of subtelomeric H3K9 methylation in the presence and absence of Hif1 could identify if this chromatin remodelling contributes to H3K9 methylation.

While I haven't been able to confirm a link between these three silencing mechanisms, the crossovers between the pathways shown in the data and mentioned above suggests that they all work together at similar regions and with similar targets. This raises the question as to why so many methods are maintained? The reason for biological redundancy has also been questioned in relation to the number of methods that plants have for producing sRNAs for RdDM (Erdmann & Picard, 2020). Is the risk of transposon mobilisation so terrible that the species has taken every possible action against it to ensure genome protection? If all three silencing mechanisms have multiple mechanisms of recruitment to target loci, including relying on the other pathways, then this may act as a failsafe to ensure that if one component is lost then the others still remain, even if this may reduce the efficiency somewhat. Between these three mechanisms, it is likely that H3K9 methylation and DNA methylation are more closely linked than either of them are with the RNAi pathway. However, future work should focus on generating a strain lacking all three major pathways to determine if there is further combinatory effect, suggesting parallel functions, and to see how they interact, and how they are established if they do rely on each other. Further exploration into epimutations, along with potential NCRIP pathways would help understand the full picture of epigenetic silencing in C. *deneoformans*.

References

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K., & Mukherjee, S. K. (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*, *67*(4), 657–685. https://doi.org/10.1128/mmbr.67.4.657-685.2003

Altschup, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aramayo, R., & Metzenberg, R. L. (1996). Meiotic Transvection in Fungi. *Cell*, *86*, 103–113. https://doi.org/10.1016/s0092-8674(00)80081-1

Aravind, L., Watanabe, H., Lipman, D. J., & Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(21), 11319–11324. https://doi.org/10.1073/pnas.200346997

Barnett, J. A. (2010). A history of research on yeasts 14: medical yeasts part 2, Cryptococcus neoformans. *Yeast*, *27*, 875–904. https://doi.org/10.1002/yea

Bennett, J. E., Kwon-Chung, K. J., & Howard, D. H. (1977). Epidemiologic differences among serotypes of Cryptococcus neoformans. *American Journal of Epidemiology*, *105*(6), 582–586. https://doi.org/10.1093/oxfordjournals.aje.a112423

Bernstein, D. A., Vyas, V. K., Weinberg, D. E., Drinnenberg, I. A., Bartel, D. P., & Fink, G. R. (2012). Candida albicans Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(2), 523–528. https://doi.org/10.1073/pnas.1118859109

Bird, A. (2007). Perceptions of epigenetics. *Nature*, *447*(7143), 396–398. https://doi.org/10.1038/nature05913

Blake Billmyre, R., Croll, D., Li, W., Mieczkowski, P., Carter, D. A., Cuomo, C. A., Kronstad, J. W., & Heitman, J. (2014). Highly recombinant VGII Cryptococcus gattii

population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. *MBio*, *5*(4), 1–16. https://doi.org/10.1128/mBio.01494-14

Boeke, J. D., Garfinkel, D. J., Styles, C. A., & Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. *Cell*, *40*(3), 491–500. https://doi.org/10.1016/0092-8674(85)90197-7

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., & Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, *351*, 329–331. https://doi.org/10.1038/351329a0

Brown, C. A., Murray, A. W., & Verstrepen, K. J. (2010). Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts. *Current Biology*, *20*(10), 895–903. https://doi.org/10.1016/j.cub.2010.04.027.Rapid

Burke, J. E., Longhurst, A. D., Natarajan, P., Rao, B., Liu, J., Sales-Lee, J., Mortensen, Y., Moresco, J. J., Diedrich, J. K., Yates, J. R., & Madhani, H. D. (2019). A non-dicer RNase III and four other novel factors required for RNAi-Mediated transposon suppression in the human pathogenic yeast cryptococcus neoformans. *G3: Genes, Genomes, Genetics*, *9*(7), 2235–2244. https://doi.org/10.1534/g3.119.400330

Byrnes, E. J., Bildfell, R. J., Frank, S. A., Mitchell, T. G., Marr, K. A., & Heitman, J. (2009). Molecular evidence that the range of the vancouver island outbreak of cryptococcus gattii infection has expanded into the pacific northwest in the united states. *Journal of Infectious Diseases*, *199*(7), 1081–1086. https://doi.org/10.1086/597306

Calo, S., Nicolás, F. E., Lee, S. C., Vila, A., Cervantes, M., Torres-Martinez, S., Ruiz-Vazquez, R. M., Cardenas, M. E., & Heitman, J. (2017). A non-canonical RNA degradation pathway suppresses RNAi-dependent epimutations in the human fungal pathogen Mucor circinelloides. *PLOS Genetics*, *13*(3). https://doi.org/10.1371/journal.pgen.1006686

Calo, S., Shertz-Wall, C., Lee, S. C., Bastidas, R. J., Nicolás, F. E., Granek, J. A., Mieczkowski, P., Torres-Martínez, S., Ruiz-Vázquez, R. M., Cardenas, M. E., & Heitman, J. (2014). Antifungal drug resistance evoked via RNAi-dependent epimutations. *Nature*, *513*(7519), 555–558. https://doi.org/10.1038/nature13575

Casadevall, A., Coelho, C., Cordero, R. J. B., Dragotakes, Q., Jung, E., Vij, R., & Wear, M. P. (2019). The capsule of Cryptococcus neoformans. *Virulence*, *10*(1), 822–831. https://doi.org/10.1080/21505594.2018.1431087

Casadevall, A., & Pirofski, L. A. (2007). Accidental virulence, cryptic pathogenesis, martians, lost hosts, and the pathogenicity of environmental microbes. *Eukaryotic Cell*, *6*(12), 2169–2174. https://doi.org/10.1128/EC.00308-07

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., Stajich, J. E., & Ramírez, L. (2016). Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLOS Genetics*, *12*(6), e1006108. https://doi.org/10.1371/journal.pgen.1006108

Catalanotto, C., Azzalin, G., Macino, G., & Cogoni, C. (2000). Gene silencing in fungi and worms. *Nature*, *404*, 245. https://doi.org/10.1038/35005169

Catalanotto, C., Azzalin, G., Macino, G., Cogoni, C., Cellulari, B., Molecolare, G., & La, R. (2002). Involvement of small RNAs and role of the qde genes in the gene silencing pathway in Neurospora. *Genes & Development*, *16*, 790–795. https://doi.org/10.1101/gad.222402.and

Catania, S., Dumesic, P. A., Pimentel, H., Nasif, A., Stoddard, C. I., Burke, J. E., Diedrich, J. K., Cook, S., Shea, T., Geinger, E., Lintner, R., Yates, J. R., Hajkova, P., Narlikar, G. J., Cuomo, C. A., Pritchard, J. K., & Madhani, H. D. (2020). Evolutionary persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase. *Cell*, *180*(2), 263–277. https://doi.org/10.1101/149385

Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: Patterns and paradigms. *Nature Reviews Genetics*, *10*(5), 295–304. https://doi.org/10.1038/nrg2540

Chow, E. W. L., Morrow, C. A., Djordjevic, J. T., Wood, I. A., & Fraser, J. A. (2012). Microevolution of cryptococcus neoformans driven by massive tandem gene amplification. *Molecular Biology and Evolution*, *29*(8), 1987–2000. https://doi.org/10.1093/molbev/mss066

Cogliati, M., D'Amicis, R., Zani, A., Montagna, M. T., Caggiano, G., de Giglio, O., Balbino, S., de Donno, A., Serio, F., Susever, S., Ergin, C., Velegraki, A., Ellabib, M. S., Nardoni, S., Macci, C., Oliveri, S., Trovato, L., Dipineto, L., Rickerts, V., … Colom, M. F. (2016). Environmental distribution of Cryptococcus neoformans and C. Gattii around the Mediterranean basin. *FEMS Yeast Research*, *16*(4), 1–12. https://doi.org/10.1093/femsyr/fow045

Cogoni, C., Irelan, J. T., Schumacher, M., Schmidhauser, T. J., Selkerl, E. U., & Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *The EMBO Journal*, *1*(12), 3153–3163. https://doi.org/10.1002/j.1460-2075.1996.tb00678.x

Cogoni, C., & Macino, G. (1999a). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature*, *399*, 166–169. https://doi.org/10.1038/20215

Cogoni, C., & Macino, G. (1999b). Posttranscriptional gene silencing in Neurospora by a RecQ DNA helicase. *Science*, *286*, 2342–2344. https://doi.org/10.1126/science.286.5448.2342

Cragle, C. E., MacNicol, M. C., Byrum, S. D., Hardy, L. L., Mackintosh, S. G., Richardson, W. A., Gray, N. K., Childs, G. v., Tackett, A. J., & MacNicol, A. M. (2019). Musashi interaction with poly(A)-binding protein is required for activation of target mRNA translation. *Journal of Biological Chemistry*, *294*(28), 10969–10986. https://doi.org/10.1074/jbc.RA119.007220

Cuerda-Gil, D., & Slotkin, R. K. (2016). Non-canonical RNA-directed DNA methylation. *Nature Plants*, *2*(11). https://doi.org/10.1038/nplants.2016.163

Czech, B., & Hannon, G. J. (2011). Small RNA sorting: Matchmaking for argonautes. *Nature Reviews Genetics*, *12*(1), 19–31. https://doi.org/10.1038/nrg2916

Dannah, N. S., Nabeel-Shah, S., Kurat, C. F., Sabatinos, S. A., & Fillingham, J. (2018). Functional analysis of Hif1 histone chaperone in Saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics*, *8*(6), 1993–2006. https://doi.org/10.1534/g3.118.200229

de Gontijo, F. A., Pascon, R. C., Fernandes, L., Machado, J., Alspaugh, J. A., & Vallim, M. A. (2014). The role of the de novo pyrimidine biosynthetic pathway in Cryptococcus neoformans high temperature growth and virulence. *Fungal Genetics and Biology*, *70*, 12–23. https://doi.org/10.1016/j.fgb.2014.06.003

Decote-Ricardo, D., LaRocque-de-Freitas, I. F., Rocha, J. D. B., Nascimento, D. O., Nunes, M. P., Morrot, A., Freire-de-Lima, L., Previato, J. O., Mendonça-Previato, L., & Freire-de-Lima, C. G. (2019). Immunomodulatory Role of Capsular Polysaccharides Constituents of Cryptococcus neoformans. *Frontiers in Medicine*, *6*. https://doi.org/10.3389/fmed.2019.00129

Ding, S. W., & Voinnet, O. (2007). Antiviral Immunity Directed by Small RNAs. *Cell*, *130*(3), 413–426. https://doi.org/10.1016/j.cell.2007.07.039

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*(5952), 544–550. https://doi.org/10.1126/science.1176945

Dumesic, P. A., Homer, C. M., Moresco, J. J., Pack, L. R., Erin, K., Coyle, S. M., Strahl, B. D., Fujimori, D. G., & Iii, J. R. Y. (2015). Product binding enforces the genomic specificity of a yeast Polycomb repressive complex. *Cell*, *160*(0), 204–218. https://doi.org/10.1016/j.cell.2014.11.039.Product

Dumesic, P. A., Natarajan, P., Chen, C., Drinnenberg, I. A., Schiller, B. J., Thompson, J., Moresco, J. J., Iii, J. R. Y., David, P., & Madhani, H. D. (2013). Stalled spliceosomes

are a signal for RNAi-mediated genome defense. *Cell*, *152*(5), 957–968. https://doi.org/10.1016/j.cell.2013.01.046.Stalled

Edman, J. C. (1992). Isolation of telomerelike sequences from Cryptococcus neoformans and their use in high-efficiency transformation. *Molecular and Cellular Biology*, *12*(6), 2777–2783. https://doi.org/10.1128/mcb.12.6.2777-2783.1992

Ellis, D. H., & Pfeiffer, T. J. (1990). Natural habitat of Cryptococcus neoformans var. gattii. *Journal of Clinical Microbiology*, *28*(7), 1642–1644. https://doi.org/10.1128/jcm.28.7.1642-1644.1990

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M. A., Jacobsen, S. E., Cooke, R., & Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes and Development*, *21*(20), 2539–2544. https://doi.org/10.1101/gad.451207

Erdmann, R. M., & Picard, C. L. (2020). RNA-directed DNA Methylation. *PLOS Genetics*, *16*(10). https://doi.org/10.1371/journal.pgen.1009034

Fausto, A., Rodrigues, M. L., & Coelho, C. (2019). The still underestimated problem of fungal diseases worldwide. *Frontiers in Microbiology*, *10*(FEB). https://doi.org/10.3389/fmicb.2019.00214

Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: Emerging patterns and implications. *Nature Reviews Genetics*, *13*(2), 97–109. https://doi.org/10.1038/nrg3142

Felsenfeld, G., & Groudine, M. (2003). Controlling the double helix. *Nature*, *421*(6921), 448–453. https://doi.org/10.1038/nature01410

Feretzaki, M., Billmyre, R. B., Clancey, S. A., Wang, X., & Heitman, J. (2016). Gene Network Polymorphism Illuminates Loss and Retention of Novel RNAi Silencing Components in the Cryptococcus Pathogenic Species Complex. *PLOS Genetics*, *12*(3), 1–25. https://doi.org/10.1371/journal.pgen.1005868

Feretzaki, M., & Heitman, J. (2013). Genetic Circuits that Govern Bisexual and Unisexual Reproduction in Cryptococcus neoformans. *PLOS Genetics*, *9*(8). https://doi.org/10.1371/journal.pgen.1003688

Ferrareze, P. A. G., Streit, R. S. A., dos Santos, F. M., Schrank, A., Kmetzsch, L., Vainstein, M. H., & Staats, C. C. (2017). sRNAs as possible regulators of retrotransposon activity in Cryptococcus gattii VGII. *BMC Genomics*, *18*(1), 1–14. https://doi.org/10.1186/s12864-017-3688-4

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, *5*(4), 103–107. https://doi.org/10.1016/0168-9525(89)90039-5

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(February), 806–811. https://doi.org/10.1038/35888

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–194. https://doi.org/10.1038/nature10947

Fraser, J. A., Huang, J. C., Pukkila-Worley, R., Alspaugh, J. A., Mitchell, T. G., & Heitman, J. (2005). Chromosomal translocation and segmental duplication in Cryptococcus neoformans. *Eukaryotic Cell*, *4*(2), 401–406. https://doi.org/10.1128/EC.4.2.401-406.2005

Fu, C., Sun, S., Billmyre, R. B., Roach, K. C., & Heitman, J. (2015). Unisexual versus bisexual mating in Cryptococcus neoformans: Consequences and biological impacts. *Fungal Genetics and Biology*, *78*, 65–75. https://doi.org/10.1016/j.fgb.2014.08.008.Unisexual

Gao, X., Hou, Y., Ebina, H., Levin, H. L., & Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Research*, *18*(3), 359–369. https://doi.org/10.1101/gr.7146408

Gillette, T. G., & Hill, J. A. (2015). Readers, writers, and erasers: Chromatin as the whiteboard of heart disease. *Circulation Research*, *116*(7), 1245–1253. https://doi.org/10.1161/CIRCRESAHA.116.303630

Goodwin, T. J. D., Butler, M. I., & Poulter, R. T. M. (2003). Cryptons: A group of tyrosine-recombinase- encoding DNA transposons from pathogenic fungi. *Microbiology*, *149*(11), 3099–3109. https://doi.org/10.1099/mic.0.26529-0

Goodwin, T. J. D., & Poulter, R. T. M. (2001). The diversity of retrotransposons in the yeast Cryptococcus neoformans. *Yeast*, *18*(9), 865–880. https://doi.org/10.1002/yea.733

Goss Tusher, V., Tibshirani, R., & Chu, G. (200 C.E.). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5116–5121. https://doi.org/10.1073/pnas.091062498

Grewal, S. I. S., & Klar, A. J. S. (1996). Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell*, *86*(1), 95–101. https://doi.org/10.1016/S0092-8674(00)80080-X

Gusa, A., Williams, J. D., Cho, J. E., Averette, A. F., Sun, S., Shouse, E. M., Heitman, J., Alspaugh, J. A., & Jinks-Robertson, S. (2020). Transposon mobilization in the human fungal pathogen Cryptococcus is mutagenic during infection and promotes drug resistance in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9973–9980. https://doi.org/10.1073/pnas.2001451117

Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., Falk, R., Parnmen, S., Lumbsch, H. T., & Boekhout, T. (2015). Recognition of seven species in the Cryptococcus gattii/Cryptococcus neoformans species complex. *Fungal Genetics and Biology*, *78*, 16–48. https://doi.org/10.1016/j.fgb.2015.02.009

Hall, B. M., Ma, C. X., Liang, P., & Singh, K. K. (2009). Fluctuation anaLysis calculator: A web tool for the determination of mutation rate using Luria-Delbück fluctuation

analysis. *Bioinformatics*, *25*(12), 1564–1565. https://doi.org/10.1093/bioinformatics/btp253

Hall, I. M., Shankaranarayana, G. D., Noma, K. ichi, Ayoub, N., Cohen, A., & Grewal, S. I. S. (2002). Establishment and maintenance of a heterochromatin domain. *Science*, *297*(5590), 2232–2237. https://doi.org/10.1126/science.1076466

Halliday, C. L., Bui, T., Krockenberger, M., Malik, R., Ellis, D. H., & Carter, D. A. (1999). Presence of α and a mating types in environmental and clinical collections of Cryptococcus neoformans var. gattii strains from Australia. *Journal of Clinical Microbiology*, *37*(9), 2920–2926. https://doi.org/10.1128/jcm.37.9.2920-2926.1999

Heitman, J. (2015). Evolution of sexual reproduction: a view from the Fungal Kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews*, *29*(2–4), 108–117. https://doi.org/10.1016/j.fbr.2015.08.002.Evolution

Heitman, J., Allen, B., & Alspaugh, J. A. (1999). On the Origins of Congenic MATalpha and MATa Strains of the Pathogenic Yeast Cryptococcus neoformans. *Fungal Genetics and Biology*, *28*, 1–5. https://doi.org/10.1006/fgbi.1999.1155

Henikoff, S., Ahmad, K., & Malik, H. S. (2001). The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*, *293*(5532), 1098–1102. https://doi.org/10.1126/science.1062939

Hoang, L. M. N., Maguire, J. A., Doyle, P., Fyfe, M., & Roscoe, D. L. (2004). Cryptococcus neoformans infections at Vancouver Hospital and Health Sciences Centre (1997-2002): Epidemiology, microbiology and histopathology. *Journal of Medical Microbiology*, *53*(9), 935–940. https://doi.org/10.1099/jmm.0.05427-0

Hong, E. J. E., Villén, J., Gerace, E. L., Gygi, S. P., & Moazed, D. (2005). A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biology*, *2*(3), 106–111. https://doi.org/10.4161/rna.2.3.2131

Horn, P. J., Bastie, J. N., & Peterson, C. L. (2005). A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes and Development*, *19*(14), 1705–1714. https://doi.org/10.1101/gad.1328005

Huang, S., Yoshitake, K., & Asakawa, S. (2021). A review of discovery profiling of piwi-interacting rnas and their diverse functions in metazoans. *International Journal of Molecular Sciences*, *22*(20). https://doi.org/10.3390/ijms222011166

Huff, J. T., & Zilberman, D. (2014). Dnmt1-Independent CG Methylation Contributes to Nucleosome Positioning in Diverse Eukaryotes. *Cell*, *156*(6), 1286–1297. https://doi.org/10.1038/jid.2014.371

Hughes, S. S., Buckley, C. O., & Neafsey, D. E. (2008). Complex selection on intron size in Cryptococcus neoformans. *Molecular Biology and Evolution*, *25*(2), 247–253. https://doi.org/10.1093/molbev/msm220

Hull, C. M., & Heitman, J. (2002). Genetics of Cryptococcus neoformans. *Annual Review of Genetics*, *36*(1), 557–615. https://doi.org/10.1146/annurev.genet.36.052402.152652

Ivanova, A. v., Bonaduce, M. J., Ivanov, S. v., & Klar, A. J. S. (1998). The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nature Genetics*, *19*(2), 192–195. https://doi.org/10.1038/566

James, T. Y., Kauff, F., Schoch, C., Matheny, P. B., Hofsetter, V., McLaughlin, D. J., Spatafora, J. W., & Vilgalys, R. (2006). Reconstructing early evolution of Fungi using a six-gene phylogeny. *Nature*, *443*, 818–822. https://doi.org/10.1038/nature05110

Janbon, G., Maeng, S., Yang, D. H., Ko, Y. J., Jung, K. W., Moyrand, F., Floyd, A., Heitman, J., & Bahn, Y. S. (2010). Characterizing the role of RNA silencing components in Cryptococcus neoformans. *Fungal Genetics and Biology*, *47*(12), 1070–1080. https://doi.org/10.1016/j.fgb.2010.10.005

Janbon, G., Ormerod, K. L., Paulet, D., Byrnes, E. J., Yadav, V., Chatterjee, G., Mullapudi, N., Hon, C. C., Billmyre, R. B., Brunel, F., Bahn, Y. S., Chen, W., Chen, Y.,

Chow, E. W. L., Coppée, J. Y., Floyd-Averette, A., Gaillardin, C., Gerik, K. J., Goldberg, J., … Dietrich, F. S. (2014). Analysis of the Genome and Transcriptome of Cryptococcus neoformans var. grubii Reveals Complex RNA Expression and Microevolution Leading to Virulence Attenuation. *PLOS Genetics*, *10*(4). https://doi.org/10.1371/journal.pgen.1004261

Jia, S., Kobayashi, R., & Grewal, S. I. S. (2005). Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. *Nature Cell Biology*, *7*(10), 1007–1013. https://doi.org/10.1038/ncb1300

Jiang, N., Yang, Y., Janbon, G., Pan, J., & Zhu, X. (2012). Identification and Functional Demonstration of miRNAs in the Fungus Cryptococcus neoformans. *PLOS ONE*, *7*(12), 20–25. https://doi.org/10.1371/journal.pone.0052734

Kainz, K., Bauer, M. A., Madeo, F., & Carmona-Gutierrez, D. (2020). Fungal infections in humans: The silent crisis. *Microbial Cell*, *7*(6), 143–145. https://doi.org/10.15698/mic2020.06.718

Kämper, J., Kahmann, R., Bölker, M., Ma, L. J., Brefort, T., Saville, B. J., Banuett, F., Kronstad, J. W., Gold, S. E., Müller, O., Perlin, M. H., Wösten, H. A. B., de Vries, R., Ruiz-Herrera, J., Reynaga-Peña, C. G., Snetselaar, K., McCann, M., Pérez-Martín, J., Feldbrügge, M., … Birren, B. W. (2006). Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. *Nature*, *444*(7115), 97–101. https://doi.org/10.1038/nature05248

Kavanaugh, L. A., Fraser, J. A., & Dietrich, F. S. (2006). Recent evolution of the human pathogen Cryptococcus neoformans by intervarietal transfer of a 14-gene fragment. *Molecular Biology and Evolution*, *23*(10), 1879–1890. https://doi.org/10.1093/molbev/msl070

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*, 617–624. https://doi.org/10.1038/nature02424

Kelly, T. J., Qin, S., Gottschling, D. E., & Parthun, M. R. (2000). Type B Histone Acetyltransferase Hat1p Participates in Telomeric Silencing. *Molecular and Cellular*

Biology, *20*(19), 7051–7058. https://doi.org/10.1128/MCB.20.19.7051-7058.2000

Kim, M. S., Kim, S. Y., Yoon, J. K., Lee, Y. W., & Bahn, Y. S. (2009). An efficient gene-disruption method in Cryptococcus neoformans by double-joint PCR with NAT-split markers. *Biochemical and Biophysical Research Communications*, *390*(3), 983–988. https://doi.org/10.1016/j.bbrc.2009.10.089

Kwon Chung, K. J., Bennett, J. E., & Theodore, T. S. (1978). Cryptococcus bacillisporus sp. nov.: serotype B C of Cryptococcus neoformans. *International Journal of Systematic Bacteriology*, *28*(4), 616–620. https://doi.org/10.1099/00207713-28-4-616

Kwon-Chung, K. J. (1975). A new genus, filobasidiella, the perfect state of Cryptococcus neoformans. *Mycologia*, *67*(6), 1197–1200. https://doi.org/10.2307/3758842

Kwon-Chung, K. J. (1976). A New Species of Filobasidiella, the Sexual State of Cryptococcus neoformans B and C Serotypes. *Mycologia*, *68*(4), 942. https://doi.org/10.2307/3758813

Kwon-Chung, K. J. (1980). Nuclear genotypes of spore chains in Filobasidiella neoformans (Cryptococcus neoformans). *Mycologia*, *72*(2), 418–422. https://doi.org/10.2307/3759266

Kwon-Chung, K. J., & Bennett, J. E. (1978). Distribution of α and α mating types of Cryptococcus neoformans among natural and clinical isolates. *American Journal of Epidemiology*, *108*(4), 337–340. https://doi.org/10.1093/oxfordjournals.aje.a112628

Kwon-Chung, K. J., Boekhout, T., Fell, J. W., & Diaz, M. (2002). (1557) Proposal to conserve the name Cryptococcus gattii against C. hondurianus and C. bacillisporus (Basidiomycota, Hymenomycetes, Tremellomycetidae). *Taxon*, *51*(4), 804–806. https://doi.org/10.2307/1555045

Kwon-Chung, K. J., Edman, J. C., & Wickes, B. L. (1992). Genetic association of mating types and virulence in Cryptococcus neoformans. *Infection and Immunity*, *60*(2), 602–605. https://doi.org/10.1128/iai.60.2.602-605.1992

Kwon-Chung, K. J., & Varma, A. (2006). Do major species concepts support one, two or more species within Cryptococcus neoformans? *FEMS Yeast Research*, *6*(4), 574–587. https://doi.org/10.1111/j.1567-1364.2006.00088.x

Laurie, J. D., Linning, R., & Bakkeren, G. (2008). Hallmarks of RNA silencing are found in the smut fungus Ustilago hordei but not in its close relative Ustilago maydis. *Current Genetics*, *53*(1), 49–58. https://doi.org/10.1007/s00294-007-0165-7

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204–220. https://doi.org/10.1038/nrg2719.Establishing

Lea, D. E., & Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics*, *49*, 264–286. https://doi.org/10.1007/BF02986080

Li, L., Zaragoza, O., Casadevall, A., & Fries, B. C. (2006). Characterization of a flocculation-like phenotype in Cryptococcus neoformans and its effects on pathogenesis. *Cell Microbiology*, *8*(11), 1730–1739. https://doi.org/10.1111/j.1462-5822.2006.00742.x.Characterization

Liebman, S. W., & Chernoff, Y. O. (2012). Prions in yeast. *Genetics*, *191*(4), 1041–1072. https://doi.org/10.1534/genetics.111.137760

Lin, X., & Heitman, J. (2006). The biology of the Cryptococcus neoformans species complex. In *Annual Review of Microbiology* (Vol. 60, pp. 69–105). https://doi.org/10.1146/annurev.micro.60.080805.142102

Lin, X., Hull, C. M., & Heitman, J. (2005). Sexual reproduction between partners of the same mating type in Cryptococcus neoformans. *Nature*, *434*(7036), 1017–1021. https://doi.org/10.1038/nature03448

Liu, H., Cottrell, T. R., Pierini, L. M., Goldman, W. E., & Doering, T. L. (2002). RNA interference in the pathogenic fungus Cryptococcus neoformans. *Genetics*, *160*(2), 463–470. https://doi.org/10.1093/genetics/160.2.463

Loftus, B. J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I. J., Fraser, J. A., Allen, J. E., Bosdet, I. E., Brent, M. R., Chiu, R., Doering, T. L., Donlin, M. J., D'Souza, C. A., Fox, D. S., Grinberg, V., … Hyman, R. W. (2005). The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. *Science*, *307*(5713), 1321–1324. https://doi.org/10.1126/science.1103773

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Luria, S. E., & Delbruck, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, *28*, 491–511. https://doi.org/10.1002/14651858.CD011168

Lyon, M. F. (1961). Gene Action in the X-chromosome of the Mouse (Mus muscufus L.). *Nature*, *190*, 372–373. https://doi.org/10.1038/190372a0

Maiti, M., Lee, H., & Liu, Y. (2007). QIP, a putative exonuclease, interacts with the. *Genes and Development*, *21*, 590–600. https://doi.org/10.1101/gad.1497607.catalytic

Marín, I., & Lloréns, C. (2000). Ty3/Gypsy retrotransposons: Description of new Arabidopsis thaliana elements and evolutionary perspectives derived from comparative genomic data. *Molecular Biology and Evolution*, *17*(7), 1040–1049. https://doi.org/10.1093/oxfordjournals.molbev.a026385

Martens, H., Jindrich, N., Oberstrass, J., Steck, T. L., Postlethwait, P., & Nellen, W. (2002). RNAi in Dictyostelium: The Role of RNA-directed RNA Polymerases and

Double-stranded RNase. *Molecular Biology of the Cell*, *13*(April), 445–453. https://doi.org/10.1091/mbc.01

Martienssen, R., & Moazed, D. (2015). RNAi and heterochromatin assembly. *Cold Spring Harbor Perspectives in Biology*, *7*(8). https://doi.org/10.1101/cshperspect.a019323

Matzke, M. A., Aufsatz, W., Kanno, T., Mette, M. F., & Matzke, A. J. M. (2002). Homology-Dependent Gene Silencing and Host Defense in Plants. *Advances in Genetics*, *46*, 235–275. https://doi.org/10.1016/s0065-2660(02)46009-9

Matzke, M. A., Primig, M., Trnovsky, J., & Matzke, A. J. M. (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *The EMBO Journal*, *8*(3), 643–649. https://doi.org/10.1002/j.1460-2075.1989.tb03421.x

Maziarz, E. K., & Perfect, J. R. (2016). Cryptococcosis. *Infectious Disease Clinics of North America*, *30*(1), 179–206. https://doi.org/10.1016/j.idc.2015.10.006

McClintock, B. (1950). The origin and behaviour of mutable loci in Maize. *Genetics*, *36*, 344–355. https://doi.org/10.1073/pnas.36.6.344

McClintock, B. (1953). Induction of Instability At Selected Loci in Maize. *Genetics*, *38*(6), 579–599. https://doi.org/10.1093/genetics/38.6.579

Mochizuki, K., & Gorovsky, M. A. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation and meiotic prophase. *Genes & Development*, *19*, 77–89. https://doi.org/10.1101/gad.1265105.types

Moore, A., Dominska, M., Greenwell, P., Aksenova, A. Y., Mirkin, S., & Petes, T. (2018). Genetic control of genomic alterations induced in yeast by interstitial telomeric sequences. *Genetics*, *209*(2), 425–438. https://doi.org/10.1534/genetics.118.300950

Motamedi, M. R., Erica Hong, E.-J., Li, X., Gerber, S., Denison, C., Gygi, S., & Moazed, D. (2008). HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic

Gene Silencing by Non-Overlapping Mechanisms. *Molecular Cell*, *32*, 778–790. https://doi.org/10.1016/j.molcel.2008.10.026

Nakagawachi, T., Soejima, H., Urano, T., Zhao, W., Higashimoto, K., Satoh, Y., Matsukura, S., Kudo, S., Kitajima, Y., Harada, H., Furukawa, K., Matsuzaki, H., Emi, M., Nakabeppu, Y., Miyazaki, K., Sekiguchi, M., & Mukai, T. (2003). Silencing effect of CpG island hypermethylation and histone modifications on O6-methylguanine-DNA methyltransferase (MGMT) gene expression in human cancer. *Oncogene*, *22*(*55*), 8835–8844. https://doi.org/10.1038/sj.onc.1207183

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., & Grewal, S. I. S. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, *292*(*5514*), 110–113. https://doi.org/10.1126/science.1060118

Nakayashiki, H., Kadotani, N., & Mayama, S. (2006). Evolution and diversification of RNA silencing proteins in fungi. *Journal of Molecular Evolution*, *63*(1), 127–135. https://doi.org/10.1007/s00239-005-0257-2

Nakayashiki, H., & Nguyen, Q. B. (2008). RNA interference: roles in fungal biology. *Current Opinion in Microbiology*, *11*(6), 494–502. https://doi.org/10.1016/j.mib.2008.10.001

Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant Cell*, *2*, 279–289. https://doi.org/10.1105/tpc.2.4.279

Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A., & Akbarian, S. (2016). Epigenetic Basis of Mental Illness. *Neuroscientist*, *22*(*5*), 447–463. https://doi.org/10.1177/1073858415608147

Nicolás, F. E., Torres-Martínez, S., & Ruiz-Vázquez, R. M. (2013). Loss and Retention of RNA Interference in Fungi and Parasites. *PLOS Pathogens*, *9*(1). https://doi.org/10.1371/journal.ppat.1003089

Nnadi, N. E., & Carter, D. A. (2021). Climate change and the emergence of fungal pathogens. *PLOS Pathogens*, *17*(4). https://doi.org/10.1371/journal.ppat.1009503

Nolan, T., Braccini, L., Azzalin, G., de Toni, A., Macino, G., & Cogoni, C. (2005). The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in Neurospora crassa. *Nucleic Acids Research*, *33*(5), 1564–1573. https://doi.org/10.1093/nar/gki300

Okagaki, L. H., Strain, A. K., Nielsen, J. N., Charlier, C., Baltes, N. J., Chrétien, F., Heitman, J. H., Dromer, F., & Nielsen, K. N. (2010). Cryptococcal cell morphology affects host cell interactions and pathogenicity. *PLOS Pathogens*, *6*(6). https://doi.org/10.1371/journal.ppat.1000953

Pathak, E., Meelam, A., & Mishra, R. (2014). Analysis of P-Loop and its Flanking Region Subsequence of Diverse NTPases Reveals Evolutionary Selected Residues. *Bioinformation*, *10*(4), 216–220. https://doi.org/10.6026/97320630010216

Pérez-Arques, C., Navarro-Mendoza, M. I., Murcia, L., Navarro, E., Garre, V., & Nicolás, F. E. (2020). A non-canonical RNAi pathway controls virulence and genome stability in Mucorales. *PLOS Genetics*, *16*(7). https://doi.org/10.1371/journal.pgen.1008611

Pidoux, A. L., & Allshire, R. C. (2005). The role of heterochromatin in centromere function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1455), 569–579. https://doi.org/10.1098/rstb.2004.1611

Pintacuda, G., Lassen, F. H., Hsu, Y. H. H., Kim, A., Martín, J. M., Malolepsza, E., Lim, J. K., Fornelos, N., Eggan, K. C., & Lage, K. (2021). Genoppi is an open-source software for robust and standardized integration of proteomic and genetic data. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22648-5

Ponger, L., & Li, W. H. (2005). Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*, *22*(4), 1119–1128. https://doi.org/10.1093/molbev/msi098

Poveda, A., Pamblanco, M., Tafrov, S., Tordera, V., Sternglanz, R., & Sendra, R. (2004). Hif1 Is a Component of Yeast Histone Acetyltransferase B, a Complex Mainly Localized in the Nucleus. *Journal of Biological Chemistry*, *279*(16), 16033–16043. https://doi.org/10.1074/jbc.M314228200

Priest, S. J., Coelho, M. A., Mixão, V., Clancey, S. A., Xu, Y., Sun, S., Gabaldón, T., & Heitman, J. (2021). Factors enforcing the species boundary between the human pathogens Cryptococcus neoformans and Cryptococcus deneoformans. *PLOS Genetics*, *17*(1), 1–33. https://doi.org/10.1371/journal.pgen.1008871

Priest, S. J., Yadav, V., Roth, C., Dahlmann, T. A., Kück, U., Magwene, P. M., & Heitman, J. (2021). Rampant transposition following RNAi loss causes hypermutation and antifungal drug resistance in clinical isolates of a human fungal pathogen. *BioRxiv*. https://doi.org/10.1101/2021.08.11.455996

Radchenko, E. A., McGinty, R. J., Aksenova, A. Y., Neil, A. J., & Mirkin, S. M. (2018). Quantitative analysis of the rates for repeat-mediated genome instability in a yeast experimental system. *Methods in Molecular Biology*, *1672*, 421–438. https://doi.org/10.1007/978-1-4939-7306-4_29

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*(8), 1896–1906. https://doi.org/10.1038/nprot.2007.261

Reedy, J. L., Bastidas, R. J., & Heitman, J. (2007). The Virulence of Human Pathogenic Fungi: Notes from the South of France. *Cell Host and Microbe*, *2*(2), *77*–83. https://doi.org/10.1016/j.chom.2007.07.004

Rgen Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*, *13*, 2513–2526. https://doi.org/10.1074/mcp

Romano, N., & Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular*

Microbiology, *6*(22), 3343–3353. https://doi.org/10.1111/j.1365-2958.1992.tb02202.x

Roy, B., & Sanyal, K. (2011). Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. *Eukaryotic Cell*. https://doi.org/10.1128/EC.05165-11

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schäfer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., & Reuter, G. (2007). Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3. *Molecular Cell*, *26*(1), 103–115. https://doi.org/10.1016/j.molcel.2007.02.025

Sarkar, S., Ma, W. T., & Sandri, G. v. H. (1992). On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, *85*(2), 173–179. https://doi.org/10.1007/BF00120324

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., & Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO Journal*, *21*(5), 1121–1131. https://doi.org/10.1093/emboj/21.5.1121

Sharma, S., Kelly, T. K., & Jones, P. A. (2009). Epigenetics in cancer. *Carcinogenesis*, *31*(1), 27–36. https://doi.org/10.1093/carcin/bgp220

Shen, L., & Pelletier, J. (2020). General and target-specific DEXD/H RNA helicases in eukaryotic translation initiation. *International Journal of Molecular Sciences*, *21*(12), 1–23. https://doi.org/10.3390/ijms21124402

Shiu, P. K. T., & Raju, N. B. (2001). Meiotic Silencing by Unpaired DNA at least one well-studied case, a double-stranded RNA. *Cell*, *107*, 905–916. https://doi.org/10.1016/S0092-8674(01)00609-2

Singh, N., Dromer, F., Perfect, J. R., & Lortholary, O. (2008). Cryptococcosis in solid organ transplant recipients: Current state of the science. *Clinical Infectious Diseases*, *47*(10), 1321–1327. https://doi.org/10.1086/592690

Spingola, M., Grate, L., Haussler, D., & Manuel, A. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA*, *5*(2), 221–234. https://doi.org/10.1017/S1355838299981682

Staab, J. F., White, T. C., & Marr, K. A. (2011). Hairpin dsRNA does not trigger RNA interference in Candida albicans cells. *Yeast*, *28*(1), 1–8. https://doi.org/10.1002/yea.1814

Sun, S., & Xu, J. (2009). Chromosomal rearrangements between serotype A and D strains in Cryptococcus neoformans. *PLOS ONE*, *4*(5). https://doi.org/10.1371/journal.pone.0005524

Sutherland, J. M., Sobinoff, A. P., Fraser, B. A., Redgrove, K. A., Siddall, N. A., Koopman, P., Hime, G. R., & McLaughlin, E. A. (2018). RNA binding protein Musashi-2 regulates PIWIL1 and TBX1 in mouse spermatogenesis. *Journal of Cellular Physiology*, *233*(4), 3262–3273. https://doi.org/10.1002/jcp.26168

Taylor, J. W. (2015). Evolutionary perspectives on human fungal pathogens. *Cold Spring Harbor Perspectives in Medicine*, *5*(9). https://doi.org/10.1101/cshperspect.a019588

Thon, G., Hansen, K. R., Altes, S. P., Sidhu, D., Singh, G., Verhein-Hansen, J., Bonaduce, M. J., & Klar, A. J. S. (2005). The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast Schizosaccharomyces pombe. *Genetics*, *171*(4), 1583–1595. https://doi.org/10.1534/genetics.105.048298

Tollefsbol, T. O. (2017). An overview of epigenetics. In *Handbook of Epigenetics: The New Molecular and Medical Genetics* (pp. 1–6). Elsevier. https://doi.org/10.1016/B978-0-12-805388-1.00001-8

Trieu, T. A., Calo, S., Nicolás, F. E., Vila, A., Moxon, S., Dalmay, T., Torres-Martínez, S., Garre, V., & Ruiz-Vázquez, R. M. (2015). A Non-canonical RNA Silencing Pathway Promotes mRNA Degradation in Basal Fungi. *PLOS Genetics*, *11*(4). https://doi.org/10.1371/journal.pgen.1005168

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO Journal*, *13*(16), 3822–3831. https://doi.org/10.1002/j.1460-2075.1994.tb06693.x

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. S., & Moazed, D. (2004). RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. *Science*, *303*(5658), 672–676. https://doi.org/10.1126/science.1093686

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., & Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, *297*(5588), 1833–1837. https://doi.org/10.1126/science.1074973

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Wang, X., Darwiche, S., & Heitman, J. (2013). Sex-induced silencing operates during opposite-sex and unisexual reproduction in Cryptococcus neoformans. *Genetics*, *193*(4), 1163–1174. https://doi.org/10.1534/genetics.113.149443

Wang, X., Hsueh, Y. P., Li, W., Floyd, A., Skalsky, R., & Heitman, J. (2010). Sex-induced silencing defends the genome of Cryptococcus neoformans via RNAi. *Genes and Development*, *24*(22), 2566–2582. https://doi.org/10.1101/gad.1970910

Wang, X., Wang, P., Sun, S., Darwiche, S., Idnurm, A., & Heitman, J. (2012). Transgene Induced Co-Suppression during Vegetative Growth in Cryptococcus neoformans. *PLOS Genetics*, *8*(8), 1–13. https://doi.org/10.1371/journal.pgen.1002885

Wang, Y., Aisen, P., & Casadevall, A. (1995). Cryptococcus neoformans Melanin and Virulence: Mechanism of Action. *Infection and Immunity*, *63*(8), 3131–3136. https://doi.org/10.1128/iai.63.8.3131-3136.1995

Wang, Y., Wei, D., Zhu, X., Pan, J., Zhang, P., Huo, L., & Zhu, X. (2016). A "suicide" CRISPR-Cas9 system to promote gene deletion and restoration by electroporation

in Cryptococcus neoformans. *Scientific Reports*, *6*(July), 1–13. https://doi.org/10.1038/srep31145

Warnecke, T., Parmley, J. L., & Hurst, L. D. (2008). Finding exonic islands in a sea of non-coding sequence: Splicing related constraints on protein composition and evolution are common in intron-rich genomes. *Genome Biology*, *9*(2). https://doi.org/10.1186/gb-2008-9-2-r29

Watkins, R. A., King, J. S., & Johnston, S. A. (2017). Nutritional requirements and their importance for virulence of pathogenic cryptococcus species. *Microorganisms*, *5*(4). https://doi.org/10.3390/microorganisms5040065

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Wickes, B. L., Mayorga, M. E., Edman, U., & Edman, J. C. (1996). Dimorphism and haploid fruiting in Cryptococcus neoformans: Association with the alpha-mating type. *Proceedings of the National Academy of Sciences*, *93*(14), 7327–7331. https://doi.org/10.1073/pnas.93.14.7327

Wickes, B. L., Moore, T. D. E., & Kwon-Chung, K. J. (1994). Comparison of the electrophoretic karyotypes and chromosomal location of ten genes in the two varieties of Cryptococcus neoformans. *Microbiology*, *140*(3), 543–550. https://doi.org/10.1099/00221287-140-3-543

Wilson, R. C., & Doudna, J. A. (2013). Molecular mechanisms of RNA interference. *Annual Review of Biophysics*, *42*(1), 217–239. https://doi.org/10.1146/annurev-biophys-083012-130404

Winzeler, E. A., Castillo-Davis, C. I., Oshiro, G., Liang, D., Richards, D. R., Zhou, Y., & Hartl, D. L. (2003). Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics*, *163*(1), 79–89. https://doi.org/10.1093/genetics/163.1.79

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., … Nurse, P. (2002). The genome sequence of Schizosaccharomyces pombe. *Nature*, *47*(9), 1215–1220. https://doi.org/10.1038/nature724

Yadav, V., Sun, S., Billmyre, R. B., Thimmappa, B. C., Shea, T., Lintner, R., Bakkeren, G., Cuomo, C. A., Heitman, J., & Sanyal, K. (2018). RNAi is a critical determinant of centromere evolution in closely related fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), 3108–3113. https://doi.org/10.1073/pnas.1713725115

Yadav, V., Sun, S., Coelho, M. A., & Heitman, J. (2020). Centromere scission drives chromosome shuffling and reproductive isolation. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(14), 7917–7928. https://doi.org/10.1073/pnas.1918659117

Yah, Z., Li, X., & Xu, J. (2002). Geographic distribution of mating type alleles of Cryptococcus neoformans in four areas of the United States. *Journal of Clinical Microbiology*, *40*(3), 965–972. https://doi.org/10.1128/JCM.40.3.965-972.2002

Zaragoza, O., Rocío, G. R., Nosanchuk, J. D., Cuenca-Estrella, M., Rodríguez-Tudela, J. L., & Casadevall, A. (2010). Fungal cell gigantism during mammalian infection. *PLOS Pathogens*, *6*(6). https://doi.org/10.1371/journal.ppat.1000945

Zhang, K., Mosch, K., Fischle, W., & Grewal, S. I. S. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nature Structural and Molecular Biology*, *15*(4), 381–388. https://doi.org/10.1038/nsmb.1406

Zhao, Y., Lin, J., Fan, Y., & Lin, X. (2019). Life Cycle of Cryptococcus neoformans . *Annual Review of Microbiology*, *73*(1), 17–42. https://doi.org/10.1146/annurev-micro-020518-120210