THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY
of EDINBURGH

Investigating the role of RNA interference in the fission yeast *Schizosaccharomyces japonicus*

Elliott Chapman

Submitted for the degree of Doctor of Philosophy

University of Edinburgh

2017

Table of Contents

Table of Figures

Table of Tables

Abbreviations

5-FOA	5-Fluoroorotic Acid
ARC	Argonaute siRNA chaperone
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CCAN	Constitutive Centromere-Associated Network
CDS	coding DNA sequence
ClrC	Cryptic loci regulator complex
CpG	CG dinucleotide
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDTA	Ethylenediaminetetraacetic acid
GFP	Green fluorescent protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMW	High molecular weight
HR	Homologous recombination
iPCR	Inverse PCR
KMT	Lysine methyltransferase

LB	Lysogeny broth
LTR	Long terminal repeat
ME	Malt extract
miRNa	microRNA
mRNA	Messenger RNA
ncRNA	Non-coding RNA
NHEJ	Non-homologous end joining
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEV	Position-effect variegation
piRNA	Piwi-interacting RNA
PMG	Pombe minimal glutamate
PNK	Polynucleotide kinase
priRNA	Primal RNA
PTGS	Post-transcriptional genes silencing
PTM	Posttranslational modification
RDRC	RNA-dependent RNA polymerase complex
RdRP	RNA-dependent RNA polymerase
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	Ribosomal RNA
SHREC	Snf2–histone deacetylase repressor complex

siRNA	short interfering RNA
SNP	Single-nucleotide polymorphism
ssRNA	Single-stranded RNA
TBE	Tris/Borate/EDTA
TBZ	Thiabendazole
TGS	Transcriptional gene silencing
TIR	Tandem inverted repeat
tRNA	Transfer RNA
TSA	Trichostatin A
TSD	Target-site duplication
UTR	Untranslated region
YE	Yeast extract

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Dr. Elizabeth Bayne, for giving me the opportunity to work on this project, and for constantly supporting my development as a scientist. I would also like to thank Prof. David Finnegan for challenging me to think critically, and for the invaluable insight into transposable elements that he provided. I am also grateful to Dr. Snezhana Oliferenko, Dr. Ying Gu and Dr. Maria Marakova at King's College London, for helping me to tame *S. japonicus*.

I extend my gratitude to the BBSRC EastBio DTP for funding this work, and to both the University of Edinburgh and the Institute of Cell Biology for giving me the opportunity to carry out my graduate studies.

I would like to thank all of my colleagues, both past and present, for their friendship and support. I am especially grateful to Dr. Francesca Taglini, Dr. Rob Van Nues, Dr. Sreerekha Pillai, Dr. Laure Verrier, Ekaterina Kapitonova, Dr. Benura Azeroglu and Dr. Sveta Makovets, for making me feel so welcome in Edinburgh, and for giving me a new-found love for basketball. Special thanks also goes to Dr. Jo Strachan; without our daily coffee break there is no way that this thesis would exist.

I would like to express my eternal gratitude to my family, whose unconditional love and support, both financial and emotional, has allowed to get to where I am today.

Finally, I would like to say thank you to Laura Tuck, for tolerating me for these four years.

Declaration

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

Signed,

Elliott Chapman

Lay Summary

Inside most cells, genetic information is contained inside the nucleus, on long strands of DNA known as chromosomes. One of the most important functions of a cell is its ability to duplicate and then equally divide these chromosomes between new daughter cells, to give genetically identical or similar offspring. To ensure that this division occurs properly, DNA must be packaged in a specialised way, wound around proteins in a complex known as chromatin. This chromatin can be further broken down depending on its function; long stretches of the chromosomes must be accessible and 'active' so that these regions can give rise to the proteins required for cellular function, this is known as 'euchromatin'. In contrast, some regions of chromatin, known as 'heterochromatin', must be kept compact and 'silent'; these regions usually contain repetitive DNA and mobile genetic elements that would negatively impact the survival of the cell if not silenced. These heterochromatic regions are also very important structurally, as in many organisms heterochromatin acts to help define the position of a discrete chromosomal region known as the centromere, which plays a key role in ensuring that chromosomes get physically separated when a cell divides. A common feature of heterochromatin is that it is formed by the action of a molecule called RNA, which is produced when the DNA sequence is 'read' to give a complementary molecule. The RNA that acts to direct heterochromatin formation does not code for protein; instead it is usually very short, and associates with specific proteins and enzymes that cause the compaction seen in heterochromatin by

modifying the chromatin itself. This process is known as 'RNA interference' (RNAi).

In biology, it is common to utilise simple, single-celled organisms for the study of biological pathways, as these organisms are often much easier to work with than complex organisms such as animals. Although these organisms, commonly known as 'model' organisms, may be simpler than their more complex multi-cellular counterparts, they still carry out many of the same biological processes using related protein factors, thus lessons learnt in simple organisms can often be applied to much more complex beings. The process of heterochromatin formation by RNAi is perhaps best understood in a single-celled species of yeast called *S. pombe*. This yeast model has centromeres that are packaged in heterochromatin, a lot like plants, flies and humans; however, unlike these higher organisms, it does not employ small RNAs to silence its mobile genetic elements called transposons. For this reason, this study has employed a new yeast model organism, called *S. japonicus*, which does use small RNA to silence transposons. The work carried out here suggests that the formation of heterochromatin by small RNAs at transposons is needed to keep *S. japonicus* alive, probably by ensuring that chromosomes get segregated properly when cells divide. I have also shown that one of the key proteins involved in RNA interference is responsible for preventing mobile genetic elements from 'jumping' throughout the genome. Taken together, this work begins to establish *S. japonicus* as a new model organism for studying RNA interference and also highlights some important differences between RNAi in *S. pombe* and *S. japonicus*, which may be important for further understanding this pathway in more complex organisms.

Abstract

RNA interference (RNAi) is a conserved pathway that plays key roles in heterochromatin formation, gene regulation and genome surveillance across a wide range of eukaryotes. One of the most utilised model organisms for studying the RNAi pathway is the fission yeast *Schizosaccharomyces pombe*. However, this species is somewhat atypical, in that it has not retained the ancestral role for RNAi in the silencing of mobile genetic elements. In contrast, the related fission yeast *S. japonicus* has a large and diverse retrotransposon complement that appears to give rise to abundant siRNAs. For this reason, we believe that *S. japonicus* may be a more suitable model for studying the role of RNAi in silencing mobile genetic elements, a function that is conserved in many higher eukaryotes.

Functional analysis of the *S. japonicus* RNAi pathway proved more challenging than expected, as it was generally not possible to recover strains bearing deletions of core RNAi components (Ago1/Clr4/Rdp1/Arb1/Arb2). This suggests that a functional RNAi pathway may be required for viability in *S. japonicus*, unlike in *S. pombe*. However, disruption mutants were isolated for the sole Dicer ribonuclease Dcr1, at very low frequency. Analysis of these mutants revealed that disruption of Dcr1 impaired the generation of retrotransposon derived siRNAs, and caused de-repression of retroelement transcript accumulation and mobilisation in an element dependent manner. Surprisingly however, Dcr1 appeared dispensable for the maintenance of H3K9me2 at transposons, suggesting that, in contrast to *S. pombe*, silencing may occur principally at the post-transcriptional level. It is also possible that

the isolated Dcr1 mutants represent rare survivors that are viable due to the presence of suppressor mutations elsewhere in the genome.

I utilised my genome wide RNA sequencing data to help improve the annotation of the *S. japonicus* genome, with a specific focus on the retrotransposon complement. From this, I identified 12 new families of LTR retrotransposon, which increased the annotated retrotransposon complement by around 40% in *S. japonicus*. Finally, I characterised the integrative preference of the *S. japonicus* retrotransposon Tj1, and found that it shares characteristics associated with the *S. cerevisiae* retrotransposons Ty1 and Ty3, mostly integrating upstream of RNA PolIII transcribed tRNA genes.

The findings of this work highlight some potentially key differences in the way the RNAi pathway functions across the fission yeast clade, both in terms of its importance for viability and its mode of action. The work undertaken here also contributes to the establishment of *S. japonicus* as a model for the study of RNA interference and genome regulation.

Chapter 1 – Introduction

1.1 – Epigenetics

The term ‘epigenetics’ was first used by Conrad Waddington in 1942 to describe “the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being” (Waddington, 1942). He followed this in 1957 with his description of the ‘Epigenetic Landscape’ which introduced the concept of cell-fate as a ball rolling down-hill, with developmental processes represented as ‘valleys’ and ‘forks’; the ultimate path of the ball was determined not only by genotype, but by cell-to-cell interactions and environmental factors (Waddington, 1957). This assertion that changes in phenotype could be achieved without changes in the underlying genotype was initially used to explain developmental processes for which mechanistic insight was lacking (Allis and Jenuwein, 2016). However, a number of ground-breaking studies, carried out both before and after Waddington coined this term, shed more light on the concept that the genetic information contained within the nucleus could be maintained in potentially plastic ‘on’ or ‘off’ states. One of the earliest studies was performed by Muller in *Drosophila melanogaster*, where X-ray irradiation of flies resulted in a variegated red-white eye colour phenotype, caused by a translocation of the white gene next to pericentric ‘silent’ heterochromatin which could encroach onto the neighbouring sequences; this phenomenon was termed position-effect variegation (PEV) (Muller, 1930). Work by Barbara McClintock in maize identified DNA elements that were able to mobilise (McClintock, 1951), which she described as being involved in the developmental regulation of gene expression (McClintock, 1965). Description of the processes of X-chromosome inactivation in mice, whereby one copy of the X-chromosome per cell is

randomly selected to be silenced in female mammals (Lyon, 1961), as well as the phenomenon of imprinting, where genes are selectively silenced depending on which parent they originated from (McGrath and Solter, 1984; Surani et al., 1984), gave clear indications for the existence of non-Mendelian inheritance. This occurs where factors other than the gene are responsible for inheritance, and is now known to be a widespread phenomenon present across many diverse species (Allis and Jenuwein, 2016).

As the field of epigenetics grew and progressed, so did the definition of the term. The prefix 'epi-' is derived from the Greek meaning 'above', thus epigenetics describes layers of regulation that exist 'above' classical genetics. However as more studies were published, the definition of this term moved away from Waddington's original descriptions of developmental processes, to encompass other fields of study in which observed phenotypes may be impacted by more than solely genotype (Deans and Maggert, 2015). Even today the definition of the term is disputed, with the main point of contention being whether epigenetic changes have to be heritable or not. One contemporary definition of the term is "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" (Berger et al., 2009). However, Adrian Bird has proposed a somewhat broader definition: "The structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). This second definition includes a number of modifications excluded by the first, such as transient modifications that are either DNA-damage or cell-cycle dependent.

Regardless of whether non-heritable changes can be considered truly 'epigenetic', it is clear that both definitions agree that epigenetics deals with alterations to chromosomes; be that alterations to the DNA itself, or to the factors associated with this DNA.

Alterations to DNA were first described around the time that epigenetics was defined (Hotchkiss, 1948), and within 32 years the role of the 5-methyl cytosine modification in gene repression had been established (Holliday and Pugh, 1975; Razin and Riggs, 1980). Work within this area found that vertebrate genomes are predominately methylated on the CpG dinucleotide, however CG hotspots (termed CpG islands) existed at a number of promoters that were devoid of methylation (Bird et al., 1985). The abnormal methylation of tumour suppressor promoters was found to be implicated carcinogenesis (Costello et al., 2000; Feinberg and Vogelstein, 1983a, 1983b; Gama-sosa et al., 1983), which along with work carried out on viroid infected plants (Wassenegger et al., 1994), firmly established DNA methylation as a key component of epigenetic regulation in diverse range of organisms (Allis and Jenuwein, 2016).

1.2 - Chromatin

Aside from direct modification of the DNA itself, the modification of those proteins associated with the DNA provides an even more complex mechanism by which epigenetic regulation can occur. The complex of DNA and its associated proteins is known as chromatin. In eukaryotes, DNA is packaged in the nucleus, wrapped around an octameric protein complex known as the nucleosome. Crystallographic studies revealed that the canonical nucleosome

contains two copies of each of the histone proteins H2A, H2B, H3 and H4, which wrap 145-147bp of DNA (Luger et al., 1997). These histone proteins consist of globular core domains, which interact with each other as well as the DNA, and long tails which project from the core nucleosome particle. Repeating nucleosome units are separated by 10-50bp of intervening linker DNA (Segal et al., 2006), thus these 'nucleosome arrays' (also known as 10nm fibres) are able to fold into higher order structures via local and long range interactions, mediated by the histone proteins themselves, as well as modified histones and accessory factors that bind chromatin. Work by Heitz (Heitz, 1928) identified two cytologically distinct regions of chromatin; he defined euchromatin as the region which was more 'open' and stained less densely, and heterochromatin as the region that showed darker staining and was more compact.

Euchromatin is generally associated with active genes, and is more accessible to the transcriptional machinery and its associated factors, whilst heterochromatin typically consists of inaccessible and highly ordered nucleosomal arrays and can be divided into two distinct types, facultative and constitutive (Grewal and Jia, 2007). Facultative heterochromatin formation is developmentally timed and occurs over genes that are not always compacted or 'silenced' (Bannister and Kouzarides, 2011); an example of this is the random inactivation of one copy of the X chromosome in mammalian cells (Lyon, 1961). On the other hand, constitutive heterochromatin is formed over regions that are permanently silenced, which are commonly highly repetitive. These loci usually constitute the centromere, which is the region responsible for accurate chromosome segregation, the telomeric regions that cap the ends

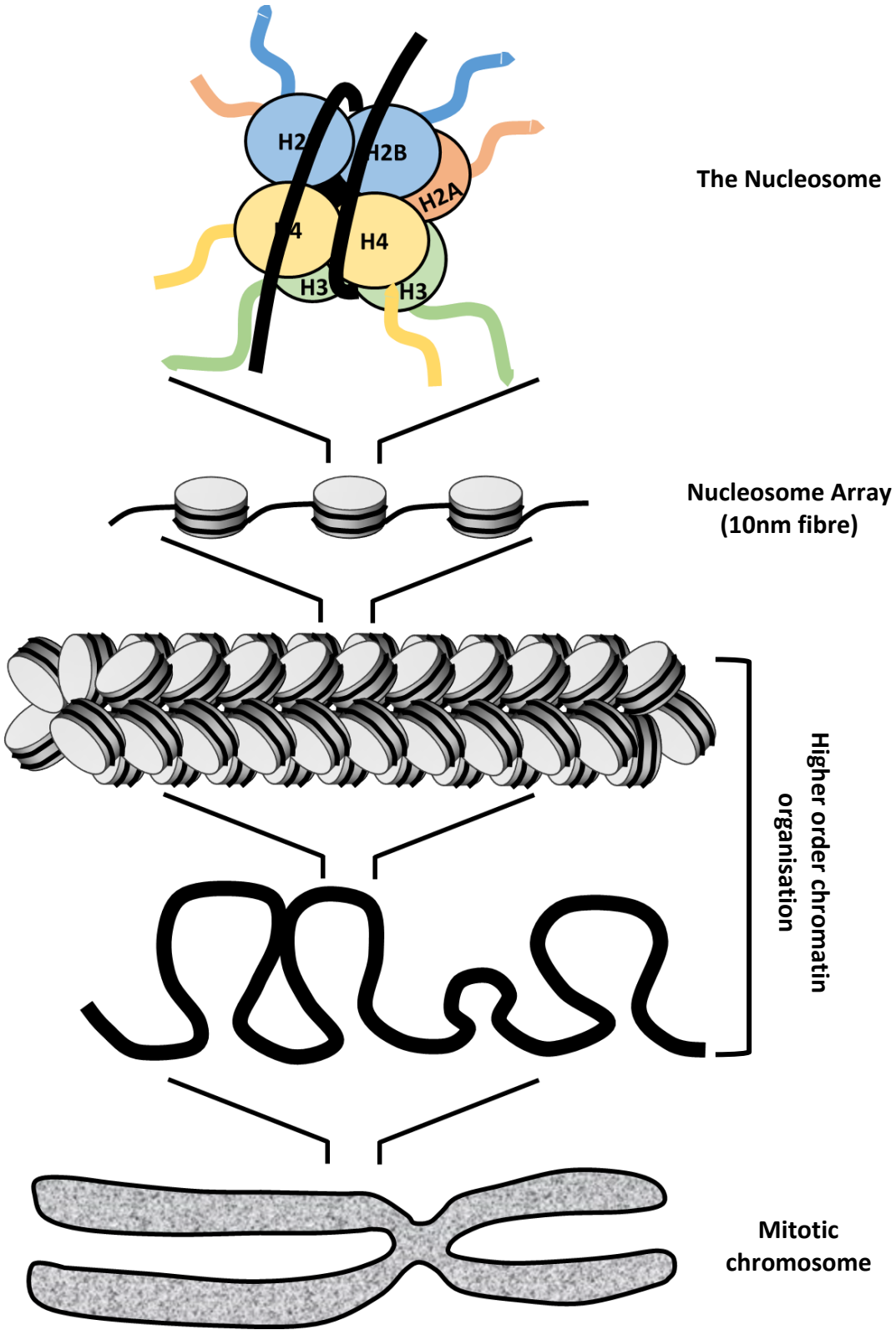


Figure 1.1 – The hierarchical organisation of chromatin

Schematic illustration of the nucleosome core particle and it's packaging to form the condensed mitotic chromosome.

of chromosomes, and mobile genetic elements such as transposons (Grewal and Jia, 2007).

As mentioned above, modification of histones is one of the key methods of mediating epigenetic changes; this may be achieved by direct modification of this histone itself, either on the tail or the globular core, or by exchange of canonical histones with specialised histone variants.

1.2.1 - Histone Post-Translational Modifications (PTMs)

Structural studies of the nucleosome revealed that highly basic histone N-terminal tails protrude from the core particle and make contact with neighbouring nucleosomes (Luger et al., 1997). Subsequent work has revealed that these tails can be modified by enzymes known as 'writers'; these modifications can be removed by enzymes termed 'erasers', and the modification state of the histone tails can be interpreted by protein factors called 'readers' (Allis and Jenuwein, 2016). These readers are responsible for modulating a number of key biological processes in response to the modification of certain nucleosomes. For a full list of histone PTMs and their reader domains, see Table 1.1.

1.2.1.1 - Acetylation

Acetylation of histones was first described in 1964 by Vincent Allfrey (Allfrey et al., 1964) however the 'writer' enzyme responsible for this modification was not identified for another 32 years. The first histone acetyltransferase (HAT) was purified from the ciliate *Tetrahymena thermophila* (Brownell et al., 1996); this was followed a month later by the discovery of the first 'eraser' histone

Modification	Modification position	Reader Domain
Methyllysine	H3Kme1	MBT
	H3Kme2	MBT
	H3K4me1	Chromo-barrel/DCD
	H3K4me2	DCD/PHD
	H3K4me3	DCD/PHD/TTD/Zf-cw
	H3K9me1	Ankyrin
	H3K9me2	Ankyrin/Chromodomain
	H3K9me3	ADD/Chromodomain
	H3K27me2	Chromodomain
	H3K27me3	Chromodomain/WD40
	H3K36me2	Chromo-barrel
	H3K36me3	Chromo-barrel/PWWP/Tudor
	H3K79me3	PWWP
	H4Kme1	MBT
	H4Kme2	MBT
	H4K20me1	PWWP
	H4K20me2	BAH/TTD
	H4K20me2	PWWP
Methylarginine	H3Rme2	Tudor
	H3R2me2	WD40
	H4Rme2	Tudor
	H4R3me2s	ADD
Acetyllysine	H2AKac	Bromodomain
	H2BKac	Bromodomain
	H3Kac	Bromodomain/DPF
	H3KacKac	DBD
	H3K56ac	Double PH
	H4Kac	Bromodomin
	H4KacKac	DBD
Phosphoserine Phosphotyrosine	H2AXS139ph (γH2AX)	Tandem BRCT
	H3T3ph	BIR
	H3S10ph	14-3-3
	H3S28ph	14-3-3

Table 1.1 – Histone posttranslational modifications and their reader domains

Table of modified histones and the protein domains that recognise them. See body text for full details.

(ADD = ATRX-Dnmt3-Dnmt3L, BAH = Bromo-Adjacent Homology, BIR = Baculovirus Inhibitor of apoptosis protein Repeat, BRCT = BRCA1 C Terminus, DBD = Double Bromodomain, DCD = Double Chromodomain, DPF = Double PHD Finger, MBT = Malignant Brain Tumor, PH = Pleckstrin Homology, TTD = Tandem Tudor Domain)

deacetylase (HDAC) (Taunton et al., 1996). The description of these opposing enzymes demonstrated that this mark was dynamic, and could be placed or removed depending on the activity of these two classes of enzyme. The HATs are split into two major families, type-A and type-B, both of which require acetyl-CoA to acetylate lysines (Bannister and Kouzarides, 2011). The type-B HATs form one group that are related to *Saccharomyces cerevisiae* HAT1; these enzymes modify free histones in the cytoplasm to ensure their proper deposition (Parthun, 2007). Type-A HATs are a more diverse class, and can be divided into 3 families: GNAT, MYST and CBP/p300 (Kimura et al., 2005; Lee and Workman, 2007). Each of these enzymes modifies multiple lysine residues on both the tail and core of histone proteins and to date modification of the following residues has been reported: H3 (K4, K9, K14, K18, K23, K27, K36 and K56), H4 (K5, K8, K12, K16, K20 and K91), H2A (K5 and K9), H2B (K5, K12, K15, K16, K20 and K120) (Musselman et al., 2012). The specificity of the enzymes may be dictated by their association with accessory proteins within large complexes, as some factors demonstrate different substrate preferences dependent on the complex with which they are associated (Bannister and Kouzarides, 2011). The opposing HDACs are organised into four classes: class I/II are related to Rpd3 and Hda1 from budding yeast, class III contains the NAD^+ dependent sirtuins, and class IV contains only one member, the human HDAC11.

The acetylation of lysines can alter chromatin in one of two ways. The addition of an acetyl moiety to a positively charged lysine acts to neutralise this charge (Turner, 1993); this in turn causes an ‘opening up’ of the chromatin as it weakens the interaction between the histone and the negatively charged DNA,

which may explain why this mark is associated with transcriptionally active chromatin (Musselman et al., 2012). This mark can also act as a binding platform for protein 'readers', and to date three motifs capable of reading this mark have been identified; the bromodomain (Dhalluin et al., 1999), the double PHD finger (Zeng et al., 2010) and the double pleckstrin homology (PH) domain (Su et al., 2012). These domains are most commonly associated with transcriptional activators, thus connecting histone acetylation to active transcription. There are also roles for these proteins in other diverse functions such as the DNA damage response (Su et al., 2012), chromatin remodelling (Garabedian et al., 2012) and DNA replication (Collins et al., 2002; Musselman et al., 2012).

1.2.1.2 - Methylation

Described in the same study as acetylation (Allfrey et al., 1964), methylation of histones is one of the most versatile and widespread post-translational modifications. This modification is more complex that acetylation, as substrates can be mono-, di- or tri- methylated in the case of lysines, or symmetrically/asymmetrically modified in the case of arginine residues. Of these two substrates, the modification of arginine is less studied, and the function is somewhat unclear (Musselman et al., 2012), thus below I will only discuss the well characterised methylation of lysine residues.

The mammalian SUV39H1 enzyme was the first lysine methyltransferase (KMT) to be discovered (Rea et al., 2000), however unlike acetylation, methylation was believed to be a permanent modification, owing to the lack of identified demethylase enzymes (Bannister and Kouzarides, 2011). This

hypothesis was disproved in 2004, with the discovery of the first lysine specific demethylase (LSD1) (Shi et al., 2004), which showed that, like acetylation, methylation was a dynamic mark. KMTs can be split into two families; those that methylate histone tails contain an enzymatic SET domain, and those that methylate the globular core of the histone lack this SET domain (Min et al., 2003a; Ng et al., 2002). Both enzymes require S-adenosyl methionine to produce methylated substrate (Bannister and Kouzarides, 2011), with modifications reported on the following residues: H3 (K4, K9, K26, K27, K36 and K79), H4K20 and H1K26 (Musselman et al., 2012). In contrast to HATs, KMTs seem to be fairly substrate-specific, and they also have the ability to specifically mono-, di- or tri- methylate their substrates (Cheng et al., 2005; Collins et al., 2005).

On the opposing side there are two main types of lysine demethylase; the initially discovered LSD1, which can only demethylate mono-/di-methylation and requires protein complexes to direct specificity, and the JmjC jumonji domain proteins (Whetstine et al., 2006) which can direct demethylation of specific targets, without the requirement for accessory factors (Mosammaparast and Shi, 2010).

As opposed to acetylation, lysine methylation does not alter the charge of this residue (Bannister and Kouzarides, 2011), thus action via altered DNA interaction is unlikely, although it does alter residue size and hydrophobicity (Musselman et al., 2012). Due to the variety of lysine methylation states available, more protein domains recognise methylation than any other modified histone substrate. To date, the domains that bind methylated lysines have been identified as the structurally related Royal superfamily consisting of

the chromo-barrel, chromodomain, double chromodomain (DCD), malignant brain tumour (MBT), Pro-Trp-Trp-Pro (PWWP), tandem Tudor domain (TTD) and Tudor domains as well as the unrelated TRX-DNMT3-DNMT3L (ADD), ankyrin, bromo-adjacent homology (BAH) , plant homeodomain (PHD), WD40 and zinc finger CW (zf-CW) domains (Musselman et al., 2012). The specificity of these domains for mono-, di- or tri-methylated lysines is dependent on size of an aromatic binding cage, while the specificity for methylated lysines in specific positions is dictated by interactions with surrounding residues present on the binding protein (Musselman et al., 2012). As the pattern of lysine methylation in cells is more complex than other histone modifications, it is not surprising that its functional significance is also more complex. Unlike acetylation, which is considered a transcriptionally activating mark, lysine methylation can be either activating or repressive; this is dependent on the residue modified, as well as the degree of methylation present. Generally H3K4 methylation is considered an activating mark, however the degree of methylation varies across active genes, with H3K4me2 enriched across gene bodies, whilst H3K4me3 occurs at transcription start sites (TSSs) (Vakoc et al., 2006). In contrast to this, H3K9 and H3K27 methylation are associated with repressive chromatin environments; these marks bind the chromodomain containing Heterochromatin Protein 1 (HP1) (Bannister et al., 2001; Lachner et al., 2001), and Polycomb (Fischle et al., 2003; Min et al., 2003b) respectively. HP1 is a key factor involved in the formation and spreading of heterochromatin (Canzio et al., 2011), whilst Polycomb is a part of the Polycomb repressive complex (PRC1) which acts to compact chromatin (Margueron and Reinberg, 2011).

1.2.1.3 - Phosphorylation and other Modifications

Histone proteins can also be targeted by phosphatases and kinases, with phosphorylation known to occur predominantly on serine, threonine and tyrosine residues (Banerjee and Chakravarti, 2011). This modification has been reported at the following sites: H3 (T3, T6, S10, T11, S28 and T45), H4S1, H2A (S1 and T120), H2BS14, and H2AXS139 (Musselman et al., 2012), and is responsible for adding a significant negative charge to the side chain (Bannister and Kouzarides, 2011). This modification is essential in the DNA damage response, where the histone variant H2AX is modified on S139 (to give γH2AX) in response to double strand breaks (Rogakou et al., 1998), which promotes binding of the checkpoint protein MDC1 via a tandem BRCT domain (Stucki et al., 2005). Phosphorylation of H3S10 has also been shown to be key in mediating faithful chromosome segregation by ejecting the H3K9me-bound HP1 protein during mitosis (Fischle et al., 2005; Mellone et al., 2003).

Histones can also be modified by ADP-ribosylation on glutamine and arginine in response to DNA damage (Hassa et al., 2006), addition of the sugar β-N-acetylglucosamine (Sakabe et al., 2010), deimination of arginine to citrulline (Cuthbert et al., 2004; Wang, 2004), and addition of the bulky protein modifiers ubiquitin and SUMO (Robzyk, 2000; Seeler and Dejean, 2003). Mono-ubiquitination of lysine residues is mostly associated with transcriptional activation, whilst SUMOylation, which can exclude ubiquitin, is associated with repression (Bannister and Kouzarides, 2011; Nathan et al., 2006).

1.2.1.4 - Histone PTM Cross-Talk

As demonstrated above, the breadth of histone PTMs is vast, with different modifications occurring on different tails in response to different stimuli. This observation led to the proposal of the 'histone code' hypothesis, which stated that 'multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions' (Strahl and Allis, 2000).

There are four main ways in which this 'histone code' code may have a downstream effect (Bannister and Kouzarides, 2011). The first is that different modifications may compete for the same residue, especially if that residue is a lysine, which is subject to acetylation, methylation and ubiquitination. An example of this is H3K27, which can be both acetylated and methylated; acetylation appears to mark active enhancer elements (Creyghton et al., 2010), whilst methylation of H3K27 is indicative of silent heterochromatin (Cao et al., 2002). Modifications present on adjacent residues may also preclude binding of specific protein factors, such as the phosphorylation of H3S10 excluding the binding of HP1 to methylated H3K9 (Fischle et al., 2005; Mellone et al., 2003). It has been demonstrated that some modifications are only placed if others are present in the vicinity, a classical example of this is the pre-requisite for H2BK123 ubiquitination for methylation of H3K4 and H3K79 *in S. cerevisiae* (Lee et al., 2007). Finally, a specific combination of modifications may act as a binding platform for protein factors. This may occur on the same histone tail, as is the case with the TAF1 transcription initiation factor which simultaneously binds H4K5ac and H4K12ac (Jacobson, 2000). Binding may also occur at two modified residues on different tails within the same

nucleosome (intranucleosomal) or on different nucleosomes (internucleosomal) (Musselman et al., 2012); intranucleosomal binding is demonstrated by the BPTF subunit of the NURF chromatin-remodelling complex, which binds both H3K4me3 via its PHD finger and H4K16ac via its bromodomain (Ruthenburg et al., 2007). The most complex manifestation of this multiple binding is exhibited by large multisubunit complexes, which contain protein 'readers' with different domains capable of recognising various marks, for example the PRC2 polycomb complex contains subunits harbouring WD40 domains, Tudor domains and PHD fingers (Margueron and Reinberg, 2011).

1.2.2 - Histone Variants

The exchange of canonical histones (H2A/H2B/H3/H4) for variants has the ability to define specialised regions of chromatin. This exchange, unlike deposition of canonical histones, is usually carried out independently of DNA replication at various points throughout the cell cycle (Henikoff and Smith, 2015). To date, variant histones related to H2A and H3 have been extensively described, whilst H2B and H4 variants do not appear to be ubiquitously expressed (Yuan and Zhu, 2012). For a list of histone variants and their conservation, see Table 1.2.

1.2.2.1 - H2A Variants

There have been four variants of the canonical histone H2A described, with varying expression patterns and tissue localisation. H2A.Z is a variant that shares ~60% sequence homology with H2A, with the largest differences

Variant	Conservation	Other names	Function
H2A.Z	Eukaryotes	-	Transcription activation/repression
H2AX	Eukaryotes (not nematodes)	-	DNA damage repair
macroH2A	Vertebrates	-	X-chromosome inactivation
H2A.Bbd	Mammals	-	Transcription activation
H3.1	Eukaryotes	H3 (*Sc, Sp*) H3.2 (*Mm, Dm, Xl*)	Replication-coupled replacement
H3.2	Eukaryotes	H3 (*Sc, Sp*) H3.1 (*Mm*)	Replication-coupled replacement
H3.3	Eukaryotes	H3 (*Sc, Sp*)	Replication-independent replacement
CENP-A	Eukaryotes (not kinetoplastida)	Cse4 (*Sc*) Cnp1 (*Sp*) CID (*Dm*) HTR12 (*At*) HCP-3 (*Ce*)	Specification of centromere

Table 1.2 – The conservation and function of the major histone variants

Table of the main H2A and H3 histone variants, showing the conservation and function of these proteins. See body text for full details.

(*At = Arabidopsis thaliana, Ce = Caenorhabditis elegans, Dm = Drosophila melanogaster, Mm = Mus musculus, Sc = Saccharomyces cerevisiae, Sp = Schizosaccharomyces pombe, Xl = Xenopus laevis*)

present in the C-terminal 'docking' domain, and the L1 loop which acts as the interface between two H2A molecules (Suto et al., 2000).This variant is essential in most organisms studied (Henikoff and Smith, 2015) with the exception of yeast, however H2A.Z deletion is synthetically lethal when combined with a number of other mutants in these single celled eukaryotes (Dhillon et al., 2006; Santisteban et al., 2000; Tada et al., 2011). H2A.Z deposition occurs via the SWI/SNF family chromatin remodeller complex SWR1 (Mizuguchi, 2004); this complex interacts with the histone chaperones Nap1 and Chz1 to mediate the exchange of chromatin loaded H2A-H2B dimers with free H2A.Z-H2B dimers in a stepwise manner (Luk et al., 2007; Straube et al., 2010). This loading of H2A.Z is antagonised by the INO80 complex, which can remove H2A.Z from chromatin (Papamichos-Chronakis et al., 2011).

H2A.Z makes up 5-10% of the total H2A protein pool in most organisms, and although this variant is widely incorporated into chromatin, its distribution is not uniform (Henikoff and Smith, 2015), and depends greatly on the organism. In budding yeast, worms, and plants, H2A.Z appears to localise at genes poised for activation (Kumar and Wigge, 2010; Whittle et al., 2008; Zhang et al., 2005), whilst in fission yeast it is enriched at the 5' end of lowly transcribed genes (Buchanan et al., 2009), and also plays a role in supressing antisense read-through transcription (Zofall et al., 2009). In mammals and flies, however, this variant is more likely to be associated with actively transcribing genes (Hardy et al., 2009; Mavrich et al., 2008). In contrast to this apparent role in transcriptional activation, H2A.Z is also deposited into silent

chromatin, and may act as boundary factor to prevent to encroachment of heterochromatin into euchromatin (Venkatasubrahmanyam et al., 2007).

Another related H2A variant is known as H2AX; this histone contains a unique C-terminal SQ(E/D)Φ (where Φ is a hydrophobic residue) motif which is phosphorylated in response to DNA double strand breaks (DSBs) to give γH2AX (Downs et al., 2000; Rogakou et al., 1998). Although not essential for the eventual repair of DSBs, γH2AX is thought to recruit and retain vital repair factors at the break site (Stucki et al., 2005) as well as stabilising the surrounding chromatin environment via recruitment of specific factors (Ünal et al., 2004).

Aside from these well characterised variants, there is also a vertebrate-specific macroH2A with an extended C-terminus (Pehrson and Fried, 1992), which has been found to localise on the inactive X chromosome and is reported to be involved in the maintenance of inactivation (Costanzi and Pehrson, 1998). Conversely the most recently discovered H2A variant, H2A.Bbd (Barr body deficient) is completely absent from the inactive X chromosome, and is implicated in gene activation (Chadwick and Willard, 2001; Soboleva et al., 2011).

1.2.2.2 - H3 Variants

The number of histone H3 variants present in an organism is dependent on species (Yuan and Zhu, 2012), however the two key histones retained across all eukaryotes are histone H3.3 and CENP-A. Histone H3.3 differs from canonical H3 by only four amino acids, three within the conserved histone fold domain and one on the N-terminal tail (Hake and Allis, 2006). These four substituted

residues are enough to discriminate histone H3.3 from canonical H3 and direct replication independent deposition via the HIRA histone chaperone, as opposed to replication dependent incorporation via CAF1 which utilises canonical H3 (Ahmad and Henikoff, 2002; Tagami et al., 2004). H3.3 is most enriched at actively transcribed genes, however it is also incorporated into silent regions such as telomeres in a pathway that is independent of the HIRA chaperone (Ahmad and Henikoff, 2002; Goldberg et al., 2010; Mito et al., 2005). It is believed that this variant may play the role of the replacement H3 histone outside of DNA replication. Active transcription is a dynamic process, and even proteins associated with silent chromatin bind with a short residency time, thus histone turnover at these loci is common. Rather than 'marking' active genes, it may be that histone H3.3 is most enriched at loci where histone turnover is most prevalent, thus it is still able to be incorporated into silent regions, but at a lower level as these regions are less dynamic (Henikoff and Smith, 2015).

Perhaps one of the most important histone H3 variants across all eukaryotes is the centromeric variant CENP-A (also known as Cse4 in *S. cerevisiae*; Cnp1 in *S. pombe*; CID in *D. melanogaster*; HRT12 in *A. thaliana* and HCP-3 in *C. elegans*) (Müller and Almouzni, 2017; Palmer et al., 1991). The centromere is one of the defining characteristics of eukaryotic chromosomes and is the site at which the mitotic spindle attaches, in order to mediate proper segregation (Henikoff and Smith, 2015). In most eukaryotes the CENP-A nucleosome is a key feature of centromeric chromatin and is absolutely essential for viability (Howman et al., 2000; Regnier et al., 2005; Stellfox et al., 2013; Stoler et al., 1995); however a number of species, such as *Trypanosoma*, butterflies, moths

and dragonflies, have lost CENP-A altogether (Drinnenberg et al., 2014; Lowell and Cross, 2004). This loss of CENP-A is associated with a switch from monocentricity, where chromosomes have a single point of microtubule attachment, to holocentricity, where spindle fibres attach along the length of the chromosome arms (Drinnenberg et al., 2014).

This histone variant has two key domains: a histone fold that is 62% identical to histone H3 and contains a CENP-A targeting domain (CATD), and an N-terminal tail that has rapidly evolved between species (Sullivan et al., 1994; Talbert and Henikoff, 2010a). The in vivo structure of the CENP-A containing nucleosome is the subject of some debate within the field (Fukagawa and Earnshaw, 2014). Some groups propose that it exists as a tetrameric hemisome (Dalal et al., 2007) that wraps DNA in a right-handed helix (Furuyama and Henikoff, 2009), and that this structure may be dynamic with the canonical octameric nucleosome structure (Bui et al., 2012; Shivaraju et al., 2012). Other groups have argued that CENP-A containing nucleosomes exhibit a structure that is very similar to the canonical nucleosome with some subtle differences (Tachiwana et al., 2011), and recent work seems to support the assertion that most CENP-A nucleosomes exist as octamers in vivo (Miell et al., 2013; Padeganeh et al., 2013; Zhang et al., 2012b). However, a consensus is yet to be reached (Fukagawa and Earnshaw, 2014).

As with other histone variants, deposition of CENP-A is uncoupled from DNA replication, however the timing of CENP-A incorporation is restricted to certain points of the cell cycle (Yuan and Zhu, 2012). This regulation occurs both temporally, via the action of cyclin dependant kinases (CDKs) (Silva et al., 2012), and spatially in some organisms, via the action of E3 ubiquitin ligases

that target non-centromeric CENP-A for removal and degradation (Hewawasam et al., 2010; Ranjitkar et al., 2010), as well as via recruitment of the deposition machinery to sites of CENP-A enrichment (Moree et al., 2011). The CENP-A deposition pathway contains a number of steps. Prior to incorporation, the centromere must be primed for CENP-A loading via the action of the Mis18 complex. This complex was first described in *S. pombe*, where *ts* mutants of Mis18, as well as the related Mis16, were found to have serious chromosome segregation defects (Hayashi et al., 2004). Homology searches revealed three homologues in human, Mis18α and Mis18β, as well as Mis18BP1^{KNL2} which were shown experimentally to be essential for CENP-A localisation (Fujita et al., 2007). The exact function of the Mis18 complex in centromere priming is still under investigation, however in *S. pombe* the *ts* mutants of Mis18 exhibit increased centromere acetylation (Hayashi et al., 2004), whilst human mutants have less acetylated lysine residues at centromeric loci (Fujita et al., 2007). The loss of Mis18 can be rescued in human cells by tethering a HAT at the centromeric loci (Ohzeki et al., 2012), indicating that the acetylation of this region is important for CENP-A loading. Although the requirement for acetylation seems to differ from yeast to humans, modification of the chromatin environment seems to be a prerequisite for CENP-A loading. Targeting of the Mis18 complex to centromeres appears to be via interaction with the Constitutive Centromere-Associated Network (CCAN), a multiprotein complex that is associated with centromeric loci throughout the cell cycle (the exact nature of which will be discussed in following sections) (Dambacher et al., 2012; Moree et al., 2011).

Like all other identified histone proteins, deposition of CENP-A requires the action of a specific histone chaperone protein, this is HJURP in humans, Scm3 in yeast and CAL1 in flies (Stellfox et al., 2013), although the CAL1 protein also seems to function in place of Mis18 in this organism (Mellone et al., 2011). These CENP-A specific chaperones bind to the histone via its unique CATD, which allows discrimination from canonical H3 (Shuaib et al., 2010; Zhou et al., 2011). HJURP binds CENP-A-H4 as a prenucleosomal complex that gets localised to centromeres (Foltz et al., 2009). This targeting of HJURP to centromeric loci appears to be via direct interaction between the tetrameric Mis18 complex and HJURP itself; direct interaction between HJURP (Scm3 in yeast) and Mis18 has been observed in both *S. pombe* (Pidoux et al., 2009; Williams et al., 2009) and human cell lines (Nardi et al., 2016). Once deposited at the centromere, CENP-A may need to be 'licenced' in order to allow for formation of the kinetochore structure required for chromosome segregation (Fukagawa and Earnshaw, 2014). This may occur via a histone PTM, as H4K20me1 has recently been shown to occur specifically on centromeric CENP-A nucleosomes (Hori et al., 2014).

It is clear that CENP-A is key for defining the centromere; not only is it required for this process, but in some systems it is sufficient to form a functional centromere at an ectopic locus (Guse et al., 2011; Mendiburo et al., 2011). Although the specification of the active centromere appears to be down to deposition of CENP-A, these regions are commonly found between blocks of pericentromeric heterochromatin (Stellfox et al., 2013). Indeed it has been shown that in *S. pombe*, disruption of this heterochromatin is enough to inhibit de novo deposition of CENP-A (Folco et al., 2008). Similarly,

nucleation of ectopic neocentromeres generally occurs in locations close to regions of constitutive heterochromatin (Ishii et al., 2008; Olszak et al., 2011), thus implying that this restrictive chromatin environment is permissive to de novo centromere formation.

1.3 - The Centromere

1.3.1 - DNA elements underlying the centromere

Centromeres are absolutely essential for all organisms to be able to faithfully segregate their genetic material, as these regions act as the major nucleation point for the microtubules that physically separate sister chromatids (Plohl et al., 2014). As mentioned above, the key determinant of a centromere seems to be the localisation of CENP-A, and it has been demonstrated that centromeres are able form on any DNA sequence (Scott and Sullivan, 2014), although there does seem to be some evolutionary preference for the underlying DNA sequence at established centromeric loci (Plohl et al., 2014). In rare cases organisms may possess holocentric centromeres, where spindle fibres attach along the whole length of the chromosome as in *C. elegans* (Dernburg, 2001). However, it is much more common to find monocentric chromosomes in eukaryotes, with one single point of microtubule attachment.

The defining feature of many eukaryotic centromeric regions is a core centromere flanked by blocks of pericentromeric heterochromatin. These discrete regions co-operate to mediate the formation of the kinetochore and attachment of the microtubules, in order to facilitate proper chromosome segregation. Although this structure is largely conserved between species, the

sequences that underlie these regions generally are not, and most centromeric loci form over 'regions' of repetitive sequence, between 10-10,000kb long.

The most well studied exception to this is the *S. cerevisiae* 'point' centromere (Figure 1.2), which consists of a conserved 125bp sequence motif made up of three conserved DNA elements (CDEI-III) (Henikoff and Henikoff, 2012). In this species the underlying DNA sequence is key to the nucleation of a functional centromere, as point mutants within these CDEs cause large scale chromosome segregation defects (Gaudet and Fitzgerald-Hayes, 1989). The AT rich sequences at these centromeres are required to interact with the CENP-A^{Cse4} histone chaperone Scm3 (Xiao et al., 2011), which deposits a single CENP-A^{Cse4} nucleosome at each centromere (Henikoff and Henikoff, 2012). There are other reports of 'repeat-free' centromeres in nature; for example in orangutan, horse, chicken and potato (Gong et al., 2012; Locke et al., 2011; Shang et al., 2010; Wade et al., 2009). However these are believed to be evolutionary new centromeres (ENCs), which have recently evolved and as yet still lack the repetitive flanking DNA required to 'stabilise' these centromeres, and as such represent "centromeres in progress" (Dumont and Fachinetti, 2017).

Most eukaryotes, from plants, to animals, to fission yeast, form their centromeres over repetitive sequences that usually consist of tandem arrays of satellite DNA repeats, mobile DNA elements such as retrotransposons, or a combination of both (Figure 1.2). In many organisms these satellite DNA repeats are very short; they are 171bp long α-satellites in humans, 120bp in mouse, 178bp in *Arabidopsis*, 156bp in maize, 340bp in pig and 359bp in *Drosophila* (Dumont and Fachinetti, 2017). These short repeats may be

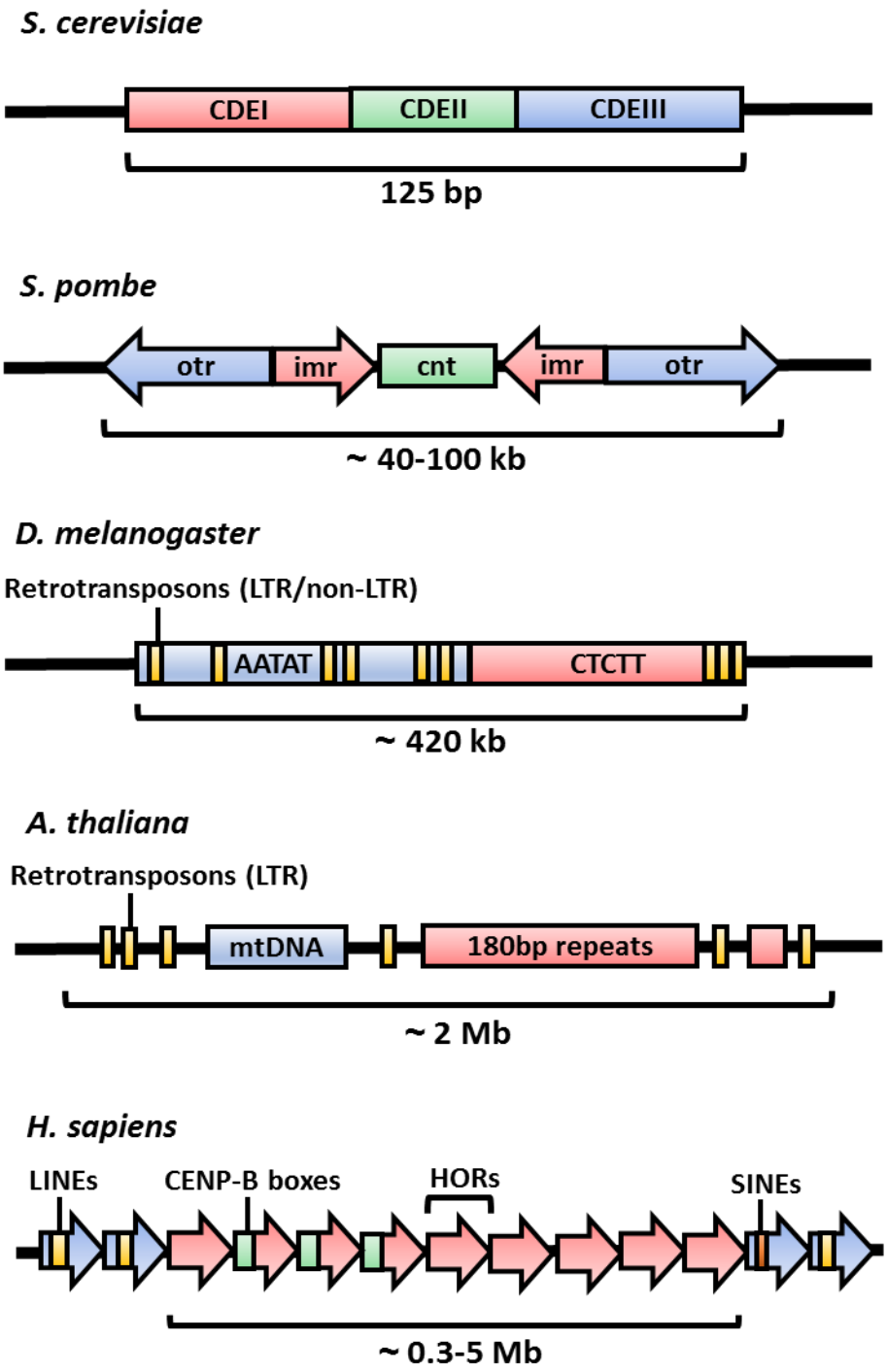


Figure 1.2 – DNA elements underlying the centromere

Schematic representation of centromeric DNA organisation in the budding yeast S. cerevisiae, the fission yeast S. pombe, the fruit fly D. melanogaster, the thale cress A. thaliana and human H. sapiens.

CDE = centromere DNA element, otr = outer repeat, imr = innermost repeat, cnt = central core, LINE = long-interspersed nuclear element, SINE = short-interspersed nuclear element, HOR = higher order repeat

organised into higher order repeats (HORs); in humans and great apes HORs are repeated hundreds to thousands of times to give rise to arrays of between 0.34-6kb, and these arrays can then form centromeres of between 0.3-5Mb (Dumont and Fachinetti, 2017). In some organisms satellite sequences are longer and less repetitive; for example, *S. pombe* centromeres contain a conserved central core sequence flanked by inverted *imr* repeats. These are then flanked on either side by 4-5kb dg/dh tandem repeats, of which there are between 2-11 copies, depending on the chromosome. The centromeres in this organism range from 35kb to 110kb long (Wood et al., 2002).

The interspersion of transposons into repetitive centromeric DNA is widespread and is known to occur in humans, flies, maize and wheat. In human cells, the functional centromere is mainly free of non α-satellite sequence, however the surrounding pericentromeric regions contain a number of LINE and Alu elements (Schueler, 2001), whilst *Drosophila* centromeres contain a number of LTR and non-LTR retrotransposons (Sun et al., 1997, 2003). This interspersion of mobile elements is even more evident in maize centromeres, which possess a small amount of the 156bp CenC satellite DNA as well as abundant CRM1-4 retrotransposons (Wolfgruber et al., 2009), and in wheat, which has centromeres almost entirely made up of CRW, Quinta and Weg retrotransposons (Li et al., 2013a).

Although the actual sequences of centromeres differs greatly, even between closely related species, the fact that a majority are formed over repetitive sequence indicates that this type of sequence may play a key role in centromere formation and may have been selected for over evolutionary time, or may be enriched within silent loci to reduce potential deleterious recombination

between repeats (Talbert and Henikoff, 2010b). One suggested role for these repetitive sequences is in the nucleation of heterochromatin, which has been shown to be a key determinant for the loading of CENP-A in both *S. pombe* and flies (Folco et al., 2008; Olszak et al., 2011). In fission yeast, pericentromeric heterochromatin is also required for sister chromatin cohesion prior to chromosome segregation (Bernard, 2001; Nonaka et al., 2002).

The second proposed function of centromeric sequences is in the generation of specific transcripts that may play a structural role in centromere formation. It has been shown that transcripts derived from human and fly centromeric repeats are required to interact with the CENP-A/HJURP complex to somehow mediate the loading of this histone variant; degradation of this transcript reduced centromeric CENP-A and increased the incidence of mitotic defects (Quénet and Dalal, 2014; Rošić and Erhardt, 2016). A similar interaction has been observed in maize, where RNA derived from the CRM retrotransposon directly interacts with CENP-A^{CENH3} (Topp et al., 2004).

It is also believed that these repetitive regions are able to preserve a unique chromosomal architecture, by mediating favourable DNA topology. Electron microscopy and NMR has recently revealed that centromeric DNA may form specific loop motifs, and the repetitiveness of these regions is a major contributor to this secondary structure, which may play a role in microtubule recognition (Aze et al., 2016; Garavís et al., 2015).

Although centromere formation seems to occur independently of specific sequence, there is evidence that specific sequence motifs may facilitate the nucleation of centromere related factors. One well-studied example is the

CENP-B box, which is a 17bp motif found regularly interspersed in α-satellite DNA (Ohzeki et al., 2002). This motif is required for the binding of the CENP-B protein, which has been identified as a key centromere associated factor that plays a role at the interface between CENP-A chromatin and the kinetochore, to facilitate accurate chromosome segregation (Fachinetti et al., 2015).

1.3.2 - The molecular components of the centromere

The ability of chromosomes to segregate depends not just on the formation of centromeric chromatin but also on the assembly of a protein complex known as the kinetochore. The kinetochore is responsible for facilitating spindle attachment; a process that is required to physically separate sister chromatids (Przewloka and Glover, 2009). The key player at the interface of the centromere and kinetochore is the constitutive centromere associated network (CCAN) (Cheeseman and Desai, 2008). This complex localises to centromeres throughout the cell cycle; in vertebrates it consists of 16 'CENP' proteins (CENP-C/H/I/K/L/M/N/O/P/Q/U/R/T/W/S/X) (Foltz et al., 2006; Okada et al., 2006), which form a number of sub-complexes with distinct functions (Gascoigne et al., 2011; Hori et al., 2008; Okada et al., 2006). The presence of a CCAN is conserved from yeast to humans (Przewloka and Glover, 2009), however some species possess a much simpler network; for example, *C. elegans* and *D. melanogaster* have a single component CCAN, consisting of CENP-C alone (Cheeseman et al., 2008; Unhavaithaya and Orr-Weaver, 2013). The CCAN is anchored to centromeres via the conserved factor CENP-C, which has been shown to directly interact with CENP-A, thus connecting the epigenetic specification of the centromere to the formation of the functional

kinetochore (Carroll et al., 2010; Guse et al., 2011). This CENP-A/CENP-C interaction precedes the formation of the full CCAN and acts as a signal for the assembly of the outer kinetochore; this extended network contains a number of key microtubule binding proteins known as the KLN1-Mis12-Ndc80 (KMN) network (Cheeseman et al., 2006; Przewloka et al., 2011; Screpanti et al., 2011). The interaction of the CCAN with the KMN network may be regulated to ensure a full kinetochore is only formed during mitosis via Aurora B kinase dependent phosphorylation of Ndc80; this modification has been shown reduce the affinity of Ndc80 for microtubules (Cheeseman et al., 2006). There is also evidence to suggest that the CCAN may be involve in resisting the forces generated by microtubules (Suzuki et al., 2014).

Another key protein complex associated with centromeres is cohesin. This complex is responsible for ensuring that sister chromatids remain physically associated with one another until they are separated during anaphase; loss of this sister chromatid 'cohesion' could cause premature dissociation prior to spindle formation, and may result in unequal segregation of genetic material to daughter cells (Peters et al., 2008). Cohesin is made up of four subunits: Smc1 (Psm1 in *S. pombe*), Smc3 (Psm3 in *S. pombe*), Scc1 (Rad21 in *D. melanogaster* and *S. pombe*), and Scc3 (Psc3 in *S. pombe*, IRR1 in *S. cerevisiae*) (Peters et al., 2008). This complex forms a distinctive V-shaped heterodimer of Smc1/3, which then binds to Scc1 and Scc3 (Haering et al., 2002, 2004). Cohesin is found to bind throughout chromosome arms at discrete sites: A/T rich cohesin attachment regions in *S. cerevisiae* (Glynn et al., 2004; Laloraya et al., 2000), transcribed regions in *Drosophila* (Misulovin et al., 2008) and at intergenic regions and introns in humans (Wendt et al.,

2008). Cohesin also binds at centromeres in all studied eukaryotic model organisms; this is perhaps its most important region of localisation as centromeres are under direct stress from the pulling of spindle fibres, which must be resisted until metaphase (Eckert et al., 2007; Tanaka et al., 2000).

In budding yeast, cohesin is recruited to the centromere via the DDK-dependent phosphorylation of Ctf19, a homologue of the CCAN component CENP-P, which provides a binding site for the Scc2/4 cohesin loading complex; this mechanism may also be conserved in vertebrate systems (Hinshaw et al., 2017; Takahashi et al., 2008). In organisms with regional centromeres surrounded by pericentromeric heterochromatin, the HP1 protein also plays a role in cohesin recruitment; in *S. pombe* Swi6 is required for cohesin enrichment at centromeres (Bernard, 2001; Nonaka et al., 2002), whilst in animals HP1 is implicated in cohesin loading (Yamagishi et al., 2008) but may function alongside parallel pathways that mediate recruitment (Koch et al., 2008; Peters et al., 2008).

During mitosis centromeres become enriched for cohesin, however this is likely due to the selective loss of this complex from chromosome arms (Waizenegger et al., 2000). Removal of cohesin depends on the phosphorylation of its subunits, either directly via the mitotically activated Polo like kinase 1 (Plk1) (Losada et al., 2002; Sumara et al., 2002), or indirectly by Aurora kinase (Giménez-Abián et al., 2004; Losada et al., 2002). The protection of cohesin at centromeres is mediated via a protein called Shugoshin (McGuinness et al., 2005), which is recruited to centromeres via interaction with phosphorylated histone H2A; this modification is mediated by the kinetochore associated kinase Bub1 (Kawashima et al., 2010). Shugosin

exists in a complex with protein phosphatase 2A (PP2A) and may act to dephosphorylate cohesin subunits (Kitajima et al., 2006), maintaining their association with centoromeric loci until such a time that all chromosomes are bi-oriented on the mitotic spindle (Peters et al., 2008).

1.4 - Transposable Elements

As introduced above, a number of eukaryotic centromeres are interspersed with mobile genetic elements known as transposons. These are an incredibly diverse class of elements that make up a large proportion of many eukaryotic genomes; in humans the genome is made up of around 50% transposons, whereas in maize this number is as high as 70% (Wessler, 2006). These elements fall broadly into two distinct classes, based on the mechanism by which they mobilise, and may either be autonomous, encoding all they need to mobilise themselves, or non-autonomous, requiring factors encoded by separate elements to mediate transposition (Levin and Moran, 2011).

1.4.1 – DNA Transposons

Class II elements are DNA elements that mobilise via non-replicative 'cut-and-paste' mechanism; transposition does not increase the copy number of these elements. Structurally, DNA transposons encode a transposase enzyme that is flanked by tandem inverted repeat (TIR) sequences; the transposase is able to recognise these TIRs and cut the transposon from the donor site, before integrating it at an acceptor site (Levin and Moran, 2011). This class also includes the more recently discovered helitron elements which mobilise via a unique rolling circle mechanism (Kapitonov and Jurka, 2007). Examples of

DNA transposons are the Tn7 element in *Escherichia coli* (Peters and Craig, 2001), which integrates at a precise attnTn7 locus via the action of its encoded TnsD DNA binding protein (Kuduvalli et al., 2001), and the *P-element* in *Drosophila,* which integrates preferentially into replication origins and promoters (Spradling et al., 2011).

1.4.2 – Retrotransposons

Class I elements consist of retrotransposons, which mobilise through an RNA intermediate; the retrotransposition of these elements increases their copy number. Retrotransposons are particularly prevalent in animals and plants and their discovery challenged the central dogma of molecular biology, by demonstrating that RNA could be converted back into DNA via reverse transcription (Finnegan, 2012). There are two main types of retrotransposons, depending on whether these elements are flanked by long terminal repeats (LTRs).

Non-LTR retrotransposons encode one or two transcribed open reading frames (ORFs); the first encodes an RNA binding protein and the second contains a repair related nuclease, reverse transcriptase and sometimes an RNaseH domain. These elements primarily integrate via target primed reverse transcription (TPRT), whereby the genomic DNA is nicked by the nuclease and the resulting ssDNA is utilised to prime reverse transcription of the retroelement. This type of mobilisation results in the integration of 5' truncated retrotransposons, thus these new copies cannot themselves mobilise (Levin and Moran, 2011). The only active elements in human cells are autonomous long interspersed nuclear elements (LINE-1) and the non-autonomous short interspersed nuclear elements (SINEs) Alu and SVA

(Dewannieux et al., 2003; Hancks et al., 2011). The integration of these non-LTR retrotransposons is dispersed throughout the genome (Lander et al., 2001), with some evidence of rare specific integration sites (Conley et al., 2005). In *Drosophila*, however, there are examples of specific non-LTR retrotransposon integration; the HeTA, TART and Tahre elements all mobilise into telomeres and act as telomere ends in the absence of specific telomeric repeats and the telomerase enzyme in this species (Levis et al., 1993).

LTR retrotransposons are divided into 3 main subclasses: the Ty1/Copia family, the Ty3/Gypsy family and the less characterised BEL/Pao elements (Levin and Moran, 2011). Ty1/Copia and Ty3/Gypsy elements are classified based on the order of their encoded proteins; both encode a viral Gag protein and a second polypeptide flanked by LTRs, however Ty1/Copia elements encode a Protease-Integrase-Reverse Transcriptase-RNase H, whilst Ty3/Gypsy elements encode Protease-Reverse Transcriptase-RNase H-Integrase (Havecker et al., 2004). Ty3/Gypsy retrotransposons are closely related to retroviruses, such as enteroviruses and the human immunodeficiency virus HIV-1 (Eickbush and Jamburuthugoda, 2008). The mechanism of mobilisation is conserved between LTR retrotransposons, and retrotransposition results in the insertion of an exact copy of the element elsewhere in the genome (Levin and Moran, 2011). First a promoter within the 5' LTR of the element drives transcription via the host RNA Polymerase II. The resulting mRNA is then translated, before it is assembled into a virus-like particle (VLP) of Gag proteins along with the element-encoded reverse transcriptase and integrase proteins. Reverse transcription then occurs via the following steps (Figure 1.3): (1) Most commonly a tRNA, or in some cases the

5’ end of the transcript primes the reverse transcription reaction. (2) This causes reverse transcription of the U5 and R portion of the 5’ LTR. (3) This ‘minus-strand strong stop DNA’ is then transferred to the 3’ end of the transcript and base pairs to the conserved R region in the 3’ LTR. (4) This primes DNA synthesis from 3’ R to 5’ R region. (5) Concurrent with reverse transcription, RNase H degrades the mRNA template, leaving an RNA fragment at the polypurine tract (PPT) to prime second strand synthesis of the 3’ LTR. (6) The template is then again switched to the 5’ LTR via complementarity with the 5’ R and U5. (7) DNA synthesis then proceeds in both directions, generating double-stranded DNA containing identical LTRs at either end. The element-encoded integrase then binds to this DNA and directs integration of the retrotransposon, usually via interaction with specific protein factors (Finnegan, 2012).

In the case of Ty1 and Ty3 in *S. cerevisiae*, integration is directed upstream of tRNA genes; for Ty1 this occurs 80-700bp from the TSS (Devine and Boeke, 1996), whilst Ty3 integrates 2bp upstream of the TSS (Chalker and Sandmeyer, 1992). Both integration events are directed by interactions between the integrase and factors of the RNA PolIII machinery (Bachman et al., 2005; Kirchner et al., 1995). Budding yeast Ty5 also displays a specific pattern of integration into heterochromatic regions via interaction with the SIR4 chromatin silencing factor (Xie et al., 2001), whilst chromoviruses mediate their integration into heterochromatin via direct interaction with modified histone tails (Gao et al., 2008).

Although transposable elements have evolved mechanisms to allow propagation whilst minimising the effect on host organism fitness, for example

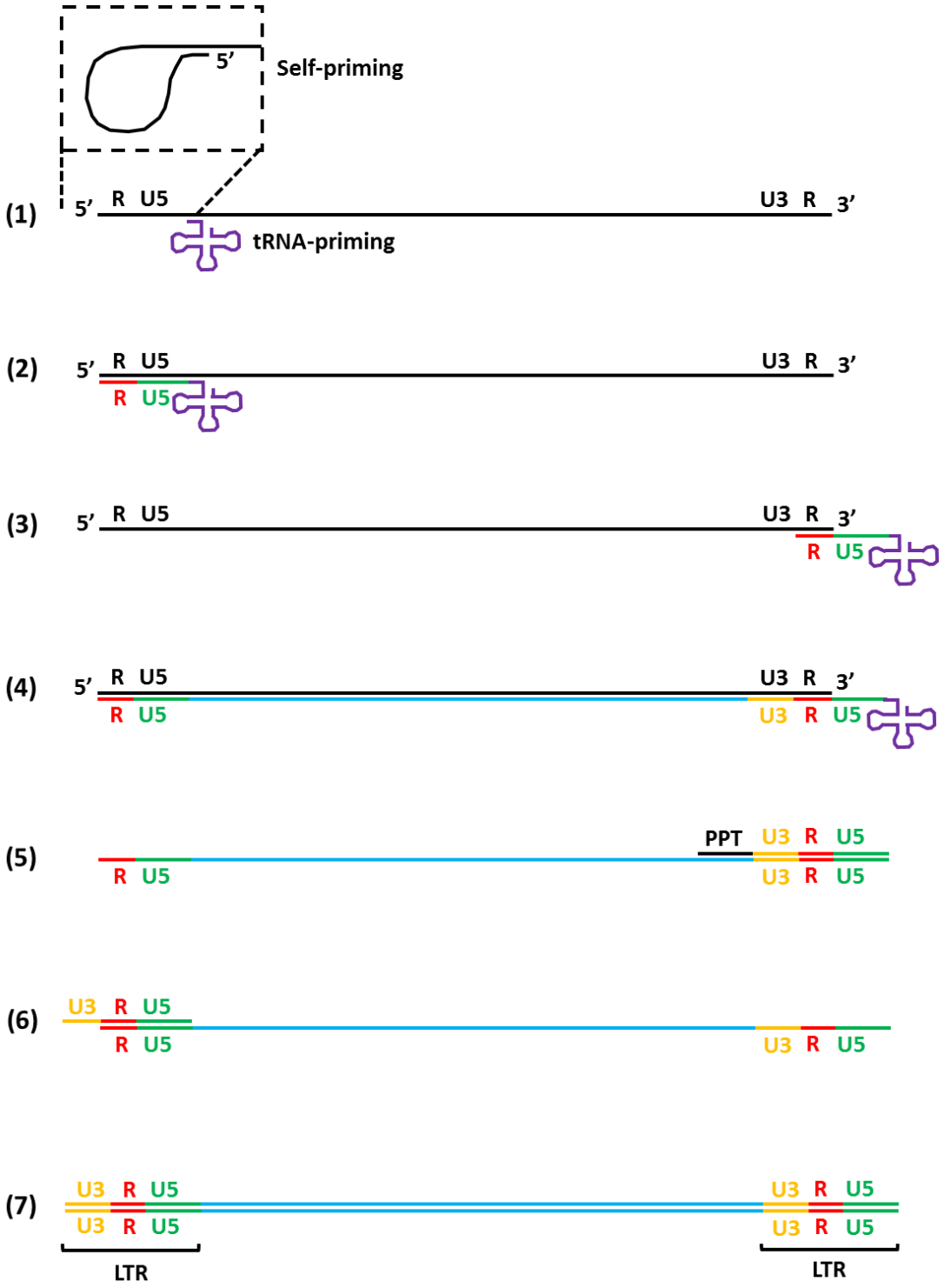


Figure 1.3 – The mechanism of LTR retrotransposon reverse transcription

Schematic diagram showing the process of self-primed and tRNA primed reverse transcription of LTR retrotransposons. See body text for full description of each step.

via integration outside of coding genes and into silent heterochromatin, uncontrolled mobilisation of these elements is undesirable. This is because retrotransposon integration into the genome may be deleterious, and as these elements can make identical copies of themselves they may also be hotspots for recombination (Slotkin and Martienssen, 2007). It is for this reason that most organisms have developed mechanisms to repress transposon activity; one of the most conserved of these pathways is based on silencing via small RNAs and is known as RNA interference (Castel and Martienssen, 2013).

1.5 - RNA Interference

The discovery of RNAi came after the observation that double stranded RNA (dsRNA) was able to robustly silence genes in both *C. elegans* and plants (Fire et al., 1998; Hamilton and Baulcombe, 1999; Waterhouse et al., 1998). These dsRNA species act to silence genes at the post-transcriptional level, either by directing degradation of the target transcript or by directly blocking translation, or at the transcriptional level, by directing chromatin modification to repress transcription (Martienssen and Moazed, 2015). The effects of RNAi are mediated by small RNAs, and these can be broadly split into 3 classes depending on their biogenesis: short interfering RNA (siRNA), micro RNA (miRNA), and Piwi-interacting RNA (piRNA) (Figure 1.4). All of these species share similar characteristics; they are 20-30nt long with a 5' phosphate and 3' hydroxyl group and they associate with specific Argonaute-clade effector proteins. For miRNAs and siRNAs the effector proteins are AGO proteins whilst piRNAs bind to PIWI proteins. Specificity is then conferred via base pairing of these effector loaded small RNAs with target loci (Moazed, 2009).

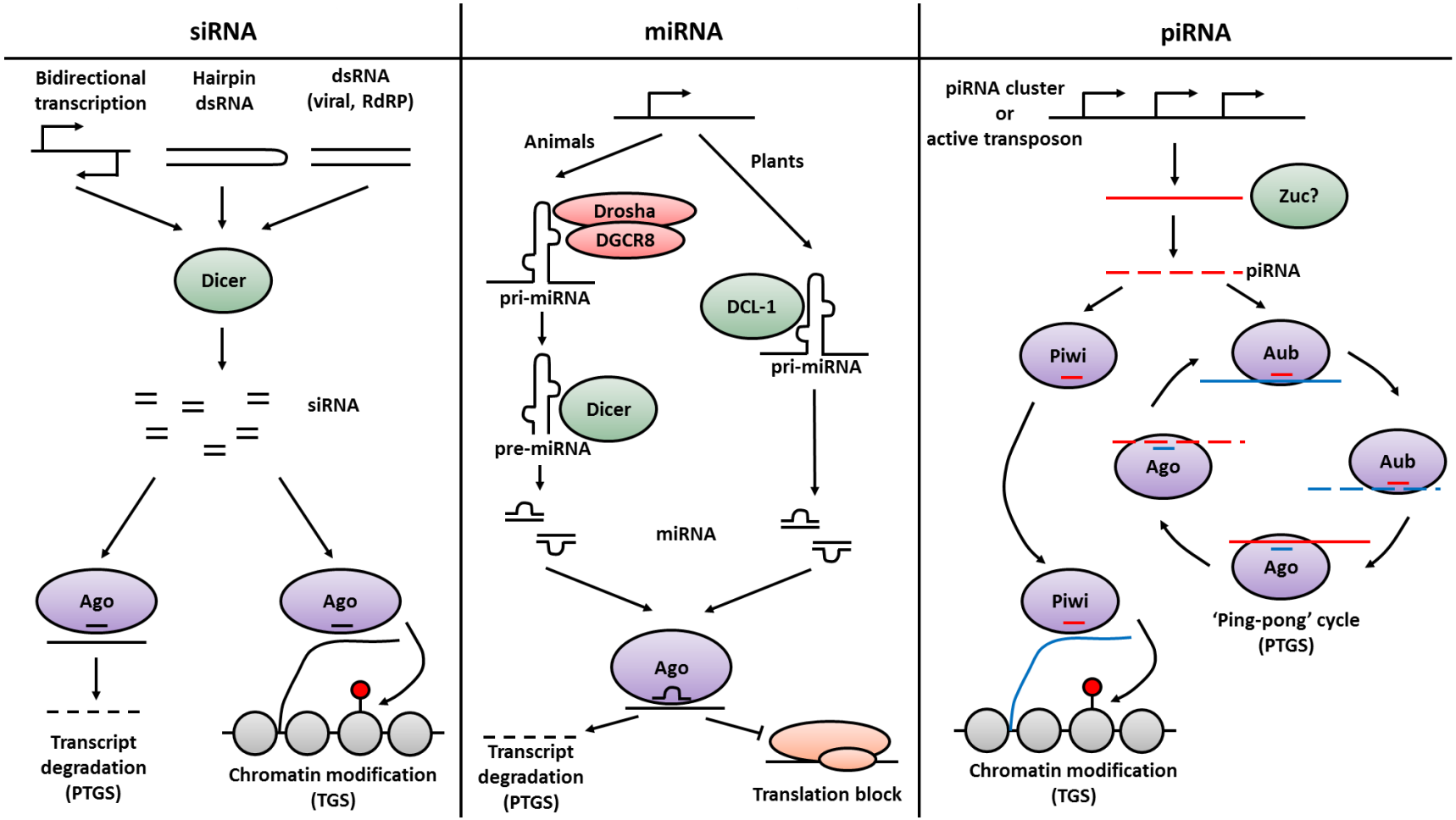


Figure 1.4 – Overview of small RNA-mediated silencing pathways

Schematic diagram showing the mechanism of silencing by siRNAs, miRNAs and piRNAs. See body text for full description of each pathway.

1.5.1 - siRNAs

The most conserved type of small RNAs are siRNAs. These are generated from perfect or near-perfectly paired dsRNA, which may be generated exogenously, from sources such as viral genomes, or endogenously, from hairpin transcripts, convergent transcripts or from base pairing interactions between transcripts of distant loci (Meister and Tuschl, 2004). These long dsRNAs are cleaved by the RNase III ribonuclease Dicer to give species 20-25nt long, with characteristic 2nt 3' overhangs (Vermeulen et al., 2005). siRNAs are well conserved from fungi to mammals, however there is a division as to whether siRNA biogenesis requires the action of an RNA dependent RNA polymerase (RdRP) in these species (Czech and Hannon, 2011).

In mammals and flies, no RdRP activity is required to generate siRNAs; *Drosophila* generates exogenous-siRNAs (exo-siRNAs) from experimentally introduced or viral dsRNA via the action of DCR-2 (Lee et al., 2004b) and the dsRNA-binding partner protein R2D2 (Liu, 2003). This organism also produces endogenous-siRNAs (endo-siRNAs) from sites of convergent transcription, inverted repeats, and transposons via the action of DCR-2 and the partner protein Loquacious-PD (Czech et al., 2008; Ghildiyal et al., 2008; Okamura et al., 2008; Zhou et al., 2009). Endo-siRNAs have been detected in mammals, where their distribution is limited to mouse oocytes and embryonic stem (ES) cells (Babiarz et al., 2008; Tam et al., 2008; Watanabe et al., 2008). Conversely, worms, plants and fission yeast require the action of RdRPs to generate siRNAs, with RdRPs in plants and fission yeast acting to synthesise long dsRNA substrates for Dicer (Czech and Hannon, 2011). In plants there are three subclasses of siRNAs. Secondary siRNAs (which include trans-acting

tasiRNAs, phased phasiRNAs and epigenetically activated easiRNAs) require miRNA directed cleavage of ncRNAs to recruit the RdRP RDR6 to produce dsRNAs; these dsRNAs are cleaved by DCL-2 or DCL-4 into 22nt or 21nt siRNAs respectively. Natural antisense short interfering (nat-siRNAs) are produced from convergent transcripts via DCL-2 to give 24nt siRNAs or DCL-1 to give 22nt siRNAs, while heterochromatic (hc-siRNAs) require RDR2 and DCL-3 to produce 24nt siRNAs (Borges and Martienssen, 2015). In *C. elegans,* however, siRNAs can be produced independently of Dicer; an initial primary siRNA produced by DCR-1 and targeted by the Ago-protein NRDE-1 to transcripts can recruit RdRP to make secondary siRNAs via de novo synthesis (Sijen et al., 2007).

siRNAs get loaded onto Ago-containing effector complexes, known as the siRNA RNA-induced transcriptional silencing complex (siRISC), and loading to specific AGO proteins can occur based on the type of siRNAs. In flies siRNAs are loaded onto AGO-2 via the action of DCR-2/R2D2 (Liu, 2003), whilst in plants AGO2 specifically loads ta-siRNAs and AGO4 loads hc-siRNAS (Czech and Hannon, 2011). These AGO proteins possess catalytic 'slicer' activity, and when loaded with siRNA duplexes, activity to cleave the duplex and eject the passenger strand (Matranga et al., 2005; Rand et al., 2005); which strand is ejected is governed by thermodynamic rules in some species (Khvorova et al., 2003; Schwarz et al., 2003). Depending on the organism, these mature complexes may either function post-transcriptionally in the cytoplasm to degrade transcripts via Argonaute slicer activity in a process known as post-transcriptional gene silencing (PTGS), or they may act to direct chromatin

modification to cause transcriptional repression in a process known as transcriptional gene silencing (TGS) (Moazed, 2009).

TGS was found to occur in the fission yeast *S. pombe* (Volpe, 2002), where RNAi nucleates heterochromatin at pericentromeric repeats via targeting of the nascent transcript generated by RNA PolII at these loci (Allshire and Ekwall, 2015) (see below). TGS also takes place in *Arabidopsis* (Mette et al., 2000), where it utilises specialised, plant-sepcific forms of RNA Polymerase (Ream et al., 2009). RNA PolIV, in combination with the RDR2 RdRP, produce long dsRNA substrates that are cleaved by DCL-3 to give 24nt siRNAs. These siRNAs get loaded to AGO4 and target transcripts emerging from RNA PolV. AGO4 then recruits the DNA methyltransferase DRM2, via the RDM1 protein, to methylate DNA. RDM1 also binds methylated DNA, thus linking siRNA amplification to sites of existing DNA methylation. DNA methylation is also linked to H3K9 methylation, a hallmark of heterochromatin, via the action of the H3K9 methyltransferase KYP/SUVH4, which itself binds methylated DNA. This H3K9 methylation also acts to recruit RNA PolIV, reinforcing the whole loop (Holoch and Moazed, 2015a).

1.5.2 - miRNAs

miRNAs are a species generated from structured hairpin precursors (Cai et al., 2004), and are conserved in plants and animals but absent in fungi and other protozoa (Bartel, 2009). The method of biogenesis differs slightly between plants and animals. In animals, miRNAs are transcribed as polycistronic units by RNA PolII to give primary miRNA (pri-miRNA) species (Lee et al., 2004a). The RNaseIII enzyme Drosha and the DGCR8/Pasha protein form the

microprocessor complex (Denli et al., 2004; Gregory et al., 2004) that cleaves pri-miRNAs to give 60-70nt pre-miRNAs (Han et al., 2006), which are then exported to the cytoplasm (Lund et al., 2004) and cleaved to give 22-23nt miRNA duplexes by the RNaseIII enzyme Dicer (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001). In mammals Dicer interacts with the accessory protein TRBP2 (Chendrimada et al., 2005) and in flies with a factor called Loquacious (Saito et al., 2005). In plants, transcription of pri-miRNAs is similar however there is no Drosha, thus pri-miRNAs are cleaved to form the miRNA duplex by the action of Dicer-Like Protein 1 (DCL1) (Reinhart et al., 2002). This cleavage occurs in several rounds and the duplexes are then methylated by the HEN1 protein (Yu, 2005) to protect them from further modification and subsequent degradation (Ramachandran and Chen, 2008).

In mammals miRNAs are sorted amongst four AGO proteins without discrimination, however in flies miRNAs get loaded specifically onto AGO1, possibly due to the presence of bulges within the miRNA duplex; this AGO1-specific loading is also observed in plants (Czech and Hannon, 2011). The *Drosophila* AGO1, along with three out of four human AGO proteins, have all lost the ability to cleave target RNA, known as 'slicer' activity (Czech and Hannon, 2011). This loss of slicer activity impacts maturation of the mature RISC complex, as the miRNA duplex must be separated and one strand ejected to facilitate target base pairing. In lieu of slicer activity, these RISC complexes may unwind their loaded duplex miRNAs via mismatches in the base-pairing interactions (Kawamata et al., 2009). Together, the miRNA-AGO complex is known as the miRNA RNA induced silencing complex (miRISC). miRISC

complexes mostly act in the cytoplasm to physically block translation of target mRNAs, due to their lack of slicer activity. Although these animal miRISC complexes are cleavage-incompetent, in plants cleavage does occur via miRNAs; this cleavage has also been implicated in the biogenesis of plant trans-acting siRNAs (ta-siRNAs) (Allen et al., 2005).

1.5.3 - piRNAs

The most recently discovered class of small RNAs are the Piwi-interacting RNAs (piRNAs). These species are abundant in metazoan germlines, and act to defend the genome against parasitic elements such as transposons (Luteijn and Ketting, 2013). piRNAs are best characterised in flies, however much work has also been done in both *C. elegans* and mammalian cells. The biogenesis of piRNAs differs greatly from miRNAs and siRNAs; piRNAs display a broader size profile (24-31nt), and are generated independently of Dicer from a single-stranded RNA (ssRNA) precursor (Brennecke et al., 2007; Vagin, 2006). Production occurs via primary and secondary pathways. The primary piRNA biogenesis pathway is initiated by bidirectional transcription of piRNA encoding loci; in *C. elegans* piRNAs (known as 21U RNAs) are individually encoded by separate genes (Billi et al., 2013), however in flies and mice these precursor transcripts are 2-100kb in length and each produce multiple piRNAs from a single piRNA 'cluster' (Brennecke et al., 2007; Li et al., 2013b). piRNAs are then trafficked into the cytoplasm (Zhang et al., 2012a) where the 5' end is specified by nucleolytic cleavage; in flies and mice this is believed to be via the nuclease Zucchini (Zuc) (Nishimasu et al., 2012), however no such factor has been identified in worms (Luteijn and Ketting, 2013). These piRNA

intermediates are then loaded onto PIWI proteins via the action of specific chaperones; in *Drosophila* piRNAs are loaded onto either PIWI or Aubergine (Aub), in mice they are loaded onto Mili or Miwi and in *C. elegans* they are loaded onto PRG1 (Thomson and Lin, 2009). Post loading, the piRNA 3' end is trimmed by a 3'-5' exonuclease known as Trimmer (Izumi et al., 2016); this processed 3' end is then methylated by HEN1 to protect it from further modification and subsequent degradation (Kawaoka et al., 2011; Saito et al., 2007). The fate of these piRNA/PIWI complexes is dependent on the PIWI protein itself; in flies Aub goes on to enter the secondary piRNA pathway, whilst PIWI enters the nucleus (Luteijn and Ketting, 2013).

The piRNA-loaded Aub protein locates transposon mRNA with complementary sequence and cleaves it via its slicer activity. This cleavage event generates new RNA species that are in the sense orientation with respect to the targeted transposon, with a processed 5' end. These species are loaded onto AGO3, and the 3' end is processed by Trimmer to generate secondary piRNAs. These piRNA/AGO3 complexes then target piRNA cluster transcripts, cleaving them to generate new piRNAs phased along the length of the transcript; these piRNAs are then loaded onto Aub. This cycle, known as the 'ping-pong' cycle (Brennecke et al., 2007), diversifies the pool of piRNAs and ensures that the amount of piRNA produced relates to the expression level of the primary piRNA targets. In the nucleus, the piRNA/PIWI complex is responsible for mediating H3K9 trimethylation over euchromatic transposable elements and piRNA clusters, possibly by interacting with the nascent transcript (Luteijn and Ketting, 2013). Interestingly, in *Drosophila*, heterochromatin marks have been found to be required for transcription at

piRNA clusters. The piRNA cluster-specific HP1 variant Rhino binds H3K9me3 (Klattenhoff et al., 2009) and recruits the transcription factor IIA (TFIIA) paralogue Moonshiner, which in turn recruits the transcription factor IID (TFIID) variant TRF2, to initiate transcription within piRNA clusters (Andersen et al., 2017).

In worms, the piRNA pathway functions slightly differently. The 21U RNAs loaded onto PRG1 bind to target RNA in the cytoplasm and recruit RdRP to generate 22G RNAs de novo from this template. These 22G RNAs are then loaded onto worm AGO-9 (WAGO-9), which enters the nucleus and triggers H3K9 trimethylation of target loci, via the action of the nuclear NRDE complexes (Luteijn and Ketting, 2013).

1.6 - Fission Yeast

The process of transcriptional genes silencing (TGS) is perhaps best characterised in the unicellular eukaryote *S. pombe* (Allshire and Ekwall, 2015). This ascomycete fission yeast is an attractive model organism, as it is stably haploid and is straightforward to genetically manipulate. Unlike the budding yeast *S. cerevisiae, S. pombe* has a full RNA interference pathway, which consists of factors conserved in higher organisms. The *S. pombe* RNAi pathway is somewhat simpler however; the genome encodes one Dicer, one Argonaute and one RdRP, whilst it also contains a solitary H3K9 methyltransferase. This simplicity is advantageous, as it makes elucidating the function of each factor more straightforward, without having to worry about redundancy between paralogues.

This species of fission yeast is also a favoured model for the study of chromosome biology; it has only 3 chromosomes that possess large, regional centromeres that share structural features and modifications with a number of higher eukaryotes (Volpe, 2002; Wood et al., 2002). The underlying DNA sequence of these centromeres is somewhat unusual however, as although these regions are repetitive, they lack both short satellite repeats and transposable elements, features conserved amongst the centromeres of species from humans to flies and plants (Plohl et al., 2014). The relatively few transposable elements that are present are dispersed throughout the chromosome arms (Wood et al., 2002) and are not silenced via the RNAi pathway (Cam et al., 2008); this is highly unusual as one of the key functions of RNAi across all kingdoms is in the regulation of mobile genetic elements (Moazed, 2009).

1.6.1 - RNAi in *S. pombe*

In *S. pombe* the RNAi pathway plays a key role in the nucleation and maintenance of heterochromatin at centromeres. Centromeres in *S. pombe* consist of a central core sequence flanked by inverted inner repeats (imr) and larger outer repeats (otr). All three centromeres are very similar in structure, differing only in the number of outer repeats. Cen1 (35kb) has two repeats, cen2 (65kb) has three and cen3 (110kb) has 11-13 (Wood et al., 2002). Interestingly, *S. pombe* lacks DNA methylation, thus the RNAi mediated silencing of these regions is mediated via histone modification. Similar to other higher eukaryotes, the pericentromeres of *S. pombe* are enriched for H3K9

methylation; this is catalysed by the sole H3 lysine 9 methyltransferase Clr4, which is a homologue of SUV39H1 (Rea et al., 2000; Volpe, 2002).

During S-phase, the repetitive pericentomeric repeats are transcribed by RNA Pol II (Chen et al., 2008). The nascent transcript then acts as a binding platform to recruit the siRNA loaded RITS complex (Shimada et al., 2016), consisting of the Argonaute protein Ago1, the GW-protein Tas3 and the chromodomain protein Chp1 (Verdel, 2004). This complex physically interacts with the RDRC complex, which contains the RdRP Rdp1, the poly-A polymerase Cid12 and the helicase Hrr1 (Motamedi et al., 2004). This complex acts to generate dsRNA, using the nascent transcript as a template, and also interacts with the Dicer protein Dcr1, which cleaves the long dsRNA substrate to give siRNA duplexes (Colmenares et al., 2007). These siRNA duplexes are first loaded onto the Arb1, Arb2 and Ago1 containing Argonaute siRNA chaperone (ARC) complex, before Ago1 slicer activity causes ejection of the siRNA passenger strand (Buker et al., 2007). This mature ssRNA loaded Ago1 is then sensed by Tas3 and forms a mature RITS complex (Holoch and Moazed, 2015b), which in turn recruits the exonuclease Triman which has been proposed to trim the 3' end of RITS loaded siRNAs to a mature length of 22nt (Marasovic et al., 2013). The fully activated RITS complex can bind the nascent transcript and via the bridging protein Stc1 (Bayne et al., 2010), recruit the ClrC complex which contains the H3K9 methyltransferase, along with the proteins Raf1, Raf2, Rik1 and Cul4 (Horn et al., 2005). This complex methylates H3K9, which creates a binding platform for the chromodomain proteins Chp2 and Swi6 (Nakayama et al., 2001); Chp2 helps to recruit the histone deacetylase (HDAC) SHREC complex to reinforce heterochromatin

formation (Motamedi et al., 2008), whilst Swi6 has been shown to modulate the loading of cohesin to pericentomeres (Nonaka et al., 2002) (Figure 1.5). The RNAi pathway in *S. pombe* contains a number of complex positive feedback loops; the RITS complex contains the chromodomain protein Chp1, which binds H3K9me and promotes the propagation of the repressive mark along chromatin (Schalch et al., 2009), whilst Clr4 itself also contains a chromodomain which causes self-reinforcement of this modification (Zhang et al., 2008). There is also coupling between heterochromatin and siRNA production, as Swi6 has been shown to stabilise the RDRC complex on the nascent transcript via the RNA silencing factor Ers1 (Hayashi et al., 2012).

There are a number of interesting questions that are yet to be fully addressed in relation to the RNAi pathway in *S. pombe,* one of which is how is the pathway initiated, i.e. where do initial RITS loaded siRNAs originate from to target pericentromeres for silencing via Clr4? Recent work has identified potential species that could fill this role. Primal RNAs (priRNAs) generated independently of Dcr1 by Ago1 and Tri1 have been shown to nucleate low levels of H3K9 methylation; these single-stranded siRNAs are derived from degraded transcript and preferentially target regions of bidirectional transcription such as pericentromeres (Halic and Moazed, 2010; Marasovic et al., 2013). There are also a small number of Dcr1-dependent, Rdp1-independent siRNAs generated, which are postulated to be formed from annealed bidirectional transcripts that may act as the initiators of this pathway (Yu et al., 2014).

Another question pertains to the apparent paradox that regions targeted for transcriptional gene silencing (TGS) require RNA PolII for efficient repression.

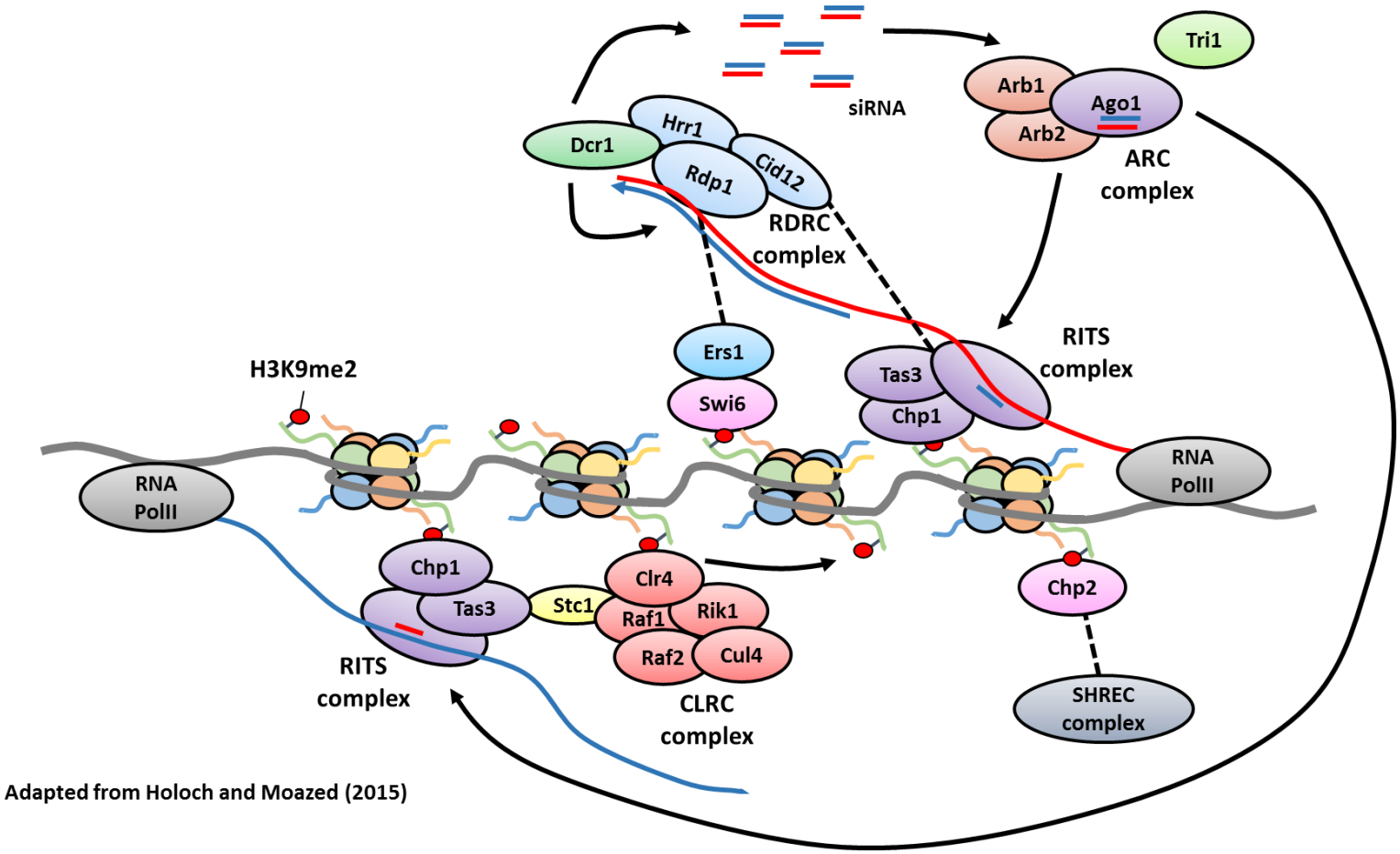


Figure 1.5 – Model of RNAi mediated heterochromatin formation in *S. pombe*

Overview of the formation of heterochromatin by the RNA interference pathway in *S. pombe*. Dotted lines indicate physical interactions not directly shown for clarity. See body text for full description of the pathway.

This is counterintuitive, however mutants in RNA PolII subunits show reduced levels of H3K9 methylation (Djupedal et al., 2005), and the nascent transcript emerging from RNA PolII is required as a platform for recruitment of the RITS complex (Shimada et al., 2016). One recent explanation for this is that H3K9 di- and tri-methylation define distinct chromatin environments. H3K9me2 also co-localises with other histone tail modifications that promote active transcription, thus H3K9 di-methylated chromatin may be a transcriptionally permissive environment. H3K9 tri-methylation is associated with complete transcriptional repression, however the kinetics of its formation are slower than for di-methylation, thus at centromeres H3K9me2 chromatin may form first to permit the siRNA and RNA PolII-dependent spreading of H3K9 di-methylation before the conversion to tri-methylation shuts down transcription (Jih et al., 2017).

1.6.2 - Silencing of retrotransposons in *S. pombe*

S. pombe possesses two families of Ty3/Gypsy LTR retrotransposons, Tf1 and Tf2. Tf2 is the more prevalent of the two, with 13 full length copies present in the genome of the lab strain 972h-, whilst Tf1 seems to be extinct in the same strain, with only solo-LTR regions from this family remaining (Esnault and Levin, 2015). These elements exhibit distinct integration patterns upon mobilisation, with Tf2 integrating predominantly via homologous recombination with existing elements in a process known as target site recycling (Hoff et al., 1998). Tf1 shows a preference for integration upstream of genes involved in stress-response pathways (Feng et al., 2013; Guo and

Levin, 2010), and this integration is able to alter the expression pattern of these genes (Feng et al., 2013).

Across the animal and plant kingdoms, the use of small RNAs to defend the genome against invasive and mobile genetic elements is highly conserved, with siRNAs and piRNAs the main mediators of this repression (Moazed, 2009). In *S. pombe* however, the RNAi pathway plays a minor role in the regulation of retrotransposons, and siRNAs from these elements are only detectable in systems where components of the exosome have been deleted (Yamanaka et al., 2013). Instead, *S. pombe* utilises a mechanism of retrotransposon silencing that employs the mammalian CENP-B homologues Abp1, Cbh1 and Cbh2 (Cam et al., 2008). These proteins are believed to have evolved from a domesticated pogo-like DNA transposase; they also appear to have retained DNA binding activity, as Abp1 has been shown to specifically recognise a 10bp AT-rich motif in the LTR of Tf1 and Tf2 (Lorenz et al., 2012). These proteins then act to recruit the HDACs Clr3 and Clr6, which in combination with the H3K4 methyltransferase Set1 and the HIRA histone chaperone, mediate transcriptional repression of the *S. pombe* retrotransposons (Anderson et al., 2009; Cam et al., 2008; Lorenz et al., 2012). These CENP-B homologue proteins also co-operate with the Pku70/80 heterodimer, which recruits the condensin complex, in order to physically cluster Tf2 elements into discrete subnuclear foci known as 'Tf bodies', localised at the nuclear periphery (Tanaka et al., 2012). Interestingly, transcriptional repression of Tf2 occurs independently of this clustering, thus 'Tf bodies' may serve a different purpose, possibly in the restriction of cDNA integration and recombination between existing elements.

1.6.3 – *S. japonicus*

In 2011 a consortium led by the Broad Institute published the sequenced genomes of the other three member of the Schizosaccharomyces clade; *S. japonicus*, *S. cryophilus* and *S. octosporus* (Rhind et al., 2011). In common with *S. pombe*, all of these species were found to have a genome of around 11.5Mb spread over three chromosomes. The GC content of the *S. cryophilus* and *S. octosporus* genomes was also similar to *S. pombe* at 36-38%, however the *S. japonicus* genome contained 44% GC. This division was also reflected in the conservation of amino acid identity between 1:1 orthologues of these species, *S. cryophilus* and *S. octosporus* were 60% identical to *S. pombe* and 85% identical to each other, whilst *S. japonicus* showed 55% identity to *S. pombe* and 50% identity to both *S. cryophilus* and *S. octosporus*. This indicated that *S. japonicus* was the earliest branching member of the clade; this divergence was calculated to have occurred 221 million years ago, whilst *S. pombe* diverged 119 million years ago, and *S. cryophilus* and *S.octosporus*, the youngest members of the clade, split 32 million years ago (Figure 4.6).

Interestingly, *S. japonicus* was found to contain a much larger and more diverse retrotransposon complement, consisting of 10 families of Ty3/Gypsy LTR retrotransposons, named Tj1-10. This is compared to two families (Tf1/Tf2) in *S. pombe,* one family (Tcry1) in *S. cryophilus* and no full-length retrotransposon families in *S. octosporus*. This apparent eradication of retrotransposons seemed to coincide with the advent of the CENP-B homologue proteins; these are absent in *S. japonicus,* whilst *S. pombe* encodes three homologues (Cam et al., 2008) and *S. octosporus* and *S. cryophilus* each encode six (Rhind et al., 2011). These genes are postulated to have arisen after

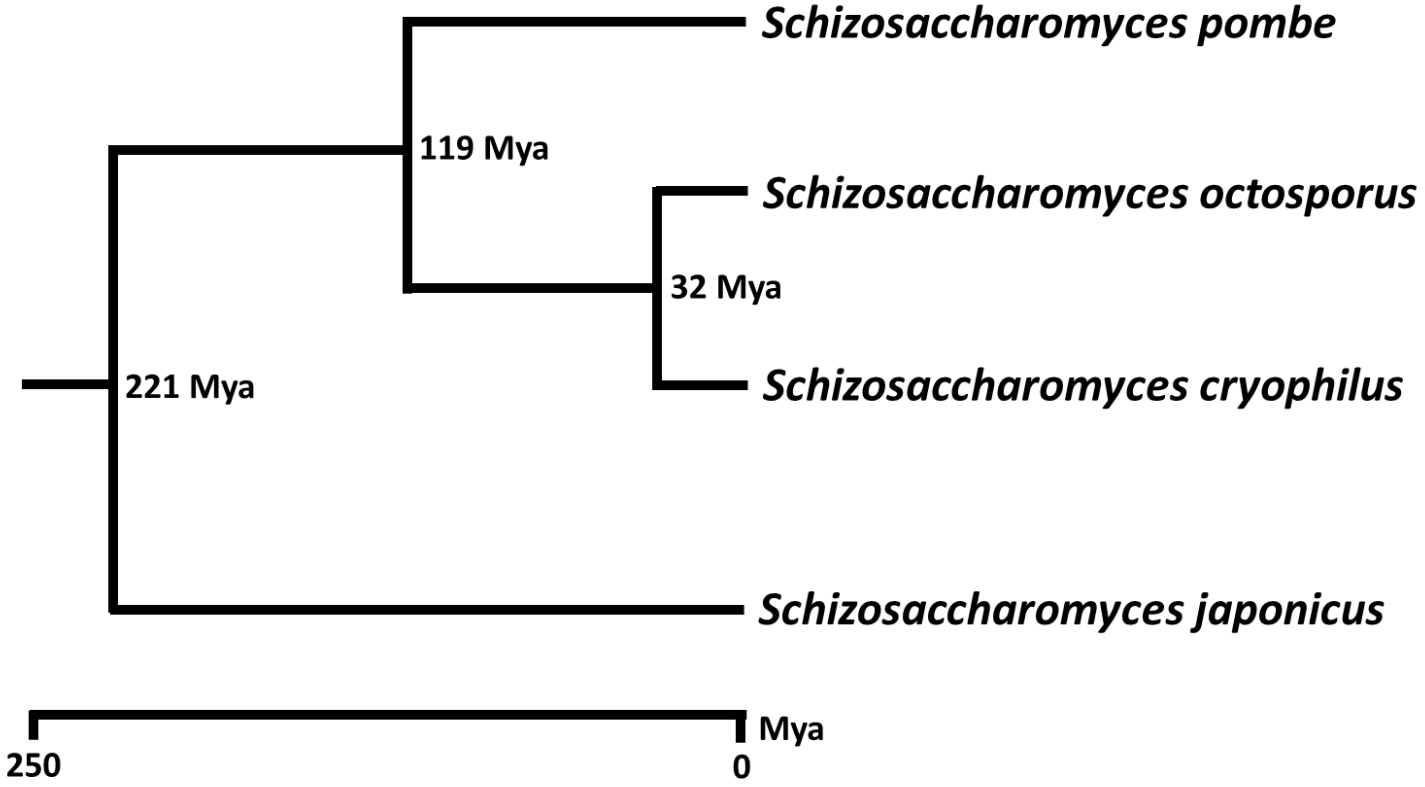
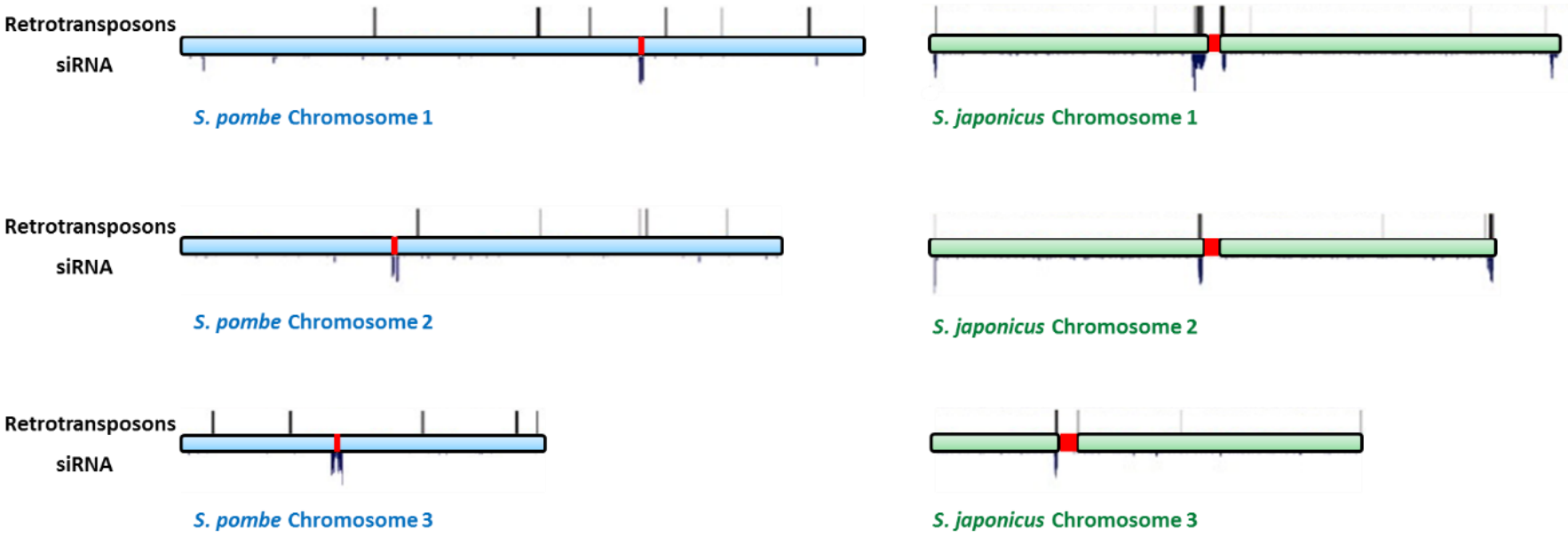


Figure 1.6– Phylogenetic analysis of the four fission yeast species

Phylogenetic tree showing the point at which the four fission yeast species branched. Evolutionary time is shown in millions of years ago (Mya).

S. japonicus diverged, and in *S. pombe* these proteins acts to repress retrotransposons. In the absence of these factors, *S. japonicus* appears to have retained the ancestral role for RNA interference in the silencing of transposable elements; indeed, 94% of sequenced small RNAs in this species map to mobile elements (Rhind et al., 2011) (Figure 1.7). The loss of small RNA-mediated retrotransposon silencing also seems to correlate with the evolution of specific pericentromeric repeat sequences and the absence of retrotransposons at these loci. *S. japonicus,* however, appears to have co-opted retrotransposons to nucleate heterochromatin at pericentromeres and telomeres, an arrangement that is reminiscent of a number of higher eukaryotes, such as maize, wheat and to some extent, human cells (Plohl et al., 2014) (Figure 1.7).



Adapted from Rhind *et al.* (2011)

Figure 1.7 – Comparison of retrotransposon and siRNA distribution in *S. pombe* and *S. japonicus*

Schematic representation of the three chromosomes from *S. pombe* and *S. japonicus*, showing the position of the annotated retrotransposons (above) and the mapped siRNAs (below). The position of the centromeres is marked in red.

1.7 – Aims

S. pombe is one of the key model organisms that has been employed to study the small RNA mediated formation of heterochromatin at repetitive regions; however, this species of fission yeast is somewhat atypical in regards to its utilisation of RNA interference, and the substrates on which this pathway acts. In *S. pombe,* retrotransposons are not targeted for regulation by RNAi and these mobile elements are dispersed throughout the genome; this is in contrast to what is observed in many higher eukaryotes, where retrotransposons make up large portions of centromeric sequence, and are also targeted for silencing by small RNAs. This localisation of RNAi controlled transposable elements at centromeres and telomeres does seen to have been conserved in the related fission yeast *S. japonicus*; thus the main aim of this project was to dissect the role of the RNA interference pathway in this species, and to specifically address whether this pathway contributes to heterochromatin formation and retrotransposon regulation.

Chapter 3 describes my efforts to construct genetic deletions for a number of core RNAi and heterochromatin factors and then assess the phenotypes of these mutants in relation to retrotransposon regulation and heterochromatin formation. Through this work it became clear that perturbation of the RNAi pathway in *S. japonicus* had a greater impact on organism fitness than equivalent deletions made in *S. pombe*. In spite of this, low frequency mutations in the siRNA generating ribonuclease Dcr1 were isolated. As this was the only deletion mutant that could be recovered, I went on to characterise the role of Dcr1, as described in Chapter 4. To achieve this, I aimed to assess the impact of losing Dcr1 on the production of small RNAs, as well as the global

changes in transcription associated with this mutation. I also wanted to investigate the impact of disrupting this factor on the formation of heterochromatin, and establish whether Dcr1 was required to restrict mobilisation of retrotransposons.

The data generated in Chapter 4 presented the opportunity to expand the annotation of the *S. japonicus* genome sequence, thus in Chapter 5 I describe the work I undertook to accomplish this. I aimed to utilise data from genome wide sequencing of small RNAs and mRNAs to identify unannotated Dcr1 regulated loci and subsequently expand the annotation of the *S. japonicus* retrotransposon complement. This would help to improve the utility of *S. japonicus* as a model for the study of RNAi and heterochromatin formation, as well as retrotransposon regulation.

From the work undertaken in Chapter 3, 4 and 5 it appeared that *S. japonicus* employed RNAi to silence its endogenous retrotransposons. This is in direct contrast to *S. pombe,* which utilises homologues of the mammalian CENP-B proteins to transcriptionally silence endogenous retroelements. As these two related species of fission yeast employ different mechanisms of silencing retrotransposons, the secondary aim of the project was to study how a retroelement from *S. japonicus* would be regulated in *S. pombe*. To do this, I aimed to introduce a plasmid-borne retrotransposon from *S. japonicus* into a number of *S. pombe* mutant backgrounds, to study how different factors impacted integration and silencing of this element. This approach would also allow me to study the integration pattern of the *S. japonicus* retrotransposon, and draw conclusions about the mechanism underpinning mobilisation.

Chapter 2 – Materials and methods

2.1 – Cloning and Fragment construction

2.1.1 – Preparation of chemically competent *E. coli*

A single fresh streaked DH5α *E. coli* colony was inoculated into 5ml of LB and grown overnight at 37°C with shaking at 250rpm. The culture was then diluted 1:200 into 100ml pre-warmed LB+20mM $MgSO_4$ and grown at 37°C with shaking at 250rpm until OD_{600} reached ~0.48. Cells were incubated on ice for 10 minutes then pelleted at 5000rpm for 5 minutes at 4°C, the pellet was gently resuspended in 40ml ice-cold TFB1 (30mM KAc, 100mM RbCl, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% glycerol, pH 5.8) and incubated on ice for 5min. Cells were pelleted at 3000rpm for 10 minutes at 4°C, the pellet was gently resuspended in 4ml TFB2 (10mM MOPS, 10mM RbCl, 75mM $CaCl_2$, 15% glycerol, pH 6.5) and incubated on ice for 15 minutes. 100µl aliquots of the cell suspension were aliquoted into pre-chilled Eppendorf tubes and frozen at -80°C.

2.1.2 – *E. coli* transformation

50µl of chemically competent DH5α cells were thawed on ice, 10-20ng of plasmid DNA was added and mixed, and the suspension was incubated on ice for 30 minutes. Cells were heat-shocked at 42°C and recovered on ice for 2min, 250µl of S.O.C media was then added and the suspension incubated at 37°C for 60min with shaking at 250rpm. Cells were plated over several volumes (150µl/100µl/50µl) onto LBAgar+50mg/ml carbenicillin and plates incubated at 37°C overnight until single colonies appeared.

2.1.3 – *E. coli* Plasmid Purification

A single antibiotic resistant DH5α colony was grown in 5ml LB+50mg/ml carbenicillin overnight at 37°C with shaking at 250rpm. Cells were harvested

at 4500rpm for 5 minutes at room temperature and plasmid DNA purified using the QIAprep Miniprep Kit (Qiagen) per the manufacturer's instructions.

2.1.4 – Pfx PCR

Amplification of fragments for cloning or for use in integrative transformation was carried out using the high-fidelity Platinum™ *Pfx* DNA Polymerase (Thermo Fisher Scientific) per the manufacturer's instructions. 1µl of purified genomic DNA (at 50ng/µl) or plasmid DNA (at 10ng/µl) was used as template with the following cycling conditions:

95°C	-	2 minutes	
95°C	-	1 minute	⎫
50°C	-	45 seconds	⎬ x35 cycles
68°C	-	1 minute/kb	⎭
68°C	-	10 minutes	

5µl of the reaction was checked by agarose gel electrophoresis, then purified using the QIAquick PCR Purification Kit (Qiagen) per the manufacturer's instructions.

2.1.5 – Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to visualise the size and presence of DNA fragments after colony PCR and amplification of fragments for cloning or transformation. As standard, 1% Agarose/TBE (90mM Tris-Borate, 2mM EDTA) gels were used, and run at 100V for 30 minutes to resolve DNA fragments. GeneRuler DNA Ladder Mix (Thermo Fisher Scientific) was used as a size marker and sample DNA was mixed with 1/6th volume of 6x loading dye (25% Ficoll-400, 1% OrangeG, 100mM Tris-HCl pH 7.5, 100mM EDTA pH 8.0).

2.1.6 – Split Marker Fusion PCR

To generate fragments for construction of targeted deletions of selected *S. japonicus* genes, the 5' and 3' portions of either a NatMX6 or *ura*$^+$ cassette were amplified using Platinum™ *Pfx* DNA Polymerase, with ~200bp of overlap between them. The 5' portion of NatMX6 was generated using the primer pair MX6_Fwd/NSR-2 and the 3' portion of NatMX6 was generated using the primer pair NSL-2/MX6_Rev, with pFA6a-NatMX6 as a template. The 5' portion of *ura*$^+$ was generated using the primer pair spUra4_F/spUra4-5Fusion_R, and the 3' portion of spUra4 was generated using the primer pair spUra4-3Fusion_F/spUra4_R, with pIRT2U as a template. To generate flanking homologies for the targeted deletion of selected *S. japonicus* genes, 1kb regions upstream of the *goi*$^+$ start codon, and downstream of the *goi*$^+$ stop codon were amplified. These fragments were generated with 20bp of overlapping homology to the 5' or 3' of either NatMX6 or *ura*$^+$. The 5' flanking homology was then fused to the 5' portion of the resistance marker, and the 3' flanking homology was fused to the 3' portion of the resistance marker via fusion PCR. Reactions were set up as follows: 100ng DNA fragment 1, 100ng DNA fragment 2, 5µl 10x *Pfx* Amplification Buffer, 1µl $MgSO_4$ (50mM), 1.5µl dNTPs (10mM), 0.5µl Platinum™ *Pfx* DNA Polymerase, to 50µl dH_2O. Initial reactions were run using the following cycling conditions:

95°C	-	2 minutes	
95°C	-	1 minute	⎫
50°C	-	45 seconds	⎬ x5 cycles
68°C	-	1 minute/kb	⎭

To each reaction, the following was then added: 5µl 10x *Pfx* Amplification Buffer, 1µl $MgSO_4$ (50mM), 1.5µl dNTPs (10mM), 5µl F Primer (10µM), 5µl R

Primer (10µM), 0.5µl Platinum™ *Pfx* DNA Polymerase, 32µl dH_2O. Reactions were run using the following cycling conditions:

95°C	-	2 minutes	
95°C	-	1 minute	⎫
50°C	-	45 seconds	⎬ x25 cycles
68°C	-	1 minute/kb	⎭
68°C	-	10 minutes	

5µl of the reaction was checked by agarose gel electrophoresis, then purified using the QIAquick PCR Purification Kit (Qiagen) per the manufacturer's instructions (See Table 2.2 for full list of primers).

2.1.7 – Tagging plasmid Construction and linearization

To target genes for C-terminal epitope tagging 1kb of homology from the 3'end of the gene ORF minus the stop codon and 1kb of homology downstream of the targeted gene 3'UTR were amplified by PCR. These fragments contained terminal restriction sites, dependent on the vector they were to be cloned in to. For *chp1*$^+$ and *stc1*$^+$ the 3' ORF homology carried a 5' XhoI and a 3' BamHI site, whilst the downstream homology contained a 5' ApaI and a 3' XhoI site. For *rik1*$^+$ the 3' ORF homology carried a 5' NdeI and a 3' BamHI site, whilst the downstream homology contained a 5' AatII and a 3' NdeI site. The *chp1*$^+$ and *stc1*$^+$ fragments were digested with the appropriate restriction enzymes and ligated into the pSO729 backbone, which contains a C-terminal GFP tag and a *ura4*$^+$ cassette. The *rik1*$^+$ fragments were digested with the appropriate restriction enzymes and ligated into the pFA6a-FLAG-NatMX6 backbone, which contains a C-terminal FLAG tag and a NatMX6. All ligations were performed at 16°C overnight using T4 DNA ligase (NEB) per the manufacturer's instructions. Ligated plasmids were transformed into chemically-competent DH5α *E. coli*. Plasmids were isolated and then sent for

Sanger sequencing. Plasmids of the correct sequence were linearised with either NdeI for the pFA6a-FLAG-NatMX6 vector or XhoI for pSO729 and used to transform *S. japonicus* (See Table 2.2 for full list of primers).

2.2 – Fission Yeast Growth

2.2.1 – *S. pombe/S. japonicus* Media and Drugs

S. pombe and *S. japonicus* were cultured using the same nutrient rich media (YES) and synthetic minimal media (PMG) as both liquid and solid media. *S. pombe* crosses were performed on Malt Extract (ME). The composition of each media is listed below:

YES Liquid: 0.5% (w/v) yeast extract, 3.0% (w/v) glucose, 200mg/l adenine, 200mg/l arginine, 200mg/l histidine, 200mg/l leucine, 200mg/l lysine, 200mg/l uracil

YES Agar: 0.5% (w/v) yeast extract, 3.0% (w/v) glucose, 2% (w/v) agar, 200mg/l arginine, 200mg/l histidine, 200mg/l leucine, 200mg/l lysine, 200mg/l uracil. For YES full adenine plates, 200mg/l adenine was added, for YES low adenine plates 20mg/l adenine was added.

PMG Liquid: 14.7mM potassium hydrogen phthalate, 15.5mM Na_2HPO_4, 3.75g/l glutamic acid, 2% (w/v) glucose, 1x salts, 1x vitamins, 1x minerals. PMG was supplemented with 200mg/l adenine, arginine, histidine, leucine, lysine and uracil as required.

PMG Agar: 14.7mM potassium hydrogen phthalate, 15.5mM Na_2HPO_4, 3.75g/l glutamic acid, 2% (w/v) glucose, 2% (w/v) agar, 1x salts, 1x vitamins, 1x minerals. PMG was supplemented with 200mg/l adenine, arginine, histidine, leucine, lysine and uracil as required.

ME Agar: 3% (w/v) Bacto-malt extract, 2% agar, 200mg/l adenine, 200mg/l arginine, 200mg/l histidine, 200mg/l leucine, 200mg/l uracil.

50x Salts: 260mM $MgCl_2$, 4.99mM $CaCl_2$, 670mM KCl, 14.1mM Na_2SO_4.

1000x Vitamins: 4.20mM pantothenic acid, 81.2mM nicotinic acid, 55.5mM inositol, 40.8µM biotin.

10,000x Minerals: 80.9mM boric acid, 23.7mM $MnSO_4$, 13.9mM $ZnSO_4$, 7.40mM $FeCl_2$, 2.47mM molybdic acid, 6.02mM KI, 1.60mM $CuSO_4$, 47.6mM citric acid.

Drugs and supplements were used at the following final concentrations:

5-Fluorootic Acid (5-FOA) (Melford Laboratories): 1g/l

Nourseothricin (clonNAT) (WERNER BioAgents): 0.1mg/ml

Geneticin (G418) (Gibco): 0.1mg/ml

Thiabendazole (TBZ) (Sigma-Aldrich): 15µg/ml in DMSO

Thiamine (Sigma Aldrich): 5µg/ml for repression of nmt1

Trichostatin A (TSA) (Selleckchem): 35µg/ml in DMSO

2.2.2 – Cell culture and harvest

S. pombe cells were cultured in Erlenmeyer flasks with a total volume five times greater than the volume of liquid culture required. Unless stated, cultures were grown at 32°C with shaking at 180rpm, cell density was calculated using a haemocytometer. Cells were decanted into 50ml Falcon tubes and harvested by centrifugation at 3,000rpm for 2 minutes at 4°C.

S. japonicus cells were cultured in Erlenmeyer flasks with a total volume five times greater than the volume of liquid culture required. Unless stated, cultures were grown at 32°C with shaking at 180rpm, due to the propensity of *S. japonicus* cells to flocculate, cell density was calculated using a Lambda 25

UV/Vis spectrophotometer (Perkin Elmer) set at a wavelength of 595nm. Cells were decanted into 50ml Falcon tubes and harvested by centrifugation at 3,500rpm for 2 minutes at 4°C.

2.2.3 – Genetic crosses

To induce conjugation and sporulation in *S. pombe,* heterothallic strains of opposite mating type (h^+ and h^-) were mixed on Malt Extract (ME) Agar plates lacking nitrogen. Crosses were incubated at 25°C for 2-3 days and ascii detected by light microscopy. A small amount of the cross was resuspended in 300µl of Glusulase enzyme extracted from the snail *Helix pomatia* diluted 1:100 in dH_2O, to digest both the ascus wall and any vegetative cells, leaving only intact spores. Digestions were incubated at 32°C overnight, centrifuged at 4,000rpm for 1 minute and spores resuspended in 300µl dH_2O. 1:100 and 1:1000 dilutions of spores were plated onto the appropriate selective plates. For multiple selective markers, replica plating was performed using velvet and correct crosses were screened by colony PCR.

2.2.4 – Long term storage of fission yeast

After confirmation of the correct genotype by colony PCR, strains were streaked as patches onto the appropriate selective plates. Plates were grown for 3 days at 32°C and the patches scraped into a 1.8ml Nunc® CryoTube® (Sigma-Aldrich) containing 50% (v/v) glycerol in dH_2O. CryoTubes were numbered according to lab convention and stored at -80°C indefinitely.

2.3 – Fission Yeast Transformation and Genotyping

2.3.1 – LiAcTE transformation of *S. pombe*

~1.5x10^8 exponentially growing *S. pombe* cells were harvested and washed once in 50ml dH_2O. The cell pellet was resuspended in 0.1M LiAcTE pH4.95 (0.1M LiAc, 10mM Tris-HCl pH8.0, 1mM EDTA) and incubated at 32°C for 1 hour with shaking at 180rpm. Cells were pelleted at 3000rpm for 2 minutes at 4°C and resuspeneded in 0.1M LiAcTE pH4.95 to a density of 1x10^9 cells/ml. 150µl of cell suspension was mixed with 1-5µg of transforming DNA and 370µl 50% PEG_{3350} in TE, then incubated at 32°C for 60min with shaking at 180rpm. Cells were heat-shocked at 42°C for 20min, then spun at 4000rpm for 1 minute to pellet, and resuspended in 500µl dH_2O. Cells were then plated at several dilutions on selective plates and incubated at 32°C until colonies appeared. For antibiotic resistant transformations, cells were added to 3ml YES and cultured at 32°C overnight with shaking at 180rpm before plating.

2.3.2 – Electroporation of *S. japonicus*

50ml of exponentially growing *S. japonicus* cells at OD_{595} 0.4-0.5 were harvested and washed twice in 50ml ice-cold dH_2O. The cell pellet was resuspended in 5ml ice-cold 1M sorbitol + 50mM DTT and incubated at 32°C for 15 minutes without shaking. Cells were pelleted at 3500rpm for 2 minutes at 4°C and resuspeneded in 5ml ice-cold 1M sorbitol. Cells were pelleted and split equally between two pre-chilled eppendorfs, then washed gently with 600µl ice-cold 1M sorbitol. Pellets were resuspended in a pre-prepared mixture of 50µl transforming DNA (4-5µg), 50µl 2M sorbitol and 0.5µl salmon sperm DNA (Life Technologies), then incubated on ice for 30min. The cell suspension was transferred to a pre-chilled electroporation cuvette (0.2cm

gap) and pulsed at 2.3kV using a Micropulser electroporator (BioRad). 1ml of ice-cold 1M sorbitol was immediately added and the cell suspension transferred to a falcon tube containing 9ml YES. Cells were then recovered at 25°C overnight with shaking at 180rpm. Cells were plated at several dilutions on selective plates and incubated at 32°C until colonies appeared.

2.3.4 – Colony PCR

A pinhead sized amount of *S. pombe/S. japonicus* cells were resuspended in 10µl 20mM NaOH using a toothpick. The cell suspension was incubated at 95°C for 1 hour and 1µl was used as template in a 20µl PCR reaction using Taq DNA Polymerase (Roche) per manufacturer's instructions. The following cycling conditions were used:

95°C	-	3 minutes	
95°C	-	30 seconds	⎫
50°C	-	30 seconds	⎬ x35 cycles
72°C	-	1 minute/kb	⎭
72°C	-	10 minutes	

4µl of Loadng dye was added to each PCR reaction and the full volume was visualised by agarose gel electrophoresis.

2.4 – Fission Yeast Nucleic Acid Methods

2.4.1 – Genomic DNA extraction

10ml of *S. pombe/S. japonicus* cells were grown to early stationary phase in YES. Cells were harvested, resuspended in 500µl SP Buffer (1.2M Sorbitol, 50mM sodium citrate, 50mM Na_2HPO_4, 40mM EDTA, pH 5.6) containing 400µg/ml Zymolyase 100T (Amsbio) and incubated at 37°C for 1 hour. Cells were pelleted, washed in 500µl SP Buffer and resuspended in 500µl TE Buffer

(10mM Tris-HCl, 1mM EDTA, pH8.0). 50µl 10% SDS was added, vortexed well and incubated at 65°C for 10min. 165µl KOAc was added, vortexed well, incubated on ice for 30 minutes then pelleted at 13,000rpm for 10 minutes at 4°C. The supernatant was mixed with 750µl isopropanol and incubated on dry ice for 10 minutes. The precipitated DNA was pelleted, resuspended in 300µl TE Buffer + 1µl (1U) RiboShredder™ RNase Blend (Epicentre) and incubated at 37°C for 1 hour. DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated by adding 1/10th volume of 3M NaOAc and 3 volumes of EtOH with incubation on dry ice for 10 minutes. The pellet was washed once in 70% EtOH and resuspended in 20µl dH_2O. DNA concentration was quantified using a NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.4.2 – Total RNA extraction

2ml of exponentially growing *S. pombe/S. japonicus* cells were harvested and total RNA extracted using the MastePure™ Yeast RNA Purification Kit (Epicentre). RNA concentration was quantified using a NanoDrop 2000 Spectrophotometer (Thermo Scientific), and RNA integrity was assessed by running 1µg of total RNA on a 1% Agarose/TBE gel.

2.4.3 – Small RNA extraction

50ml of exponentially growing *S. pombe/S. japonicus* cells were harvested and washed in H_2O. Pellets were resuspended in 500µl extraction buffer (50mM Tris-HCl pH 7.5, 10mM EDTA pH 8.0, 100mM NaCl, 1% SDS), 500µl phenol:chloroform (5:1), with 500µl 425-600µm acid-washed glass beads. Cells were lysed for 2x2min on a Mini-Beadbeater (BioSpec), with 1min on ice between cycles. Cell suspensions were centrifuged at 13,000rpm for 5 minutes

at 4°C and the supernatant extracted once with an equal volume phenol:chloroform (5:1), once with an equal volume phenol:chloroform:isoamyl alcohol (25:24:1), then once with pure chloroform. High molecular weight (HMW) RNA was precipitated by adding 100µl 40% PEG_{8000} and 50µl 5M NaCl with incubation on ice for 30min. The precipitation was centrifuged at 13,000rpm for 20min at 4°C and the pellet (HMW RNA) was discarded. 3 volumes of 100% EtOH was added to the supernatant and this was incubated at -20°C overnight to precipitate small RNA. RNA was pelleted via centrifugation at 13,000rpm for 30 minutes at 4°C, and this pellet was washed in 95% EtOH, before it was resuspended in 35µl DEPC treated dH_2O. RNA concentration was quantified using a NanoDrop 2000 Spectrophotometer (Thermo Scientific).

2.4.4 – Reverse Transcription

1µg of MasterPure extracted total RNA was mixed with 1µl 10x Turbo DNase Buffer and 1µl of Turbo DNase in a final volume of 10µl. Samples were incubated at 37°C for 1 hour. 2µl dNTPs (10mM), 2µl random hexamers (100ng/µl) and 12µl dH_2O were added and samples incubated at 65°c for 5 minutes, followed by 5 minutes on ice. 8µl 5x SuperScript III Buffer, 2µl DTT (0.1M) and 2µl dH_2O were added, mixed gently, and split into two 19µl aliquots. To one aliquot 1µl SuperScript III RT enzyme was added and both samples were incubated at 25°C for 5min, 50°C for 60min and 70°C for 15min. For qPCR, reverse transcription reactions and -RT controls were diluted 1:4 in dH_2O.

2.4.5 – qPCR

qPCR was performed in 96-well plates on a LightCycler® 96 system (Roche) using LightCycler® 480 SYBR Green I Mastermix. Reactions were set up as follows: 5 µl LightCycler Mastermix, 0.5µl Forward primer (10µM), 0.5µl Reverse primer (10µM), 1µl dH_2O and 3µl template DNA. The following cycling conditions were used:

95°C - 2 minutes
95°C - 20 seconds ⎫
55°C - 20 seconds ⎬ x45 cycles
72°C - 20 seconds ⎭

Data was analysed using the LightCycler® 96 software (Roche). Data were generated from 3 biological replicates, and error bars represent standard deviation. The p-value was calculated using the Student's t-test, and asterisks were used to denote statistical significances as follows: * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$. (See Table 2.3 for full list of qPCR primers).

2.4.6 – Sanger Sequencing

Sequencing was performed using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and reactions were set up as follows: 2µl BigDye™ Terminator v3.1 Sequencing Buffer, 3.68µl dH_2O, 0.32µl primer (10µM), 2µl BigDye™ Terminator v3.1 Ready Reaction Mix, 2µl DNA (300-400ng). Reactions were run using the following cycling conditions:

95°C - 5 minutes
95°C - 30 seconds ⎫
50°C - 20 seconds ⎬ x25 cycles
60°C - 4 minutes ⎭
60°C - 1 minute

Samples were sent to Edinburgh Genomics for sequence analysis using a 3730*xl* DNA Analyzer (Thermo Fisher Scientific).

2.4.7 – Southern Blot

10-20µg of purified genomic DNA was digested with HindIII overnight at 37°C in a final volume of 25µl. Samples were mixed with 5µl Bromophenol Blue loading dye [30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol] and run on a 0.8% Agarose/TBE gel at 50V for 6 hours. The gel was depurinated for 30 minutes in depurinating solution (1% HCl) until the blue dye-front turned yellow. The gel was washed twice in dH_2O then soaked in denaturing solution (0.5M NaOH, 1.5M NaCl) for 30 minutes, until the dye-front turned from yellow back to blue. The separated DNA fragments were blotted onto an Amersham Hybond N+ membrane (GE Healthcare) via capillary transfer. A piece of 3MM Whatmann paper larger that the gel was saturated with denaturing solution, and the stack constructed on top as follows: 3x3MM saturated in denaturing solution, gel (face down), Amersham Hybond N+ soaked in denaturing solution, 3x3MM saturated in denaturing solution, a stack of paper towels, weight (such as a book). The edge of the gel was sealed with parafilm to avoid short-circuiting and the stack left to transfer overnight. After unstacking, the membrane was crosslinked at 2x1200J in a CX-2000 Crosslinker (UVP) and stored at room temperature until required. Prior to probing, the membrane was pre-hybridised at 65°C in Church buffer (500mM sodium phosphate pH 7.2, 7% SDS, 1mM EDTA, 1% BSA). Probes were constructed from 400-500bp PCR products complementary to the 5' end of the NeoR cassette, labelled by random priming with [α-^{32}P] dCTP using High Prime solution (Roche) as follows: 13µl boiled PCR product (25ng), 3µl [α-^{32}P] dCTP (30µCi), 4µl High Prime solution, with incubation at 37°C for 30 minutes. Reactions were stopped by adding 80µl 25mM EDTA, the probe was

boiled for 5 minutes then added to 20ml of fresh Church buffer. This hybridisation solution was added to the pre-hybridised membranes and incubated at 65°C overnight. Membranes were washed 2x with room temperature wash buffer (40mM sodium phosphate pH 7.2, 1% SDS, 1mM EDTA) and 2x with 65°C wash buffer, and then exposed to a phosphor screen for > 6 hours and the radioactive signal visualised on a Typhoon Phosphorimager (GE Healthcare).

2.4.8 – Northern Blot

10-100µg of extracted small RNA was mixed with an equal volume of 2xFDE sample buffer (1mg/ml xylene cyanol, 1mg/ml bromophenol blue, 5mM EDTA, 100% deionised formamide), denatured at 95°C for 3 minutes then separated on a 12% Urea/TBE polycrylamide mini-gel made from SequaGel UreaGel components (National Diagnostics) as follows: 4.8ml Concentrate, 4.2ml Diluent, 1ml Buffer, 80µl 10% APS and 4µl TEMED. Electrophoresis was performed with the Mini-PROTEAN Tetra Cell (Bio-Rad) in 0.5x TBE at 200V, after a pre-run at 200V for 10 minutes. The separated RNA species were blotted onto an Amersham Hybond N+ membrane using the semi-dry transfer apparatus (Hoeffer). The stack was assembled as follows: 3x3MM saturated in 0.5xTBE, Amersham Hybond N+ soaked in 0.5xTBE, gel (face down), 3x3MM saturated in 0.5xTBE, and transferred at 250mA for 35 minutes. After unstacking, the membrane was crosslinked by placing face up on a piece of Whatmann 3MM saturated with crosslinking solution (0.16M EDC, 0.13M methylimidazole, pH 8.0), sealed in foil and incubated at 50-60°C for 2 hours. Membranes were stored at -20°C until required.

Prior to probing, the membrane was pre-hybridised at 42°C in Church buffer (500mM sodium phosphate pH 7.2, 7% SDS, 1mM EDTA). Probes were constructed from oligonucleotides complementary to abundant *S. japonicus* retrotransposon-derived siRNAs or snoRNA58, end-labelled with [γ-^{32}P] dATP using T4 PNK (NEB) as follows:

S. japonicus retrotransposon siRNA probes:

4pmol oligonucleotide, 1µl 10x PNK Buffer, 1µl T4 PNK, 5µl [γ-^{32}P] dATP, 2µl dH_2O

S. japonicus snoRNA58 probe:

50pmol oligonucleotide, 1µl 10x PNK Buffer, 1µl T4 PNK, 1µl [γ-^{32}P] dATP, 6µl dH_2O

Reactions were incubated at 37°C for 1 hour, 40µl of dH_2O was added and unincorporated nucleotides were removed by passing the probe over a pre-prepared G50 ProbeQuant spin column (Amersham). The pre-hybridisation buffer was poured off and 25ml fresh Church buffer was added, along with the both purified probes. Membranes were probed at 42°C overnight. Membranes were washed 2x with 42°C 2xSSC, 0.2% SDS, and then exposed to a phosphor screen for >6 hours and the radioactive signal visualised on a Typhoon Phosphoimager (GE Healthcare).

2.4.9 – Retrotransposon integration sequencing

5µg of genomic DNA purified from a strain harbouring a single copy insertion of Tj1-NeoR was digested with 50U of HindIII overnight at 37°C. Each digest was then diluted to 1ng/µl in dH_2O, and 15 ligation reactions were set up in parallel as follows: 17µl digest (1ng/µl), 2µl 10x T4 DNA Ligase Buffer, 1µl T4 DNA Ligase (20U). Ligations were incubated at 16°C overnight. The ligation

reactions were pooled and purified using a QIAquick PCR Purification Kit (Qiagen) per manufacturer's instructions. Inverse PCR reactions were set up as follows: 5µl 10x *Pfx* Amplification Buffer, 1µl $MgSO_4$ (50mM), 1.5µl dNTPs (10mM), 2.5µl Tj1_NeoR_iPCR_F Primer (10µM), 2.5µl Tj1_NeoR_iPCR_R Primer (10µM), 0.5µl Platinum™ *Pfx* DNA Polymerase, 10µl ligation reaction, 27µl dH_2O. Reactions were run with the following cycling conditions:

95°C	-	2 minutes	
95°C	-	1 minute	⎫
50°C	-	45 seconds	⎬ x35 cycles
68°C	-	4 minutes	⎭
68°C	-	10 minutes	

The whole PCR reactions was analysed by agarose gel electrophoresis and bands of the correct size were extracted using a sharp scalpel. PCR products were purified from gel slices using a QIAquick Gel Extraction Kit (Qiagen) per manufacturer's instructions. Purified DNA was then sent for Sanger sequencing with primers Tj1_NeoR_iPCR_F and Tj1_NeoR_iPCR_R.

2.4.10 – RNA-Seq

Purified total RNA was sent to Edinburgh Genomics for library preparation and sequencing. Libraries were prepared using a TruSeq Stranded mRNA Library Prep Kit (Illumina), and sequenced on a HiSeq System (Illumina). Bioinformatic analysis was performed using tools available on the Galaxy web server (Afgan et al., 2016). Reads were trimmed of adapters and primers using Trim Galore! (Krueger, 2016) and aligned to the *S. japonicus* SJ5 genebuild using TopHat (Trapnell et al., 2009). Read counts for each gene were calculated using FeatureCounts (Liao et al., 2014) and differential expression calculated using DESeq2 (Love et al., 2014). BAM files were converted to

bigWig format using bamCoverage (Ramírez et al., 2014) and normalised to 1x sequencing depth (also known as Reads Per Genomic Content (RPGC)) using an effective genome size of 11,135,996bp for *S. japonicus*. bigWig files were visualised using the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013).

2.4.11 – siRNA-Seq

Purified total RNA was sent to Edinburgh Genomics for library preparation and sequencing. Libraries were prepared using a TruSeq Small RNA Library Preparation Kit (Illumina), and sequenced on a HiSeq System (Illumina). Bioinformatic analysis was performed using tools available on the Galaxy web server (Afgan et al., 2016). Reads were trimmed of adapters and primers using Trim Galore! (Krueger, 2016) and reads outside the rage 14-35nt were discarded using the Filter FASTQ tool (Blankenberg et al., 2010). Reads were then aligned to the *S. japonicus* SJ5 genebuild using Bowtie2 (Langmead and Salzberg, 2012). Read counts for each gene were calculated using FeatureCounts (Liao et al., 2014) and differential expression calculated using DESeq2 (Love et al., 2014). BAM files were converted to bigWig format using bamCoverage (Ramírez et al., 2014) and normalised to 1x sequencing depth (also known as Reads Per Genomic Content (RPGC)) using an effective genome size of 11,135,996bp for *S. japonicus*. bigWig files were visualised using the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013). Read length distribution and 5' nucleotide bias were calculated using FastQC (Andrews, 2010).

2.4.12 – Genome Re-sequencing

Purified genomic DNA was sent to Edinburgh Genomics for library preparation, sequencing and bioinformatics analysis. Libraries were prepared using the TruSeq Nano gel free (350bp insert) kit (Illumina), and sequenced on a MiSeq System (Illumina). Reads were trimmed of adapters and primers using cutadapt (Martin, 2011) and aligned to the *S. japonicus* SJ5 genebuild using BWA-MEM (Li and Durbin, 2010). Variant calling was performed using GATK pipeline, SNPs and indels were called using HaplotypeCaller (Van der Auwera et al., 2013; DePristo et al., 2011; McKenna et al., 2010). Unique variants were found using the GATK SelectVariant tool and common variants were found using vcf-isec from VCFtools (Danecek et al., 2011).

2.5 – Fission Yeast Protein Methods

2.5.1 – Chromatin Immunoprecipitation (ChIP)

50ml of *S. japonicus* cells at OD_{595} of 0.8-1.0 were fixed with 1% paraformaldehyde (PFA) for 15 minutes and room temperature with shaking, then washed twice in 50ml ice-cold PBS. Cells were split equally into two screw cap tubes and pelleted to remove any residual PBS. Pellets were resuspended in 350µl lysis buffer (50mM HEPES-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) containing 1mM PMSF and 1x Yeast protease inhibitors (Roche), with 500µl 425-600µm acid-washed glass beads. Cells were lysed for 2x2min on a Mini-Beadbeater (BioSpec), with 1 minute on ice between cycles. To collect the lysate, the bottom of the tube was pierced with a 25G needle, placed into a clean Eppendorf and centrifuged at 1,000rpm at 4°C for 1 minute. Chromatin was sonicated using a Bioruptor

Twin (Diagenode) for 20 cycles of 30 seconds ON/30 seconds OFF using the 'high' power setting. Sheared chromatin was clarified by centrifugation at 13,000rpm at 4°C for 15 minutes and the supernatant was pre-cleared by adding 25µl Protein-G Agaose (Roche) prepared as a 1:1 slurry of beads:lysis buffer, with incubation at 4°C for 1 hour with rotation. 10µl of pre-cleared chromatin was frozen at -20°C as an 'input' control, and the rest of the pre-cleared lysate was incubated with the appropriate antibody (Table 2.4) and 25µl of Protein-G Agarose:lysis buffer slurry at 4°C overnight with rotation. Beads were washed once in lysis buffer briefly, once in lysis buffer high salt (50mM Hepes-KOH pH7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) for 10 minutes at 4°C, once in wash buffer (10mM Tris-HCl pH8.0, 1mM EDTA, 0.25M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) for 10 minutes at 4°C, and once in TE buffer briefly. To purify immunoprecipitated DNA, 100µl of 10% (w/v) Chelex-100 resin (BioRad) was added to the washed beads and the 10µl 'input' sample, and tubes were incubated at 100°C for 12 minutes. 2.5µl of 10mg/ml Proteinase K (Roche) was added and the samples incubated at 55°C for 30 minutes with shaking at 1000rpm on a Thermomixer (Eppendorf). The Proteinase-K was inactivated via incubation at 100°C for 10 minutes, tubes were cooled and spun briefly, and 50µl of supernatant was transferred to a fresh tube. For qPCR, 'input' samples were diluted 1:80 in dH_2O, whilst IP samples were diluted 1:8 in dH_2O.

2.5.2 – Protein extraction

50ml of *S. japonicus* cells at OD_{595} of 0.8-1.0 were washed twice in 50ml ice-cold dH_2O. Pellets were resuspended in 500µl lysis buffer (50mM HEPES-

NaOH pH7.6, 75mM KCl, 1mM $MgCl_2$, 1mM EGTA, 0.1% Triton X-100) containing 1mM PMSF and 1x Yeast protease inhibitors (Sigma P8215), with 500µl 425-600µm acid-washed glass beads. Cells were lysed for 2x2min on a Mini-Beadbeater (BioSpec), with 1 minute on ice between cycles. To collect the lysate, the bottom of the tube was pierced with a 25G needle, placed into a clean Eppendorf and centrifuged at 1,000rpm at 4°C for 1 minute. DTT was added to the lysate to a final concentration of 500µM, and the supernatant was clarified by two rounds of centrifugation at 13,000rpm for 15 minutes at 4°C.

2.5.3 – Immunoprecipitation

Crude lysates were pre-cleared by adding 25µl Protein-G Agaose (Roche) prepared as a 1:1 slurry of beads:Lysis buffer, with incubation at 4°C for 1 hour with rotation. 25µl of pre-cleared lysate was frozen at -20°C as an 'input' control, and the rest was incubated with the appropriate antibody (Table 2.4) and 25µl of Protein-G Agarose:lysis buffer slurry at 4°C for 3 hours with rotation. Beads were washed three times in Lysis buffer and twice in PBS. 20µl of SDS sample buffer (2% SDS, 50mM Tris-HCl pH6.8, 2mM EDTA, 10% glycerol, 0.03% Bromophenol Blue, 2% β-Mercaptoethanol) was added to the washed beads and the 25µl 'input' sample, and samples were boiled at 100°C for 5 minutes. Samples were then used immediately for SDS-PAGE analysis, or were frozen at -20°C until required.

2.5.4 – Western Blot

Boiled protein samples in loading buffer were separated by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis), on a 1.5mm thick 10% polyacrylamide gel cast in the Mini-PROTEAN Tetra Cell apparatus as follows:-

10% resolving gel: 2.5ml 1.5M Tris-HCl pH8.8, 3.3ml 30% Acrylamide/Bis-acrylamide (Sigma), 100µl 10% SDS, 100µl 10% APS, 10µl TEMED, 3.99ml dH_2O.

Stacking gel: 250µl 1M Tris-HCl pH6.8, 340µl 30% Acrylamide/Bis-acrylamide (Sigma), 20µl 10% SDS, 1.39ml dH_2O.

Gels were run at 200V in 1x Running Buffer (for 1L of 5x: 30g Tris Base, 144g Glycine, 5g SDS) for 45 minutes, PageRuler Prestained Protein Ladder (Thermo Scientific) was used as a size marker. Proteins were blotted onto 0.45µm pore Protran nitrocellulose membrane (GE Healthcare) using the semi-dry transfer apparatus (Hoeffer). The stack was assembled as follows: 3x3MM saturated in Transfer Buffer (20ml 5x Running Buffer, 60ml dH_2O, 20ml MeOH), 0.45µm pore Protran nitrocellulose membrane (GE Healthcare) soaked in Transfer Buffer, gel (face down), 3x3MM saturated in Transfer Buffer, and transferred at 65mA per gel for 90 minutes. The membrane was rinsed in dH_2O and the transfer efficiency was assessed via incubation with Ponceau S for 5 minutes. To avoid-nonspecific background binding of the antibody to the membrane, the membrane itself was incubated in Blocking Buffer (5% Marvel milk powder, 0.2% Tween-20 in 1x PBS) for 1 hour at room temperature. The primary antibody (Table 2.4) was diluted in fresh blocking buffer and incubated with the membranes overnight at 4°C. The membrane was washed twice with PBS-Tween (0.2% Tween-20 in 1x PBS), and the appropriate IRDye® conjugated secondary antibody (LI-COR Biosciences) (Table 2.4) diluted in Blocking Buffer was added to the membrane. Secondary antibody incubations were carried out at room temperature for 3 hours in the

dark; membranes were then washed twice in PBS-Tween. These membranes were scanned using an Odyssey® CLx Imaging System (LI-COR Biosciences).

2.6 – Fission Yeast Genetic Assays

2.6.1 – Spot Tests

A small amount of cells were resuspended in 100µl dH_2O in one well of a 96-well plate. This suspension was mixed well, and 10-fold serially diluted into 90µl of water in the adjacent well using a multichannel pipette. This was repeated until a total of five times, giving a range of dilutions down to 10^{-6} fold. A 48-well pinner (or frogger) was sterilised by flaming in EtOH and used to transfer a small amount of culture from each well onto the desired selective plate(s).

2.6.2 – TSA Assay

S. japonicus cells were cultured to an OD_{595} of ~0.8.in YES. Cultures were split into two and diluted to an OD_{595} of 0.1 in 15ml of YES, then 30µl of TSA (final concentration 35µg/ml) or 30µl of DMSO was added to each. Cells were cultured overnight until OD_{595} reached ~1.0. 500µl of culture was taken from each, and the cells were pelleted, washed, and resuspended in 100µl dH_2O. This resuspension was used as the top dilution in a spotting assay, on both non-selective YES and YES + 15µg/ml TBZ.

2.6.3 – Retrotransposition Assay

In order to evaluate the integration pattern of the *S. japonicus* Tj1 element, the plasmid borne copy of Tj1 (pHL2861) was transformed into *S. pombe* strains using lithium acetate and transformants selected on PMG-Ura plates containing 5µg/ml Thiamine, to select for cells that contained the plasmid

whilst simultaneously repressing transcription of Tj1. These Ura$^+$ colonies were grown for two days and then replica plated onto PMG-Ura plates lacking thiamine, to maintain the plasmid whilst inducing transcription of Tj1 from the *nmt1* promoter. After four days growth, these colonies were then replica plated onto PMG+Ura+5-FOA plates containing 5µg/ml Thiamine. These plates repressed transcription of Tj1, whilst simultaneously driving, and then selecting for, loss of the *URA3* plasmid. After two days growth, these colonies were finally replica plated to YES+5-FOA+G418 plates; these plates selected against retention of the plasmid whilst selecting for Tj1-*neoR*, only those cells with copies of Tj1 integrated into the genome should grow on these plates. From this, the transposition rate of Tj1 was calculated by dividing the number of colonies that grew on YES+5-FOA+G418 (those that carried integrated Tj1) by the number of colonies that grew on PMG+Ura+5-FOA (cells that had lost the Tj1-neoR plasmid), this rate was then expressed as the % of cells with an integrated copy of Tj1-neoR.

Table 2.1 – Fission Yeast Strains

Strain Number	Strain Genotype	*Species*
2742	h− mat-P2028 ura4-D3 ade6(sj)-domE	*S. japonicus*
2743	h+ mat-2017 ura4-D3 ade6(sj)-domE	*S. japonicus*
2730	h- mat-P2028 dcr1::NatR ura4-D3 ade6(sj)-domE	*S. japonicus*
3054	h− mat-P2028 dcr15'::NatR ura4-D3 ade6(sj)-domE	*S. japonicus*
2666	h+ mat-2017 rik1-FLAG-NatMX6 ura4-D3 ade6(sj)-domE	*S. japonicus*
2667	h+ mat-2017 rik1-FLAG-NatMX6 ura4-D3 ade6(sj)-domE	*S. japonicus*
2668	h+ mat-2017 chp1-GFP-ura4+ ura4-D3 ade6(sj)-domE	*S. japonicus*
2669	h+ mat-2017 chp1-GFP-ura4+ ura4-D3 ade6(sj)-domE	*S. japonicus*
2670	h+ mat-2017 stc1-GFP-ura4+ ura4-D3 ade6(sj)-domE	*S. japonicus*
2671	h+ mat-2017 stc1-GFP-ura4+ ura4-D3 ade6(sj)-domE	*S. japonicus*
3254	h− mat-P2028 sjPku70::spUra4 ura4-D3 ade6(sj)-domE	*S. japonicus*
3256	h+ mat-2017 sjPku70::spUra4 ura4-D3 ade6(sj)-domE	*S. japonicus*
3258	h− mat-P2028 sjPku80::spUra4 ura4-D3 ade6(sj)-domE	*S. japonicus*
3260	h+ mat-2017 sjPku80::spUra4 ura4-D3 ade6(sj)-domE	*S. japonicus*
9	h+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18	*S. japonicus*
10	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18	*S. pombe*
129	h- clr4:NATR ade6-210 leu1-32 ura4-D18 arg3-D his3D	*S. pombe*
131	h- dcr1:NATR ade6-210 leu1-32 ura4-D18 arg3-D his3D	*S. pombe*
2060	h- abp1::leu2 ura4-D18 leu1-32 ade6-M216	*S. pombe*
2062	h- cbh1::leu2 ura4-D18 leu1-32 ade6-M216	*S. pombe*
2064	h- cbh2::leu2 ura4-D18 leu1-32 ade6-M216	*S. pombe*
tEC023	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrII(464524):Tj1-NeoR	*S. pombe*
tEC105	h+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrIII(1068021):Tj1-NeoR	*S. pombe*
tEC106	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrIII(1068021):Tj1-NeoR	*S. pombe*
tEC109	h+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrII(744846):Tj1-NeoR	*S. pombe*
tEC111	h+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrIII(331310):Tj1-NeoR	*S. pombe*
tEC112	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrIII(331310):Tj1-NeoR	*S. pombe*
tEC113	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrII(2069592):Tj1-NeoR	*S. pombe*
tEC114	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrII(2069592):Tj1-NeoR	*S. pombe*
tEC118	h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18 ChrI(5232997):Tj1-NeoR	*S. pombe*

Table 2.2 – DNA oligonucleotides

Primer Name	Sequence (5'-3')
Molecular cloning	
MX6_Fwd	GACATGGAGGCCCAGAATACC
NSR-2	AACTCCGTCGCGAGCCCCATCAAC
NSL-2	AAGGTGTTCCCCGACGACGAATCG
MX6_Rev	CAGTATAGCGACCAGCATTCAC
spUra4_F	TTAGCTACAAATCCCACTGG
spUra4_R	AGCTTGTGATATTGACGAAAC
spUra4-5Fusion_R	TTATGTAGTCGCTTTGAAGG
spUra4-3Fusion_F	AATCCGAAATCTTAGAATTGG
jDcr1_Disruption_Deletion_F	TGTGTATTCGCCCATGTACG
jDcr1_disdel_MX6Fusion5_R	GTATTCTGGGCCTCCATGTCAGGCGTAGTTGTGATAGTAGG
jDcr1_dis_MX6Fusion3_F	GAATGCTGGTCGCTATACTGTTGCTTGTTATGAACACAGG
jDcr1_Disruption_R	AACCTTATTAGTGACATCAATTCC
jDcr1_del_MX6Fusion3_F	GAATGCTGGTCGCTATACTGAAAGACGAAGAGACTAAAACG
jDcr1_Deletion_R	AGAGATTCCAGTCCAGAAGC
jClr4_Deletion_F	AGGGTTGCATGTCGCAGTGC
jClr4_del_MX6Fusion5_R	GTATTCTGGGCCTCCATGTCGCTGTGTTTTGGCAGTGTCAACC
jClr4_del_MX6Fusion3_F	GAATGCTGGTCGCTATACTGTCTCTCGCCGCTTATGAAGTC
jClr4_Deletion_R	ATCGGAAACGCTTCATCATGG
jAgo1_Deletion_F	TGGAGAATTTCACAACACATGC
jAgo1_del_MX6Fusion5_R	GTATTCTGGGCCTCCATGTCACTCTGCTGTGGGGATGTCTG
jAgo1_del_MX6Fusion3_F	GAATGCTGGTCGCTATACTGAAGATGTGGTTCATGTGATGG
jAgo1 Deletion R	TCTGATGTCGGGACTGTTGC
jArb1_del-5_F	TTGGCACCGTTAATGTTTCC
jArb1_del_Ura4Fusion5_R	CCAGTGGGATTTGTAGCTAAGCACCACAACCTATTGAACC
jArb1_del_Ura4Fusion3_F	GTTTCGTCAATATCACAAGCTGCGTTTGGTCTAACTAACAC
jArb1_del-3_R	CTTGCGTTCAAAGTCCTACG
jArb2_del-5_F	TCTTACAGTGATTTGAAAGC
jArb2_del_Ura4Fusion5_R	CCAGTGGGATTTGTAGCTAAATGGATGATACACAGACTGG
jArb2_del_Ura4Fusion3_F	GTTTCGTCAATATCACAAGCTACTCTGAACTCGTTACAAGC
jArb2_del-3_R	TTGTCCGACTTGTCTGTAGC
jRdp1_del-5_F	TCCAAGAGTAGTCAGTAAGG
jRdp1_del_Ura4Fusion5_R	CCAGTGGGATTTGTAGCTAACAAAACCCATATACAATTGC
jRdp1_del_Ura4Fusion3_F	GTTTCGTCAATATCACAAGCTTGCTACTCAATAAGGATAAGG
jRdp1_del-3_R	TTGTAATGGTTGTCAATTCC
jTri1_del-5_F	TGATGTCTCGATGAATAACG
jTri1_del_MX6Fusion5_R	GTATTCTGGGCCTCCATGTCGAGCACAGGGAGAATATAAG
jTri1_del_MX6Fusion3_F	GAATGCTGGTCGCTATACTGCCGACATCTGGAACTAAAAC
jTri1_del-3_R	TGTCACTTCTGGTTACAAGG

jPku70-5_F	AAATGCATACGAAGGAAGC
jPku70-5_Ura4_R	CCAGTGGGATTTGTAGCTAAAAAACATGACAGTCTGTGAG
jPku70-3_Ura4_F	GTTTCGTCAATATCACAAGCTGGTCTCGTTCTTTTCTTTTAG
jPku70-3_R	GGTGCGATACTAATATTGTCG
jPku80-5_F	GTCTAGTTACGATATATGCATGC
jPku80-5_Ura4_R	CCAGTGGGATTTGTAGCTAACACCAATTGTAAATTCGTACG
jPku80-3_Ura4_F	GTTTCGTCAATATCACAAGCTCAGAGAAATCTATCGTTTTGC
jPku80-3_R	CACCGGTATAATAGAATACAAGG
NdeI_jRik13ORF_BamHI_F	ATATCATATGCGTCAGTGGAAGACTATTGG
NdeI_jRik13ORF_BamHI_R	TTAAGGATCCGTGCTACACTGGATAAATCAAC
AatII_jRik1DS-3UTR_NdeI_F	TAATGACGTCCTAAGAATGATGGCTTCACTTCC
AatII_jRik1DS-3UTR_NdeI_R	AATTCATATGTCAATGCCAAACCATCTTGTTCG
XhoI_jChp13ORF_BamHI_F	ATATCTCGAGAGCCCCTTTTGGATCGTTTTGC
XhoI_jChp13ORF_BamHI_R	TTAAGGATCCGCCAAGACCGACAGTAGACAT
ApaI_jChp1DS-3UTR_XhoI_F	TTAAGGGCCCTTGGACATATAGTGTAGTGCCTG
ApaI_jChp1DS-3UTR_XhoI_R	TAATCTCGAGCTGTGGGACCGTTTGTTTGC
XhoI_jStc13ORF_BamHI_F	ATATCTCGAGTTTCTTCTGTTCGTCTTGTTCC
XhoI_jStc13ORF_BamHI_R	TTAAGGATCCAGTGTTCTTGTTCGCACTATAT
ApaI_jStc1DS-3UTR_XhoI_F	TAATGGGCCCTCTTGATTGACCGGCTTATGG
ApaI_jStc1DS-3UTR_XhoI_R	AATTCTCGAGTTTGATTCGTGGGAGCAACG
qPCR	
q_Tj1_LTR_F	AGCGACCAGTACGTTGATCT
q_Tj1_LTR_R	GGCGGAGGCCGATATAAAGA
q_Tj1_ORF_F	TATTATTGTGCCGAGCGAAGG
q_Tj1_ORF_R	GATGACCACCGTTCACTGAG
q_Tj2_LTR_F	ACAGAGGAAACACATCCCCG
q_Tj2_LTR_R	CCGAACAGGGCGATTTACCT
q_Tj2_ORF_F	GGAGACAATCCCGTTGGCTA
q_Tj2_ORF_R	ACTGGATGTCCGGGTCAAAC
q_Tj3_LTR_F	ACACCGAGAGCTCCAGCGTT
q_Tj3_LTR_R	AGGCTTTCGTAGCGTAGTGC
q_Tj3_ORF_F	AACCGTCCACATCGCCGTAT
q_Tj3_ORF_R	CGGTCCAGCAACTCGTCAAT
q_Tj4_LTR_F	GGGAAACCCTCGGAATCCTC
q_Tj4_LTR_R	ATCCCCTTTTGTGGCTTGCT
q_Tj4_ORF_F	GGTACCACCAGAGGGTGAAC
q_Tj4_ORF_R	TCCTCCGTTTAGGGGATCGT
q_Tj5_LTR_F	AGCAACACCAAGAACGCTGC
q_Tj5_LTR_R	TAACGCGCAGCTTAACGTGG
q_Tj5_ORF_F	GCAGTGTTCAAGCGTCTTCGG
q_Tj5_ORF_R	GGAAGTTGACGCGCTTCCTG
q_Tj6_LTR_F	TCCAACTGCTGAGCAACCTT
q_Tj6_LTR_R	ACGGAGGTTTCCGATTTAGGT

q_Tj6_ORF_F	GAGCAAAGGAGTGGCAGAGT
q_Tj6_ORF_R	TTCGTCTGCGGTCTGTAACC
q_Tj7_LTR_F	TCACTTACAGAACCCACAGC
q_Tj7_LTR_R	TTCTGTTGCTAATTTCCTATGG
q_Tj7_ORF_F	AAGAACCGATACGCGCTGC
q_Tj7_ORF_R	CTTGCTGAAGACGCGTGC
q_Tj8_LTR_F	AGCCTAAATCTGGGCGAGCA
q_Tj8_LTR_R	TGCAGCTCGTTGCAGTTCGT
q_Tj8_ORF_F	AGCCGTCGAATTCGCCCTA
q_Tj8_ORF_R	TTGAGGCCGCGGTAGTCAAT
q_Tj9_LTR_F	CGCTACAAGGCCTAGGGAGT
q_Tj9_LTR_R	GTAAGCGCTGCGTGGAGTTC
q_Tj9_ORF_F	AACGGGCGATGAGGAAACAA
q_Tj9_ORF_R	TTCCGTCGCTGAGGTCTTCC
q_Tj10_LTR_F	TACGATCGCCAACTTCAAAC
q_Tj10_LTR_R	ATGGTATGACTAAGCGTGGC
q_Tj10_ORF_F	CATTCGGACTCACCAATGCG
q_Tj10_ORF_R	TGTTGATCGCGGTCCTTTGA
q_jap_act1_F	GTTTTGCTGGCGACGACG
q_jap_act1_R	ATACCACGCTTGCTTTGAG
q_jap_fba1_F	GAGATGGAGATCGGTATCAC
q_jap_fba1_R	GCTTGACGTTGTTGGTCTTC
q_jap_his3_F	TACAGACCCTAAAATTGCCC
q_jap_his3_R	CAAGCTTGGACGTAGGATTA
q_jap_crm1_F	AGTCTTATCACCGCAACTTT
q_jap_crm1_R	ATTACGGAATGCAGGAACTT
q_act1_F	GGTTTCGCTGGAGATGATG
q_act1_R	ATACCACGCTTGCTTTGAG
Tf2_LTR_F	TGATAGGTAACATTATAACCCAGT
Tf2_LTR_R	ACGCAGTTTGGTATCTGATT
Tf2_ORF_F	GGTAGGCAGTTTATGTGCTC
Tf2_ORF_R	AGAACAGCCTCGTATGGTAA
Southern blot probes	
NeoR_Probe_F	GCATTTCTTTCCAGACTTGTTC
NeoR_Probe_R	ATATTCAACGGGAAACGTCTTG
Northern blot probes	
Tj1A	GTTTGCCGCACTTCAACCCGAAA
Tj1B	CGCGTTCTTCTGTCAACTCCTGCG
Tj1C	TGAGGCCGTTCAGTCATTCTTGC
Tj2A	ACCGTTGTTCTTCGTTATCCTTA
Tj2B	CCGTCCTCGTCTCCGTGCCA
Tj2C	CCGTCCTGTTCGTTCGAAAGAA
Tj3A	TCAGGTGTTCGTCTTCCGTCTCTG

Tj3B	TAACTCGTTCAGCTGTCGGTCCA
Tj3C	TTGAACCGTCCACATCGCCGTA
Tj4A	ATGCCTTCTCCACATTCCGGACA
Tj4B	GCGTTCTGTATTGTTCTTCATA
Tj4C	CTCGTCTTGTTCCGTCCGACATGT
Tj5A	TGGCGCTCCAGTTCTGTTCGTTCG
Tj5B	TTTGCGTTTTCGCTCGCGTCCGTA
Tj5C	CAGCTGTCGTCGTCCTCCGTGTA
Tj6A	CTCCATTGTTCTTTCTTCCGCTT
Tj6B	TCTCCAAAGCCGTCTGCCCGCAA
Tj6C	GATACCACTGCTTCGCGTCCATGT
Tj7A	CGCTTGCTTGCTCCACGTCCGCG
Tj7B	TGATGACAAGAACGACGGCGTCACC
Tj7C	GATCGTTCTCTCGCTCTGCCCATC
Tj8A	CGCCGCGCCCAGCCCTACGTCA
Tj8B	CATGTCAGTCGTTCTTCGTCGTCGA
Tj8C	ACTACCGTCCTTCTTCCTCACGAA
Tj9A	TACCGCTCCACTGTGCCCGAGTC
Tj9B	ATCGCGTCTTCCAATCCTCGTT
Tj9C	ATCGTCCGTTCAGACTGTCCGT
Tj10A	CACCGTTTTGTCCAGATCTTCA
Tj10B	TCGCTCCACTTGCATTTGTCCG
Tj10C	GCTTTGCTCGTCTAGTCCGCAA
jap_snoRNA58	CTGCTAAATCAGAAGTCTAGCATC
iPCR Sequencing	
Tj1_NeoR_iPCR_F	CCGTTGAATATGGCTCATAAC
Tj1_NeoR_iPCR_R	GAACAAGTCTGGAAAGAAATGC

Table 2.3 – Plasmids

Plasmid Name	Source
pFA6a-NatMX6	(Bähler et al., 1998)
pIRT2U	(Hindley et al., 1987)
pFA6a-FLAG-NatMX6	(Noguchi et al., 2008)
pSO729	Gift from S. Oliferenko
pFA6a-jRik1-FLAG-NatMX6	This study
pSO729-jChp1-GFP	This study
pSO729-jStc1-GFP	This study
pHL2861	(Guo et al., 2015)

Table 2.4 – Antibodies

Antibody	Species	Source	Dilution
α-H3K9me2	Mouse monoclonal	(Nakagawachi et al., 2003)	1µl/sample for ChIP
α-H3	Rabbit polyclonal	Abcam (ab1791)	2µl/sample for ChIP
α-H4	Rabbit monoclonal	Merck-Millipore (05-858)	1.5µl/sample for ChIP
α-Cnp1	Sheep polyclonal	Gift from A. Pidoux and R. Allshire	10µl/sample for ChIP
α-Rbp1	Mouse monoclonal	Abcam (8WG16)	5µl/sample for ChIP
α-GFP	Rabbit polyclonal	Thermo Fisher Scientific (A11122)	1.5µl/sample for ChIP
α-FLAG	Mouse monoclonal	Sigma-Aldrich (M2)	1:1000 dilution for WB
IRDye® 800CW	Goat anti-Mouse	Li-Cor (926-32210)	1:20000 dilution for WB
IRDye® 680RD	Goat anti-Rabbit	Li-Cor (926-68171)	1:20000 dilution for WB

Chapter 3 – Investigating the role of RNAi and heterochromatin in *S. japonicus*

3.1 - Introduction

Over the past 60 years, the fission yeast *Schizosaccharomyces pombe* has become one of the key model organisms employed for the study of eukaryotic cell biology, and is now one of two major yeast model systems along with the budding yeast *Saccharomyces cerevisiae* (Hoffman et al., 2015). Over this period, *S. pombe* has made key contributions to our understanding of a number of core biological processes, with seminal discoveries made concerning the nature of the cell cycle (Lee and Nurse, 1987; Nurse, 1975; Nurse et al., 1976), the dynamic nature of mitotic chromosome segregation (Hirano et al., 1986; Samejima et al., 1993), and the interplay between chromatin states and the epigenetic control of gene expression (Allshire et al., 1994, 1995; Ekwall et al., 1996).

Of particular interest within the lab is the contribution made by *S. pombe* to our understanding of heterochromatin formation via RNA interference; fission yeast are a particularly good model for studying this process, as they are single celled eukaryotes with large, regional centromeres that are similar in architecture to many higher organisms, as they are made up of a discrete centromeric region flanked by blocks of pericentromeric heterochromatin. They are also stably haploid, making the phenotypic study of mutant alleles fairly simple, whilst their ability to integrate exogenous linear DNA via homologous recombination facilitates generation of these mutant alleles.

Looking specifically at the RNAi and heterochromatin formation pathway, *S. pombe* is unusual amongst eukaryotes as it possesses a single copy of each of the core components of this pathway; the ribonuclease Dicer (Dcr1), the PIWI family protein Argonaute (Ago1), the RNA-dependent RNA-Polymerase

(Rdp1) and the Lysine 9 methyltransferase (Clr4). This is in contrast to *Drosophila melanogaster* which contains 5 Argonaute proteins, *Arabisopsis thaliana* which encodes 10 Argonaute proteins, and *Caenorhabditis elegans* which contains at least 26 Argonaute genes (Hutvagner and Simard, 2008). Deletion mutants for these core components are viable in *S. pombe,* making this species an ideal model for genetic analysis of this pathway (Martienssen et al., 2005; Wood et al., 2002)

Although *S. pombe* has been crucial in the study of RNA interference as it relates to heterochromatin formation, there are a number of things that make this organism somewhat atypical. Firstly, *S. pombe* does not utilise its RNAi pathway to silence retrotransposons under normal conditions, instead it has developed a separate pathway that relies on domesticated transposon related proteins, that enact transcriptional silencing independently of RNAi (Cam et al., 2008). This is somewhat unusual, as the silencing of mobile genetic elements by RNAi is a highly conserved function of the pathway (Buchon and Vaury, 2006; Castel and Martienssen, 2013). Secondly, retrotransposons in *S. pombe* are not enriched at pericentromeric and telomeric loci, but are instead distributed throughout the chromosome arms. This organisation of transpososable elements is distinct from many higher eukaryotes, such as *D. melanogaster*, *A. thaliana*, *Oryza sativa* (rice), *Zea mays* (maize) *Triticum* spp. (wheat), and mammals, which contain large, complex centromeres and pericentromeres, interspersed with various transposable elements (TEs) (Plohl et al., 2014; Slotkin and Martienssen, 2007; Wong and Choo, 2004).

Recent work by the Broad Institute to sequence the genomes of the other three members of the fission yeast clade; *S. japonicus, S. octosporus* and *S.*

cryophilus, revealed a number of intriguing differences between these yeast and *S. pombe* (Rhind et al., 2011). In terms of centromeric organisation and the RNAi pathway, the most striking difference was observed in the earliest branching member of the clade, *S. japonicus*. This species contains a far larger retrotransposon complement than *S. pombe* and these TEs also appear to cluster at the presumed centromeres and telomeres, in a similar way to the higher eukaryotes mentioned above. Deep sequencing of small RNAs revealed that these elements produce a majority of the siRNAs in *S. japonicus*, indicating that this fission yeast may have retained the conserved ancestral role for RNAi in the silencing of transposable elements. It is for these reasons, coupled with the fact that *S. japonicus* retains the characteristics that have made *S. pombe* a favoured model organism for the study of eukaryotic cell biology (Klar, 2013; Niki, 2014), that we hypothesise that *S. japonicus* may prove to be a valuable new model for the study of RNA interference in the context of transposable element regulation.

In this chapter I present the work I have undertaken to begin genetic analysis of the RNAi pathway in *S. japonicus*, as it pertains to the silencing of mobile genetic elements and the nucleation of heterochromatin.

3.2 - Annotation and Sequence analysis of *S. japonicus* RNAi genes

The first task was to annotate the RNAi genes in *S. japonicus*, as not all key factors had been identified on the Ensembl Fungi resource (Kersey et al., 2016). To do this, I carried out searches using known *S. pombe* RNAi and heterochromatin genes (obtained from PomBase (Wood et al., 2012) using the Gene Otology (Blake et al., 2015) terms GO:0031048 - chromatin silencing by small RNA and GO:0030702 - chromatin silencing at centromere) (Figure 3.1). Examples of unannotated RNAi genes in S. japonicus were *ago1*$^+$, *dcr1*$^+$ and *stc1*$^+$. The *S. japonicus stc1*$^+$ gene had previously been located and used for sequence alignment (He et al., 2013), however the location of the core factors *ago1*$^+$ and *dcr1*$^+$ remained unidentified. Using OrthoDB (Zdobnov et al., 2017) to perform an orthologue search using the *S. pombe ago1*$^+$ (SPCC736.11) and *dcr1*$^+$ (SPCC188.13c) sequences revealed that the orthologue of *S. pombe ago1*$^+$ in *S. japonicus* was SJAG_02621 and the orthologue of *S. pombe dcr1*$^+$ in *S. japonicus* was SJAG_03689.

To validate whether these *S. japonicus* genes contained the conserved characteristic motifs of their *S. pombe* orthologues, the primary sequences were run through the NCBI Conserved Domain Search (Marchler-Bauer et al., 2017). SJAG_02621 contained both a PAZ Domain and a PIWI Domain that are hallmarks of Argonaute-family proteins (Höck and Meister, 2008). Similarly SJAG_03689 contained the N-terminal helicase/ATPase domain and tandem RNaseIII domains, yet lacked a conserved PAZ domain, similar to the Dcr1 of *S. pombe* (MacRae et al., 2006, 2007) (Figure 3.1A). Although gene

S. pombe Gene Name	*S. pombe* Gene ID	*S. japonicus* Gene Name	*S. japonicus* Gene ID	*S. japonicus* Gene Description	% Identity	% Similarity	% Gaps
***ago1*⁺**	SPCC736.11	*N/A*	SJAG_02621	Hypothetical protein	62.4	78.0	1.7
***arb1*⁺**	SPAC140.03	*N/A*	SJAG_02594	Argonaute binding protein 1	40.9	55.3	17.2
***arb2*⁺**	SPAC13G7.07	*N/A*	SJAG_04061	Argonaute binding protein 2	30.3	53.3	5.2
***chp1*⁺**	SPAC18G6.02c	*chp1*	SJAG_03916	Chromodomain protein Chp1	31.5	49.3	21.1
***chp2*⁺**	SPBC16C6.10	*N/A*	SJAG_01579	Chromodomain protein 2	29.3	43.5	27.6
***cid12*⁺**	SPCC663.12	*cid12*	SJAG_02725	Poly(A) polymerase Cid12	47.8	68.4	6.4
***clr4*⁺**	SPBC428.08c	*clr4*	SJAG_04574	Histone H3 methyltransferase Clr4	53.8	66.9	12.3
***dcr1*⁺**	SPCC188.13c	*N/A*	SJAG_03689	Hypothetical protein	35.5	53.6	7.9
***ers1*⁺**	SPCC1393.05	*ers1*	SJAG_04253	RNA-silencing factor Ers1	23.3	42.0	24.8
***hrr1*⁺**	SPCC1739.03	*hrr1*	SJAG_03193	Helicase Required for RNAi-mediated heterochromatin assembly Hrr1	42.9	61.3	12.6
***pcu4*⁺**	SPAC3A11.08	*N/A*	SJAG_01540	Cullin 4	48.2	68.9	0.7
***raf1*⁺**	SPCC613.12c	*raf1*	SJAG_00756	Rik1-associated factor Raf1	42.3	57.9	9.5
***raf2*⁺**	SPCC970.07c	*raf2*	SJAG_02119	Rik1-associated factor Raf2	32.7	50.4	12.0
***rdp1*⁺**	SPAC6F12.09	*rdp1*	SJAG_03356	RNA-directed RNA polymerase Rdp1	50.7	70.5	2.1
***rik1*⁺**	SPCC11E10.08	*rik1*	SJAG_03680	Silencing protein Rik1	28.7	49.2	5.8
***stc1*⁺**	SPBP8B7.28c	*N/A*	SJAG_02409	Hypothetical protein	26.5	41.0	23.1
***swi6*⁺**	SPAC664.01c	*swi6*	SJAG_01601	Chromodomain protein Swi6	36.9	48.5	29.9
***tri1*⁺**	SPBC29A10.09c	*N/A*	SJAG_01836	CAF1 family ribonuclease	30.3	47.8	17.0
***tas3*⁺**	SPBC83.03c	*N/A*	SJAG_03042	RITS complex subunit 3	31.4	45.4	18.7

Table 3.1 – Identity and sequence comparison of RNAi genes in *S. pombe* and *S. japonicus*

Table showing the Gene Name and Gene ID of RNAi genes in *S. pombe* and the corresponding homologues in *S. japonicus.* The sixth column shows the percentage of residues that are identical, the seventh the percentage of residues that are similar and the eighth shows the percentage of gaps between the *S. pombe* and *S. japonicus* proteins. Identities were calculated using Clustal Omega (Sievers et al., 2011)

synteny between *S. pombe* and *S. japonicus* is not well conserved (Rhind et al., 2011), a number of core RNAi factors are expressed from convergent genes, and this local gene order appeared to be conserved. The *ago1*$^+$ gene ran convergent to *mmi1*$^+$ (SJAG_02622), *dcr1*$^+$ to *spn6*$^+$ (SJAG_03688) and *clr4*$^+$ to *meu6*$^+$ (SJAG_04574), as is the case in *S. pombe* (Figure 3.1B).

Comparison of the primary protein sequence of the core RNAi proteins in *S. pombe* and *S. japonicus* revealed that the highest degree of identity occurred within the core catalytic components of the pathway. The core RITS component Ago1 showed the highest degree of conservation, with 62.4% identity and 78.0% similarity, followed by the core ClrC subunit Clr4, which was 53.8% identical with 66.9% similarity. The RNA-direct RNA polymerase Rdp1 also showed a high degree of identity, 50.7%, whilst the lowest level of conservation was found in genes that act to recruit other factors, such as Chp1, Rik1 and Stc1, which were 31.5%, 28.7% and 26.5% identical respectively (Table 3.1).

3.3 - Most core RNAi and heterochromatin factor deletions cannot be recovered in *S. japonicus*

In order to gain insight into the mechanism and function of the RNAi pathway in *S. japonicus*, a number of key RNAi and heterochromatin factors were targeted for genetic deletion via homologous recombination (HR) with a cassette (NatMX6) that conferred resistance to the antibiotic Nourseothricin.

The RNAi and heterochromatin components chosen for deletion were the ribonuclease *dcr1*$^+$ (SJAG_03689), the argonaute protein *ago1*$^+$ (SJAG_02621), the sole lysine 9 methyltransferase *clr4*$^+$ (SJAG_04574), the RNA dependent RNA polymerase *rdp1*$^+$ (SJAG_03356) and the ARC complex

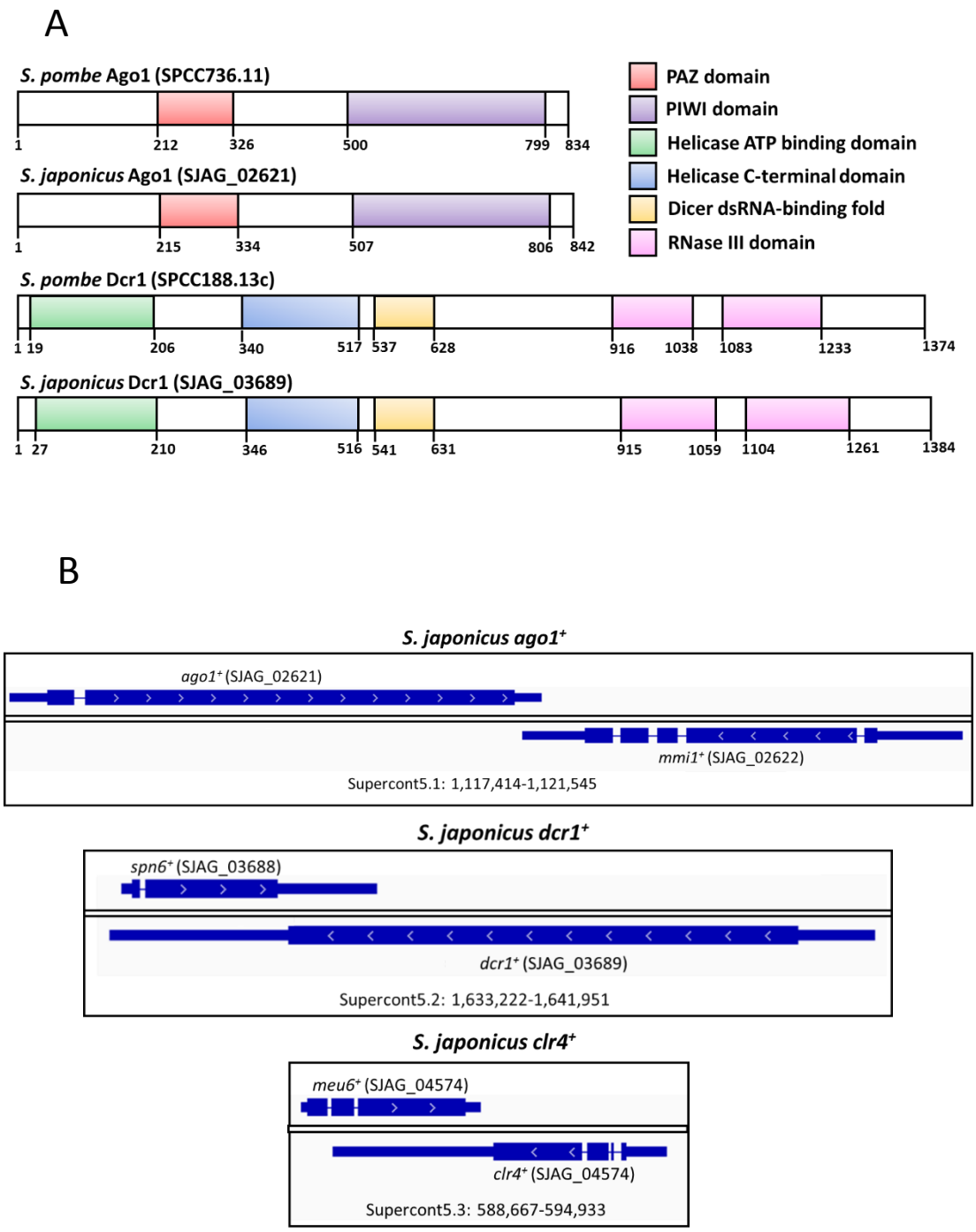


Figure 3.1 – Conservation of key domains and local gene order in core *S. japonicus* RNAi genes

(A) Schematic representation of the Dcr1 and Ago1 genes from *S. pombe* and *S. japonicus,* showing conserved domains. Position of conserved domains is marked below, numbers indicate amino acid locations.
(B) Schematic representation of the genomic loci of *S. japonicus ago1*$^{+}$and *mmi1*$^{+}$, *spn6*$^{+}$ and *dcr1*$^{+}$ and *meu6*$^{+}$ and *clr4*$^{+}$

subunits *arb1*$^{+}$ (SJAG_02594) and *arb2*$^{+}$ (SJAG_04061). All of these genes have been shown to play a key role in the formation of transcriptionally silent heterochromatic domains at repeat elements in *S. pombe* (Buker et al., 2007; Ekwall et al., 1996; Motamedi et al., 2004; Nakayama et al., 2001; Volpe, 2002).

S. japonicus is much less easily transformed than the lab strain of *S. pombe*, as standard lithium acetate based protocols are not applicable, due to the decreased viability of *S. japonicus* in LiOAc (Aoki et al., 2010). Initial attempts to transform *S. japonicus* using published electroporation protocols (Aoki and Niki, 2017; Aoki et al., 2010; Furuya and Niki, 2009) proved unsuccessful, with no recovery of antibiotic resistant colonies. Extensive optimisation of the protocol revealed a number of critical steps required for successful transformation, most of which related to the specific growth phase of cells required. Firstly, very fresh cells were essential for inoculating liquid cultures for transformation, cells that spent longer than 2-3 days on plates post streaking from a glycerol stock were not able to be transformed. Cells also had to be grown to a precise density, however due to the tendency of *S. japonicus* to flocculate when grown in liquid media (Aoki et al., 2017) counting exact numbers using a haemocytometer was not practical, and instead density was recorded using a spectrophotometer set at a wavelength of 595nm. For *S. japonicus,* cultures at OD_{595} of 0.4-0.5 worked well, and corresponded to cells in early log-phase. Once the transformation procedure was established, and resistant colonies isolated for genotyping, very high proportions of incorrectly integrated 'false-positive' background colonies were obtained for all transformations.

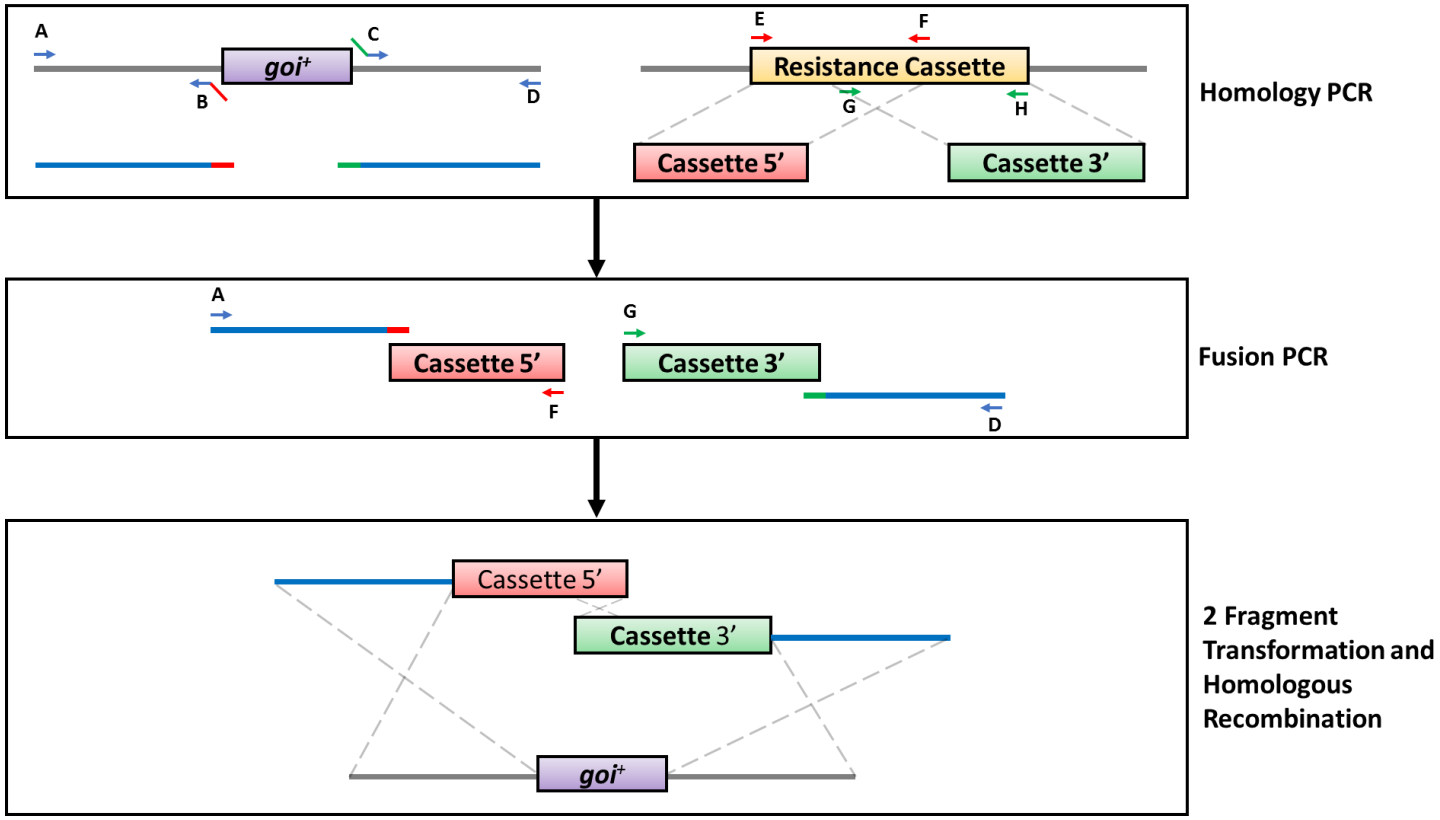


Figure 3.2 – Overview of 'Split-Marker' Transformation Strategy

Illustration of the steps involved in generating 'split-marker' fragments for targeted gene manipulation. The first stage involves PCR amplification of homologies, a 5' homology (using primers A & B) and a 3' homology (using primers C & D) upstream and downstream of the region to be targeted. The desired resistant marker is also amplified as two truncated fragments (using primers E & F and G & H). The second step involves fusion PCR of the 5' genome homology with the 5' resistance cassette fragment (using primers A & F) and the 3' resistance cassette fragment with the 3'genome homology (using primers G & D). These two fragments are then both used directly for transformation of the desired yeast strain.

To increase targeting efficiency and to reduce this problematic background, a 'split-marker' strategy developed for use in *Cryptococcus Neoformans* (Fu et al., 2006; Kim et al., 2009) was adapted for use in *S. japonicus (*Figure 3.2). Briefly, the selectable marker was split into two parts, with 200bp overlap between fragments. These split marker fragments were then fused to 1kb fragments homologous to regions immediately upstream and downstream of the targeted gene open reading frame (ORF). The long homologies, 1kb compared to 80bp traditionally used for yeast transformation, were designed to increase specific targeting efficiency (Krawchuk and Wahls, 1999) whilst the split marker was utilised to reduce background, as three homologous recombination events were required to generate a strain with a correctly integrated marker (Fu et al., 2006; Kim et al., 2009).

Although a large number of Nat^R colonies were isolated for each knockout over multiple transformation experiments using this strategy, none were positive for the specific gene deletion (with the exception of 2 *dcr1*$^+$ clones which will be discussed in Chapter 4) (Table 3.2A).

A possible explanation for the inability to recover deletions in RNAi and heterochromatin factors is that these loci may be inefficiently targeted. To attempt to overcome this, deletions of the non-homologous end joining (NHEJ) genes *pku70*$^+$ (SJAG_05372) or *pku80*$^+$ (SJAG_04372) were constructed. It has been shown in various species (Kooistra et al., 2004; Li et al., 2010; Näätsaari et al., 2012; Nayak et al., 2006; Ninomiya et al., 2004; Pöggeler and Kück, 2006; Villalba et al., 2008), including *S. pombe* (Fennessy et al., 2014), that deletion of these proteins can increase the efficiency of gene

A

S. japonicus Gene Target	*S. japonicus* Gene ID	Resistant Colonies Screened	Positive Colonies	% Positive
ago1⁺	SJAG_02621	119	0	0.00
arb1⁺	SJAG_02594	16	0	0.00
arb2⁺	SJAG_04061	16	0	0.00
clr4⁺	SJAG_04574	54	0	0.00
dcr1⁺	SJAG_03689	122	2	1.64
rdp1⁺	SJAG_03356	16	0	0.00
tri1⁺	SJAG_01836	16	12	75.00

B

Deletion Background	*S. japonicus* Gene Target	*S. japonicus* Gene ID	Resistant Colonies Screened	Positive Colonies	% Positive
pku70Δ	*ago1⁺*	SJAG_02621	6	0	0.00
pku80Δ	*ago1⁺*	SJAG_02621	5	0	0.00
pku70Δ	*clr4⁺*	SJAG_04574	5	0	0.00
pku80Δ	*clr4⁺*	SJAG_04574	6	0	0.00
pku70Δ	*dcr1⁺*	SJAG_03689	6	0	0.00
pku80Δ	*dcr1⁺*	SJAG_03689	16	0	0.00

Table 3.2 – Knockout frequencies of core RNAi genes in wild-type and *pku70/80Δ* backgrounds

(A) Table showing the total number of resistant colonies screened for each deletion candidate, and the percentage of those colonies that had a correct deletion by PCR, in a wild-type background.
(B) Table showing the total number of resistant colonies screened for each deletion candidate, and the percentage of those colonies that had a correct deletion by PCR, in either a *pku70Δ* or *pku80Δ* background

targeting by homologous recombination, thus making the recovery of specifically targeted integrants more likely.

However, even in the *pku70Δ* or *pku80Δ* backgrounds no *RNAiΔ* mutants could be recovered, although a number of resistant colonies were isolated (Figure 3.4B). This suggests that failure to recover these deletion mutants is unlikely to be due to inefficient targeting of these loci.

3.4 - A factor involved solely in heterochromatin establishment can be deleted in *S. japonicus*

In order to establish whether it was only deletions in 'core' RNAi factors that could not be recovered, an RNAi accessory factor was chosen for analysis. *tri1*$^{+}$ (SPBC29A10.09c) is a CAF1 family ribonuclease that has been shown in *S. pombe* to exclusively play a role in the establishment of heterochromatin, but is not required for maintenance of the repressive state (Marasovic et al., 2013). I therefore targeted the *S. japonicus tri1*$^{+}$ homologue SJAG_01836 for deletion.

Interestingly, this factor could be deleted with high efficiency; of the 16 Nat^{R} colonies, 12 had the correctly integrated NatMX6 cassette by PCR (Figure 3.4A). This *tri1Δ* mutant was spotted onto plates containing 15µg/ml Thiabendazole (TBZ). TBZ is a microtubule destabilising drug that is used to assay for defects in centromeric heterochromatin formation; cells that are unable to properly nucleate heterochromatin at the pericentromere display hypersensitivity to this drug. The isolated *tri1Δ* strain did not exhibit sensitivity to TBZ under normal 'maintenance' conditions (Figure 3.3A), which is consistent with the phenotype exhibited by this mutant in *S. pombe*

To assess whether *tri1*$^+$ exhibited a phenotype in a heterochromatin establishment situation, cells were cultured in media containing the Histone Deacetylase (HDAC) inhibitor Trichostatin A (TSA). TSA acts to inhibit histone deacetylation, thus precluding subsequent histone methylation, leading to a loss of H3K9me2 at pericentromeric regions (Nakayama et al., 2001; Rea et al., 2000; Tamaru and Selker, 2001). Cells that are unable to re-establish heterochromatin following recovery from TSA treatment should exhibit TBZ sensitivity, due to the lack of proper pericentromeric heterochromatin formation. Indeed, the *tri1Δ* strain did exhibit a growth defect on TBZ after culture in TSA when compared to wild-type cells, however this strain also showed a comparable reduction in viability when plated onto non-selective media (Figure 3.3B). This indicates that disruption of heterochromatin may impact viability of *S. japonicus,* in the absence of factors that may act to re-establish the mark.

3.5 - RNAi factors can be genetically tagged with no impact on function

Given that core RNAi factors could not be deleted, it was decided to assess whether certain RNAi gene loci were amenable to another form of genetic manipulation – integration of an epitope-tag. The RNAi and heterochromatin factors *dcr1*$^+$, *clr4*$^+$ and *ago1*$^+$ are known to only tolerate an N-terminal tag in *S. pombe*. As yet a system to engineer N-terminally tagged genes under endogenous promoter control has not been fully established in *S. japonicus*, thus three genes known to tolerate C-terminal epitope tags were selected: *chp1*$^+$ (SJAG_02621), *rik1*$^+$ (SJAG_02594) and *stc1*$^+$ (SJAG_04061). Chp1 is a

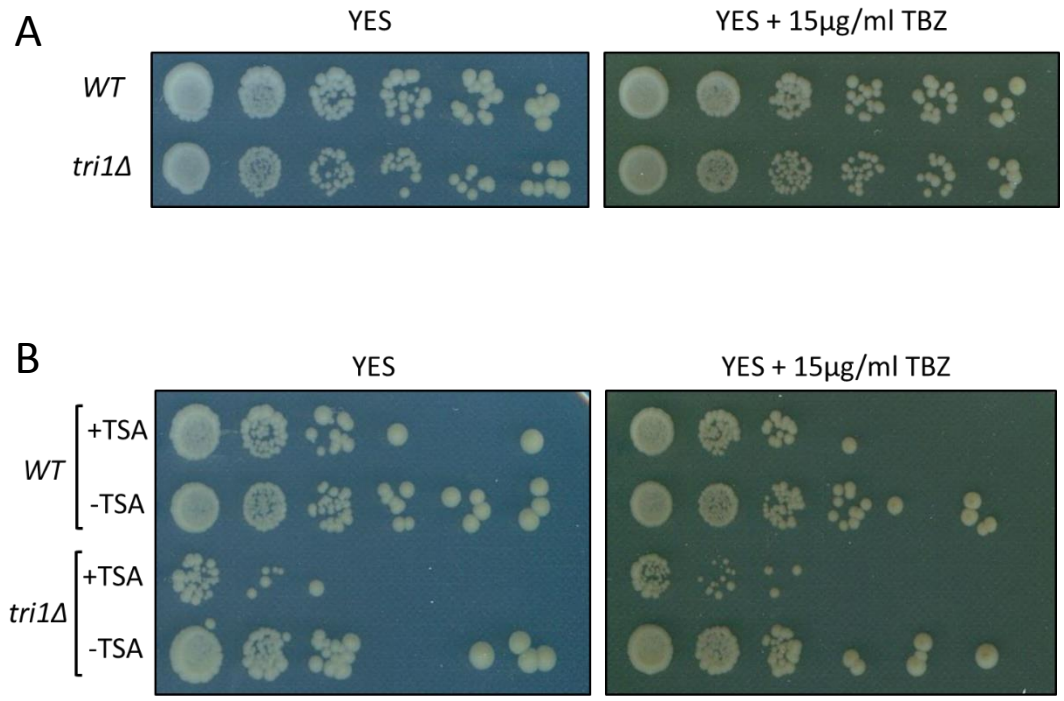


Figure 3.3 – TBZ sensitivity of *S. japonicus tri1Δ* strains in heterochromatin maintenance and establishment scenarios

(A) Spotting assay of *wild-type* and *tri1Δ* strains in a heterochromatin maintenance situation. 10-fold serial dilutions of logarithmically growing yeast were spotted onto non-selective YES, or YES plates containing 15µg/ml TBZ

(B) Spotting assay of *wild-type* and *tri1Δ* strains in a heterochromatin establishment situation. 10-fold serial dilutions of logarithmically growing yeast treated with TSA or DMSO (mock) were spotted onto non-selective YES, or YES plates containing 15µg/ml TBZ

chromodomain protein that forms part of the RITS complex, Rik1 is a component of the ClrC complex and Stc1 acts as a bridge between RITS and ClrC in *S. pombe*. Disruption of all three of these factors has been shown to impact pericentromeric heterochromatin formation (Bayne et al., 2010; Ekwall et al., 1996; Nakayama et al., 2001; Sadaie et al., 2004), thus these proteins should only retain the ability to nucleate heterochromatin if the epitope tags do not impair their function.

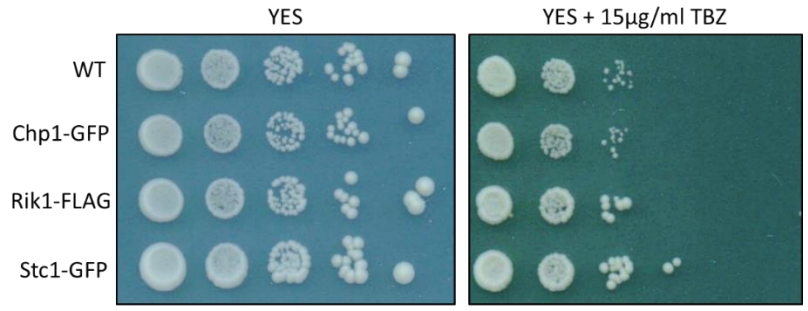
These three genes were targeted for epitope tag integration in a slightly different way to the deletion constructs; in this case 1kb of homology from the 3'end of the gene ORF minus the stop codon and 1kb of homology downstream of the targeted gene 3'UTR were cloned into plasmids carrying either FLAG-NaMX6 (pFA6a-FLAG-NatMX6) or GFP-sjUra4 (pSO729). These plasmids were then linearised with either NdeI for the pFA6a vector or XhoI for pSO729, and transformed into *S. japonicus* via electroporation.

Targeting efficiencies for these transformations were very high, ranging from 75-100% recovery of correctly integrated tags (Figure 3.4A). To validate whether these tagged proteins were stably expressed, these strains were subject to either FLAG-IP (for Rik1-FLAG) or GFP-IP (for Chp1-GFP and Stc1-GFP). All tagged proteins were detectable by western blot at roughly the expected size; Rik1-FLAG = ~120kDa, Chp1-GFP = ~150kDa and stc1-GFP = ~60kDa (Figure 3.4C). To assess the impact of the tag on protein functionality, these strains were spotted onto plates containing 15ug/ml TBZ. None of the strains showed any sensitivity to TBZ, indicating that the addition of an epitope tag did not impair their ability to nucleate heterochromatin at pericentromeres

A

S. japonicus Gene Target	*S. japonicus* Gene ID	Tagging cassette	Colonies Screened	Positive Colonies	% Positive
chp1⁺	SJAG_02621	C-terminal GFP-Ura4	12	9	75.00
rik1⁺	SJAG_02594	C-terminal FLAG-NatMX6	4	4	100.00
stc1⁺	SJAG_04061	C-terminal GFP-Ura4	6	5	83.33

B



C

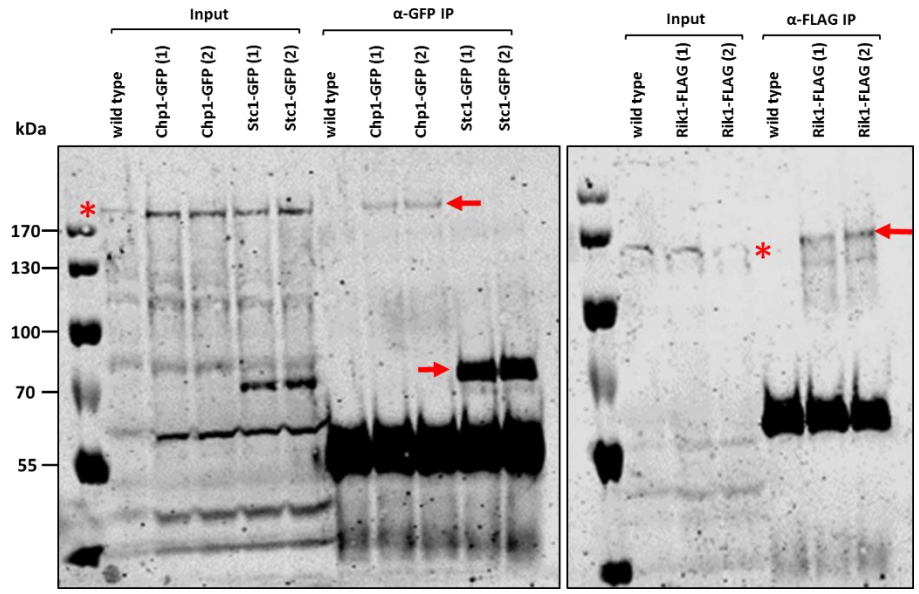


Figure 3.4 – Core RNAi proteins can be tagged, and tagged strains do not show chromosome segregation defects

(A) Table showing the total number of resistant colonies screened for each tagging candidate, and the percentage of those colonies that had a correct integration by PCR

(B) Spotting assay looking at chromosome segregation defects in strains with epitope-tagged RNAi components by assessing sensitivity to the microtubule destabilising drug TBZ

(C) Western blot analysis of affinity purified GFP-tagged Chp1 and Stc1 or FLAG-tagged Rik1. Arrows indicate immunoprecipitated proteins of interest. Asterisks (*) indicate the position of non-specific protein bands.

(Figure 3.4B). Thus integrations at RNAi gene loci can be recovered when these do not impair gene function.

3.6 - Chp1-GFP co-localises with H3K9me2 at retrotransposons

To quantitatively assess whether the tagging of these RNAi proteins had any effect on their ability to nucleate heterochromatin, functional analysis of the Chp1-GFP strain was carried out. Chp1 is a component of the RITS complex that has been shown to bind methylated histones via its chromodomain. Deletion of Chp1 in *S. pombe* abolishes siRNA production and largely reduces H3K9 methylation at centromeres (Bayne et al., 2010; Partridge et al., 2002; Sadaie et al., 2004; Schalch et al., 2009; Volpe, 2002). Due to its direct interaction with chromatin via the chromodomain, it can be efficiently enriched by standard ChIP. Both wild-type (containing untagged Chp1) and Chp1-GFP strains were used for ChIP with antibodies raised against the GFP tag and against H3K9me2. As retrotransposons are postulated to constitute the pericentromeric repeats in *S. japonicus* qPCR primers were designed to amplify either the LTR or ORF of each individual retroelement specifically, in order to include solo-LTRs as well as more complete elements in the analysis and to identify any differential modification that may occur across elements.

Chp1 was found to be specifically enriched at all retrotransposons, with little difference between elements or between LTR and ORF (Figure 3.5). This enrichment of Chp1 also seemed to mirror the level of H3K9me2 enrichment seen at each element, which is not surprising given that it directly binds H3K9me2 via its chromodomain. It was also evident that the addition of the GFP tag did not interfere with Chp1's presumed role in heterochromatin

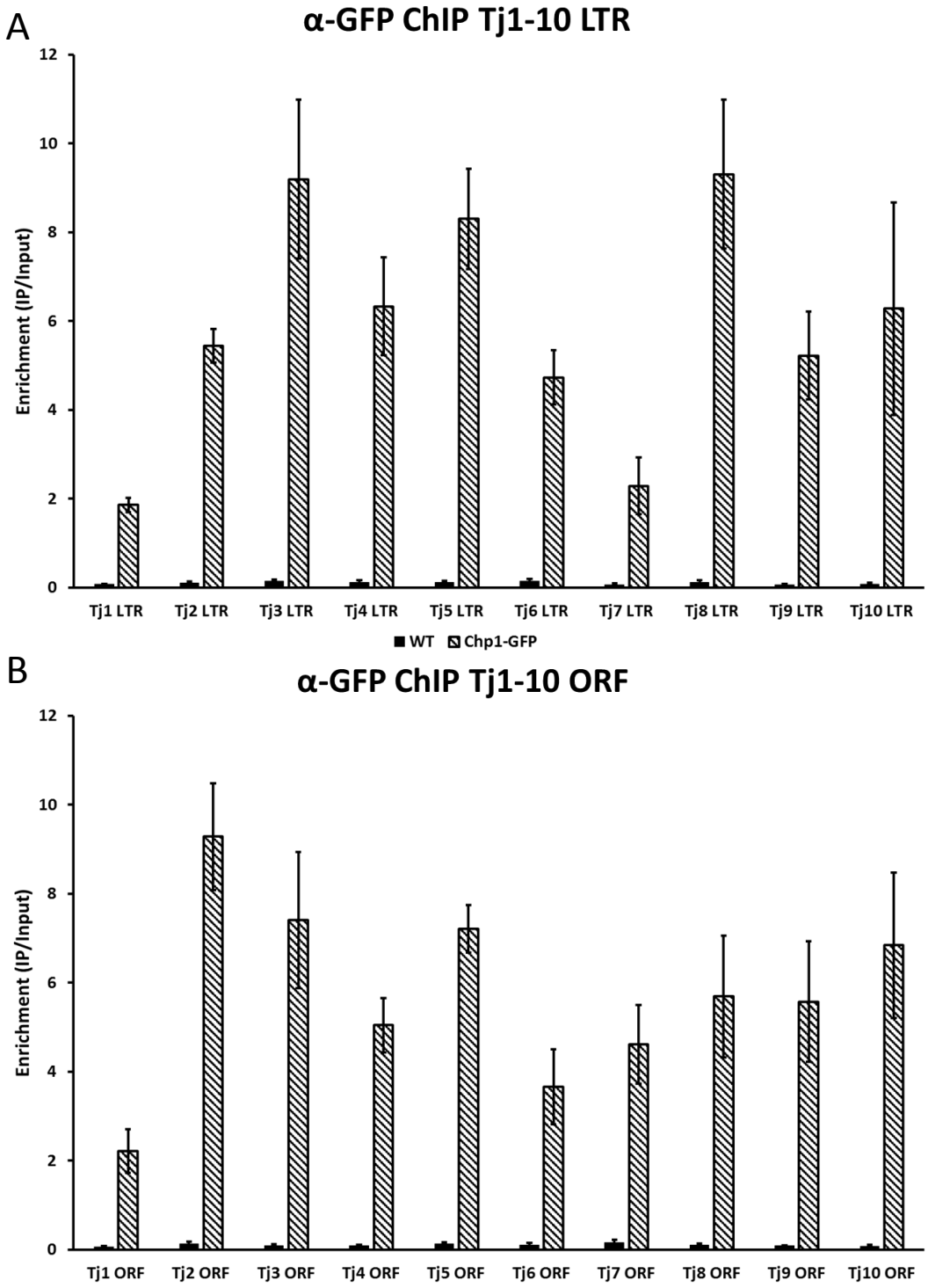


Figure 3.5 – Chp1-GFP localises at retrotransposons in *S. japonicus*

(A) α-GFP ChIP-qPCR assay to assess the enrichment of Chp1-GFP at *S. japonicus* retrotransposon LTRs
(B) α-GFP ChIP-qPCR assay to assess the enrichment of Chp1-GFP at *S. japonicus* retrotransposon ORF sequences

Error bars represent 1 S.D. from 3 biological replicates.

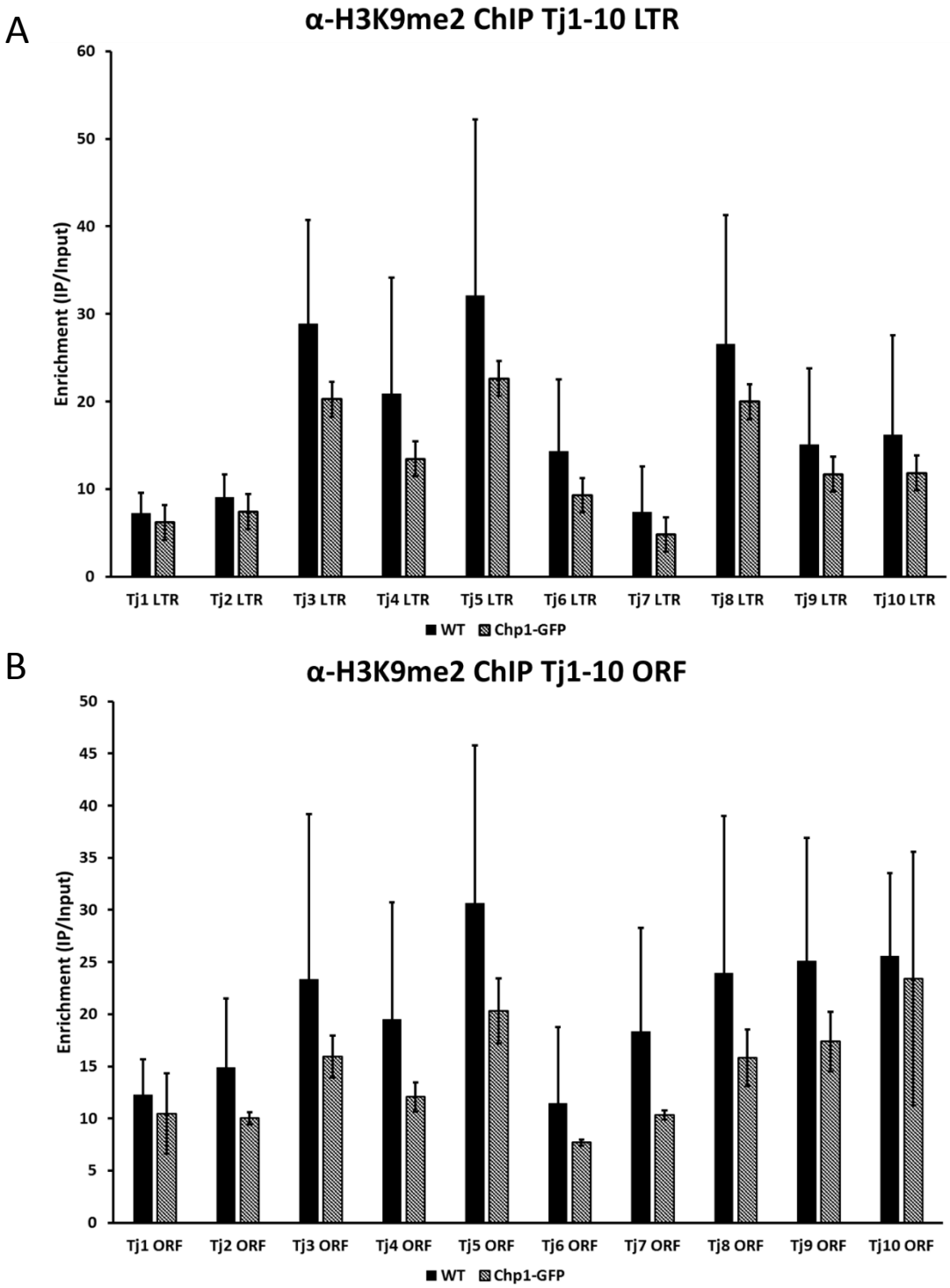


Figure 3.6 – Tagging of Chp1 with GFP does not affect H3K9me2 at retrotransposons

(A) α-H3K9me2 ChIP-qPCR assay to assess the enrichment of H3K9me2 at *S. japonicus* retrotransposon LTRs
(B) α-H3K9me2 ChIP-qPCR assay to assess the enrichment of H3K9me2 at *S. japonicus* retrotransposon ORF sequences

Error bars represent 1 S.D. from 3 biological replicates.

formation, as levels of H3K9me2 were the same in strains carrying either untagged or GFP-tagged Chp1 (Figure 3.6).

3.7 - Discussion

In this chapter I have presented the work I undertook to begin the characterisation of the RNAi pathway in *S. japonicus*.

Initial work focused on fully annotating all conserved components of the *S. japonicus* RNAi pathway, as the genome annotation for this species is still sparse in comparison to *S. pombe*. This is mainly due to that fact that the data has only been available since 2011 (Rhind et al., 2011), 9 years less that the *S. pombe* genome sequence (Wood et al., 2002), whilst far fewer laboratories are using *S. japonicus* as a model organism compared to *S. pombe*. This is also the first study that has attempted to understand how the RNAi pathway functions in *S. japonicus*.

Identification of the Ago1 and Dcr1 homologues in *S. japonicus* revealed that the main functional domains of these proteins are well conserved at the amino acid level, whilst expansion of this analysis to all conserved pathway members showed that there is a high degree of sequence identity between *S. pombe* and *S. japonicus* homologues of the core enzymatic components Ago1, Clr4 and Rdp1, whilst factors involved in the recruitment of complexes showed lower degrees of identity.

Local gene order around RNAi genes was also found to be conserved, despite the fact that gene synteny between *S. pombe* and *S. japonicus* is globally very low (Rhind et al., 2011). In both *S. pombe* and *S. japonicus*, RNAi genes run convergent with, and often overlap, genes on the opposite strand; this convergence is postulated to induce transient heterochromatin to silence these genes at certain points of the cell cycle (Gullerova et al., 2011).

Taken together, the conservation of all core RNAi factors, as well as the preservation of the local convergent gene order, indicates that the RNAi pathway itself may function, and be regulated in a similar way in *S. pombe* and *S. japonicus*. This is in contrast to observations made with two later branching members of the fission yeast clade, *S. octosporus* and *S. cryophilus*, which appear to have lost the conserved RITS complex component Tas3, and also possess truncated versions of Chp1 that are unable to rescue silencing of centromeric repeats in *S. pombe* (Upadhyay et al., 2017).

In order to investigate whether the RNAi pathways of *S. pombe* and *S. japonicus* function in analogous ways, I attempted to construct deletion mutants of a number of core RNAi factors, in order to deduce their function within the *S. japonicus* pathway. Attempts to isolate these mutants in a wild-type background were largely unsuccessful (with the exception of 2 Dcr1 mutants, which will be discussed in depth in chapter 4), with all transformants isolated carrying incorrectly integrated resistance markers. This occurred even when utilising a 'split-marker' strategy that had been shown in *C. Neoformans* to increase targeting efficiency whilst reducing false-positive background (Fu et al., 2006).

From here I decided to employ a system that had been shown to increase targeting efficiency even further, in an attempt to rule out the possibility that these RNAi gene loci were difficult to target. Deletions were made in the NHEJ proteins Pku70 and Pku80, which form a heterodimeric complex that bind DNA double-strand breaks, licencing the DNA ends for repair by NHEJ (Manolis et al., 2001; Walker et al., 2001). Deletion of these genes increases targeting efficiency by removing Ku-mediated NHEJ as a repair pathway, thus

driving repair by homologous recombination, theoretically increasing the chance of isolating correctly integrated resistance markers.

Even in this background deletions were not isolated; without a functional Ku-mediated NHEJ pathway the fragments were still integrated in non-homologous regions, indicating that another Ku-independent pathway may operate to integrate fragments, such microhomology-mediated end joining (MHMEJ) (Decottignies, 2007). This inability to recover functional mutants of RNAi genes in *S. japonicus*, even when attempted in genetic backgrounds that have been shown to favour integration by homologous recombination, may indicate that the RNAi pathway itself may be an essential process. This is in stark contrast to *S. pombe*, where deletion mutants of all core components of the RNAi machinery are viable, and a majority have no detectable impact on growth phenotype.

In order to try to establish whether all RNAi related factors may be essential, or whether only the core proteins were required for viability, an RNAi accessory factor was selected for deletion. The factor *tri1*$^{+}$ had been shown to be solely involved in the establishment of heterochromatin domains in *S. pombe* (Marasovic et al., 2013), thus deletion of this factor was predicted to have no effect on the maintenance of heterochromatin. Indeed, transformants deleted for *tri1*$^{+}$ were readily recovered, and when the *tri1Δ* strain was spotted onto TBZ, there was no evident growth defect, indicating the pericentromeric heterochromatin was still intact. However when treated with TSA, which removes H3K9 methylation by blocking deacetylation, and plated onto TBZ, these cells showed a reduction in viability when compared to a strain carrying wild type *tri1*$^{+}$, consistent with a loss of heterochromatin at centromeres.

Strikingly there was also an apparent loss of viability in TSA treated *tri1Δ* cells plated onto non-selective media, which may indicate that the presence of heterochromatin, and by extension the factors responsible for establishing and maintaining the mark, is essential for viability in *S. japonicus*. Caution should be exercised when drawing conclusions from this assay, as the perceived loss of viability of the *tri1Δ* strain under non-selective conditions may be attributable to differential loading of the wild-type and mutant cells; thus further work will need to be undertaken to corroborate this observation.

In order to see if RNAi loci could be manipulated at all, I attempted to epitope tag a number of core factors. Correct integration of the tag was achieved with high efficiency in all cases, and the tagged proteins could be detected by western blot, yet showed no detrimental effect upon organism fitness in either the presence or absence of TBZ. In support of this, ChIP analysis of the Chp1-GFP strain, found no impact of the tag on H3K9me2 levels. Interestingly, Chp1-GFP was also shown to co-localise with H3K9me2 at all families of *S. japonicus* retrotransposon, suggesting that the Ago1 containing RITS complex may be present at these loci. Although it has previously been shown that H3K9me2 is present at centromeric loci, and that a majority of siRNAs are generated from transposable elements (Rhind et al., 2011), this is the first demonstration that H3K9me2 is present at all 10 families or retrotransposon, and the first evidence of a physical link between components of the RNAi pathway and retrotransposons in *S. japonicus*.

Thus far evidence to suggest that RNAi is essential in *S. japonicus* is circumstantial, as no direct proof has been yet obtained. One drawback to *S. japonicus* is that it has only been adopted as model organism relatively

recently, and a number of genetic tools available for *S. pombe* do not exist or are not yet optimised for use with *S. japonicus*. More work is required to generate methods that will allow the study of essential processes in *S. japonicus*, the extent of this will be addressed in Discussion chapter 7.1.

Chapter 4 – Investigating the role of *dcr1*$^{+}$ in *S. japonicus*

4.1 - Introduction

In the previous chapter I presented data to support the hypothesis that the RNAi pathway and the formation of heterochromatin may be essential processes in *S. japonicus*. Despite the fact that deletion mutants of a majority of core RNAi genes could not be isolated, two mutants of the ribonuclease Dcr1 were isolated at very low frequency. In *S. pombe* this gene is essential for the production of siRNAs from centromeric repeats, and mutants of *dcr1*$^{+}$ are known to show reduced levels of H3K9 methylation, increased levels of centromeric transcript accumulation and defects in chromosome segregation (Bühler et al., 2008; Cam et al., 2005; Provost et al., 2002; Volpe, 2002). As retrotransposon sequences in *S. japonicus* are postulated to act in an analogous way to *S. pombe* centromeric repeats (Rhind et al., 2011), I set out to investigate the impact of perturbing *dcr1*$^{+}$ upon retrotransposon silencing. To do this, I utilised a number of techniques to assess changes in the pattern of histone modification, transcript accumulation and small RNA production from these loci, whilst also evaluating the impact on genome organisation. Taking this further, I extended analysis to evaluate the impact of losing Dcr1 on the organism as a whole, whilst also looking to address why these *dcr1*$^{+}$ mutants were able to survive, whilst other RNAi components seem to be essential for viability in *S. japonicus*.

4.2 – Deletion or disruption mutants of *dcr1*$^+$ exhibit differing growth phenotypes

In the previous chapter it was reported that deletion mutants for a number of core RNAi factors, as well as factors essential for the maintenance of heterochromatin, could not be recovered in *S. japonicus*. Despite this, two mutants for the Ribonuclease Dcr1 were isolated, at very low frequency. The first of these mutants was constructed with a fragment that replaced the entire ORF of *dcr1*$^+$ with a NatMX6 cassette. During isolation and subsequent genotyping of this mutant, it was noted that colonies had a tendency to grow invasively into agar plates, even on rich media. Microscopic analysis of this mutant revealed that the cells grew exclusively as invasive hyphae (Figure 4.1A), and had lost the ability to grow vegetatively as yeast.

S. japonicus is known to exhibit a dimorphic growth phenotype under certain conditions such as nutrient or genomic stress (Furuya and Niki, 2010; Sipiczki et al., 1998), however the inability to revert to yeast growth in rich media indicated a growth phenotype that may have been due to the deletion of the *dcr1*$^+$ ORF. Analysis of the locus revealed that the 3' end of the *dcr1*$^+$ ORF overlapped with the last 729bp of the *spn6*$^+$ 3' UTR which is encoded on the opposite strand. In order to assess whether it was deletion of *dcr1*$^+$ or disruption of the *spn6*$^+$ 3'UTR that caused this dimorphic switch, a shorter disruption mutant of the *dcr1*$^+$ ORF was constructed, that would not impact the *spn6*$^+$ transcript. This consisted of a NatMX6 cassette that replaced 269bp of the 5'UTR and 108bp of the CDS of *dcr1*$^+$ (Figure 4.1B). A single positive clone was isolated for the minimal disruption mutant; this mutant appeared to have a mixed 'semi-hyphal' phenotype although a large proportion of cells

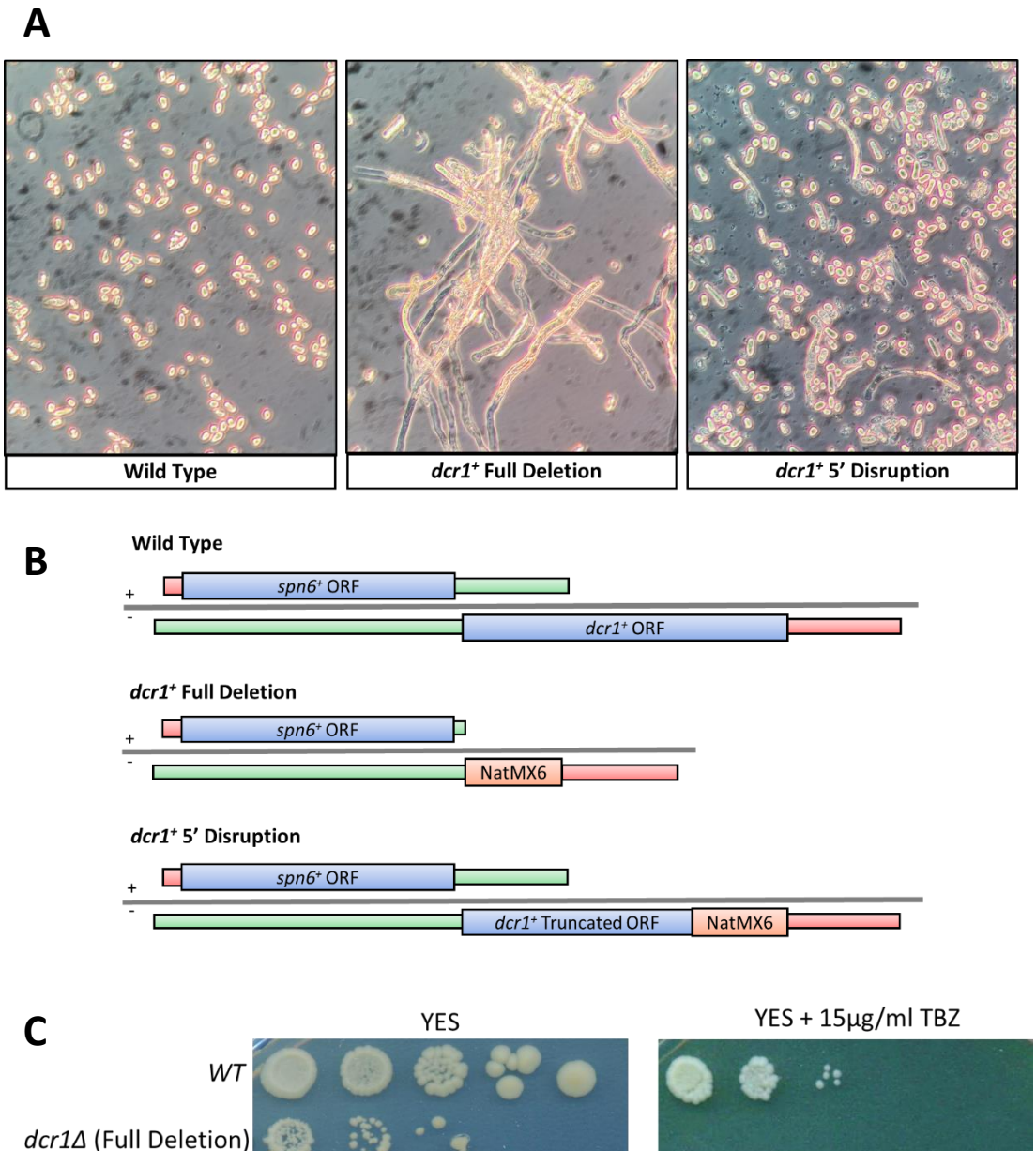


Figure 4.1 – Two mutants of *dcr1⁺* in *S. japonicus* give rise to differing growth states but identical chromosome segregation defects
(A) Microscopic analysis of wild-type, *dcr1⁺* deletion (*dcr1Δ)* and *dcr1⁺* 5' disruption (*dcr1Δ5')* strains grown in YES at 32°C
(B) Schematic representation of the *dcr1⁺/spn6⁺* locus in the wild-type and mutant strains
(C) Spotting assay looking at chromosome segregation defects in *dcr1Δ* strains by assessing sensitivity to the microtubule destabilising drug TBZ

grew vegetatively as yeast, indicating that the severe hyphal phenotype observed in the full *dcr1*$^{+}$ ORF deletion was due to interruption of the *spn6*$^{+}$ mRNA. As well as impacting cell morphology, disruption of the *dcr1*$^{+}$ ORF also negatively affected growth rate, with generation time slowed from 110 minutes in wild-type, to 155 minutes in the *dcr1*$^{+}$ disruption strain, grown in YES at 32°C.

4.3 - Disruption of *dcr1*$^{+}$ appears to impact functional centromere formation

As RNAi is postulated to silence retrotransposons, which constitute a majority of the presumed *S. japonicus* centromeres (Rhind et al., 2011), the *dcr1Δ* mutant strains were assayed for defects in centromeric heterochromatin formation by testing for sensitivity to the microtubule destabilising drug TBZ. The *dcr1Δ* strains exhibited a hypersensitivity to TBZ (Figure 4.1C), indicating that disruption of *dcr1*$^{+}$ likely impaired formation of centromeric heterochromatin, which in turn impacted the fidelity of chromosome segregation. The disruption mutant of *dcr1*$^{+}$ exhibited the same phenotype on TBZ as the full deletion, indicating that the minimal deletion of the 5' portion of the gene is enough to perturb function in a comparable way to full deletion of the ORF. As there was no difference in TBZ sensitivity between these two mutants, all further molecular analysis of the *dcr1*$^{+}$ mutant was carried out in the disruption background, which from this point will be known as *dcr1Δ5'*.

4.4 - Disruption of *dcr1*$^+$ dramatically changes the small RNA pool

In *S. pombe*, deletion of Dcr1 completely abrogates production of centromeric siRNAs and causes an accumulation of centromeric repeat transcript. This loss of siRNAs is accompanied by a marked decrease in H3K9me2 levels at centromeres, and an increased incidence of mis-segregated chromosomes (Provost et al., 2002; Volpe, 2002). In order to assess whether the disruption of *dcr1*$^+$ lead to a complete loss of siRNAs, as evident in *S. pombe*, a northern blot was carried out. From the previously published siRNA-Seq dataset (Rhind et al., 2011), antisense oligonucleotides corresponding to the three most abundant siRNA reads for each element were created, and end labelled for use as probes. A probe against snoRNA58 was used as a loading control (Figure 4.2). For Tj1/Tj3/Tj4/Tj5/Tj6/Tj10, the disruption of *dcr1*$^+$ appeared to completely abolish siRNA production from these elements, as would be expected for Dcr1 dependent siRNAs that feed into the RNAi pathway. However for a number of elements, namely Tj2, Tj7, Tj8 and Tj9, disruption of *dcr1*$^+$ caused a change in the small RNA species generated. For both Tj2 and Tj9 this change corresponded to a large reduction but not complete loss of small RNAs, with the remaining small RNA species exhibiting a slightly larger size for Tj2 and slightly smaller for Tj9. The most dramatic change occurred for Tj7 and Tj8, whose small RNA pools were altered but not reduced. For Tj7 there was a larger smear of small RNA species with most appearing to be smaller than observed in a wild type situation, whilst for Tj8 the level of small RNA species was comparable to wild type, however these species appeared to exhibit a slightly larger size profile.

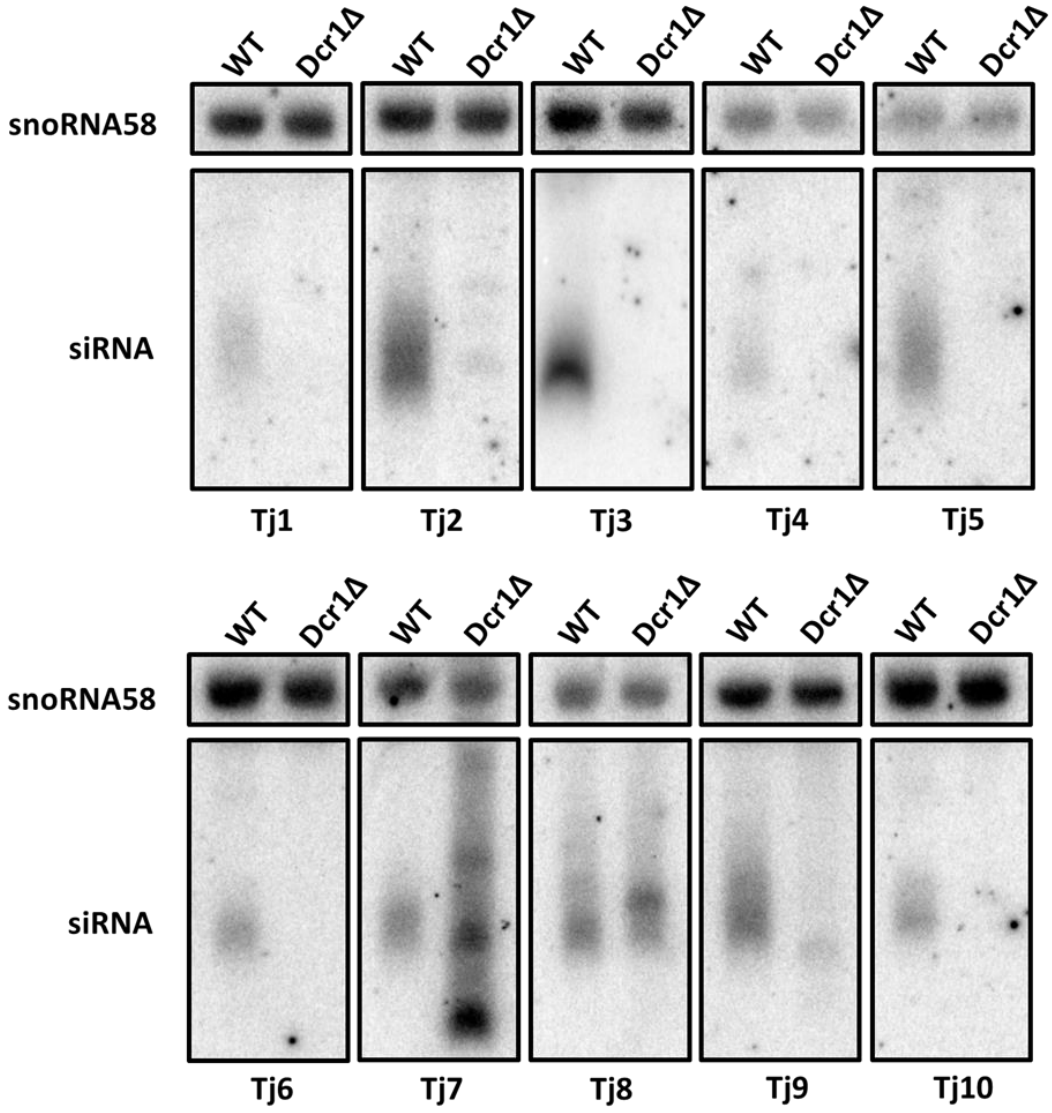


Figure 4.2 – Disruption of *dcr1*$^{+}$ causes retrotransposon derived small RNAs to be lost from all elements except Tj2, Tj7, Tj8 and Tj9

Northern blot analysis of small RNAs derived from each of the annotated *S. japonicus* retrotransposon families in wild-type and *dcr1Δ5'* strains. Antisense probes were created for each of the three most abundant siRNAs for each element, using published small-RNA sequencing data (Rhind et al., 2011). A probe against snoRNA58 was used as a loading control.

In order to specifically assess the extent to which disrupting *dcr1*$^{+}$ altered the small RNA pool, deep sequencing of small RNAs was carried out from total RNA extracted from logarithmically growing wild-type and *dcr1Δ5' S. japonicus*. Trimmed sequence reads in the rage of 14-35nt were aligned to the *S. japonicus* SJ5 genebuild for analysis.

The mean number of mapped sequence reads over two replicates in the wild-type was 12,993,353 and in the Dcr1 mutant 12,266,630. There was no significant difference in the total number of reads between wild type and the mutant, neither was there a difference in the number of reads mapping to annotated structured ncRNA (tRNA, rRNA, snoRNA, snRNA) or cDNA. There was however a surprising increase in the number of reads mapping to annotated retrotransposons Tj1-10, with the *dcr1Δ5'* mutant having nearly four times as many retrotransposon derived small RNAs as the wild type (Figure 4.3). This was unexpected, as deletion of *dcr1*$^{+}$ in *S. pombe* completely abrogates siRNA production and I have shown via northern blot that disruption of *dcr1*$^{+}$ in *S. japonicus* caused a reduction or complete loss of retrotransposon derived siRNAs for all elements, except Tj7 and Tj8.

Looking generally at the presumed telomeres and centromeres, the amount of small RNA produced in the *dcr1*$^{+}$ mutant was generally reduced (Figures 4.4 & 4.5), however there were certain regions that seemed to retain small RNA signal, and even appeared to generate a greater number of small RNA species in the *dcr1Δ5'* background.

In order to investigate whether specific elements were losing or gaining small RNA species in the *dcr1Δ5'* background, trimmed reads of 14-35nt were aligned to each retrotransposon family individually. The number of reads that

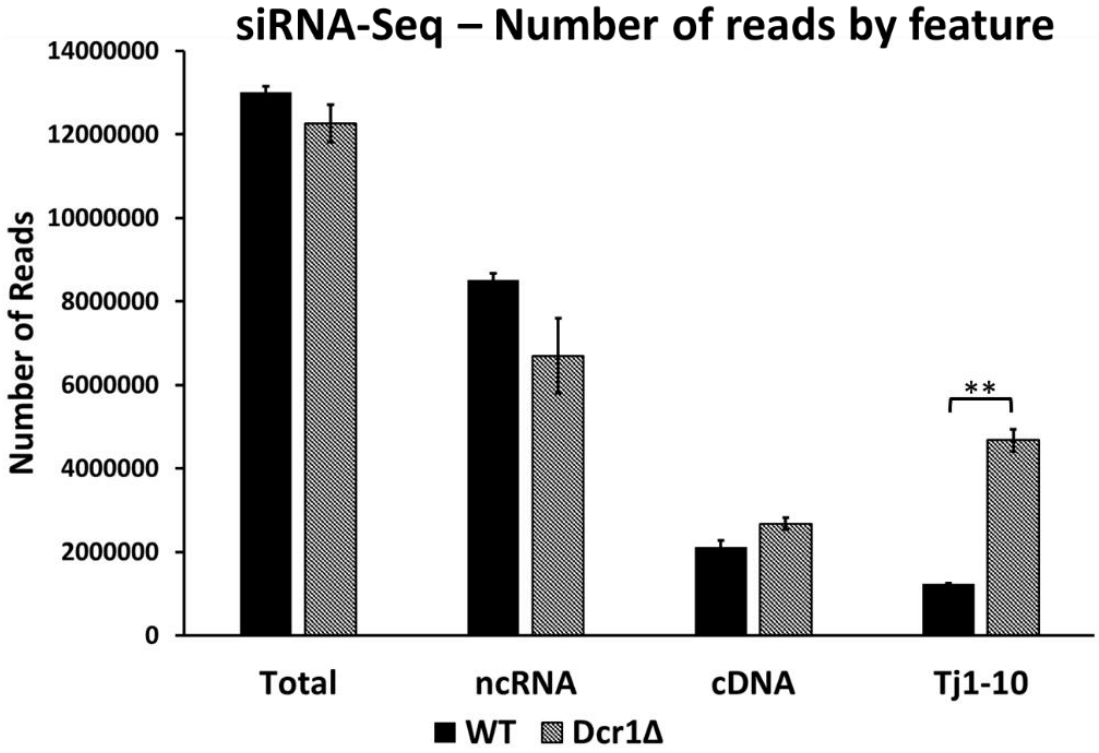


Figure 4.3 – Significantly more small RNA reads map to annotated retrotransposons in the *dcr1Δ5'* mutant than the wild-type

Graph showing the total number of reads, and the number of these that mapped to annotated ncRNA, cDNA and Tj1-10 retrotransposons in the wild type and *dcr1Δ5'* mutant

Error bars represent 1 S.D. from 2 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

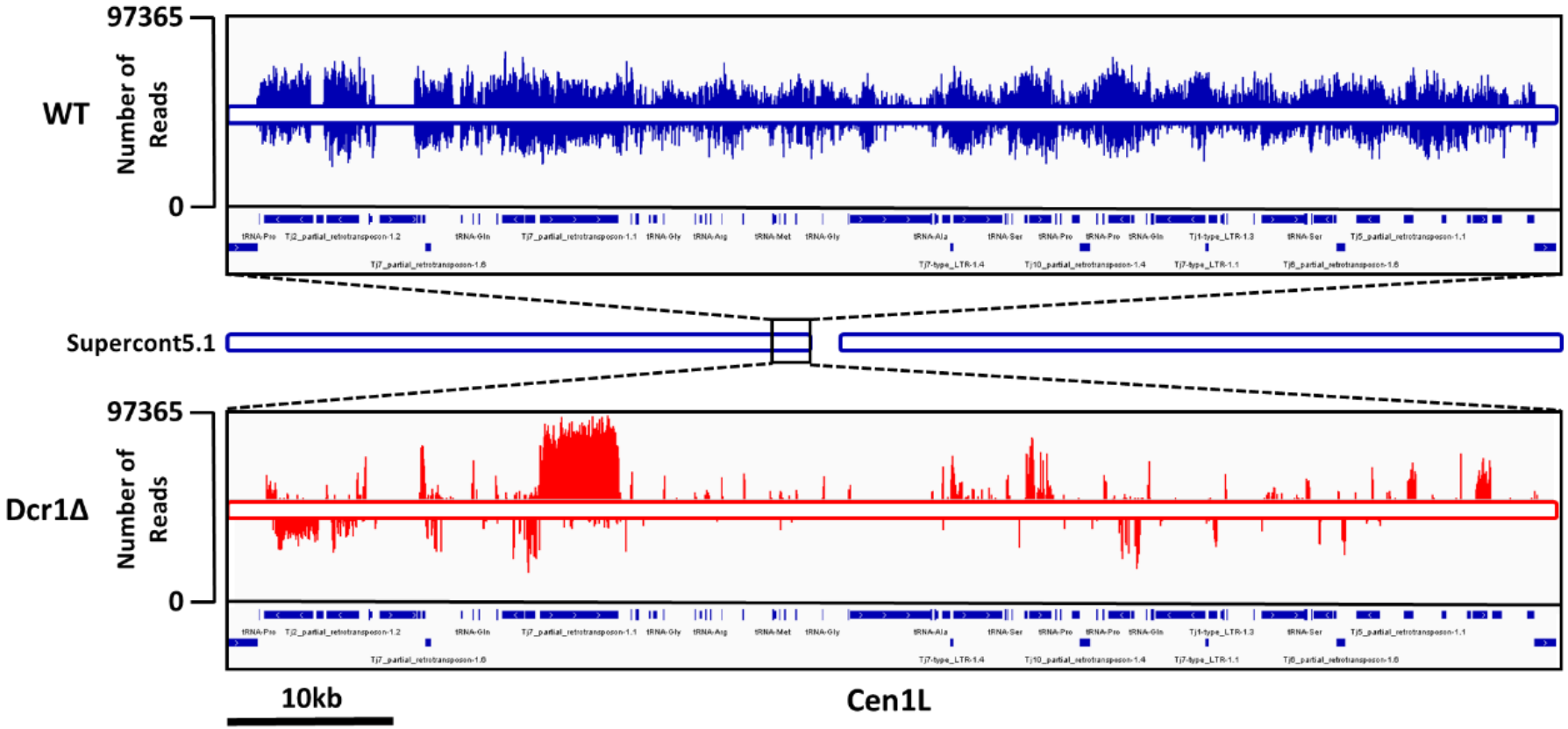


Figure 4.4 – Centromeric regions mostly lose siRNAs in the absence of *dcr1⁺*, however some loci retain small RNA signal

Genome browser view showing the number of small RNA reads that align to the Cen1L region in wild type and *dcr1Δ* strains. Blue bars below represent annotated retrotransposon sequences

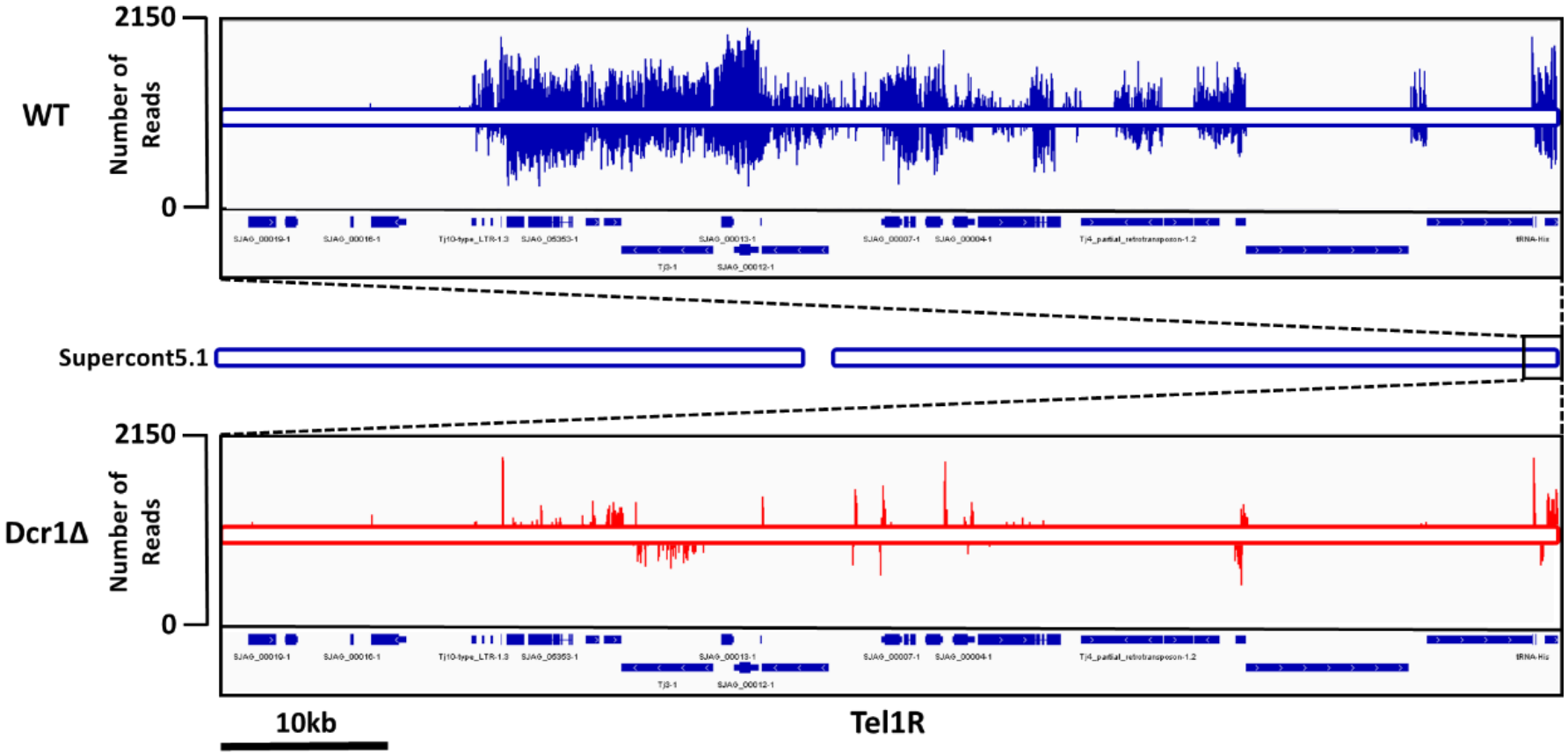


Figure 4.5 – Telomeric regions mostly lose siRNAs in the absence of *dcr1*$^{+}$, however some loci retain small RNA signal

Genome browser view showing the number of small RNA reads that align to the Tel1R region in wild type and *dcr1Δ* strains. Blue bars below represent annotated retrotransposon sequences and subtelomeric genes

mapped to each element were calculated for the wild type and *dcr1Δ5'* mutant, in order to assess the distribution of reads across different retrotransposon families (Figure 4.6).

In the wild type strain, small RNA reads were fairly evenly distributed across all elements, with the smallest fraction (2%) aligning to Tj1 elements and the largest fraction (20%) aligning to Tj10 family elements. However, in the *dcr1Δ5'* mutant, this distribution was dramatically changed. 91% of retrotransposon small RNA reads mapped to Tj7 family elements, whilst 6% mapped to Tj9. 1% mapped to each of Tj2, Tj3 and Tj5, whilst Tj1, Tj4, Tj6, Tj8 and Tj10 accounted for less than 0.5% of mapped reads.

As well as the distribution, the total number of reads mapping to each element also changed remarkably, with all elements except Tj7 and Tj9 generating less small RNA species in the absence of Dcr1. The number of reads that mapped to Tj9 elements increased from 205,375 in the wild type to 283,868 in the *dcr1Δ5'* strain, however the change for Tj7 elements was much greater, increasing from 138,216 reads in the wild type to 4,302,841 reads in the *dcr1Δ5'* mutant, an increase of over 30-fold.

In order to assess whether the small RNA species generated in the *dcr1Δ5'* mutant shared characteristics of wild type siRNAs, the size profile of mapped reads for each retrotransposon family was calculated. In wild-type cells, the most abundant small RNAs were between 21-24nt, with a peak between 22-23nt. In the *dcr1Δ5'* mutant this size profile was shifted towards smaller sequences, with a broader profile enriched between 14-20nt, and a peak at 18-19nt (Figure 4.7B).

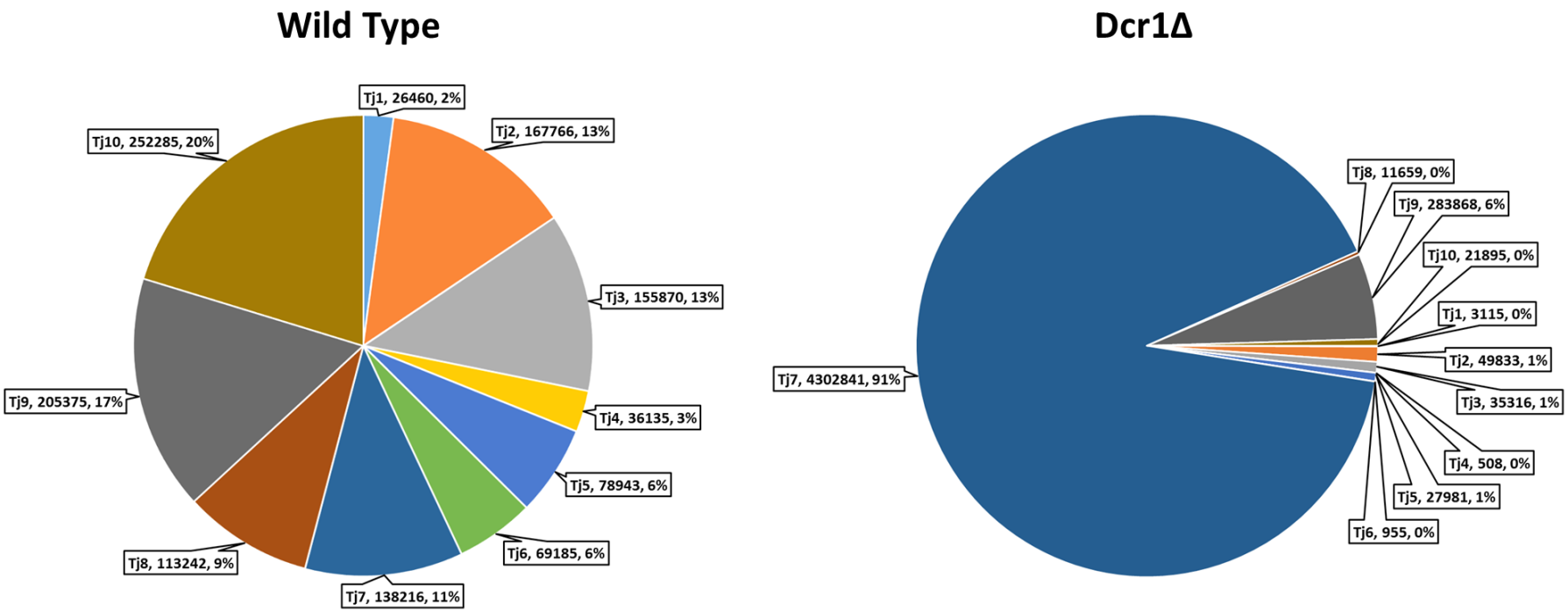


Figure 4.6 – Most *S. japonicus* retroelements lose siRNAs in the absence of *dcr1*$^{+}$, however Tj7 and Tj9 generate more small RNAs

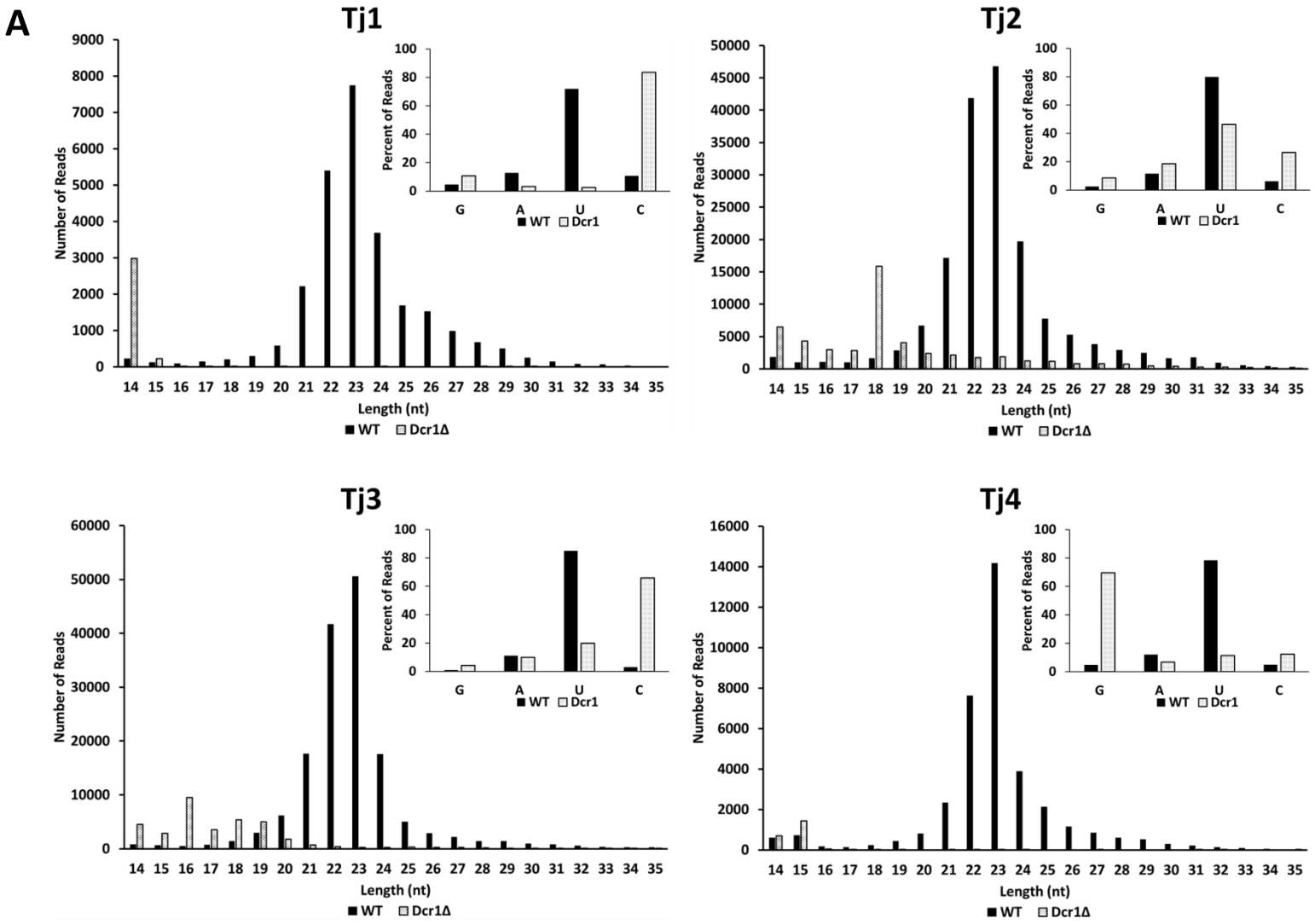
Pie charts showing the distribution of retrotransposon-mapped small RNA reads in the wild type and *dcr1Δ* mutants. Both the percentage and number of reads that map to each retrotransposon family

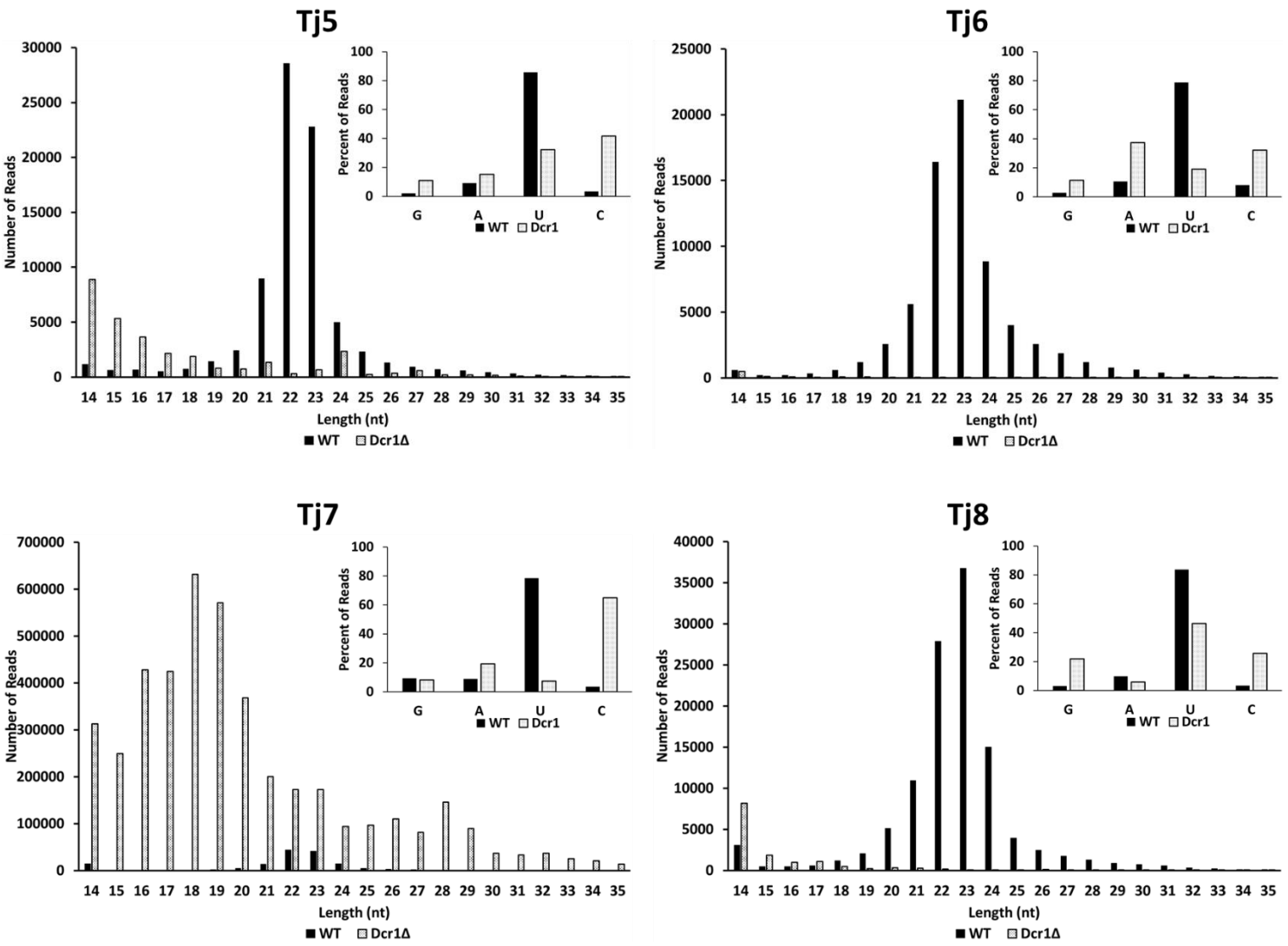
Breaking this down further to look at individual elements (Figure 4.7A), in wild type cells all retrotransposon derived small RNAs exhibited the typical Dcr1 dependent size profile, with a peak at 22-23nt. For those elements that appeared to lose siRNAs upon *dcr1*$^{+}$ disruption, the mapped small RNAs peaked at around 14-15nts, whereas for those elements that retained siRNAs (Tj7 and Tj9) the size profile was much broader, with a peak at 18nts for these elements. As Tj7 contributed to a majority of the small RNA reads in the *dcr1Δ* mutant, the combined profile of all elements was dominated by Tj7 derived sequences.

Another characteristic of centromeric siRNAs in *S. pombe* is the strong bias towards a Uracil base at the 5' position (Bühler et al., 2008; Djupedal et al., 2009). To assess whether the small RNAs generated in *S. japonicus* possessed this same characteristic, the composition of retrotransposon mapping small RNAs was examined. Overall, for elements Tj1-10 there was a very strong bias toward a 5' U in wild-type cells (Figure 4.7B), with 81% of reads carrying a U in the first position. In contrast 5'G was strongly selected against, with only 3% of reads starting with a Guanine. In contrast, deletion of *dcr1*$^{+}$ caused a dramatic shift in the 5' bias of mapped small RNA reads. Across Tj1-10, 63% of reads began with a cytosine, up from just 4% in wild-type, whilst the 5' U bias dropped from 81% to 9%. The bias towards a 5' G or A was increased over wild type, from 3% to 8.5% for G and from 11% to 19% for A, however these still remained much lower in comparison to a 5' C. Breaking this bias down over the individual elements (Figure 4.7A), it was evident that this overall shift from a 5' U to a 5' C was mainly contributed to by reads mapping to Tj7 or Tj9, as these elements accumulated more small RNAs in the *dcr1Δ5'* mutant, whilst

the other 8 elements all lost small RNAs. Although the relative abundance of certain 5' nucleotide biased reads was elevated for these 8 elements, the absolute number of reads was greatly reduced. Taken together, the fairly distinct size profile and 5' nucleotide bias of small RNA species generated from Tj7 and Tj9 in the *dcr1Δ5*' mutant, indicated that these species may be generated by a factor with a defined product specificity.

In order to assess the extent of the changes in siRNA levels across the genome, differential expression analysis was performed, using DESeq2 (Love et al., 2014) in combination with the published *S. japonicus* gene coordinates and the retrotransposon coordinates listed in Supplementary Table 1. From this it was evident that retrotransposons could be divided into two groups; those that gained small RNAs (Supplementary Table 2) upon deletion of *dcr1*$^{+}$ and those that lost siRNAs (Supplementary Table 3). This divide was not solely dictated by the retrotransposon family, but instead by the individual element itself. Although those elements that produced more siRNAs upon deletion of Dcr1 were generally either Tj2 family, Tj7 family or Tj9 family, only a few specific elements displayed this increase, whilst a majority of the Tj2, Tj7 and Tj9 elements actually lost small RNA signal without Dcr1. In addition, a single Tj5 and Tj6 element also seemed to gain small RNAs when Dcr1 was absent. It was also possible to assess which strand these small RNAs were derived from, again using DESeq2. From the analysis, it was evident that those elements that lost small RNAs in the absence of Dcr1 lost both sense and antisense derived species; 133 elements displayed a statistically significant ($p<0.05$) decrease in sense derived small RNAs, whilst 157 elements displayed a statistically significant decrease in antisense small RNAs. In contrast, 18 retroelements


A
Tj1
Tj2
Tj3
Tj4
Number of Reads
Length (nt)
Percent of Reads
G
A
U
C
WT
Dcr1
Dcr1Δ


Tj5
Number of Reads
Length (nt)
Percent of Reads
G A U C
WT Dcr1
WT Dcr1Δ
Tj6
Number of Reads
Length (nt)
Percent of Reads
G A U C
WT Dcr1
WT Dcr1Δ
Tj7
Number of Reads
Length (nt)
Percent of Reads
G A U C
WT Dcr1
WT Dcr1Δ
Tj8
Number of Reads
Length (nt)
Percent of Reads
G A U C
WT Dcr1
WT Dcr1Δ

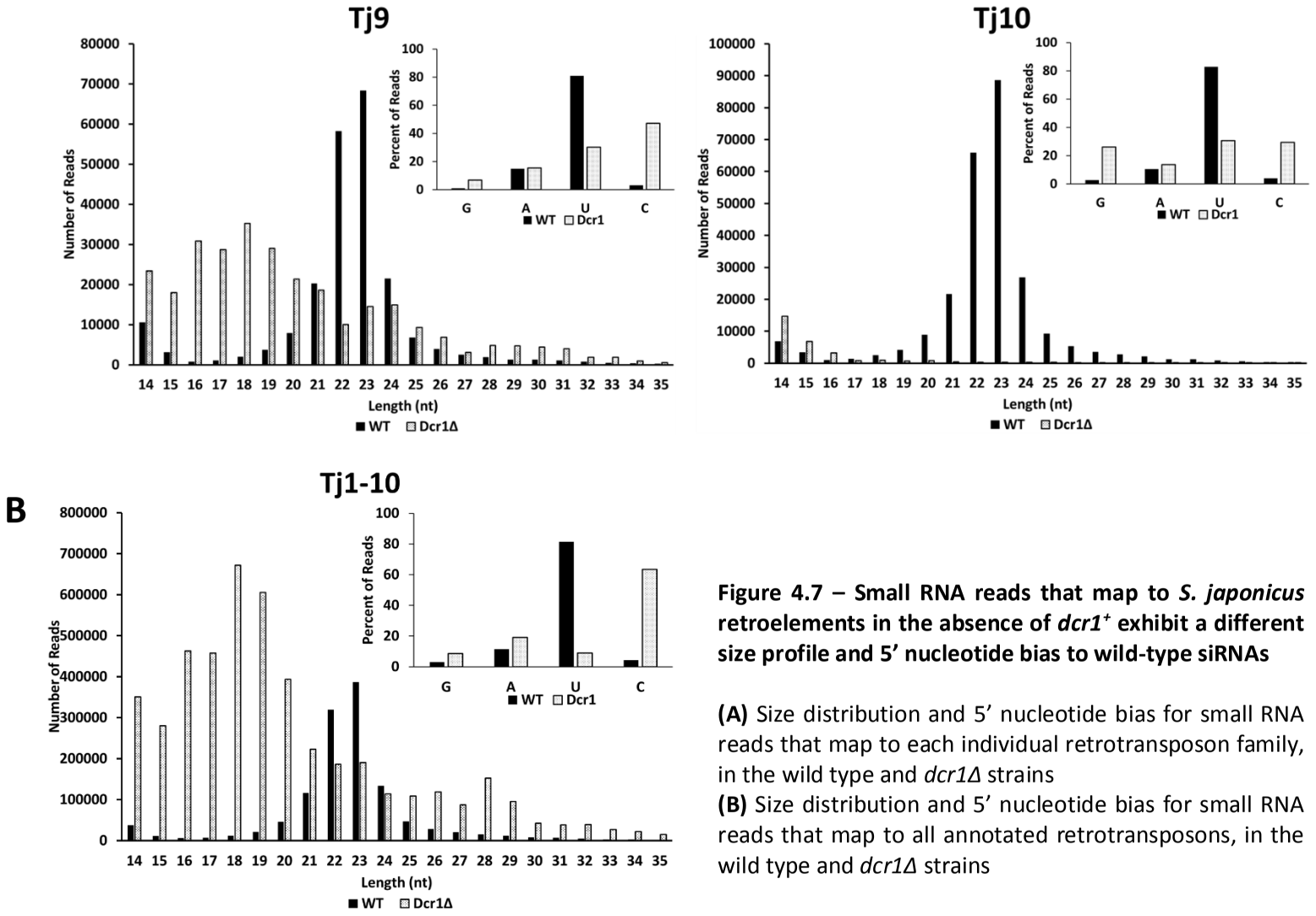


Figure 4.7 – Small RNA reads that map to *S. japonicus* retroelements in the absence of *dcr1*$^{+}$ exhibit a different size profile and 5' nucleotide bias to wild-type siRNAs

(A) Size distribution and 5' nucleotide bias for small RNA reads that map to each individual retrotransposon family, in the wild type and *dcr1Δ* strains
(B) Size distribution and 5' nucleotide bias for small RNA reads that map to all annotated retrotransposons, in the wild type and *dcr1Δ* strains

showed a statistically significant increase in sense derived small RNAs without Dcr1, however there were no elements that accumulated more antisense small RNA in this *dcr1Δ5*' mutant. This indicated that if an element accumulated small RNA species in the absence of Dcr1, these RNAs were derived exclusively from sense transcript (Supplementary Tables 4 and 5).

4.5 – Disruption of *dcr1*$^{+}$also alters small RNA production from non-retrotransposon loci

Aside from retrotransposons, there were a number of other genes that lost siRNAs upon disruption of *dcr1*$^{+}$. Amongst the genes that appeared to lose the largest amount of small RNAs without Dcr1 were the *S. japonicus* homologues of the *S. pombe* Tlh1/2 telomeric helicases (SJAG_00013/SJAG_05353/ SJAG_05105/SJAG_05173/SJAG_06642), as well as genes directly up or downstream (SJAG_00012/SJAG_05354/SJAG_06640) of these helicases (Supplementary Table 3). This indicated that in *S. japonicus* Dcr1 was responsible for processing transcripts from these telomeric genes into siRNAs in the wild type situation, as is the case in *S. pombe* (Djupedal et al., 2009).

4.6 - Disruption of *dcr1*$^{+}$ reduces H3K9me2 levels at specific retrotransposons

In *S. pombe*, repeat units that define the pericentromere are enriched for nucleosomes that contain a methylated lysine 9 on histone H3 (Nakayama et al., 2001). In this fission yeast, deletion of *dcr1*$^{+}$ abolished siRNA production and reduces H3K9me2 to between 20-40% of wild-type at pericentromeric heterochromatin (Bayne et al., 2008, 2010), indicating that Dcr1 is partially required to maintain wild type levels of histone methylation at these loci. As disruption of *dcr1*$^{+}$ abrogated legitimate siRNA production in *S. japonicus,* I wanted to see whether the *dcr1Δ5'* mutant exhibited a similar loss of H3K9me2, as exhibited in *S. pombe*. To do this, ChIP-qPCR was carried out using an antibody raised against H3K9me2 and qPCR primers specific for each individual retroelement. Previous studies (Rhind et al., 2011), in combination with data I presented in the previous chapter, have shown that these elements are associated with heterochromatin in a wild-type background.

Surprisingly, disruption of *dcr1*$^{+}$ did not impact H3K9me2 levels at most retrotransposons, as a majority of elements retained methylation to wild type levels upon loss of *dcr1*$^{+}$. There were some notable exceptions, with specific elements appearing to lose the H3K9me2 mark quite dramatically (Figure 4.8).

At Tj2 and Tj8, methylation decreased to around 40% of wild type levels, however only the loss at the LTRs of both elements was considered statistically significant. The methylation at the LTR of Tj9 was reduced to around a quarter of wild type, whilst the methylation level within the ORF changed much less significantly. The most striking overall loss of methylation occurred at Tj7, with

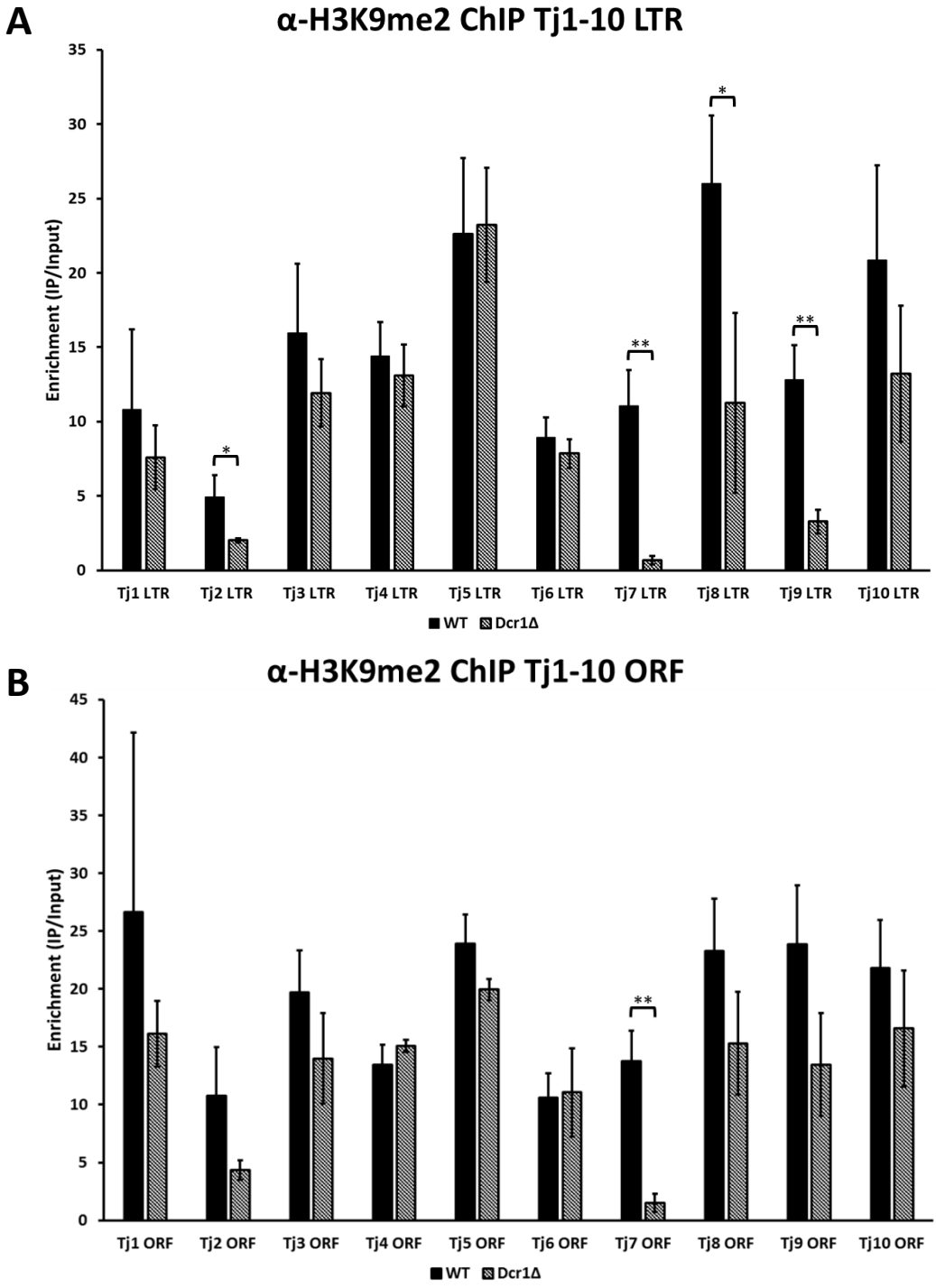


Figure 4.8 – Disruption of *dcr1*$^{+}$causes H3K9me2 to be lost from only a subset of retrotransposon elements

(A) α-H3K9me2 ChIP-qPCR assay to assess the enrichment of H3K9me2 at *S. japonicus* retrotransposon LTRs
(B) α-H3K9me2 ChIP-qPCR assay to assess the enrichment of H3K9me2 *S. japonicus* retrotransposon ORF sequences
Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

the LTR losing around 94% of methylation, whilst methylation within the ORF was reduced by around 90%. This indicated that each retrotransposon family may be differentially regulated by Dcr1, with further differences occurring between LTRs and coding sequence of elements.

4.7 - Histone H3 is also lost from retroelements that lose H3K9me2

The loss of H3K9me2 at specific retrotransposon loci could either be due to a defect in placing the K9-dimethyl mark on histone H3, or it could have been caused by the loss of histone H3 from these loci. To investigate whether it was K9 dimethylation that was lost specifically, ChIP against total histone H3 was carried out (Figure 4.9). For those elements that lost H3K9me2 (Tj2, Tj7 and Tj9), Histone H3 levels were also reduced by a similar factor. Tj2 lost 56% and 57% of its total H3 from LTRs and ORFs respectively. Tj7 lost 96% of H3 from LTRs and 93% of H3 from ORFs, whilst the LTR of Tj9 lost 77% of H3 upon *dcr1*$^{+}$ disruption.

In order to establish whether the loss of H3K9me2 was solely down to the loss of the H3 nucleosomes, the ratio of modified to unmodified H3 was calculated. For all elements, there was no difference in this ratio between wild type and the *dcr1Δ5*' strain (Figure 4.10). This indicated that the loss of H3K9 dimethylation was not due to a defect in placing the mark, but rather as a direct result of bulk loss of the Histone H3 substrate itself. Interestingly Tj1 appeared to lose H3 occupancy but not H3K9me2, indicating that this element may have preferentially lost unmodified nucleosomes upon Dcr1 disruption.

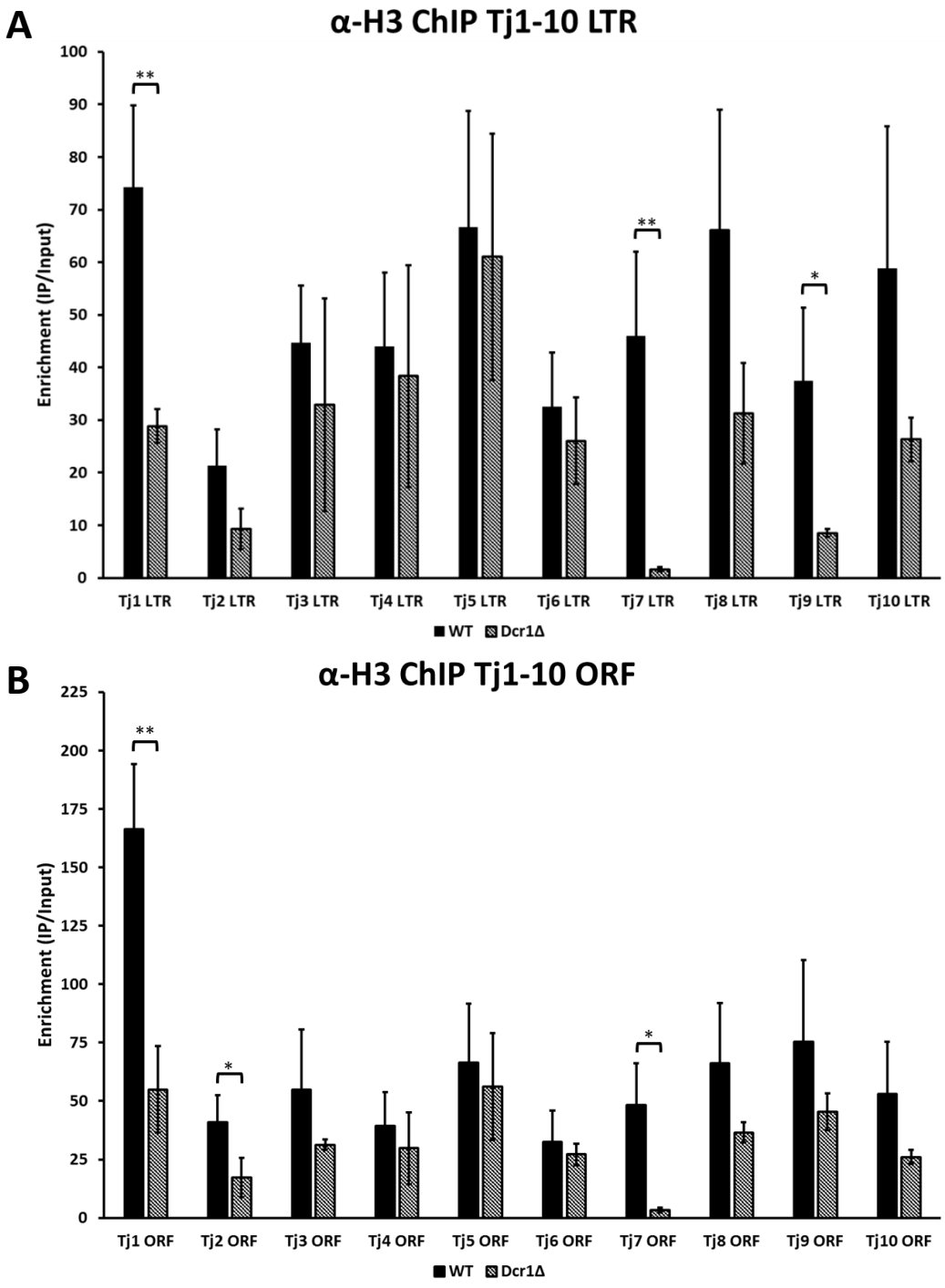


Figure 4.9 – Disruption of *dcr1*$^{+}$causes a decreases in Histone H3 occupancy at subset of retrotransposon elements

(A) α-H3 ChIP-qPCR assay to assess the enrichment of H3 at *S. japonicus* retrotransposon LTRs
(B) α-H3 ChIP-qPCR assay to assess the enrichment of H3 *S. japonicus* retrotransposon ORF sequences
Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

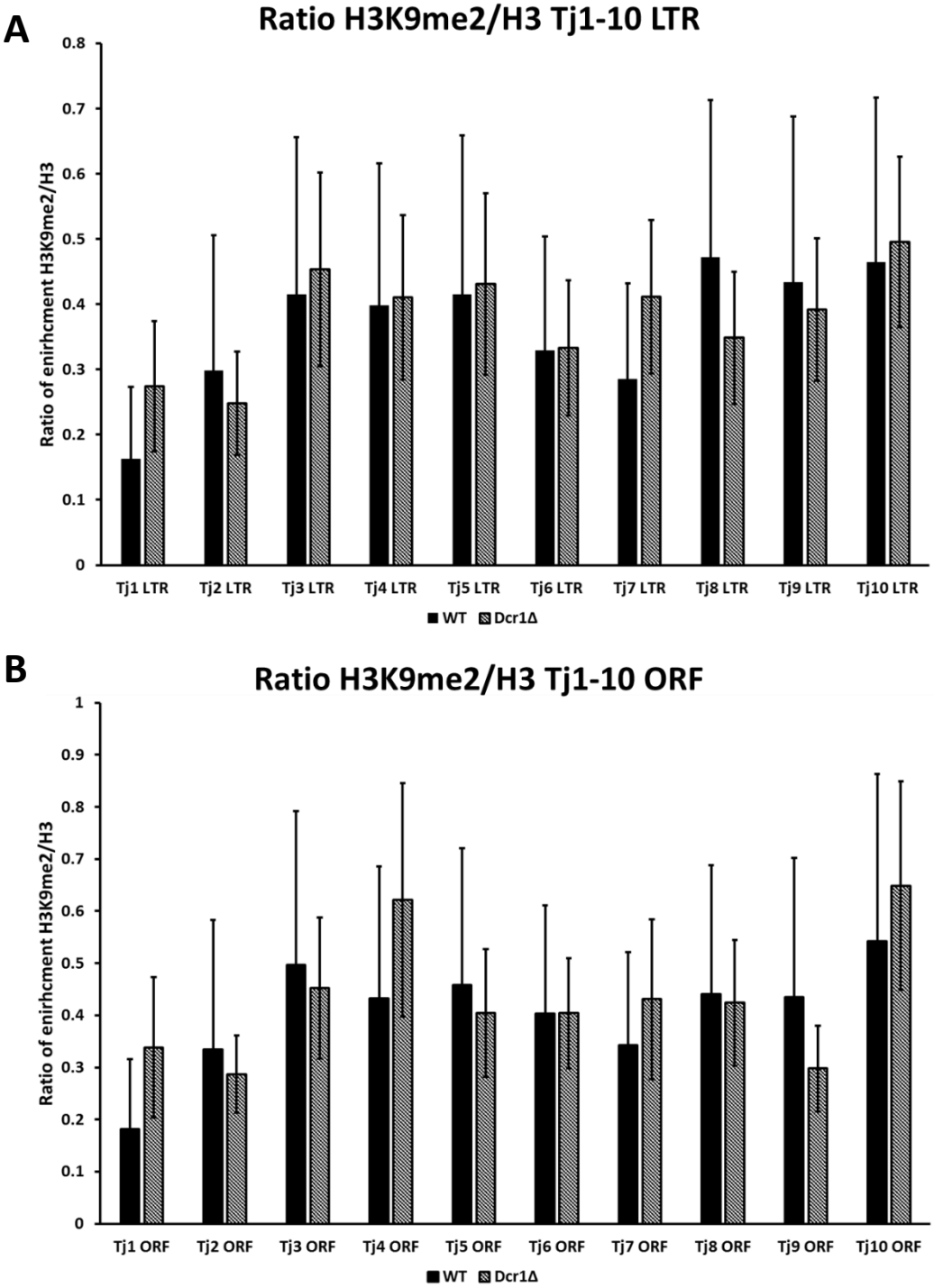


Figure 4.10 – H3 histones that remain at retrotransposons in the *dcr1⁺* mutant remain methylated to the same degree as in wild-type cells

(A) Ratio of H3K9me2/H3 enrichment at *S. japonicus* retrotransposon LTRs
(B) Ratio of H3K9me2/H3 enrichment at *S. japonicus* retrotransposon ORF sequences

Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

4.8 - Overall nucleosome occupancy is reduced at elements that lose H3K9me2

From the previous ChIP experiment, it was evident that at some specific retrotransposons histone H3 was depleted upon disruption of *dcr1*$^{+}$. I wanted to establish whether this was due to a loss of nucleosome density as a whole, or whether histone H3 specifically was being lost from the nucleosome and possibly replaced. To do this, an antibody against the invariant histone H4 was used for ChIP and immunoprecipitated chromatin was analysed by qPCR as before. This allowed me to assess whether histone H4 occupancy (and by extension nucleosome occupancy) changed over each element upon Dcr1 disruption. In general, for those elements that lost H3 occupancy, there was a corresponding decrease in H4 occupancy, indicating that nucleosome density was reduced at these loci (Figure 4.11). One exception to this was Tj2, which lost both H3K9me2 and H3 occupancy, but did not exhibit a corresponding loss of H4. It was also evident that for some elements, namely Tj7 and the LTR of Tj9, the reduction in H4 occupancy was not great enough to account for the extent of H3K9me2 and H3 loss by a reduction in total nucleosome density.

Assessing the ratio of H3 to H4 occupancy for each retrotransposon revealed that the composition of nucleosomes at the elements that showed a reduction in H3K9me2 was altered (Figure 4.12). For those that showed no change in methylation levels, the ratio of H3:H4 was unaltered upon Dcr1 deletion, however the elements that lost H3K9me2 (Tj2, Tj7 and Tj9) also seemed to lose H3 from the nucleosomes as well, indicating that the global loss of H3K9me2 at these loci might be due, at least in part, to the exchange of histone H3 with a non-methylatable H3 variant.

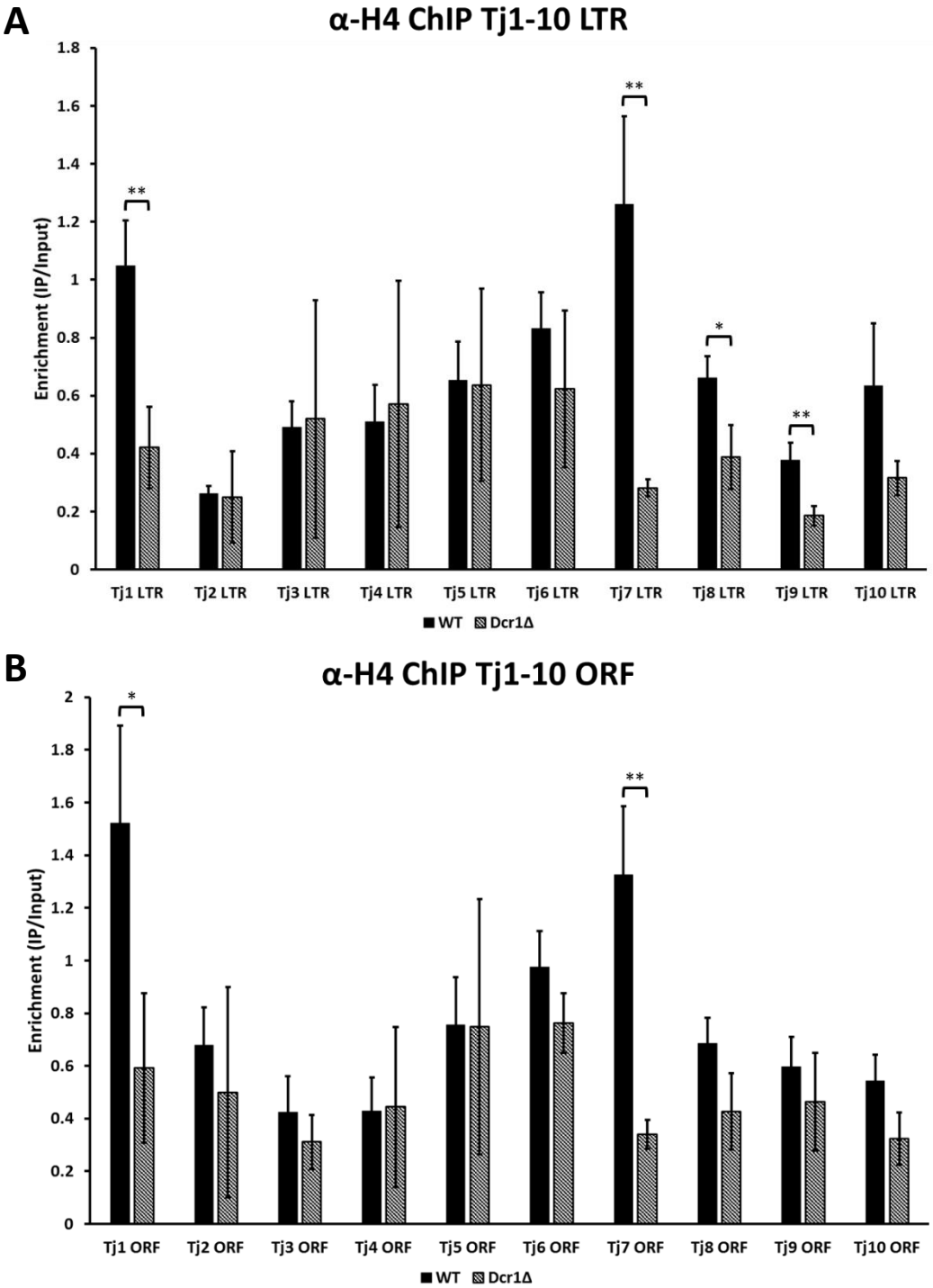


Figure 4.11 – Disruption of *dcr1*$^{+}$causes a decreases in Histone H4 occupancy at a subset of retrotransposon elements

(A) α-H4 ChIP-qPCR assay to assess the enrichment of H4 at *S. japonicus* retrotransposon LTRs
(B) α-H4 ChIP-qPCR assay to assess the enrichment of H4 at *S. japonicus* retrotransposon ORF sequences
Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

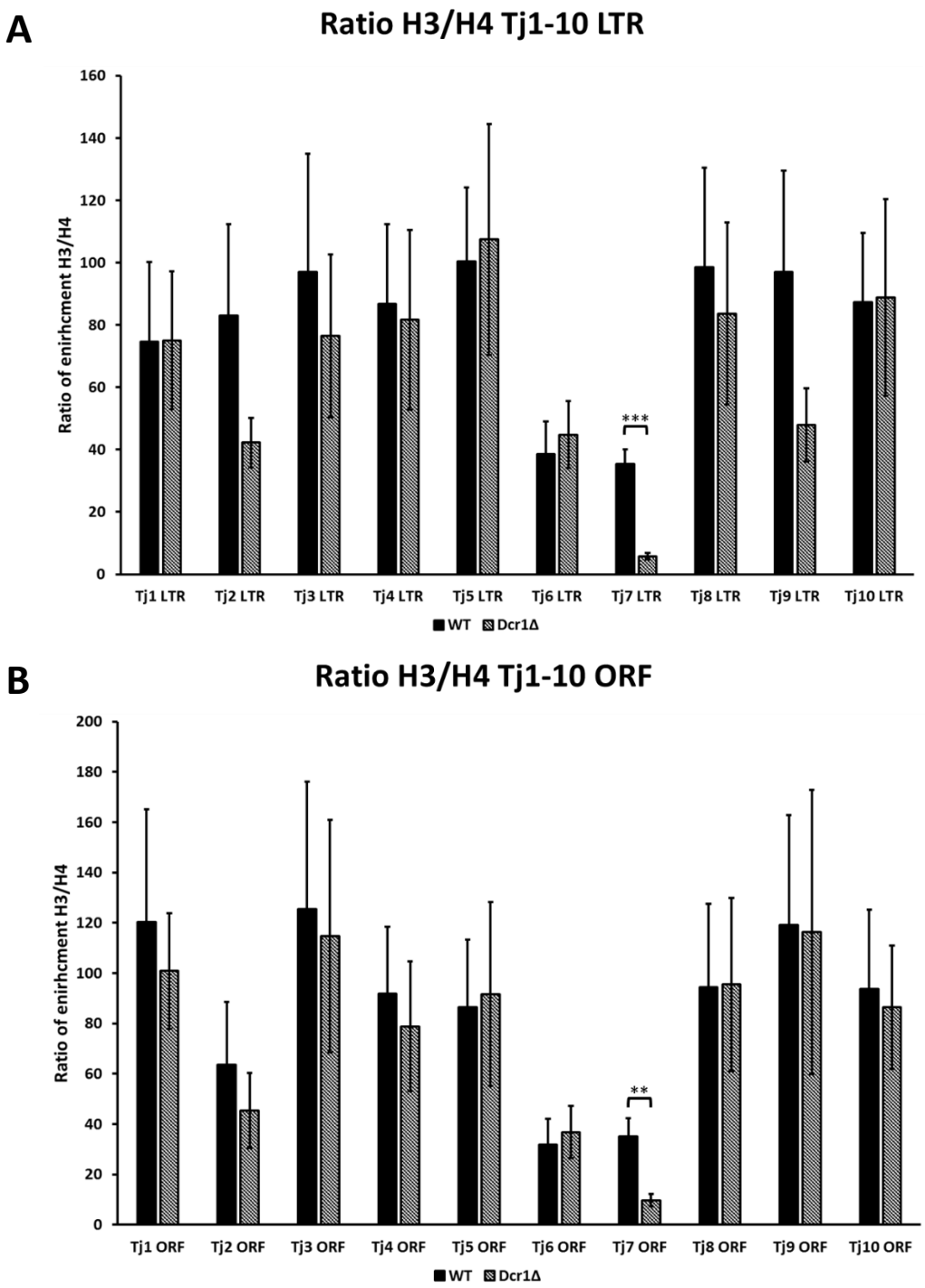


Figure 4.12 – Disruption of *dcr1*$^{+}$ does not significantly change nucleosome composition for any retrotransposon element except Tj7

(A) Ratio of H3/H4 enrichment at *S. japonicus* retrotransposon LTRs
(B) Ratio of H3/H4 enrichment at *S. japonicus* retrotransposon ORF sequences

Error bars represent 1 S.D. from 3 biological replicates.
* $p<0.05$, ** $p<0.01$, *** $p<0.005$

4.9 - Elements that lose H3K9me2 are loaded with CENP-A^{Cnp1}

Retrotransposons in *S. japonicus* define the pericentromeric region, however a conserved centromeric 'central core' sequence has yet to be identified. In the absence of this sequence, it is possible that retrotransposons may fulfil this role in *S. japonicus*; it is therefore also possible that some may be loaded with the H3 variant Cnp1, the fission yeast homologue of the human CENP-A. This histone variant is responsible for epigenetically defining the position of the centromere and mediating formation of the kinetochore, to allow for faithful chromosome segregation (Chen et al., 2003; Folco et al., 2008; Mendiburo et al., 2011; Takahashi et al., 2000). An anti-Cnp1 antibody was used for ChIP to assess whether the loss of H3 from nucleosomes at certain retrotransposons was due to a corresponding loading of the CENP-A^{Cnp1} variant.

For all elements except Tj6, regardless of their methylation status, there was an increase in CENP-A^{Cnp1} occupancy (Figure 4.13). The most striking increases occurred at those elements that lost H3K9me2 to the largest degree, Tj2 LTR, Tj7 and Tj9 LTR. For Tj2, CENP-A^{Cnp1} loading was increased around 3.3 fold, whereas the increase at Tj7 was around 6.7 fold. The largest increase in CENP-A^{Cnp1} occupancy occurred at Tj9 LTRs, which increased 116 fold over wild type. Tj6 was the only element to lose CENP-A^{Cnp1} upon Dcr1 deletion, with a decrease of around 2 fold observed. Interestingly, Tj6 and Tj7 appeared to be enriched for CENP-A^{Cnp1} in the wild type (Figure 4.14A), indicating that these two elements may function as the 'central core' sequences in *S. japonicus*, epigenetically defining the position of the centromere.

As 6 out of 10 retrotransposon families were increasingly enriched for CENP-A^{Cnp1} in a *dcr1*$^{+}$ disruption background, I wanted to investigate whether Dcr1

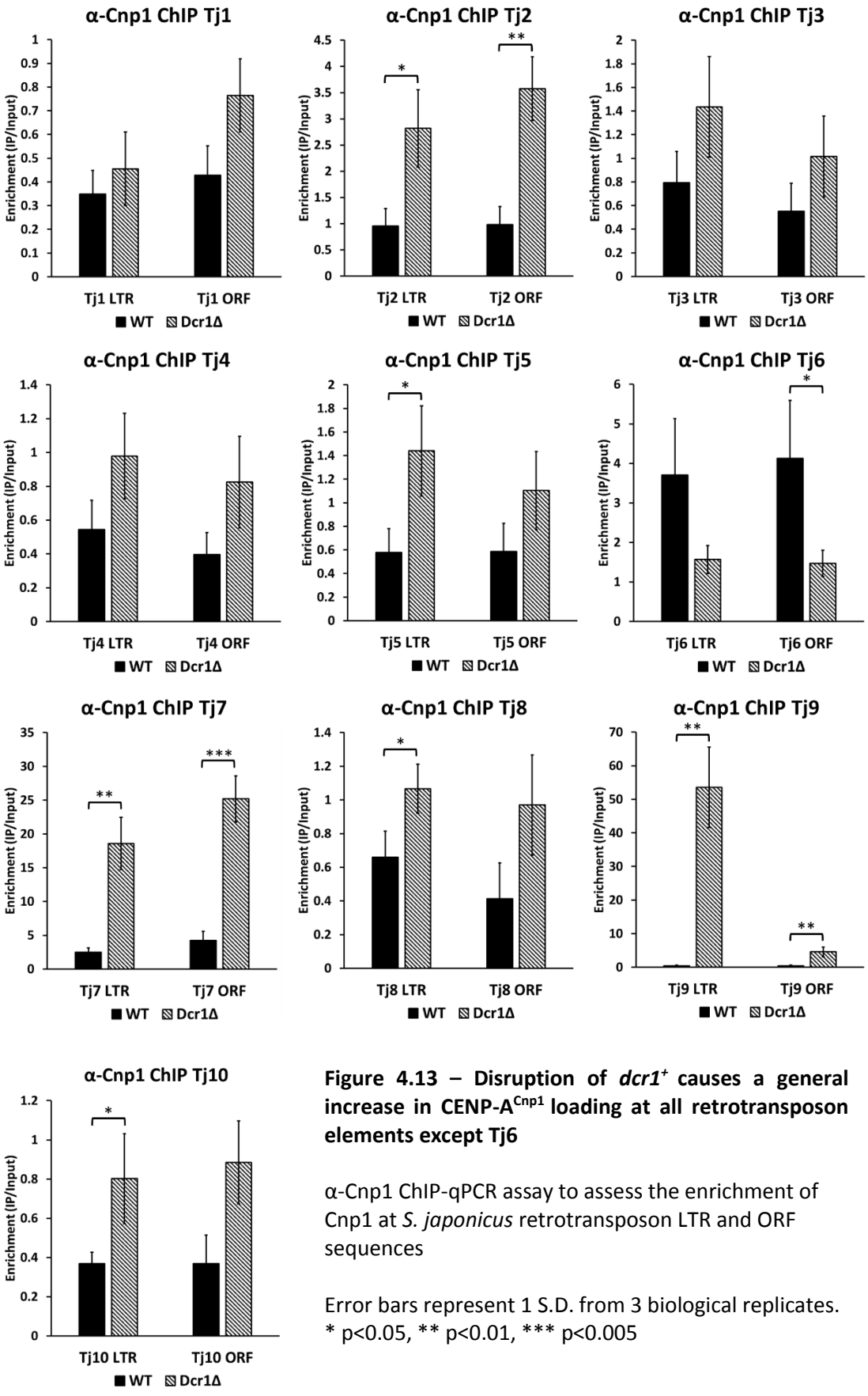


Figure 4.13 – Disruption of *dcr1*$^{+}$ causes a general increase in CENP-A^{Cnp1} loading at all retrotransposon elements except Tj6

α-Cnp1 ChIP-qPCR assay to assess the enrichment of Cnp1 at *S. japonicus* retrotransposon LTR and ORF sequences

Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

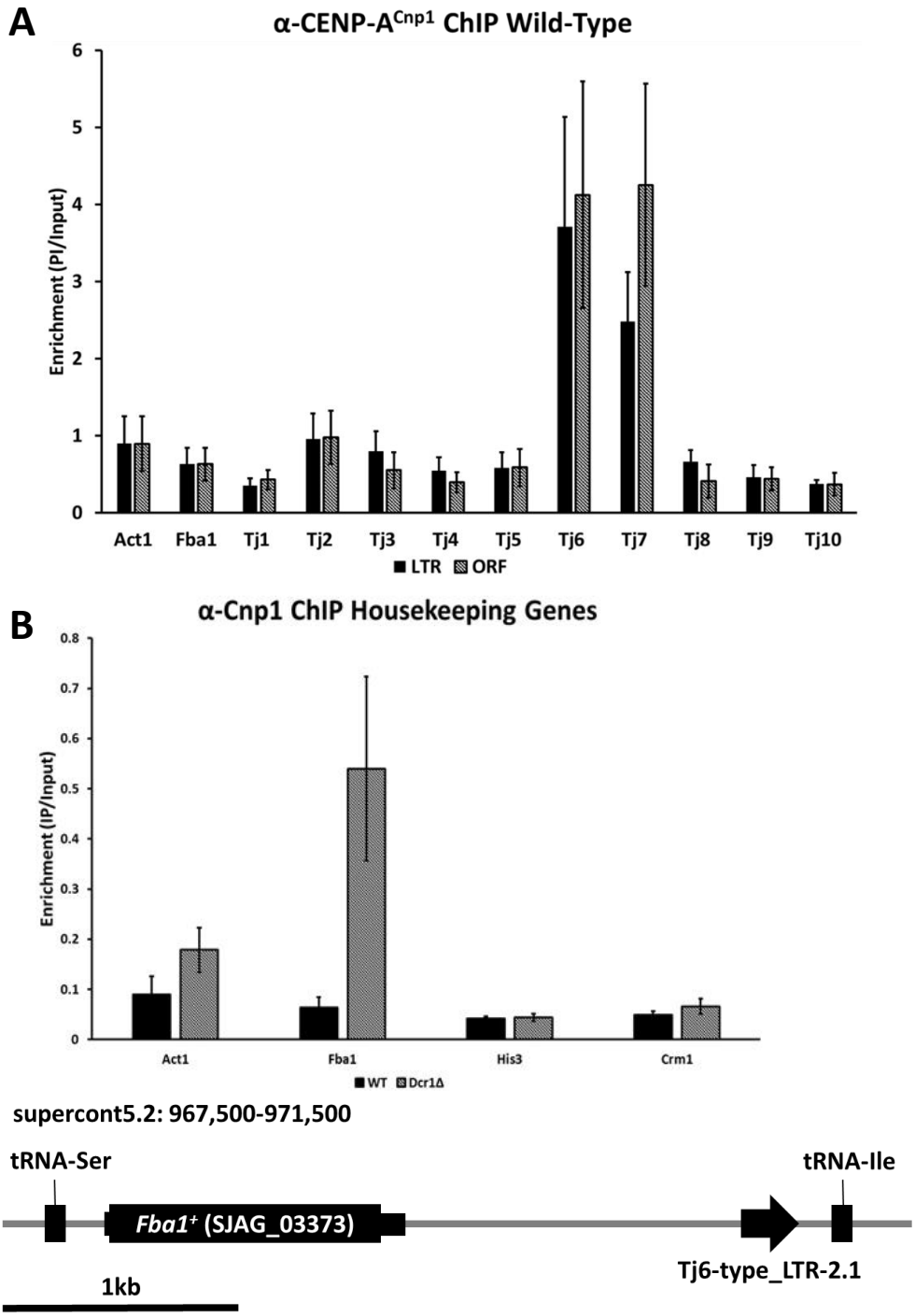


Figure 4.14 – Disruption of *dcr1*$^{+}$causes CENP-A^{Cnp1} loading at a centromere distal location, adjacent to an element that exhibits CENP-A^{Cnp1} enrichment in wild type

(A) α-Cnp1 ChIP-qPCR assay to assess the enrichment of Cnp1 at *S. japonicus* retrotransposon in the wild-type situation
(B) α-Cnp1 ChIP-qPCR assay to assess the enrichment of Cnp1 at four *S. japonicus* housekeeping genes. Shown below is a schematic representation of the *fba1*$^{+}$ locus

Error bars represent 1 S.D. from 3 biological replicates.

was responsible for restricting CENP-A^{Cnp1} localisation genome wide, or whether its effect was limited to retrotransposon loci. To do this, I combined ChIP of CENP-A^{Cnp1} with qPCR primers specific for four different euchromatic genes; *act1*$^+$ (SJAG_03145), *fba1*$^+$ (SJAG_03373/SJAG_01310), *his3*$^+$ (SJAG_03055) and *crm1*$^+$ (SJAG_01358). There was no significant enrichment of CENP-A^{Cnp1} at *act1*$^+$, *his3*$^+$ or *crm1*$^+$, however at *fba1*$^+$ (which has two copies with ~98% identity) there was an increase in CENP-A^{Cnp1} occupancy of around 8.5-fold (Figure 4.14B). Upon closer inspection of the *fba1*$^+$ loci it was noted that one copy, SJAG_03373, sits around 1.5kb away from a solo Tj6 LTR, which may explain the increase in CENP-A^{Cnp1} loading upon Dcr1 deletion. This revealed that upon disruption of *dcr1*$^+$, CENP-A^{Cnp1} may get redistributed into the chromosome arms, away from centromeres.

4.10 - Retrotransposon transcript accumulates upon *dcr1*$^+$ disruption

As a number of elements appeared to lose H3, and therefore subsequently H3K9me2 upon deletion of *dcr1*$^+$, the next logical step was to assess whether these elements lost silencing and were transcribed at higher levels. To assess the impact of *dcr1*$^+$ disruption on global retrotransposon transcript levels, qRT-PCR was carried out using the same retrotransposon specific qPCR primers used for ChIP-qPCR. mRNA levels were normalised to the housekeeping genes *act1*$^+$/*fba1*$^+$/*his3*$^+$ (for clarity only *act1*$^+$ is shown, but normalisation to all housekeeping genes gave comparable results).

A majority of retrotransposons showed a statistically significant increase in retrotransposon transcript levels, regardless of the changes in histone

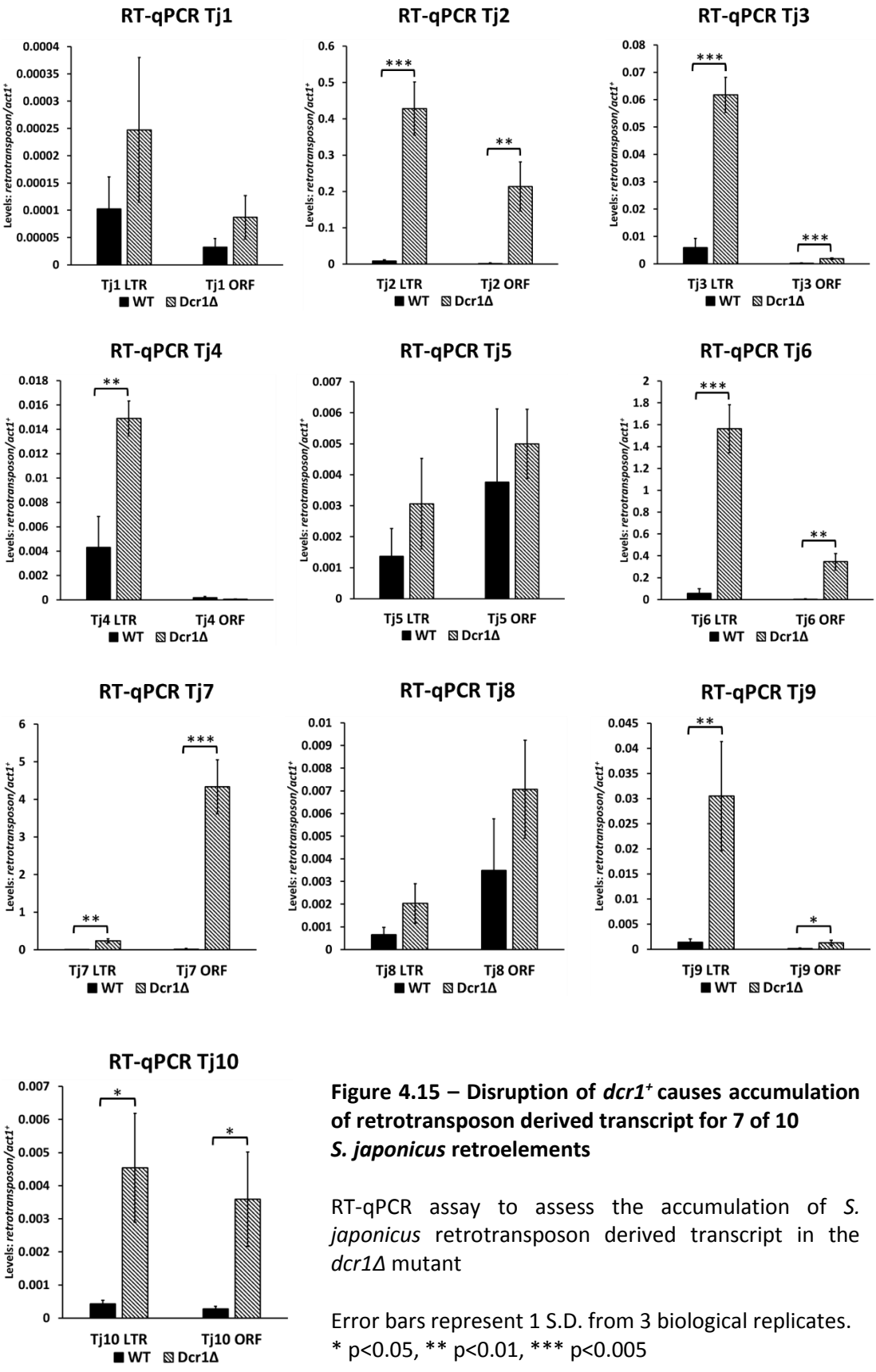


Figure 4.15 – Disruption of *dcr1⁺* causes accumulation of retrotransposon derived transcript for 7 of 10 *S. japonicus* retroelements

RT-qPCR assay to assess the accumulation of *S. japonicus* retrotransposon derived transcript in the *dcr1Δ* mutant

Error bars represent 1 S.D. from 3 biological replicates.
* p<0.05, ** p<0.01, *** p<0.005

modifications upon *dcr1*$^{+}$ disruption (Figure 4.15). Tj1, Tj4, Tj5 and Tj8 did not appear to accumulate transcript to levels above that of the wild type, however significant increases were evident for Tj2, Tj3, Tj6, Tj7, Tj9 and Tj10. The smallest increases were evident for Tj3 and Tj10, which accumulated around 10-fold more transcript than in the wild type, whilst Tj9 increased 10-20 fold. The largest increases in transcript were evident for Tj2, Tj6 and Tj7, where levels were up to 100-fold greater than wild type for Tj2 and Tj6, and up to 550-fold greater than wild type for Tj7. Of the elements that displayed elevated transcript accumulation, Tj2, Tj7 and Tj9 showed the greatest decrease in H3K9me2 levels and the greatest increase in CENP-A^{Cnp1} loading, indicating that these retrotransposons may be subject to the strongest Dcr1-dependent regulation, both in terms of the modification of associated histones and in regulation of transcript accumulation.

RT-qPCR gives an indication of global changes, however due to the repetitive nature of retrotransposons, conclusions about the accumulation of specific element transcripts are not always possible. To address this, RNA-Seq was performed on polyA-enriched RNA from wild type and *dcr1Δ5'* cells. Analysis of the RNA-Seq data revealed that centromeric and telomeric regions were generally transcriptionally upregulated in the absence of Dcr1 (Figure 4.16). Of the 266 identified *S. japonicus* retrotransposon elements (Rhind et al., 2011) (See Supplementary Table 1 for co-ordinates of annotated retrotransposon elements) 151 exhibited a statistically significant change in expression upon disruption of Dcr1, with all 151 of these elements transcriptionally upregulated. Included in these 151 transcriptionally upregulated elements were all 13 full length retrotransposons, 118 partial retrotransposons and 20 solo-LTRs

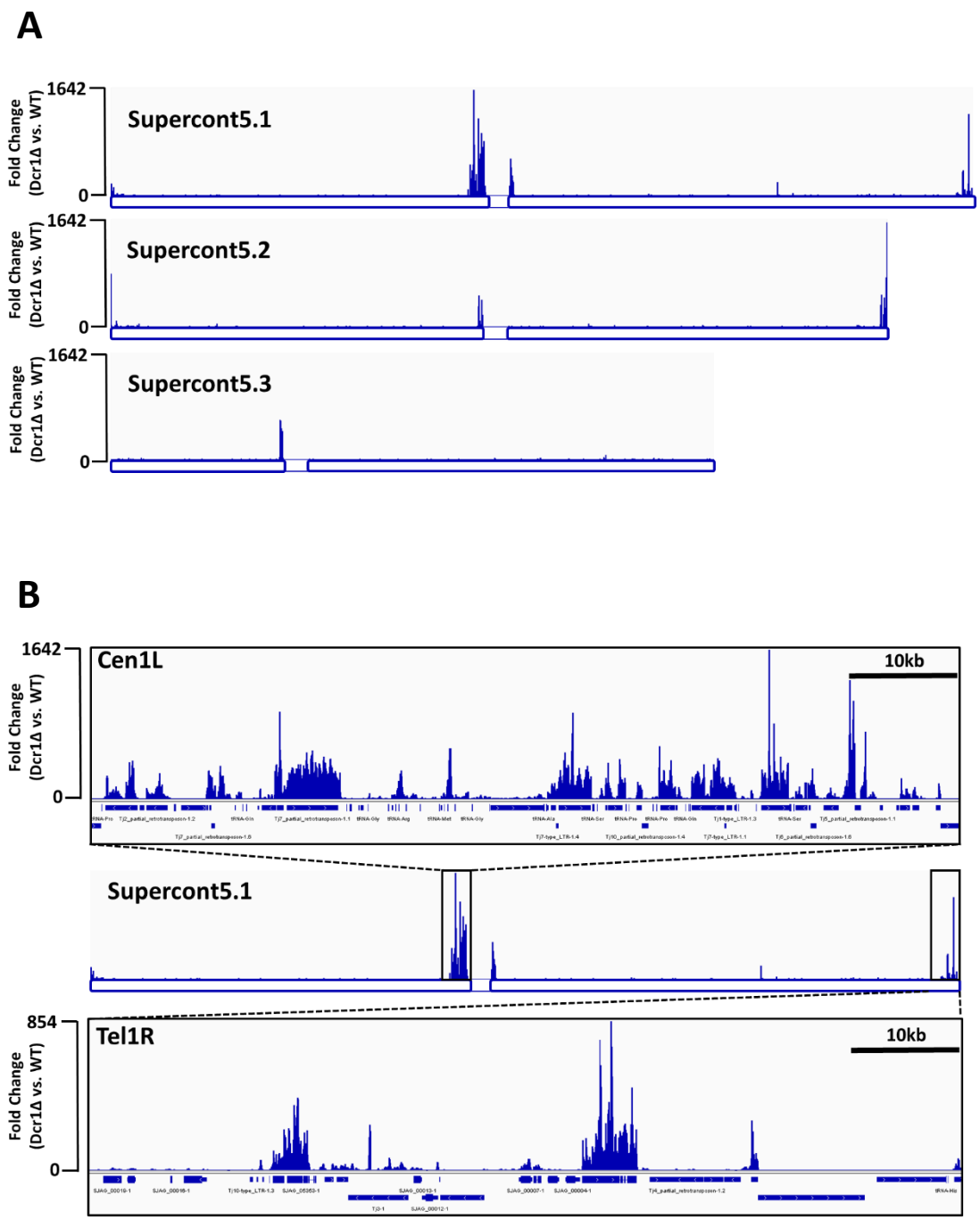


Figure 4.16 – Disruption of *dcr1⁺* causes transcriptional upregulation of centromeric and telomeric loci *in S. japonicus*

(A) Diagram of the three chromosomal *S. japonicus* Supercontigs, showing the fold change in transcript levels upon disruption of *dcr1⁺*

(B) Zoomed in view of Cen1L and Tel1R, showing fold change in transcript accumulation upon disruption of *dcr1⁺*. Blue boxes underneath represent annotated retrotransposons and subtelomeric genes.

(Supplementary Table 6). Elevated transcript levels ranged from 1.5-fold greater than wild type (Tj5-type_LTR-1.3) up to 194-fold greater than wild type (Tj3_partial_retrotransposon-1.1). For each element family, the percentage of elements that showed a significant increase in transcript was calculated. Tj1, Tj3, Tj4, Tj5 and Tj10 had relatively few transcriptionally upregulated elements (23-45%), whilst the majority (60-84%) of Tj2, Tj6, Tj7, Tj8 and Tj9 elements exhibited an increase (Table 4.1). This pattern roughly correlated with the increases seen by qRT-PCR, where Tj2, Tj6 and Tj7 displayed the greatest increase in retrotransposon transcript. There was a slight position effect associated with which retrotransposons accumulated transcript, with 82.2% of centromere associated elements and 36.9% of telomeric elements significantly upregulated upon deletion of Dcr1 (Table 4.1B). This does not factor in the 71 upregulated elements that are as-yet unassociated with a specific genomic region, due to the incompleteness of the genome assembly. As 120 of the 128 retrotransposon elements currently associated with one of the three main chromosomal contigs (supercont5.1, 5.2 and 5.3) are located at the annotated centromeres and telomeres, it is logical to believe that a majority of the 138 unmapped elements would also map to centromeric or telomeric loci, thus in reality the number of centromeric and telomeric retrotransposons that accumulate transcript upon Dcr1 disruption is expected to be greater than the number currently annotated. The greater transcriptional upregulation of centromeres compared to telomeres indicated that although Dcr1 plays a role in silencing both loci, telomeres may possess additional pathways to maintain genomic integrity, as is evident in *S. pombe* (Hansen et al., 2006; Kanoh et al., 2005).

A

Element	Full Length		Partial		Solo-LTR		Total	
	Number upregulated	% upregulated	Number upregulated	% upregulated	Number upregulated	% upregulated	Number upregulated	% upregulated
Tj1	N/A	N/A	1/2	50.0%	3/7	42.8%	4/9	44.4%
Tj2	1/1	100.0%	7/8	87.5%	3/9	33.3%	11/18	61.1%
Tj3	3/3	100.0%	13/30	43.3%	0/2	0.0%	16/35	45.7%
Tj4	1/1	100.0%	1/8	12.5%	1/4	25.0%	3/13	23.1%
Tj5	2/2	100.0%	9/23	39.1%	1/5	20.0%	12/30	40.0%
Tj6	2/2	100.0%	19/20	95.0%	0/3	0.0%	21/25	84.0%
Tj7	N/A	N/A	12/14	85.7%	4/8	50.0%	16/22	72.7%
Tj8	N/A	N/A	14/22	63.6%	1/3	33.3%	15/25	60.0%
Tj9	3/3	100.0%	30/38	78.9%	6/10	60.0%	39/51	76.5%
Tj10	1/1	100.0%	12/23	52.2%	1/12	8.3%	14/36	38.8%

B

Locus	Number of Retrotransposon Elements	Number of Retrotransposon Elements Upregulated by RNA-Seq	% of Elements Upregulated
Tel1L	9	4	44.4
Cen1L	40	28	70.0
Cen1R	12	12	100.0
Tel1R	12	5	41.7
Tel2L	7	2	28.6
Cen2L	10	7	70.0
Cen2R	0	N/A	N/A
Tel2R	20	14	70.0
Tel3L	0	N/A	N/A
Cen3L	9	8	88.9
Cen3R	0	N/A	N/A
Tel3R	1	0	0.00
Chromosome Arm	8	0	0.00
Unplaced	138	71	51.5

Table 4.1 – Disruption of *dcr1*$^+$ causes transcriptional upregulation of all *S. japonicus* retrotransposon families, regardless of whether these elements are centromeric or telomeric

(A) Table showing the proportion of each kind of retrotransposon element upregulated upon disruption of *dcr1*$^+$ for each individual retrotransposon family in *S. japonicus*
(B) Table showing the number of retrotransposon elements associated with each centromere or telomere and the percentage of these upregulated in the absence of *dcr1*$^+$

Combining this RNA-Seq dataset with the small RNA-Seq dataset generated previously, allowed me to correlate those elements that exhibited altered transcript accumulation and those that exhibited altered small RNA levels. The retrotransposable elements fell into one of two categories. 13 elements seemed to gain small RNAs in the absence of Dcr1 and also accumulated transcript (Supplementary Table 8), indicating that this increase in small RNA complement was due to Dcr1-independent processing of an increased pool of full length transcript into shorter species. 105 elements lost siRNAs without Dcr1 and saw a subsequent increase in transcript levels (Supplementary Table 9). These elements accumulated transcript that was not processed in the absence of Dcr1, indicating that in a wild-type situation Dicer was responsible for processing this transcript into small RNA, however this transcript was not processed in a Dcr1-independent manner in the dcr1Δ5' strain.

No retroelements exhibited a decrease in transcript levels (Supplementary Table 7), thus it is reasonable to conclude that deletion of Dcr1 globally deregulated transcriptional silencing of retrotransposons, regardless of what subsequently happened to the full length transcript. This pattern did not apply to all elements, as there were a proportion that exhibited no change in transcript accumulation in the absence of Dcr1. However, these elements appear to represent those that are not able to be transcribed, either due to mutation within, or the absence of a promoter proximal to the element, and are thus not targets of Dcr1.

4.11 – Disruption of *dcr1*$^+$ alters transcript levels of non-retrotransposon genes

Although a majority of the most upregulated genes were retrotransposons (Figure 4.17A), disruption of *dcr1*$^+$ caused a number of other transcriptional changes, with 1225 genes (excluding retrotransposons) significantly upregulated and 1307 significantly downregulated ($p<0.05$). Introduction of a cut-off for Log2(FC) values of +/- 0.6, which corresponded to an increase/decrease in transcript of 1.5-fold, decreased the number of genes classified as having altered expression to 385 upregulated (Supplementary Table 6) and 364 downregulated (Supplementary Table 7). Utilising an even more stringent cut-off of +/- 1.0, corresponding to transcript levels that double or halve in the absence of Dcr1, reduced this further to 178 upregulated and 114 downregulated genes (Figure 4.17B).

Of the 385 genes upregulated >1.5-fold, 128 were classified as uncharacterised 'hypothetical proteins', which have no associated name or description within Ensembl. Within these hypothetical genes were 3 out of the 6 annotated Tlh1/2 telomeric helicase homologues in *S. japonicus* (SJAG_05173, SJAG_05353, and SJAG_00013). Of these, SJAG_05173 is the most upregulated non-retrotransposon element upon Dcr1 disruption, indicating that other telomere associated elements aside from retrotransposons were transcriptionally upregulated in the absence of Dcr1.

I used BioMart (Smedley et al., 2009) to look for conserved protein domains within the rest of the 'hypothetical proteins', and of 128 upregulated genes, 45 had conserved domains (Supplementary Table 6). The most abundant domains were Zinc fingers, which were present in 9 of the genes, 5 of these

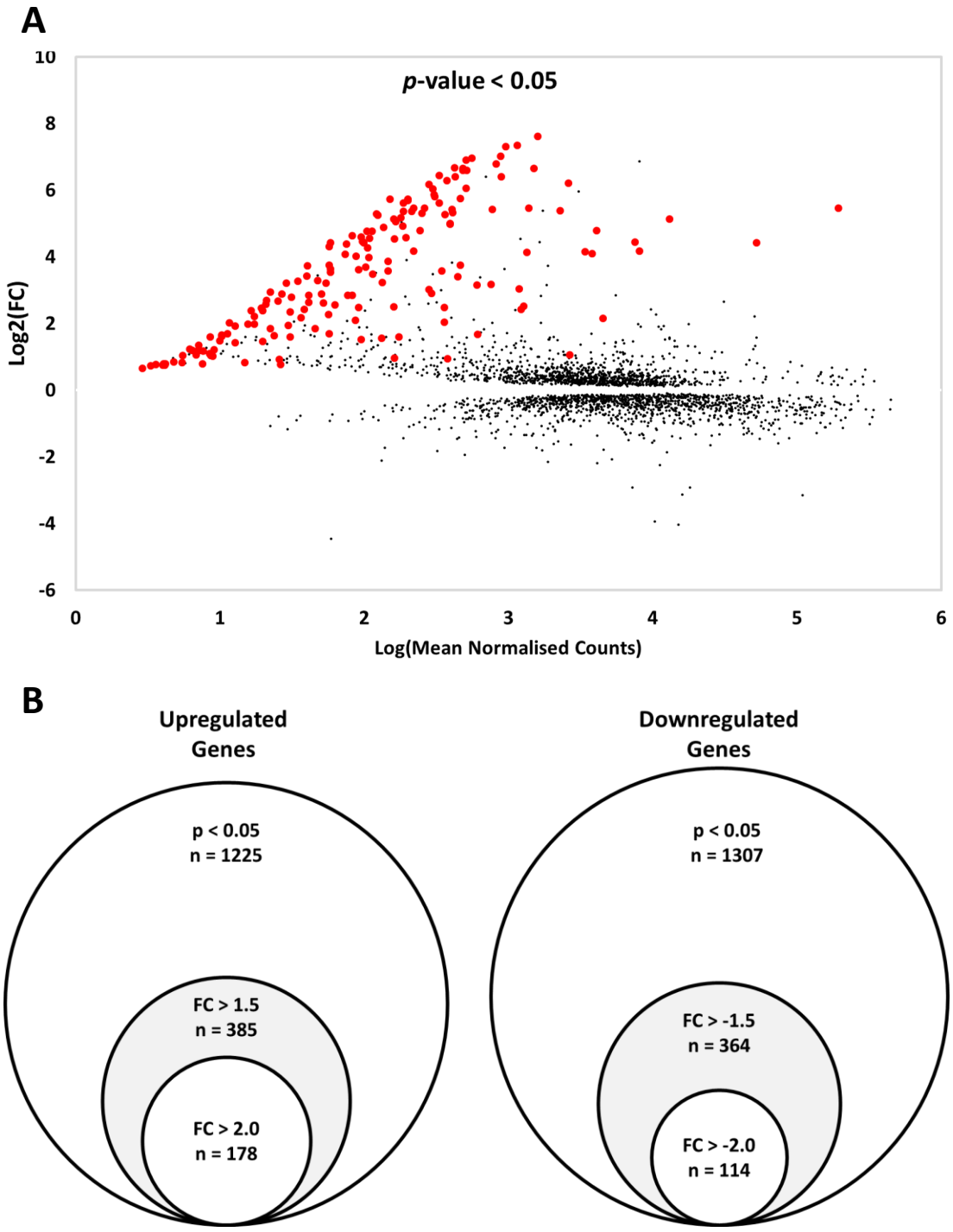


Figure 4.17 – Retrotransposons make up a majority of the most upregulated elements in the absence of *dcr1*$^{+}$

(A) Graph showing Log2(FC) vs. Log(Mean Normalised Counts) for all genes and retrotransposons that showed a statistically significant (p<0.05) change in transcript level upon disruption of *dcr1*$^{+}$. Highlighted in red are annotated retrotransposons.
(B) Venn-diagrams showing the total number of genes significantly altered (p<0.05) in the absence of *dcr1*$^{+}$, as well as the number of those genes that exhibited changes greater than 1.5 or 2.0-fold.

were C2H2-type, whilst 4 were CCHC-type. 8 proteins contained RTA1 domains; the RTA1 family are fungal proteins with multiple transmemembrane regions. There were also a number of proteins that contained retrotransposons associated domains, such as aspartic proteases, gag proteins and chromodomains. The prevalence of specific protein domains within these upregulated genes indicated that there may be conservation of the genes families transcriptionally regulated by Dcr1.

In order to assess the impact of *dcr1*$^{+}$ disruption upon the expression of generalised pathways, g:Profiler (Reimand et al., 2016) was used to search for enriched GO terms within a given list of genes, in this case those transcriptionally up or downregulated 1.5-fold upon *dcr1*$^{+}$ disruption. As extensive analysis of gene function in *S. japonicus* was not available, GO terms assigned to *S. pombe* homologues were used. For those genes upregulated 1.5-fold, there were no enriched GO terms, however the list of genes downregulated 1.5-fold contained 83 significantly enriched terms (Supplementary Table 10). All enriched GO terms related to various aspects of metabolism, with the most significantly enriched terms 'oxoacid metabolic process' (GO:0043436, $p = 2.26 \times 10^{-16}$), 'organic acid metabolic process' (GO:0006082, $p = 2.85 \times 10^{-16}$), 'carboxylic acid metabolic process' (GO:0019752, $p = 4.44 \times 10^{-16}$) and 'glycolytic process' (GO:0006096, $p = 2.70 \times 10^{-15}$). It is evident that disruption of *dcr1*$^{+}$ caused transcriptional reprogramming to alter cellular metabolism, however whether this is due to a direct role played by Dcr1 in regulating metabolism or whether it is an indirect product of the altered growth state exhibited when functional Dcr1 is absent remains unclear.

As isolation of the *dcr1Δ5'* mutant was a rare event, it was possible that a change in transcriptional pattern occurred to allow this particular mutant isolate to survive, via transcriptional up/downregulation of one or more factors that could compensate for the loss of the Dcr1 protein. In order to assess whether this *dcr1Δ5'* strain exhibited altered transcriptional regulation of other functionally related genes, the list of genes up/downregulated by 1.5-fold were searched for occurrence of particular GO terms. GO terms were selected for those that contained the keywords, 'siRNA', 'Small RNA', 'Silencing', 'Centromere', or 'Transposon' (see Supplementary Table 11 for full list of terms). From this analysis a number of genes were found to be transcriptionally altered, 39 genes were upregulated, whilst 30 genes were downregulated (Table 4.2). Of the 39 upregulated genes, only six (tRNA isopentenyltransferase, ww domain binding protein 11, YEATS family histone acetyltransferase Yaf9, SNF2 family helicase Ino80, ubiquitin-like protein modifier SUMO Pmt3 and TATA-binding protein) exhibited greater than a 1.5-fold increase in transcript, and one (tRNA isopentenyltransferase) showed a ~2-fold increase in transcript. Of the downregulated genes, six out of 30 (STE/STE7/MEK1 protein kinase Byr1, ubiquitin conjugating enzyme Ubc7/UbcP3, Clr6 histone deacetylase complex subunit Pst1, Rad6, repressor-RNA polymerase III Maf1, and argonaute binding protein Arb1) exhibited greater than a 1.5-fold decrease in transcript, and none showed more than a 2-fold decrease. The modest nature of these changes, coupled with the lack of obvious mechanism by which these changes could suppress the loss of Dcr1, makes inferring the biological significance difficult. Although a number of RNAi related genes were altered in the absence of Dcr1, only the level of *arb1*$^{+}$

GeneID	Gene description	log2(FC)
SJAG_01575	tRNA isopentenyltransferase	1.00219101
SJAG_00363	ww domain binding protein 11	0.87647346
SJAG_04097	YEATS family histone acetyltransferase subunit Yaf9	0.82500855
SJAG_02872	SNF2 family helicase Ino80	0.76194589
SJAG_03843	ubiquitin-like protein modifier SUMO Pmt3	0.73965468
SJAG_00421	TATA-binding protein	0.66005905
SJAG_02697	ubiquitinated histone-like protein Uhp1	0.55225606
SJAG_02409	LIM-like protein linking chromatin modification to RNAi Stc1	0.54985196
SJAG_01290	cell wall protein Gas1	0.51921146
SJAG_03325	COP9/signalosome complex subunit Csn2	0.46431033
SJAG_02768	Set1C complex subunit Swd1	0.4458687
SJAG_00327	DNA replication pre-initiation complex subunit Cdc45	0.43037726
SJAG_03081	heterotrimeric G protein beta subunit Git5	0.42841897
SJAG_01075	Sir2 family histone deacetylase Hst2	0.40702401
SJAG_03047	histone H2A variant H2A.Z	0.39640874
SJAG_00913	hira protein	0.38285893
SJAG_00300	ubiquitin conjugating enzyme Ubc15	0.37025781
SJAG_01579	chromodomain protein 2 Chp2	0.36230624
SJAG_03844	histone deacetylase Clr3	0.35333116
SJAG_03685	telomere maintenance protein Ccq1	0.33599435
SJAG_03160	CDC7 protein kinase Hsk1	0.32773243
SJAG_02905	ubiquitin family	0.32409728
SJAG_02119	Rik1-associated factor Raf2	0.31390786
SJAG_01301	Sm snRNP core protein Smd3	0.30887112
SJAG_00618	DNA polymerase alpha accessory factor Mcl1	0.30177314
SJAG_02074	MCM loader	0.28369467
SJAG_03664	jmjc domain chromatin associated protein Epe1	0.28292521
SJAG_04574	histone H3 methyltransferase Clr4	0.28232318
SJAG_03689	hypothetical protein	0.27300255
SJAG_03006	ubiquitin-protein ligase E3 Brl1	0.27032845
SJAG_01745	LEM domain-containing protein Heh1/Lem2	0.23274315
SJAG_03128	FACT complex component Pob3	0.21900538
SJAG_01204	RTT109 family histone lysine acetyltransferase	0.20633294
SJAG_00676	histone acetyltransferase Hat1	0.20581742
SJAG_00184	SAGA complex subunit Sgf29	0.17658979
SJAG_00188	ubiquitin-protein ligase E3	0.17103028
SJAG_04062	WD repeat protein Crb3	0.16856055
SJAG_03694	actin-like protein Arp6	0.16705452
SJAG_03454	damaged DNA binding protein Ddb1	0.16310205

GeneID	Gene description	log2(FC)
SJAG_00993	STE/STE7/MEK1 protein kinase Byr1	-0.870584206
SJAG_04612	ubiquitin conjugating enzyme Ubc7/UbcP3	-0.756668599
SJAG_02846	Clr6 histone deacetylase complex subunit Pst1	-0.69093006
SJAG_00293	Rad6	-0.671423236
SJAG_02898	repressor-RNA polymerase III Maf1	-0.641644468
SJAG_02594	argonaute binding protein 1 Arb1	-0.641414086
SJAG_01836	CAF1 family ribonuclease Tri1	-0.588419589
SJAG_01610	RNA polymerase II associated Paf1 complex	-0.524819873
SJAG_00457	Gtr1/RagA G protein Gtr1	-0.458659812
SJAG_00995	transcription factor TFIIH complex subunit Rad15	-0.420533561
SJAG_02800	origin recognition complex subunit Orc2	-0.413861006
SJAG_04581	kinetochore protein Mal2	-0.388118553
SJAG_04589	ATP-dependent RNA helicase Uap56	-0.378429453
SJAG_01718	STE/STE11 protein kinase Byr2	-0.374580913
SJAG_00448	RNA polymerase II transcription termination factor	-0.356273291
SJAG_03126	serine/threonine protein phosphatase PP1	-0.348647563
SJAG_04547	NASP family CENP-A chaperone (sim3)	-0.336290466
SJAG_01146	histone lysine methyltransferase Set9	-0.326527744
SJAG_03255	histone deacetylase complex subunit Cti6	-0.308759707
SJAG_02725	poly(A) polymerase Cid12	-0.303014547
SJAG_04253	RNA-silencing factor Ers1	-0.264231665
SJAG_00934	cryptic loci regulator Clr1	-0.251705701
SJAG_01653	ATP-dependent DNA helicase Hrp3	-0.250580046
SJAG_00160	hypothetical protein clr5	-0.246041809
SJAG_02209	transcription elongation factor complex subunit Iws1	-0.245944274
SJAG_01777	TRF Taz1	-0.231189249
SJAG_00581	histone deacetylase Clr6	-0.22189385
SJAG_00883	nucleolar protein Dnt1	-0.18756435
SJAG_04186	Sir2 family histone deacetylase Sir2	-0.169651445
SJAG_01601	chromodomain protein Swi6	-0.161647435

Table 4.2 – Genes involved in silencing, small RNA biogenesis and transposon regulation are transcriptionally altered in the absence of *dcr1⁺*

List of genes up or downregulated upon the disruption of *dcr1⁺* that had any of the associated GO terms listed in Supplementary Table 11

was downregulated by more than 1.5-fold, thus it is unlikely that transcriptional reprogramming compensates for the loss of Dcr1.

In *S. pombe*, a number of genetic deletions have been found to rescue the *dcr1*$^{+}$ null phenotype in a double-mutant background (Reddy et al., 2011; Reyes-Turcu et al., 2011; Tadeo et al., 2013; Wang et al., 2014). Overexpression of the chromodomain protein Swi6 has also been shown to have the same effect (Tadeo et al., 2013). To see if any of these genes were transcriptionally altered in the *S. japonicus dcr1Δ5'* mutant, the list of significantly up/downregulated genes ($p < 0.05$) was searched for the occurrence of these specific 15 genes (Table 4.3A), obtained from PomBase (Wood et al., 2012). Of these, 4 were found to be downregulated in the absence of Dcr1, whilst 4 were transcriptionally upregulated (Table 4.3B). However, as was the case above, these transcriptional changes were very modest, with no gene upregulated more than 1.5-fold, and no gene downregulated more than 1.5-fold. Thus, it is unlikely that transcriptional reprogramming of specific genes contributed to the survival of this specific *dcr1*$^{+}$ mutant.

4.12 - Retrotransposon transcript accumulation is not caused by increased transcription

As a number of retroelements accumulated transcript in a *dcr1Δ5'* mutant, it was important to differentiate whether this was due to an increase in the rate of transcription or whether it was due to a lack of transcript processing. To do this, ChIP-qPCR using a pan-specific antibody against the DNA-directed RNA polymerase II subunit Rpb1 was performed, to assess any changes in Pol II occupancy at retrotransposons when Dcr1 was disrupted. Somewhat

A

spGeneID	Gene Name	Gene Description	Evidence for Phenotypic Suppression
SPAC17G8.13c	mst2	histone acetyltransferase Mst2	Reddy et al., 2011
SPAC2F7.04	med1	mediator complex subunit Med1	Reddy et al., 2011
SPBC17D11.04c	nto1	histone acetyltransferase complex PHD finger subunit Nto1	Reddy et al., 2011
SPBC16G5.13	ptf2	Mst2 histone acetytransferase acytyltransferase complex subunit	Reddy et al., 2011
SPBC1D7.04	mlo3	RNA binding protein Mlo3	Reyes-Turcu et al., 2011
SPAC20H4.03c	tfs1	transcription elongation factor TFIIS	Reyes-Turcu et al., 2011
SPBC1778.02	rap1	telomere binding protein Rap1	Tadeo et al., 2013
SPAC16A10.07c	taz1	human TRF ortholog Taz1	Tadeo et al., 2013
SPAC19G12.13c	poz1	Pot1 associated protein Poz1	Tadeo et al., 2013
SPCC188.07	ccq1	telomere maintenance protein Ccq1	Tadeo et al., 2013
SPAC6F6.16c	tpz1	Tppl homolog Tpz1	Tadeo et al., 2013
SPAC26H5.06	pot1	telomere end-binding protein Pot1	Tadeo et al., 2013
SPAC31G5.18c	sde2	silencing defective protein Sde2	Tadeo et al., 2013
SPAC664.01c	swi6	HP1 family chromodomain protein Swi6	Tadeo et al., 2013
SPAC1D4.01	tls1	splicing factor Tls1	Wang et al., 2014

B

GeneID	log2(FC)	P-value	Gene Description
SJAG_01717	-0.418963212	7.87E-06	RNA binding protein Mlo3
SJAG_01378	-0.288193843	7.33E-03	Tppl homolog Tpz1
SJAG_01777	-0.231189249	2.92E-03	human TRF ortholog Taz1
SJAG_01601	-0.161647435	2.68E-02	chromodomain protein Swi6
SJAG_02691	0.115050619	4.42E-02	transcription elongation factor TFIIS
SJAG_02391	0.235992731	1.31E-03	mediator complex subunit Med1
SJAG_02905	0.324097279	1.76E-04	silencing defective protein Sde2
SJAG_03685	0.335994351	6.37E-06	telomere maintenance protein Ccq1

Table 4.3 - Factors known to phenotypically suppress *dcr1Δ* in *S. pombe* are modestly transcriptionally altered in *S. japonicus*

(A) List of genes whose deletion or overexpression is known to phenotypically suppress loss of *dcr1*$^{+}$ in *S. pombe*

(B) List of *S. japonicus* orthologues transcriptionally changed in a *dcr1Δ* background

surprisingly, there was no difference in Pol II occupancy at any retrotransposon in the absence of Dcr1 (Figure 4.18A&B). This indicated that the increase in retrotransposon transcript in the mutant may be due solely to a lack of processing by Dcr1, and not an increase in transcriptional rate of the elements.

There was a significant 5.4 fold decrease in Pol II occupancy at *fba1*$^{+}$ in a *dcr1Δ5'* mutant (Figure 4.18C). This may be related to the increased loading of CENP-A^{Cnp1} at this locus, causing transcriptional changes associated with this centromere specific histone variant.

4.13 – Specific retroelements appear to mobilise in the *dcr1Δ5'* strain

As a number of retrotransposons accumulated transcript in the *dcr1Δ5'* mutant, I wanted to determine whether this mRNA got reverse transcribed and re-integrated into the genome, completing the retrotransposon life cycle. To assess whether the de-regulation of retrotransposon transcripts in the absence of Dcr1 lead to a subsequent expansion of retrotransposon copy number, qPCR was performed on genomic DNA extracted from wild type and *dcr15'Δ* mutant cells, using primers specific for each retroelement. The levels of retrotransposon DNA were normalised to a single copy gene (*act1*$^{+}$) that should stay constant regardless of variation in retrotransposon copy number. From this analyses, it was clear that a number of retroelements actively mobilised upon disruption of Dcr1 (Figure 4.19). The most striking increase was in the complement of Tj7 retrotransposons, which showed around a 30-fold increase over wild type. Tj2 also appeared to mobilise, increasing around

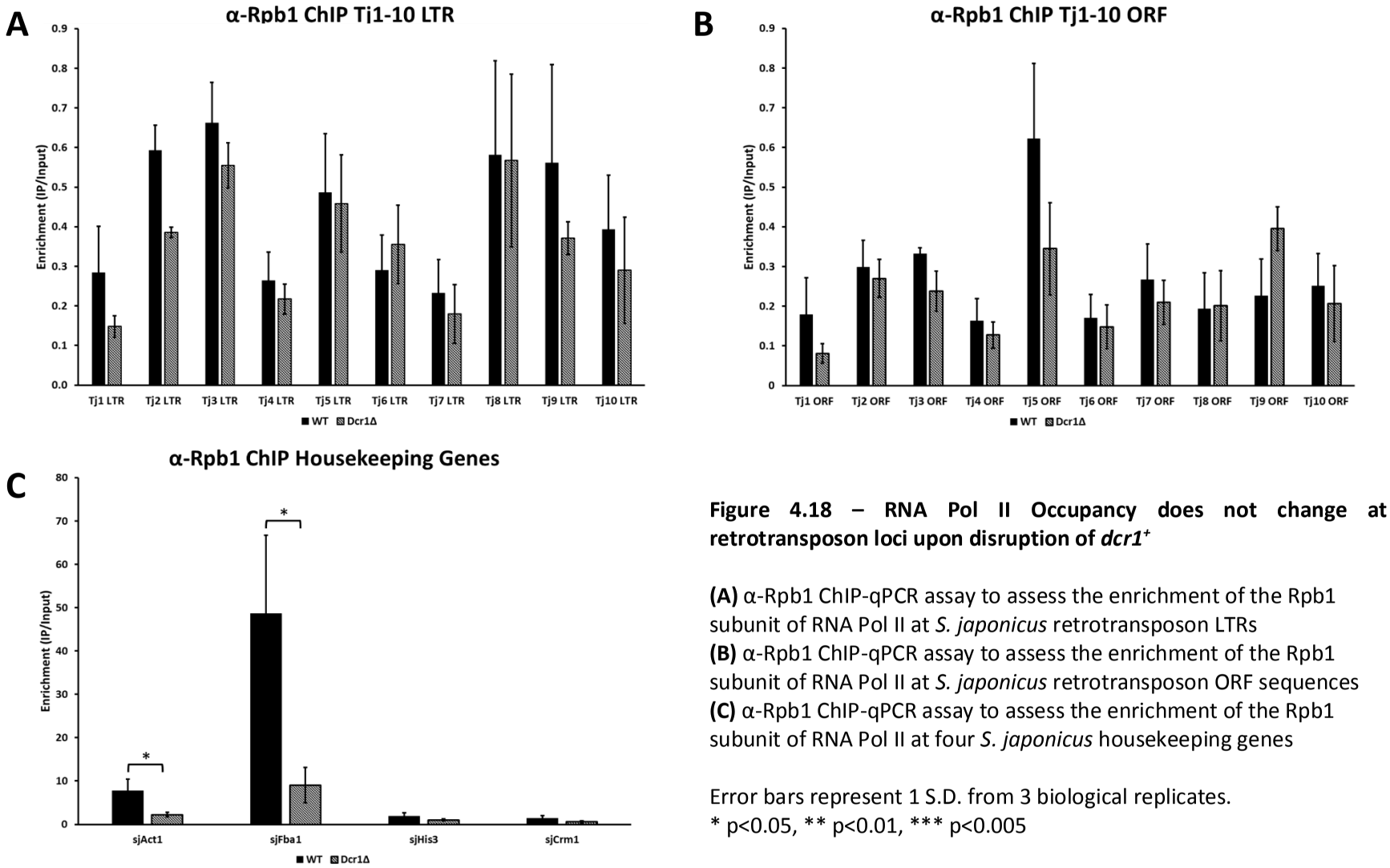


Figure 4.18 – RNA Pol II Occupancy does not change at retrotransposon loci upon disruption of *dcr1*$^{+}$

(A) α-Rpb1 ChIP-qPCR assay to assess the enrichment of the Rpb1 subunit of RNA Pol II at *S. japonicus* retrotransposon LTRs
(B) α-Rpb1 ChIP-qPCR assay to assess the enrichment of the Rpb1 subunit of RNA Pol II at *S. japonicus* retrotransposon ORF sequences
(C) α-Rpb1 ChIP-qPCR assay to assess the enrichment of the Rpb1 subunit of RNA Pol II at four *S. japonicus* housekeeping genes

Error bars represent 1 S.D. from 3 biological replicates.
* $p<0.05$, ** $p<0.01$, *** $p<0.005$

2-fold. The LTR of Tj9 appeared to increase in copy number 3-fold whilst the ORF did not change; this was unusual, as mobilisation of a retrotransposon requires the whole element to retrotranspose, and mobilisation of LTRs as discrete elements is not possible. The reason for the specific mobilisation of Tj2, Tj7 and Tj9 is unclear, as is the reason for the increased activity of Tj7 over other mobilised elements. It is possible that these differences are down to underlying variations in the coding sequence of each element, and that Tj2, Tj7 and Tj9 are the only element family to properly encode all factors required to mediate retrotransposition.

4.14 - Suppressor mutations may allow the *dcr1Δ5'* strain to survive

In the previous chapter it was proposed that RNA interference in *S. japonicus* is an essential process, owing to the inability thus far to isolate deletion mutants for most of the core protein coding genes involved in the pathway. However, this conclusion was challenged by the fact that single deletion and disruption mutants of *dcr1*$^{+}$ had been isolated, albeit at very low frequency. One possible explanation for the recovery of these specific *dcr1Δ* mutants is that spontaneous suppressor mutations may have arisen elsewhere in the genome, allowing survival of these strains. To investigate this, the genomic DNA of both a wild-type and the *dcr1Δ5'* mutant were sequenced and unique SNPs and indels were called as variants from the published genome sequence (Figure 4.20A). A majority of these SNPs or indels occurred at the repetitive regions at the telomeres, centromeres or tRNA arrays. Within the chromosome arms there were 6 unique SNPs in the *dcr1Δ5'* mutant, and of these 2 induced changes in the coding sequence of genes (Figure 4.20B). The first of these was

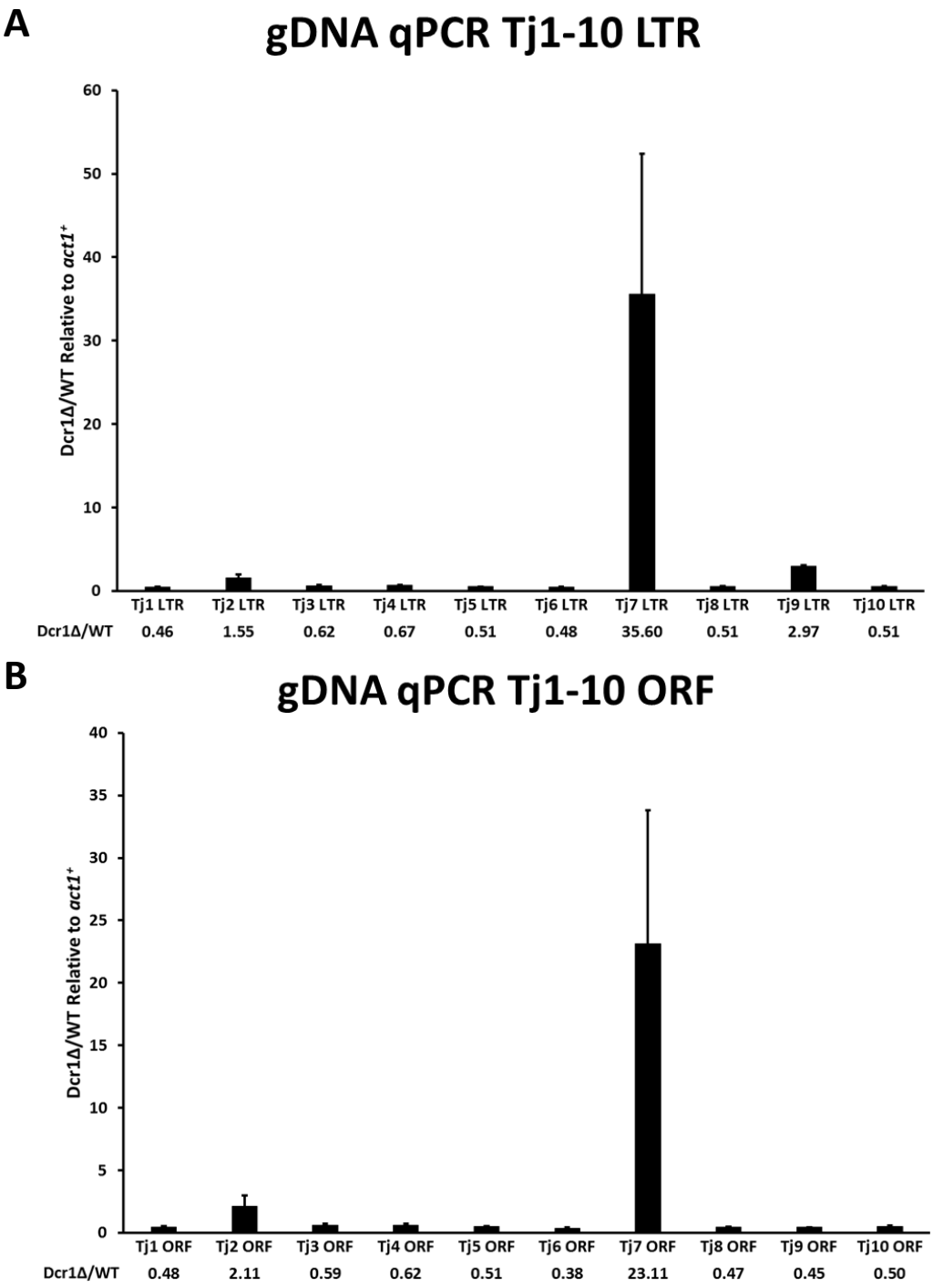
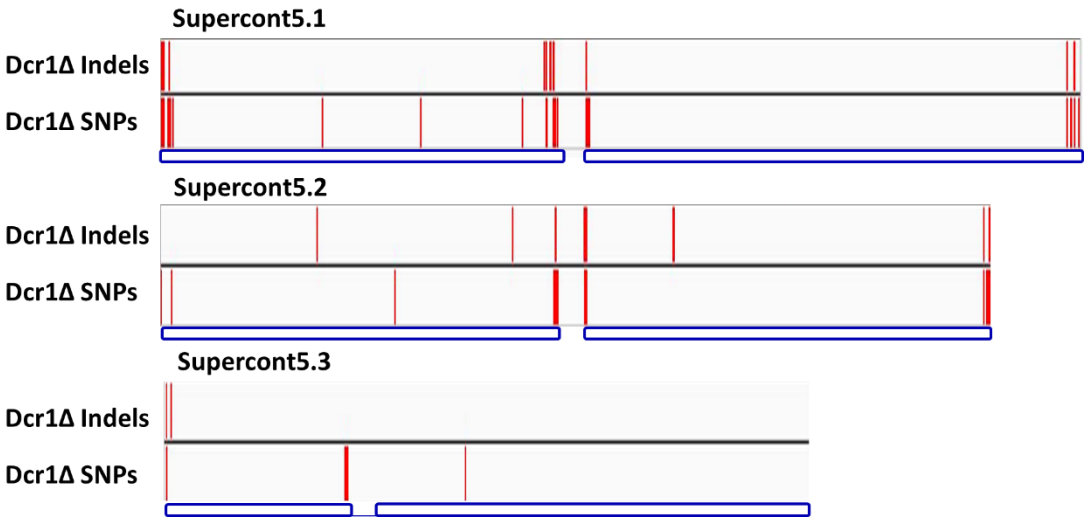


Figure 4.19 – A subset of retrotransposons appear to mobilise in the absence of *dcr1⁺*

(A) qPCR against retrotransposon LTRs, using genomic DNA extracted from wild type and *dcr1Δ* strains. Differences in copy number are shown relative to wild-type, normalised to the single-copy gene *act1⁺*
(B) qPCR against retrotransposon ORF sequences, using genomic DNA extracted from wild type and *dcr1Δ* strains. Differences in copy number are shown relative to wild-type, normalised to the single-copy gene *act1⁺*
Error bars represent 1 S.D. from 3 biological replicates.

a substitution mutation in *gpa2*$^{+}$ (SJAG_02444). This gene encodes for a heterotrimeric G-protein alpha subunit, and the SNP in the *dcr1*$^{+}$ mutant strain caused a V328F substitution mutation which may impact upon guanine nucleotide binding. The second coding SNP occurred within the mRNA cleavage ubiquitin-protein ligase E3 *mpe1*$^{+}$ (SJAG_02668). This mutation caused an R77W substitution, which sits within the conserved DWNN (domain with no name) domain. How either of these mutations might potentially compensate for the loss of *dcr1*$^{+}$ is not obvious, as neither factors are known to be directly involved in the nucleation of heterochromatin by RNAi in *S. pombe*.

A



B

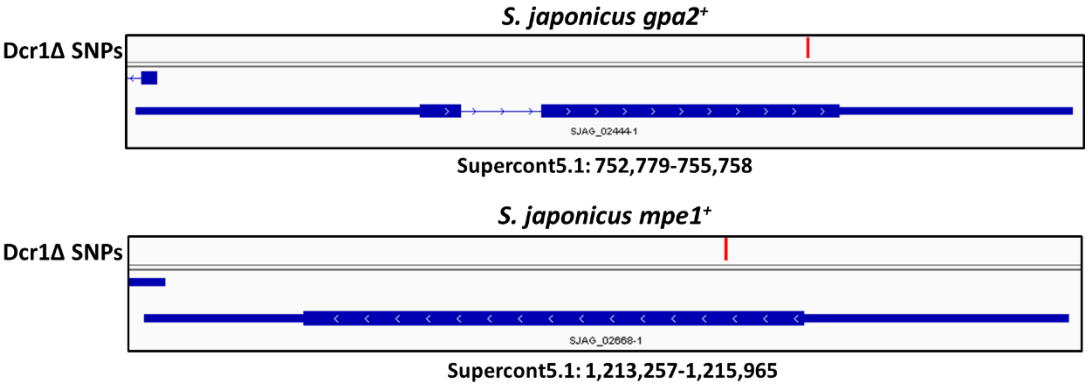


Figure 4.20 – The isolated *dcr1Δ5'* strain carries coding mutations in the *gpa2*$^{+}$and *mpe1*$^{+}$ genes

(A) Position of *dcr1Δ5'* specific Indels and SNPs that differ from the published *S. japonicus* genome assembly SJ5, for the 3 gene containing supercontigs

(B) Location of the *dcr1Δ5'* specific SNPs within the coding sequence of *gpa2*$^{+}$ and *mpe1*$^{+}$

4.15 - Discussion

In this chapter I have presented the work I undertook to investigate the impact of disrupting Dcr1 upon retrotransposon regulation in *S. japonicus*.

Of all tested core RNAi components in *S. japonicus*, *dcr1*$^{+}$ was the only gene that could be mutated, with two separate functional deletions isolated. The first of these mutants was a full deletion of the *dcr1*$^{+}$ ORF, which conferred a permanent and irreversible hyphal growth phenotype to *S. japonicus* cells. It was found that this mutant also truncated the 3'UTR of the mRNA that was encoded on the opposite strand that encoded the septin gene *spn6*$^{+}$. In *S. pombe spn6*$^{+}$ encodes a septin protein involved in meiotic spore formation, however orthologues in yeasts such as *S. cerevisiae* and *C. albicans* are known to have an effect on cell septation and hyphal morphology (Adams and Pringle, 1984; Haarer and Pringle, 1987; Li et al., 2012). The permanent hyphal phenotype observed in the *dcr1Δ5'* mutant may have either been caused by a legitimate *dcr1*$^{+}$ dependent effect, or may have been due to the truncation of the *spn6*$^{+}$ 3'UTR. To investigate this I constructed a partial deletion mutant of *dcr1*$^{+}$, in which a NatMX6 cassette was integrated at the 5' end of the *dcr1*$^{+}$ ORF, 3.4kb downstream of the *spn6*$^{+}$ 3'UTR. These cells grew in a far more 'yeast-like' way, however there were still a proportion that exhibited hyphal growth under normal conditions. This 'semi-hyphal' phenotype may be down to altered expression of *spn6*$^{+}$ upon *dcr1*$^{+}$ disruption, as the proper expression of convergent genes is proposed to contribute to their regulation (Gullerova et al., 2011).

Phenotypic analysis of these two *dcr1*$^{+}$ mutants showed no difference between the two, in terms of their sensitivity to TBZ. Both strains exhibited

hypersensitivity to this microtubule-destabilising drug, indicating chromosome segregation defects. This identical phenotype, coupled with initial H3K9me2 ChIP experiments and siRNA northern blots (data not shown), indicated that this 5' interruption mutant of *dcr1*$^{+}$ conferred the same phenotype at centromeric loci as the full ORF deletion, without the associated permanent hyphal growth. This interruption mutant also partially deleted the N terminal ATP binding domain of Dcr1, which has been shown to be required for siRNA generation in *S. pombe* (Colmenares et al., 2007). As the interruption mutant grew in a more 'yeast-like' way it was far easier to generate material for downstream analysis via growth in liquid media for this mutant than for the full deletion, which due to its obvious septation defects grew far more slowly and tended to flocculate in a more severe way. It was for this reason that the 5' disruption mutant of *dcr1*$^{+}$ was used for all subsequent experiments.

Looking first at retrotransposons, disruption of *dcr1*$^{+}$ caused complex changes at retrotransposon loci which could broadly be split into two distinct outcomes, depending on whether or not the element appeared to retrotranspose in the absence of Dcr1.

The elements that did not appear to mobilise were Tj1, Tj3, Tj4, Tj5, Tj6, Tj8 and Tj10. ChIP analysis revealed that these elements largely retained wild type levels of H3K9me2 in the absence of Dcr1, and with the exception of Tj1 and the LTR of Tj8, retained nucleosome occupancy. These elements largely accumulated transcript, with Tj3, Tj4, Tj6 and Tj10 showing the largest increase in transcript accumulation by RT-qPCR, whilst RNA-Seq analysis revealed that all families of retrotransposon were at least partially upregulated

in the absence of Dcr1. ChIP analysis of the major RNA Pol II subunit Rbp1 showed that this transcript accumulation was not caused by increased transcription of the element, as Pol II occupancy remained the same.

These elements gave rise to very few small RNA reads in the absence of Dcr1; analysis of the size and 5' nucleotide bias of these reads revealed that most were very short (~14nt) and were most likely a product of random transcript degradation. Tj8 exhibited an unusual pattern; by northern Blot it appeared to retain small RNAs that were slightly larger than the wild type species, however deep sequencing of small RNAs did not reveal any significant Tj8 derived species. This may be due to the increased specificity of deep sequencing over probe-based detection of small-RNAs, as any cross reactivity of the probe with reads derived from other elements would give rise to a spurious signal.

These non-mobile elements showed very modest, if any increase in CENP-A^{Cnp1} occupancy in the *dcr1*$^{+}$ mutant. Tj6 is unusual in that itactually appeared to lose CENP-A^{Cnp1} loading without Dcr1. This was interesting, as Tj6 was the only non-mobile element to accumulate transcript to the level of the mobile elements. Tj6 also appeared to be one of the two elements loaded with CENP-A^{Cnp1} in the wild-type background. Although disruption of *dcr1*$^{+}$ appeared to reduce CENP-A^{Cnp1} enrichment at Tj6, this reduction reflected the combined status of all Tj6 elements as ChIP-qPCR cannot distinguish specific loci. It is therefore feasible that CENP-A^{Cnp1} levels may in fact have increased on a solo Tj6 element located on a chromosome arm, possibly contributing to increased CENP-A^{Cnp1} at the housekeeping gene *fba1*$^{+}$. This enrichment of CENP-A^{Cnp1} within the chromosome arm is unusual, and may affect the fidelity of chromosome segregation in the *dcr1Δ5'* mutant.

The remaining elements Tj2, Tj7 and Tj9 did appear to mobilise in the absence of Dcr1, to varying degrees. Tj2 and Tj9 exhibited a modest increase in copy number, around 2-3 fold over wild-type, whilst Tj7 increased 25-35 fold. Interestingly it was only the LTR of Tj9 that appeared to mobilise whilst the ORF did not. This is unusual, as retrotransposition proceeds via mobilisation of a complete element, and solo-LTRs are usually the product of mobilisation followed by recombination between two adjacent LTR's. All of the elements that mobilised showed an overall reduction in the level of H3K9 dimethylation, associated with a loss of histone H3 rather than removal of the mark specifically. Tj7 and Tj9 also showed a marked overall reduction in nucleosome occupancy, as measured by analysing the enrichment of histone H4 at these loci. These changes in nucleosome occupancy were accompanied by the largest increases in transcript accumulation of all the elements, however these massive increases were independent of PolII occupancy, indicating that the accumulation was not due to an increase in the rate of transcription. Perhaps most interestingly these elements actually accumulated a large amount of small RNA-species in the absence of Dcr1, with more reads mapping to Tj7 and Tj9 in the *dcr1Δ5'* mutant than the wild-type. Analysis of these Dcr1-independent reads revealed that, as with the non-mobile elements, all reads were derived from sense transcript. However unlike for non-mobile elements, these sequences were incredibly abundant, and exhibited a more specific size profile and 5' nucleotide bias. These small RNAs showed a slightly broader size distribution than those derived via Dcr1 mediated cleavage, with a peak at around 18-19nt. The 5' nucleotide bias also displayed a dramatic shift, from a 5' U in the wild-type to a predominant 5' C in the mutant. This specific pattern

of small RNAs indicated that this degradation was not random, as was the case with non-mobile elements, but instead may have been mediated by a specific factor.

These three mobile elements, Tj2, Tj7 and Tj9, were the ones that showed the greatest increase in CENP-A^{Cnp1} enrichment when Dcr1 was absent. Tj7 seemed to be loaded with CENP-A^{Cnp1} in the wild-type, as did Tj6, indicating that these elements under normal conditions may define the position of the centromere. For the three mobile elements, Tj2, Tj7 and Tj9, the amplification in copy number could account for the observed increase in CENP-A^{Cnp1} loading, as it is possible that newly integrated elements were preferentially loaded with nucleosomes containing the centromeric CENP-A^{Cnp1} variant, rather than the canonical H3 histone. I will explore why this may occur in discussion chapter 7.2.

The correlation between element mobilisation, loss of H3K9me2, increase in transcript, accumulation of Dcr1-independent small RNAs and loading of CENP-A^{Cnp1}, indicated that these processes may be interconnected, and the ability of an element to mobilise may dictate transcript processing and locus modification. The extent of this relationship will be fully explored in discussion chapter 7.2.

The increase in copy number of certain elements also changes the way we must look at results from genome wide and qPCR studies in the *dcr1Δ5'* mutant. As the genome of this mutant appears to be plastic the copy number of certain elements changes from cell to cell. For experiments where an input normalisation calculation is performed, such ChIP-qPCR or ChIP-Seq, copy number is controlled for, as the amount of immunoprecipitated DNA is

normalised to the amount in the input fraction. This is not an issue if all elements are modified in the same way, however if highly similar or identical regions carry different modifications, for example if endogenous elements carry H3K9 methylation yet the more abundant newly mobilised elements are loaded with CENP-A^{Cnp1}, this will look like an overall decrease in H3K9me2 across all elements, when in reality the absolute amount of H3K9 methylation on the original element is the same as in the wild-type.

For RT-qPCR, RNA levels at loci of interest are commonly normalised to the levels of housekeeping genes such as *act1*$^{+}$ or *fba1*$^{+}$, which in theory should not change across different mutants. If there is an increase in copy number of a certain element, and the newly inserted elements are transcribed, it would appear as though the original element locus is massively transcriptionally upregulated, when in fact the transcript that accumulates would be generated from an expanded number of identical or near-identical elements.

A similar situation arises for RNA-Seq/siRNA-Seq experiments, where reads are mapped to a reference genome without an input normalisation. If copy number expansion occurs, reads generated from the multiple identical new copies of an element will all map back to the original element as the newly integrated elements are not present in the reference sequence, making it appear as though RNA levels are massively increased at this single element, when in fact the increase is likely spread across all elements. This becomes particularly evident in situations such as the one observed for Tj7, whereby transcript and small RNAs are massively enriched at Tj7-partial retrotransposon-1.1, however there are actually 30 other newly integrated copies of this element that are unaccounted for in the reference.

Aside from retrotransposons, Dcr1 also appeared to regulate subtelomeric genes. Of the most upregulated genes in the absence of Dcr1, a large proportion were unannotated 'hypothetical' proteins. Included in these were 3 of the 6 known *S. pombe* telomeric helicase homologues, whilst there were a number of zinc-finger containing genes, as well as transmembrane RTA1 family proteins. It is known that these gene families are enriched at subtelomeres in other species (Mefford and Trask, 2002; Riethman et al., 2004); their massive upregulation in the *dcr1Δ5'* background indicates that these loci are normally silenced in a Dcr1 dependent way. Whether this silencing is at the transcriptional or post-transcriptional level is currently unknown, as the methylation status of elements at these subtelomeric loci has yet to be assessed. The implication of Dcr1 in silencing of subtelomeric genes, and the observation that a number of these genes produce siRNAs, suggests that the RNAi pathway in *S. japonicus* may act on a more diverse set of substrates than in *S. pombe*. In *S. pombe*, the telomeric helicase genes contain a region homologous to the centromeric dg/dh elements (called cenH), however the *S. japonicus* Tlh genes do not carry this region and also do not seem to be homologous to known *S. japonicus* retrotransposons. Therefore how these sequences are targeted as RNAi substrates is unclear.

It should also be noted that disruption of Dcr1 downregulated a number of genes, which were enriched for those that are involved in metabolism. Whether this is due to a direct Dcr1 dependent effet is not known, as the mechanism by which Dcr1 could downregulate genes is unclear. It cannot be ruled out that the 'semi-hyphal' phenotype that is evident in the *dcr1Δ5'* mutant is linked to transcriptional reprogramming of metabolism genes, especially seeing as

induction of hyphal growth normally occurs upon nutrient starvation (Sipiczki et al., 1998).

Although two functional deletions of Dcr1 had been isolated, it was not possible to recover deletions in any of the other core RNAi factors in *S. japonicus*. It was for this reason that I postulated that the RNAi pathway is essential. It is possible therefore that there may have been some sort of compensatory mechanism in place, either via changes in gene expression or changes in protein coding sequences, that suppressed the loss of Dcr1 in the deletion mutant strains. Interrogation of the RNA-Seq data to look for genes that were up or downregulated in the *dcr1Δ5'* mutant revealed that a few genes involved in nucleating heterochromatin were upregulated (*clr4*$^+$, *stc1*$^+$, *chp2*$^+$ and *raf2*$^+$) whilst genes involved in small RNA processing were downregulated (*arb1*$^+$, *ers1*$^+$, *cid12*$^+$ and *tri1*$^+$), however the degree of change was very small (less than 1.5-fold), thus the biological relevance of this is doubtful. Similarly, some genes that are known to phenotypically suppress a *dcr1Δ* strain when deleted in *S. pombe* also displayed altered transcript levels, however as with most up/downregulated genes in the *dcr1Δ* mutant, the changes were fairly modest. In order to assess whether there were compensatory mutations in protein coding genes that allowed propagation of cells even in the absence of Dcr1, the genome of the mutant was sequenced. This revealed two potential SNPs in the coding sequence of genes, the first of which was a V328F mutation in the G-Protein Gpa2. Interestingly, in *S. pombe*, V327A gives constitutive catalytic activity (Douglas Ivey et al., 2010). There is little to suggest how this mutation could compensate for the loss of Dcr1, however Epistasis Mapping of the *S. pombe* genome (Ryan et al., 2012) revealed that *gpa2*$^+$ (SPAC23H3.13c)

exhibits a positive genetic interaction with RNAi-related genes *ago1*$^{+}$, *tas3*$^{+}$ and *ccr4*$^{+}$. A positive genetic interaction is defined as a 'mutation or deletion in separate genes, each of which alone causes a minimal phenotype, but when combined in the same cell results in a less severe fitness defect', and typically indicates genes that act together in the same pathway. The second mutation was an R77W mutation in mRNA cleavage ubiquitin-protein ligase E3 *mpe1*$^{+}$. This gene is essential in *S. pombe* (Kim et al., 2010) and is predicted to be involved with mRNA cleavage and polydenylation; however, how this mutation may affect the protein, and how this may rescue a Dcr1 deletion mutation, is unclear.

In this chapter I have presented results addressing the role of Dcr1 in the regulation of retrotransposons in *S. japonicus*. I have found that Dcr1 plays a major role in regulating these elements, however the mechanism of this regulation differs from that employed at repeat regions by *S. pombe*. In *S. japonicus* Dcr1 appears to function post-transcriptionally, acting to degrade retrotransposon transcripts to prevent accumulation and subsequent retrotransposition, independently of the H3K9me2 status of the element. Dcr1 is required to suppress mobilisation of a subset of retrotransposons, which in the absence of the ribonuclease increase in copy number and alter the composition of the *S. japonicus* genome.

Chapter 5 – Re-annotation of the *S. japonicus* genome

5.1 - Introduction

Sequencing the genome of a new species is an incredibly informative and useful undertaking, however there are a number of issues associated with building a complete genome sequence, especially one of a eukaryote. Mapping of repetitive regions can prove difficult (Eichler et al., 2004), and genebuilds only improve iteratively over time, with the application of experimental evidence, coupled with re-sequencing experiments and the advent of new technologies (Goodwin et al., 2016; Wang et al., 2009). Although the *S. pombe* genome was sequenced in 2002 (Wood et al., 2002) the exact number of the centromeric repeats is still not known, due to the difficulties associated with building a contig of repetitive sequence, despite work undertaken to address the issue (Ellermeier et al., 2010; He et al., 2016; Wood et al., 2002). In the case of *S. japonicus*, this issue is even more evident, as only a single genome sequencing experiment has been carried out (Rhind et al., 2011), and there are thus far no published studies looking at the regulation of *S. japonicus* repetitive regions. The repetitive regions in *S. japonicus* also appear to be far less ordered than in *S. pombe*, consisting of a mixture of full length and partial retrotransposons, as well as solo-LTRs and tRNA arrays (Rhind et al., 2011).

This study has generated the first mRNA and small RNA sequencing libraries, in a mutant that impacts repeat unit silencing in *S. japonicus*. Using this data it is therefore possible to re-assess the current *S. japonicus* genebuild, and to search for unannotated regions that are regulated by Dcr1. These regions may contain previously unidentified retrotransposons or other gene families, furthering our understanding of the *S. japonicus* repetitive sequence complement. This, coupled with the fact that the published genome annotation

of *S. japonicus* was competed using a previous genebuild SJ4 (which consists of 33 scaffolds), allows for the re-annotation of *S. japonicus* retrotransposon and repetitive sequences for the newest iteration of the assembly SJ5 (which consists of 32 scaffolds).

5.2 - Deep Sequencing of *S. japonicus* RNAs reveals a number of unannotated Dcr1 regulated regions

This study generated the first small RNA-Seq and mRNA-Seq datasets in an *S. japonicus* RNAi mutant. As well as revealing RNAi-mediated regulation of annotated repeat regions, combining these two datasets gave me the opportunity to identify previously unannotated regions of the *S. japonicus* genome that are regulated by Dcr1. As these regions carried no annotation, it made automating discovery difficult, thus these regions were identified manually. IGV (Thorvaldsdóttir et al., 2013) was used to visualise the locations of mapped small RNA and mRNA reads from wild type and *dcr1Δ* mutant strains, in relation to the known annotated coding genes, ncRNAs and retrotransposons (Rhind et al., 2011). The coordinates of any unannotated regions that showed a change in small RNA enrichment, coupled with an increase in transcript levels, were noted and the sequence exported. I identified 42 such 'regions of interest', listed in Figure 5.1. An example of one such 'region of interest' is shown in Figure 5.1. These sequences were then systematically searched for any conserved protein coding domains using the NCBI Conserved Domain resource (Marchler-Bauer et al., 2015); this tool gives an indication of any conserved protein domains contained within a primary nucleotide sequence, along with the frame in which it is encoded. Of the 42 'regions of interest', 22 contained no known protein domain, 2 contained DEAD-box helicase domains, whilst 18 contained protein domains associated with retrotransposable elements (Gag, Protease, Reverse Transcriptase, RNaseH, Integrase or Chromodomain). An example of the output for a region in the latter category is shown in Figure 5.2.

Supercontig	Start	End	Length (bp)	Features
5.1	1030	6023	4994	RT/RNaseH/Integrase/Chromodomain
5.2	1694	2753	1060	gag
5.2	3241	3716	476	RNaseH
5.2	1838570	1840390	1821	N/A
5.2	3843713	3844776	1064	RNaseH
5.2	3860154	3865461	5308	Chromodomain/Integrase/RNaseH/RT/Protease
5.3	860528	863503	2976	Integrase/RNaseH/RT/Protease
5.3	3032039	3035380	3342	N/A
5.4	415	1941	1527	N/A
5.4	5693	7125	1433	RT/RNaseH
5.4	7126	8401	1276	gag
5.4	38535	43081	4547	gag/Protease/RT/RNaseH/Integrase/Chromodomain
5.4	68073	68576	504	N/A
5.4	89488	91204	1717	Integrase/Chromodomain
5.4	132386	133351	966	N/A
5.4	139103	142759	3657	Protease/RT
5.5	601	3527	2927	Protease/RT
5.5	5759	7729	1971	Zinc finger
5.5	40415	43928	3514	Chromodomain/Integrase/RNaseH/RT/Protease
5.5	50162	51980	1819	N/A
5.5	84872	87264	2393	Protease/gag
5.5	92507	93163	657	N/A
5.6	4952	5750	799	N/A
5.6	40373	42441	2069	N/A
5.7	1	1041	1041	N/A
5.7	1224	2524	1301	N/A
5.7	10038	11315	1278	N/A
5.8	14207	17853	3647	N/A
5.8	18897	21967	3071	N/A
5.11	1184	1832	649	N/A
5.13	1	1444	1444	DEAD-box helicase
5.13	2074	7990	5917	N/A
5.13	8077	9552	1476	N/A
5.13	9734	9987	254	N/A
5.14	8230	8945	716	N/A
5.16	1095	5711	4617	Protease/RT/RNaseH/Integrase
5.17	1	458	458	N/A
5.20	3318	4770	1453	Integrase/Chromodomain
5.25	1	2800	2800	gag/Protease
5.26	1002	2799	1798	DEAD-box helicase
5.27	1	917	917	N/A
5.27	1562	2650	1089	N/A

Table 5.1 – List of potential Dcr1 regulated 'regions of interest'

Table showing manually identified 'regions of interest' that displayed altered small RNA and/or transcript accumulation upon disruption of *dcr1*$^{+}$, yet were not annotated with any known features. Conserved protein domains within these regions were discovered using the NCBI Conserved Domain Database (Marchler-Bauer et al., 2015) and are listed in the final column. (N/A = no conserved protein domain)

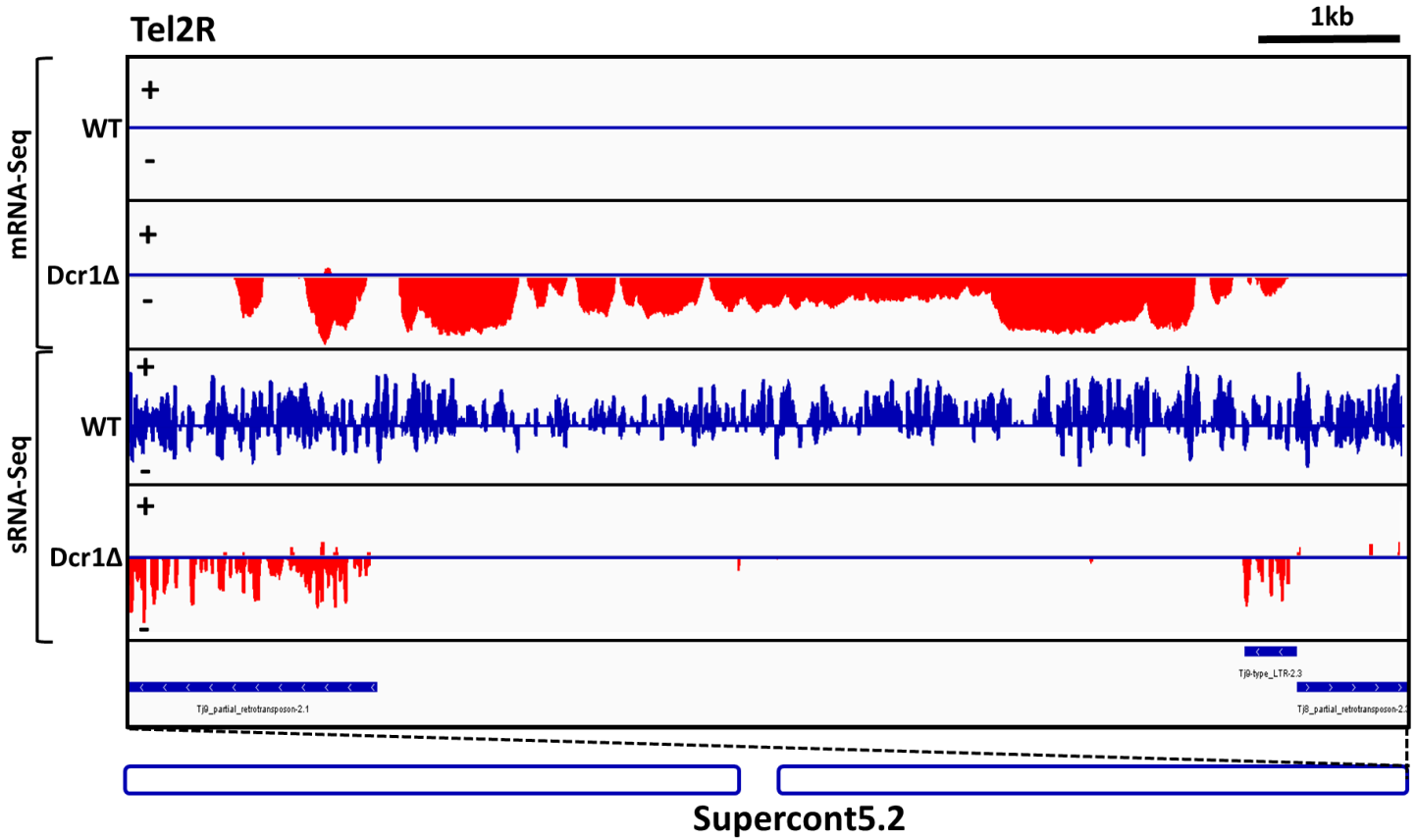
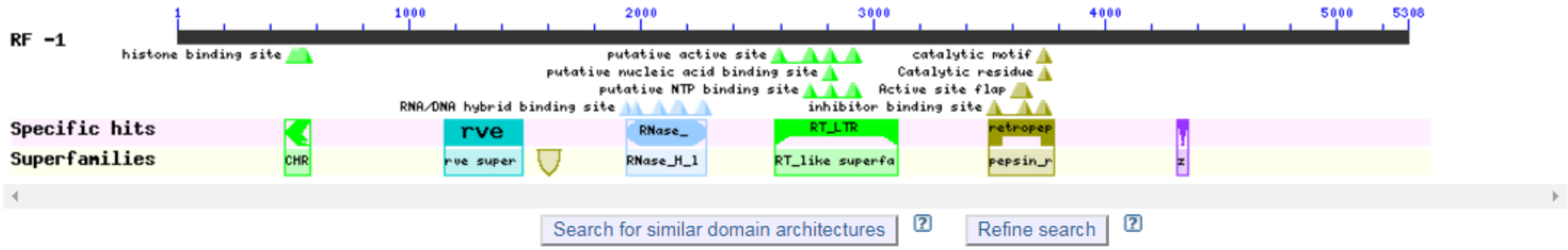


Figure 5.1 – Unannotated regions of the *S. japonicus* genome appear to be regulated by Dcr1

Genome browser view showing the number of mRNA and small RNA reads that align to supercont5.2:3860154-3865461 in wild type and *dcr1Δ* strains. Blue bars below represent annotated retrotransposon sequences



List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	RT_LTR	cd01647	RT_LTR: Reverse transcriptases (RTs) from retrotransposons and retroviruses which have long ...	2579-3103	4.65e-92
[+]	RNase_HI_RT_Ty3	cd09274	Ty3/Gypsy family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H) ...	1934-2281	2.15e-56
[+]	retropepsin_like	cd00303	Retropepsins; pepsin-like aspartate proteases; The family includes pepsin-like aspartate ...	3500-3781	3.33e-13
[+]	rve	pfam00665	Integrase core domain; Integrase mediates integration of a DNA copy of the viral genome into ...	1148-1486	6.47e-12
[+]	CHROMO	cd00024	Chromatin organization modifier (chromo) domain is a conserved region of around 50 amino acids ...	461-568	3.31e-06
[+]	zf-CCHC	pfam00098	Zinc knuckle; The zinc knuckle is a zinc binding motif composed of the the following ...	4313-4360	6.35e-05
[+]	zf-H2C2 super family	cl07828	His(2)-Cys(2) zinc finger; This domain binds to histone upstream activating sequence (UAS) ...	1550-1654	3.22e-03

Figure 5.2 – A number of unannotated Dcr1 regulated regions contain conserved retrotransposon related protein domains

Screenshot of the output from the NCBI Conserved Domain Database search tool (Marchler-Bauer et al., 2015) using the sequence from supercont5.2:3860154-3865461 as input. The panel at the top graphically illustrates the location of the conserved protein domains, RF indicates the frame of these domains (in this case -1). The bottom panel indicates the co-ordinates of these conserved domains, as well the E-value for each 'hit'.

5.3 - *S. japonicus* contains 22 discrete families of retrotransposon

For those regions of interest that contained retrotransposon-related domains, the nucleotide sequence of each discrete predicted protein domain was aligned against the equivalent domains from the 10 already-identified retrotransposon elements in *S. japonicus* using Clustal Omega (Sievers et al., 2011). This allowed me to ascertain whether these unannotated retrotransposon related protein domains belonged to any of the 10 known retrotransposon families, or whether these elements represented novel *S. japonicus* retrotransposons. From this sequence analysis, it was evident that all of the unannotated retrotransposon related sequences were derived from previously undiscovered retrotransposable elements, due to the level of sequence conservation between these unannotated elements and the 10 known *S. japonicus* retrotransposon families. Those elements that belong within the same family typically show >99% sequence identity, however the unannotated retrotransposons range from 40-70% identical to Tj1-10, which is the same degree of identity shown when the annotated Tj1-10 retrotransposons are aligned. Comparison of these unannotated retrotransposons revealed that in addition to the 10 previously discovered retroelement families, the *S. japonicus* genome contained both full length and partial remnants of an additional 12 retrotransposon families, which I have named Tj11-22. The original phylogenetic clustering of Tj1-10 was done by analysis of the Reverse Transcriptase sequences for each element (Rhind et al., 2011), thus I expanded this to include the newly discovered elements Tj11-22 (Figure 5.3). From this it was evident that these elements fell into one of the two existing lineages in *S. japonicus*. Tj13, Tj17 and Tj21 fell

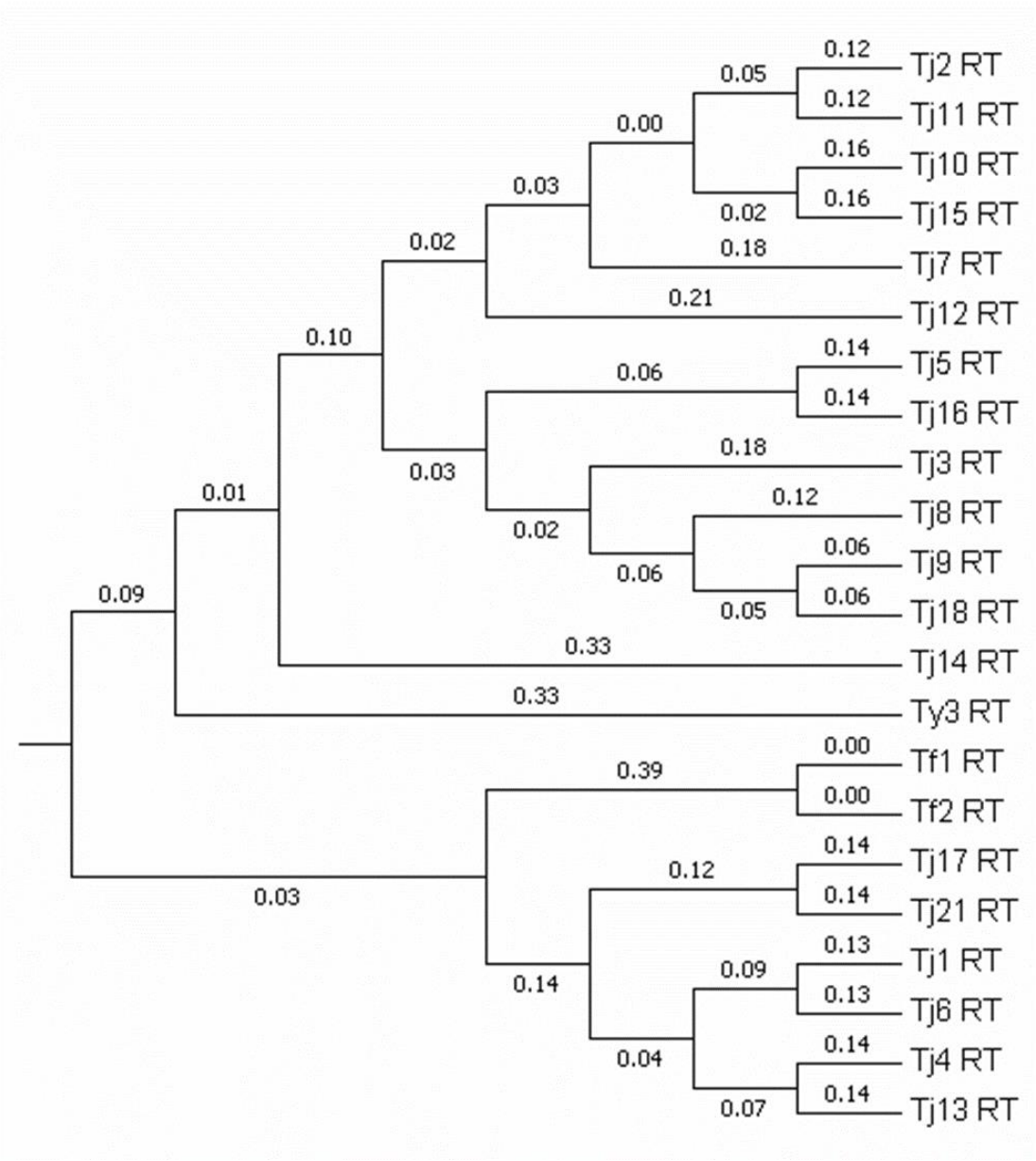


Figure 5.3 – Phylogenetic analysis of *S. japonicus* retrotransposon Reverse Transcriptase sequences reveals that newly discovered retroelements fall into one of the two existing lineages

Core *S. japonicus* Reverse Transcriptase sequences were isolated using co-ordinates generated from the NCBI Conserved Domain Database (Marchler-Bauer et al., 2015) and phylogenetic analysis was carried out using MEGA7 (Kumar et al., 2016). Sequences were first aligned using the ClustalW algorithm (Clustalw et al., 2003), the evolutionary history was inferred using the UPGMA method (Seath and Sokal, 1973). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site.

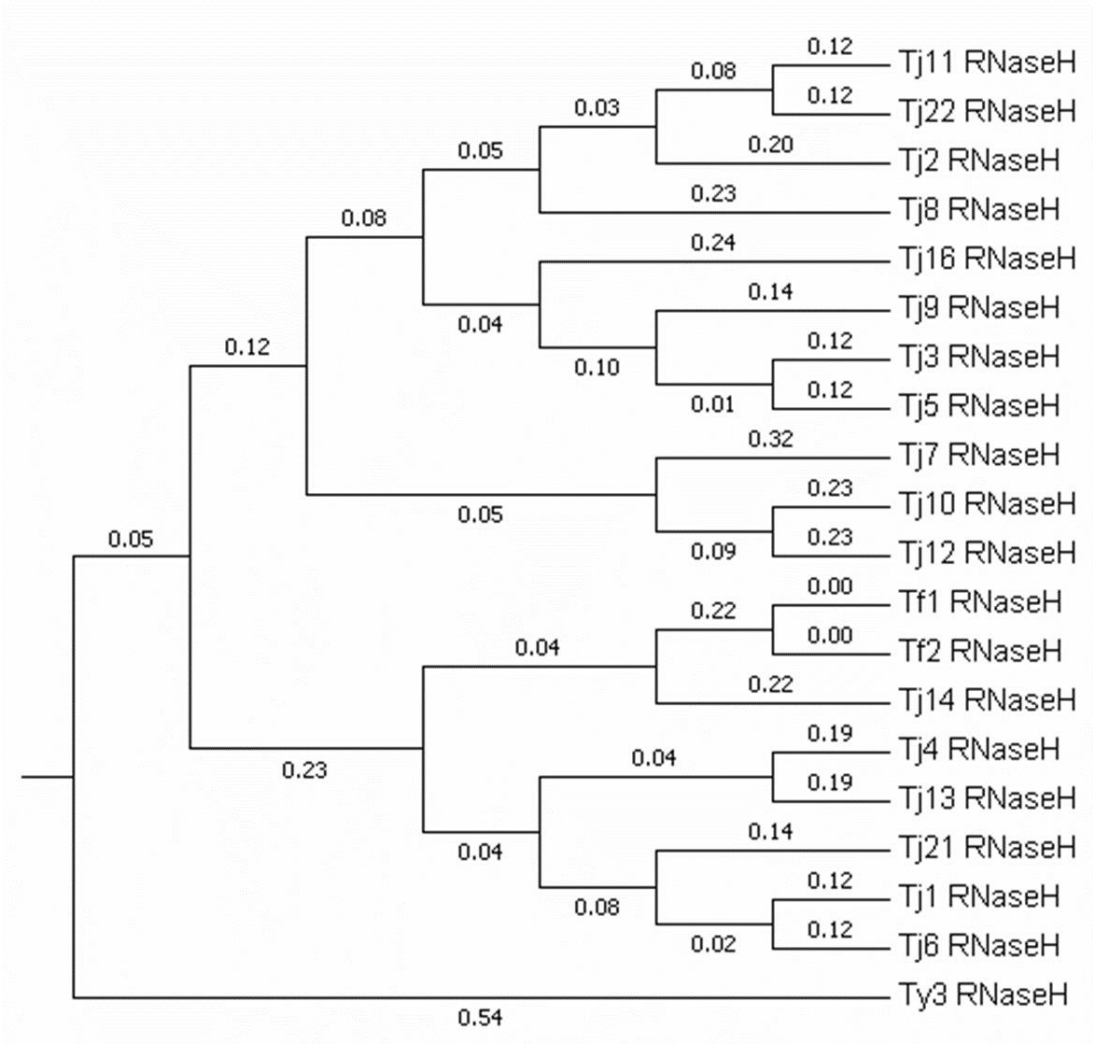


Figure 5.4 – Phylogenetic analysis of *S. japonicus* retrotransposon RNaseH sequences reveals that Tj22 is related to tRNA-primed elements

Core *S. japonicus* RNaseH sequences were isolated using co-ordinates generated from the NCBI Conserved Domain Database (Marchler-Bauer et al., 2015) and phylogenetic analysis carried out using MEGA7 (Kumar et al., 2016). Sequences were first aligned using the ClustalW algorithm (Clustalw et al., 2003), the evolutionary history was inferred using the UPGMA method (Seath and Sokal, 1973). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site.

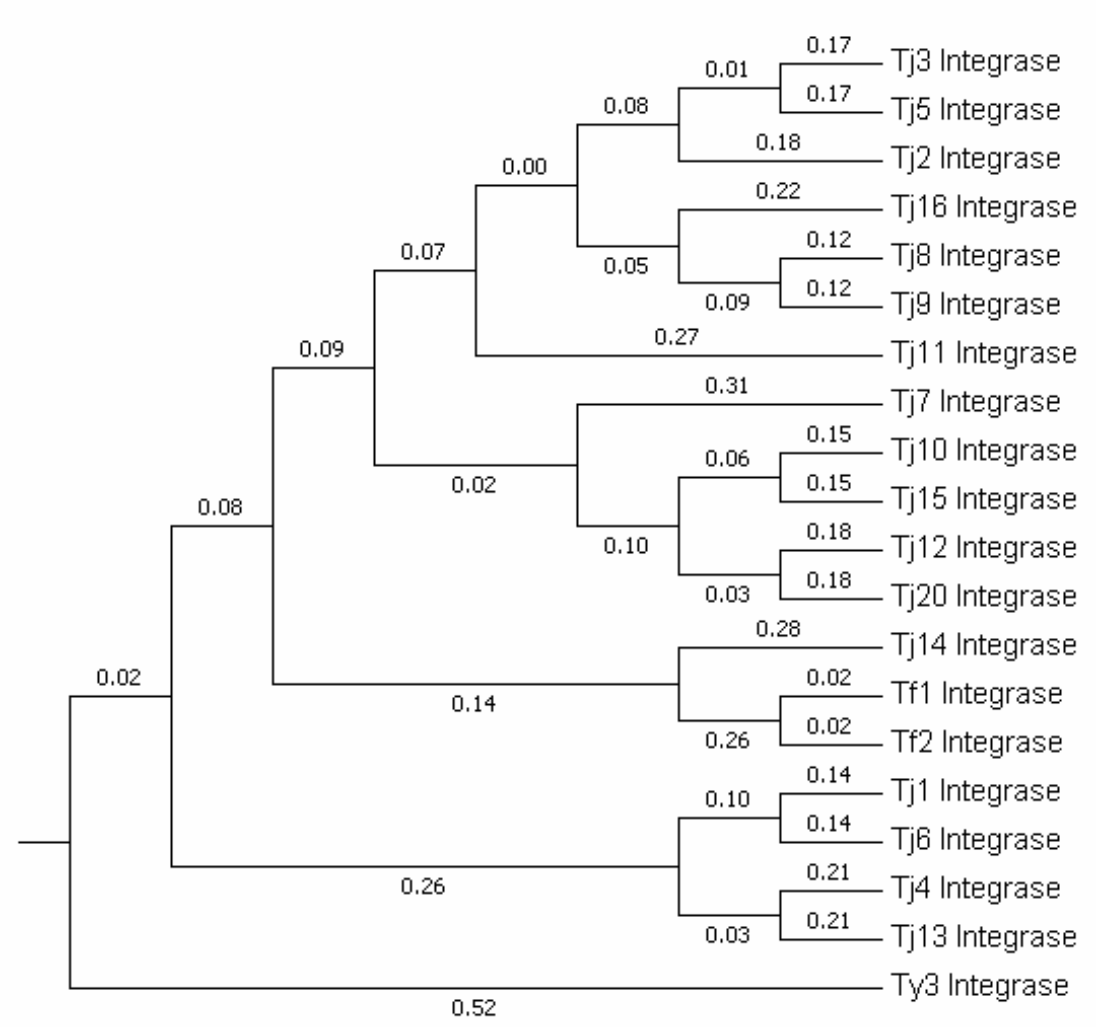


Figure 5.5 – Phylogenetic analysis of *S. japonicus* retrotransposon Integrase sequences reveals Tj20 is related to tRNA-primed elements

Core *S. japonicus* Integrase sequences were isolated using co-ordinates generated from the NCBI Conserved Domain Database (Marchler-Bauer et al., 2015) and phylogenetic analysis carried out using MEGA7 (Kumar et al., 2016). Sequences were first aligned using the ClustalW algorithm (Clustalw et al., 2003), the evolutionary history was inferred using the UPGMA method (Seath and Sokal, 1973). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site.

into the same clade as Tj1, Tj4 and Tj6, as well as *S. pombe* Tf1 and Tf2, probably self-priming their own reverse transcription. Tj11, Tj12, Tj14, Tj15, Tj16, Tj17 and Tj18 fell into the same clade as Tj2, Tj3, Tj5, Tj7, Tj8, Tj9 and Tj10, as well as Ty3, as elements that utilise tRNAs to prime their reverse transcription. Tj19, Tj20 and Tj22 did not have any identifiable Reverse Transcriptase sequence present within the genome, thus they were excluded from this analysis; however phylogenetic analysis of the RNAseH and Integrase sequences (Figure 5.4 & 5.5) indicated that both Tj20 and Tj22 were related to the Ty3 tRNA primed clade. Tj19 sequences contained only solo-LTRs and conserved gag-proximal zinc finger motifs, thus it was not possible to phylogenetically classify this element.

5.4 - Two new full-length LTR retrotransposons identified in *S. japonicus*: Tj11 and Tj12

Three of the 18 regions of interest that contained retrotransposon-associated domains contained the domains in the correct order and on the appropriate strand to constitute a full-length retrotransposon (see Figure 5.2 for example). These regions were then scanned upstream of the gag and downstream of the Integrase/Chromodomain for direct repeat sequences beginning TGT and ending ACA which could constitute LTRs. From this, three full length elements belonging to two families were discovered: Tj11-1, Tj12-1 and Tj12-2. Tj11-1 was found to be 5550bp long and flanked by two near-identical 278bp LTRs. Tj12-1 was 6044bp long, flanked at the 5' end by a 349bp LTR and at the 3' end by a 387bp LTR that was identical aside from a 38bp insert, 198bp from the start of the repeat. Tj12-2 was 6006bp, and was flanked at the 5' and 3' by near-

identical LTRs of 351bp and 349bp respectively. All of these elements contained a chromodomain at the end of the integrase, indicating that integration may be targeted to heterochromatic loci. Both of these retrotransposon families were phylogenetically classified as belonging to the Ty3 tRNA primed family, and analysis of the full length elements revealed sequences downstream of the 5' LTR that contained a conserved TGG motif (Figure 5.6A) which would allow them to anneal to the highly conserved CCA at the 3' end of tRNAs. Of these elements, Tj11-1 appeared to contain a number of frameshifts and premature stop codons within the pol sequence, whilst Tj12-1 contained a single frameshift in the middle of the predicted reverse transcriptase domain. Tj12-1 encoded a single ORF that included the Protease, Reverse Transcriptase, Integrase and Chromodomain sequences, indicating that this element may encode all components necessary to mediate mobilisation (Figure 5.6B).

5.5 - The *S. japonicus* centromeres and telomeres are more densely populated by retrotransposons than initially described

The 12 new retrotransposon families were identified by manual analysis of genomic regions that exhibited changed small RNA and/or transcript signatures upon deletion of Dcr1. This method is quite crude, as only large regions without annotation were immediately obvious to select for analysis. In order to establish whether there were also smaller partial retrotransposons or solo-LTRs related to these newly identified elements elsewhere in the genome, the 'regions of interest' (Table 5.1) were used as input sequences for BLAST (Altschul et al., 1990) searches against the *S. japonicus* genome. These

A

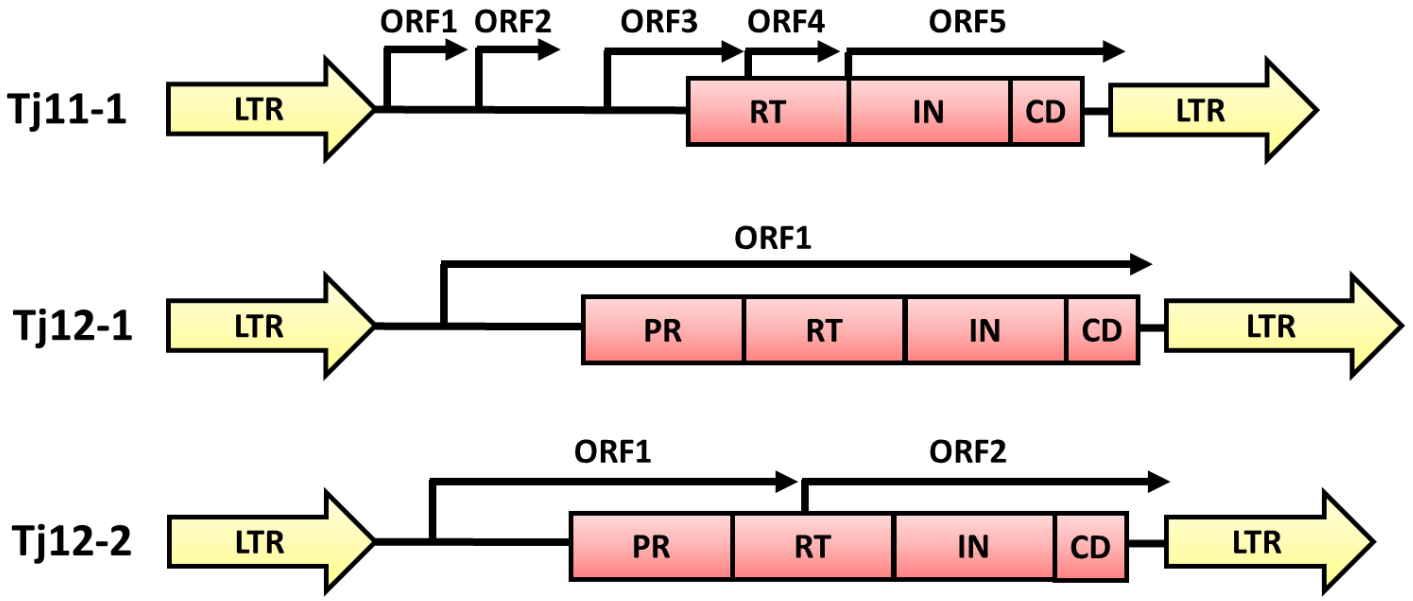
Element	Priming Site	Priming
Tj1	GTATTCGAC	Self
Tj2	TT**TGG**TTGCGAGGCAAAACTT	tRNA
Tj3	TT**TGG**TTGCGAGGCTCTTCTGATC	tRNA
Tj4	AAAGTGAACGACT	Self
Tj5	CT**TGG**TTGCGAGGCTTTAAC	tRNA
Tj6	TTGTTCGAC	Self
Tj7	CT**TGG**TTGCGAGGCTTACAACAA	tRNA
Tj8	CT**TGG**TTGCGAGGCTTCTGAAAA	tRNA
Tj9	CT**TGG**TTGCGAGGCTCATCAAAT	tRNA
Tj10	GG**TTG**CGAGTTCTCTGCCCTGT	tRNA
Tj11	CA**TGG**AACCGAG	tRNA
Tj12	TT**TGG**TTGCGAGGC	tRNA

Figure 5.6 – Tj11 and Tj12 appear to use tRNAs to prime their reverse transcription, however only 1 of the 3 newly discovered full-length retrotransposons expresses its polypeptide without any frame-shift mutations

(A) Table of potential Reverse Transcription priming sites within the elements Tj1-12. Underlined TGG motifs indicate the potential annealing site of the conserved 3' CCA of tRNAs

(B) Schematic illustration of the full length Tj11 and Tj12 elements. Right facing arrows indicate the position of ORFs within pol sequence (starting from the first ATG downstream of a stop codon) (PR = Protease, RT = Reverse Transcriptase, IN = integrase, CD = chromodomain)

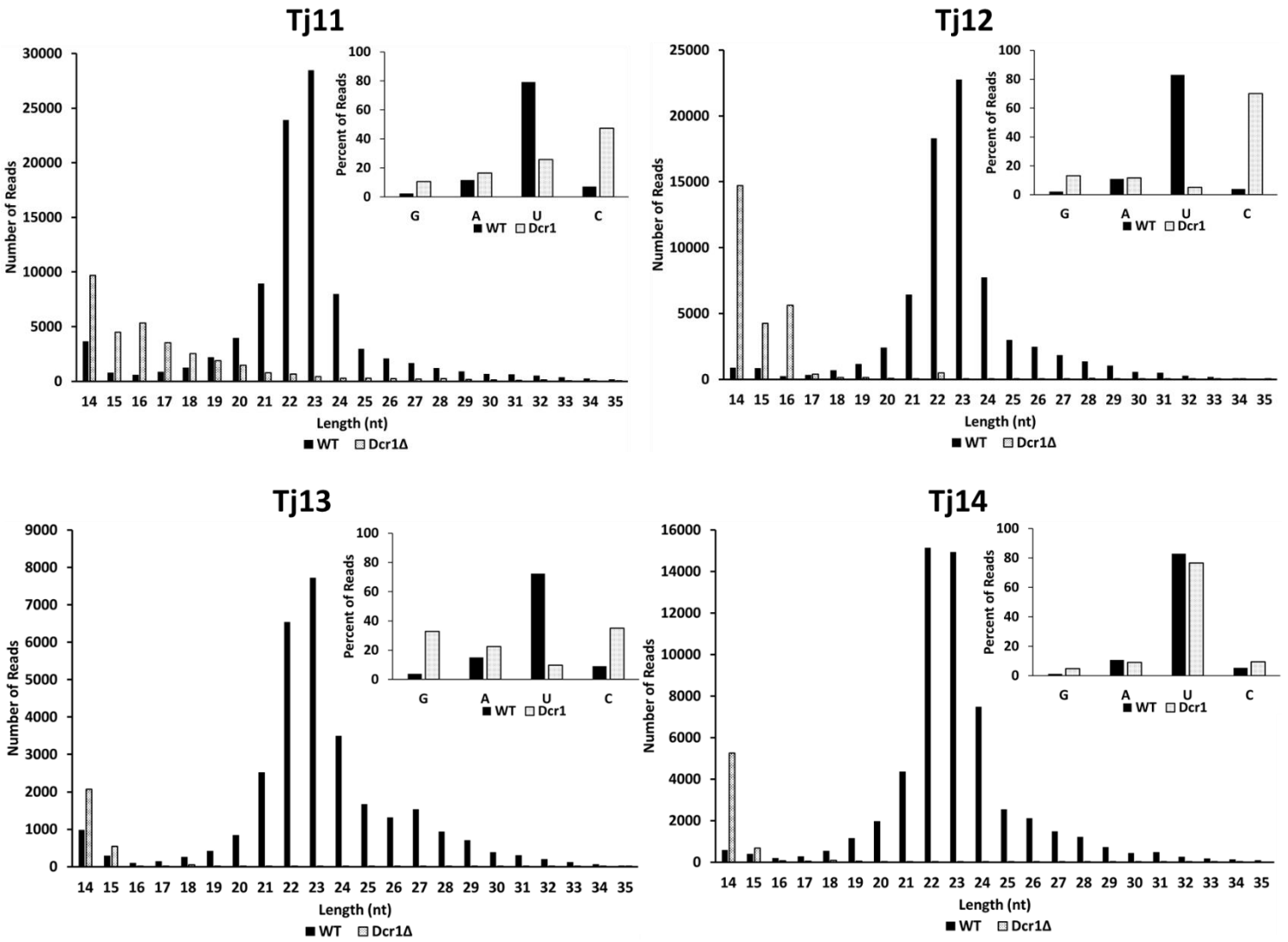
B

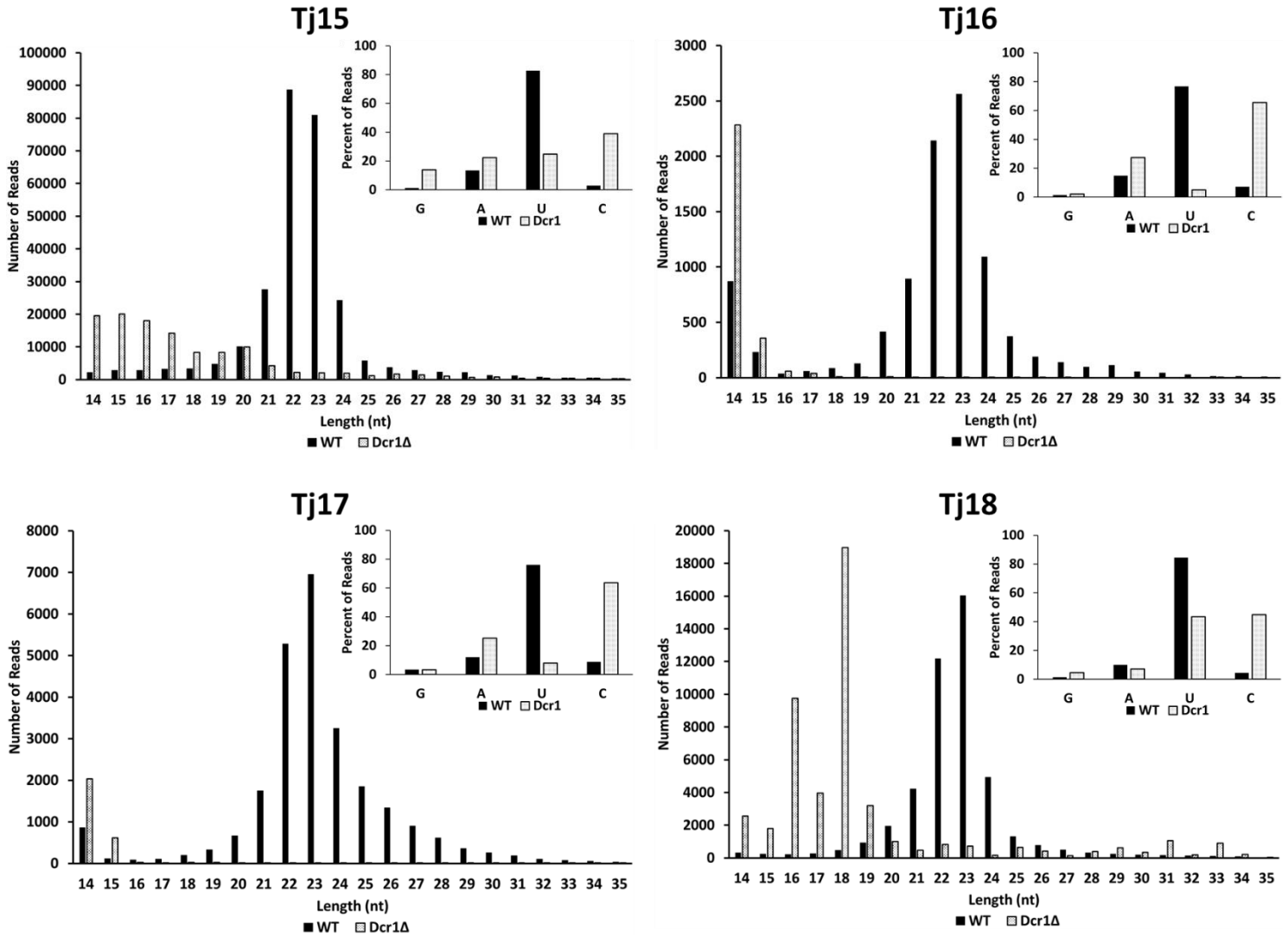


searches revealed homology between the 'regions of interest' and loci elsewhere within the genome, indicating that multiple partial copies these retroelements exist throughout the genome in the same way as for Tj1-10. These BLAST searches located 130 new retrotransposons, partial retrotransposons and solo LTRs (Supplementary Table 12) and of these 49 mapped to one of the three main chromosomal supercontigs. A majority of these 49 elements were located at either the presumed telomeres or centromeres, filling in previously unannotated gaps between Tj1-10 retrotransposon elements. This extended annotation revealed that there was actually very little 'non-retrotransposon' sequence that constituted centromeres and telomeres; instead, most of these regions are made up of irregular repeats of retrotransposon sequence, arranged consecutively in either the sense or antisense orientation. The remaining 81 elements mapped to the other 29 unplaced scaffolds, which are presumed to be located within the repetitive centromeres or telomeres, lending further weight to the conclusion the retrotransposable elements are targets for silencing at centromeres and telomeres. Of the 11.7Mb *S. japonicus* genome, 0.31Mb consists of retroelelements of the Tj1-10 family. The newly discovered elements Tj11-22 spanned a combined 0.11Mb, increasing the total portion of the *S. japonicus* genome annotated as made up of retrotransposons from 2.6% to 3.6%.

5.6 - A large proportion of newly discovered elements are regulated by Dcr1

As Tj1-10 seemed to be regulated by Dcr1, I wanted to assess whether these new elements were regulated in the same way. I also wanted to investigate whether these elements accumulated transcript and lost siRNAs on disruption of *dcr1*$^{+}$, as is consistent with elements that do not mobilise, or whether they accumulated both transcript and siRNA in the *dcr1Δ* background, as is the case for retrotransposons that appeared to mobilise throughout the genome. Log fold changes in both small RNA signal and mRNA were calculated for these elements using featureCounts (Liao et al., 2014) and DEseq2 (Love et al., 2014). The list of elements that showed significant changes in transcript accumulation was then compared to the list of elements that showed significant changes in small RNA signal. As these elements were visually identified by their apparent changes in small RNA abundance, it is not surprising that a majority (83/130) showed significant loss of small RNAs in the absence Dcr1, whilst 64 showed a significant increase in transcript (Supplementary Table 13 & 14) Of these, 50 of the elements showed both a decrease in siRNAs and an increase in transcript levels, consistent with those elements that were transcriptionally deregulated upon loss of Dcr1 but did not mobilise. Notably all 3 of the newly identified full length retrotransposons were seen to lose siRNAs and accumulate transcript, indicating that these may not be mobilisation competent, if they follow the pattern exhibited by Tj2, Tj7 and Tj9. There were no examples of elements that accumulated both transcript and small RNAs, indicating that none of the newly identified elements were capable of mobilisation.


Tj11
Number of Reads
Length (nt)
WT Dcr1Δ
Percent of Reads
G A U C
WT Dcr1
Tj12
Number of Reads
Length (nt)
WT Dcr1Δ
Percent of Reads
G A U C
WT Dcr1
Tj13
Number of Reads
Length (nt)
WT Dcr1Δ
Percent of Reads
G A U C
WT Dcr1
Tj14
Number of Reads
Length (nt)
WT Dcr1Δ
Percent of Reads
G A U C
WT Dcr1


Tj15
Number of Reads
Length (nt)
WT
Dcr1Δ
Percent of Reads
G
A
U
C
WT
Dcr1
Tj16
Tj17
Tj18

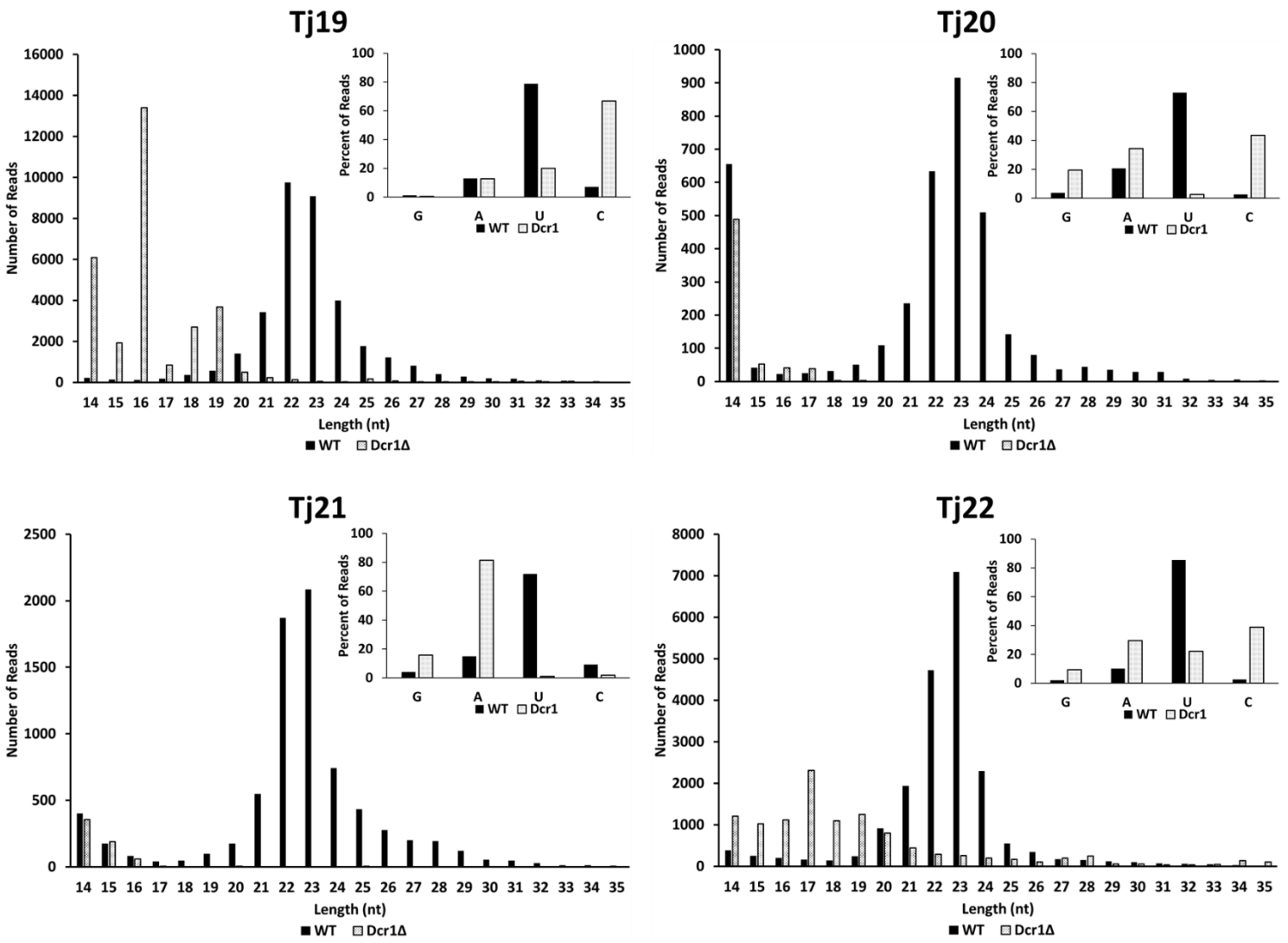


Figure 5.7 - Small RNA reads that map to newly annotated *S. japonicus* retroelements Tj11-22 exhibit a typical siRNA profile

Size distribution and 5' nucleotide bias for small RNA reads that map to each individual retrotransposon family Tj11-22, in the wild type and *dcr1Δ* strains.

Analysis of the size profile and 5' bias of small RNA reads mapping to the newly identified retrotransposon loci (Figure 5.7) in a wild-type situation revealed that these small RNA species were mostly 22-23nt long, with a strong preference for a Uracil at the 5' position. These characteristics confirmed that these small RNA species were legitimate siRNAs; this size pattern and 5' nucleotide bias are shared by siRNAs in *S. pombe* (Djupedal et al., 2009) and *S. japonicus* (this study). In the *dcr1Δ* mutant the size profile of the remaining small RNAs was changed for all elements; most small RNA species were found to be much smaller at 14-16nt. The exception to this was Tj18, which seemed to retain reads of around 18nt. This pattern is similar to the mobilisation competent elements Tj7 and Tj9, however the origin of these Tj18 derived reads is unclear, as no full length Tj18 element has been identified and thus it is unlikely to retrotranspose. Phylogenetic analysis did show that this element is closely related to Tj2 and Tj9, two elements that retain some small RNA signal and do appear to mobilise in the absence of Dcr1, thus some Tj2/9 derived reads may map to Tj18.

The bias towards a 5' Uracil is also mainly lost in small RNAs in the *dcr1Δ* mutant, with a 5' C the predominant motif associated with small RNAs generated from these elements in the absence of Dcr1. This pattern echoes that seen with the previously annotated Tj1-10 elements, indicating that these 12 new elements may be regulated by Dcr1 in a similar way.

5.7 - Identification of two new Telomeric Helicase homologues

As mentioned in Section 5.2, analysis of Dcr1 regulated 'regions of interest' revealed two that encoded DEAD-box helicases. In *S. pombe*, the subtelomeric

genes Tlh1 and Tlh2 both encode RecQ family DEAD box helicases that are known to generate siRNAs via their conserved centromeric-like cenH sequences (Cam et al., 2005) and establish heterochromatin via RNAi (Kanoh et al., 2005). In *S. japonicus*, there are six identified homologues of Tlh1/2, five of which contain at least partial DEAD-box helicase domains. To see whether the unannotated DEAD box helicase domains were related to the known telomeric helicase genes, the core helicase domains from the unannotated regions were aligned to the core helicase domains from SJAG_05353, SJAG_05173, SJAG_05105, SJAG_06597 and SJAG_06642. From this analysis it was evident that the unannotated helicase found on supercontig 5.26 was related to SJAG_05353 and SJAG_05173, showing 97-100% identity to these elements. The helicase on supercontig 5.13 showed less identity to conserved telomeric helicases, however this partial domain was only 231bp long, yet still had 60% identity to SJAG_05173. This indicated that these newly discovered DEAD-box helicase containing genes were likely to also be homologues of the *S. pombe* Tlh1/2 telomeric helicases and were therefore presumed to be located within subtelomeric regions in *S. japonicus*.

5.8 – Over half of Dcr1 regulated loci carry no conserved features

By far the largest portion of Dcr1 regulated 'regions of interest' carried no conserved protein domains at all. These regions ranged from 254bp to as large as 5917bp (Table 5.1). These regions were selected for the reason that they lost small RNA signal upon disruption of *dcr1*$^{+}$, indicating that these loci were in some way regulated by Dcr1 (Figure 5.8). A number also showed evidence of increased transcript accumulation in the absence of Dcr1.

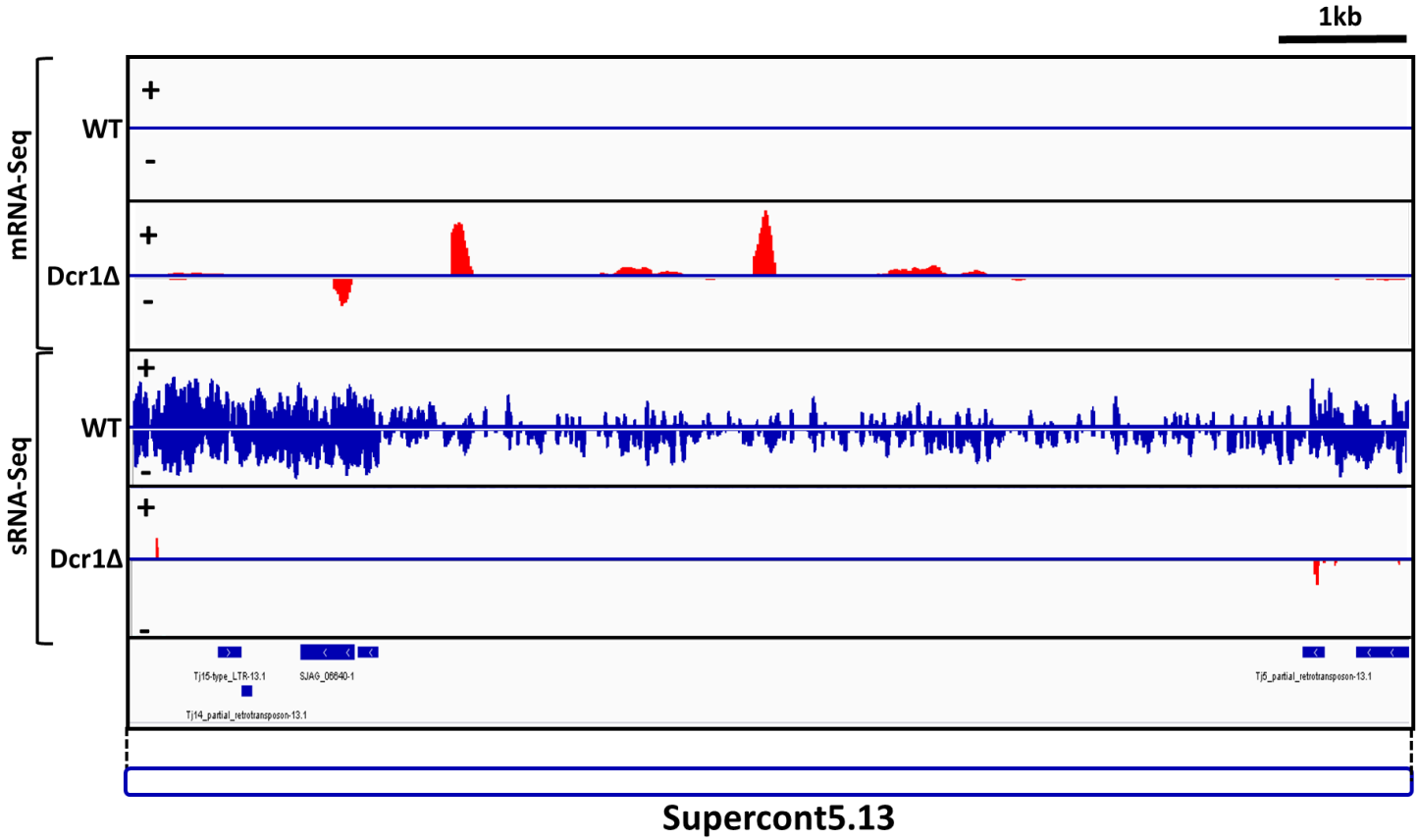


Figure 5.8 – A majority of Dcr1 regulated regions of the *S. japonicus* genome contain no conserved protein coding domains and remain unannotated

Genome browser view showing the number of mRNA and small RNA reads that align to supercont5.13 in wild type and *dcr1Δ* strains. Blue bars below represent annotated retrotransposon and coding gene sequences

Initial evaluation of these domains using BLAST (Altschul et al., 1990) to look for repetition of these sequences throughout the genome found that a majority did have some homology to other regions (data not shown). These were mostly short stretches that may possibly correspond to solo-LTRs and other retrotransposon derived sequences that do not contain known protein-coding domains. Due to time constraints, full characterisation of these unannotated regions was not possible, however these remain of interest for future investigation.

5.9 – Discussion

In this chapter I have presented the work I undertook to extend annotation of Dcr1 regulated loci in *S. japonicus*, utilising the genome-wide RNA-sequencing datasets I generated during the course of this study.

I manually identified a number of regions that appeared to lose siRNAs and/or increase in transcript abundance in the absence of Dcr1, that were not yet annotated with any feature. Bioinformatic analysis of these regions revealed that a number contained known retrotransposon related domains, such as Proteases, Reverse Transcriptases, RNase H and Integrase domains. Phylogenetic analysis of these domains uncovered 12 new retrotransposon families in *S. japonicus* that were evolutionarily distinct from the previously annotated Tj1-10 elements. Of these newly discovered elements, three appeared to be full length retrotransposons, encoding a Protease, Reverse Transcriptase and Integrase domain, as well as a C-terminal Chromodomain, flanked by full LTRs. I identified potential tRNA priming sites for reverse transcription in all three of these elements, however only one element appeared to have maintained the ability to mobilise, as both other elements contained frameshift mutations within the polypeptide coding sequence. These full length elements may not have been discovered during original analysis (Rhind et al., 2011) as they lie at the very end of larger supercontigs; these contigs may have had sequence added to them between genebuild Sj4 and Sj5, thus extending coverage to the full retroelements. Originally annotated partial retrotransposon elements were identified via BLAST searches using full length retrotransposon sequence; any truncated elements

were not used as inputs, thus excluding Tj13-22 from annotation. Programs such as RepeatMasker (Tarailo-Graovac and Chen, 2009) are commonly used to identify known transposon sequences within a genome, however this tool requires a library of known repeats to search against, something that is not available for unannotated retroelements.

Analysis of the newly discovered retroelements revealed that in the wild-type, all produced canonical Dcr1 dependent siRNAs, of 22-23nt with a dominant 5' Uracil. Upon disruption of Dcr1 this distinctive pattern was altered, with a loss of siRNAs observed for most elements. Any small RNAs that were generated were very short (14-16nt), consistent with those generated via degradation of transcript. The disruption of *dcr1*$^{+}$ also increased transcript accumulation for roughly half of all newly identified elements, thus these elements appear to be regulated by Dcr1 in the same way as Tj1-10. Whether this regulation occurs transcriptionally or post-transcriptionally in unknown, as the methylation status of these elements has not yet been investigated. The fact that Dcr1 seems to function exclusively in a post-transcriptional way to regulate Tj1-10 elements suggests that this silencing mechanism may extend to all retrotransposons.

Aside from retrotransposon derived sequences, I also identified two new RecQ-family helicases that were related to the known telomeric helicases in both *S. pombe* and *S. japonicus*. Both of these elements seem to be regulated by Dcr1, and add further weight to the argument that Dcr1 acts to repress transcription of a diverse range of elements in *S. japonicus*. This discovery of two new presumed telomeric associated genes, both in as-yet unplaced scaffolds, also aids in locating these scaffolds, suggesting that they are likely telomeric rather

than centromeric sequences. Further identification of position specific elements could in future be used to guide placement of unknown contigs.

Finally, I identified a number of regions that appear to be regulated by Dcr1 but contain no known protein domains. Time constraints did not allow me to explore these in detail, but investigation of these regions may reveal other types of Dcr1 regulated loci, with as-yet unknown function.

This re-annotation of retrotransposons highlighted just how incomplete the published *S. japonicus* genome annotation was, and potentially still is. Given that 29 of the 32 scaffold contigs of the SJ5 genome assembly are not associated with any of the 3 main chromosomal contigs, it is possible that there are still a number of previously unidentified retroelements, either associated with sequence gaps, or spanning smaller contigs that have yet to be joined. Combining data from just 2 genome wide experiments increased the retrotransposon annotation by 40%, and revealed new functionally active elements, not just partial pieces and solo-LTRs. This exercise demonstrated the value of combining computational with experimental approaches, and continuously updating genome annotations as new experimental evidence becomes available, especially when working with uncommonly studied model organisms.

Chapter 6 – Expression of the *S. japonicus* retrotransposon Tj1 in *S. pombe*

6.1 - Introduction

In Chapter 4 I showed evidence that in *S. japonicus* the RNAi pathway acts to constrain retrotransposons and prevent accumulation of their transcripts and subsequent transposition. This is in contrast to the mechanism employed by *S. pombe*, whereby RNA interference does not regulate retrotransposons under normal conditions; instead a set of domesticated transposon factors, homologous to the human CENP-B proteins, work in combination with HDACs and the Ku heterodimer to restrict transcription and mobilisation of endogenous retrotransposons.

As *S. pombe* and *S. japonicus* employ different mechanisms of silencing retrotransposons, I wanted to introduce a retroelement from *S. japonicus* into *S. pombe,* to study how this 'foreign' element would be handled. This allowed me to address whether RNAi mediated targeting of this retrotransposon was inherent to the element or to the host organism, and also facilitated identification of the integration pattern of this *S. japonicus* retrotransposon.

Much work has been carried out using the recently extinct S. *pombe* retrotransposon Tf1 to study integration patterns and element repression (Cam et al., 2008; Guo and Levin, 2010; Hickey et al., 2015; Levin and Boeke, 1992; Lorenz et al., 2012), however thus far no work has been carried out to examine the silencing of an exogenous retroelement in *S. pombe*. In this chapter I present work carried out to profile the pattern of Tj1 integration events in *S. pombe* and elucidate the mechanism of silencing employed by *S. pombe* when challenged with an *S. japonicus* retroelement that has been shown to elicit an RNAi response in this related fission yeast species.

6.2 - Tj1 is able to actively retrotranspose in *S. pombe*

In order to establish whether Tj1 could actively mobilise in *S. pombe,* the plasmid pHL2861 was obtained from the lab of Henry Levin (Guo et al., 2015). This plasmid carried the full-length Tj1 element marked with a Neomycin resistance cassette, under transcriptional control on the thiamine-regulatable *nmt1* promoter, as well as the *URA3* gene for selection of the plasmid. Tj1 was shown in Chapter 4 to be regulated by Dcr1 in *S. japonicus*; deletion of this gene caused loss of associated siRNAs and an increase in transcript accumulation. It was also one of the only elements within the SJ5 assembly that did not appear to have a frameshift within the gag or pol sequences that could potentially prevent retrotransposition. As no full length copy of Tj1 had been annotated within the *S. japonicus* genome assembly (SJ5) this plasmid was constructed by taking the longest partial Tj1 element (which is 4514bp long and contains a partial 5' LTR and a full length 3' LTR) and fusing the self-priming sequence in the 5′ LTR to the *nmt1* promoter. A neomycin resistance (*neoR*) cassette was then integrated upstream of the polypurine tract, so that integration of the element could be tracked by screening for resistance to Geneticin (G418).

In order to evaluate the integration pattern of this element, the plasmid borne copy of Tj1 was transformed into the wild-type *S. pombe* strain and a transposition assay was performed (See Figure 6.1 for assay overview). Following transformation, a number of ura^{-}, G418^{R} colonies were isolated, indicating that Tj1 was capable of mobilising into the genome of *S. pombe*. As I wanted to ascertain where these Tj1 elements had integrated within the *S. pombe* genome, it was first important to isolate single copy insertion events to

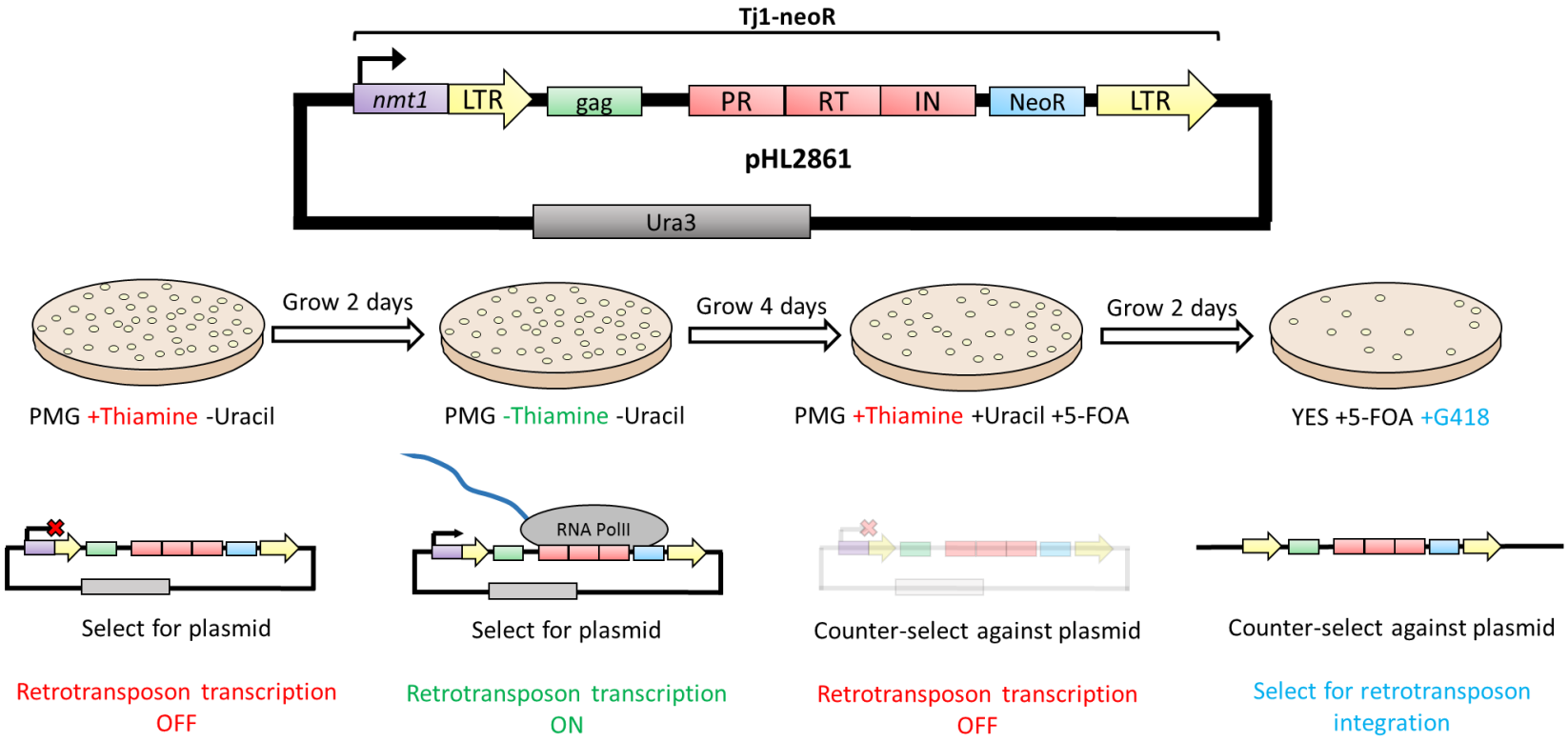


Figure 6.1 – Overview of the Tj1 retrotransposition assay

The plasmid pHL2861 encodes a copy of Tj1 under the control of the *nmt1*$^{+}$ promoter, with a *neoR* cassette integrated upstream of the polypurine tract, on a *Ura3*$^{+}$ backbone. This plasmid was transformed into *S. pombe* and transformants were first plated on PMG-Ura+Thi to select for the plasmid and to transcriptionally repress Tj1 expression, before being replica plated on to PMG-Ura-Thi to drive Tj1 expression. Ura resistant colonies were then replica plated on to PMG+Thi+Ura+5-FOA to both repress Tj1 transcription and to drive loss of the plasmid itself. These 5-FOA resistant colonies were then replica plated on to YES+5-FOA+G418 to select for cells that carried genomically integrated copies of Tj1-*neoR*

facilitate Sanger sequencing of the loci. To do this, southern blot analysis was carried out; genomic DNA was extracted from 20 *neoR* colonies, this was then digested overnight with HindIII and Southern blotting performed. A probe that hybridised to the *neoR* cassette was used, the reason being that there was a HindIII site at the 5' end of the *neoR* cassette, thus it was predicted that the downstream HindIII site would sit somewhere within the *S. pombe* genome, depending on where the Tj1 element integrated. Probing against *neoR* would therefore give a unique banding pattern depending on the location of integration, making it possible to detect strains where only a single copy of Tj1 was integrated (Figure 6.2A).

Of the 20 selected colonies, nine were found to have single copy insertions by Southern blot (colonies 5, 7, 13, 14, 15, 16, 17, 19 and 20) (Figure 6.2B). These clones were therefore taken forward for insertion sequencing.

6.3 - Tj1 integrates upstream of PolIII transcribed genes

In order to sequence these single copy Tj1 insertions, inverse PCR was used (see Figure 6.2A for sequencing strategy overview). Of the 9 single colonies selected, 6 were successfully sequenced. Of these 6 integration events, 5 Tj1 elements integrated within 11bp of a tRNA gene, with the remaining copy of Tj1 integrating 4bp upstream of a 28S rRNA gene (Figure 6.3A). There was no obvious bias towards a specific tRNA insertion locus, with 2 integrations occurring upstream of tRNA-Asp, and the other 3 occurring upstream of either a tRNA-Gly, tRNA-Ser of tRNA-Phe. The integrated copies of Tj1 were also spread across the arms of all 3 chromosomes, with no apparent bias towards specific chromosomal loci as sites of insertion. All integrations occurred on the

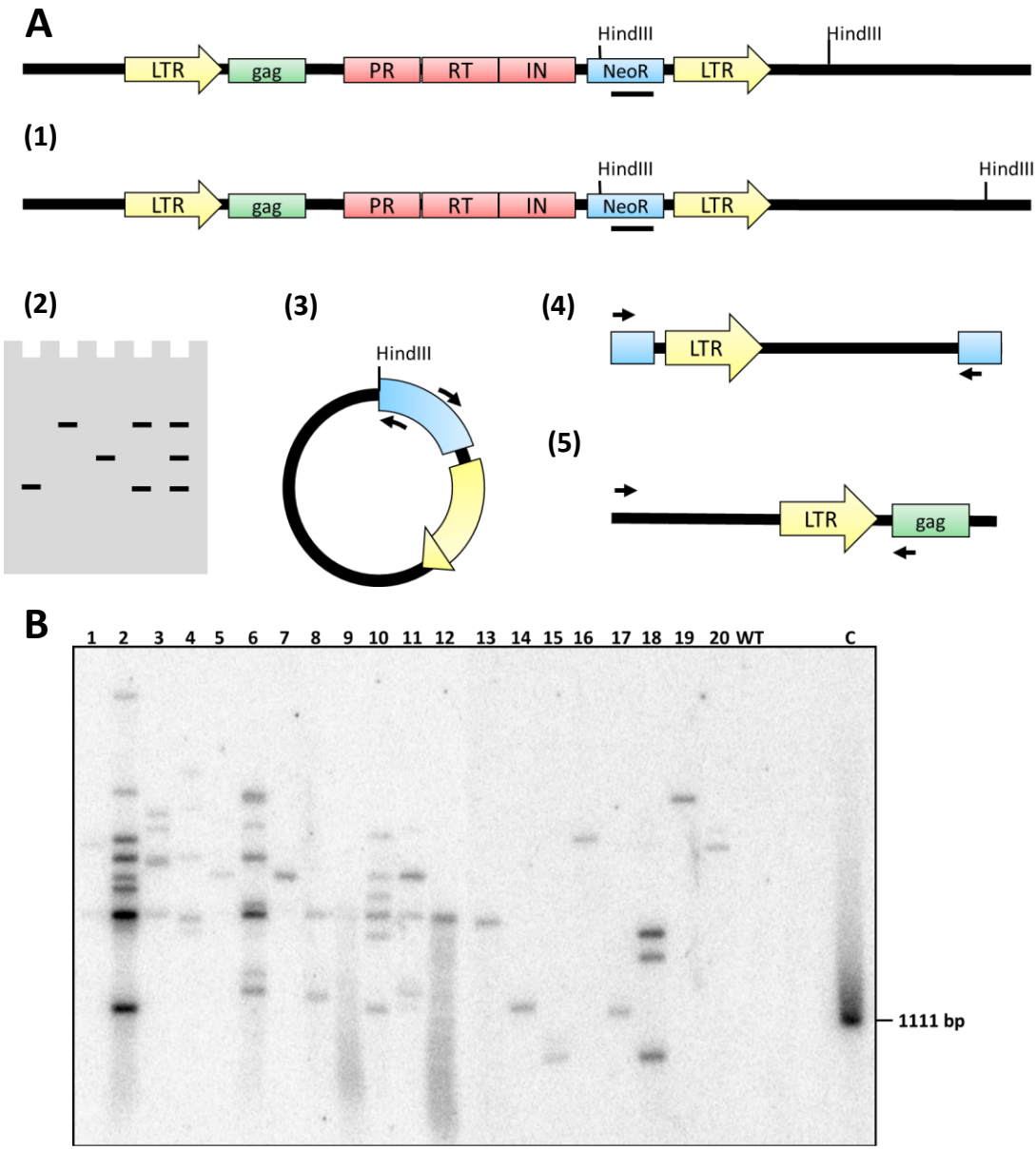


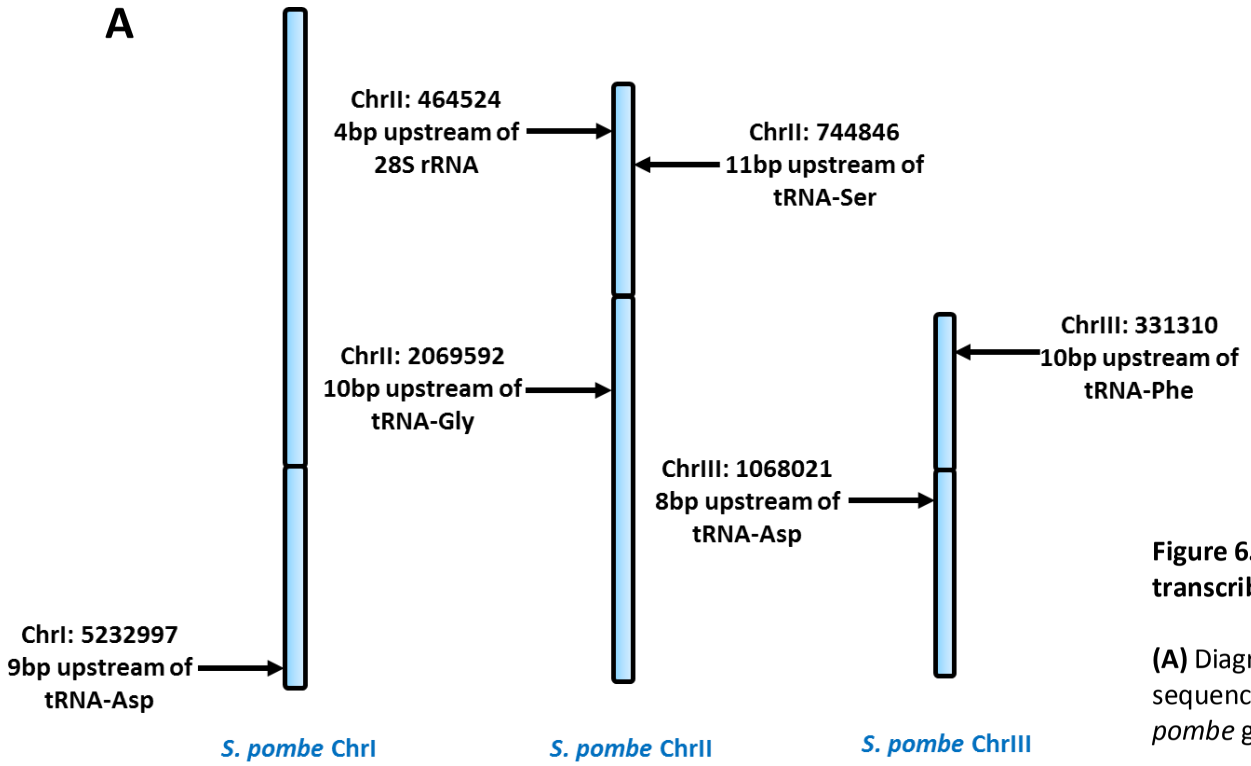
Figure 6.2 – Overview of the Tj1 sequencing strategy

(A) Genomic DNA extracted from G418^{R} colonies was digested with HindIII, and this DNA was then used for Southern blotting. The precise location of the Tj1 integration event determined the size of the HindIII fragment (1), thus a ^{32}P-labelled PCR product spanning the end of the *neoR* cassette was used (black bar) to probe the membrane (2). After identification of clones carrying single Tj1 insertions, this HindIII genomic DNA was subjected to intramolecular ligation and inverse PCR (3). Products were sequenced using primers (black arrows) specific to the 3' end of *neoR* (4) to locate sequence downstream of the integration, primers specific to the 5' end of the integration were then designed and used for confirmation of the insertion locus (5).
(B) Southern blot showing the number of Tj1 copies in 20 individual wild-type clones. All gDNA samples were HindIII digested, pHL2861 was used as a control (Lane C).

forward strand in relation to the downstream ncRNA, and all also had 5-bp Target Site Duplications (TSD) at either end of the element (Figure 6.3B). These TSDs are a hallmark of de-novo retrotransposition, and are caused by the staggered cuts generated by the element encoded integrase (Levin and Moran, 2011). The defining characteristic of Tj1 integration loci seemed to be that most were upstream of genes transcribed by RNA Pol III; 83% of integration events occurred at these loci , whilst the other integration event occurred upstream of the RNA Pol II transcribed 28S rRNA. This may indicate that interaction with the RNA Pol III transcription machinery is a key step in the integration of Tj1 elements.

6.4 - Tj1 integration is repressed by the CENP-B homologue pathway

As Tj1 is regulated by RNAi in *S. japonicus*, I wanted to investigate whether this element could elicit a similar RNAi response in *S. pombe*, which utilises a separate CENP-B homologue mediated pathway to silence its endogenous retrotransposons. The plasmid containing Tj1-*neoR* was transformed into *S. pombe* strains containing deletions of the RNAi component *dcr1*$^+$, the histone H3 lysine 9 methyltransferase *clr4*$^+$ or the CENP-B homologue proteins *abp1*$^+$, *cbh1*$^+$ or *cbh2*$^+$. These mutants were then put through the transposition assay as described above (Figure 6.1). From this, the transposition rate of Tj1 in the various mutant backgrounds was calculated by dividing the number of colonies that grew on YES+5-FOA+G418 (those that carried integrated Tj1) by the number of colonies that grew on PMG+Ura+5-FOA (all those that had lost the



B

Insertion Locus	Target Site Duplication?
ChrI:5232997	GGTCA
ChrII:464524	AACTG
ChrII:744846	TGACA
ChrII:2069592	ATATC
ChrIII:331310	ATAAC
ChrIII:1068021	AGTCA

Figure 6.3 – Tj1 integrates upstream of RNA Pol III transcribed genes

(A) Diagram showing the integration sites of the six sequenced single-copy Tj1 insertions across the *S. pombe* genome.

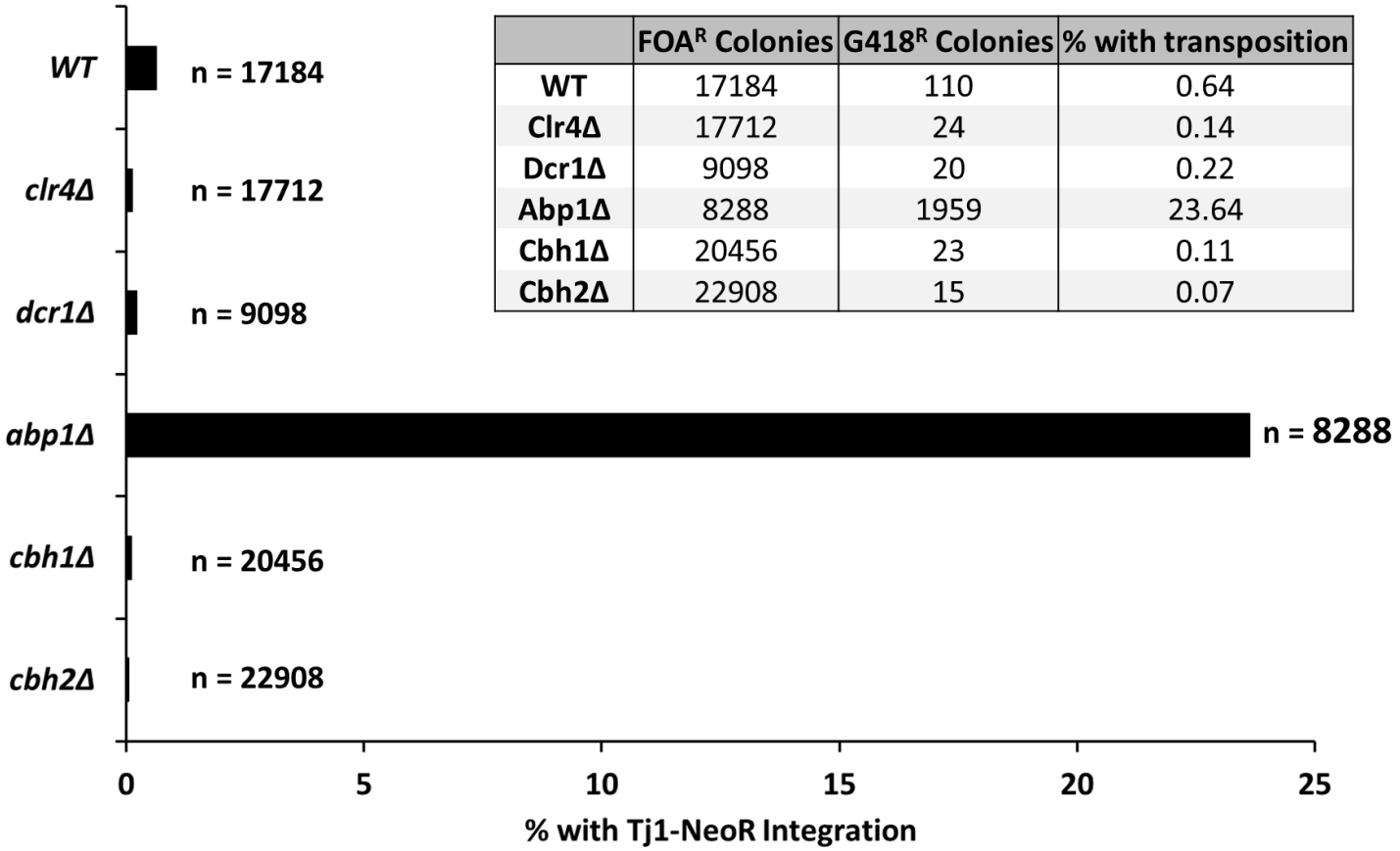
(B) 5-bp target site duplications (TSDs) identified up and downstream of the single-copy Tj1 LTR sequences

Tj1-*neoR* plasmid); this rate was then expressed as the % of cells with an integrated Tj1-*neoR* (Figure 6.4).

It was clear from this analysis that the only factor that seemed to impact Tj1 integration rates was Abp1, the core protein component of the CENP-B homologue Host Genome Surveillance pathway. In the strain carrying a deletion of Abp1, 23.64% of cells had a subsequent Tj1-*neoR* integration event, compared to 0.64% in wild-type. In contrast, deletion of the RNAi factor Dcr1, and the H3K9 methyltransferase Clr4, or the CENP-B homologue factors Cbh1 and Cbh2 had no impact on the rate of Tj1 integration, indicating that it was Abp1 alone that acted to suppress the integration of Tj1 into the *S. pombe* genome.

6.5 – A copy of Tj1 integrated upstream of the 28S rRNA is partially transcriptionally repressed by Abp1

In *S. pombe,* Abp1 acts co-operatively with Cbh1 and Cbh2 to transcriptionally repress the endogenous Tf2 and exogenous Tf1 elements (Cam et al., 2008). To ascertain whether these CENP-B homologues were capable of transcriptionally silencing an integrated copy of Tj1, a wild type strain carrying Tj1-NeoR integrated 4bp upstream of a 28S rRNA gene on ChrII:464524 was crossed to strains carrying single deletions of each of the CENP-B homologues, *abp1*$^{+}$, *cbh1*$^{+}$ and *cbh2*$^{+}$, as well as the core RNAi protein *dcr1*$^{+}$ and the H3K9 methyltransferase *clr4*$^{+}$. This allowed me to separate whether silencing of Tj1-neoR in *S. pombe* was enacted by the CENP-B homologues, as is the case with the endogenous Tf2, or whether it was mediated via RNAi and chromatin modification, as is the case in *S. japonicus*. Crossed strains were assessed by



	FOA^R Colonies	G418^R Colonies	% with transposition
WT	17184	110	0.64
Clr4Δ	17712	24	0.14
Dcr1Δ	9098	20	0.22
Abp1Δ	8288	1959	23.64
Cbh1Δ	20456	23	0.11
Cbh2Δ	22908	15	0.07

Figure 6.4 – Abp1 represses Tj1 integration

Graph showing the rate of Tj1 integration into the *S. pombe* genome in wild-type, *clr4Δ*, *dcr1Δ, abp1Δ, cbh1Δ* or *cbh2Δ* strains. The inset table gives the number of FOA^R colonies (those that had lost pHL2861) and the number of G418^R colonies (those that subsequently carried a genomically integrated copy of Tj1)

Southern blot to ensure that that they still carried single copies of Tj1-neoR, as any changes in Tj1 copy number may influence transcript levels in the resulting strain (Figure 6.5A). Tj1-neoR was found to exist as a single copy element in all crossed strains, except for one *abp1Δ* clone; this indicated that the Abp1 protein may play a role in repressing mobilisation of Tj1 once integrated. For downstream analysis this clone was excluded, as comparing single to multiple insertion events would skew assessment of transcript levels. qPCR was performed using primers specific for the Tj1 LTR or ORF; primers for the Tf2 LTR and ORF were also used as internal controls to monitor silencing of the endogenous *S. pombe* retroelements. From this it was evident that the integrated copy of Tj1 was not fully transcriptionally silenced when compared to the endogenous Tf2 elements. Only deletion of Abp1 caused any change in transcript levels, increasing the amount of Tj1 transcript by around 1.5-2.0-fold. This was in sharp contrast to Tf2, which accumulated transcript to levels between 30 and 100-fold greater than wild-type in the absence of Abp1 (Figure 6.5B). These results indicate that transcriptional silencing of retrotransposons is mediated mainly by Abp1 in *S. pombe,* however this repression is far less efficient for the exogenous Tj1 element than for the endogenous Tf2.

6.6 – Partial transcriptional repression is mediated by Abp1 at all Tj1 insertion loci

In order to assess whether this partial repression of Tj1-neoR integrated upstream of the 28S rRNA was due to locus specific effects, the 5 other wild-type strains carrying single copy insertions of Tj1 were crossed to a strain carrying a deletion at the *abp1*$^{+}$ locus. The resultant strains were confirmed to

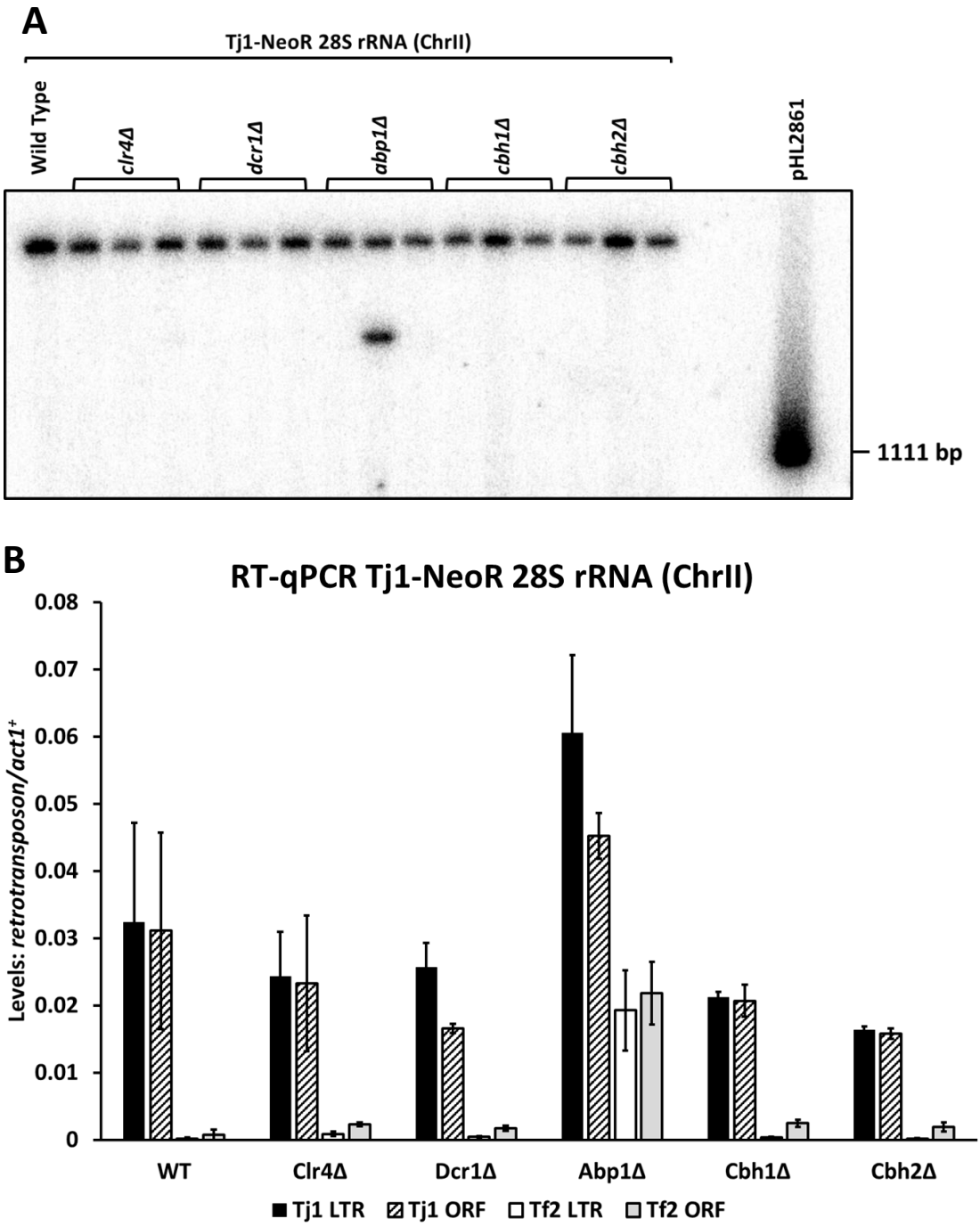


Figure 6.5 – Tj1 integrated upstream of the 28SrRNA is partially transcriptionally silenced by Abp1

(A) Southern blot analysis of the progeny of crosses between mutant strains and wild-type strain containing Tj1 integrated upstream of the 28S rRNA (ChrII). For each cross, three progeny clones were screened by Southern Blot

(B) RT-qPCR assay to assess the accumulation of Tj1 and Tf2 retrotransposon derived transcript in the wild-type and mutant backgrounds

carry single copies of Tj1 by southern blot (Figure 6.6), before Tj1 transcript levels were assessed by qPCR.

In wild-type cells the transcript levels of Tj1 did not vary greatly between different integration sites, however transcript levels for Tj1 were consistently much greater than for Tf2, which is known to be transcriptionally silenced in wild-type *S. pombe*. Tf2 has 13 copies, compared to the single copy of Tj1 integrated, so this indicates that transcriptional silencing of Tj1 in a wild-type background may not be as efficient as for the endogenous Tf2. Upon deletion of Abp1, levels of both Tj1 and Tf2 increased, however the extent of the increase was markedly different. Levels of Tj1 transcript increased roughly 3-4 fold over wild type, whilst levels of Tf2 transcript increased up to 238-fold over wild type (Figure 6.7). This increase in Tj1 transcript was due to legitimate transcriptional de-repression and not an increase in copy number, as all *abp1Δ* strains still only carried a single copy of Tj1 as assessed by Southern blot (Figure 6.6). These results indicated that the transcriptional silencing of Tj1 mediated by Abp1 in *S. pombe* is far less efficient than at Tf2 for all exogenous Tj1 elements, regardless of the Tj1 integration site

Although Tj1 and Tf2 are fairly closely related retrotransposons (see Figure 5.4, 5.5 and 5.6) I wanted to investigate why Abp1 may transcriptionally repress one more strongly than the other. Previous work has shown that the LTRs of Tf1 and Tf2 contain two conserved motifs, termed A1 and A2, which play a role in the recruitment of Abp1 (Lorenz et al., 2012). In order to see whether these motifs were present in the LTR of Tj1, I aligned the Long Terminal Repeat sequences of Tj1, Tf1 and Tf2 using Clustal Omega (Sievers et al., 2011) (Figure 6.8). From this, it was evident that both the A1 motif (TAATATAATA), which

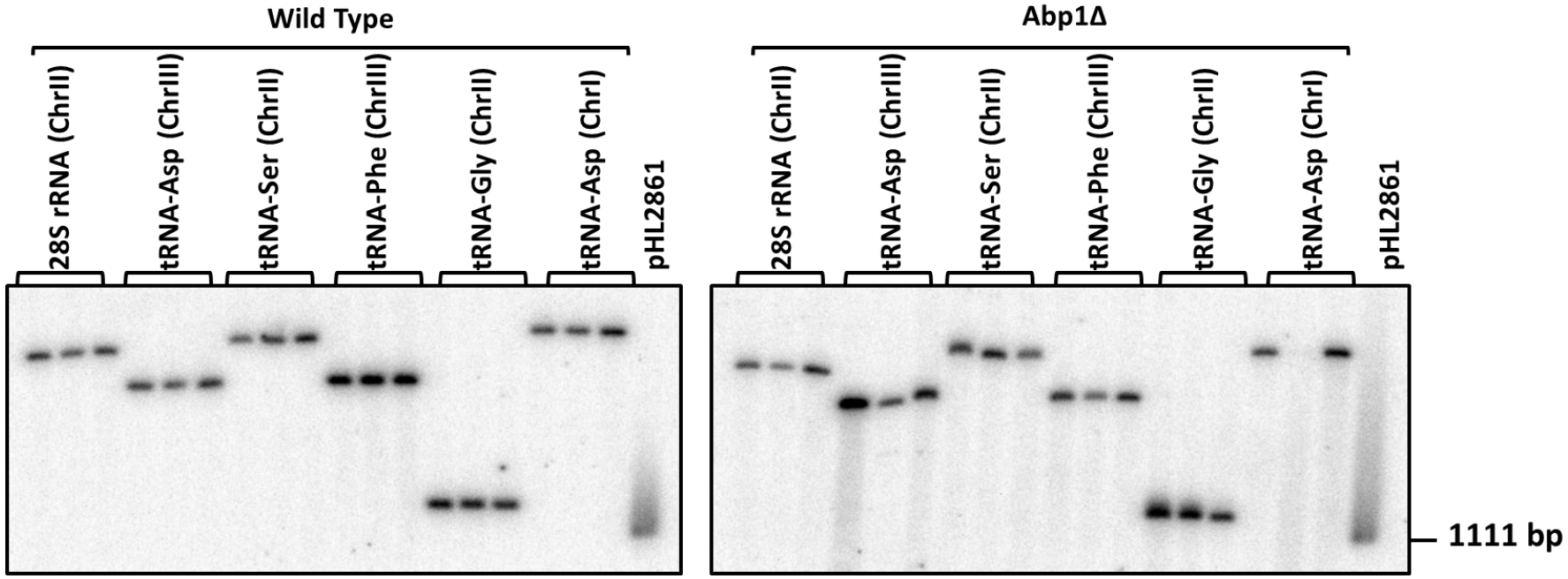


Figure 6.6 – Confirmation that single copy Tj1 integrants crossed into an *abp1Δ* background remained as single copies

Southern blot analysis of the progeny of crosses between wild-type strains containing the previously sequenced Tj1 single-copy integrants and an *abp1Δ* strain. Wild-type parental strains carrying the single-copy insertions were assayed in the left panel, while three progeny clones for each cross were analysed in the right panel. All DNA was digested using HindIII

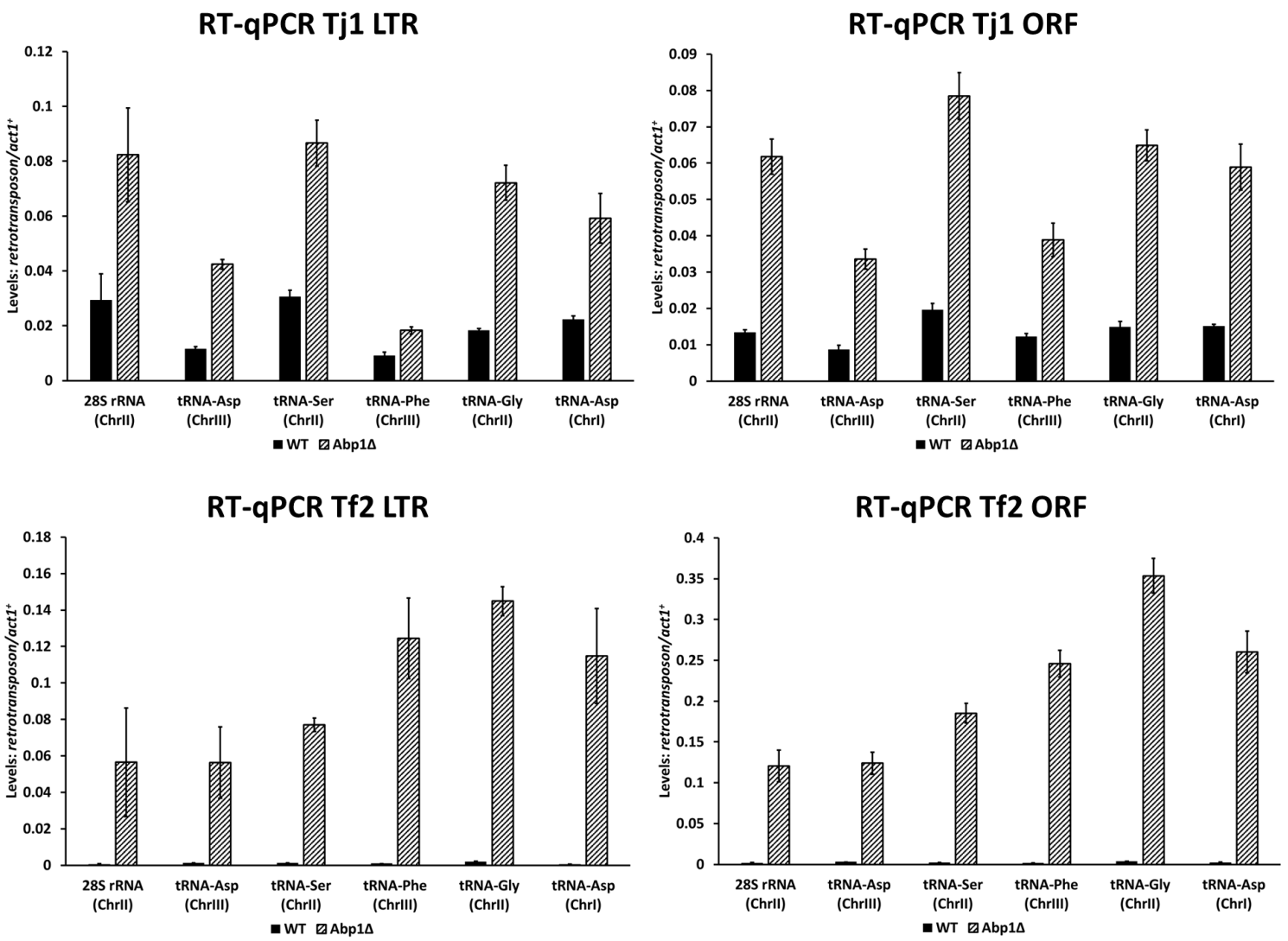


Figure 6.7 – Tj1 is partially silenced by Abp1, regardless of integration locus

RT-qPCR assay to assess the accumulation of Tj1 and Tf2 retrotransposon derived transcript in the wild-type and *abp1Δ* mutant backgrounds

has been shown to be critical for Abp1 binding at Tf2 (Lorenz et al., 2012), and the A2 motif (TAATACAATA) were not-conserved in Tj1 (Figure 6.8). The LTR of Tj1 did however carry a related motif (CAATAAGATA), which differed from A1 and A2 by 3 nucleotides. It is possible that Abp1 could bind this motif with lower affinity that the more AT-rich A1 and A2 sequences, thus mediating partial transcriptional silencing.

Figure 6.8 – The LTR of Tj1 does not contain the conserved A1 and A2 motifs required for Abp1 binding to Tf1 and Tf2 in *S. pombe*

Sequence alignment of the Tj1, Tf1 and Tf2 LTR sequences. Alignments were performed using Clustal Omega (Sievers et al., 2011), and asterisks below the sequence denote the position of conserved residues. Red boxes indicate the position of the conserved A1 and A2 motifs (Lorenz et al., 2012).

6.7 – Discussion

In this chapter I have presented the work I undertook to characterise the behaviour of the *S. japonicus* retrotransposon Tj1 when ectopically introduced into *S. pombe*.

I found that this element is able to mobilise, and integration into the genome occurs via de novo retrotransposition. This retrotransposition occurs specifically upstream of elements transcribed by RNA Pol III, such as tRNA genes. This pattern is very similar to that exhibited by the Ty1 and Ty3 retrotransposons of *S. cerevisiae,* which both integrate upstream of RNA Pol III transcribed genes, with differing specificity. Ty1 integrates within 100-700bp upstream of target genes (Devine and Boeke, 1996) and is mainly guided via interactions with the RNA Pol III subunit AC40 (Bridier-Nahmias et al., 2015). Ty3 exhibits a more specific integration pattern, within 1-4 nucleotides of the transcription start site (TSS) of the downstream gene, which has been shown to be determined by interactions between the Ty3 integrase and the TFIIIB transcription factor (Aye et al., 2001; Kirchner et al., 1995; Yieh et al., 2000). It is therefore possible that the integration of Tj1 upstream of RNA Pol III transcribed genes proceeds in a similar manner, via interactions between the Tj1 integration complex and the Pol III transcription machinery, however the exact factor that may mediate this integration is still to be elucidated.

Whilst this work was being undertaken, the lab of Henry Levin published a similar dataset that characterised 17 Tj1 integration events in *S. pombe* (Guo et al., 2015). This study found a very similar pattern of integration, although no sites were shared between the data they generated and the data I generated.

Thus my data, which increases the number of integration sites profiled by ~40%, further strengthens the conclusion that Tj1 integrates with a preference for RNA Pol III genes. This also agrees with the proposed model of centromere evolution in *S. japonicus,* whereby Tj1 elements were targeted to the pericentromeric regions, via integration upstream of existing tRNA genes (Guo et al., 2015).

To expand on this work, I looked at the method of repression employed by *S. pombe* when faced with the exogenous Tj1 element. Integration of Tj1 was strongly repressed by Abp1, a component of the CENP-B homologue pathway that *S. pombe* utilises to repress its endogenous transposable elements. In contrast, deletion of RNAi factors that are expected to control retotransposition in *S. japonicus* appeared to have no effect on mobilisation. To see whether Tj1, once integrated, was transcriptionally repressed in the same way as endogenous Tf2 elements, I assessed transcript levels for a copy of Tj1 integrated upstream of the 28S rRNA in wild-type, *clr4Δ*, *dcr1Δ*, *abp1Δ*, *cbh1Δ* and *cbh2Δ* backgrounds. From this I found that only Abp1 acts to transcriptionally repress Tj1, however not to the same extent as Tf2. This pattern was observed at all single copies of Tj1 in *S. pombe*, regardless of the integration site. This partial repression of Tj1 transcription by Abp1 may be due to the location of Tj1 insertions; as tRNA/rRNA genes are very highly expressed, the presence of the transcriptional machinery at these loci may preclude clustering into transcriptionally silent 'Tf bodies' by Abp1. It may also be down to the element itself, which is missing the conserved A1 motif that has been shown to be essential for Abp1 binding to LTRs in *S. pombe*. Tj1 does however contain a sequence related to the conserved A1 and A2 regions, which

may be conserved enough to bind Abp1 more weakly that Tf2, thus mediating weaker transcriptional silencing.

Taken together, these results indicated that the retrotransposons of *S. japonicus* are not inherently substrates for the RNAi pathway based on sequence or structure alone; rather the mechanism of retrotransposon regulation depends solely upon the organism in which the element is being silenced, and the pathways that have evolved to cope with mobile elements. Although Tj1 was utilised for this study, subsequent work on the regulation of endogenous elements in *S. japonicus* has revealed that this element is fairly inert; although it is modestly transcriptionally upregulated in the absence of RNAi, it does not mobilise throughout the *S. japonicus* genome. There are however a number of elements that do appear to actively mobilise, such as Tj2, Tj7 and Tj9, thus extending this study to include those elements may provide more insight into the behaviour of mobilisation competent *S. japonicus* retrotransposons in *S. pombe*.

Chapter 7 – Discussion

7.1 - RNAi in *S. japonicus*: an essential process?

At the outset of this study I aimed to dissect the function of the RNAi pathway in *S. japonicus,* and whether it contributed to heterochromatin formation and retrotransposon regulation. My initial plan was to generate genetic deletions of a number of key RNAi and heterochromatin factors which would then allow me to elucidate the function of each of these factors, and would permit comparison of pathway function between other fission yeasts as well as higher eukaryotes.

In *S. pombe*, deletions of all RNAi factors are viable, and very few impact growth under normal conditions. In these mutants, both siRNAs and H3K9 methylation are largely lost with no obvious impact on organism fitness, unless challenged with external agents such as the microtubule poison TBZ (Bühler et al., 2008; Buker et al., 2007; Cam et al., 2005; Provost et al., 2002; Volpe, 2002; Volpe et al., 2003). It was therefore somewhat surprising that in *S. japonicus* deletion of homologous genes proved very challenging. No viable deletion mutants could be recovered of the Argonaute protein *ago1*$^+$, the H3K9 methyltransferase *clr4*$^+$, the RNA-dependent RNA polymerase *rdp1*$^+$, or the ARC complex subunits *arb1*$^+$ and *arb2*$^+$. In spite of this inability to delete most of the targeted RNAi and heterochromatin genes, two individual isolates bearing deletions in the ribonuclease *dcr1*$^+$ were recovered. The recovery of these mutants represented very rare events, indicating that the tolerance of *S. japonicus* for this mutation may have arisen due to spontaneous suppressor mutations elsewhere in the genome.

In lieu of functional deletions, a number of core RNAi and heterochromatin proteins were genetically tagged. As in *S. pombe*, tagged versions of *chp1*$^+$,

rik1^{+} and *stc1*$^{+}$ were well tolerated (Bayne et al., 2010; Horn et al., 2005; Sadaie et al., 2004), and did not impact functionality of these genes. This ability to tag but not delete core RNAi and heterochromatin genes opens up the possibility that these genes may be required for viability in *S. japonicus*, unlike in *S. pombe*. This then begs the question as to why these factors may be required, and why deletion of these genes could possibly be lethal.

As RNAi is hypothesised to regulate retrotransposons in *S. japonicus* (Rhind et al., 2011; this study) it is possible that disruption of the RNAi pathway could lead to massive deregulation of retrotransposons. These elements might then mobilise throughout the genome, altering the coding sequence or transcriptional pattern of other genes. There is precedent for this, as Tf1 elements in *S. pombe* have been shown to impact gene expression via mobilisation into gene promoters (Feng et al., 2013; Guo and Levin, 2010). There are also examples of transposon mobilisations that may cause cancers, via mobilisation into the coding sequence of key tumour suppressors in humans (Goodier and Kazazian, 2008; Iskow et al., 2010; Kinzler et al., 1992). Another mechanism by which mobile element deregulation could impact organism fitness occurs when transposon derived proteins are targeted to specific genomic loci and cause DNA cleavage without transposon insertion; this persistent DNA damage is then either not repaired, or repaired in a deleterious manner (Gasior et al., 2006).

In the isolated *dcr1Δ5'* mutant in *S. japonicus*, there is a global deregulation of retrotransposons. Retrotransposon-associated siRNAs are lost and transcripts accumulate, and a subset of these elements also mobilise. For one element in particular, Tj7, this mobilisation causes an increase in copy number of around

30-fold; however, viability under normal growth conditions is not largely affected. This indicates that these newly mobilised Tj7 elements do not affect transcription or cause deletions within essential *S. japonicus* genes. This loss of retrotransposon silencing but not viability argues against a model whereby deletion of RNAi and heterochromatin factors causes deleterious retrotransposition. Interestingly, the *S. japonicus* retroelements, although deregulated, still retain H3K9me2, indicating that heterochromatin is still present at these regions. It is therefore possible to separate the processes of retrotransposon transcript processing and heterochromatin maintenance at these loci. Deletion of the heterochromatin establishment factor *tri1*$^{+}$ also gave an indication that H3K9 methylation may be required for viability in *S. japonicus,* as in the presence of the HDAC inhibitor TSA, *tri1Δ* mutants exhibited a potential loss of viability. This assay was not conclusive, however repetition of this assay in combination with ChIP for H3K9me2 may allow correlation of methylation levels with viability. Together these observations support the possibility that the maintenance of heterochromatin, rather than the silencing of retrotransposons per se, is essential for viability in *S. japonicus*.

In *S. pombe*, deletion of RNAi factors and perturbation of pericentromeric heterochromatin causes chromosome segregation defects such as lagging chromosomes; however, these are sensed by the organism, which slows spindle elongation to allow these lagging chromosomes to 'catch up'. These mitotic defects are postulated to be triggered by a loss of cohesion at pericentromeres (Pidoux and Allshire, 2005), which is caused by an associated loss of the chromodomain HP1 homologue Swi6 (Bernard, 2001; Nonaka et al., 2002),

which in turn is caused by a loss of its H3K9me docking site (Nakayama et al., 2001). Swi6 is responsible for nucleating high concentrations of the cohesion protein at centromeres, which is required to ensure that sister kinetochores face away from each other, facilitating chromosome bi-orientation (Bernard, 2001). In the absence of Swi6, the cohesin proteins Rad21 and Psc3 are lost specifically from centromeres but not chromosome arms; the remaining arm cohesion is postulated to be sufficient for accurate chromosome segregation (Pidoux and Allshire, 2005). Temperature sensitive (*ts)* mutants of the *rad21*$^{+}$ and *psc3*$^{+}$ cohesin subunits exhibit residual loading of cohesin at the permissive temperature, which although not sufficient to prevent lagging chromosomes, does maintain viability. A double mutant containing *swi6Δ* and *rad21-ts* or *psc3-ts* at the permissive temperature is inviable however; it is believed that the absence of both centromeric and arm cohesin causes such a severe chromosome segregation defect that combination of these two mutants is synthetically lethal (Pidoux and Allshire, 2005). Loss of cohesion has also been proposed to be incompatible with viability in *S. cerevisiae*, where a targeted loss of cohesion on ChrIII caused a 4000-fold increase in the frequency of chromosome mis-segregation; this is significant and would be expected to cause lethality if spread across all chromosomes (Eckert et al., 2007). Loss of cohesion has also been shown to halt mitotic progression in *Drosophila* (Vass et al., 2003) and *Xenopus* (Losada et al., 1998).

The reason for the reduction in viability upon the loss of cohesion is that failure to segregate chromosomes correctly leads to the generation of aneuploid cells (Compton, 2011). Aneuploids are cells that carry an abnormal number of chromosomes, which can either be due to chromosome loss or gain. This

alteration in chromosome dosage is usually catastrophic to the cell, as either loss of essential genes or changes in gene dosage can lead to mitotic arrest and cell death (Potapova and Gorbsky, 2017).

How well aneuploidy is tolerated depends on the number of chromosomes possessed by the organism in question. For example, *S. cerevisiae*, which has 16 chromosomes, tolerates up to 5 extra chromosomes. In *S. pombe* however, which only has 3 chromosomes, the only tolerated aneuploid is the highly unstable disomy of chromosome 3; all other possible aneuploids between n and 2n are inviable (Tange et al., 2012).

In *S. pombe* the chromosome segregation defects associated with the loss of RNAi and heterochromatin only impact viability in the presence of the microtubule destabilising drug TBZ, which causes chromosome segregation defects that render cells inviable (Volpe et al., 2003). It is possible that in *S. japonicus* this reduced viability phenotype manifests in the absence of TBZ, and may be evident under normal conditions; thus perturbation of heterochromatin in *S. japonicus* may be sufficient to completely abrogate chromosome cohesion and cause catastrophic aneuploidy. The question is, what may lead to this exaggerated phenotype, which is not displayed in *S. pombe*? One possible explanation is that not all factors involved in chromosome segregation are conserved from *S. pombe* to *S. japonicus*, thus disrupting the pathway at the level of heterochromatin formation may lead to different, more severe, effects. Looking at genes involved in chromosome segregation (GO:0007059), there are 212 factors in *S. pombe*, and of these, 16 do not have identifiable homologues in *S. japonicus* (Table 7.1). A number of the genes absent from *S. japonicus* are involved in meiotic chromosome

S. pombe Gene ID	*S. pombe* Gene Name	*S. pombe* Gene Description
SPAC11H11.05c	*fta6*$^+$	Mis6-Sim4 complex Fta6
SPAC15E1.07c	*moa1*$^+$	meiotic cohesin complex associated protein (Meikin) Moa1
SPAC19G12.06c	*hta2*$^+$	histone H2A beta
SPAC27D7.05c	*apc14*$^+$	anaphase-promoting complex subunit Apc14
SPAC3G9.01	*nsk1*$^+$	Clp1-interacting, microtubule plus-end binding Nsk1
SPAC4D7.07c	*csi2*$^+$	mitotic chromosome segregation protein Csi2
SPAC5D6.08c	*mes1*$^+$	meiosis II protein Mes1
SPAC9E9.10c	*cbh1*$^+$	CENP-B homolog Cbh1
SPBC1105.04c	*abp1*$^+$	CENP-B homolog Abp1
SPBC14F5.12c	*cbh2*$^+$	CENP-B homolog Cbh2
SPBC27.02c	*ask1*$^+$	DASH complex subunit Ask1
SPBC27B12.02	*mis19*$^+$	centromere protein Mis19/Eic1
SPBC2G2.14	*csi1*$^+$	mitotic chromosome segregation protein Csi1
SPBC776.16	*mis20*$^+$	centromere protein Mis20/Eic2
SPBP35G2.03c	*sgo1*$^+$	inner centromere protein, shugoshin, Sgo1
SPCC1753.03c	*rec7*$^+$	meiotic recombination protein Rec7

Table 7.1 – Table of *S. pombe* chromosome segregation genes that have no *S. japonicus* homologue

Genes annotated to be involved in chromosome segregation (GO:0007059) in *S. pombe* were searched against the *S. japonicus* genome to find homologues. Those genes that are present in *S. pombe* but not *S. japonicus* are listed above.

segregation in *S. pombe*, such as *mes1*$^{+}$, *rec7*$^{+}$ and *mei2*$^{+}$ (Kimata et al., 2011; Sakuno et al., 2011; Watanabe and Yamamoto, 1994). As any heterochromatin deficient *S. japonicus* cells constructed by transformation would only undergo mitosis, it is unlikely that the absence of a meiosis specific factor could cause the apparent exacerbated viability phenotype observed in this yeast. Two of the non-conserved factors, *csi1*$^{+}$ and *csi2*$^{+}$, are involved in microtubule dynamics, and play a role in the formation of the mitotic spindle and its interaction with the kinetochore (Costa et al., 2014; Hou et al., 2012); also absent in *S. japonicus* are factors involved in CENP-A^{Cnp1} deposition, such as *fta6*$^{+}$, *mis19*$^{+}$ and *mis20*$^{+}$ (Shiroiwa et al., 2011; Subramanian et al., 2014). The CENP-B homologue genes *abp1*$^{+}$, *cbh1*$^{+}$ and *cbh2*$^{+}$ are predicted to have evolved after the evolutionary split between *S. japonicus* and *S. pombe* (Cam et al., 2008; Rhind et al., 2011); as well as silencing retrotransposons in *S. pombe* (Cam et al., 2008), these also play a role in mitotic chromosome segregation. Double mutants of *abp1Δ/cbh1Δ/cbh2Δ* display gross chromosome segregation defects (Irelan et al., 2001); whilst triple mutants of the CENP-B homologues cannot be recovered in *S. pombe,* conditional deletions of *abp1*$^{+}$ in a *cbh1Δ cbh2Δ* background gives rise to elongated cells that arrest during the cell cycle (Locovei et al., 2006). This indicates that these factors have evolved to play a key role in chromosome segregation in *S. pombe*.

The absence of a number of factors related to chromosome segregation suggests that *S. japonicus* may have evolved a kinetochore or mitotic chromosome segregation pathway distinct from that in *S. pombe*, in terms of composition, structure and formation. This may explain why it has thus far been impossible to perturb heterochromatin in *S. japonicus*, as methylated

H3K9, or possibly the proteins that recognise this mark such as Swi6 or Chp2, may have evolved to play a more essential role in mitotic chromosome segregation in this species.

Up to this point, evidence that heterochromatin maintaining genes are essential in *S. japonicus* is purely circumstantial; in order to prove this hypothesis I will have to expand the genetic toolbox available in *S. japonicus*. In *S. pombe*, demonstrating that a gene is essential for viability is relatively straightforward. First, one copy of the gene of interest (*goi*$^{+}$*)* is knocked out with a resistance cassette in a diploid. This diploid is then sporulated and the asci dissected; failure to generate a resistant haploid indicates that the *goi*$^{+}$ is essential. Unfortunately this system is not applicable in *S. japonicus*, since although stable diploids can be generated via interallelic complementation of two adenine auxotrophs (in an analogous way to the ade6-210/216 system in *S. pombe*), these diploids cannot be transformed (Furuya and Niki, 2011). To get around this issue, genetic manipulation will have to be employed. One way to prove that these genes are essential is to incorporate the ORF of the *goi*$^{+}$ at an ectopic locus under the control of a regulateable promoter, then knock out the endogenous copy. Shutting off transcription of the ectopic gene copy should then result in cell death, if the gene is indeed required for viability. This approach has been successfully employed to show that CaDcr1 is essential in the fungal pathogen *C. albicans* (Bernstein et al., 2012). Initial work (data not shown) has indicated that the *S. japonicus nmt1*$^{+}$ locus may be an ideal target for ectopic gene integration, as expression can be repressed by over 350-fold in the presence of thiamine; this repression is much tighter than achieved when using plasmid-born *nmt1*$^{+}$ promoter elements in *S. pombe* (Basi et al.,

1993). This difference is probably due in part to the presence of long range regulatory elements not present when using the minimal promoter, as is the case with the *S. pombe urg1*$^{+}$ promoter (Watson et al., 2011; Watt et al., 2008). Another similar approach would be to integrate the same regulatable promoter upstream of potential genes of interest, thus modulating their expression in response to thiamine. The disadvantage of these transcriptional repression systems is that they are inherently leaky, and not all genes respond to transcriptional repression, especially those with endogenously low-levels. There is also the issue that ectopic promoters are usually very strong, causing massive overexpression and potential side-effects, however this can be addressed via the construction of attenuated promoters, such as nmt41 and nmt81 in *S. pombe* (Basi et al., 1993). Another system that is of interest is the Anchor-Away (AA) technique, which relies on the heterodimerisation of the human FKBP12 domain to the FRB domain of human mTOR that occurs in the presence of rapamycin (Haruki et al., 2008). In this system, the FRB domain is fused to the *goi*$^{+}$*,* whilst FKBP12 domain is fused to a highly expressed cytoplasmic protein 'anchor'. Upon induction of binding, the cytoplasmic protein physically restricts the FRB tagged protein from entering the nucleus, thus anchoring it away from its site of action (Ding et al., 2014). This would be appropriate for use with RNAi and heterochromatin proteins, as these factors are postulated to work in the nucleus.

As I have hypothesised that RNAi may have evolved to play a more essential role in *S. japonicus,* it is possible that factors within this pathway have different or additional functions to those described in *S. pombe*. To investigate this, epitope tagged versions of conserved RNAi and heterochromatin proteins

could be utilised for IP mass spectrometry (IP-MS) analysis, in order to evaluate their interacting partners. This analysis may reveal new, uncharacterised complexes that may go some way to explain why it has not yet been possible to isolate deletion mutants for a number of these key factors.

7.2 - Dcr1 and the regulation of retrotransposons in *S. japonicus*

The fact that most tested RNAi and heterochromatin factors could not be deleted, has led me to develop the hypothesis that the RNAi pathway and the formation of pericentromeric heterochromatin is an essential process in *S. japonicus*. Despite this assertion, two deletion mutants of the sole Dicer ribonuclease *dcr1*$^{+}$ have been recovered, and both of these mutants are viable. In *S. pombe* Dcr1 is an essential component of the RNAi pathway and is responsible for generating siRNAs from target loci, which then feed into the Ago1-containing effector complex RITS to direct heterochromatin formation (Motamedi et al., 2004). The recovery of the *dcr1*$^{+}$ deletions is somewhat surprising, as if the RNAi pathway truly is essential in *S. japonicus*, it would be expected that the gene responsible for generating siRNAs would also be essential. This leaves a number of possibilities as to why these two *dcr1*$^{+}$ deletion mutants are viable. The first is that Dcr1 functions differently in *S. japonicus* to *S. pombe*, potentially acting entirely in a post-transcriptional capacity; the second is that Dcr1 independent small RNAs are sufficient to maintain heterochromatin in *S. japonicus*; the third is that suppressor mutations have arisen in these strains to maintain heterochromatin in the absence of Dcr1. I will address each of these possibilities below.

The fact that deletion of *dcr1*$^{+}$ does not seem to reduce H3K9me2 levels at retrotransposons, whilst retrotransposon transcript levels increase and levels of siRNAs is reduced, may indicate that Dcr1 functions differently in *S. japonicus* when compared to *S. pombe*. Specifically, it would imply that the generation of siRNAs is dispensable for the maintenance of heterochromatin, and that Dcr1 may function to affect post transcriptional gene silencing (PTGS). Although common in other eukaryotes, this mechanism of silencing is not widely employed in *S. pombe*; instead RNAi functions to silence centromeric repeats both transcriptionally via H3K9 methylation and co-transcriptionally via RNAi mediated degradation of heterochromatin associated transcripts (Bühler et al., 2006; Volpe, 2002).

In higher eukaryotes, Dicer proteins can generate siRNAs or miRNAs that get loaded on to Argonaute-containing RISC complexes, which then go on to silence transcripts via Argonaute-mediated cleavage or by directly blocking translation (Martienssen and Moazed, 2015). In these species, the Dicer protein acts alongside an accessory dsRNA binding protein such as R2D2 (Drosophila) or TRBP2 (human), to promote formation of a mature RISC, which can enact PTGS in the cytoplasm (Czech and Hannon, 2011). If Dcr1 does function post-transcriptionally in *S. japonicus*, it may be expected that it would require a partner dsRNA binding protein, which is absent from *S. pombe*. Analysis of genes in *S. japonicus* that have no detectable *S. pombe* homologues returned 468 genes, and of these there is one (SJAG_00473) that contains a dsRNA binding domain (dsRBD_dom, InterPro:IPR014720). This protein is 242 amino acids long, and contains a single dsRNA binding motif; this is similar to both R2D2, which is 311 amino acids and contains a tandem

dsRBD (Liu, 2003), and TRBP2, which is 366 amino acids and contains 3 dsRBD domains (Chendrimada et al., 2005). This gene is therefore particularly interesting for future investigations of Dcr1 function, particularly to see whether or not it physically associates with Dcr1 in *S. japonicus*.

The theory that transcriptional gene silencing (TGS) occurs independently of Dcr1 in *S. japonicus* is challenged by 2 observations. The first is that although *dcr1*$^{+}$ deletion mutants have been isolated, these are very rare events. I have shown previously that deregulation of retrotransposons does not reduce viability, thus if *dcr1*$^{+}$ does truly function exclusively post-transcriptionally, why has it not been possible to isolate more *dcr1*$^{+}$ deletion mutants? Secondly, *dcr1*$^{+}$ is the only RNAi or heterochromatin factor deletion mutant that has been successfully isolated in *S. japonicus*. If heterochromatin maintenance is in fact essential, the inability to recover *clr4*$^{+}$ deletions is understandable, as it is the sole H3K9 methyltransferase. However the inability to recover *ago1*$^{+}$, *rdp1*$^{+}$, *arb1*$^{+}$ and *arb2*$^{+}$ deletions seems to suggest that siRNAs are required for viability (and hence heterochromatin formation) in *S. japonicus*, as all of these factors are involved in either the biogenesis, processing or function of siRNAs at centromeres in *S. pombe* (Buker et al., 2007; Volpe et al., 2003).

One way that Dcr1 could be a dispensable factor for heterochromatin maintenance via RNAi would be if there were a population of *dcr1*$^{+}$ independent small RNAs that were able to feed into the RNAi pathway and direct methylation via the CLRC complex.

In *S. pombe*, a population of small RNAs are generated independently of Dcr1 via Ago1 and the 3'-5' exonuclease Tri1 (Marasovic et al., 2013). These RNA species have been shown to be generated from abundant coding transcripts,

and show a similar size profile and 5' nucleotide bias to Dcr1 dependent siRNAs. These priRNAs (primary siRNAs) are loaded on to Ago1 and target regions that are bi-directionally transcribed, such as the dg and dh centromeric repeats, for silencing via H3K9 methylation. It is postulated that these priRNAs act to establish heterochromatin domains, and the level of H3K9 methylation they induce is very low (Halic and Moazed, 2010). It is therefore possible that in *S. japonicus* similar species may exist, which could possibly maintain H3K9 methylation in the absence of a functional copy of Dcr1. These priRNAs may be able to target retrotransposon loci, due to the non-regular arrangement of these elements, with partial retrotransposons and promoter-containing solo LTRs interspersed on both strands throughout the presumed centromeric and telomeric regions, which may promote bi-directional transcription. However, analysis of both RNA-Seq and small RNA-Seq datasets does not support this hypothesis, as neither bi-directional transcripts, nor small RNA species matching the profile of priRNAs (22-23nt, 5' Uracil) were readily detected in *dcr1Δ5'* mutant cells. In *S. pombe* priRNAs are of low abundance (Halic and Moazed, 2010; Marasovic et al., 2013) and thus it may be possible that the detection of these species may be masked by the massive amounts of small RNAs generated from the mobilisation competent retroelements in *S. japonicus*; this issue will be discussed further below.

Although the *dcr1*$^{+}$ deletion and disruption mutants have been shown to true loss of function mutants, the recovery of these were undoubtedly rare events, and the efficiency of isolation was not improved even when utilising methods that have been previously shown to promote manipulation of inefficiently targeted loci (Fennessy et al., 2014). Analysis of the transcriptional profile in

the absence of Dcr1 did not reveal any changes that could have contributed to the rescue of the *dcr1Δ5'* strain, thus it was for this reason that I undertook genomic DNA re-sequencing of this mutant. This would allow me to identify any underlying alterations in coding sequence that may have resulted in either loss or gain of function mutations within other genes. From these analyses, I identified two potential candidate mutations within coding sequence that altered the amino acid composition of two separate genes. The first of these was a V328F mutation in the alpha G-protein Gpa2, which caused a potential gain of function mutation (Douglas Ivey et al., 2010). This gene is involved in signalling cascades initiated by G-protein coupled receptors (GPCRs), mainly in the process involved in the activation of adenylate cyclase in response to glucose levels (Welton and Hoffman, 2000). There is no obvious link between this process and the maintenance of heterochromatin domains, however studies which constructed an Epistasis map of the *S. pombe* genome did reveal genetic interactions between gpa2^{+} and a number of RNAi and heterochromatin factors (Ryan et al., 2012), indicating a possible functional link. There is some literature evidence that links G-protein signalling to chromatin modification, for example β-arrestin signalling in human cells can regulate nuclear histone acetylation (Kang et al., 2005), whilst activation of the AT1 receptor has been shown to cause methylation of histone H2A which then binds the Swi6 homologue HP1α (Jagannathan et al., 2010). How this may relate to the maintenance of heterochromatin in *S. japonicus* however is unclear. The second mutation was an R77W mutation in the mRNA cleavage E3 ubiquitin-protein ligase Mpe1. The function of this mutation is unknown, however Mpe1 is a conserved factor in *S. cerevisiae* and human cells that is

known to play a role in mRNA maturation (Vo et al., 2001), thus there is a potential link between this factor and RNA interference. Although neither of these mutations provide a satisfactory explanation as to why this particular *dcr1Δ5'* mutant might be viable, it may be possible to investigate whether these genomic changes are sufficient to supress *dcr1*$^{+}$ loss, by first recreating these mutations in a wild type strain, before knocking out *dcr1*$^{+}$ and assessing how efficiently this mutation is recovered. I could also sequence the genome of the other *dcr1Δ* isolate, as well as attempting to isolate and sequence more *dcr1Δ* strains, to see if a pattern emerges with regards to the potential suppressor mutations that exist to compensate for the loss of Dcr1.

Taken together, it seems as though the most feasible explanation for why *dcr1Δ* is the only RNAi or heterochromatin mutant that has been isolated is that the two recovered strains constitute unique events. In these strains other factors, possibly Gpa2 or Mpe1, function to supress the loss of this gene, with regards to viability but not retrotransposon silencing. Thus, under normal conditions, Dcr1 may in fact be required for viability in *S. japonicus*.

Regardless of whether *dcr1*$^{+}$ is actually required for heterochromatin formation in *S. japonicus*, it plays a role in keeping retrotransposon transcript accumulation repressed, whilst generating siRNAs from these elements. In the absence of this factor, retrotransposons of the Tj2, Tj7 and Tj9 family appear to mobilise, whilst Tj1, Tj3, Tj4, Tj5, Tj6, Tj8 and Tj10 family elements do not increase in copy number. What determines whether a retrotransposon mobilises is unclear, however most full length elements appear to contain a number of frameshifts within the coding sequence, which may introduce premature stop codons within the polypeptide sequence, preventing the

translation of the proteins required for retrotransposition. It is not clear whether these frameshifts actually exist in vivo as the genome assembly is poor; these frameshifts also seem to be more prevalent in the later assembly versions, possibly due to the inclusion of lots of short read sequencing (Rhind et al., 2011). It is interesting to note however that for those elements that do mobilise, namely Tj2, Tj7 and Tj9, there are no complete, frameshift-free elements annotated within the *S. japonicus* genebuild SJ5. This indicates that the assembly itself is lacking, as I have shown that these elements are in fact mobilisation competent.

The mobilisation competent elements show a common set of characteristics in the absence of Dcr1; they get loaded with the centromere specific histone variant CENP-A^{Cnp1} and they give rise to discrete, sense derived, small RNA species that are distinct from canonical Dcr1 dependent siRNAs.

The reason for the loading of CENP-A^{Cnp1} on to these elements is not clear, however there are a number of possibilities. These elements may mobilise specifically into or near regions that are already loaded with CENP-A^{Cnp1} as all of the mobilisation competent elements seem to encode chromodomains, which may direct their integration into heterochromatin adjacent to an existing CENP-A domain (Levin and Moran, 2011). These naïve elements may then get loaded with this histone variant, due to the fact that CENP-A^{Cnp1} is able to direct its own loading via interactions with the CENP-A^{Cnp1} deposition machinery (Stellfox et al., 2013). A second possibility is that these elements carry some sequence specific motif that directs their loading with CENP-A^{Cnp1}, as is the case for loading of this variant at *S. cerevisiae* point centromeres (Xiao et al., 2011). This would make sense for Tj7, which is loaded with CENP-A^{Cnp1}

in the wild-type situation, thus expansion of the copy number may explain the increased loading in the mutant, however both Tj2 and Tj9 do not show this same pattern of wild-type loading. It is also possible that increased accumulation of the transcript could play a role, as in some species, such as maize and humans, retrotransposon transcript plays a structural role in the recruitment of CENP-A^{Cnp1} and the formation of the kinetochore (Chueh et al., 2009; Topp et al., 2004). Thus in the absence of Dcr1, and hence the possible co-transcriptional degradation of the transcript, the accumulated transcript could play an analogous role.

Interestingly, it is the increase in CENP-A^{Cnp1} loading that may explain the chromosome segregation defects observed in the Dcr1 mutants in the presence of TBZ, as deletion of Dcr1 does not seem to lead to perturbation of centromeric heterochromatin, which is known to cause TBZ sensitivity due to loss of sister chromatid cohesion in *S. pombe* (Bernard, 2001; Nonaka et al., 2002). The reason for the chromosome segregation defects in *S. japonicus* could possibly be due to the mobilisation of retrotransposons into chromosome arms, which then get loaded with CENP-A^{Cnp1}. As this histone variant is known to define the position of the centromere (Stellfox et al., 2013), it is possible that this could nucleate a neocentromere, causing multipoint attachment of the mitotic spindle to both centromeres and loci within the chromosome arms. There is some indication that this may occur in the *S. japonicus dcr1Δ5'* mutant, as I have shown evidence of CENP-A^{Cnp1} loading in the chromosome arms, at the coding gene *fba1*$^{+}$. It is also possible that these elements mobilise back into the centromere, thus increasing the size or disturbing the array structure of this region, again interfering with kinetochore formation.

In order to assess the exact cause of these segregation defects, I would first need to confirm that the TBZ sensitivity is caused by a legitimate defect in chromosome segregation, as recent work has shown that drug sensitivity can in some cases be caused by altered expression of membrane transporter genes (Ard et al., 2014), which may potentially occur in the *dcr1Δ5'* mutant. To do this, I will need to cytologically assess chromosome segregation defects, via the visualisation of lagging chromosomes (Pidoux et al., 2000). To achieve this, mitotic cells are fixed, the DNA and microtubules are fluorescently stained, and the cells are then viewed under a fluorescent microscope. This analysis could be combined with staining for CENP-A^{Cnp1}, in order to visualise the position of CENP-A^{Cnp1} domains in the absence of Dcr1, and potentially assess where in the genome the mobile elements are inserted. This assessment of retrotransposon insertion preference could be done in a more precise way, by utilising the same method as employed in this study for Tj1, involving element expression in *S. pombe* and sequencing of insertions. This may prove challenging as the cloning of repetitive elements is problematic, however to circumvent the need for a cloned retroelement it may also be possible to directly assess insertions in *S. japonicus* bioinformatically. Programs such as RelocaTE2 (Chen et al., 2017) or LoRTE (Filee and Disdero, 2016) could be employed to analyse genomic DNA re-sequencing data to locate sites of de-novo retrotransposition, by scanning for TSDs and differences from the reference sequence. As the repetitive regions of the *S. japonicus* genome are difficult to assemble using short read sequencing data generated by this study, this method may be more useful for picking up potential insertions in to the chromosome arms, however the use of low coverage long read sequencing may

get around this issue (Filee and Disdero, 2016). It may then be possible to combine this updated genome assembly with ChIP-Seq, to assess global patterns of both CENP-A^{Cnp1} and H3K9me2 in the absence of Dcr1. This may have some caveats however, as the likelihood of the active Tj7 element mobilising into the same locus in all cells is very low, thus this genome sequencing may reveal a pattern of Tj7 insertions that represents a population of many individual mobilisation events across many individual cells. As these elements are very recently mobilised it is also likely that they will all be identical, making the mapping of sequencing reads to specific elements impossible, which may prove problematic if different elements are associated with different histone variants or modifications, as appears to be the case for Tj7 family elements.

Aside from differential CENP-A^{Cnp1} loading, mobilisation competent elements give rise to a large amount of small RNA species in the absence of Dcr1, which are not generated by those elements that do not mobilise. These RNAs are distinct from siRNAs, as they exhibit a smaller, slightly broader size profile, with an altered 5' nucleotide bias. These species also seem to be derived solely from sense transcript, and do not appear to be generated as a product of double-stranded RNA. This distinct pattern, coupled with the fact that these small RNAs are derived from mobilisation competent elements exclusively, indicates that these may be generated by a specific factor that may be coupled to the process of retrotransposition. An obvious candidate for this factor is the RNase H subunit of the reverse transcriptase enzyme. During retrotransposition, this polymerase centre containing enzyme acts to generate cDNA, using the RNA transcript as a template; this reverse transcription

occurs co-operatively with the degradation of the transcript template via the RNase H (Hughes, 2015). Much work has been carried out to assess the processivity of the reverse transcriptase enzyme, with most studies focused on the RT of HIV-1, which is a retrovirus related to retrotransposons. This RNase H has found to not be totally processive, and to function as an endonuclease in vivo, specifically cleaving the RNA component of the RNA:DNA hybrid along its length in a stepwise manner (DeStefano et al., 1991; Krug and Berger, 1989). Interestingly, the distance between the RT and RNase H centres in most retroviruses is 18bp (Sarafianos et al., 2001), which correlates with the size of the cleavage products made by HIV-1 RNase H and with the size of the fragments generated from retrotransposon transcript in the *S. japonicus* Dcr1 mutant. The fact that these small RNA species show a bias in the 5' end nucleotide indicates that these are generated by an enzyme that exhibits a cleavage site preference, and indeed both HIV-1 and M-MLV RNase H enzymes have been shown to have a preference for certain nucleotides either side of the cleavage site (Schultz and Champoux, 2008). This preference, although different in nature to that exhibited buy the small RNAs in *S. japonicus*, may suggest that the mechanism of cleavage is conserved. It may also account for the broader size profile of these Dcr1-independent species when compared to siRNAs, as the RNase H enzyme responsible for generation of these small RNA in *S. japonicus* may promote cleavage within a window (form 15-20 nucleotides) with preference for a proximal cytosine residue. This work may be the first demonstration of in vivo specificity of an LTR retrotransposon RNAse H on full length mRNA targets.

Once generated, it is not clear whether or not these Dcr1-independent, retrotransposon-derived small RNA species get loaded on to Ago1. There is precedent for the loading of single stranded RNA species onto Argonaute family proteins, such as priRNAs in *S. pombe* (Halic and Moazed, 2010; Marasovic et al., 2013) and piRNAs in animal cells (Luteijn and Ketting, 2013); however, the small RNA species generated in the S. japonicus *dcr1Δ5'* mutant do not show the typical profile associated with Argonaute loaded sRNAs (Bühler et al., 2008). Argonaute-associated small RNAs are usually 22-24nt with a 5' Uracil; this has been shown to be a structural preference exhibited by the Argonaute proteins, mediated by specific residues in the MID-domain (Frank et al., 2010). It is not clear whether it is possible for those species generated independently of Dcr1 in *S. japonicus* to be Argonaute loaded, however even if they were, this association would likely lead to silencing incompetent complexes. The loaded ssRNA species would be derived exclusively from the sense strand of retrotransposon transcript, thus they would not be able to base pair with nascent transcripts as part of a RITS or RISC complex. In order to assess whether these species are Argonaute associated, an epitope tagged version of Ago1 would need to be generated in the *dcr1Δ* mutant background, and sequencing of Ago1-associated small RNAs carried out. This would give a true indication of whether these small RNAs could be expected to direct silencing, as unless they are loaded onto an Argonaute containing complex, they are not be expected to be physiologically relevant. This sequencing strategy would also allow assessment of whether any Dcr1 independent priRNAs exist in *S. japonicus*.

From these experiments, a model emerges for the regulation of retrotransposons by Dcr1 in *S. japonicus* (Figure 7.1). In the wild-type situation, retrotransposons associated with H3K9me2 are transcribed; this mRNA is then either co-transcriptionally or post-transcriptionally degraded by Dcr1 into siRNAs, which represses retrotransposon mobilisation. In the absence of Dcr1, this degradation does not occur, thus retrotransposon transcript accumulates. This accumulation is independent of RNA Pol II occupancy and H3K9me2 status, which remain unchanged in the *dcr1Δ5'* mutant. If the full length transcript arises from a mobilisation incompetent element, it accumulates and may be non-specifically degraded via the exosome or another RNA processing pathway. If the mRNA is derived from a mobilisation competent element however, this transcript may get translated to produce the Gag protein and full length polypeptide encoding the protein components required for mobilisation. This element is assembled into Gag VLPs, where reverse transcription occurs and cDNA is made. Concurrent to this, the sense mRNA is degraded via the action of the RT RNase H, which gives rise to small RNA species that may get loaded on to Ago1. The retrotransposon cDNA is then directed to specific genomic loci via the encoded integrase, whereby it integrates, creating a new, identical copy. These new copies are then loaded with CENP-A^{Cnp1}, either within existing retrotransposon arrays, or at new sites in chromosome arms. This altered localisation of CENP-A^{Cnp1} may lead to chromosome segregation defects in the presence of TBZ (Subramanian et al., 2014). The newly inserted retrotransposable elements may then get transcribed and repeat the mobilisation process, massively altering the retrotransposon complement of this mutant.

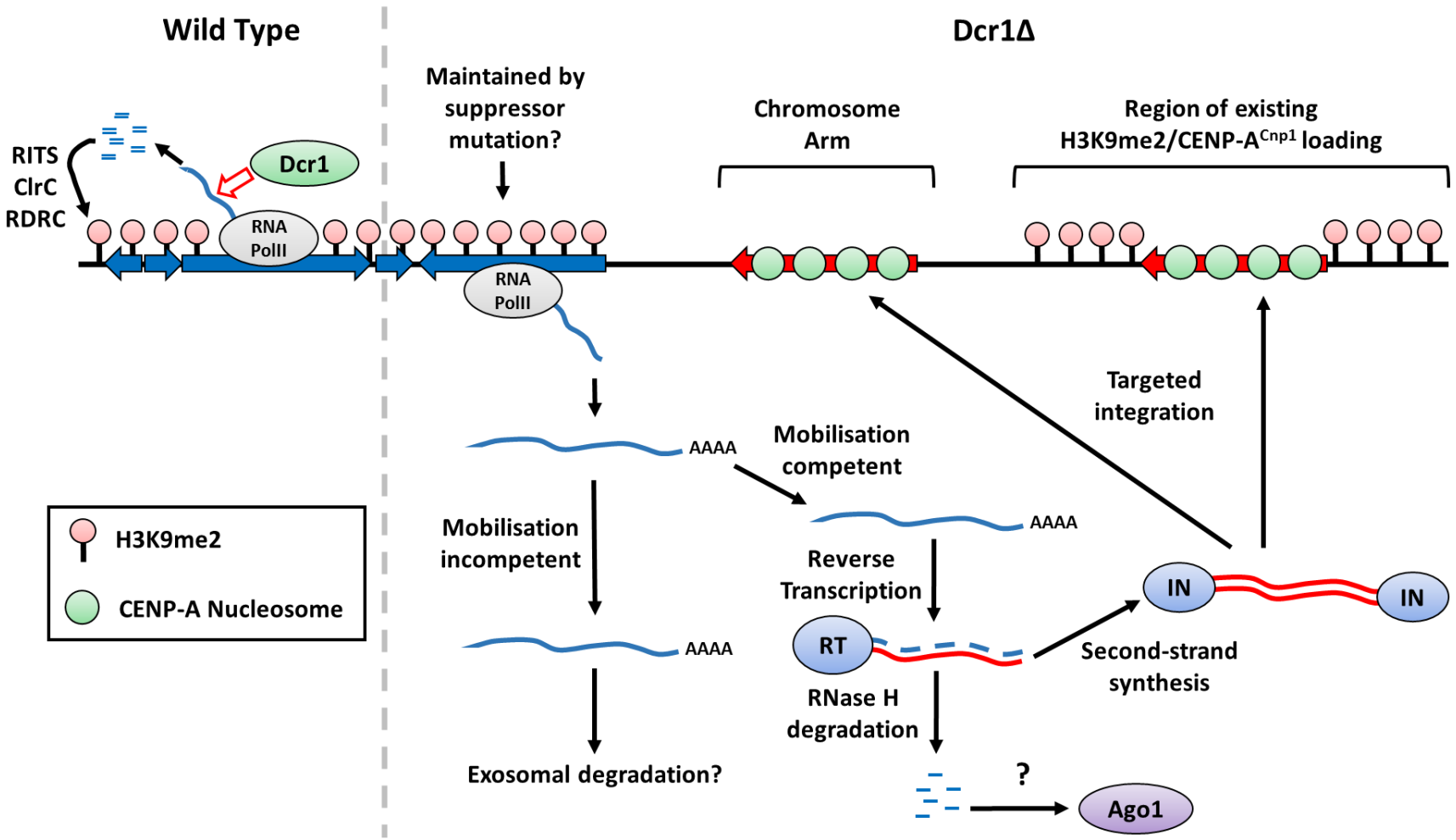


Figure 7.1 – Model for the role of Dcr1 in the regulation of retrotransposons in *S. japonicus*

7.3 - Is *S. japonicus* a useful model organism?

At the outset of this work, I aimed to not only characterise the function of the RNAi pathway in *S. japonicus*, but I also wanted to establish whether this organism would be a suitable model for the study of RNA interference as it pertains to the regulation of retrotransposons. Looking first at the structure of the presumed centromeres in this organism, in comparison to *S. pombe*, this yeast has more in common with higher eukaryotes in regards to its use of retrotransposons within pericentromeres. The pericentromeric regions of *S. japonicus* appear to be more disordered than the same loci in *S. pombe*, with interspersed retrotransposon elements encoded on both strands; a unique centromeric central core sequence is also yet to be identified in *S. japonicus*. The lack of large stretches of annotated satellite DNA, as seen in *Drosophila*, mouse and humans, make these centromeres more like maize and rice, in that they appear to be almost exclusively retrotransposon based (Plohl et al., 2014). That is not to say that such satellites do not exist in *S. japonicus*; this study has revealed a number of Dcr1 regulated loci that seem to be devoid of retrotransposon sequence, however the poor sequencing coverage of repetitive regions has thus far made assembly of complete centromeric and telomeric loci impossible. If the newly identified Dcr1-dependent regions do prove to behave as retrotransposon independent DNA satellites, this may give *S. japonicus* a centromere structure more akin to human and fly, where short repeat sequences are interspersed with retrotransposons (Plohl et al., 2014). To determine whether these regions are repetitive satellites, either homology based or de novo identification of repeats should be performed, using

bioinformatic tools such as RepeatMasker (Tarailo-Graovac and Chen, 2009) or RECON (Bao and Eddy, 2003).

A subset of *S. japonicus* centromeric retrotransposons are also loaded with CENP-A^{Cnp1}, which again is similar to rice and maize (Topp et al., 2004; Zhong et al., 2002), although such loading may also occur in human cells (Chueh et al., 2009). The basis for the differential CENP-A loading between different retrotransposon families is unclear; further investigation is needed to unpack the mechanism of CENP-A^{Cnp1} locus determination in *S. japonicus*. It is possible that the co-localisation of CENP-A^{Cnp1} with the most active of all retrotransposons may be a mechanism to nucleate neocentromeres upon mobilisation; this process is hypothesised to actively drive speciation (Chueh et al., 2009; Marshall et al., 2008).

Although these characteristics relate *S. japonicus* centromeres to higher eukaryotes perhaps more-so than *S. pombe*, there are a number of inherent issues with *S. japonicus* as a model organism. Looking specifically at the study of the RNAi pathway, the apparent inability to recover functional deletions for a majority of the key factors and effector proteins makes investigation of the function of this pathway somewhat challenging. *S. pombe* has a well-established genetic toolbox that facilitates manipulation of both essential and non-essential genes, however many of these techniques will need to be optimised before they can be effectively implemented in *S. japonicus*. For example, there are currently fewer auxotrophic markers available for use in *S. japonicus* than *S. pombe*, with auxotrophy limited to Ura- and Ade-. Other commonly used markers such as Leu- and Arg- do not work as well in *S. japonicus* as *S. pombe* (S. Oliferenko, personal communication), thus

commonly used *S. pombe* plasmids and cassettes will have to be modified. The ability to N-terminally tag protein coding genes will be key to understanding the function of RNAi in *S. japonicus* via the investigation of interacting protein factors and associated RNA species. Most of the core RNAi and heterochromatin proteins in *S. pombe* do not tolerate C-terminal epitope tags; as these genes are postulated to be essential in *S. japonicus,* the generation of N-terminally tagged clones is more challenging, as interruption of upstream sequence elements would likely have a detrimental impact on viability. Systems do exist to facilitate this, via the integration of both a selectable marker and a regulatable promoter upstream of the essential *goi*$^{+}$ coding sequence, followed by their removal via the action of the CRE/LoxP recombination system leaving an in-frame N-terminal tag (Werler et al., 2003). However as this system is optimised for use in *S. pombe*, it utilises counter-selectable markers not appropriate for *S. japonicus*, so this must be addressed. As long flanking homologies are required for efficient recombination in *S. japonicus,* the generation of integration cassettes is also challenging, especially for larger epitope tags.

As mentioned previously, the current *S. japonicus* genebuild is fairly poor, especially with regards to repetitive regions of the genome. This makes drawing conclusions about the apparent RNAi-mediated regulaton of these regions especially challenging, as both qPCR and deep sequencing experiments run into the issue of how to distinguish between differentially regulated regions that may share a high degree of sequence identity. Couple this with the fact that these repetitive regions appear to be mobile, and this adds another layer of complexity that makes *S. japonicus* a somewhat

challenging model to work with, especially in the context of RNAi and repeat silencing. The only way to begin to address this issue is to improve the genome sequencing coverage; recent technological advances with long-read sequencers, such as PacBio or the Oxford Nanopore (Goodwin et al., 2016) may be useful in future, as although these technologies generate lower sequencing depth, single long reads are able to cover more repetitive sequence, thus aiding in the placement of retrotransposons with centromeric and telomeric arrays.

In order to investigate RNAi mediated silencing in *S. japonicus* independently of the poorly characterised respective retrotransposon sequences, it may be advantageous to utilise a transgene system, whereby an ectopic dsRNA producing hairpin is introduced, which can then act to silence a transgene of complementary sequence, such as GFP (Sigova et al., 2004; Simmer et al., 2010). This system would eliminate the issue of repetitiveness, and would also allow investigation of silencing mechanisms in S. *japonicus*, free of the complications associated with retrotransposon loci. This assay format would also address whether RNAi mediated silencing in S. *japonicus* is cis-restricted as in *S. pombe* (Bühler et al., 2006), or whether this species can silence genes in trans, distant from the site of siRNA generation.

Overall, the apparent structure of *S. japonicus* centromeres, coupled with the role of RNAi in silencing these loci, and the hypothesised requirement of RNAi for viability in this species, make it an exciting organism in which to study RNAi mediated regulation. At this point however the utility of this species as a model organism is somewhat limited, first and foremost by the quality of the genome assembly and secondly by the lack of established tools for genetic manipulation. The only way that these concerns will be addressed is with more

studies; from my work, two RNA sequencing experiments have helped to increase the retrotransposon annotation by 40%, whilst I have also developed a number of strains, such as the *pku70/80*$^{+}$ deletions, which will be useful for further work. At this time, *S. japonicus* shows a lot of promise as an emerging model system, however it still has a long way to go before it challenges *S. pombe* as a favoured model for the study of centromere biology and RNA interference.

References

1. Adams, A.E.M., and Pringle, J.R. (1984). Relationship of actin and tubulin distribution to bud growth in wild-type and morphogenetic-mutant *Saccharomyces cerevisiae*. J. Cell Biol. *98*, 934–945.
2. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. *44*, W3–W10.
3. Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell *9*, 1191–1200.
4. Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207–221.
5. Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. Proc. Natl. Acad. Sci. *51*, 786–794.
6. Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500.
7. Allshire, R.C., and Ekwall, K. (2015). Epigenetics regulation of chromatin states in *Schizosaccharomyces pombe*. Cold Spring Harb Perspect. Biol. *7*, 1–25.
8. Allshire, R.C., Javerzat, J.P., Redhead, N.J., and Cranston, G. (1994). Position effect variegation at fission yeast centromeres. Cell *76*, 157–169.
9. Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., and Cranston, G. (1995). Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev. *9*, 218–233.
10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
11. Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54–59.
12. Anderson, H.E., Wardle, J., Korkut, S. V., Murton, H.E., Lopez-Maury, L., Bahler, J., and Whitehall, S.K. (2009). The Fission Yeast HIRA Histone Chaperone Is Required for Promoter Silencing and the Suppression of Cryptic Antisense Transcripts. Mol. Cell. Biol. *29*, 5158–5167.
13. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Babraham Bioinforma.
14. Aoki, K., and Niki, H. (2017). Transformation of *Schizosaccharomyces japonicus*. Cold Spring Harb. Protoc. *2017*, 996–998.
15. Aoki, K., Nakajima, R., Furuya, K., and Niki, H. (2010). Novel episomal vectors and a highly efficient transformation procedure for the fission yeast *Schizosaccharomyces japonicus*. Yeast *27*, 1049–1060.
16. Aoki, K., Furuya, K., and Niki, H. (2017). *Schizosaccharomyces japonicus*: A distinct dimorphic yeast among the fission yeasts. Cold Spring Harb. Protoc. *2017*, 963–973.
17. Ard, R., Tong, P., and Allshire, R.C. (2014). Long non-coding RNA-

mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. Nat. Commun. *5*, 5576.
18. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinforma.
19. Aye, M., Dildine, S.L., Claypool, J. a, Jourdain, S., and Sandmeyer, S.B. (2001). A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. Mol. Cell. Biol. *21*, 7839–7851.
20. Aze, A., Sannino, V., Soffientini, P., Bachi, A., and Costanzo, V. (2016). Centromeric DNA replication reconstitution reveals DNA loops and ATR checkpoint suppression. Nat. Cell Biol. *18*, 684–691.
21. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. Genes Dev. *22*, 2773–2785.
22. Bachman, N., Gelbart, M.E., Tsukiyama, T., and Boeke, J.D. (2005). TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to *S. cerevisiae* tDNAs. Genes Dev. *19*, 955–964.
23. Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. Yeast *14*, 943–951.
24. Banerjee, T., and Chakravarti, D. (2011). A peek into the complex realm of histone phosphorylation. Mol Cell Biol *31*, 4858–4873.
25. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.
26. Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E. a, Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.
27. Bao, Z., and Eddy, S.R. (2003). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. *13*, 1269–1276.
28. Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215–233.
29. Basi, G., Schmid, E., and Maundrell, K. (1993). TATA box mutations in the *Schizosaccharomyces pombe* nmt1 promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. Gene *123*, 131–136.
30. Bayne, E.H., Portoso, M., Kagansky, A., Kos-Braun, I.C., Urano, T., Ekwall, K., Alves, F., Rappsilber, J., and Allshire, R.C. (2008). Splicing Factors Facilitate RNAi-Directed Silencing in Fission Yeast. Science. *322*, 602–606.
31. Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.L., Kim, D.U., Park, H.O., Ponting, C.P., Rappsilber, J., et al. (2010). Stc1: A Critical Link between RNAi and Chromatin Modification

Required for Heterochromatin Integrity. Cell *140*, 666–677.
32. Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. 781–783.
33. Bernard, P. (2001). Requirement of Heterochromatin for Cohesion at Centromeres. Science. *294*, 2539–2542.
34. Bernstein, D. a, Vyas, V.K., Weinberg, D.E., Drinnenberg, I. a, Bartel, D.P., and Fink, G.R. (2012). Candida albicans Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. Proc. Natl. Acad. Sci. *109*, 523–528.
35. Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.
36. Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K. (2013). A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched Expression of *Caenorhabditis elegans* piRNAs. PLoS Genet. *9*.
37. Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.
38. Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.
39. Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., et al. (2015). Gene ontology consortium: Going forward. Nucleic Acids Res. *43*, D1049–D1056.
40. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Team, G. (2010). Manipulation of FASTQ data with galaxy. Bioinformatics *26*, 1783–1785.
41. Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 1–15.
42. Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. Cell *128*, 1089–1103.
43. Bridier-Nahmias, A., Tchalikian-Cosson, A., A. Baller, J., Menouni, R., Fayol, H., Flore, A., Saïb, A., Werner, M., F. Voytas, D., and Lesage, P. (2015). An RNA polymerase III subunit determines sites of retrotransposon integration. Science. *348*, 585–588.
44. Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.
45. Buchanan, L., Durand-Dubief, M., Roguev, A., Sakalar, C., Wilhelm, B., Strålfors, A., Shevchenko, A., Aasland, R., Shevchenko, A., Ekwall, K., et al. (2009). The *Schizosaccharomyces pombe* JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains. PLoS Genet. *5*.
46. Buchon, N., and Vaury, C. (2006). RNAi: a defensive RNA-silencing against viruses and transposable elements. Heredity (Edinb). *96*, 195–202.

47. Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. Cell *125*, 873–886.
48. Bühler, M., Spies, N., Bartel, D.P., and Moazed, D. (2008). TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. Nat. Struct. Mol. Biol. *15*, 1015–1023.
49. Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. Cell *150*, 317–326.
50. Buker, S.M., Iida, T., Bühler, M., Villén, J., Gygi, S.P., Nakayama, J.-I., and Moazed, D. (2007). Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. Nat. Struct. Mol. Biol. *14*, 200–207.
51. Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957–1966.
52. Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., and Grewal, S.I.S. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat. Genet. *37*, 809–819.
53. Cam, H.P., Noma, K., Ebina, H., Levin, H.L., and Grewal, S.I.S. (2008). Host genome surveillance for retrotransposons by transposon-derived proteins. Nature *451*, 431–436.
54. Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Mol. Cell *41*, 67–81.
55. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science. *298*, 1039–1043.
56. Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. J. Cell Biol. *189*, 1143–1155.
57. Castel, S.E., and Martienssen, R. a (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat. Rev. Genet. *14*, 100–112.
58. Chadwick, B.P., and Willard, H.F. (2001). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J. Cell Biol. *152*, 375–384.
59. Chalker, D.L., and Sandmeyer, S.B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. *6*, 117–128.
60. Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore–microtubule interface. Nat. Rev. Mol. Cell Biol. *9*, 33–46.
61. Cheeseman, I.M., Chappie, J.S., Wilson-Kubalek, E.M., and Desai, A. (2006). The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. Cell *127*, 983–997.

62. Cheeseman, I.M., Hori, T., Fukagawa, T., and Desai, A. (2008). KNL1 and the CENP-H/I/K complex coordinately direct kinetochore assembly in vertebrates. Mol Biol Cell *19*, 587–594.
63. Chen, E.S., Saitoh, S., Yanagida, M., and Takahashi, K. (2003). A cell cycle-regulated GATA factor promotes centromeric localization of CENP-A in fission yeast. Mol. Cell *11*, 175–187.
64. Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature *451*, 734–737.
65. Chen, J., Wrightsman, T.R., Wessler, S.R., and Stajich, J.E. (2017). RelocaTE2: a high resolution transposable element insertion site mapping tool for population resequencing. PeerJ *5*, e2942.
66. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740–744.
67. Cheng, X., Collins, R.E., and Zhang, X. (2005). Structural and Sequence Motifs of Protein (Histone) Methylation Enzymes. Annu. Rev. Biophys. Biomol. Struct. *34*, 267–294.
68. Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H.A., and Wong, L.H. (2009). LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. PLoS Genet. *5*.
69. Clustalw, U., To, C., and Multiple, D.O. (2003). ClustalW and ClustalX. Options *Chapter 2*, 1–22.
70. Collins, N., Poot, R.A., Kukimoto, I., García-Jiménez, C., Dellaire, G., and Varga-Weisz, P.D. (2002). An ACF1–ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. Nat. Genet. *32*, 627–632.
71. Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J. Biol. Chem. *280*, 5563–5570.
72. Colmenares, S.U., Buker, S.M., Buhler, M., Dlakić, M., and Moazed, D. (2007). Coupling of Double-Stranded RNA Synthesis and siRNA Generation in Fission Yeast RNAi. Mol. Cell *27*, 449–461.
73. Compton, D.A. (2011). Mechanisms of aneuploidy. Curr. Opin. Cell Biol. *23*, 109–113.
74. Conley, M.E., Partain, J.D., Norland, S.M., Shurtleff, S. a, and Kazazian, H.H. (2005). Two independent retrotransposon insertions at the same site within the coding region of BTK. Hum. Mutat. *25*, 324–325.
75. Costa, J., Fu, C., Khare, V.M., and Tran, P.T. (2014). Csi2P Modulates Microtubule Dynamics and Organizes the Bipolar Spindle for Chromosome Segregation. Mol. Biol. Cell *25*, 3900–3908.
76. Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599–601.
77. Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., et

al. (2000). Aberrant CpG-island methylation has non-random and tumour-type–specific patterns. Nat. Genet. *24*, 132–138.
78. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.
79. Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., et al. (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545–553.
80. Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet *12*, 19–31.
81. Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. Nature *453*, 798–802.
82. Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric structure of centromeric nucleosomes in interphase *Drosophila* cells. PLoS Biol. *5*, 1798–1809.
83. Dambacher, S., Deng, W., Hahn, M., Sadic, D., Fröhlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., and Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus *3*, 101–110.
84. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
85. Deans, C., and Maggert, K.A. (2015). What do you mean, “Epigenetic”? Genetics *199*, 887–896.
86. Decottignies, A. (2007). Microhomology-mediated end joining in fission yeast is repressed by Pku70 and relies on genes involved in homologous recombination. Genetics *176*, 1403–1415.
87. Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231–235.
88. DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
89. Dernburg, A.F. (2001). Here, there, and everywhere: Kinetochore function on holocentric chromosomes. J. Cell Biol. *153*.
90. DeStefano, J., Buiser, R., Mallaber, L., Bambara, R., and Fay, P. (1991). Human immunodeficiency virus reverse transcriptase displays a partially processive 3’ to 5’ endonuclease activity. J. Biol. Chem. *266*, 24295–24301.
91. Devine, S.E., and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev. *10*, 620–633.
92. Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. *35*, 41–48.

93. Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, a K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.
94. Dhillon, N., Oki, M., Szyjka, S.J., Aparicio, O.M., and Kamakaka, R.T. (2006). H2A . Z Functions To Regulate Progression through the Cell Cycle. Mol. Cell. Biol. *26*, 489–501.
95. Ding, L., Laor, D., Weisman, R., and Forsburg, S.L. (2014). Rapid regulation of nuclear proteins by rapamycin-induced translocation in fission yeast. Yeast *31*, 253–264.
96. Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. *19*, 2301–2306.
97. Djupedal, I., Kos-Braun, I.C., Mosher, R.A., Söderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., et al. (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J. *28*, 3832–3844.
98. Douglas Ivey, F., Taglia, F.X., Yang, F., Lander, M.M., Kelly, D.A., and Hoffman, C.S. (2010). Activated alleles of the *Schizosaccharomyces pombe* gpa2+ Gα gene identify residues involved in GDP-GTP exchange. Eukaryot. Cell *9*, 626–633.
99. Downs, J. a, Lowndes, N.F., and Jackson, S.P. (2000). A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. Nature *408*, 1001–1004.
100. Drinnenberg, I.A., deYoung, D., Henikoff, S., and Malik, H.S. ingh (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife *3*.
101. Dumont, M., and Fachinetti, D. (2017). DNA Sequences in Centromere Formation and Function. In Progress in Molecular and Subcellular Biology, pp. 305–336.
102. Eckert, C.A., Gravdahl, D.J., and Megee, P.C. (2007). The enhancement of pericentromeric cohesin association by conserved kinetochore components promotes high-fidelity chromosome segregation and is sensitive to microtubule-based tension. Genes Dev. *21*, 278–291.
103. Eichler, E.E., Clark, R. a, and She, X. (2004). An assessment of the sequence gaps: unfinished business in a finished human genome. Nat. Rev. Genet. *5*, 345–354.
104. Eickbush, T.H., and Jamburuthugoda, V.K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. *134*, 221–234.
105. Ekwall, K., Nimmo, E.R., Javerzat, J.P., Borgstrøm, B., Egel, R., Cranston, G., and Allshire, R. (1996). Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. J. Cell Sci. *109 (Pt 1*, 2637–2648.
106. Ellermeier, C., Higuchi, E.C., Phadnis, N., Holm, L., Geelhood, J.L., Thon, G., and Smith, G.R. (2010). RNAi and heterochromatin repress centromeric meiotic recombination. Proc. Natl. Acad. Sci. *107*, 8701–

8705.
107. Esnault, C., and Levin, H.L. (2015). The Long Terminal Repeat Retrotransposons Tf1 and Tf2 of *Schizosaccharomyces pombe*. Microbiol. Spectr. *3*, 1–13.
108. Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev. Cell *33*, 314–327.
109. Feinberg, A.P., and Vogelstein, B. (1983a). Hypomethylation of ras oncogenes in primary human cancers. Biochem. Biophys. Res. Commun. *111*, 47–54.
110. Feinberg, A.P., and Vogelstein, B. (1983b). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89–92.
111. Feng, G., Leem, Y.E., and Levin, H.L. (2013). Transposon integration enhances expression of stress response genes. Nucleic Acids Res. *41*, 775–789.
112. Fennessy, D., Grallert, A., Krapp, A., Cokoja, A., Bridge, A.J., Petersen, J., Patel, A., Tallada, V.A., Boke, E., Hodgson, B., et al. (2014). Extending the *Schizosaccharomyces pombe* molecular genetic toolbox. PLoS One *9*.
113. Filee, J., and Disdero, E. (2016). LoRTE: Detecting transposon-induced genomic variants using low coverage PacBio long read sequences. bioRxiv 73551.
114. Finnegan, D.J. (2012). Retrotransposons. Curr. Biol. *22*, 432–437.
115. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806–811.
116. Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.
117. Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.
118. Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science. *319*, 94–97.
119. Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006). The human CENP-A centromeric nucleosome-associated complex. Nat. Cell Biol. *8*, 458–469.
120. Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. Cell *137*, 472–484.
121. Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818–822.
122. Fu, J., Hettler, E., and Wickes, B.L. (2006). Split marker

transformation increases homologous integration frequency in Cryptococcus neoformans. Fungal Genet. Biol. *43*, 200–212.
123. Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of Centromere for CENP-A Recruitment by Human hMis18α, hMis18β, and M18BP1. Dev. Cell *12*, 17–30.
124. Fukagawa, T., and Earnshaw, W.C. (2014). The centromere: Chromatin foundation for the kinetochore machinery. Dev. Cell *30*, 496–508.
125. Furuya, K., and Niki, H. (2009). Isolation of heterothallic haploid and auxotrophic mutants of *Schizosaccharomyces japonicus*. Yeast *26*, 221–233.
126. Furuya, K., and Niki, H. (2010). The DNA damage checkpoint regulates a transition between yeast and hyphal growth in *Schizosaccharomyces japonicus*. Mol. Cell. Biol. *30*, 2909–2917.
127. Furuya, K., and Niki, H. (2011). Construction of diploid zygotes by interallelic complementation of ade6 in *Schizosaccharomyces japonicus*. Yeast *28*, 747–754.
128. Furuyama, T., and Henikoff, S. (2009). Centromeric Nucleosomes Induce Positive DNA Supercoils. Cell *138*, 104–113.
129. Gama-sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. *11*, 6883–6894.
130. Gao, X., Hou, Y., Ebina, H., Levin, H.L., and Voytas, D.F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res. *18*, 359–369.
131. Garabedian, M. V., Noguchi, C., Ziegler, M.A., Das, M.M., Singh, T., Harper, L.J., Leman, A.R., Khair, L., Moser, B.A., Nakamura, T.M., et al. (2012). The double-bromodomain proteins Bdf1 and Bdf2 modulate chromatin structure to regulate S-phase stress response in *Schizosaccharomyces pombe*. Genetics *190*, 487–500.
132. Garavís, M., Méndez-Lago, M., Gabelica, V., Whitehead, S.L., González, C., and Villasante, A. (2015). The structure of an endogenous *Drosophila* centromere reveals the prevalence of tandemly repeated sequences able to form i-motifs. Sci. Rep. *5*, 13307.
133. Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. Cell *145*, 410–422.
134. Gasior, S.L., Wakeman, T.P., Xu, B., and Deininger, P.L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. J. Mol. Biol. *357*, 1383–1393.
135. Gaudet, A., and Fitzgerald-Hayes, M. (1989). Mutations in CEN3 cause aberrant chromosome segregation during meiosis in *Saccharomyces cerevisiae*. Genetics *121*, 477–489.
136. Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., et al. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. Science. *320*, 1077–1081.
137. Giménez-Abián, J.F., Sumara, I., Hirota, T., Hauf, S., Gerlich, D., De

La Torre, C., Ellenberg, J., and Peters, J.M. (2004). Regulation of sister chromatid cohesion between chromosome arms. Curr. Biol. *14*, 1187–1193.
138. Glynn, E.F., Megee, P.C., Yu, H.G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. PLoS Biol. *2*.
139. Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell *140*, 678–691.
140. Gong, Z., Wu, Y., Koblizkova, A., Torres, G.A., Wang, K., Iovene, M., Neumann, P., Zhang, W., Novak, P., Buell, C.R., et al. (2012). Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. Plant Cell *24*, 3559–3574.
141. Goodier, J.L., and Kazazian, H.H. (2008). Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. Cell *135*, 23–35.
142. Goodwin, S., Mcpherson, J.D., and Mccombie, W.R. (2016). Coming of age : ten years of next- generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.
143. Gregory, R.I., Yan, K., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.
144. Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nat. Rev. Genet. *8*, 35–46.
145. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.
146. Gullerova, M., Moazed, D., and Proudfoot, N.J. (2011). Autoregulation of convergent RNAi genes in fission yeast. Genes Dev. *25*, 556–568.
147. Guo, Y., and Levin, H.L. (2010). High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. Genome Res. *20*, 239–248.
148. Guo, Y., Singh, P.K., and Levin, H.L. (2015). A long terminal repeat retrotransposon of *Schizosaccharomyces japonicus* integrates upstream of RNA pol III transcribed genes. Mob. DNA *6*, 19.
149. Guse, A., Carroll, C.W., Moree, B., Fuller, C.J., and Straight, A.F. (2011). In vitro centromere and kinetochore assembly on defined chromatin templates. Nature *477*, 354–358.
150. Haarer, B.K., and Pringle, J.R. (1987). Immunofluorescence localization of the *Saccharomyces cerevisiae* CDC12 gene product to the vicinity of the 10-nm filaments in the mother-bud neck. Mol. Cell. Biol. *7*, 3678–3687.
151. Haering, C.H., Löwe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol. Cell *9*, 773–788.
152. Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., and Löwe, J. (2004). Structure and stability of cohesin's Smc1-kleisin

interaction. Mol. Cell *15*, 951–964.
153. Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: The "H3 barcode hypothesis." Proc. Natl. Acad. Sci. *103*, 6428–6435.
154. Halic, M., and Moazed, D. (2010). Dicer-Independent Primal RNAs Trigger RNAi and Heterochromatin Formation. Cell *140*, 504–516.
155. Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. *286*, 950–2.
156. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. Cell *125*, 887–901.
157. Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E., and Kazazian, H.H. (2011). Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum. Mol. Genet. *20*, 3386–3400.
158. Hansen, K.R., Ibarra, P.T., and Thon, G. (2006). Evolutionary-conserved telomere-linked helicase genes of fission yeast are repressed by silencing factors, RNAi components and the telomere-binding protein Taz1. Nucleic Acids Res. *34*, 78–88.
159. Hardy, S., Jacques, P.É., Gévry, N., Forest, A., Fortin, M.È., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet. *5*.
160. Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. Mol. Cell *31*, 925–932.
161. Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? Microbiol. Mol. Biol. Rev. *70*, 789–829.
162. Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biol. *5*, 225.
163. Hayashi, A., Ishida, M., Kawaguchi, R., Urano, T., Murakami, Y., and Nakayama, J. -i. (2012). Heterochromatin protein 1 homologue Swi6 acts in concert with Ers1 to regulate RNAi-directed heterochromatin assembly. Proc. Natl. Acad. Sci. *109*, 6159–6164.
164. Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell *118*, 715–729.
165. He, C., Pillai, S.S., Taglini, F., Li, F., Ruan, K., Zhang, J., Wu, J., Shi, Y., and Bayne, E.H. (2013). Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. Proc. Natl. Acad. Sci. *110*, E1879-88.
166. He, H., Zhang, S., Wang, D., Hochwagen, A., and Li, F. (2016). Condensin Promotes Position Effects within Tandem DNA Repeats via the RITS Complex. Cell Rep. *14*, 1018–1024.
167. Heitz, E. (1928). Das Heterochromatin der Moose. Jahrbücher Für Wissenschaftliche Bot. *69*, 762–818.
168. Henikoff, S., and Henikoff, J.G. (2012). "Point" centromeres of

Saccharomyces harbor single centromere-specific nucleosomes. Genetics *190*, 1575–1577.
169. Henikoff, S., and Smith, M.M. (2015). Histone variants and epigenetics. Cold Spring Harb. Perspect. Biol. *7*, 1–25.
170. Hewawasam, G., Shivaraju, M., Mattingly, M., Venkatesh, S., Martin-Brown, S., Florens, L., Workman, J.L., and Gerton, J.L. (2010). Psh1 Is an E3 Ubiquitin Ligase that Targets the Centromeric Histone Variant Cse4. Mol. Cell *40*, 444–454.
171. Hickey, A., Esnault, C., Majumdar, A., Chatterjee, A.G., Iben, J.R., McQueen, P.G., Yang, A.X., Mizuguchi, T., Grewal, S.I.S., and Levin, H.L. (2015). Single-nucleotide-specific targeting of the Tf1 retrotransposon promoted by the DNA-Binding protein sap1 of *Schizosaccharomyces pombe*. Genetics *201*, 905–924.
172. Hindley, J., Phear, G., Stein, M., and Beach, D. (1987). Sucl+ encodes a predicted 13-kilodalton protein that is essential for cell viability and is directly involved in the division cycle of *Schizosaccharomyces pombe*. Mol Cell Biol *7*, 504–511.
173. Hinshaw, S.M., Makrantoni, V., Harrison, S.C., and Marston, A.L. (2017). The Kinetochore Receptor for the Cohesin Loading Complex. Cell *171*, 72–84.e13.
174. Hirano, T., Funahashi, S., Uemura, T., and Yanagida, M. (1986). Isolation and characterization of *Schizosaccharomyces pombe* cut mutants that block nuclear division but not cytokinesis. EMBO J. *5*, 2973–2979.
175. Höck, J., and Meister, G. (2008). The Argonaute protein family. Genome Biol. *9*, 210.
176. Hoff, E.F., Levin, H.L., and Boeke, J.D. (1998). *Schizosaccharomyces pombe* retrotransposon Tf2 mobilizes primarily through homologous cDNA recombination. Mol. Cell. Biol. *18*, 6839–6852.
177. Hoffman, C.S., Wood, V., and Fantes, P.A. (2015). An ancient yeast for young geneticists: A primer on the *Schizosaccharomyces pombe* model system. Genetics *201*, 403–423.
178. Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science. *187*, 226–232.
179. Holoch, D., and Moazed, D. (2015a). RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. *16*, 71–84.
180. Holoch, D., and Moazed, D. (2015b). Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. Nat. Struct. Mol. Biol. *22*, 328–335.
181. Hori, T., Okada, M., Maenaka, K., and Fukagawa, T. (2008). CENP-O class proteins form a stable complex and are required for proper kinetochore function. Mol. Biol. Cell *19*, 843–854.
182. Hori, T., Shang, W.H., Toyoda, A., Misu, S., Monma, N., Ikeo, K., Molina, O., Vargiu, G., Fujiyama, A., Kimura, H., et al. (2014). Histone H4 Lys 20 Monomethylation of the CENP-A Nucleosome Is Essential for Kinetochore Assembly. Dev. Cell *29*, 740–749.
183. Horn, P.J., Bastie, J.N., and Peterson, C.L. (2005). A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. Genes Dev. *19*, 1705–1714.

184. Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J. Biol. Chem. *175*, 315–332.
185. Hou, H., Zhou, Z., Wang, Y., Wang, J., Kallgren, S.P., Kurchuk, T., Miller, E.A., Chang, F., and Jia, S. (2012). Csi1 links centromeres to the nuclear envelope for centromere clustering. J. Cell Biol. *199*, 735–744.
186. Howman, E. V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H.A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc. Natl. Acad. Sci. *97*, 1148–1153.
187. Hughes, S.H. (2015). Reverse Transcription of Retroviruses and LTR Retrotransposons. Microbiol. Spectr. *58*, 25.
188. Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. *9*, 22–32.
189. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. Science. *293*, 834–838.
190. Irelan, J.T., Gutkin, G.I., and Clarke, L. (2001). Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. Genetics *157*, 1191–1203.
191. Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y., and Takahashi, K. (2008). Heterochromatin Integrity Affects Chromosome Reorganization After Centromere Dysfunction. Science. *321*, 1088–1091.
192. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. Cell *141*, 1253–1261.
193. Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., Katsuma, S., and Tomari, Y. (2016). Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. Cell *164*, 962–973.
194. Jacobson, R.H. (2000). Structure and Function of a Human TAFII250 Double Bromodomain Module. Science. *288*, 1422–1425.
195. Jagannathan, R., Kaveti, S., Desnoyer, R.W., Willard, B., Kinter, M., and Karnik, S.S. (2010). At1 receptor induced alterations in histone H2A reveal novel insights into GPCR control of chromatin remodeling. PLoS One *5*, 1–13.
196. Jih, G., Iglesias, N., Currie, M.A., Bhanu, N. V., Paulo, J.A., Gygi, S.P., Garcia, B.A., and Moazed, D. (2017). Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. Nature 1–26.
197. Kang, J., Shi, Y., Xiang, B., Qu, B., Su, W., Zhu, M., Zhang, M., Bao, G., Wang, F., Zhang, X., et al. (2005). A nuclear function of beta-arrestin1 in GPCR signaling: regulation of histone acetylation and gene transcription. Cell *123*, 833–847.
198. Kanoh, J., Sadaie, M., Urano, T., and Ishikawa, F. (2005). Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. Curr. Biol. *15*, 1808–1819.
199. Kapitonov, V. V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic

rolling-circle transposons. Trends Genet. *23*, 521–529.
200. Kawamata, T., Seitz, H., and Tomari, Y. (2009). Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. Nat. Struct. Mol. Biol. *16*, 953–960.
201. Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' End Formation of PIWI-Interacting RNAs In Vitro. Mol. Cell *43*, 1015–1022.
202. Kawashima, S.A., Yamagishi, Y., Honda, T., Ishiguro, K. -i., and Watanabe, Y. (2010). Phosphorylation of H2A by Bub1 Prevents Chromosomal Instability Through Localizing Shugoshin. Science. *327*, 172–177.
203. Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., et al. (2016). Ensembl Genomes 2016: More genomes, more complexity. Nucleic Acids Res. *44*, D574–D580.
204. Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.
205. Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat. Biotechnol. *28*, 617–623.
206. Kim, M.S., Kim, S.Y., Yoon, J.K., Lee, Y.W., and Bahn, Y.S. (2009). An efficient gene-disruption method in Cryptococcus neoformans by double-joint PCR with NAT-split markers. Biochem. Biophys. Res. Commun. *390*, 983–988.
207. Kimata, Y., Kitamura, K., Fenner, N., and Yamano, H. (2011). Mes1 controls the meiosis I to meiosis II transition by distinctly regulating the anaphase-promoting complex/cyclosome coactivators Fzr1/Mfr1 and Slp1 in fission yeast. Mol. Biol. Cell *22*, 1486–1494.
208. Kimura, A., Matsubara, K., and Horikoshi, M. (2005). A decade of histone acetylation: Marking eukaryotic chromosomes with specific codes. J. Biochem. *138*, 647–662.
209. Kinzler, K.W., Vogelstein, B., Horii, A., Miyoshi, Y., and Nakamura, Y. (1992). Disruption of the APC Gene by a Retrotransposal Insertion of LI Sequence in a Colon Cancer. Cancer Res. *52*, 643–645.
210. Kirchner, J., Connolly, C.M., and Sandmeyer, S.B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science. *267*, 1488–1491.
211. Kitajima, T.S., Sakuno, T., Ishiguro, K., Iemura, S., Natsume, T., Kawashima, S.A., and Watanabe, Y. (2006). Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature *441*, 46–52.
212. Klar, A.J.S. (2013). *Schizosaccharomyces japonicus* yeast poised to become a favorite experimental organism for eukaryotic research. G3 (Bethesda). *3*, 1869–1873.
213. Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. Cell *138*, 1137–1149.
214. Koch, B., Kueng, S., Ruckenbauer, C., Wendt, K.S., and Peters, J.M.

(2008). The Suv39h-HP1 histone methylation pathway is dispensable for enrichment and protection of cohesin at centromeres in mammalian cells. Chromosoma *117*, 199–210.
215. Kooistra, R., Hooykaas, P.J.J., and Steensma, H.. Y. (2004). Efficient gene targeting in Kluyveromyces lactis. Yeast *21*, 781–792.
216. Krawchuk, M.D., and Wahls, W.P. (1999). High-efficiency gene targeting in *Schizosaccharomyces pombe* using a modular, PCR-based approach with long tracts of flanking homology. Yeast *15*, 1419–1427.
217. Krueger, F. (2016). Trim Galore!
218. Krug, M.S., and Berger, S.L. (1989). Ribonuclease H activities associated with viral reverse transcriptases are endonucleases. Proc. Natl. Acad. Sci. *86*, 3539–3543.
219. Kuduvalli, P.N., Rao, J.E., and Craig, N.L. (2001). Target DNA structure plays a critical role in Tn7 transposition. EMBO J. *20*, 924–932.
220. Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in *Arabidopsis*. Cell *140*, 136–147.
221. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. msw054.
222. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.
223. Laloraya, S., Guacci, V., and Koshland, D. (2000). Chromosomal addresses of the cohesin component Mcd1p. J. Cell Biol. *151*, 1047–1056.
224. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.
225. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.
226. Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.
227. Lee, M.G., and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature *327*, 31–35.
228. Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007). Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. Cell *131*, 1084–1096.
229. Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. EMBO J. *23*, 4051–4060.
230. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69–81.
231. Levin, H.L., and Boeke, J.D. (1992). Demonstration of retrotransposition of the Tf1 element in fission yeast. EMBO J. *11*, 1145–1153.
232. Levin, H.L., and Moran, J. V. (2011). Dynamic interactions between

transposable elements and their hosts. Nat. Rev. Genet. *12*, 615–627.
233. Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F. miin (1993). Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell *75*, 1083–1093.
234. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
235. Li, B., Choulet, F., Heng, Y., Hao, W., Paux, E., Liu, Z., Yue, W., Jin, W., Feuillet, C., and Zhang, X. (2013a). Wheat centromeric retrotransposons: The new ones take a major role in centromeric structure. Plant J. *73*, 952–965.
236. Li, L., Zhang, C., and Konopka, J.B. (2012). A Candida albicans temperature-sensitive cdc12-6 mutant Identifies roles for septins in selection of sites of germ tube formation and hyphal morphogenesis. Eukaryot. Cell *11*, 1210–1218.
237. Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., et al. (2013b). An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. Mol. Cell *50*, 67–81.
238. Li, Z.H., Du, C.M., Zhong, Y.H., and Wang, T.H. (2010). Development of a highly efficient gene targeting system allowing rapid genetic manipulations in penicillium decumbens. Appl. Microbiol. Biotechnol. *87*, 1065–1076.
239. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
240. Liu, Q. (2003). R2D2, a Bridge Between the Initiation and Effector Steps of the *Drosophila* RNAi Pathway. Science. *301*, 1921–1925.
241. Locke, D.P., Hillier, L.W., Warren, W.C., Worley, K.C., Nazareth, L. V., Muzny, D.M., Yang, S.-P., Wang, Z., Chinwalla, A.T., Minx, P., et al. (2011). Comparative and demographic analysis of orang-utan genomes. Nature *469*, 529–533.
242. Locovei, A.M., Spiga, M.-G., Tanaka, K., Murakami, Y., and D'Urso, G. (2006). The CENP-B homolog, Abp1, interacts with the initiation protein Cdc23 (MCM10) and is required for efficient DNA replication in fission yeast. Cell Div. *1*, 27.
243. Lorenz, D.R., Mikheyeva, I. V., Johansen, P., Meyer, L., Berg, A., Grewal, S.I.S., and Cam, H.P. (2012). CENP-B Cooperates with Set1 in Bidirectional Transcriptional Silencing and Genome Organization of Retrotransposons. Mol. Cell. Biol. *32*, 4215–4225.
244. Losada, A., Hirano, M., and Hirano, T. (1998). Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. Genes Dev. *12*, 1986–1997.
245. Losada, A., Hirano, M., and Hirano, T. (2002). Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. Genes Dev. *16*, 3004–3016.
246. Love, M.I., Anders, S., and Huber, W. (2014). Differential analysis of count data - the DESeq2 package.
247. Lowell, J.E., and Cross, G.A.M. (2004). A variant histone H3 is enriched at telomeres in Trypanosoma brucei. J. Cell Sci. *117*, 5937–5947.

248. Luger, K., Mäder, a W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.
249. Luk, E., Vu, N.D., Patteson, K., Mizuguchi, G., Wu, W.H., Ranjan, A., Backus, J., Sen, S., Lewis, M., Bai, Y., et al. (2007). Chz1, a Nuclear Chaperone for Histone H2AZ. Mol. Cell *25*, 357–368.
250. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear Export of MicroRNA Precursors. Science. *303*, 95–98.
251. Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.
252. Lyon, M. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus L.*). Nature *192*, 963–964.
253. MacRae, I., Zhou, K., Li, F., Repic, A., Brooks, A., Cande, Z., Adams, P., and Doudna, J. (2006). Structural Basis for Double-Stranded RNA Processing by Dicer. Science. *311*, 195–198.
254. MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. Nat Struct Mol Biol *14*, 934–940.
255. Manolis, K.G., Nimmo, E.R., Hartsuiker, E., Carr, A.M., Jeggo, P.A., and Allshire, R.C. (2001). Novel functional requirements for non-homologous DNA end joining in *Schizosaccharomyces pombe*. Embo J *20*, 210–221.
256. Marasovic, M., Zocco, M., and Halic, M. (2013). Argonaute and triman generate dicer-independent prirnas and mature siRNAs to initiate heterochromatin formation. Mol. Cell *52*, 173–183.
257. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.
258. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. *45*, D200–D203.
259. Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.
260. Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H.A. (2008). Neocentromeres: New Insights into Centromere Structure, Disease Development, and Karyotype Evolution. Am. J. Hum. Genet. *82*, 261–282.
261. Martienssen, R., and Moazed, D. (2015). RNAi and heterochromatin assembly. Cold Spring Harb. Perspect. Biol. *7*.
262. Martienssen, R.A., Zaratiegui, M., and Goto, D.B. (2005). RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. Trends Genet. *21*, 450–456.
263. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10.
264. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into

Ago2-containing RNAi enzyme complexes. Cell *123*, 607–620.
265. Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature *453*, 358–362.
266. McClintock, B. (1951). Chromosome organization and genic expression. Cold Spring Harb. Symp. Quant. Biol. *16*, 13–47.
267. McClintock, B. (1965). Components of action of the regulators Spm and Ac. Carnegie Inst. Wash. Yearb. *64*, 527–534.
268. McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179–183.
269. McGuinness, B.E., Hirota, T., Kudo, N.R., Peters, J.M., and Nasmyth, K. (2005). Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. In PLoS Biology, pp. 0433–0449.
270. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
271. Mefford, H.C., and Trask, B.J. (2002). The Complex Structure and Dynamic Evolution of Human Subtelomeres. Nat. Rev. Genet. *3*, 91–102.
272. Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343–349.
273. Mellone, B.G., Ball, L., Suka, N., Grunstein, M.R., Partridge, J.F., and Allshire, R.C. (2003). Centromere Silencing and Function in Fission Yeast Is Governed by the Amino Terminus of Histone H3. Curr. Biol. *13*, 1748–1757.
274. Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I., and Karpen, G.H. (2011). Assembly of *Drosophila* centromeric chromatin proteins during mitosis. PLoS Genet. *7*.
275. Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A., and Heun, P. (2011). *Drosophila* CENH3 Is Sufficient for Centromere Formation. Science. *334*, 686–690.
276. Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. *19*, 5194–5201.
277. Miell, M.D.D., Fuller, C.J., Guse, A., Barysz, H.M., Downes, A., Owen-Hughes, T., Rappsilber, J., Straight, A.F., and Allshire, R.C. (2013). CENP-A confers a reduction in height on octameric nucleosomes. Nat. Struct. Mol. Biol. *20*, 763–765.
278. Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003a). Structure of the catalytic domain of human Dot1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711–723.
279. Min, J., Zhang, Y., and Xu, R.M. (2003b). Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. *17*, 1823–1828.
280. Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D., et al.

(2008). Association of cohesin and Nipped-B with transcriptionally active regions of the *Drosophila melanogaster* genome. Chromosoma *117*, 89–102.
281. Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090–1097.
282. Mizuguchi, G. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science. *303*, 343–348.
283. Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413–420.
284. Moree, B., Meyer, C.B., Fuller, C.J., and Straight, A.F. (2011). CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J. Cell Biol. *194*, 855–871.
285. Mosammaparast, N., and Shi, Y. (2010). Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. Annu. Rev. Biochem. *79*, 155–179.
286. Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789–802.
287. Motamedi, M.R., Hong, E.J.E., Li, X., Gerber, S., Denison, C., Gygi, S., and Moazed, D. (2008). HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms. Mol. Cell *32*, 778–790.
288. Muller, H.J. (1930). Types of visible variations induced by X-rays in *Drosophila*. J. Genet. *22*, 299–334.
289. Müller, S., and Almouzni, G. (2017). Chromatin dynamics during the cell cycle at centromeres. Nat. Rev. Genet. *18*, 192–208.
290. Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227.
291. Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F.S., and Glieder, A. (2012). Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One *7*.
292. Nakagawachi, T., Soejima, H., Urano, T., Zhao, W., Higashimoto, K., Satoh, Y., Matsukura, S., Kudo, S., Kitajima, Y., Harada, H., et al. (2003). Silencing effect of CpG island hypermethylation and histone modifications on O6-methylguanine-DNA methyltransferase (MGMT) gene expression in human cancer. Oncogene.
293. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science. *292*, 110–113.
294. Nardi, I.K., Zasadzińska, E., Stellfox, M.E., Knippler, C.M., and Foltz, D.R. (2016). Licensing of Centromeric Chromatin Assembly through the Mis18α-Mis18β Heterotetramer. Mol. Cell *61*, 774–787.
295. Nathan, D., Ingvarsdottir, K., Sterner, D.E., Bylebyl, G.R., Dokmanovic, M., Dorsey, J.A., Whelan, K.A., Krsmanovic, M., Lane, W.S.,

Meluh, P.B., et al. (2006). Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. Genes Dev. *20*, 966–976.
296. Nayak, T., Szewczyk, E., Oakley, C.E., Osmani, A., Ukil, L., Murray, S.L., Hynes, M.J., Osmani, S.A., and Oakley, B.R. (2006). A versatile and efficient gene-targeting system for Aspergillus nidulans. Genetics *172*, 1557–1566.
297. Ng, H.H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev. *16*, 1518–1527.
298. Niki, H. (2014). *Schizosaccharomyces japonicus*: the fission yeast is a fusion of yeast and hyphae. Yeast *31*, 83–90.
299. Ninomiya, Y., Suzuki, K., Ishii, C., and Inoue, H. (2004). Highly efficient gene replacements in Neurospora strains deficient for nonhomologous end-joining. Proc. Natl. Acad. Sci. *101*, 12248–12253.
300. Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., et al. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature *491*, 284–287.
301. Noguchi, C., Garabedian, M. V., Malik, M., and Noguchi, E. (2008). A vector system for genomic FLAG epitope-tagging in *Schizosaccharomyces pombe*. Biotechnol. J. *3*, 1280–1285.
302. Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I.S., and Watanabe, Y. (2002). Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. Nat. Cell Biol. *4*, 89–93.
303. Nurse, P. (1975). Genetic control of cell size at cell division in yeast. Nature *256*, 547–551.
304. Nurse, P., Thuriaux, P., and Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. Mol. Gen. Genet. MGG *146*, 167–178.
305. Ohzeki, J., Bergmann, J.H., Kouprina, N., Noskov, V.N., Nakano, M., Kimura, H., Earnshaw, W.C., Larionov, V., and Masumoto, H. (2012). Breaking the HAC Barrier: Histone H3K9 acetyl/methyl balance regulates CENP-A assembly. EMBO J. *31*, 2391–2402.
306. Ohzeki, J. ichirou, Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J. Cell Biol. *159*, 765–775.
307. Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., Desai, A., and Fukagawa, T. (2006). The CENP-H–I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat. Cell Biol. *8*, 446–457.
308. Okamura, K., Chung, W.-J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803–806.
309. Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., Saccani, S., and Heun, P. (2011). Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nat. Cell Biol. *13*, 799–808.

310. Padeganeh, A., Ryan, J., Boisvert, J., Ladouceur, A.M., Dorn, J.F., and Maddox, P.S. (2013). Octameric CENP-A Nucleosomes Are Present at Human Centromeres throughout the Cell Cycle. Curr. Biol. *23*, 764–769.
311. Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. *88*, 3734–3738.
312. Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200–213.
313. Parthun, M.R. (2007). Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319–5328.
314. Partridge, J.F., Scott, K.S.C., Bannister, A.J., Kouzarides, T., and Allshire, R.C. (2002). cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. Curr. Biol. *12*, 1652–1660.
315. Pehrson, J.R., and Fried, V. a (1992). MacroH2A, a core histone containing a large nonhistone region. Science. *257*, 1398–1400.
316. Peters, J.E., and Craig, N.L. (2001). Tn7: smarter than we thought. Nat Rev Mol Cell Biol *2*, 806–814.
317. Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev. *22*, 3089–3114.
318. Pidoux, A.L., and Allshire, R.C. (2005). The role of heterochromatin in centromere function. Philos Trans R Soc L. B Biol Sci *360*, 569–579.
319. Pidoux, A.L., Uzawa, S., Perry, P.E., Cande, W.Z., and Allshire, R.C. (2000). Live analysis of lagging chromosomes during anaphase and their effect on spindle elongation rate in fission yeast. J. Cell Sci. *113 Pt 23*, 4177–4191.
320. Pidoux, A.L., Choi, E.S., Abbott, J.K.R., Liu, X., Kagansky, A., Castillo, A.G., Hamilton, G.L., Richardson, W., Rappsilber, J., He, X., et al. (2009). Fission Yeast Scm3: A CENP-A Receptor Required for Integrity of Subkinetochore Chromatin. Mol. Cell *33*, 299–311.
321. Plohl, M., Meštrović, N., and Mravinac, B. (2014). Centromere identity from the DNA point of view. Chromosoma *123*, 313–325.
322. Pöggeler, S., and Kück, U. (2006). Highly efficient generation of signal transduction knockout mutants using a fungal strain deficient in the mammalian ku70 ortholog. Gene *378*, 1–10.
323. Potapova, T., and Gorbsky, G. (2017). The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. Biology (Basel). *6*, 12.
324. Provost, P., Silverstein, R. a, Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., Wright, A., Samuelsson, B., Radmark, O., and Ekwall, K. (2002). Dicer is required for chromosome segregation and gene silencing in fission yeast cells. Proc. Natl. Acad. Sci. *99*, 16648–16653.
325. Przewloka, M.R., and Glover, D.M. (2009). The Kinetochore and the Centromere: A Working Long Distance Relationship. Annu. Rev. Genet. *43*, 439–465.
326. Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez,

M., and Glover, D.M. (2011). CENP-C is a structural platform for kinetochore assembly. Curr. Biol. *21*, 399–405.
327. Quénet, D., and Dalal, Y. (2014). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. Elife *3*, e03254.
328. Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a Family of Exoribonucleases in *Arabidopsis*. Science. *321*, 1490–1492.
329. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*.
330. Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621–629.
331. Ranjitkar, P., Press, M.O., Yi, X., Baker, R., MacCoss, M.J., and Biggins, S. (2010). An E3 Ubiquitin Ligase Prevents Ectopic Localization of the Centromeric Histone H3 Variant via the Centromere Targeting Domain. Mol. Cell *40*, 455–464.
332. Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science. *210*, 604–610.
333. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.
334. Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Paša-Tolić, L., and Pikaard, C.S. (2009). Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. Mol. Cell *33*, 192–203.
335. Reddy, B.D., Wang, Y., Niu, L., Higuchi, E.C., Marguerat, S.B., Bähler, J., Smith, G.R., and Jia, S. (2011). Elimination of a specific histone H3K14 acetyltransferase complex bypasses the RNAi pathway to regulate pericentric heterochromatin functions. Genes Dev. *25*, 214–219.
336. Regnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W., and Brown, W. (2005). CENP-A Is Required for Accurate Chromosome Segregation and Sustained Kinetochore Association of BubR1. Mol. Cell. Biol. *25*, 3967–3981.
337. Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 1–7.
338. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626.
339. Reyes-Turcu, F.E., Zhang, K., Zofall, M., Chen, E., and Grewal, S.I.S. (2011). Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. Nat. Struct. Mol. Biol. *18*, 1132–1138.
340. Rhind, N., Chen, Z., Yassour, M., Thompson, D.A., Haas, B.J., Habib, N., Wapinski, I., Roy, S., Lin, M.F., Heiman, D.I., et al. (2011). Comparative functional genomics of the fission yeasts. Science. *332*, 930–936.
341. Riethman, H., Ambrosini, A., Castaneda, C., Finklestein, J., Hu, X.L.,

Mudunuri, U., Paul, S., and Wei, J. (2004). Mapping and initial analysis of human subtelomeric sequence assemblies. Genome Res *14*, 18–28.
342. Robzyk, K. (2000). Rad6-Dependent Ubiquitination of Histone H2B in Yeast. Science. *287*, 501–504.
343. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858–5868.
344. Rošić, S., and Erhardt, S. (2016). No longer a nuisance: Long non-coding RNAs join CENP-A in epigenetic centromere regulation. Cell. Mol. Life Sci. *73*, 1387–1398.
345. Ruthenburg, A.J., Li, H., Patel, D.J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983–994.
346. Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., et al. (2012). Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Mol. Cell *46*, 691–704.
347. Sadaie, M., Iida, T., Urano, T., and Nakayama, J. (2004). A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. EMBO J. *23*, 3825–3835.
348. Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. PLoS Biol. *3*, 1202–1212.
349. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. Genes Dev. *21*, 1603–1608.
350. Sakabe, K., Wang, Z., and Hart, G.W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc. Natl. Acad. Sci. *107*, 19915–19920.
351. Sakuno, T., Tanaka, K., Hauf, S., and Watanabe, Y. (2011). Repositioning of aurora b promoted by chiasmata ensures sister chromatid mono-orientation in meiosis i. Dev. Cell *21*, 534–545.
352. Samejima, I., Matsumoto, T., Nakaseko, Y., Beach, D., and Yanagida, M. (1993). Identification of seven new cut genes involved in *Schizosaccharomyces pombe* mitosis. J. Cell Sci. *105 (Pt 1*, 135–143.
353. Santisteban, M.S., Kalashnikova, T., and Smith, M.M. (2000). Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. Cell *103*, 411–422.
354. Sarafianos, S.G., Das, K., Tantillo, C., Clark, A.D., Ding, J., Whitcomb, J.M., Boyer, P.L., Hughes, S.H., and Arnold, E. (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J. *20*, 1449–1461.
355. Schalch, T., Job, G., Noffsinger, V.J., Shanker, S., Kuscu, C., Joshua-Tor, L., and Partridge, J.F. (2009). High-Affinity Binding of Chp1 Chromodomain to K9 Methylated Histone H3 Is Required to Establish Centromeric Heterochromatin. Mol. Cell *34*, 36–46.
356. Schueler, M.G. (2001). Genomic and Genetic Definition of a Functional Human Centromere. Science. *294*, 109–115.

357. Schultz, S.J., and Champoux, J.J. (2008). RNase H activity: Structure, specificity, and function in reverse transcription. Virus Res. *134*, 86–103.
358. Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.
359. Scott, K.C., and Sullivan, B.A. (2014). Neocentromeres: A place for everything and everything in its place. Trends Genet. *30*, 66–74.
360. Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E., and Musacchio, A. (2011). Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. Curr. Biol. *21*, 391–398.
361. Seath, A., and Sokal, R.. (1973). Numerical Taxonomy. The Principles and Practice of Numerical Classification. Systamatic Zool. *24*, 263–268.
362. Seeler, J.-S., and Dejean, A. (2003). Nuclear and unclear functions of SUMO. Nat. Rev. Mol. Cell Biol. *4*, 690–699.
363. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.
364. Shang, W.H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., and Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. *20*, 1219–1228.
365. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.
366. Shimada, Y., Mohn, F., and Bühler, M. (2016). The RNA-induced transcriptional silencing complex targets chromatin exclusively via interacting with nascent transcripts. Genes Dev. *30*, 1–10.
367. Shiroiwa, Y., Hayashi, T., Fujita, Y., Villar-Briones, A., Ikai, N., Takeda, K., Ebe, M., and Yanagida, M. (2011). Mis17 is a regulatory module of the Mis6-Mal2-Sim4 centromere complex that is required for the recruitment of CenH3/CENP-a in fission yeast. PLoS One *6*.
368. Shivaraju, M., Unruh, J.R., Slaughter, B.D., Mattingly, M., Berman, J., and Gerton, J.L. (2012). Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. Cell *150*, 304–316.
369. Shuaib, M., Ouararhni, K., Dimitrov, S., and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc. Natl. Acad. Sci. *107*, 1349–1354.
370. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*.
371. Sigova, A., Rhind, N., and Zamore, P.D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in *Schizosaccharomyces pombe*. Genes Dev. 2359–2367.
372. Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H.A. (2007). Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class. Science. *315*, 244–247.

373. Silva, M.C.C., Bodor, D.L., Stellfox, M.E., Martins, N.M.C., Hochegger, H., Foltz, D.R., and Jansen, L.E.T. (2012). Cdk Activity Couples Epigenetic Centromere Inheritance to Cell Cycle Progression. Dev. Cell *22*, 52–63.
374. Simmer, F., Buscaino, A., Kos-Braun, I.C., Kagansky, A., Boukaba, A., Urano, T., Kerr, A.R.W., and Allshire, R.C. (2010). Hairpin RNA induces secondary small interfering RNA synthesis and silencing in trans in fission yeast. EMBO Rep. *11*, 112–118.
375. Sipiczki, M., Takeo, K., Yamaguchi, M., Yoshida, S., and Miklos, I. (1998). Environmentally controlled dimorphic cycle in a fission yeast. Microbiology *144*, 1319–1330.
376. Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272–285.
377. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A. (2009). BioMart--biological queries made easy. BMC Genomics *10*, 22.
378. Soboleva, T.A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G.A., and Tremethick, D.J. (2011). A unique H2A histone variant occupies the transcriptional start site of active genes. Nat. Struct. Mol. Biol. *19*, 25–30.
379. Spradling, A.C., Bellen, H.J., and Hoskins, R.A. (2011). *Drosophila* P elements preferentially transpose to replication origins. Proc. Natl. Acad. Sci. *108*, 15948–15953.
380. Stellfox, M.E., Bailey, A.O., and Foltz, D.R. (2013). Putting CENP-A in its place. Cell. Mol. Life Sci. *70*, 387–406.
381. Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. *9*, 573–586.
382. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.
383. Straube, K., Blackwell, J.S., and Pemberton, L.F. (2010). Nap1 and Chz1 have separate Htz1 nuclear import and assembly functions. Traffic *11*, 185–197.
384. Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226.
385. Su, D., Hu, Q., Li, Q., Thompson, J.R., Cui, G., Fazly, A., Davies, B.A., Botuyan, M.V., Zhang, Z., and Mer, G. (2012). Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106. Nature *483*, 104–107.
386. Subramanian, L., Toda, N.R.T., Rappsilber, J., and Allshire, R.C. (2014). Eic1 links Mis18 with the CCAN/Mis6/Ctf19 complex to promote CENP-A assembly. Open Biol. *4*, 140043–140043.
387. Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J. Cell Biol. *127*, 581–592.
388. Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., and Peters, J.M. (2002). The dissociation of cohesin from

chromosomes in prophase is regulated by polo-like kinase. Mol. Cell *9*, 515–525.
389. Sun, X., Wahlstrom, J., and Karpen, G. (1997). Molecular structure of a functional *Drosophila* centromere. Cell *91*, 1007–1019.
390. Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional *Drosophila* centromere. Genome Res. *13*, 182–194.
391. Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nat. Educ. *308*, 548–550.
392. Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat. Struct. Biol. *7*, 1121–1124.
393. Suzuki, A., Badger, B.L., Wan, X., DeLuca, J.G., and Salmon, E.D. (2014). The Architecture of CCAN Proteins Creates a Structural Integrity to Resist Spindle Forces and Achieve Proper Intrakinetochore Stretch. Dev. Cell *30*, 717–730.
394. Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232–235.
395. Tada, K., Susumu, H., Sakuno, T., and Watanabe, Y. (2011). Condensin association with histone H2A shapes mitotic chromosomes. Nature *474*, 477–483.
396. Tadeo, X., Wang, J., Kallgren, S.P., Liu, J., Reddy, B.D., Qiao, F., and Jia, S. (2013). Elimination of shelterin components bypasses RNAi for pericentric heterochromatin assembly. Genes Dev. *27*, 2489–2499.
397. Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. Cell *116*, 51–61.
398. Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science. *288*, 2215–2219.
399. Takahashi, T.S., Basu, A., Bermudez, V., Hurwitz, J., and Walter, J.C. (2008). Cdc7-Drf1 kinase links chromosome cohesion to the initiation of DNA replication in Xenopus egg extracts. Genes Dev. *22*, 1894–1905.
400. Talbert, P.B., and Henikoff, S. (2010a). Histone variants — ancient wrap artists of the epigenome. Nat. Rev. Mol. Cell Biol. *11*, 264–275.
401. Talbert, P.B., and Henikoff, S. (2010b). Centromeres convert but don't cross. PLoS Biol. *8*, 1–5.
402. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534–538.
403. Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature *414*, 277–283.
404. Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. *101*, 11030–11035.

405. Tanaka, A., Tanizawa, H., Sriswasdi, S., Iwasaki, O., Chatterjee, A.G., Speicher, D.W., Levin, H.L., Noguchi, E., and Noma, K.I. (2012). Epigenetic Regulation of Condensin-Mediated Genome Organization during the Cell Cycle and upon DNA Damage through Histone H3 Lysine 56 Acetylation. Mol. Cell *48*, 532–546.
406. Tanaka, T., Fuchs, J., Loidl, J., and Nasmyth, K. (2000). Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. Nat. Cell Biol. *2*, 492–499.
407. Tange, Y., Kurabayashi, A., Goto, B., Hoe, K.L., Kim, D.U., Park, H.O., Hayles, J., Chikashige, Y., Tsutumi, C., Hiraoka, Y., et al. (2012). The CCR4-NOT complex is implicated in the viability of aneuploid yeasts. PLoS Genet. *8*.
408. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinforma.
409. Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. Science. *272*, 408–411.
410. Thomson, T., and Lin, H. (2009). The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. Annu. Rev. Cell Dev. Biol. *25*, 355–376.
411. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.
412. Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc. Natl. Acad. Sci. *101*, 15986–15991.
413. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.
414. Turner, B.M. (1993). Decoding the nucleosome. Cell *75*, 5–8.
415. Ünal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J.E., and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Mol. Cell *16*, 991–1002.
416. Unhavaithaya, Y., and Orr-Weaver, T.L. (2013). Centromere proteins CENP-C and CAL1 functionally interact in meiosis for centromere clustering, pairing, and chromosome segregation. Proc. Natl. Acad. Sci. *110*, 19878–19883.
417. Upadhyay, U., Srivastava, S., Khatri, I., Nanda, J.S., Subramanian, S., Arora, A., and Singh, J. (2017). Ablation of RNA interference and retrotransposons accompany acquisition and evolution of transposases to heterochromatin protein CENPB. Mol. Biol. Cell *28*, 1132–1146.
418. Vagin, V. V. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science. *313*, 320–324.
419. Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of Histone Lysine Methylation across Transcribed Mammalian Chromatin. Mol. Cell. Biol. *26*, 9185–9195.
420. Vass, S., Cotterill, S., Valdeolmillos, A.M., Barbero, J.L., Lin, E., Warren, W.D., and Heck, M.M.S. (2003). Depletion of Drad21/Scc1 in

Drosophila cells leads to instability of the cohesin complex and disruption of mitotic progression. Curr. Biol. *13*, 208–218.
421. Venkatasubrahmanyam, S., Hwang, W.W., Meneghini, M.D., Tong, A.H.Y., and Madhani, H.D. (2007). Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. Proc. Natl. Acad. Sci. *104*, 16609–16614.
422. Verdel, A. (2004). RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Science. *303*, 672–676.
423. Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W.S., Karpilow, J., and Khvorova, A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. RNA *11*, 674–682.
424. Villalba, F., Collemare, J., Landraud, P., Lambou, K., Brozek, V., Cirer, B., Morin, D., Bruel, C., Beffa, R., and Lebrun, M.H. (2008). Improved gene targeting in Magnaporthe grisea by inactivation of MgKU80 required for non-homologous end joining. Fungal Genet. Biol. *45*, 68–75.
425. Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol. Cell. Biol. *21*, 8346–8356.
426. Volpe, T. (2002). Regulation of heterochromatic silencing and histone H3 Lysine-9 by RNAi. Science. *297*, 1833–1837.
427. Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C. (2003). RNA interference is required for normal centromere function in fission yeast. Chromosom. Res. *11*, 137–146.
428. Waddington, C.H. (1942). The epigenotype. Endeavour 18–20.
429. Waddington, C.H. (1957). The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology (Alen & Unwin).
430. Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al. (2009). Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. Science. *326*, 865–867.
431. Waizenegger, I.C., Hauf, S., Meinke, a, and Peters, J.M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell *103*, 399–410.
432. Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607–614.
433. Wang, Y. (2004). Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. Science. *306*, 279–283.
434. Wang, J., Tadeo, X., Hou, H., Andrews, S., Moresco, J.J., Yates, J.R., Nagy, P.L., and Jia, S. (2014). Tls1 regulates splicing of shelterin components to control telomeric heterochromatin assembly and telomere length. Nucleic Acids Res. *42*, 11419–11432.
435. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.
436. Wassenegger, M., Heimes, S., Riedel, L., and Sänger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567–576.

437. Watanabe, Y., and Yamamoto, M. (1994). *S. pombe* mei2+encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell *78*, 487–498.
438. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539–543.
439. Waterhouse, P.M., Graham, M.W., and Wang, M.-B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. *95*, 13959–13964.
440. Watson, A.T., Werler, P., and Carr, A.M. (2011). Regulation of gene expression at the fission yeast *Schizosaccharomyces pombe* urg1 locus. Gene *484*, 75–85.
441. Watt, S., Mata, J., López-Maury, L., Marguerat, S., Burns, G., and Bähler, J. (2008). urg1: A uracil-regulatable promoter system for fission yeast with short induction and repression times. PLoS One *3*.
442. Welton, R.M., and Hoffman, C.S. (2000). Glucose monitoring in fission yeast via the Gpa2 galpha, the git5 Gbeta and the git3 putative glucose receptor. Genetics *156*, 513–521.
443. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.
444. Werler, P.J.H., Hartsuiker, E., and Carr, A.M. (2003). A simple Cre-loxP method for chromosomal N-terminal tagging of essential and non-essential *Schizosaccharomyces pombe* genes. Gene *304*, 133–141.
445. Wessler, S.R. (2006). Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. *103*, 17600–17601.
446. Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., et al. (2006). Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. Cell *125*, 467–481.
447. Whittle, C.M., McClinic, K.N., Ercan, S., Zhang, X., Green, R.D., Kelly, W.G., and Lieb, J.D. (2008). The genomic distribution and function of histone variant HTZ-1 during *C. elegans* embryogenesis. PLoS Genet. *4*.
448. Williams, J.S., Hayashi, T., Yanagida, M., and Russell, P. (2009). Fission Yeast Scm3 Mediates Stable Assembly of Cnp1/CENP-A into Centromeric Chromatin. Mol. Cell *33*, 287–298.
449. Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., et al. (2009). Maize centromere structure and evolution: Sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet. *5*.
450. Wong, L.H., and Choo, K.H.A. (2004). Evolutionary dynamics of transposable elements at the centromere. Trends Genet. *20*, 611–616.
451. Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. Nature *415*, 871–880.
452. Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan,

B.W., Staines, D.M., Aslett, M., Lock, A., Bahler, J., Kersey, P.J., et al. (2012). PomBase: A comprehensive online resource for fission yeast. Nucleic Acids Res. *40*.
453. Xiao, H., Mizuguchi, G., Wisniewski, J., Huang, Y., Wei, D., and Wu, C. (2011). Nonhistone Scm3 Binds to AT-Rich DNA to Organize Atypical Centromeric Nucleosome of Budding Yeast. Mol. Cell *43*, 369–380.
454. Xie, W., Gai, X., Zhu, Y., Zappulla, D.C., Sternglanz, R., and Voytas, D.F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol. Cell. Biol. *21*, 6606–6614.
455. Yamagishi, Y., Sakuno, T., Shimura, M., and Watanabe, Y. (2008). Heterochromatin links to centromeric protection by recruiting shugoshin. Nature *455*, 251–255.
456. Yamanaka, S., Mehta, S., Reyes-Turcu, F.E., Zhuang, F., Fuchs, R.T., Rong, Y., Robb, G.B., and Grewal, S.I.S. (2013). RNAi triggered by specialized machinery silences developmental genes and retrotransposons. Nature *493*, 557–560.
457. Yieh, L., Kassavetis, G., Geiduschek, E.P., and Sandmeyer, S.B. (2000). The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. J. Biol. Chem. *275*, 29800–29807.
458. Yu, B. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. Science. *307*, 932–935.
459. Yu, R., Jih, G., Iglesias, N., and Moazed, D. (2014). Determinants of Heterochromatic siRNA Biogenesis and Function. Mol. Cell *53*, 262–276.
460. Yuan, G., and Zhu, B. (2012). Histone variants and epigenetic inheritance. Biochim. Biophys. Acta - Gene Regul. Mech. *1819*, 222–229.
461. Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simao, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E. V. (2017). OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. *45*, D744–D749.
462. Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258–262.
463. Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012a). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871–884.
464. Zhang, H., Roberts, D.N., and Cairns, B.R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell *123*, 219–231.
465. Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I.S. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol. *15*, 381–388.
466. Zhang, W., Colmenares, S.U., and Karpen, G.H. (2012b). Assembly of *Drosophila* Centromeric Nucleosomes Requires CID Dimerization. Mol. Cell *45*, 263–269.

467. Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K. (2002). Centromeric Retroelements and Satellites Interact with Maize Kinetochore Protein CENH3. Plant Cell *14*, 2825–2836.
468. Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. RNA *15*, 1886–1895.
469. Zhou, Z., Feng, H., Zhou, B.-R., Ghirlando, R., Hu, K., Zwolak, A., Miller Jenkins, L.M., Xiao, H., Tjandra, N., Wu, C., et al. (2011). Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. Nature *472*, 234–237.
470. Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D., and Grewal, S.I.S. (2009). Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature *461*, 419–422.