Accelerated Adaptation through Stimulated Copy Number Variation in *Saccharomyces cerevisiae*

Epigenetics Department, The Babraham Institute, Cambridge

Homerton College, University of Cambridge

Ryan Matthew Hull

April 2018

This thesis is submitted for the degree of Doctor of Philosophy

Declaration of Originality

This work was carried out at the Babraham Institute under the supervision of Dr Jon Houseley between October 2014 and April 2018.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text.

During the course of my PhD, a selection of the data presented in this thesis, along with additional data produced by colleagues and not included in this thesis, contributed to a first author publication: Hull, R. M. *et al.* (2017) 'Environmental change drives accelerated adaptation through stimulated copy number variation', *PLoS Biology*. A copy of this publication can be found at the end of this thesis in Appendix A.

This dissertation does not exceed 60,000 words.

Ryan Hull

Acknowledgements

Firstly, I would like to thank my supervisor Jon Houseley for his guidance and great scientific input into my project these last three and a half years. He has been a wonderful supervisor, colleague and friend, and has helped me develop both scientifically and professionally.

Secondly, I would like to thank my assessor Wolf Reik and my mentor Cris Cruz-Bolland, for your ideas and feedback on where to take my project. I would also like to thank Simon Cook and all the members of the Cook lab, in particular Matt Sale, for helping start me off with cancer cell work, so I can try to translate my yeast research into humans. Thanks must also go to Anne Segonds-Pichon for her invaluable statistical advice, and to the Gottschling lab for providing the Mother Enrichment yeast strains used during my PhD.

Thank you to all the members of the Houseley lab, both past and present, and to all my cohort of PhD students for the good times we have shared. Thank you to everyone at the Babraham Institute and associated Facilities for helping to make my PhD go more smoothly and be an enjoyable experience.

Thank you also to my funders, the BBSRC, MRC, and Homerton College for financing my PhD and making it a possibility.

I would also like to thank my family for being there for me during my entire education. Finally, and most importantly, I would like to thank my beautiful fiancée Iva, who has been by my side for the entire PhD. You have shared all the good and bad moments that came during my PhD and I am so grateful that I got to spend it with you.

Table of acknowledgement of assistance received during course of thesis

1) Initial training in techniques and laboratory practice and subsequent mentoring:

My supervisor Dr Jon Houseley and my mentor Dr Cris Cruz-Bolland conducted my initial lab training.

Dr Jon Houseley, Dr Cris Cruz-Bolland and my assessor Professor Wolf Reik conducted my subsequent mentoring.

2) Data obtained from a technical service provider (e.g. DNA sequencing, illustrations, simple bioinformatics information etc)

All new plasmids used in this study had their DNA sequence confirmed, with the DNA sequencing performed by Beckman Coulter Genomics.

3) Data produced jointly (e.g. where it was necessary or desirable to have two pairs of hands)

Dr Anne Segonds-Pichon helped in the design of the statistical approach to analyse the results of the Cell Fate Analysis experiment. Actual analysis was performed by myself.

4) Data/materials provided by someone else (e.g. one-off analysis, bioinformatics analysis, where parallel data or technical provision in a very different area is needed to provide a connected account in the thesis)

The Mother Enrichment Program (MEP) wild-type strain was kindly provided by the Gottschling lab.

Dr Jon Houseley performed all bioinformatics analysis on my RNAseq libraries for Fig. 3.1b.

Dr Jon Houseley performed the Northern blot analysis for Fig. 3.6b.

Abstract

Repetitive regions of the genome, such as the centromeres, telomeres and ribosomal DNA account for a large proportion of the genetic variation between individuals. Differences in the number of repeat sequences between individuals is termed copy number variation (CNV) and is rife across eukaryotic genomes. CNV is of clinical importance as it has been implicated in many human disorders, in particularly cancers where is has been associated with tumour growth and drug resistance.

The copper-resistance gene *CUP1* in *Saccharomyces cerevisiae* is one such CNV gene. *CUP1* is transcribed from a copper inducible promoter and encodes a protein involved in copper detoxification. In this work I show that yeast can regulate their repeat levels of the *CUP1* gene through a transcriptionally stimulated CNV mechanism, as a direct adaptation response to a hostile environment. I characterise the requirement of the epigenetic mark Histone H3 Lysine 56 acetylation (H3K56ac) for stimulated CNV and its limitation of only working at actively transcribed genes. Based upon my findings, I propose a model for how stimulated CNV is regulated in yeast and show how we can pharmacologically manipulate this mechanism using drugs, like nicotinamide and rapamycin, to stimulate and repress a cell's ability to adapt to its environment. I further show that the model is not limited to high-copy *CUP1* repeat arrays, but is also applicable to low-copy systems. Finally, I show that the model extends to other genetic loci in response to different challenging environments, such as formaldehyde stimulation of the formaldehyde-resistance gene *SFA1*.

To the best of our knowledge, this is the first example of any eukaryotic cell undergoing genome optimisation as a novel means to accelerate its adaptation in direct response to its environment. If conserved in higher eukaryotes, such a mechanism could have major implications in how we consider and treat disorders associated with changes in CNV.

Contents

List of Figures

List of tables

Table of abbreviations

Abbreviation	Meaning
ARS	Autonomous replicating sequence
AU	Arbitrary units
BER	Base excision repair
BIR	Break induced replication
cDNA	Reverse transcribed mRNA
CMG	Cdc45-Mcm2-7-GINS complex
CNV	Copy number Variation
CO	Crossover event
CUT	Cryptic unstable transcript
DHFR	Dihydrofolate reductase
dHJ	Double Holliday Junction
D-loop	Displacement loop
dmin	Double minute
dNTPs (dATP/dCTP/dGTP/dTTP)	Deoxyribonucleotide triphosphate
DSB	DNA double-strand break
DSBR	Double-strand break repair
eccDNA	Extrachromosomal circular DNA
E. coli	*Escherichia coli*
ERC	Extrachromosomal ribosomal DNA circle
FA	Formaldehyde
FGFR2	Fibroblast growth factor receptor 2
FKBP	FK506-binding protein
FoSTeS	Fork stalling and template switching
FPC	Fork protection complex
FSHD	Facioscapulohumeral muscular dystrophy
G0/G1/G2 phase	Growth phases 0/I/II
Gal	Galactose
GCR	Gross chromosomal rearrangement
GINS	go-ichi-ni-san complex
Glu	Glucose
H3K56ac	Histone H3 Lysine 56 acetylation
HDAC	Histone deacetylase
HR	Homologous recombination
IGS1 and IGS2	Ribosomal DNA intergenic spacer regions 1 and 2
IR	Ionising radiation
Kan	Kanamycin
LOH	Loss of heterozygosity
LTR	Long terminal repeat
M phase	Mitosis phase
MEK1/2	Mitogen-activated or extracellular signal-regulated protein kinase kinases 1 and 2
MEP	Mother Enrichment Program
MMBIR	Microhomology-mediated break-induced replication
MMEJ	Microhomology-mediated end joining
MMIR	Microhomology/microsatellite-induced replication
MMR	DNA mismatch repair
MRX	Mre11-Rad50-Xrs2 complex
NAHR	Non-allelic homologous recombination
Nat	Nourseothricin
NCO	Non-crossover event

ncRNA	Non-coding RNA
NDP	Nitrogen discrimination pathway
NER	Nucleotide excision repair
Nic	Nicotinamide
OD	Optical density
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PIK	Phosphatidylinositol kinase
Pol I/II/III	RNA Pol I/II/III
Pol α/δ/ε/ζ	DNA polymerase α/δ/ε/ζ
Poly(dA/T)	Continuous sequence of the base A/T
rDNA	Ribosomal DNA
RFB	Replication fork barrier
RFC	Replication factor C complex
RFS	Replication fork stalling site
RNAi	RNA interference
RNase H1/2	Ribonuclease H1 and H2
rNTPs (rATP/rCTP/rGTP/rUTP)	Ribonucleotide triphosphates
RPA	Replication protein A
rRNA	Ribosomal RNA
RTG	Retrograde response
S phase	Synthesis phase
SD	Segmental duplication
SDSA	Synthesis-dependent strand annealing
SFDA	Small fragment-driven DNA amplification
SSA	Single-strand annealing
SSB	DNA single-strand break
ssDNA	Single-stranded DNA
TAM	Transcription-associated mutagenesis
TAR	Transcription-associated recombination
TLS	Translesion DNA synthesis
TOR	Target of rapamycin
TORC1 and TORC2	TOR complex 1 and 2
Trp	Tryptophan
UAS	Upstream activating sequence
UV	Ultraviolet radiation
WS	Williams-Beuren syndrome

1. Introduction

1.1 Adaptation for survival

Adaptation is a fundamental process that all organisms must undertake in order to remain competitive in their surrounding environment. Within any given population, there is an underlying level of random genetic variation, and cells or organisms with beneficial mutations for their environment gain a fitness advantage and an increased likelihood of propagating. As such, beneficial mutations spread through the population by natural selection. Neo-Darwinian theory stipulates that the underlying rate of random genetic variation is constant and must remain unaffected by changes in the surrounding environment (Mayr, 1982). Cells need not however, generate these random mutations by accident. In fact, genome-wide induction of mutation has been well characterised in response to conditions of stress in bacteria (Cairns and Foster, 1991; McKenzie, Lombardo and Rosenberg, 1998). Similar mutations have also been reported in yeast, as part of the stress response (Shor, Fox and Broach, 2013), and comparable interactions of the environment on the genome have been postulated in higher eukaryotes (Metzgar and Wills, 2000). These studies have shown that some mutations can be directly induced in response to the environment and therefore not all mutations follow neo-Darwinian theory.

1.2 Mutation rates

When cells experience a change in environment, in particular exposure to a challenging environment, such as exposure of bacteria to a virus, the net effect is the appearance of a novel variant strain characterised by a mutation(s) that confer an adaptive advantage in the environment, such as resistance to the virus (Luria and Delbrück, 1943). There are two alternate hypotheses for the mechanism by which the mutation is thought to first arise: randomly pre-existing mutations and acquired mutations.

The randomly pre-existing mutation hypothesis states that the mutation occurs by random in the population prior to the change in environment, but is rare as the mutation is not being selected for (Luria and Delbrück, 1943). However, in the new environment, where the mutation confers an adaptive advantage, the mutation is now selected for and ultimately becomes the dominant allele in the population (Luria and Delbrück, 1943). Alternatively, the acquired mutation hypothesis states that the mutation does not pre-exist in the population prior to the environmental change, but instead is acquired by a small number of cells that survived the initial change in environment (Luria and Delbrück, 1943). Again the small number of cells that acquired the beneficial mutation become selected for in the new environment and the mutation ultimately becomes the dominant allele in the population (Luria and Delbrück, 1943).

'Acquired' mutation is a general term for any mutations that are induced by the environment, which can be further subdivided into the terms 'adaptive' and 'directed' mutation, which are better defined. An 'adaptive' mutation is the term given to mutations that arise in response to an environment, which can be both beneficial or detrimental to the cell, but only mutations conferring an adaptive benefit become selected for in the environment (Delbrück and Bailey, 1946; Rosenberg, 2001). 'Adaptive' mutation must not be confused with the term 'directed' mutation, which refers to mutations that confer a fitness advantage in the environment, being preferentially induced by the environment (Rosenberg, 2001). The main difference between 'adaptive' and 'directed' mutations is that 'directed' mutations require a foreknowledge of what gene to mutate and what type of mutation is best to introduce for the given environment, whereas 'adaptive' mutations do not already know the outcome of the mutation prior to introducing it.

In a classic study, Luria and Delbrük designed a fluctuation test (Fig. 1.1) to determine whether mutations in bacteria that made them resistant to a viral infection were random pre-existing or acquired mutations (Luria and Delbrück, 1943). The principle behind the fluctuation test in the study was to inoculate a small amount *Escherichia coli* (*E. coli*) into separate pre-cultures and, after a

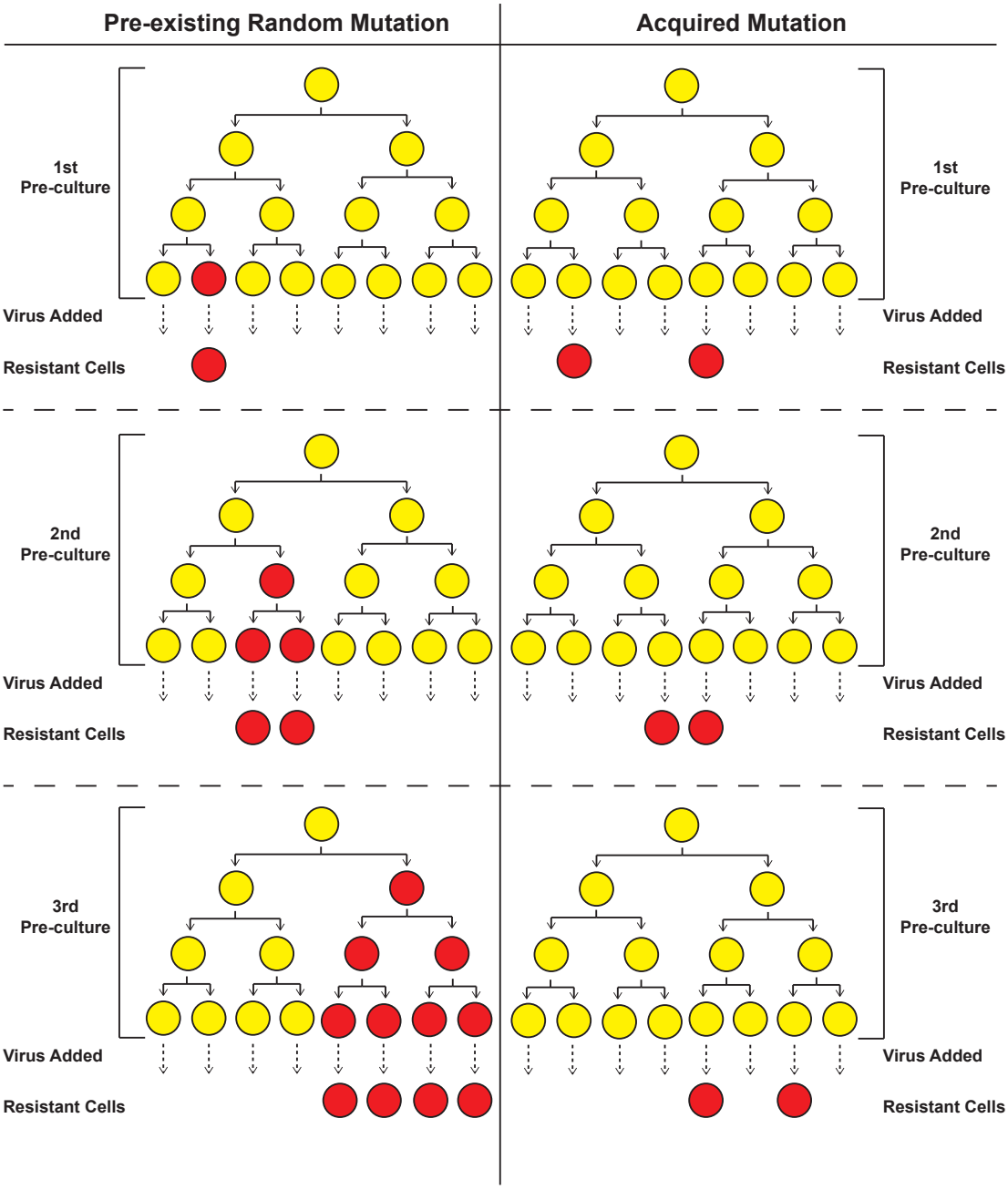


Figure 1.1: Schematic of the two possible scenarios for the fluctuation test devised by Luria and Delbrück, 1943. A small number of bacteria were inoculated into separate pre-cultures and grown for several generations, and then plated onto media containing a virus. Yellow and red circles represent sensitive and virus-resistant bacteria respectively. Pre-existing random mutation (left pathway) produces a variable number of "clones of resistant cells", depending on when the random mutation first arose in the pre-culture. Pre-existing random mutation follows a Luria-Delbrück distribution, with a much greater variance than mean number of resistant cells per plate. Acquired mutation (right pathway) produces resistant cells in response to virus exposure. No resistant cells pre-exist in the pre-cultures. Acquired mutation follows a Poisson distribution, with the mean equal to the variance.

period of growth, plate an equal volume of each culture onto agar plates containing the virus, and compare the number of virus-resistant bacterial colonies that grow on each of the plates. In the situation where the mutation existed prior to the addition of the virus, the number of “clones of resistant bacteria” in each pre-culture at the time of challenging with the virus, is dependent upon how early on in the culture growth the resistance mutation first arose. As such, each pre-culture is predicted to contain a variable number of resistant bacteria, with the variance of resistant cells per plate being considerably greater than the mean, and following a probability distribution that became known as the Luria-Delbrück distribution.

In the alternative scenario where the mutation is acquired, the pre-cultures do not contain any ‘resistant’ cells at the point of virus exposure, but a similar, small number of random bacteria are predicted to survive the initial viral infection from each pre-culture and introduce a heritable virus-resistant mutation. The number of resistant cells per plate from the acquired resistance hypothesis follows a Poisson distribution, with the mean equal to the variance (Luria and Delbrück, 1943).

Luria and Delbrück observed a much greater variance than mean, in the number of virus-resistant bacteria per plate, and therefore concluded that mutation in bacteria is randomly pre-existing and not acquired in response to the environment, in this case the virus (Luria and Delbrück, 1943). Further classical papers have also shown that mutations can arise independently of the environment (Newcombe, 1949; Lederberg and Lederberg, 1952). However, these classic papers do not consider mutations that may be induced by contact with a selective environment (Rosenberg, 2001), and cases of adaptive mutation have been reported in non-growing or slow growing bacteria (McKenzie, Lombardo and Rosenberg, 1998; Hastings *et al.*, 2000). These bacterial studies of adaptive mutation utilise a *lac* +1 frameshift mutation reversion assay, which selects for revertant Lac^+ cells that can now grow on lactose (Cairns and Foster, 1991). Most Lac^+ revertant colonies have adaptive point mutations, but revertant Lac^+ bacteria with 20-50 copy amplifications of 7-40kb direct repeats, containing the weakly functional mutant *lac* allele, were also observed, showing that gene dosage compensation can be an adaptive mutation (Foster and Trimarchi, 1994;

Rosenberg *et al.*, 1994; Hastings *et al.*, 2000). The revertants in the *lac* +1 frameshift studies are definitely not 'directed' mutations, as mutations were not specific to the *lac* gene or even the same chromosome (Foster, 1997; Torkelson *et al.*, 1997). Instead the revertant mutations are believed to arise as a result of genome-wide hyper-mutation in a small subpopulation of starved cells, thereby following a random Darwinian process of mutation (Foster, 1997; Torkelson *et al.*, 1997; Rosenberg, 2001).

Adaptive mutation has also been described in non-dividing yeast, using *lys2* frameshift mutation reversion assays, equivalent to the *lac* reversion system in bacteria (Steele and Jinks-Robertson, 1992; Baranowska, Policinska and Jachymczyk, 1995; Heidenreich and Wintersberger, 1997). The frequency of adaptive mutation has also been linked to the proofreading activity of DNA polymerase δ, with a marked increase in the emergence of adaptive mutations when proofreading activity is impaired under restrictive conditions, such as growth at 37°C (Baranowska, Policinska and Jachymczyk, 1995).

The importance of adaptive mutation in human somatic cells has also been recognised in regards to cancer (Strauss, 1992; Rosenberg *et al.*, 1994), with the major implication being the acquisition of mutations that enable a cell to proliferate in a growth-limited state (Rosenberg, 2001). Therefore there is growing evidence that cells do not always adapt using mechanisms blind to the environment, as proposed by neo-Darwinian theory. Instead, specific challenging environments can promote genome-wide induction of mutation in a small subpopulation of cells, which include point mutations and genomic rearrangements. These environmentally induced mutations may or may not provide an adaptive advantage to the cell in the environment, and so they still follow a random Darwinian process of mutation and environmental adaptation, and therefore must be strongly distanced from 'directed' mutation arguments.

1.3 Copy Number Variation

Gene dosage compensation through amplification of the weakly functional mutant *lac* allele was an adaptive mutation event observed in the *lac* +1 frameshift studies in bacteria (Hastings *et al.*, 2000). However, repetitive

elements are found naturally and are highly abundant in some eukaryotes, accounting for more than 50% of the human genome (Richard, Kerrest and Dujon, 2008). The study classified all repetitive elements as falling into one of two major families: 1) tandem repeats and 2) dispersed repeats. The family of tandem repeats comprises of tandem genes, the ribosomal DNA (rDNA) repeat array, and satellite DNA, which is itself subdivided into satellites, minisatellites, and microsatellites. The dispersed repeats family is made up of transposons, tRNA genes, and gene paralogues (Richard, Kerrest and Dujon, 2008).

A change in the number of repeats is termed copy number variation (CNV) and can be an amplification or a contraction event (Fig. 1.2), with the human genome showing 5-10% variation in CNV between individuals (Iafrate *et al.*, 2004; Sebat *et al.*, 2004; Zarrei *et al.*, 2015). Traditionally, CNV was believed to be caused by meiotic recombination, but recent studies have shown CNV profiles to vary between organs and tissue types within an individual, evident of CNV also forming by mitotic recombination (Dieckhoff *et al.*, 2004; Bruder *et al.*, 2008; Piotrowski *et al.*, 2008). Also CNV is not limited to single repeat changes, as DNA repair is not limited to using the adjacent repeats, suggesting that higher-order DNA structures are also important for repair resulting in CNV (Lupski, 1998).

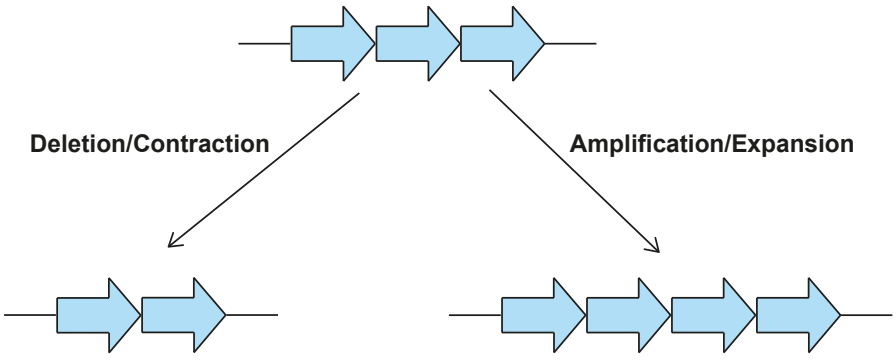


Figure 1.2: CNV at a multi-copy locus. The CNV gene (pale blue arrow), shown here as a 3-repeat array, can lose a copy (pathway left arm) to become a 2-repeat array, or gain a copy (pathway right arm) to become a 4-repeat array, as a result of errors in DNA replication or DNA repair.

CNV of genes does not always have to occur on chromosomes, but can also occur through the less studied extrachromosomal circular DNAs (eccDNAs) (Moller *et al.*, 2015). In yeast, eccDNAs accumulate in Mother cells with age and are generally enriched for eccDNAs arising from highly repetitive regions, such as the ribosomal DNA (rDNA) (Sinclair and Guarente, 1997), telomeres (Horowitz and Haber, 1985), and tandemly repeated genes (Moller *et al.*, 2015). However, eccDNAs have also been detected arising from regions with short or no repetitive sequence (Moller *et al.*, 2015). In humans, eccDNAs have been identified as recombination products of immunoglobulin class switch rearrangement (von Schwedler, Jack and Wabl, 1990) and as double minutes (dmin) in cancer (Cox, Yuncken and Spriggs, 1965; Radloff, Bauer and Vinograd, 1967; Storlazzi *et al.*, 2010).

1.4 CNV in human health and disease

CNV of protein coding genes has been implicated in many human disorders (Craddock *et al.*, 2010; Stankiewicz and Lupski, 2010), with some specific genetic syndromes being directly attributed to changes in copy number (Francke, 1999; van der Maarel, Tawil and Tapscott, 2011). One of the most common CNV disorders in humans is Williams-Beuren syndrome (WS), affecting between 1 in 7500 and 1 in 20,000 people at birth (Martens, Wilson and Reutens, 2008). WS is caused by a heterozygous deletion of up to 2 Mb of chromosomal sub-band 7q11.23 (Francke, 1999). Patients with WS often present specific physical, cognitive, and behavioural abnormalities, such as cardiovascular and dental irregularities, and gross deficiencies in visual-spatial processing (Francke, 1999; Tassabehji *et al.*, 1999). Known genes present in the deletion region on chromosome 7 include *ELN*, *RFC2*, *LIMK1*, *FZD1*, *STX1A*, *EIF4H*, *GTF2I*, *CYLN2*, *FKBP6*, *WBSCR9*, *BCL7B*, *WS-bHLH*, *TBL2*, *CPETR1*, *CPETR2*, and *GTF2IRD1* (Francke, 1999). Haploinsufficiency of *ELN*, which encodes the connective-tissue protein elastin, has been shown to be responsible for the cardiological features, but not the facial abnormalities, in patients with WS (Tassabehji *et al.*, 1999). *LIMK1*, which encodes a protein tyrosine kinase expressed in the developing

brain, and *STX1A*, which encodes a component of the synaptic apparatus, have been considered good candidate genes responsible for the cognitive and behavioural differences observed in patients with WS (Tassabehji *et al.*, 1999). However, a study of patients hemizygous for just the *LIMK1* or *STX1A* gene suggested neither gene is likely to contribute to any WS phenotypes (Tassabehji *et al.*, 1999). Nevertheless, WS remains an important example of the importance of maintaining the correct copy number of genes in the genome.

One of the most well characterised human diseases caused by an abnormal CNV event is facioscapulohumeral muscular dystrophy (FSHD). It is the third most commonly inherited muscular dystrophy, initially affecting the shoulder and facial muscles, and later progressing downward to the lower extremities (Pandya, King and Tawil, 2008). FSHD is caused by a contraction in the D4Z4 microsatellite repeat (de Greef, Frants and van der Maarel, 2008). There is no correlation between the D4Z4 repeat array size and the age of onset, nor the severity of the disease, but patients with repeat arrays of just 1 to 3 units usually experience infantile onset and rapid disease progression (Lunt *et al.*, 1995). Loss of D4Z4 repeats causes transcriptional up-regulation of closely located genes, such as *FRG1, FRG2* and *ANT1* (Gabellini, Green and Tupler, 2002). Therefore CNV is important in gene expression regulation. Recent evidence in *Saccharomyces cerevisiae* has also found that transcripts from multi-copy loci form double stranded RNA more efficiently than transcripts from an equivalently expressed single-copy loci, therefore CNV can directly influence gene expression through RNA interference (RNAi) (Cruz and Houseley, 2014).

Copy number variation of protein coding genes need not always be a detrimental process for cells. In fact, there are numerous examples, particularly in response to environmental stresses, where novel CNV alleles actually enhance a cell's ability to adapt to its environment. In yeast, novel CNVs have been discovered during evolution studies that can bestow drug resistance, enhance growth ability during prolonged nutrient starvation, and even complement genetic defects (Huang and Campbell, 1995; Dunham *et al.*, 2002; Libuda and Winston, 2006; Payen *et al.*, 2008; Gresham *et al.*, 2010).

In humans, the copy number of the *AMY1* gene, encoding amylase, the enzyme responsible for starch hydrolysis, is positively correlated with salivary amylase protein levels, and also with individuals from populations considered to have high-starch diets (Lebenthal, 1987; Perry *et al.*, 2007). This suggests that individuals from populations with a high-starch diet have undergone a strong positive selection towards higher *AMY1* copy numbers as a direct adaptation to their environment, likely driven by an improved digestion of starchy foods and/or to counteract any detrimental effect of intestinal disease (Perry *et al.*, 2007).

In cancer, albeit to the detriment of the host, copy number amplification of oncogenes has been shown to enhance the proliferation rate of cancer cells. A meta-analysis of over 3000 cancers, from more than two-dozen cancer types, found that amplification of the cyclin-dependent kinase 4 (*CDK4*) oncogene activates Cdk4, a master regulator of cell cycle progression from G1 (Beroukhim *et al.*, 2010). Another example is the copy number amplification of the fibroblast growth factor receptor 2 (*FGFR2*) gene, which in gastric, lung, breast, ovarian, and endometrial cancers, induces aberrant FGFR2 signalling activation (Katoh, 2008).

CNVs have not only been linked to tumour growth, but have also been implicated in tumour drug resistance. Methotrexate resistance in human squamous cell carcinomas has been attributed to an increased dihydrofolate reductase (DHFR) content, as a result of an increased *DHFR* gene copy number (Frei *et al.*, 1984). Similarly, copy number amplification of the driving oncogene in BRAF V600E or KRAS G13D mutant colorectal cancers can overcome inhibition of the mitogen-activated or extracellular signal-regulated protein kinase kinases 1 and 2 (MEK1/2) caused by the cytostatic drug AZD6244, also known as Selumetinib (Corcoran *et al.*, 2010; Little *et al.*, 2011).

Double minutes (dmin) are also a cytogenetic hallmark of genomic amplifications in cancer and have a major effect on the copy number of genes, although the function of dmin in cancer is largely unknown (Storlazzi *et al.*, 2010). Sequenced eccDNAs from HeLa cancer cells found that some eccDNAs contain unique sequence and so are not always formed from repetitive regions (van Loon, Miller and Murnane, 1994).

CNV arises from errors in DNA replication and repair, which can lead to deregulation of gene expression and dysfunction associated with their genetic diseases. However, these errors need not always be harmful to the cell and in some specific instances, often in response to a strong environmental pressure, CNV of protein coding genes can be beneficial, giving a cell a survival advantage in its environment. In the case of cancer, this is altogether to the detriment of the host, but does highlight the importance of CNV in environmental adaptation.

1.5 DNA damage and double-strand break repair

CNV has long been proposed to occur through meiotic recombination between mismatched repeats (Lupski, 1998). Non-allelic homologous recombination (NAHR) between chromosomes (interchromosomal), between chromatids of the same chromosome (intrachromosomal or interchromatid), within the same chromatid (intrachromatid), or by sister chromatid exchange can occasionally result in deletions and duplications through unequal crossing over (Kornreich, Bishop and Desnick, 1990; Rüdiger *et al.*, 1991; Marcus *et al.*, 1993; Olds *et al.*, 1993; Pousi *et al.*, 1994; Liu *et al.*, 2012). NAHR favours deletions over duplications, as deletions can occur both in *cis* and in *trans*, whereas duplications can only result from crossovers in *trans* (Liu *et al.*, 2012).

However, CNV mutations are not limited to the germline, as proven by monozygotic twin studies where each twin displays a different CNV profile (Bruder *et al.*, 2008). CNV has also been shown to vary between different organs and tissues of an individual (Piotrowski *et al.*, 2008). Therefore CNV is also possible in somatic cells and somatic CNV can arise as a result of errors in DNA replication (Hastings *et al.*, 2009). DNA double-strand breaks (DSBs) appear to be the initiating event for CNV rearrangements in both somatic and germline cells, with error-prone DSB repair pathways having the potential to result in a CNV event.

1.5.1 Endogenous vs. exogenous DNA damage

DNA damage can be both endogenous and exogenous. Endogenous genetic damage to a cell is an inevitable consequence of many bi-products formed from normal cellular processes, such as hydrolysis, oxidation, alkylation, and mismatch of DNA bases (Hakem, 2008). In reality the genome of a cell is continuously being damaged (Bernstein *et al.*, 2013). For instance, oxidant by-products of normal metabolism cause extensive DNA damage, with approximately 74,000-100,000 incidents of oxidative damage to DNA per cell per day in rats and 10,000-12,000 in humans (Fraga *et al.*, 1990; Ames, Shigenaga and Hagen, 1993; Helbock *et al.*, 1998). During unperturbed replication, the cellular DNA repair pathways usually cope with all the endogenous damage, despite its high frequency, often by pausing the DNA replication machinery at the site of damage until it has been repaired (Maher, Branagan and Morrical, 2011; Nam and Cortez, 2011). Alternatively, DNA damage tolerance mechanisms, like translesion DNA synthesis (TLS), bypass the damaged site by using specialized polymerases to synthesise across, or extend from, the site of damage (Prakash, Johnson and Prakash, 2005; Gali *et al.*, 2017). TLS polymerases have lower fidelity than replicative polymerase, since they lack proof-reading activity, meaning that damage bypass can be error-free or error-prone depending on whether or not the correct nucleotide is inserted opposite a lesion (Sale, 2013). Despite the best efforts of cells to resolve all DNA damage, unresolved damage can accumulate in non-dividing cells, contributing to aging and age-related cellular defects (Holmes, Bernstein and Bernstein, 1992; Hoeijmakers, 2009).

In contrast, exogenous DNA damage is caused by external environmental factors, such as ionizing radiation (IR), ultraviolet (UV) radiation, and various chemicals agents (Hakem, 2008). Exogenous DNA damage is often more harmful to cells than endogenous damage, most likely because exogenous damage causes higher levels of DNA damage and overloads the DNA repair pathways, and/or induces more novel types of DNA damage that are harder and take longer to repair (Bernstein *et al.*, 2013). Exogenous DNA damage is just as damaging to replicating and non-replicating cells, but in replicating

cells there is an increased risk of developing cancer (Bernstein *et al.*, 2013). This is because damage-induced mutations in oncogenes, tumour suppressor genes, genes involved in genomic stability, or other driver mutation events, can generate a clonal cell population with a distinct advantage in proliferation (Basu, 2018). Endogenous and exogenous DNA damage can result in easily repaired single strand breaks or more problematic double strand breaks.

1.5.2 Single-strand break repair

Cells have various strategies to protect and repair its genome from high levels of endogenous and exogenous DNA damage. For repair of single strand breaks (SSBs), these include base excision repair (BER), nucleotide excision repair (NER) and DNA mismatch repair (MMR) (Caldecott, 2008). BER and MMR repair small, non-helix-distorting damaged bases, whilst NER deals with bulky, helix-distorting DNA adducts (Reardon and Sancar, 2001; Kim and Wilson, 2012). These SSB repair mechanisms all follow four general steps (Fig. 1.3). 1) Recognition of single-stranded DNA (ssDNA) damage, such as an incorrect or damaged base, an abasic site, or loss of a run of one or more nucleotides. 2) If necessary, marking and removal of just the damaged nucleotide or a short run of 2-20 nucleotides that includes the damaged site. 3) DNA polymerase repair with the correct sequence, using the template DNA strand with an undamaged complementary sequence. 4) Ligation of the DNA phosphodiester bonds by DNA ligase to seal the nick between the repaired sequence and the rest of the DNA strand (Caldecott, 2008).

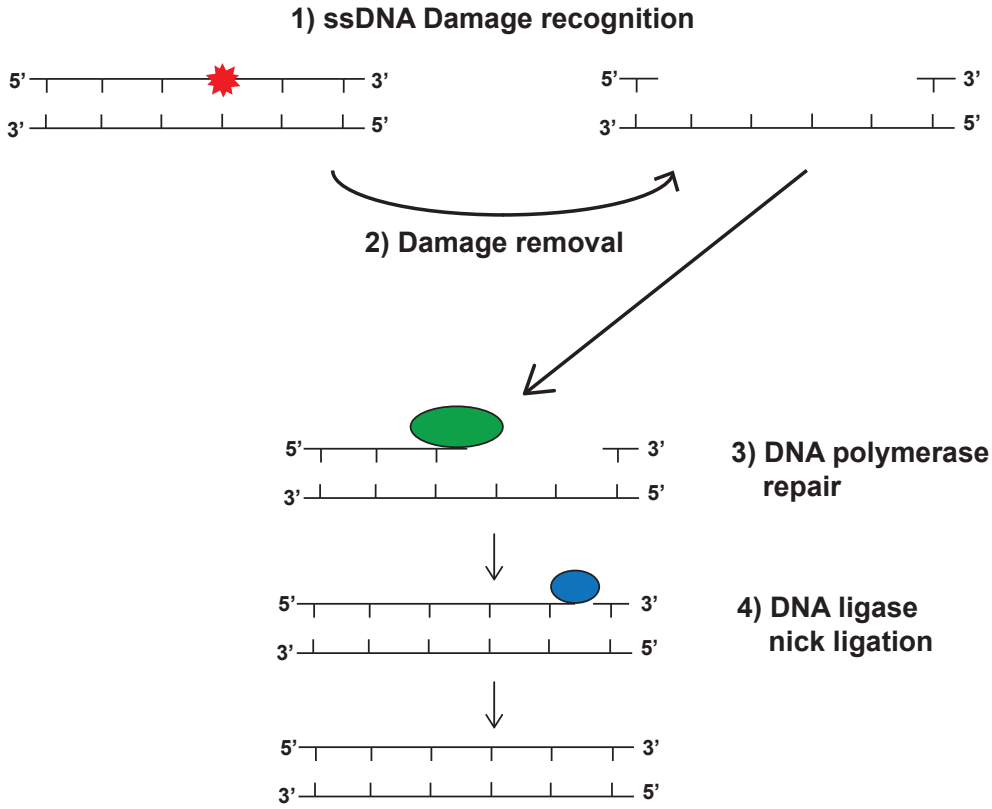


Figure 1.3: Single-strand break DNA damage repair. The cell can repair all single-strand breaks (SSBs) by base excision repair (BER), nucleotide excision repair (NER), or mismatch repair (MMR). All SSB repair begins with **1)** a recognition system for the type of ssDNA damage. If necessary, followed by **2)** removal of the damaged site and/or a run of 2-20 nucleotides. **3**) synthesis of the correct sequence by a DNA polymerase using the undamaged template DNA strand, and finishes by **4)** DNA ligase sealing the remaining gap in the phosphodiester DNA backbone.

1.5.3 Double-strand break repair

DNA double-strand breaks (DSBs) are a much greater issue for the cell, as failure to repair a DSB can lead to genomic instability and cell death (Ceccaldi, Rondinelli and D'Andrea, 2018). Also incorrect repair of DSBs can have deleterious consequences for the cell, such as genomic rearrangements, deletions, translocations and DNA fusions, all of which are commonly found in cancerous cells (Aplan, 2006; Cannan and Pederson, 2016). Most DSBs caused by endogenous damage are produced during DNA

replication when the DNA polymerase stalls upon encountering a SSB in the template strand, which causes the replication fork to collapse and a subsequent DSB to occur (Pfeiffer, Goedecke and Obe, 2000; Syeda, Hawkins and McGlynn, 2014). Replication forks can also stall at DNA binding proteins, DNA secondary structures, and transcription complexes, all of which can result in a DSB (Prado and Aguilera, 2005; Mirkin and Mirkin, 2007). However, fork stalling alone is not enough to guarantee that the fork will collapse, and at the replication fork barrier (RFB) in the rDNA repeats, the majority of stalled forks are stably maintained until they converge with another oncoming fork (Kobayashi *et al.*, 1998; Tsang and Carr, 2008; Irmisch *et al.*, 2009).

Cells have various mechanisms to repair DSBs (Fig. 1.4), which can broadly be split into homologous and non-homologous repair pathways, which differ in the requirement or not for an intact homologous sequence as a template for repair (Aguilera and Gómez-González, 2008; Huertas, 2010). There is also a third DSB repair pathway, known as microhomology-mediated end joining (MMEJ), which uses parts of both the homologous and non-homologous repair (McVey and Lee, 2008).

Non-homologous end joining (NHEJ) is the only full member of the non-homologous DNA repair branch. In addition to its role in DNA repair, vertebrate cells use NHEJ for V(D)J recombination and class switch recombination to produce a novel and diverse immunoglobulin repertoire (Lieber, 2010). NHEJ primarily operates in G0 and G1 phases of the cell cycle, and for haploid cells in G0 or G1 NHEJ is currently believed to be the only way to repair DSBs, as the broken chromosome has no homologue for repair by homologous recombination (HR) (Gao *et al.*, 2016). NHEJ does not require any strand resection and works by directly ligating the two broken sides of the DSB, often benefitting from a complementary overhang of just 1 or 2 base pairs between the break termini, known as a region of microhomology (Pannunzio *et al.*, 2014). In *Saccharomyces cerevisiae*, the Mre11-Rad50-Xrs2 (MRX) complex is recruited to the DSB ends, where it removes damaged or mismatched nucleotides to be re-filled by DNA polymerase, leaving the 3’ OH and 5’ phosphate termini accessible for the

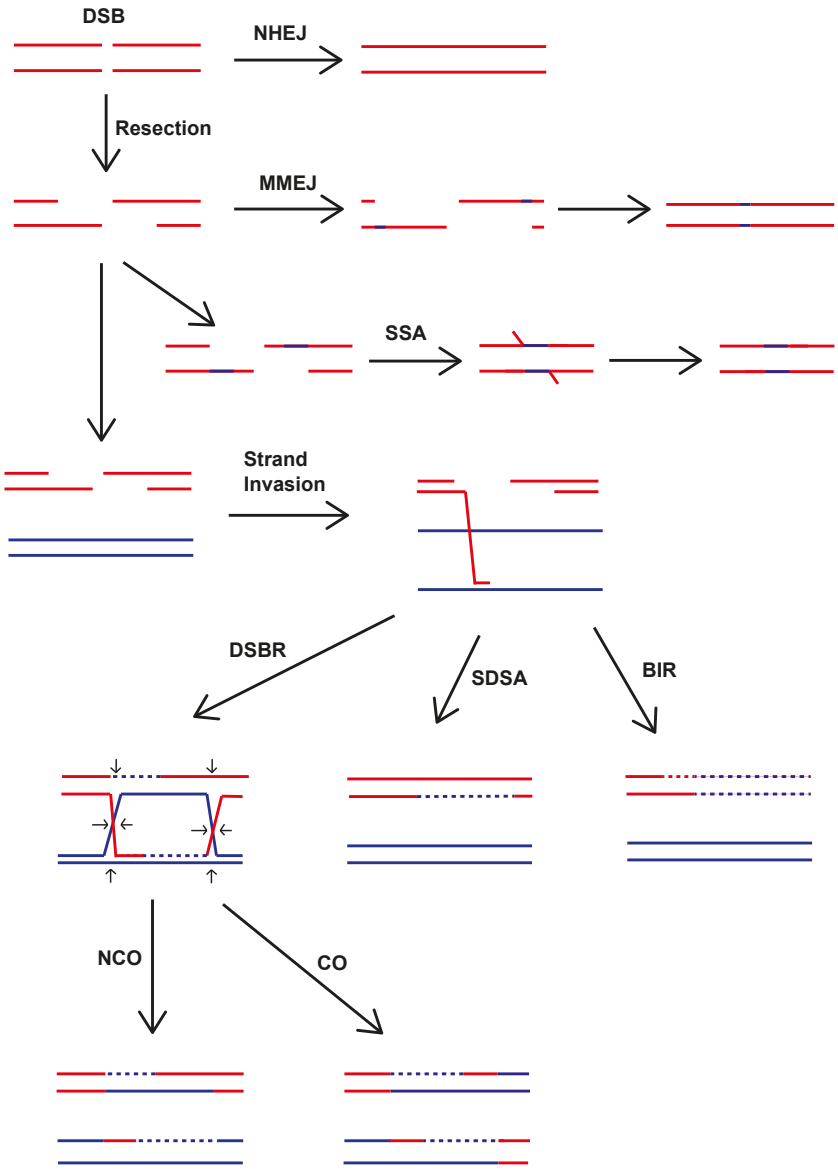


Figure 1.4: Mechanisms of DNA double-strand break repair. Cells can repair a double-strand break (DSB) without the need for a homologous template sequence by simply ligating the two broken ends of the DSB in non-homologous end joining (NHEJ). Alternatively, the 5' ends of the DSB are resected and repair proceeds using a homologous template. In single-strand annealing (SSA), the resection reveals a region of homology either side of the DSB for annealing and flap removal. For double-strand break repair (DSBR), synthesis-dependent strand annealing (SDSA), and break induced replication (BIR), a 3' overhang invades the homologous template. In DSBR, the other 5' end is captured and forms a double Holliday junction (dHJ) that can resolve with (CO) or without (NCO) crossovers. In SDSA, the invading 3' end does limited synthesis before being displaced and captured by the other 5' end. In BIR, the invading 3' end synthesises to the end of the chromosome. Microhomology-mediated end joining (MMEJ), utilizes both HR-dependent and non-HR-dependent processes, with continued resection of the DSB until a 2-3bp region of microhomology is exposed for ligating the two DSB ends.

Dnl4-Lif1 DNA ligase complex to resolve the double-strand break and merge the two broken ends (Chen *et al.*, 2001). NHEJ can be further stimulated by the Yku70p/Yku80p complex, which is the yeast homolog of the mammalian Ku70/Ku80 heterodimer (Chen *et al.*, 2001). NHEJ is highly efficient at DSB repair, but the repair is often more mutagenic than HR-dependent repair pathways (Huertas, 2010). This is because NHEJ has no recognition system to detect whether the two sequences either side of the DSB that are being repaired were originally contiguous, potentially leading to nucleotide loss, inversions and translocations (Huertas, 2010).

DSB repair by homologous recombination (HR) can be sub-divided into four repair pathways: 1) single-strand annealing (SSA), 2) double-strand break repair (DSBR), 3) synthesis-dependent strand annealing (SDSA), and 4) break-induced replication (BIR) (Huertas, 2010). HR-dependent repair, although involved in DSB repair, is believed to primarily function in resolving stalled or collapsed replication forks (Aguilera and Gómez-González, 2008). All HR repair mechanisms require a region of perfect, or near perfect, sequence homology on an undamaged, complementary template strand to drive the repair process (Krogh and Symington, 2004). All HR repair mechanisms also begin with 5' to 3' degradation of the DNA strands either side of the DSB, which leaves long stretches of overhanging 3' single-stranded DNA that is then bound by the single-stranded DNA binding complex Replication protein A (RPA) (Krogh and Symington, 2004). After strand resection and RPA coating, the HR pathways diverge.

For SSA there must be a homologous region in both of the single-stranded sequences either side of the DSB, which is primarily true for repeated sequences. The homologous sequences first anneal, then any DNA overhang is cleaved, and finally the strands are re-ligated (Lin, Sperle and Sternberg, 1984, 1985; Pâques and Haber, 1999). SSA results in a small deletion of the region between the end of DSB and the sites of homology (Huertas, 2010). If the homologous region used for SSA is at least 400bp it is nearly 100% efficient in yeast, but this drops to just 5% of cells surviving the DSB if there is only 60bp of homology (Sugawara and Haber, 1992).

In DSBR, SDSA and BIR, the HR-protein Rad52 targets Rad51 to displace RPA on the coated single-stranded DNA, forming the Rad51 nucleoprotein

filament (Song and Sung, 2000; Sugiyama and Kowalczykowski, 2002). Extension of the Rad51 nucleoprotein filament and RPA displacement is catalysed by Rad55 and Rad57 (Krogh and Symington, 2004). The Rad51 nucleoprotein filament seeks out a region of homology on another chromatid by strand invasion, which is catalysed by Rad54 to promote chromatin remodelling, DNA unwinding and strand annealing between the donor DNA and the incoming Rad51 nucleoprotein filament (Solinger and Heyer, 2001; Solinger, Kiianitsa and Heyer, 2002; Krogh and Symington, 2004). After strand invasion and displacement loop (D-loop) formation, the DSBR, SDSA and BIR pathways diverge.

During DSBR, DNA synthesis begins on the invading 3' strand, displacing one of the donor DNA duplex strands for the other overhanging 3' end of the DSB to anneal to and begin a second round of leading strand synthesis (Orr-Weaver, Szostak and Rothstein, 1981; Szostak *et al.*, 1983; Symington, Rothstein and Lisby, 2014). The newly synthesized stretches of DNA are then re-ligated to 5' ends of the DSB, forming a double Holliday junction intermediate (dHJ) (Orr-Weaver, Szostak and Rothstein, 1981; Szostak *et al.*, 1983; Symington, Rothstein and Lisby, 2014). Resolution of the dHJ is performed by a helicase and topoisomerase, which does not cause any crossover events (NCO) (Orr-Weaver, Szostak and Rothstein, 1981; Szostak *et al.*, 1983; Symington, Rothstein and Lisby, 2014). Alternatively endonucleases can resolve the dHJ, which can produce crossovers (COs) (Orr-Weaver, Szostak and Rothstein, 1981; Szostak *et al.*, 1983; Symington, Rothstein and Lisby, 2014).

The SDSA model only produces NCOs and is predicted to work by both of the 3' single-stranded DNA overhangs invading the template DNA duplex and undergoing partial DNA synthesis (Nassif *et al.*, 1994; Symington, Rothstein and Lisby, 2014). Then the partially repaired strands are displaced by a DNA helicase, with annealing of the nascent complementary strands, gap filling and ligation (Nassif *et al.*, 1994; Symington, Rothstein and Lisby, 2014). Alternatively only one of the two 3′ single-stranded DNA overhangs invades the homologous DNA duplex, undergoes limited DNA synthesis before dissociating and annealing to the 3′ single-stranded DNA tail at the other side

of the DSB (Ferguson and Holloman, 1996; Symington, Rothstein and Lisby, 2014).

Finally, in BIR a single 3' end of a break invades the template DNA duplex and proceeds to replicate the DNA to the end of the chromosome (Kraus, Leung and Haber, 2001; Llorente, Smith and Symington, 2008; Symington, Rothstein and Lisby, 2014). As such, BIR results in extensive loss of heterozygosity (LOH) events (Symington, Rothstein and Lisby, 2014). On top of this, BIR often proceeds through several rounds of strand invasion, synthesis, and dissociation, which causes chromosome rearrangements when the dissociation and re-invasion occur within dispersed repeated sequences (Smith, Llorente and Symington, 2007; Ruiz, Gomez-Gonzalez and Aguilera, 2009). Therefore BIR has the potential to be highly mutagenic and contribute to genome evolution, adaptation and/or disease (Symington, Rothstein and Lisby, 2014).

Under conditions of stress, where there is a depletion of available Rad51 for HR-dependent repair, but unaffected Rad52 levels, an alternative, error-prone form of classical BIR can occur, known as microhomology-mediated break-induced replication (MMBIR) (Hastings, Ira and Lupski, 2009). Stalled replication forks repaired by MMBIR have low-processivity, with multiple-template switching events that cause complex rearrangements suggested to be the origin of CNV (Hastings, Ira and Lupski, 2009). The switch to error-prone repair by MMBIR is believed to be induced by Rad51 depletion, and in tumour cells undergoing hypoxia, where *RAD51* expression is repressed, there is an increase in genomic instability (Coquelle *et al.*, 1998).

Yeast can even undergo a form of BIR repair that is independent of Rad51 and Rad52 proteins, known as microhomology/microsatellite-induced replication (MMIR) (Payen *et al.*, 2008). MMIR is much less efficient, and is believed to not even require a DSB, but instead utilises a template switch between microsatellites or microhomologous sequences (Payen *et al.*, 2008).

Perhaps surprisingly, cells can choose to switch from high-fidelity to error-prone DSB repair mechanisms, such as MMBIR and MMIR, in a controlled manner in response to stress, showing a clear parallel between the environment and the chosen DNA DSB repair strategy (Ponder, Fonville and Rosenberg, 2005).

The last DSB repair mechanism is microhomology-mediated end joining (MMEJ), which displays characteristics of both HR-dependent and non-HR-dependent repair (Truong *et al.*, 2013). MMEJ requires an initial end resection step, like all HR-dependent pathways, and mainly operates in S-phase, but is otherwise analogous to NHEJ (Truong *et al.*, 2013; Cannan and Pederson, 2016). In MMEJ, the 5' DNA strands either side of the DSB are resected until they reveal a microhomology of 5-25 complementary base pairs on both sides of the DSB (Cannan and Pederson, 2016). The end resection step in MMEJ and HR are both catalysed by the MRX complex (Wang and Xu, 2017), but in MMEJ this end resection is independent of the long range resection factors Sgs1 and Exo1, with *exo1*Δ *sgs1*Δ mutants undergoing a significantly increased frequency of MMEJ (Deng *et al.*, 2014; Wang and Xu, 2017). Following end resection, the two sides of the DSB anneal via the revealed region of microhomology, non-annealed DNA ends are removed, and the remaining gaps filled and ligated (Cannan and Pederson, 2016). MMEJ is even more error-prone than NHEJ, always resulting in deletions, as it loses all DNA sequence between the break site and the region of microhomology (McVey and Lee, 2008). MMEJ has often been considered a 'back-up' DSB repair mechanism for when NHEJ and other mechanisms fail (Wang *et al.*, 2003; Lieber *et al.*, 2004). However, such assumptions contribute to the idea of MMEJ being irrelevant under normal physiological conditions, whereas MMEJ is surprisingly robust, especially in V(D)J recombination or when exposed to genomic stresses, and actively competes with NHEJ and all HR-dependent repair pathways (Truong *et al.*, 2013; McVey and Lee, 2008). Therefore MMEJ is not just a back-up repair pathway for resolving DSBs and maintaining cell viability.

1.6 Mechanisms effecting the rate of CNV

CNV can arise as a result of errors in DNA recombination, replication, and DSB repair (Hastings *et al.*, 2009; Weischenfeldt *et al.*, 2013; Bai *et al.*, 2016). Multiple processes are known to alter the rates of CNV and some of these are responsive to cellular control mechanisms (Hastings *et al.*, 2009; Arlt, Wilson

and Glover, 2012; Wilson *et al.*, 2015; Zheng *et al.*, 2016). Therefore there is a lot of evidence to suggest that CNV events are not simply random mutations, and could even be an adaptive process important for the rapid evolution of cells in response to challenging environments (Zheng *et al.*, 2016).

The rDNA repeat array in *Saccharomyces cerevisiae* is a highly specialised locus with multiple, well-characterised mechanisms in place to maintain optimum copy number (Szostak and Wu, 1980; Kobayashi *et al.*, 1998, 2004; Takeuchi, Horiuchi and Kobayashi, 2003; Houseley and Tollervey, 2011; Ide, Saka and Kobayashi, 2013; Jack *et al.*, 2015). As such the budding yeast rDNA has become a model for the study of CNV. The rate of CNV at the rDNA is influenced by factors, such as local transcription, epigenetic regulation by the Sir2-family of histone deacetylases (HDACs), and TOR signalling (Takeuchi, Horiuchi and Kobayashi, 2003; Kobayashi *et al.*, 2004; Jack *et al.*, 2015).

1.6.1 The Ribosomal DNA model

The budding yeast ribosomal DNA (rDNA) has been used extensively for studies on CNV (Fig. 1.5). The rDNA repeat array is maintained at approximately 150 copies, each containing a 9.1kb sequence that encodes the ribosomal RNAs, and undergoes frequent CNV (Szostak and Wu, 1980). By comparison, humans have between 400 and 600 rDNA repeats, split across 5 chromosomes (Krystal *et al.*, 1981). Within each rDNA repeat of *Saccharomyces cerevisiae* there is a replication fork barrier (RFB) from which almost all recombination events are initiated (Keil and Roeder, 1984). The nucleolar protein Fob1 is required for both stalling of the processive replication machinery at the rDNA RFB and homologous recombination at the HOT1 recombination hotspot, with *fob1Δ* mutants being unable to recombine their rDNA (Kobayashi and Horiuchi, 1996). The RFB is located at the 3' end of the 35S pre-RNA gene and arrests approximately 90% of leftward-moving replication forks, thereby blocking head-on collisions between the replication and transcription machineries (Brewer, Lockshon and Fangman, 1992). In

fob1Δ cells, where replication fork blocking activity at the RFB is lost, there is an increase in the number of collisions between the transcription unit and the replication fork travelling in the opposite direction to transcription (Takeuchi, Horiuchi and Kobayashi, 2003). This transcription-dependent fork collision induces recombination and has been proposed as a mechanism for CNV of highly-transcribed single-copy genes (Takeuchi, Horiuchi and Kobayashi, 2003).

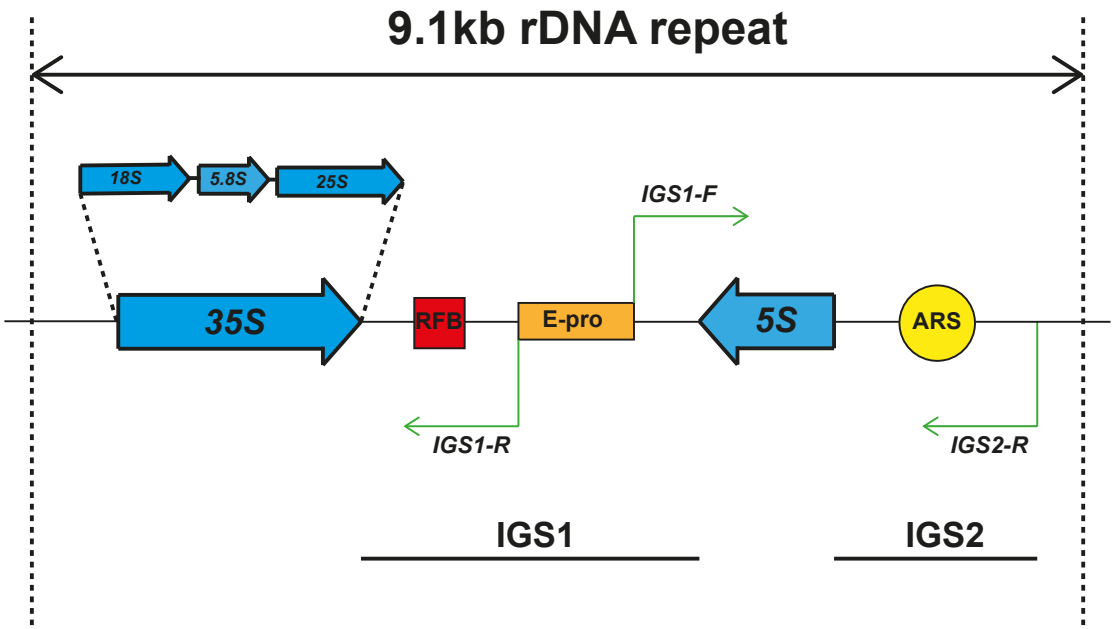


Figure 1.5: Schematic of the rDNA locus. Each 9.1kb rDNA repeat contains the 35S pre-rRNA, the 5S rRNA, a replication fork barrier (RFB), an autonomous replicating sequence (ARS), two intergenic spacer regions (IGS1 and IGS2) and an E-pro promoter from which the *IGS1-F* and *IGS1-R* non-coding RNAs (ncRNAs) are transcribed. A further ncRNA, *IGS2-R*, is transcribed from IGS2.

The stalled forks at the rDNA RFB are considered to be relatively stable. This is most likely because both the leading and lagging strands stall close to each other, and therefore there is a lack of available ssDNA at the forks to promote recombination and checkpoint activation (Gruber, Wellinger and Sogo, 2000). As such, the replication forks can therefore remain stalled for a prolonged period of time, only to resume replication once the barrier is removed (Labib

and Hodgson, 2007). Alternatively, if the forks collapses, DNA repair through a recombination event is required before replication can restart (Tsang and Carr, 2008). This potentially explains the elevated recombination rate observed at the rDNA locus (Irmisch *et al.*, 2009).

1.6.2 Recombination at the rDNA

When a stalled replication fork collapses at the rDNA RFB, a DSB often forms nearby (Weitao *et al.*, 2003). Replication restart now requires a DSB repair event. The DSB repair pathway used at the rDNA appears to be the HR-dependent BIR mechanism, as long as the interaction with the sister chromatid is not blocked (Houseley and Tollervey, 2011). As a result of the homology between rDNA repeats, a lot of the BIR events search out a mismatched repeat on the sister chromatid and repair via NAHR, resulting in CNV (Dittwald *et al.*, 2013). Interestingly, CNV at the rDNA can still occur independently of the essential HR proteins Rad52, Rad51 and Rad59, showing that alternate HR-independent repair pathways can also resolve the rDNA DSB when BIR with the sister chromatid is blocked (Houseley and Tollervey, 2011).

Rad52-independent CNV could be the result of replication Fork Stalling and Template Switching (FoSTeS) (Fig. 1.6). The FoSTeS model proposes that the lagging strand at a collapsed fork in one repeat disengages and anneals to a neighbouring repeat template sequence that has an active replication fork (Lee, Carvalho and Lupski, 2007). The template sequence is copied and then can disengage again and repeat the process, or ligate to the newly synthesised strand by the same mechanism as for Okazaki fragments (Lee, Carvalho and Lupski, 2007). FoSTeS, along with MMBIR, have been proposed as being responsible for complex genomic rearrangements and therefore are potentially very important in the evolution of genomes (Zhang, Carvalho and Lupski, 2009).

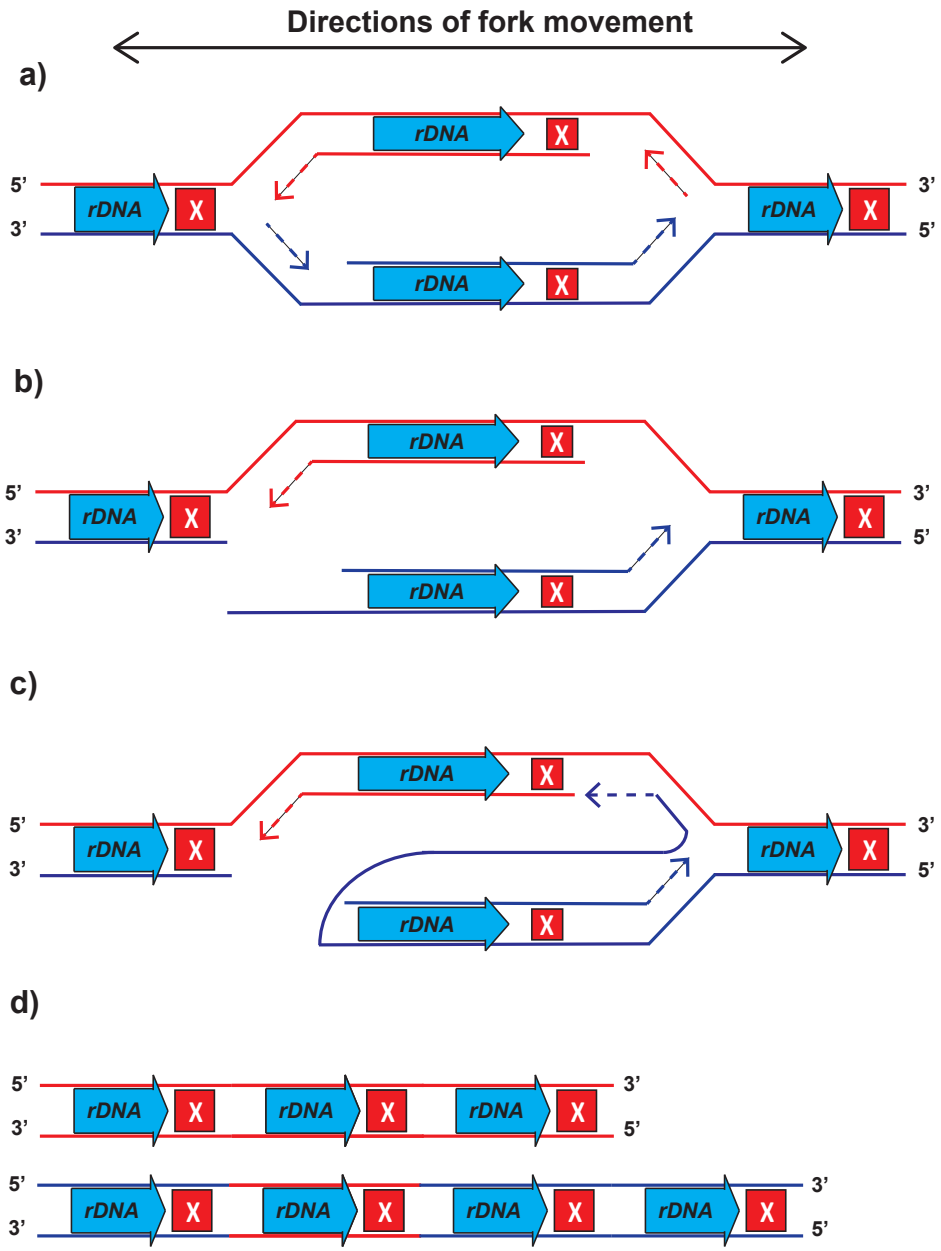


Figure 1.6: Rad52-independent CNV at the rDNA. **a**) A replication bubble at the rDNA. **b**) At one end of the replication bubble, the stalled replication fork collapses at the rDNA RFB, leaving a ssDNA region from lagging strand synthesis. **c**) The region of ssDNA shares homology and anneals to the exposed ssDNA region of another replication fork and is extended and ligated to the leading strand by the same mechanism as an Okazaki fragment. **d**) Continued lagging strand synthesis at the invaded fork primes leading strand synthesis in the new fork, the collapsed replication fork restarts by an as yet unknown mechanism, and the chromosome is replicated with the addition of an rDNA repeat in one of the replicated chromosomes. Blue arrows represent rDNA repeats. A white cross in a red square represents the rDNA RFB. Solid red and blue lines represent the Watson and Crick DNA strands of chromosome XII respectively. Dashed arrows represent newly synthesised DNA, with the same colour coding for the Watson and Crick DNA strands.

1.6.3 Transcription and CNV

Replication fork stalling is not a process unique to the rDNA RFB. Fragile sites are loci that are prone to breakage and include sites of replication fork stalling. Fragile sites that arise from replication fork stalling have been found genome wide in *Saccharomyces cerevisiae*, as identified by γH2A, the early response marker of DNA damage (Szilard *et al.*, 2010). These sites of aberrant replication are also more prone to CNV mutations, most likely because of the elevated level of DNA repair and recombination needed at these loci (Slack *et al.*, 2006; Liu *et al.*, 2012; Sima and Gilbert, 2014). Interestingly, about half the γH2A enriched sites map to repressed protein-coding genes, and are dependent upon histone deacetylases for the formation of γH2A at the loci (Szilard *et al.*, 2010). Given that collisions between the replication and transcription machinery are particularly mutagenic (Aguilera and Garcia-Muse, 2013; Sankar *et al.*, 2016), protein-coding genes containing fragile sites may have evolved to be repressed under normal growth conditions, so as to avoid regularly undergoing potentially detrimental recombination events. Alternatively, natural selection acting on randomly located fragile sites genome-wide, would retain only the fragile sites at protein-coding genes that are repressed under normal growth conditions, as fragile sites that regularly undergo potentially detrimental recombination events would be rapidly lost. It is also possible that in a challenging environment, where the normally repressed protein-coding genes containing fragile sites are induced, increased recombination at these genes confers a fitness benefit, which is why the fragile site is retained over evolution. This fitness benefit could be increased population heterogeneity, resulting from increased errors associated with DNA recombination such as CNV events, which may aid the adaptation of a population of cells to the challenging environment. Transcription alone can also have an effect on the mutation rate, irrespective of collisions with stalled replication forks, with highly transcribed loci being more prone to spontaneous mutation (Thomas and Rothstein, 1989; Datta and Jinks-Robertson, 1995). Taken together this means that normally repressed genes containing fragile sites experience a large amount of

genome instability at the locus upon high transcriptional induction, thus making them vastly more prone to CNV. In addition, bacterial studies have shown transcription to directly cause replisome dissociation (Mangiameli *et al.*, 2017). They have also shown that head-on collisions resulting from transcription orientated against the direction of replication, result in higher mutation rates than observed for co-directional conflicts (Paul *et al.*, 2013). Another way transcription can interfere with replication is through R-loop formation (Sollier and Cimprich, 2015). R-loops are three-stranded nucleic acid structures, involving an RNA-DNA hybrid, where the transcribed RNA strand invades the DNA duplex and binds to the complementary DNA sequence, causing a partial displacement of the DNA duplex (Thomas, White and Davis, 1976). R-loops are involved in replication and recombination, where they can be harmful structures, but can also act as important intermediates in specific cellular processes (Aguilera and García-Muse, 2012).

R-loops can expose a region of single-stranded DNA, during formation of the RNA-DNA hybrid, which is more susceptible to transcription-associated mutagenesis (TAM) and transcription-associated recombination (TAR) (Skourti-Stathaki and Proudfoot, 2014). However, the exact mechanism of action of R-loops in causing genome instability is unknown, but it has been predicted that spontaneous DNA damage, such as deaminations on the unpaired DNA template strand, lead to DSBs and recombination (Aguilera, 2002; Li and Manley, 2006; Aguilera and Garcia-Muse, 2012). As such, sites of R-loop accumulation make the locus more prone to mutagenesis, or a protein that recognises R-loops initiates mutagenesis at the locus (Skourti-Stathaki and Proudfoot, 2014). Either way, R-loops are an important product of transcription that can influence genome stability.

R-loops are resolved by Ribonuclease H (RNase H) enzymes, RNA/DNA helicases like Sen1, and topoisomerases that relax the DNA negative supercoiling that encourages persistent R-loop formation (Drolet, Bi and Liu, 1994; Drolet *et al.*, 1995; Cerritelli and Crouch, 2009; Tuduri *et al.*, 2009; El Hage *et al.*, 2010; Mischo *et al.*, 2011). R-loops can also be prevented from forming in *rad51* and *rad52* homologous recombination mutants, suggesting that these proteins are required for strand exchange in the formation of the

RNA-DNA hybrid (Wahba, Gore and Koshland, 2013). Other mRNA biogenesis and processing proteins can also prevent R-loop formation, such as the RNA-binding protein Npl3, and the THO/TREX complex (Huertas and Aguilera, 2003; Dominguez-Sanchez *et al.*, 2011; Castellano-Pozo, Garcia-Muse and Aguilera, 2012; Santos-Pereira *et al.*, 2013).

The Bacteriophage T4 initiates replication, immediately after bacterial infection, using T4 RNase H processed R-loops as free 3' ends for lagging strand synthesis (Kreuzer and Brister, 2010). In *E. coli*, R-loops have been proposed to initiate replication independent of *oriC*, through RNase H1 processing to produce 3' ends that become extended by DNA Pol I (Itoh and Tomizawa, 1980; Aguilera and García-Muse, 2012). Similarly in eukaryotes, mitochondrial DNA replication has been shown to be primed by an RNA molecule produced by the mitochondrial RNA polymerase (Baldacci, Cherif-Zahar and Bernardi, 1984; Xu and Clayton, 1996; Pohjoismaki *et al.*, 2010). R-loops have also evolved as a natural source of Ig class-switch recombination in vertebrate B cells (Yu *et al.*, 2003; Aguilera and García-Muse, 2012). In *Schizosaccharomyces pombe*, R-loops formed from non-coding RNAs (ncRNAs) have also been shown to mediate RNAi-directed heterochromatin formation (Nakama *et al.*, 2012). Therefore it is unsurprising that transcription can affect CNV, as transcription has many possible mechanisms to increase local genome instability.

Local transcription also appears to regulate the CNV rate at the model rDNA locus, by interfering with stalled replication forks at the RFB (Kobayashi *et al.*, 1998; French *et al.*, 2003; Jack *et al.*, 2015). Within each rDNA repeat, RNA polymerase I (Pol I) transcribes the 35S pre-rRNA gene containing the 18S, 5.8S and 25S rRNAs, RNA polymerase III (Pol III) transcribes the 5S rRNA gene, and RNA polymerase II (Pol II) transcribes the *IGS1-F*, *IGS1-R* and *IGS2-R* ncRNAs from the two intergenic spacer (IGS) regions. Recombination and therefore CNV at the rDNA is dependent upon Pol I transcription initiating from the recombination-stimulating sequence HOT1 (Stewart and Roeder, 1989). In addition, expression of the *IGS1-F* and *IGS1-R* ncRNAs from their shared E-pro bi-directional promoter, stimulates the dissociation of cohesin, sister-chromatid recombination and NAHR (Kobayashi and Ganley, 2005). *IGS1-F* is a stable ncRNA (SUT), whereas *IGS1-R* is a cryptic unstable

transcript (CUT), a class of noncoding RNA that is degraded instantly after transcription (Wyers *et al.*, 2005; Houseley *et al.*, 2007). *IGS1-R* is transcribed through the rDNA RFB, potentially causing interference with the stalled replication machinery, and is believed to be an important element for CNV at the rDNA (Houseley *et al.*, 2007).

1.6.4 Sir2-family of histone deacetylases

The Sir2-family of NAD+-dependent histone deacetylases (HDACs), also known as the sirtuins, consist of Sir2, Hst1, Hst2, Hst3, and Hst4 (Brachmann *et al.*, 1995; Imai *et al.*, 2000; Landry *et al.*, 2000). Sir2 was first identified as a silencing factor at the *Saccharomyces cerevisiae* mating type loci (Rine and Herskowitz, 1987), and then also at the telomeres through the upstream activating sequence (UAS)-binding protein Rap1 (Moretti *et al.*, 1994). Sir2 has also been shown to silence transcription at the rDNA (Smith and Boeke, 1997), where it is recruited to the rDNA by the RENT nucleolar complex (Straight *et al.*, 1999). The transcriptional silencing by Sir2 also inhibits recombination at the rDNA (Huang *et al.*, 2006), with *sir2* mutants displaying a 10- to 15-fold increase in mitotic and meiotic intrachromosomal recombination rates (Gottlieb and Esposito, 1989). Therefore Sir2 is involved in silencing rDNA transcription (Li, Mueller and Bryk, 2006) and inhibition of recombination at the rDNA (Gottlieb and Esposito, 1989), which may be independent or interconnected functions of Sir2.

The sirtuins Hst3 and Hst4 play a major role in preserving genome integrity by maintaining low levels of histone H3 lysine 56 acetylation (H3K56ac) outside of S phase (Fig. 1.7), with their loss resulting in increased spontaneous DNA damage, chromosome loss, thermosensitivity, and acute sensitivity to genotoxic agents (Celic *et al.*, 2006). Hst3 and Hst4 are selectively down-regulated by DNA damage checkpoint-mediated repression systems, essential for cells to survive DNA damage (Maas *et al.*, 2006; Miller, Maas and Toczyski, 2006). The DNA damage sensor Mec1 activates the DNA damage response through phosphorylation of Hst3, targeting it for ubiquitin-

mediated proteolysis, enabling H3K56ac accumulation at DNA damage sites (Thaminy *et al.*, 2007). Effects of Hst3 and Hst4 loss include the up-regulation of DNA-damage inducible genes and DNA damage markers such as Rad53, and synthetic lethality when combined with other non-lethal mutations in DNA replication and double-strand break repair, all of which appear to be H3K56ac-dependent phenotypes (Celic, Verreault and Boeke, 2008). Interestingly, the *hst3Δ hst4Δ* phenotypes can be suppressed in combination with replication factor C (RFC) complex mutations, such as overexpression of the PCNA clamp loader large subunit Rfc1, or in *ctf4Δ* replisome mutants that have defective sister chromatid cohesion (Celic, Verreault and Boeke, 2008). This suggests that H3K56ac causes DNA damage through interference with normal replisome functions.

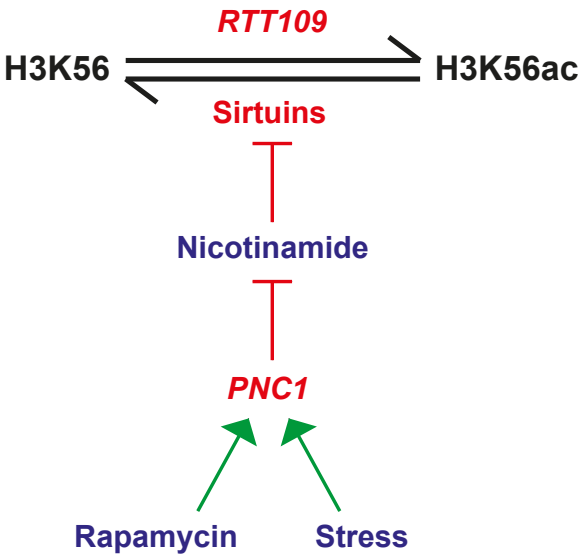


Figure 1.7: Regulation of Histone H3 Lysine 56 acetylation. Histone H3 Lysine 56 (H3K56) is acetylated by the acetyltransferase Rtt109 and deacetylated by the sirtuins. Nicotinamide is a natural repressor of the sirtuins and is turned over by the nicotinamidase Pnc1. Inhibition of the TOR pathway by rapamycin, or in response to stress, up-regulates *PNC1*, which indirectly activates the sirtuins by reducing nicotinamide levels.

1.6.5 HDAC inhibitor nicotinamide

The histone deacetylase reaction of the sirtuins is mechanistically coupled to the hydrolysis of the essential co-factor NAD+ to a single molecule of *o*-acetyl-ADP ribose and nicotinamide (Landry, Slama and Sternglanz, 2000; Tanny and Moazed, 2001). Nicotinamide is then converted into nicotinic acid by the nicotinamidase Pnc1 (Ghislain, Talla and Francois, 2002), and continues through the NAD+ salvage pathway (Fig. 1.8), to ultimately reform a molecule of NAD+ (Sandmeier *et al.*, 2002).

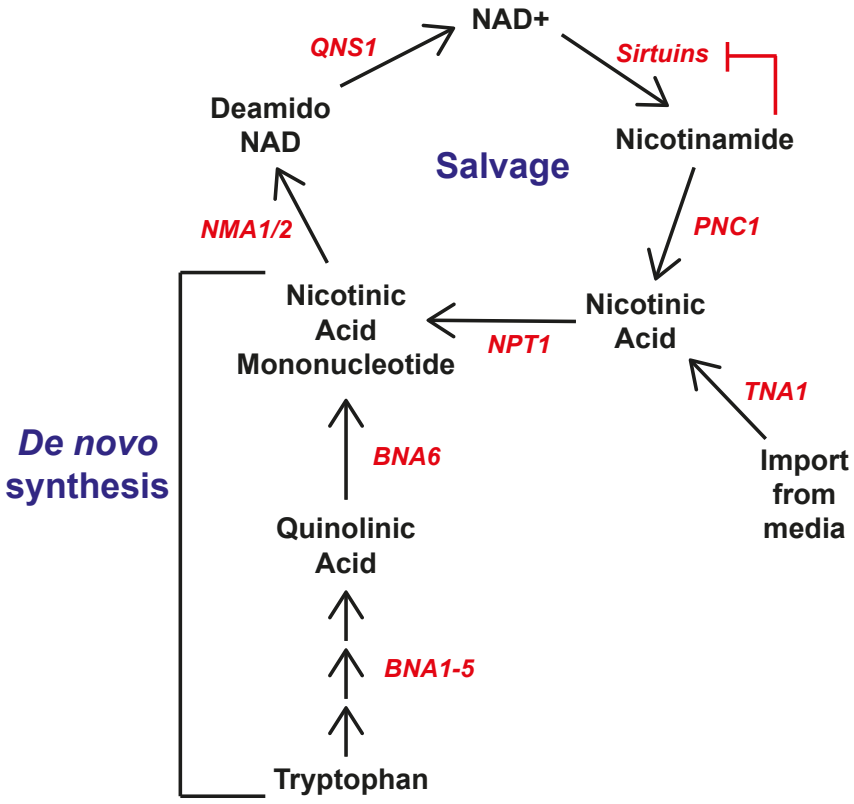


Figure 1.8: NAD+ *de novo* synthesis and salvage pathways. The essential co-factor NAD+ can be synthesised *de novo* starting from tryptophan. NAD+ is converted to nicotinamide as part of the sirtuin deacetylation reaction. Nicotinamide is a feedback inhibitor of the sirtuins, regulating its own production. Alternatively, nicotinamide is converted to nicotinic acid by Pnc1 and carries on around the NAD+ salvage pathway to regenerate NAD+. Nicotinic acid can also be imported into the cell directly by Tna1.

Nicotinamide is a non-competitive inhibitor of the sirtuins (Fig. 1.7 and 1.8), enabling nicotinamide to self-regulate its own production (Landry, Slama and Sternglanz, 2000; Bitterman *et al.*, 2002). Inhibition of the sirtuins by nicotinamide has been shown to strongly inhibit transcriptional silencing, decrease genome stability, increase rDNA recombination, and shorten replicative life span to levels observed in *sir2Δ* yeast (Bitterman *et al.*, 2002). Overexpression of *PNC1* can counter the effects of nicotinamide, by preventing the accumulation of nicotinamide and therefore positively regulating Sir2-mediated silencing and longevity (Gallo, Smith and Smith, 2004).

Caloric restriction has been shown to extend lifespan in yeast, worms, flies, mice, and potentially in primates and humans (Weindruch *et al.*, 1986; Lakowski and Hekimi, 1998; Roth, Ingram and Lane, 1999; Lin, Defossez and Guarente, 2000; Rogina and Helfand, 2004). The lifespan extension effect of caloric restriction has been shown to be dependent on *PNC1*, and that nicotinamide depletion and thus sirtuin activation is what is required for longevity (Anderson *et al.*, 2003). However the importance of Sir2 in longevity has been disputed in a more recent paper, showing that media from glucose-restricted cells is what is necessary and sufficient to produce lifespan extension, independent of Sir2 to produce or respond to the activity (Mei and Brenner, 2015). *PNC1* is also induced under low-intensity stress, preventing the accumulation of the inhibitory nicotinamide, showing the importance of the sirtuins in regulating critical cellular processes in response to changes in the environment (Anderson *et al.*, 2003; Gallo, Smith and Smith, 2004).

Repression of the Sir2-family HDACs by nicotinamide, significantly increases the H3K56ac level in all regions of the rDNA locus (Ha and Huh, 2011). Nicotinamide treatment also stimulates rapid rDNA repeat amplifications in a low-repeat strain, which could be blocked by overexpression of *PNC1* (Jack *et al.*, 2015). As such, H3K56ac is likely important for CNV at the rDNA, with the rate of CNV being controlled by sirtuin activity in response to the level of nicotinamide in the environment.

1.6.6 Rapamycin and TOR signalling

Repression of the Target of Rapamycin (TOR) signalling pathway can also overexpress *PNC1*, increasing Sir2 activity and stabilising the rDNA (Medvedik *et al.*, 2007). The TOR proteins are members of the phosphatidylinositol kinase (PIK)-related kinases family of proteins that sense and respond to nutrient availability and cellular stresses (Kuruvilla and Schreiber, 1999; Shamji, Kuruvilla and Schreiber, 2000; Schneper, Duvel and Broach, 2004). In yeast, there are two structurally and functionally similar, but not identical, TOR proteins encoded by *TOR1* and *TOR2* (Heitman, Movva and Hall, 1991; Helliwell *et al.*, 1994). *In vivo* the TOR proteins combine with other proteins to form two functionally distinct complexes, TOR complex 1 (TORC1) and TOR complex 2 (TORC2) (Loewith *et al.*, 2002). TORC1 can contain either *TOR1* or *TOR2* and is sensitive to rapamycin, whereas TORC2 always contains *TOR2* and is rapamycin insensitive (Loewith *et al.*, 2002). TORC1 and TORC2 are involved in completely independent cellular processes. TORC1 functions include translation initiation, ribosome biogenesis, autophagy, nutrient uptake, and metabolism (Wei and Zheng, 2011). TORC2 is important for actin cytoskeleton dynamics and polarised cell growth (Schmidt, Kunz and Hall, 1996; Jacinto *et al.*, 2004).

The TOR signalling pathway controls various cellular processes, such as the nitrogen discrimination pathway (NDP) that becomes activated when glutamine levels are low, the retrograde (RTG) response that is activated by faulty TCA cycle enzymes or mitochondria, cell growth and proliferation, and ribosome biogenesis (Schneper, Duvel and Broach, 2004). TOR is able to regulate many unique transcriptional and physiological processes at the same time through a kinase signalling cascade (Duvel and Broach, 2004). Multiple proteins, at various steps in the kinase signalling cascade, are regulated by phosphatases to direct TOR activation to the appropriate physiological response for the environment (Duvel and Broach, 2004).

Rapamycin is a lipophilic macrolide immunosuppressive drug that binds to the FK506-binding protein (FKBP) proline rotamase, encoded by *FPR1*, which is essential for the toxicity of rapamycin (Heitman *et al.*, 1991; Heitman, Movva

and Hall, 1991). FKBP was revealed to be non-essential for growth (Heitman *et al.*, 1991; Koltin *et al.*, 1991; Tanida *et al.*, 1991; Wiederrecht *et al.*, 1991) and later shown to only act as a co-factor or receptor for rapamycin (Loewith and Hall, 2011). Cells that develop resistance to rapamycin were found to have mutations in either *FPR1*, *TOR1*, or *TOR2*, confirming that the target of FKBP-rapamycin is TOR (Heitman, Movva and Hall, 1991; Cafferkey *et al.*, 1993). Upon rapamycin binding, FKBP is hijacked or corrupted to now interact with *TOR1* and *TOR2* (Loewith and Hall, 2011). The FKBP-rapamycin complex targets a conserved serine residue in the conserved PI kinase-related domain of the yeast Tor1 (Ser1972) and Tor2 (Ser1975) proteins (Stan *et al.*, 1994; Lorenz and Heitman, 1995). Cells treated with rapamycin arrest growth in early G1-phase and display inhibited translation initiation that is characteristic of starved cells entering stationary phase, suggesting that rapamycin mimics the effects of nutrient depletion (Barbet *et al.*, 1996). Repression of TOR signalling by rapamycin promotes the re-localisation of the transcription factors Msn2 and Msn4 from the cytoplasm to the nucleus, where they up-regulate the expression of *PNC1* (Fig. 1.7), thus reducing the nicotinamide pool, activating Sir2 and stabilising the rDNA (Gallo, Smith and Smith, 2004; Medvedik *et al.*, 2007). In keeping with this, inhibition of TOR pharmacologically with rapamycin or naturally through nitrogen starvation, both increased the association of Sir2 with the rDNA, caused deacetylation of rDNA histones, silencing of Pol II transcripts, and reduced homologous recombination as evidenced by decreased rDNA CNV and reduced formation of extrachromosomal rDNA circles (ERCs) (Ha and Huh, 2011).

1.7 Connections between replication and CNV

BIR is a well-characterised mechanism of DNA repair that has long been implicated in rDNA replication and CNV; therefore DNA replication may also influence CNV. DNA replication is a highly regulated process that involves numerous proteins as part of the replisome complex (Fig. 1.9). DNA replication is tightly controlled in time and space to maintain genome integrity (Symeonidou, Taraviras and Lygerou, 2012). In addition, cells have evolved

checkpoint pathways, which can arrest the cell cycle in response to changes in the external environment, such as genotoxic stresses (Siddiqui, On and Diffley, 2013). Checkpoint arrest of the cell cycle provides cells time to correct mutations and lesions, instead of proceeding with the segregation of chromosomes containing errors to daughter cells (Harper and Elledge, 2018). The replisome complex is composed of numerous specialised proteins and protein assemblies that function together to support DNA replication by the DNA polymerases (Leman and Noguchi, 2013). The replicative helicase mini-chromosome maintenance proteins (Mcm2-7) unwind the DNA ahead of the replication fork and is required throughout S-phase for DNA replication (Chong *et al.*, 2000; Labib, Tercero and Diffley, 2000; Pacek and Walter, 2004). Mcm2-7 interacts with the DNA replication initiation factor Cdc45 and the go-ichi-ni-san (GINS) complex to form the Cdc45-Mcm2-7-GINS (CMG) complex for full replicative helicase activity (Leman and Noguchi, 2013). The chromatin-associated protein Ctf4 in *Saccharomyces cerevisiae*, or the human orthologue And1, connects the CMG complex to the lagging strand DNA polymerase α (Pol α) (Gambus *et al.*, 2009; Bermudez *et al.*, 2010). Pol α synthesises the lagging strand primer (Pellegrini, 2012). DNA polymerase δ (Pol δ) then synthesises DNA on the lagging strand in a discontinuous manner, initiated from the lagging strand primer (Pellegrini, 2012). The S-phase checkpoint protein Mrc1, or the human orthologue Claspin, connects the Mcm2-7 helicase to the leading strand DNA polymerase ε (Pol ε) that synthesises the leading strand continuously (Lou *et al.*, 2008; Petermann, Helleday and Caldecott, 2008; Leman and Noguchi, 2013). Pol δ and Pol ε are tethered securely to the DNA by the proliferating cell nuclear antigen (PCNA), which is loaded by the replication factor C (RFC) complex (Cai *et al.*, 1998; Leman and Noguchi, 2013). RFC can recognise primer-template junctions, where it loads PCNA (Tsurimoto and Stillman, 1991; Podust *et al.*, 1995). Mrc1 also associates with the Fork Protection Complex (FPC), which comprises the topoisomerase I-interacting factor (Tof1) and the replication fork associated factor (Csm3) in *Saccharomyces cerevisiae*, or the human orthologues Timeless and Tipin, to stabilise the replication machinery at the replication fork (Katou *et al.*, 2003; Bando *et al.*, 2009). Replication protein A

(RPA) is also found as part of the replisome, where it stabilises the region of single-stranded DNA at the replication fork (Wold and Kelly, 1988; Alani *et al.*, 1992; Leman and Noguchi, 2013).

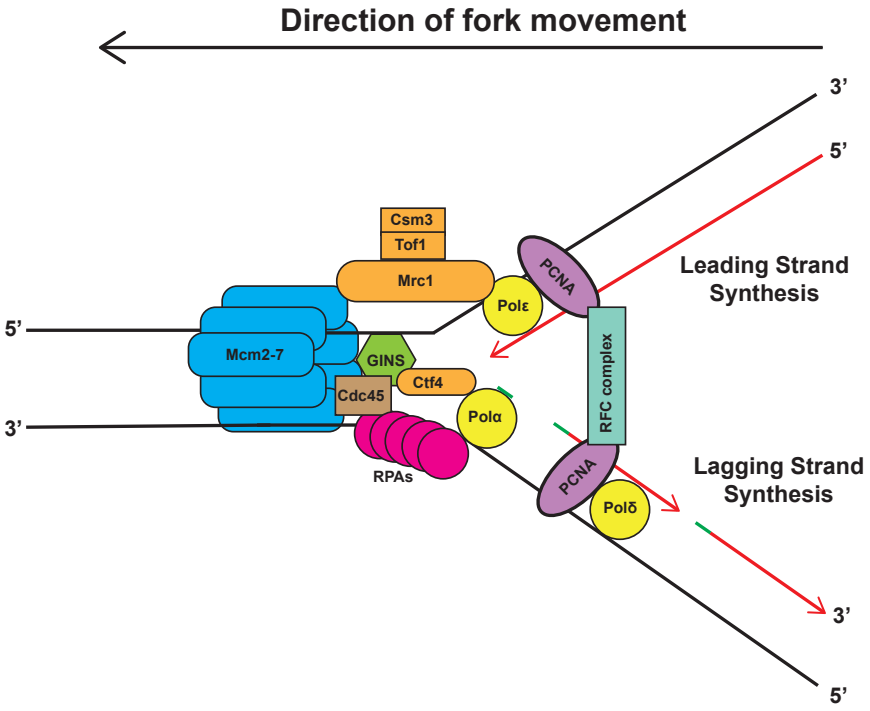


Figure 1.9: Schematic of the major components of the replisome complex in *Saccharomyces cerevisiae*. The replicative helicase mini-chromosome maintenance proteins (Mcm2-7) unwind the DNA ahead of the replication fork. Mcm2-7 interacts with the DNA replication initiation factor Cdc45 and the go-ichi-ni-san (GINS) complex to form the Cdc45-Mcm2-7-GINS (CMG) complex. The chromatin-associated protein Ctf4 connects the CMG complex to the lagging strand DNA polymerase α (Pol α) that synthesises the lagging strand primer. DNA polymerase δ (Pol δ) synthesises DNA on the lagging strand in a discontinuous manner, initiated from the lagging strand primer. The S-phase checkpoint protein Mrc1 connects the Mcm2-7 helicase to the leading strand DNA polymerase ε (Pol ε) that synthesises the leading strand continuously. Pol δ and Pol ε are tethered securely to the DNA by the proliferating cell nuclear antigen (PCNA), which is loaded by the replication factor C (RFC) complex. Mrc1 also associates with the Fork Protection Complex (FPC), which comprises the topoisomerase I-interacting factor (Tof1) and the replication fork associated factor (Csm3), to stabilise the replication machinery at the replication fork. Replication protein A (RPA) stabilises the region of single-stranded DNA at the replication fork.

1.7.1 Histone H3K56 acetyltransferase Rtt109

Recent work from our group identified H3K56 acetylation as a critical effector of TOR activity on rDNA copy number regulation (Jack, 2014). Rtt109 was first identified as a regulator of Ty1 transposition (Scholes *et al.*, 2001), and later shown to be important for the acetylation of H3K56, along with the histone chaperone Asf1, and independent of the putative cell cycle specific acetyltransferase Spt10 (Schneider *et al.*, 2006). Rtt109 was confirmed by *in vitro* assays to be the acetyltransferase responsible for acetylation of H3K56, catalysing the acetylation in an Asf1-stimulated manner (Fig. 1.7) (Driscoll, Hudson and Jackson, 2007). The human p300/CBP histone acetyltransferase has been proposed to be a distant homologue of the yeast Rtt109 (Bazan, 2008; Wang *et al.*, 2008).

Rtt109 is important for genome stability and resistance to DNA damaging agents (Driscoll, Hudson and Jackson, 2007). The H3K56 acetyltransferase activity of Rtt109 is also important for replisome integrity and the regulation of DNA replication (Han, Zhou, Horazdovsky, *et al.*, 2007). The Rtt109-Asf1 complex has a crucial role in maintaining expression homeostasis during DNA replication through its H3K56 acetylation activity (Voichek, Bar-Ziv and Barkai, 2016). H3K56ac histones become selectively incorporated on the newly synthesised DNA strand, reducing the transcription efficiency of the early S-phase replicated genes and thus maintaining expression homeostasis throughout the DNA replication process (Voichek, Bar-Ziv and Barkai, 2016). The Rtt109-Asf1 complex also appears to be important for a cell to recover from a DSB, which is believed to be via a chaperone-independent function of Asf1 in facilitating the dephosphorylation of Rad53 following DNA repair, thus signalling its completion (Tsabar *et al.*, 2016).

Tandem-affinity purification of Rtt109 also recovered another histone chaperone Vps75, which stabilises the acetyltransferase *in vivo* (Jessulat *et al.*, 2008). Loss of either Rtt109 or Vps75 resulted in reduced efficiency of DNA repair by NHEJ and hyper-sensitivity to DNA-damaging agents, suggesting an important role for Rtt109 in DNA damage repair (Jessulat *et al.*, 2008). However, no defect in HR DNA repair was observed in *rtt109* mutants,

whereas *VPS75* is still important for DSB repair using HR (Jessulat *et al.*, 2008). Rtt109 has been shown to also acetylate H3K9 *in vivo*, but only when in a complex with both histone chaperones Asf1 and Vps75 (Fillingham *et al.*, 2008), suggesting that Vps75 targets the Rtt109-Asf1 complex to H3K9, and away from the default H3K56 site. Gnc5 is the only other known histone acetyltransferase for H3K9 (Fillingham *et al.*, 2008).
At the rDNA, *rtt109* mutants exhibit a 3-fold hyper-amplification phenotype in the number of rDNA repeats (Ide, Saka and Kobayashi, 2013). The hyper-amplification phenotype of *rtt109* mutants at the rDNA was also observed in a low-copy strain, which in addition was shown to overcome the inhibitory effect of rapamycin on rDNA amplification (Jack, 2014). Despite deletion of *RTT109* relieving rapamycin-induced repression of rDNA amplifications, rapamycin treated *rtt109Δ* cells were still more heterogeneous in rDNA copy number than wild-type cells, potentially as a result of the continued effects of rapamycin on Sir2 (Jack, 2014). Therefore Rtt109 is involved in regulation of DNA replication, transcription, and DNA repair, all of which could be responsible for the stabilising role of Rtt109 on CNV at the rDNA.

1.7.2 S-phase checkpoint protein Mrc1

The important replisome component Mrc1 is an S-phase checkpoint protein that connects the Mcm2-7 helicase to the leading strand DNA Pol ε (Lou *et al.*, 2008). Mrc1 senses replication stress and through Rad53, delays entry into mitosis, whilst promoting DNA repair and completion of DNA replication (Fig. 1.10a) (Alcasabas *et al.*, 2001). The mammalian homologue of Mrc1 is claspin (Chini and Chen, 2004; Petermann, Helleday and Caldecott, 2008). Mrc1 initially binds to early replicating sequences and follows the replication fork along the chromatin (Osborn and Elledge, 2003). When the replication fork stalls, Mrc1 along with another S-phase checkpoint protein Tof1, interact directly with the replication machinery to form a stable pausing structure that serves to anchor subsequent DNA repair events (Katou *et al.*, 2003). The stalled fork recruits the genome integrity checkpoint protein Mec1, which phosphorylates Mrc1, thus activating the Rad53 checkpoint response (Osborn

and Elledge, 2003). Mrc1 interacts directly with Pol2, the catalytic subunit of DNA Pol II that participates in leading-strand synthesis during DNA replication and DNA repair, stabilising Pol2 at stalled forks and potentially coupling polymerisation to the unwinding of DNA at the replication fork (Lou *et al.*, 2008). The signal for the end of the S-phase checkpoint appears to be the degradation of Mrc1 by the origin-binding F-box protein Dia2 (Fong, Arumugam and Koepp, 2013).

Under osmostress Mrc1 is phosphorylated by the Hog1 mitogen-activated protein kinase involved in osmoregulation (Fig. 1.10b), but on different sites to those phosphorylated by Mec1 in response to replication stress (Duch *et al.*, 2013). Phosphorylation by Hog1 delays early and late origin firing by preventing binding of the DNA replication initiation factor Cdc45, as well as slowing down progression of the replication-complex (Duch *et al.*, 2013). Also given that Hog1 regulation of Mrc1 is independent of Rad53, it appears to be a specific stress-induced S-phase checkpoint response, distinct from the usual DNA damage checkpoint response (Duch *et al.*, 2013).

Similarly to the osmostress response, Mrc1 is also phosphorylated on the Hog1-specific phosphorylation sites (T169, S215, and S229) in response to heat and oxidative stresses, and show delayed S-phase progression caused by altered DNA replication (Duch *et al.*, 2018). The study proposed that the Mrc1-dependent delay in replication fork progression in response to environmental stress is to prevent replication-transcription conflicts. Replication-transcription conflicts in non-phosphorylatable *mrc1*^{3A} mutants could be mediated when combined with mutations in the Msn2 and Msn4 transcription factors (Duch *et al.*, 2018). Since repression of TOR signalling by rapamycin promotes the re-localisation of Msn2 and Msn4 from the cytoplasm to the nucleus, rapamycin may also promote stability at the rDNA through Mrc1 checkpoint activity (Gallo, Smith and Smith, 2004; Medvedik *et al.*, 2007). In addition to the checkpoint activation role of Mrc1 in response to replication and environmental stresses, Mrc1 is also important for telomere capping and protection from the exonuclease Exo1, involved in the initial strand resection step for HR-dependent DSB repair (Tsolou and Lydall, 2007). Therefore Mrc1 is important in the regulation of DNA repair, as well as DNA replication.

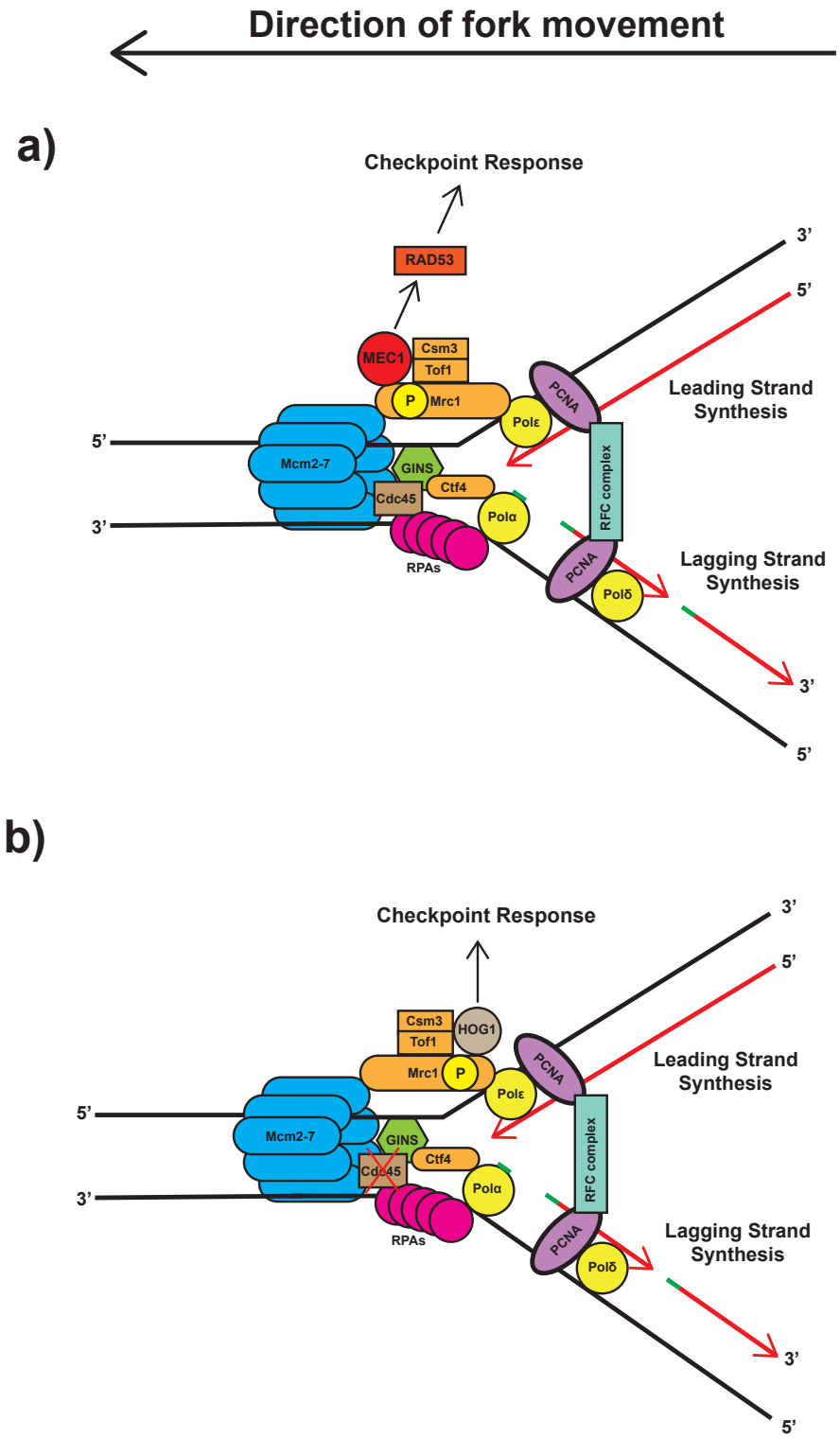


Figure 1.10: Checkpoint response at stalled replication forks. **a**) During replication stress, Mrc1 associates with Tof1 to stabilise the replication machinery at the replication fork. Together they recruit Mec1, which phosphorylates Mrc1, which in turn can then activate Rad53 and initiate the DNA damage checkpoint response. **b**) During osmostress, heat and oxidative stresses, Hog1 phosphorylates Mrc1 on different sites to Mec1 phosphorylation, which prevents Cdc45 binding and initiates the stress-induced S-phase checkpoint response.

1.7.3 Pol32

Another replisome component Pol32, encoded by the *POL32* gene also known as *REV5* (Lawrence, Krauss and Christensen, 1985), is a non-essential subunit of DNA Pol δ (Gerik *et al.*, 1998). The *Saccharomyces cerevisiae* Pol δ comprises of three subunits in a 1:1:1 stoichiometric ratio: Pol3 (125 kDa), Pol31 (55 kDa), and Pol32 (40 kDa) (Gerik *et al.*, 1998; Johansson, Majka and Burgers, 2001). In addition to these three Pol δ subunits, *Schizosaccharomyces pombe* Pol δ contains two more subunits, whilst the human Pol δ has one additional subunit (Zuo *et al.*, 1997; Liu *et al.*, 2000; Podust *et al.*, 2002).

Pol32 binds the proliferating cell nuclear antigen (PCNA) that functions as a sliding replication clamp and processivity factor for Pol δ (Fig. 1.9) (Gerik *et al.*, 1998; Paunesku *et al.*, 2001). Deletion of *POL32* causes a cold sensitivity growth phenotype and an increased sensitivity to DNA damaging agents (Gerik *et al.*, 1998). *pol32Δ* yeast are weak anti-mutators and are defective for damage-induced mutagenesis (Gerik *et al.*, 1998). Pol32 appears to be important in genome stability, since *pol32Δ* mutants have an increased frequency of deletion and duplication events at sequences flanked by short direct repeats (Huang *et al.*, 2002). This function of Pol32 is unlikely to involve homologous recombination as it is Rad52-independent (Huang *et al.*, 2002). The increased instability in *pol32Δ* strains, and therefore heightened level of DNA damage, is probably why *pol32Δ* mutations are lethal when combined with mutations in either of the other two Pol δ subunits *POL3* and *POL31*, or in the PCNA encoding gene *POL30* (Johansson, Garg and Burgers, 2004). Deletion of *POL32* significantly reduces the efficiency of BIR, but does not prevent all BIR events (Deem *et al.*, 2008). Furthermore, the study found that the BIR events associated with *pol32Δ* cells formed half-crossovers that arise from aberrant processing of BIR intermediates (Deem *et al.*, 2008). In *pol32Δ* cells, BIR events can still initiate strand invasion, but are defective for DNA synthesis, where the stalled intermediates are cleaved to produce half-crossovers (Smith, Lam and Symington, 2009).

Segmental duplications (SDs) are a major source of CNV polymorphisms between individuals (Redon *et al.*, 2006). Pol32 is required for all SDs, with duplications arising from replication-based mechanisms and not unequal crossovers (Payen *et al.*, 2008). BIR and MMIR could both produce SDs in Pol32-dependent mechanisms, only differing in the requirement of additional HR proteins, such as Rad52, Rad51, and Rad1 (Payen *et al.*, 2008). As such, Pol32 is important in the evolution of CNV events.

Outside of its function as part of Pol δ, when Pol32 combines with Pol31 and two other proteins Rev3 and Rev7, it forms the four subunit polymerase ζ (Pol ζ) complex (Makarova, Stodola and Burgers, 2012). Pol ζ has an important role in PCNA-dependent DNA translesion synthesis (TLS) and mutagenesis in all eukaryotes. Pol32 is indispensable for Pol ζ formation and also for PCNA binding, with either property of Pol32 potentially explaining the defective mutagenesis observed in *pol32Δ* cells (Makarova, Stodola and Burgers, 2012).

Pol32 has two important, but separable functions as part of the replisome complex. The first function of Pol32 is to ensure error-free DNA synthesis during replication fork restart, as an important subunit of the Pol δ protein complex. However, the second function of Pol32 is in error-prone TLS, as an indispensable subunit of the Pol ζ complex. Therefore Pol32 has both mutagenic and anti-mutagenic properties dependent upon the polymerase complex.

1.8 Common sites of gene amplification in yeast

The rDNA is not the only known CNV locus in the *Saccharomyces cerevisiae* genome, with CNV having been reported at many other loci during growth in a selective environment (Hansche, Beres and Lange, 1978; Brown, Todd and Rosenzweig, 1998; Dunham *et al.*, 2002; Libuda and Winston, 2006; Gresham *et al.*, 2008), but also in the absence of selection (Argueso *et al.*, 2008; Vernon, Lobachev and Petes, 2008; Hoang *et al.*, 2010; McCulley and Petes, 2010). CNV is believed to have played a major role in the evolution of eukaryotic genomes (Koszul *et al.*, 2004). CNV genes have been detected

following long-term evolution studies that typically cover 50 to 500 generations of selective growth (Gorter de Vries, Pronk and Daran, 2017), but have also been observed on a much faster scale, such as genomic *CUP1* amplifications in the presence of copper (Giusy Manuela Adamo *et al.*, 2012), or the production of extrachromosomal circular DNAs (eccDNAs) (Moller *et al.*, 2015).

1.8.1 Long-term evolution of CNV

Compared to haploid yeast, diploids and tetraploids undergo significantly faster adaptation, which provides evidence that copy number is important in environmental adaptation (Selmecki *et al.*, 2015). Duplication of a gene has been suggested to contribute to evolution in three ways (Taylor and Raes, 2004; Hahn, 2009). Firstly, duplications can cause gene dosage effects, such as increased gene expression that confers an adaptive advantage in the environment (Kondrashov *et al.*, 2002). Secondly, genetic redundancy caused by a duplication event may relieve one allele from purifying selection and allow it to develop a novel function, known as neofunctionalisation (Voordeckers *et al.*, 2012). Thirdly, the functions and regulation of the ancestral gene may become split across the duplicated genes, thereby developing specialised functions in each gene, known as subfunctionalisation (Hughes, 1994; Force *et al.*, 1999; Voordeckers *et al.*, 2012).
Long-term evolution studies have identified CNV events that arise spontaneously during prolonged growth in a selective environment, which confer an adaptive advantage in the environment (Hansche, Beres and Lange, 1978; Brown, Todd and Rosenzweig, 1998; Dunham *et al.*, 2002; Libuda and Winston, 2006; Gresham *et al.*, 2008). Duplications of the linked acid phosphatase genes *PHO3* and *PHO5* arise spontaneously in asexually reproducing populations of *Saccharomyces cerevisiae* during selection in phosphate-limited chemostat populations, over 100 generations (Hansche, Beres and Lange, 1978). Amplification of *PHO3* and *PHO5* genes occurred at a rate of 10^{-11} to 10^{-12} duplications per mitosis (Hansche, Beres and Lange, 1978). Strains with a duplication of *PHO3* and *PHO5* can produce 20- to 40-

fold more enzyme than a strain without the duplication, during growth on phosphate-limited medium, thereby conferring an adaptive advantage (Field and Schekman, 1980). Similarly, spontaneous amplification of the structural genes *ADH2* and *ADH4*, encoding yeast alcohol dehydrogenases, caused resistance to antimycin A, when grown at 15°C, at a rate of 10^{-10} amplifications per cell per generation (Dorsey *et al.*, 1992).

In limited glucose environments, multiple tandem duplications have been detected in the high-affinity hexose transport loci *HXT6* and *HXT7*, over a 450 generation evolution study (Brown, Todd and Rosenzweig, 1998). Most aneuploidy regions were the result of translocations, with the genome rearrangements bounded by transposon-related sequences at the breakpoints (Dunham *et al.*, 2002). CNV mutants that have adapted to nutrient-limited environments display a 5-50% increase in fitness over their ancestor (Gresham *et al.*, 2008). However, Wenger *et al*. found that the adaptation benefit was an environment-specific trade-off, as yeast adapted to a carbon-limited environment were outcompeted by their ancestor in carbon-rich environments (Wenger *et al.*, 2011).

Yeast of the *Saccharomyces* clade were found to amplify the *SUL1* or *SUL2* sulphur transporter following approximately 200 to 500 generations of continual culture in a sulphate-limited environment (Sanchez *et al.*, 2017). However, the choice of which sulphur transporter gene was amplified showed differential sub-fractionalisation across the *Saccharomyces* species, dependent upon recent changes in noncoding sequence (Sanchez *et al.*, 2017). This provides evidence for the genome being optimised to be more prone to CNV at certain loci in certain environments based upon the genomic sequence.

CNV can also rescue lost phenotypes caused by other genetic mutations (Koszul *et al.*, 2004). Yeast in which the only reported lactate transporter, *JEN1*, is deleted are unable to grow on media containing lactose (de Kok *et al.*, 2012). Evolution of *jen1Δ* yeast to grow on lactose in serial batch culture was achieved through a spontaneous amplification of the *AD2Y* gene encoding an acetate transporter (de Kok *et al.*, 2012). Also in uracil auxotrophic cells, with a mutant *ura2-15-30-72* allele, duplication of a segment

of the mutant allele that corresponds to the ATCase coding domain recovered Ura$^+$ revertants (Schacherer *et al.*, 2005, 2007). Similarly, deletion of *RPL20A*, encoding the ribosomal 60S subunit protein L20A, causes a slow-growth phenotype that can be recovered by duplication of its homologous *RPL20B* gene (Koszul *et al.*, 2004; Payen *et al.*, 2008). There is also evidence of gene dosage compensation by formation of a circular chromosome (Libuda and Winston, 2006). Histones H2A and H2B in *Saccharomyces cerevisiae* are encoded by two gene pairs, *HTA1-HTB1* and *HTA2-HTB2*, and when *HTA1-HTB1* is deleted, *HTA2-HTB2* compensates via amplification as a new small circular chromosome (Libuda and Winston, 2006). Similarly the *GAP1* gene, encoding the general amino acid permease, is frequently amplified as a self-propagating extrachromosomal DNA circle (eccDNA), in combination with deletion of the chromosomal *GAP1* gene, during an evolution experiment in nitrogen-limited chemostat cultures over at least 250 generations (Gresham *et al.*, 2010). Interestingly, the *GAP1* eccDNA is rapidly lost after just a few generations in rich media, suggesting that retention of *GAP1* eccDNAs confers a strong selective disadvantage in a nitrogen-rich environment (Gresham *et al.*, 2010).

Experimental evolution studies have also reported gene loss when not maintained under selection (Naseeb *et al.*, 2017). The *IFA38* gene encodes an elongase enzyme required for very-long chain fatty acid synthesis, and is important for maintaining membrane fluidity and for resistance to ethanol and other stressors (Tehlivets, Scheuringer and Kohlwein, 2007; Ding *et al.*, 2009; Naseeb *et al.*, 2017). Artificial duplication of *IFA38* conferred a fitness advantage in fermentable media, both immediately after duplication, and in an evolved population after 500 generations, through an increased level of gene expression (Naseeb *et al.*, 2017). However, strains evolved in non-fermentable glycerol during respiratory growth, where *IFA38* is not being actively selected for, the duplicate gene becomes selectively lost in 80% of cases after as little as 25 generations (Naseeb *et al.*, 2017). Interestingly, this was only the case when the duplicate gene was inserted at a non-tandem position to the original *IFA38* gene, whereas all the strains carrying a tandem copy of *IFA38* retained both repeats (Naseeb *et al.*, 2017). Naseeb *et al*. postulated that this discrepancy between tandem and non-tandem repeats is

due to some duplicate pairs not being functionally equivalent, or biased towards duplicate loss based on genome location, as has been previously reported (Scannell *et al.*, 2006; Makino and McLysaght, 2012; Naseeb *et al.*, 2017).

CNV events have also been detected in the absence of selection (Argueso *et al.*, 2008; Vernon, Lobachev and Petes, 2008; Hoang *et al.*, 2010; McCulley and Petes, 2010). In *Saccharomyces cerevisiae*, the *MEC1* and *TEL1* genes, homologues of the mammalian *ATR* and *ATM* genes respectively, encode proteins that are involved in the detection of DNA damage and in the regulation of telomere length (Vernon, Lobachev and Petes, 2008; McCulley and Petes, 2010). Haploid and diploid *tel1 mec1* cells are genetically unstable, producing a high rate of chromosomal aberrations, including duplications, deletions, and translocations, following 200 generations of culturing in a non-selective environment (Vernon, Lobachev and Petes, 2008; McCulley and Petes, 2010).

Long-term evolution studies have shown that spontaneous CNV events arise during prolonged growth in challenging environments, where the novel CNV allele confers an adaptive advantage in the environment. There is evidence that the novel CNV becomes dominant in the population as long as the selection pressure for the allele remains, but is rapidly lost once the selection pressure is removed. Spontaneous CNV has also been detected in the absence of selection in genetically unstable strains. Finally, differential evolution of the *SUL1* and *SUL2* amplification in a sulphate-limited environment across the *Saccharomyces* clade, as determined by changes in the non-coding sequence, provides strong evidence for certain genomic loci being more prone to CNV in certain environments based upon the genomic sequence. As such, CNV is a major contributor to the formation of an evolved population, which has specifically adapted to the surrounding environment.

1.8.2 Copper resistance *CUP1* gene

CUP1 is a well-characterised gene that can rapidly change copy number in an environment-specific context (Giusy Manuela Adamo *et al.*, 2012). The *CUP1*

gene was identified as the gene responsible for mediating resistance to the toxicity of copper (Fogel and Welch, 1982). In trace amounts, copper is an essential macronutrient for budding yeast, but at high concentrations it can become toxic to cells (Rutherford and Bird, 2004). $CuSO_4$ concentrations of 0.5mM and above are inhibitory to the growth of most *Saccharomyces cerevisiae* (Butt, E. Sternberg, *et al.*, 1984). Yeast require just a single copy of *CUP1* to survive under normal growth conditions, but in high copper concentrations, strains with higher *CUP1* copy numbers and enhanced resistance are selected for (Fig. 1.11) (Fogel and Welch, 1982). Isolated copper resistant strains with high *CUP1* copy numbers have a 50-fold enhanced transcription rate upon exposure to 0.5mM $CuSO_4$ (Butt, E. Sternberg, *et al.*, 1984).

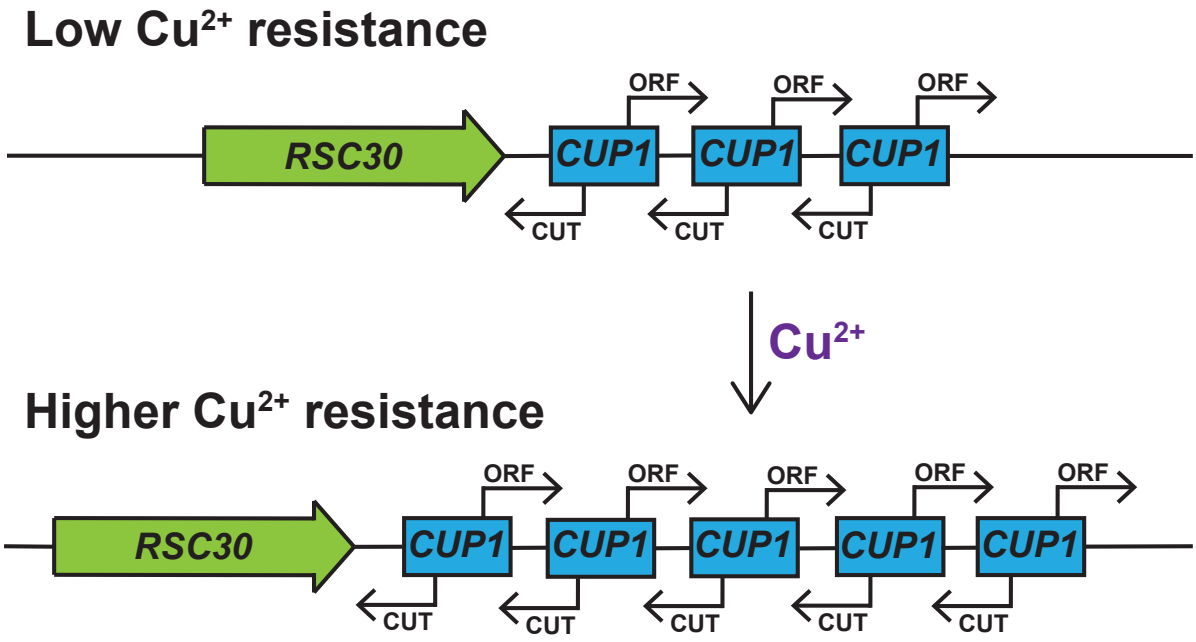


Figure 1.11: Copper resistance through *CUP1* copy number amplification. Yeast grown in the presence of copper select for strains with higher copy number of the copper-resistance gene *CUP1* (blue rectangle). The direction of the *CUP1* ORF and CUT are shown as black arrows. *RSC30* (green arrow) is a chromatin-remodelling gene of unknown function that flanks the *CUP1* repeat array.

Cellular uptake of copper ions across the plasma membrane is typically more labile as lower valence species (Yamaguchi-Iwai *et al.*, 1997). In *Saccharomyces cerevisiae* the *FRE1* gene product, a ferric and cupric reductase, is the primary reductase responsible for reducing Cu^{2+} ions to the more membrane labile Cu^{+} ions (Dancis *et al.*, 1992; Hassett and Kosman, 1995; Yamaguchi-Iwai *et al.*, 1997). A multimeric plasma membrane protein CTR1, encoded by the *CTR1* gene, is believed to then bind Cu^{+} ions in a pocket for subsequent internalisation and translocation of the Cu^{+} ion into the cytosol in a process that has not yet been fully characterised (Dancis, Haile, *et al.*, 1994; Dancis, Yuan, *et al.*, 1994; Knight *et al.*, 1996; Yamaguchi-Iwai *et al.*, 1997). Both *FRE1* and *CTR1* transcription is dependent upon copper, therefore reduction of extracellular Cu^{2+} and cellular uptake of Cu^{+} is regulated homeostatically (Dancis, Yuan, *et al.*, 1994; Hassett and Kosman, 1995; Yamaguchi-Iwai *et al.*, 1997).

CUP1 encodes a metallothionein, a heavy metal ion-binding protein often involved in detoxification (Winge *et al.*, 1985; Ecker *et al.*, 1986). Cup1 has been shown to bind several metal ions *in vitro* (Winge *et al.*, 1985), but only copper and cadmium ions are also bound *in vivo* (Ecker *et al.*, 1986; Jeyaprakash, Welch and Fogel, 1991). Cup1 can bind up to 8 Cu^{+} ions through 12 cysteine residues per molecule, thus removing excess Cu^{+} ions from causing damage inside the cell (Winge *et al.*, 1985). The *CUP1* regulatory sequence has a specific transcriptional requirement for copper (Butt, E. J. Sternberg, *et al.*, 1984) and increased resistance through metallothionein production is a combined effect of *CUP1* gene amplification and induced transcription (Karin *et al.*, 1984). *CUP1* complementation on a high copy number plasmid restored copper resistance to a sensitive *cup1Δ* strain, but interestingly it had no effect on cadmium resistance (Jeyaprakash, Welch and Fogel, 1991). This shows that the *CUP1* confers resistance to copper and cadmium in a separable way, and that resistance is not just attributable to an increased rate of transcription, but also requires a genomic amplification of the *CUP1* gene.

Positive transcriptional regulators of the *CUP1* gene, other than high concentrations of copper, include *cis*-acting upstream activation sequences (Butt, E. J. Sternberg, *et al.*, 1984; Thiele and Hamer, 1986) and *trans*-acting

transcriptional regulatory factors (Hamer, Thiele and Lemontt, 1985; Gorman *et al.*, 1986). Through genetic approaches to identify the *trans*-acting factors, the *ACE1* gene was found to activate *CUP1* expression upon exogenous copper exposure and whose loss makes cells copper sensitive (Thiele, 1988). The *ACE1* gene was found to map to the same locus as another gene *CUP2* (Mortimer and Schild, 1985; Thiele, 1988), and was later confirmed to be the same gene (Buchman *et al.*, 1989). Cup2 is a copper-regulated DNA-binding protein that acts as the primary sensor of intracellular Cu^+ ions, by refolding from its metal-free apoprotein conformation upon binding of Cu^+ ions to its cysteine-rich DNA-binding domain (Buchman *et al.*, 1989). Cup2 either directly binds the upstream control region of *CUP1*, or indirectly regulates the synthesis or activity of another protein that binds there (Welch *et al.*, 1989). In *cup2* mutants, with or without the addition of copper, only very low levels of *CUP1* mRNA can be produced and the *CUP1* promoter-binding factor is lost, confirming that *CUP2* functions as a Cu^+-regulated transcriptional activator of *CUP1* (Buchman *et al.*, 1989; Welch *et al.*, 1989). As a result, *cup2Δ* yeast display a decrease in copper resistance of two orders of magnitude (Welch *et al.*, 1989).

The *CUP2* gene also has a paralogue *HAA1* in *Saccharomyces cerevisiae* that arose from the whole genome duplication (Byrne and Wolfe, 2005). Just like *CUP2*, *HAA1* is also characterised as a transcriptional activator, supporting the idea of a regulatory role for *CUP2*, but unlike *CUP2*, *HAA1* is not involved in copper adaptation, but instead it plays an essential role in weak acid sensing and adaptation (Keller *et al.*, 2001; Fernandes *et al.*, 2005).

Besides *CUP2*, the only other known *trans*-regulator of *CUP1* is the trimeric heat shock transcription factor Hsf1, which activates *CUP1* expression in response to heat shock, glucose starvation and oxidative stresses through binding to a heat shock element located within the *CUP1* promoter upstream regulatory region (Tamai *et al.*, 1994; Mager and De Kruijff, 1995). Therefore it is possible that Cup1 may have additional functions in the general stress response, beyond just copper resistance. In particular, Cup1 appears to exhibit some copper-dependent antioxidant activity (Tamai *et al.*, 1993; Liu and Thiele, 1996). Deletion of the superoxide dismutase gene *SOD1* in yeast,

which is required for protection from oxygen-related toxicity, is lethal to cells grown on the respiratory carbon source lactate (Tamai *et al.*, 1993). However, *sod1*Δ yeast supplemented with copper remain viable (Tamai *et al.*, 1993). In addition, *CUP1* mRNA levels become dramatically up-regulated in conditions of oxidative stress, further suggesting a potential antioxidant role for *CUP1*, as a substitute for superoxide dismutase, in cellular defence against oxidative stress (Tamai *et al.*, 1993). This antioxidant response, through up-regulated *CUP1* transcription, was found to be strongly activated by the superoxide anion generator menadione, and to require both the *CUP1* promoter heat shock element and Hsf1 (Liu and Thiele, 1996). Therefore *CUP1* appears to have two distinct transcriptional activation mechanisms in place, which have been perfected to respond either specifically to increases in copper toxicity levels or more generally as part of a cellular stress response.
In addition to the multi-copy *CUP1* locus, *Saccharomyces cerevisiae* have a second copper-binding metallothionein that is transcribed from the single copy *CRS5* gene (Culotta, Howard and Liu, 1994). It was discovered through a genomic screen for genes conferring high copper resistance in a *cup1Δ* background (Culotta, Howard and Liu, 1994). Crs5 is another cysteine-rich metallothionein that is regulated by copper, oxidative stress and *CUP2*, but unlike *CUP1*, *CRS5* is specific to just copper resistance, remaining sensitive to cadmium toxicity and sharing very little homology with *CUP1* (Culotta, Howard and Liu, 1994). *CUP1* plays a much more dominant role in copper detoxification than *CRS5*, which is at least in part due to its stronger promoter (Jensen *et al.*, 1996). A promoter exchange experiment found that copper resistance via *CRS5* increased when driven by the strong *CUP1* promoter, and conversely the *CUP1*-driven copper resistance was reduced when replaced by the weaker *CRS5* promoter (Jensen *et al.*, 1996).

1.8.3 Formaldehyde resistance *SFA1* gene

The *SFA1* gene is responsible for a cell's resistance to formaldehyde (Wehner and Brendel, 1993; Wehner, Rao and Brendel, 1993). Formaldehyde is a weak mutagen and recombinagen to wild-type budding yeast, with

concentrations of formaldehyde of 1mM or higher proving to be cytostatic or cytotoxic to haploid wild-type cells (Wehner and Brendel, 1993). First annotated in 1989 (Mortimer *et al.*, 1989), the *SFA1* gene encodes a glutathione-dependent formaldehyde dehydrogenase, which oxidises formaldehyde to S-formylglutathione for its detoxification (Wehner, Rao and Brendel, 1993; Grey, Schmidt and Brendel, 1996). The role of *SFA1* in formaldehyde resistance was first established through its hyper-resistance phenotype when expressed in cells on a high copy number plasmid (Fig. 1.12) (Wehner, Rao and Brendel, 1993). Hyper-resistant cells maintain the high-copy *SFA1* plasmid as long as they are being cultured in 5mM formaldehyde, a lethal concentration to untransformed yeast, showing the importance of maintaining high levels of Sfa1 for formaldehyde detoxification (Wehner and Brendel, 1993). The genomic *SFA1* is a single copy, non-essential gene, but it does still display hyper-sensitivity to formaldehyde in null mutants (Gompel-Klein, Mack and Brendel, 1989).

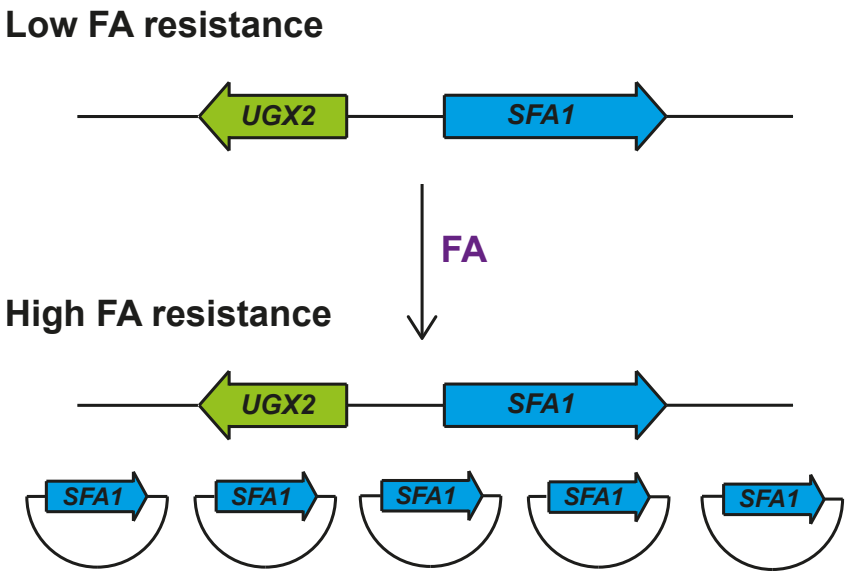


Figure 1.12: Formaldehyde resistance through *SFA1* copy number amplification. Growth of yeast containing the *SFA1* gene (blue arrow) expressed on a high copy number plasmid showed much greater resistance to formaldehyde (FA) than its untransformed counterpart with a single genomic *SFA1* repeat. *SFA1* shares a bi-directional promoter with its upstream divergent *UGX2* gene (green arrow) of unknown function.

SFA1 shares an 868bp divergent promoter with the *UGX2* gene of unknown function (Wehner, Rao and Brendel, 1993). Promoter deletion studies found that the *SFA1* upstream sequence experiences negative interference controlling the expression of *SFA1*, but the region between positions -145 and -172 is totally or partially responsible for control of *SFA1* inducibility by formaldehyde (Wehner, Rao and Brendel, 1993).

SFA1 is a member of the class III alcohol dehydrogenase family of enzymes that show bi-functional activity as glutathione-dependent formaldehyde dehydrogenases and alcohol dehydrogenases, but within the cell it appears that the general role of Sfa1 is in formaldehyde detoxification and not the turnover of long-chain alcohols (Fernandez *et al.*, 1995). Sfa1 has been shown to localise to both the cytoplasm (Huh *et al.*, 2003) and the mitochondria (Sickmann *et al.*, 2003).

The *SFA1* gene is expressed in response to chemicals such as formaldehyde, ethanol and methyl methanesulfonate (Wehner, Rao and Brendel, 1993), but it is also under the negative regulation of Sko1, a basic leucine zipper transcription factor of the ATF/CREB family (Nehlin, Carlberg and Ronne, 1992), which in turn is regulated by the mitogen activated protein kinase Hog1 involved in osmoregulation (Rep *et al.*, 2001). *SFA1* is induced in *sko1Δ* yeast, where its on-going repression is lifted, but also in response to osmotic stress through the Hog1 kinase acting on Sko1 and alleviating its repression of *SFA1* (Rep *et al.*, 2001). In addition, *SFA1* becomes induced under oxidative stress (DeRisi, Iyer and Brown, 1997; Rep *et al.*, 2001). This response in dependent upon another basic leucine zipper transcription factor Yap1 (DeRisi, Iyer and Brown, 1997; Rep *et al.*, 2001). Therefore it has been suggested that the osmoregulation and oxidative stress responses are independent processes, with Sko1 influencing Yap1 promoter access or activity in some way during the oxidative stress response (Rep *et al.*, 2001). It also means that the *SFA1* gene has two distinct transcriptional activation profiles: the first as a specific response to cellular formaldehyde levels, and the second as a more general induction response to various stresses.

SFA1 is the cell's primary route for formaldehyde detoxification, but in its absence another alcohol dehydrogenase Adh1 can still produce a hyper-resistance phenotype, albeit much less pronounced, most likely via a

reductive step that produces methanol (Grey, Schmidt and Brendel, 1996). Yeast also have an essential esterase, encoded by the *YJL068C* gene, which can also function as an S-formylglutathione hydrolase that may potentially also be involved in the detoxification of formaldehyde into S-formylglutathione (Degrassi *et al.*, 1999; Vandenbol and Portetelle, 1999).
SFA1 has also been used in combination with *CUP1* for CNV assays (Narayanan *et al.*, 2006; Zhang *et al.*, 2013). The *CUP1 SFA1* system allows for detection of cells with gross chromosomal rearrangements (GCRs) with amplified *CUP1* and *SFA1* genes by selecting for dual copper and formaldehyde resistance (Narayanan *et al.*, 2006). The *CUP1 SFA1* reporter system used by Narayanan *et al*. in haploid cells enabled the detection of 5- to 13-fold extrachromosomal amplifications of *CUP1* and *SFA1*, but also observed intrachromosomal amplifications, which were highly unstable and often produced secondary rearrangements upon propagation (Narayanan *et al.*, 2006). Zhang *et al*. optimised the assay for selection of low-order amplifications, using the gene dosage-dependent phenotypes of *CUP1* and *SFA1* to now separate cells that have amplified from one to two copies in haploids, or from two to three copies in diploid cells (Zhang *et al.*, 2013). Although genomic CNV has not been observed at *SFA1*, the experimental evidence for *SFA1* being a high-copy enhancer of formaldehyde resistance, and its involvement in CNV assays, makes *SFA1* a likely candidate for endogenous CNV events.

1.8.4 Extrachromosomal DNA circles

So far, we have only discussed CNV arising from changes in the number of repeats of chromosomal sequences. However, cells also have an alternative and far less understood form of undergoing CNV that involves extrachromosomal DNA circles (eccDNAs) (Møller *et al.*, 2015). eccDNAs are circular DNA species, believed to arise from genome excision through various recombination mechanisms (Fig. 1.13). In *Saccharomyces cerevisiae*, they range in size from 1kb up to 38kb and cover 23% of the genome, which accounts for thousands of genes (Møller *et al.*, 2015). eccDNAs are of

medical interest, as they have been shown to accumulate in cancer cells, where they are referred to as double minutes (Cox, Yuncken and Spriggs, 1965; Radloff, Bauer and Vinograd, 1967). These double minutes contribute to the intracellular genetic heterogeneity of tumours, with particular importance placed on their amplification of oncogenes and drug resistance genes (Paulsen *et al.*, 2018).

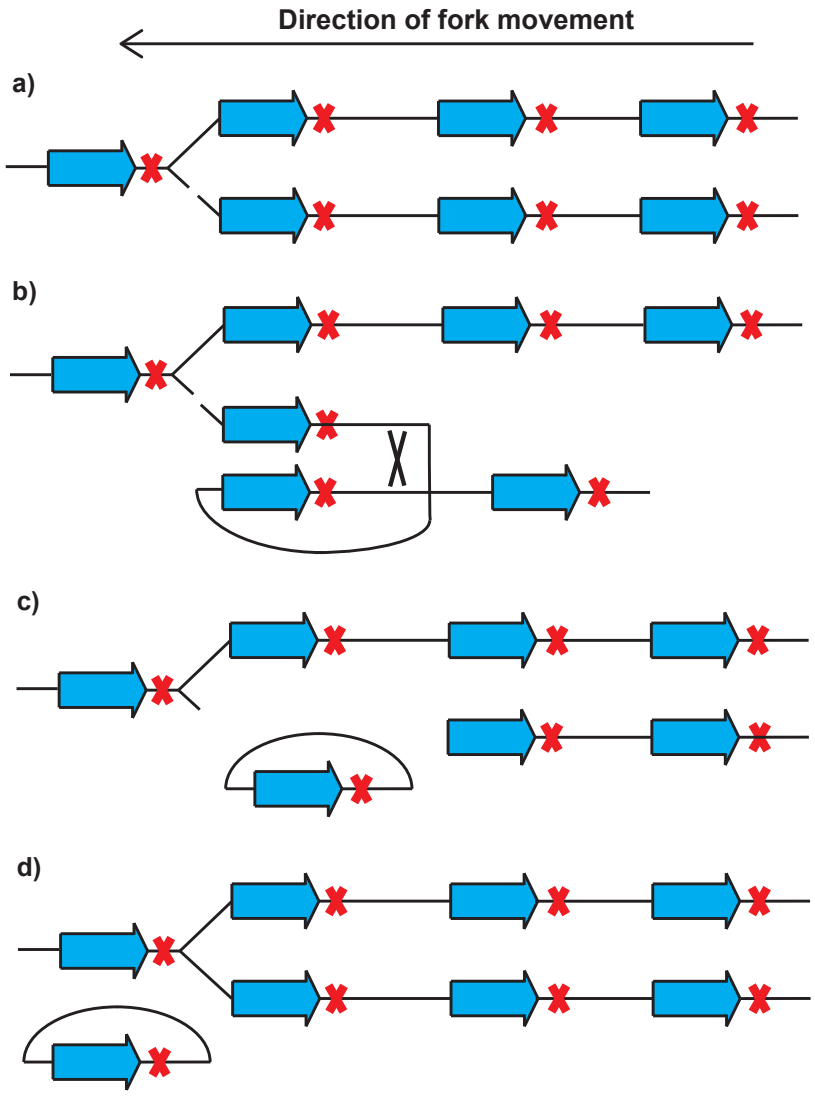


Figure 1.13: Proposed recombination model for eccDNA formation. **a**) A collapsed replication fork at a replication fork stalling site (red cross) induces a DSB to begin DNA repair. **b**) During the DNA repair process, a genomic repeat (blue arrow) loops out and aligns with a homologous repeat on the same DNA strand and **c**) becomes excised from the genome to form an eccDNA. **d**) Genomic repair proceeds as normal, re-synthesising the lost repeat using the homologous chromatid.

The largest producers of eccDNAs are the rDNA repeats (where they are known as ERCs), telomeric repeats, transposon remnants, and tandemly repeated genes like *CUP1*, *HXT6/7*, and *ENA1* (Sinclair and Guarente, 1997; Møller *et al.*, 2015). Many recombination models have been proposed for eccDNA formation, with the simplest arising from intrachromosomal ectopic homologous recombination between long terminal repeats (LTRs) that flank the genomic sequence (Gresham *et al.*, 2010). The rate at which this happens also appears to be controlled by the environment, potentially implicating it in environmental adaptation (Gresham *et al.*, 2010). Other known mechanisms of eccDNA formation that require a DSB include incorrect NHEJ (van Loon, Miller and Murnane, 1994; Storlazzi *et al.*, 2010), and MMEJ (Vogt *et al.*, 2004; Shibata *et al.*, 2012), but DSB-independent mechanisms also exist such as replication errors near short inverted repeats (Brewer *et al.*, 2011), or by small fragment-driven DNA amplification (SFDA) (Mukherjee and Storici, 2012).

eccDNAs are actively accumulated in mother cells, as part of an asymmetric division that potentially works by selective attachment of eccDNAs to structures retained in the mother, analogous to the association of nucleolar fragments with the nuclear periphery of aging cells (Sinclair and Guarente, 1997; Sinclair, Mills and Guarente, 1997). eccDNAs are only passed to daughter cells during division of very old mother cells, where there is most likely a breakdown in the asymmetry of inheritance (Sinclair and Guarente, 1997). ERCs have been hypothesised to contribute to aging by sequestration of the DNA replication machinery, thereby triggering permanent cell cycle arrest (Sinclair and Guarente, 1997). As such, it is believed that mother cells selectively retain the harmful eccDNAs as a protective mechanism for daughter cells.

Accumulation of ERCs has been implicated as a cause and not simply a consequence of aging. In *fob1Δ* cells, in which ERC formation is slowed, cells benefit from a 30-60% extension in lifespan (Defossez *et al.*, 1999). However ERCs can still accumulate to high levels in aged yeast, just at a reduced rate compared to wild-type cells, showing that ERC formation is not strictly dependent of Fob1 (Lindstrom *et al.*, 2011). As Fob1 is important generally for all recombination events at the rDNA, the lifespan extension phenotype in

fob1Δ cells may be the result of general reduced rDNA instability and not decreased ERC accumulation (Ganley *et al.*, 2018).

It is believed that most eccDNAs are replicated with each cell cycle, with 80% of circles containing a consensus sequence for an autonomously replicating sequence (ARS) element, enabling them to replicate independently of the chromosomes (Møller *et al.*, 2015). Therefore eccDNAs exponentially accumulate in mother cells with age. eccDNA accumulation causes massive variation in CNV within the genomes of any given population at any time, which potentially has many roles in adaptation. Aged yeast have even been shown to display a fitness advantage over their young counterparts when grown on non-optimal carbon sources (Frenk *et al.*, 2017). This enhanced fitness in aged yeast could be because of their eccDNA accumulation, increasing population heterogeneity and thus providing an adaptive advantage. Therefore it may be that under normal growth conditions, eccDNAs are detrimental to cells, but if the environment switches, in particular to a challenging environment, eccDNA retention provides a greater genomic diversity and therefore improved adaptive ability. However, current evidence for eccDNAs being transcriptionally active is minimal. The Tyler group recently observed a 8-fold and 30-fold reduction in transcription of the *NTS1* and *NTS2* rDNA ncRNAs in aged *fob1Δ* yeast, from which they concluded that ERCs can be transcribed by RNA Pol II and the reduction in transcription of the rDNA ncRNAs was due to the decreased accumulation of ERCs in *fob1Δ* cells (Pal *et al.*, 2018).

Importantly for eccDNAs, they can arise from genomic regions containing short or no repetitive sequence, as well as from multi-copy loci (Møller *et al.*, 2015). This means that excision and then re-integration of an eccDNA into the genome at its original genomic locus is one hypothesised mechanism for gene duplication events from single copy (Møller *et al.*, 2015).

1.9 General model for CNV

Combining all the current literature on CNVs, replication fork stalling, transcriptional interference, models of DNA repair, and in particular the

specialised rDNA copy number amplification model (Jack *et al.*, 2015), we propose a model for stimulated CNV at other loci in the yeast genome (Fig. 1.14). Firstly, we predict the importance of replication fork stalling at a site upstream of the candidate CNV gene, which can be identified by the DNA damage marker γH2A. Secondly, we propose that candidate CNV genes are primarily repressed under normal growth conditions and that they become strongly induced in response to a specific challenging environment. Finally, based on the importance of the Pol II transcribed ribosomal ncRNAs in regulating CNV at the rDNA, we predict that other candidate CNV genes would also use bi-directional promoters.

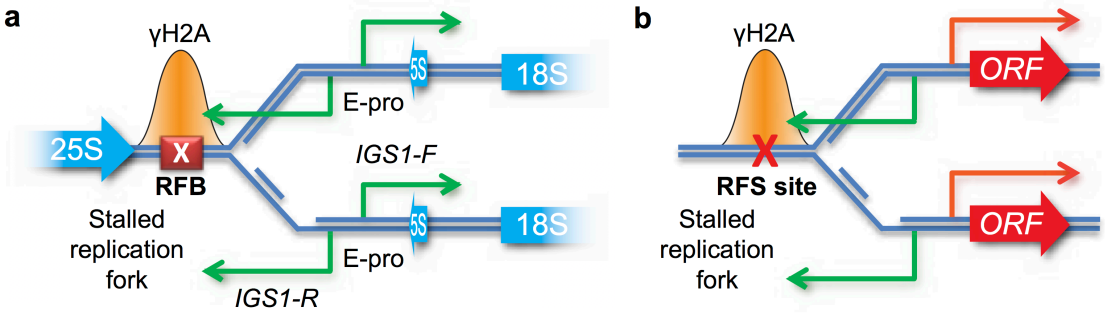


Figure 1.14: Predicted general model for stimulated CNV. Replication fork stalling model for the **a**) known specialised rDNA locus, and **b**) predicted general genomic locus. The replication machinery is predicted to stall at a replication fork stalling site (RFS), just upstream of our general CNV gene (red arrow), which is marked by the early DNA damage response marker γH2A. The CNV gene is transcribed from a bi-directional promoter with one direction transcribing into the RFS.

Our model predicts that the replication machinery pauses at the fork stalling sites of CNV genes, and will simply restart if not interfered with. However, in specific challenging environments, we propose that the appropriate CNV gene(s) would become highly expressed, and that transcription of either its sense or its antisense transcript would interfere with the replication machinery at the stalled fork, causing the fork to collapse. Next, a double strand break must be introduced at the site of the collapsed fork, before any DNA repair

process can begin. DNA DSB repair would then proceed using the intact sister chromatid as a template for BIR. If the sequence found for homologous repair is the matching repeat, there will not be any copy number change. However, if repair occurs off a mismatched repeat, a CNV event will result. For any possibility of mismatching the repeat during DSB repair and thus CNV, the genome must contain a minimum of two parental repeats.

2. Materials and Methods

2.1 Yeast culture

2.1.1 Yeast strains

The *Saccharomyces cerevisiae* strains used in this study are listed in Table 2.1. All strains used are derivatives of S288C (Brachmann *et al.*, 1998). Yeast deletion strains were created by standard methods and verified by PCR. Plasmids used in this study are listed in Table 2.2 and were verified by restriction digest and/or sequencing. Oligonucleotide pairs used in this study are listed in Table 2.3.

Table 2.1: Yeast strains used in this study.

Strain ID	Strain Name	Genotype	Reference
YRH21	BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(Brachmann *et al.*, 1998)
YJH108	*rrp6Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rrp6Δ::NatMX6*	(Houseley and Tollervey, 2006)
YRH79	P_{GAL1}-*HA*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ade2::MET25 cup1::17x[P_{GAL1}-3HA]-ADE2 pRS316-CUP1* Note: All chromosomal *CUP1* repeats were replaced with a *3HA* ORF under the control of the *GAL1* promoter and amplified to 17 copies upon transformation. A single *ADE2* marker is present at the telomere proximal end.	(Hull *et al.*, 2017)
YRH34	MEP wild-type diploid	*ade2::hisG/ade2::hisG his3/his3 leu2/leu2 lys2/+ met15D::ADE2/+ ura3D0/ura3Δo trp1D63/trp1D63 hoD::SCW11pr-Cre-EBD78-NatMX/hoD::SCW11pr-Cre-EBD78-NatMX loxP-UBC9-loxP-LEU2/lOXp-UBC9-loxP-LEU2 loxP-CDC20-Intron-loxP-HPHMX/loxP-CDC20-Intron-loxP-HPHMX*	(Lindstrom and Gottschling, 2009)
YRH35	MEP P_{GAL}-*HA* heterozygote	*ade2::hisG/ade2::MET15 met15::ADE2/met15Δ0 trp1Δ63 his3 leu2 ura3Δ0 hoD::SCW11pr-Cre-EBD78-KanMX/hoD::SCW11pr-Cre-EBD78-NatMX loxP-UBC9-loxP-LEU2 loxP-CDC20-Intron-loxP-HPHMX cup1::17xP_{GAL1}-HA-ADE2/CUP1*	(Hull *et al.*, 2017)
YRH77	*rtt109Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rtt109Δ::KanMX4*	(Giaever *et al.*, 2002)
YCJ3	*hst3Δ hst4Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 hst3Δ::KanMX6 hst4Δ::LEU2*	(Jack *et al.*, 2015)
YJH87	*sir2Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sir2Δ::HpHMX6*	(Houseley *et al.*, 2007)
YJH81	*fob1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 fob1::HpHMX6*	(Jack *et al.*, 2015)
Deletion Collection	*csm1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 csm1::KanMX4*	(Giaever *et al.*, 2002)
YJH247	*asf1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 asf1::HIS3*	(Houseley and Tollervey, 2011)
YCJ225	*ctf4Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ctf4::LEU2*	(Jack *et al.*, 2015)

YCJ221	*mms22Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mms22::LEU2*	(Jack *et al.*, 2015)
YCJ272	*mms1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mms1:KanMX4*	(Giaever *et al.*, 2002)
YJH366	*dun1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 dun1Δ:KanMX4*	(Giaever *et al.*, 2002)
YRH78	*mrc1Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mrc1Δ:KanMX4*	(Giaever *et al.*, 2002)
YJH363	*pol32Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pol32Δ:KanMX4*	(Giaever *et al.*, 2002)
YRH86	MEP *rtt109Δ* diploid	*ade2::hisG his3 leu2 lys2/+ met15::ADE2/+ ura3Δ0 trp1Δ63 hoD::SCW11pr-Cre-EBD78-NatMX loxP-UBC9-loxP-LEU2 loxP-CDC20-Intron-loxP-HPHMX rtt109::TRP1*	(Hull *et al.*, 2017)
YRH82	*P_{GAL}-HA rtt109Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ade2::MET25 17x[P_{GAL1}-3HA]-ADE2 rtt109::KanMX6 pRS316-CUP1*	(Hull *et al.*, 2017)
YRH23/24	3x*CUP1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3xCUP1*	(Hull *et al.*, 2017)
YRH42	3x*CUP1 rtt109Δ*	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3xCUP1 rtt109Δ::KanMX6*	(Hull *et al.*, 2017)
YRH28	3x[*CUP1-NodA*]	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3x[CUP1-NodA]*	This study
YRH30	3x[*CUP1-Ttef*]	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3x[CUP1-Ttef]*	This study
YRH33	3x[*GFP-CUP1*]	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3x[GFP-CUP1]*	This study
YRH108/109 /110	3x*CUP1 trp1Δ*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-3xCUP1 trp1Δ::NatMX6*	(Hull *et al.*, 2017)
YRH149	2x*CUP1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-2xCUP1*	This study
YRH87	1x*CUP1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-CUP1*	This study
YRH151	1x*CUP1 met17Δ*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 cup1Δ::ADE2-CUP1 met17Δ::KanMX6*	This study
YRH161	1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int.	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 met17Δ::KanMX6 cup1Δ::ADE2*-2x[*CUP1*]-*MET25-GFP* Note: A plasmid containing a single *CUP1* repeat, a *MET25-GFP* marker, and no other origin of replication or centromere integrated alongside a single genomic *CUP1* repeat. A single *ADE2* marker is present at the telomere proximal end.	This study
YRH89	3x*SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-3xSFA1*	(Hull *et al.*, 2017)
YRH100	3x*SFA1 rtt109Δ*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-3xSFA1 rtt109Δ::KanMX6*	(Hull *et al.*, 2017)
YRH127	7x*SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-7xSFA1*	This study
YRH129	6x*GFP-$_{LAG}$P-SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-ugx2Δ::*6x*GFP-$_{1LAG}$P-SFA1*	This study
YRH131	7x*UGX2-P_{GAL}-GFP-SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-7xUGX2-P_{GAL1}-GFP-SFA1*	This study
YRH130	10x*UGX2-P_{GAL}-GFP-SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-10xUGX2-P_{GAL1}-GFP-SFA1*	This study
YRH159	12x*UGX2-$_{LAG}$P-GFP-SFA1*	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ade2Δ::LEU2 sfa1::NatMX6 cup1Δ::ADE2-CUP1-12xUGX2-$_{1LAG}$P-GFP-SFA1*	This study
YRH145	MEP 1x*CUP1*-1x*SFA1* heterozygote	*ade2::hisG/ade2::hisG his3/his3 leu2/leu2 lys2/+ met15D::ADE2/+ ura3D0/ura3Δo trp1D63/trp1D63 hoD::SCW11pr-Cre-EBD78-NatMX/hoD::SCW11pr-Cre-EBD78-NatMX loxP-UBC9-loxP-LEU2/lOXp-UBC9-loxP-LEU2 loxP-CDC20-Intron-loxP-HPHMX/loxP-CDC20-Intron-loxP-HPHMX sfa1::KanMX6/sfa1::KanMX6 cup1::Trp1/cup1::ADE2-SFA1 V208I-CUP1*	This study
YRH146	MEP 1x*CUP1*-1x*SFA1* Diploid	*ade2::hisG/ade2::hisG his3/his3 leu2/leu2 lys2/+ met15D::ADE2/+ ura3D0/ura3Δo trp1D63/trp1D63 hoD::SCW11pr-Cre-EBD78-NatMX/hoD::SCW11pr-Cre-EBD78-NatMX loxP-UBC9-loxP-LEU2/lOXp-UBC9-loxP-LEU2 loxP-CDC20-Intron-loxP-HPHMX/loxP-CDC20-Intron-loxP-HPHMX sfa1::KanMX6/sfa1::KanMX6 cup1:: ADE2-SFA1 V208I-CUP1 /cup1::ADE2-SFA1 V208I-CUP1*	This study

Table 2.2: Plasmids used in this study.

Plasmid ID	Plasmid Name	Construction	Reference
pJH252	pRS316-*CUP1*	*CUP1* PCR cloned into pRS316 by *Cla*I *Not*I	(Hull *et al.*, 2017)
pJH254	pBS-*CUP1*	*CUP1* PCR cloned into pBS II SK- by *Eco*RI *Cla*I	(Hull *et al.*, 2017)
pJH278	pBS-P_{GAL1}-*3HA*	Multi-fragment construction: pBS-CUP1 *Spe*I *Eco*RI P_{GAL1} PCR *Spe*I *Pac*I *3HA* PCR *Pac*I *Nhe*I *CUP1* PCR 3' *Nhe*I *Eco*RI	(Hull *et al.*, 2017)
pJH264	p*ADE2-CUP1*-flanks	Multi-fragment construction: pBSII SK- *Kpn*I *Sac*I *CUP1* flank PCR *Kpn*I *Cla*I *ADE2* PCR *Cla*I *Spe*I *CUP1* flank PCR *Spe*I *Sac*I	(Hull *et al.*, 2017)
pJH280	p*ADE2* 3x P_{GAL1}-*3HA*	Multi-fragment construction: pADE2-*CUP1*-flanks *Cla*I *Eco*RI pBS-P_{GAL1}-*3HA Cla*I *Sal*I pBS-P_{GAL1}-*3HA Xho*I *Ngo*MIV pBS-P_{GAL1}-*3HA Xma*I *Eco*RI	(Hull *et al.*, 2017)
pRH9	p*ADE2-CUP1*	Multi-fragment construction: pBS II SK- *Xho*I *Sac*I p*ADE2-CUP1*-Flanks *Eco*RI *Sac*I *RSC30* right flank PCR *Xho*I *Eco*RI	(Hull *et al.*, 2017)
pRH12	p*ADE2*-3x*CUP1*	Multi-fragment construction: p*ADE2-CUP1 Sal1 EcoR1* pBS-*CUP1 Xho1 BglII* pBS-*CUP1 BamH1 EcoR1*	(Hull *et al.*, 2017)
pJH295	pBS *CUP1-NodA*	*CUP1* PCR cloned into pBS II SK- by *Cla*I *Not*I	This study
pJH296	pBS *CUP1-Ttef*	Multi-fragment construction: pBS II SK *Cla*I *Not*I PCR 1083-1262 on BY4741 gDNA *Cla*I *Nhe*I PCR 1263-1264 on pFA6a-*KanMX6 Nhe*I *Not*I	This study
pJH285	pBS *GFP-CUP1*	pJH254 *Xma*I *Eco*RV backbone pRH28 *Xma*I *Bgl*II (blunt) insert	This study
pRH22	p*ADE2*-2x*CUP1*	*Sal*I *Eco*RI fragment of p*ADE2-CUP1* cloned into *Xho*I *Eco*R1 sites of pBS-*CUP1*	This study
pJH97	pRS411	-	(Cost and Boeke, 1996; Brachmann *et al.*, 1998)
pJH353	p*CUP1-MET25-GFP*	PCR on YSF53 gDNA *Bgl*II *Not*I cloned into *Bgl*II *Not*I pBS-*CUP1*	This study
pJH294	pBS *SFA1*	*SFA1* PCR cloned into pBS II SK- by *Xho*I *Not*I	(Hull *et al.*, 2017)
pJH310	p*ADE2-CUP1*-1x*SFA1*	*Xho*I *Eco*RI fragment of pJH294 cloned into *Sal*I *Eco*RI sites of pRH9	(Hull *et al.*, 2017)
pJH311	pBS 2x*SFA1*	*Xho*I *Eco*RI fragment of pJH294 cloned into *Sal*I *Eco*RI sites of pJH294	(Hull *et al.*, 2017)
pJH312	p*ADE2-CUP1*-3x*SFA1*	*Xho*I *Eco*RI fragment of pJH311 cloned into *Sal*I *Eco*RI sites of pJH310	(Hull *et al.*, 2017)
pJH337	pGEM *SFA1* part	PCR 1460-1234 on BY4741 gDNA in pGEM-Teasy	(Hull *et al.*, 2017)
pJH338	pJET P_{GAL1}-*GFP*	Multi-fragment construction: PCR 1485-1110 on BY4741 gDNA *Pac*I digest PCR pFA6a F1-1486 on pFA6a-*GFP-TRP1 Pac*I digest pJET *Eco*RV	(Hull *et al.*, 2017)
pJH340	pGEM *ugx2::*$_{1LAG}$*P-GFP-SFA1*	pJH337 *Nhe*I *Pst*I + pJH338 *Nhe*I *Pst*I	(Hull *et al.*, 2017)

pJH347	pGEM 2x*ugx2::$_{1LAG}$P-GFP-SFA1*	pJH340 *Xho*I *Xma*I + pJH340 *Xho*I *Ngo*MIV	(Hull *et al.*, 2017)
pJH348	p*ADE2-CUP1 ugx2::$_{1LAG}$P-GFP-SFA1*	pRH9 *Sal*I *Eco*RI + pJH340 *Xho*I *Eco*RI	(Hull *et al.*, 2017)
pJH350	p*ADE2-CUP1 3xugx2::$_{1LAG}$P –GFP-SFA1*	pJH348 *Sal*I *Eco*RI + pJH347 *Xho*I *Eco*RI	(Hull *et al.*, 2017)
pJH336	pGEM *SFA1* F1R5	pGEM T-easy + PCR *SFA1* on BY4741 gDNA	This study
pJH341	pJET *SFA1* F6R1	pJET 2.1 + PCR *SFA1* on BY4741 gDNA	This study
pJH342	pJET *SFA1* F1R5 F6R1	*Cla*I *Sph*I pJH341 + *Cla*I *Sph*I pJH336	This study
pJH343	pGEM *P_{GAL1}-GFP* NS	pGEM T-easy + PCR *P_{Gal}-GFP* on YJH281 gDNA *Nhe*I *Sph*I	This study
pJH344	pJET *UGX2 P_{GAL1}-GFP SFA1*	*Nhe*I *Sph*I pJH342 + *Nhe*I *Sph*I pJH343 (1.3kb fragment)	This study
pJH345	pJET 2x*UGX2 P_{GAL1}-GFP SFA1*	*Xba*I *Ngo*MIV pJH344 + *Xba*I *Xma*I pJH344	This study
pJH346	p*ADE2-CUP1 1xUGX2 P_{GAL1}-GFP SFA1*	*Eco*RI *Sal*I pRH9 + *Xho*I *Eco*RI pJH344	This study
pJH349	p*ADE2-CUP1 3xUGX2 P_{GAL1}-GFP SFA1*	*Sal*I *Eco*RI pJH346 + *Xho*I *Eco*RI pJH345	This study
pRH25	pJET-*P_{Gal10}*	PCR *P_{Gal10}* on BY4741 gDNA and blunt end ligate into pJET 1.2	This study
pRH28	p*UGX2-$_{1LAG}$P-GFP-SFA1*	Multi-fragment construction: *Eco*RI *Sal*I pRH25 *Nhe*I *Sph*I pJH344 *Sph*I *Pac*I pJH344	This study
pRH31	p*Ade2-CUP1 UGX2-$_{1LAG}$P-GFP-SFA1*	pRH9 *Eco*RI *Sal*I + pRH28 *Xho*I *Eco*RI	This study
pRH34	p2x*UGX2-$_{1LAG}$P-GFP-SFA1*	pRH28 *Xba*I *Ngo*MIV + pRH28 *Xba*I *Xma*I	This study
pRH37	p*ADE2-CUP1-* 3x*UGX2-$_{1LAG}$P-GFP-SFA1*	pRH34 *Xho*I *Eco*RI + *Sal*I *Eco*RI pRH31	This study
pRH15	pBS *SFA1* V208I	V208I substitution mutation in *SFA1* in pBS *SFA1*	This study
pRH19	pBS *ADE2-CUP1-RSC30-SFA1* V208I	*Eco*RI *Sal*I fragment of p*ADE2-CUP1* cloned into *Eco*RI *Xho*I sites of pBS *SFA1* V208I	This study

Table 2.3: Oligonucleotide pairs used in this study.

Oligonucleotide sequences	Description
AGTTTCGATAGTTTAAACTCCGACTATATCTGAGACGAACA…TATGAAATCTGGCGCGCCTGGTGGTTGGCAAATGAC	*MET25* – on BY4741
ATTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTATATAGAATTCGAGCTCGTTTAAAC…AGTATATCATCTCATTTCCGTAAATACCAAATGTATTATATATTGCGGATCCCCGGGTTAATTAAG	Deletion cassette for *ADE2* – on pFA6a-*MET25* plasmid
CACTGTTACCATCGATACTCGAGGATCCCGGGATTCATGGTACCCGCTGCTGAA…AACTGACCATGCGGCCGCGAATTCAGATCTGTCGACGCCGGCTTTTTTTTTTTTTTTATTCGA	*CUP1* repeat for pBS *CUP1* and pRS316 *CUP1* – on BY4741 gDNA
TCGTATTAGTACTAGTTTATATTGAATTTTCAAAAATTCTTAC…ATTTTGGAAGTTAATTAATTCGCTGAACATTATAGTTTTTTCTCCTTGACGTT	P_{GAL1} – on BY4741 gDNA
CGGATCCCCGGGTTAATTAAG…TATGGTCTCTGCTAGCTCAGCACTGAGCAGCGTAATC	*3HA* – on pFA6a-*3HA-KanMX6*
GTGTTCAACTGCTAGCAACGAATAGTCTTTAATATATTCATCTAAC…AACTGACCATGCGGCCGCGAATTCAGATCTGTCGACGCCGGCTTTTTTTTTTTTTTTTATTCGA	*CUP1* 3' – on BY4741 gDNA
TACGCTCAGTGGTACCGAGCTCTCAATGACCCTATTCAATAAGCA…GAACTCAGTCATCGATGAAGACTGACCTAGAAGCGAATG	*CUP1* flank – on BY4741 gDNA
CCATCGATTTACTAGTGCATATGTATTAATCCTAAAATGTATTAT…TTAACTAGGAGAGCTCTAGCGAGTCAGAAGCTGTCAAG	*CUP1* flank – on BY4741 gDNA
TAGTGCTAGTATCGATTGAATCGAATTCCACCATAGATCTGAATTAATTCTTGA…GGTCGATCCTACTAGTCACACCGCATAGATCTTATGTATG	*ADE2* for 3xP_{GAL1}-*3HA* plasmid – on BY4741
ATTTTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATAGAATTCGAGCTCGTTTAAAC…TTTTTGAAAAAAATGTATTACTCAAGACATTCGCTTCTAGGTCAGTCTTCCGGATCCCCGGGTTAATTAAG	Deletion cassette for *CUP1* repeats – on pFA6a-*LEU2* plasmid
CAGTAGAGTTAAAAGGTCAATTCAACCGGTCTTCAATAAGACATGCGGATCCCCGGGTTAATTAAG…CGATGCTACATACGTGTACTAAATAATAAATATCAATATGTATCAGAATTCGAGCTCGTTTAAAC	Deletion cassette for *RTT109* - on pFA6a-*TRP1* plasmid
AATTCTAAGAAAGGCAAGGTTGA…GAACTCAGTCATCGATGAAGACTGACCTAGAAGCGAATG	*CUP1* flank probe (single copy target) – on BY4741 gDNA
AACTGACCATGCATGCTAGTTAGAAAAAGACATTTTTGCTGT…GGATCCTAATACGACTCACTATAGGGAGAGGATCATTTCCCAGAGCAGCATGA	*CUP1* specific probe (multi-copy target) – on BY4741 gDNA
TCGTATTAGTACTAGTTTATATTGAATTTTCAAAAATTCTTAC…GGATCCTAATACGACTCACTATAGGGAGAGGATCAGCACTGAGCAGCGTAATC	P_{GAL1}-*3HA* specific probe (multi-copy target) – on pBS-P_{GAL1}-*3HA* gDNA
ATGTTCAGCGAATTAATTAACTTCCA…GGATCCTAATACGACTCACTATAGGGAGAGGATCATTTCCCAGAGCAGCATGA	*CUP1* RNA probe – on BY4741 gDNA
GGATCCTAATACGACTCACTATAGGGAGAGGAGCGTATCCTTTTTACGAGATGAAA…CCGTATAAACCTATACACATATA	*CUP1* CUT RNA probe – on BY4741 gDNA
TGATAAATCCCCGTTAAGTCGTA…GCCAACCTGAGCAGTAGAGTAA	*rtt109::KanMX4* – on YJH237 gDNA

CACTGTTACCATCGATACTCGAGGATCCCGGGATTCATGGTACCCGCTGC TGAA…AACTGACCATGCTAGCTCATTTCCCAGAGCAGCAT	*CUP1* ORF and no terminator sequence
AACTGACCATGCTAGCTCAGTACTGACAATAAAAAGATTCTTG…AACTGA CCATGCGGCCGCGAATTCAGATCTGTCGACGCCGGCTTTTTTTTTTTTTT TTTGGATGGCGGCGTTAGTATC	*Ttef* sequence from pFA6a-*KanMX6*
TGATTATTTTCAGGGGTGTCCGAGTCCACCTCTACAACATCCACCCGGAT CCCCGGGTTAATTAAG…TTCCAGAAAATTTGAGTCATGCTTACTTAGTTT AATTAAGTACTC GAATTCGAGCTCGTTTAAAC	*SFA1* deletion – on pFA6a-*NatMX6*
CACTGTTACCATCGATACTCGAGTCTAGACCCGGGTGCCCGGTTTGATAC CTGTA…AACTGACCATGCGGCCGCGAATTCACTAGTCGACGCCGGCTTTT TTTTTTTTTTTTCAAAAGTGACAAAGCTAACGT	*SFA1* cloning – on BY4741 gDNA
CGTATGATGCGAAGAAACCA…CCTGTGAGTGATAAATTCTTCGA	*SFA1* probe – on BY4741 gDNA
TACGCTCAGTCTCGAGGAGCTCTCAATGACCCTATTCAATAAGCA…AACT GACCATGCGGCCGCGAATTCAGATCTGTCGACGCCGGCTTTTTTTTTTTTT TTTTATTCGA	*RSC30* right flank - on BY4741 gDNA
CGTATGATGCGAAGAAACCA…GGATCCTAATACGACTCACTATAGGGAGA GGACGCCTTTTTGCACATTAGCT	*SFA1* ORF probe (Southern) – on BY4741 gDNA
GGATCCTAATACGACTCACTATAGGGAGAGGAATCCTTATCAAATGATAC GCTCTT…ATAGGCCAACAAATTCTGACACT	*SFA1* upstream CUT probe – on BY4741 gDNA
GGTCGTCAGATACATAGATACAATTCTATTACCCCCATCCATACACGGAT CCCCGGGTTAATTAAG…AGAGAAAGTAGGTTTATACATAATTTTACAACT CATTACGCACACGAATTCGAGCTCGTTTAAAC	*MET17* deletion on pFA6a-*KanMX6*
TCGTATTAGTGCTAGCTTATATTGAATTTTCAAAAATTCTTAC…ATTTTG GAAGTTAATTAATTCGCTGAACATTATAGTTTTTTCTCCTTGACGTT	P_{GAL1} for P_{GAL1}-*GFP SFA1* construct (1485-1110) – on BY4741 gDNA
AATCGCTTCAGCTAGCGATTTAAGGCGGTAAGAAGGA…AACTGACCATGC GGCCGCGAATTCACTAGTCGACGCCGGCTTTTTTTTTTTTTTTCAAAAG TGACAAAGCTAACGT	*SFA1* for P_{GAL1}-*GFP SFA1* construct (1460-1234) – on BY4741 gDNA
CGGATCCCCGGGTTAATTAA…AATCGCTTCACCCGGGTGATCTATATTAC CCTGTTATCCCTAG	*GFP* for P_{GAL}-*GFP SFA1* construct (pFA6a F1-1486) – on pFA6a-*GFP-TRP1*
CACTGTTACCATCGATACTCGAGTCTAGACCCGGGTGCCCGGTTTGATAC CTGTA…AATCGCTTCAGCTAGCACCCACACACCGCATGTCTT	*SFA1* for pJH336 on BY4741 gDNA
AATCGCTTCAGCATGCGGTCTTTTTACCTCTGTAAGTTACGT… AACTGACCATGCGGCCGCGAATTCACTAGTCGACGCCGGCTTTTTTTTTTT TTTTTTCAAAAGTGACAAAGCTAACGT	*SFA1* for pJH341 on BY4741 gDNA
TCGTATTAGTGCTAGCTTATATTGAATTTTCAAAAATTCTTAC… AATCGCTTCAGCATGCTGATCTATATTACCCTGTTATCCCTAG	P_{GAL1}-*GFP* for pJH343 on YJH281 gDNA
GCTAGCGTTTTTTCTCCTTGACGTTAAAGT…TTAATTAACATTTATATTG AATTTTCAAAAATTCTTAC	P_{Gal10} promoter for pRH25 on BY4741 gDNA
CTGCGGGACTATCGGACTCTCCG…CCAAATACTGCAACGGTATC	*SFA1* V208I site directed mutagenesis on pJH294

2.1.2 Liquid media

All experiments were performed in yeast nitrogen base media supplemented with CSM amino acids, unless otherwise stated, and 2% Glucose or Galactose. All media components were purchased from Formedium. Cells were grown at 30°C with shaking at 200rpm, unless stated otherwise. Yeast nitrogen base media contains 250nM $CuSO_4$. Nicotinamide (Sigma 47865-U) was added to the media at 5mM. Rapamycin (Santa Cruz Biotechnology sc-3504) was added to the media at 25nM. $CuSO_4$ and formaldehyde (103999 MERCK Formaldehyde solution min. 37% free from acid, or 11586711 Thermo Scientific™ Pierce™ Methanol-free Formaldehyde Ampules) were added to the media at the concentrations stated.

2.1.3 Solid media

Yeasts were grown on 25ml solid YPD agar or YNB agar supplemented with the appropriate amino acid drop-out media in standard 90mm Petri Dishes. For negative selection of *URA3* containing plasmids, yeasts were grown on YNB agar with the appropriate amino acid supplements and 1mg ml^{-1} 5-FOA (Formedium 5FOA10) and 50µg ml^{-1} Uracil (Formedium DOC0212). For 1x*CUP1* 1x*SFA1* amplification analysis, cells were plated on YNB agar lacking $CuSO_4$ and supplemented with the appropriate amino acids and $CuSO_4$ was added to the media at 200µM and formaldehyde at 1.6mM.

2.1.4 Spectrophotometric quantification of yeast cultures

Yeast culture cell density was approximated using Amersham Biosciences Ultrospec 10 Cell Density Meter and OD_{600nm} readings were converted to actual cell density using widely available conversion tables. For OD values between 0.05 and 1, the conversion relationship is roughly linearly proportional to cell number. For cultures with OD values greater than 1, the

spectrophotometric sample was diluted (usually 1 in 10) to give a more accurate reading within the linear range.

2.1.5 Yeast transformation

For transformation, 4ml yeast culture was grown to mid-log and all $7x10^6$ cells were washed first with 4ml water and secondly with 1ml water. Cells were re-suspended in 240µl 50% w/v PEG, 34µl genomic DNA transformation template or 31µl water with 3µl plasmid template, 50µl 2mg ml^{-1} salmon sperm DNA (Sigma, boiled prior to use) and 36µl 1M LiOAc. Transformations were incubated for 40 minutes at 42°C, then pelleted and re-suspended in 50µl water. Transformation cells were then plated either directly onto the appropriate selective auxotrophic plate and given 2 to 3 days growth at 30°C, or when using antibiotic selection, first plated onto a YPD plate for overnight growth at 30°C and then replica plated onto a YPD plate containing the selective antibiotic markers and grown for 2 to 3 days at 30°C.

2.2 DNA isolation for genotyping

A single large colony of yeast was picked up from an agar plate and re-suspended in 200µl DNA Fast Lysis buffer (2% Triton-X100, 1% SDS, 100mM NaCl, 1mM EDTA, and 10mM Tris pH8) with 50µl acid-washed glass beads. DNA was phenol-chloroform extracted by vortexing, then centrifugation at top speed for 5 minutes at 4°C. The upper phase contained the DNA for proceeding directly into a genotyping PCR.

2.3 Polymerase Chain Reaction (PCR)

Various polymerase enzymes were used for PCR, dependent upon the required downstream applications. Actual PCR reaction compositions and run conditions are shown in Table 2.4. Takara LA® polymerase was the preferred polymerase for amplifying deletion cassette markers for transformation.

Takara LA® PCRs were performed in 40µl reactions with the total sample being gel extracted prior to transformation, using either the QIAGEN or NEB gel extraction kits as per manufacturer's protocol. Phusion® was the preferred polymerase for all cloning reactions because of its high fidelity and reactions were performed in 50µl. For probe template production, Phire II® polymerase was preferred and reactions were performed in 50µl. All probe templates were gel extracted, as above, before use in a random primed probing reaction. Genotyping reactions were performed using One*Taq*® DNA polymerase (NEB) in 12.5µl reactions.

Table 2.4: Polymerase chain reaction components and run conditions.

DNA Polymerase	Reaction Composition	Run Conditions
Takara LA®	4µl 10x LA PCR buffer II 6.4µl 2.5mM Takara dNTPs 0.2µl 100µM each primer 0.8µl template DNA 0.4µl Takara LA® Water to 40µl	94°C for 2 minutes 98°C for 10 seconds 50°C for 30 seconds 68°C for 1 minute per kb 68°C for 5 minutes
Phusion®	10µl 5x Phusion buffer 1µl 10mM dNTPs 0.2µl 100µM each primer 1µl template DNA 0.5µl Phusion® Water to 50µl	98°C for 30 seconds 98°C for 10 seconds 50°C for 10 seconds 72°C for 15-30 seconds per kb 72°C for 5 minutes 35 cycles
Phire II®	10µl 5x Phire II buffer 1µl 10mM dNTPs 1µl 10µM each primer 1µl template DNA 1µl Phire II® Water to 50µl	98°C for 30 seconds 98°C for 5 seconds 50°C for 5 seconds 72°C for 10-15 seconds per kb 72°C for 1 minute 35 cycles
One*Taq*®	6.25µl 2x One*Taq*® Master Mix 0.25µl 10µM each primer 0.25µl upper phase DNA Water to 12.5µl	94°C for 30 seconds 94°C for 15 seconds 50°C for 15 seconds 68°C for 1 minute per kb 68°C for 5 minutes 35 cycles

2.4 DNA gel electrophoresis

DNA was resolved on 1xTBE agarose gels ranging from 0.8% to 1% for between 20 minutes and 1 hour, depending upon the expected DNA fragment size. Gels contained 2.5µl Sybr® Safe (Invitrogen) or were post-stained with 500ml of 0.1µg ml^{-1} ethidium bromide in water for 15 minutes with shaking and 5 minutes de-staining in 500ml water with shaking. Gel visualization was performed using a Gel Doc™ XR+ System (BioRad), or if the DNA was to be used for any downstream processes, using a blue-light-transilluminator.

2.5 DNA sequencing

All created and used plasmids were verified through restriction digest and/or DNA sequencing. All DNA sequencing was performed by Beckman Coulter Genomics, with the sequencing data analysed using DNASTAR Lasergene.

2.6 DNA extraction and Southern blotting

Cells from a 2ml saturated culture were washed with 50mM EDTA then spheroplasted with 250µl 0.34U ml^{-1} lyticase (Sigma L4025) in 1.2M sorbitol, 50mM EDTA, and 10mM DTT at 37°C for 45 minutes. After centrifuging at 1,000g, cells were gently re-suspended in 400µl of 0.3% SDS, 50mM EDTA, and 100µg ml^{-1} RNase A (Sigma R4875) and incubated at 37°C for 30 minutes. 4µl of 20mg ml^{-1} proteinase K (Roche 3115801) was added, and samples were mixed by inversion and heated to 65°C for 30 minutes. 160µl 5M KOAc was added and samples were mixed by inversion and then chilled on ice for 30 minutes. After 10 minutes of centrifuging at 20,000g, the supernatant was poured into a new tube containing 500µl phenol:chloroform (pH 8) and samples were mixed on a wheel for 30 minutes. Samples were centrifuged for 10 minutes at 10,000g, and the upper phase was extracted using cut tips and precipitated with 400µl isopropanol. Pellets were washed with 70% ethanol, air-dried, and left overnight at 4°C to dissolve in 20µl TE.

10µl of dissolved DNA was digested with 20U restriction enzyme (NEB), as stated, for 3 hours at 37°C. Digested DNA was then phenol-chloroform extracted, ethanol precipitated and re-suspended in 20µl TE. DNA concentration quantification was performed using a NanoDrop 2000 (ThermoFisher Scientific) and equivalent DNA amounts were loaded. For analysis of *CUP1* DNA, electrophoresis was performed on a 25cm 0.8% or 1% 1xTBE gel in 1xTBE at 90-120V for 16-24 hours, depending on the number of *CUP1* repeats in the parental strain. For *SFA1* analysis DNA was run on a 1% 0.5xTBE gel in a Bio-Rad CHEF DR-III system at 6 V cm^{-1}, 15°C, 0.5–1.5 second switch, 120° included angle, in 0.5xTBE.

Gels were stained for 15 minutes in 500ml of 0.1µg ml^{-1} ethidium bromide in water with gentle shaking and then de-stained for 5 minutes in 500ml water. An ethidium bromide staining image was acquired, using either the Gel Doc™ XR+ System (BioRad) or the Typhoon FLA 9500 (GE). Gels were then depurinated for 15 minutes in 0.25N HCl, then denatured for 40 minutes in 0.5N NaOH, followed by washing twice in neutralising 1.5M NaCl, 0.5M Tris (pH7.5). DNA was then given at least 16 hours to transfer to a HyBond N+ membrane (GE) by capillary transfer in 6xSSC.

The HyBond N+ membrane with transferred DNA was cross-linked using a Stratagene 1800 Stratalinker UV Crosslinker. Pre-hybridisation of the cross-linked membrane was at 42°C in 10ml Ultrahyb buffer (Life Technologies) for 1 hour. The probe template was a 100-1000bp PCR product, gel-extracted prior to use. 25ng probe template was diluted in 37µl water and boiled for 5 minutes at 100°C, then placed on ice for 5 minutes. To the probe, 10µl of 5x labelling buffer (250µl NEBuffer 2; 0.825µl of each 100mM dATP, dGTP, and dTTP; 12.5µl of 1mg ml^{-1} random 9-mers (NEB S1254S dissolved in 33µl water to 1mg ml^{-1}); and water to 500µl), 1µl Klenow 5'-3' exo-enzyme (NEB), and 2µl (20µCi) [α-^{32}P]dCTP 3000Ci $mmol^{-1}$ were added and mixed by pipetting. Probe mix was incubated 1 hour at 37°C, and then cleaned using a miniQuick Spin column (Roche) by centrifuging at 3000rpm for 4 minutes. Probe mix was then boiled at 100°C for 5 minutes, then transferred to ice for 2 minutes, and added to the pre-hybridising membrane at 42°C. The hybridisation reaction was incubated for at least 16 hours.

After hybridization, the probe was discarded and the membrane was washed twice in 50ml of 6xSSC, and then incubated for 20 minutes at 42°C in 50ml of 6xSSC. The membrane was then incubated twice, for 20 minutes each, in 50ml of 0.1xSSC 0.1% SDS at 42°C. The membrane was then scrubbed in 0.1xSSC 0.1% SDS at room temperature to remove background probe. Probed membranes were exposed to a FUJIFILM imaging plate for 3 hours and then a short exposure image was taken using a Typhoon FLA 9500 (GE) at 1000PMT and 200µm Pixel size. Dependent upon the signal to noise of the short exposure image, the probed membrane was then re-exposed to an imaging plate for 1-4 days and a long exposure image was taken at 500 or 700PMT and 100µm Pixel size. Bands were quantified using ImageQuant (GE) and data analysed using GraphPad Prism v6.05.

2.7 Rapid DNA extraction and PFGE for Cell Fate Analysis

Individual colonies grown on non-selective plates, following the Cell Fate Analysis experiment, were cultured to stationary phase in 100µl synthetic complete medium in 96-well plates for 3 days at 30°C. 50µl culture from four outgrowths were pooled per sample and re-suspended in 50µl 50mM EDTA containing 17U lyticase (Sigma L2524) and incubated at 37°C for 45 minutes. 1.6µl of 10% SDS and 1µl of 20mg ml^{-1} Proteinase K were added and incubated at 65°C for 30 minutes. 32µl of 5M KOAc was added and samples were kept on ice for 30 minutes before centrifuging for 10 minutes at 20,000g. The supernatant was decanted and DNA precipitated with 100µl isopropanol and 1µl glycogen, centrifuging at 20,000g for 15 minutes at 4°C. The DNA pellet was washed with 70% ethanol and air-dried. The DNA pellet was then incubated overnight at 37°C in 20µl 1x NEB CutSmart buffer with 20U *Eco*RI-HF (NEB). DNA was quantified using PicoGreen (Thermo Fisher Scientific) and equivalent amounts of DNA were loaded on a 1% 0.5xTBE gel in a Bio-Rad CHEF DR-III system at 6 V cm^{-1}, 15°C, 0.5–1.5 second switch, 120° included angle, in 0.5xTBE for 16 to 20 hours, as specified. Gels were

ethidium bromide imaged, washed, blotted and probed as above for DNA extraction and Southern blotting.

2.8 Cu^{2+} sensitivity and growth curve assays

For the Cu^{2+} sensitivity assays, 2.5µl saturated culture was diluted to 200µl in synthetic complete in each well of a 96-well flat-bottomed cell culture plate, with concentrations of $CuSO_4$ up to 3mM. For Cu^{2+} sensitivity assays involving the MEP wild-type, 0.5mM ascorbic acid was also added, since ascorbate increases the cellular uptake of copper (Hassett and Kosman, 1995), thus increasing the effective toxicity of the copper used. This allowed for the measurement of small changes in resistance in cells with high *CUP1* copy numbers to be observed, without having to use such high copper concentrations that copper precipitates out of the media. Plates were covered with a gas-permeable membrane and grown at 30°C for 3 days in the dark, since ascorbate is light sensitive. Cells were re-suspended by pipetting, and the OD_{600nm} was measured using a BD FLUOstar Omega plate reader. Area-under-curve measurements were calculated for each sample and compared by 1-way ANOVA. For 3x*CUP1* cells and derivatives, the assay was performed with lower concentrations of $CuSO_4$ and under normal light, without ascorbic acid addition.

For growth curves, saturated pre-cultures were diluted 1:1,000 into 200µl synthetic complete per well with or without $CuSO_4$ at the required concentration. Plates were sealed as above and grown at 30°C with shaking in a BD FLUOstar Omega plate reader. The OD_{600nm} measurements were taken every 15 minutes and smoothed growth curves were generated by taking a nine time-point rolling average. Derivatives were calculated using GraphPad Prism.

2.9 Competition assay

Six 3x*CUP1* wild-type and six 3x*CUP1 trp1*Δ::*NatMX6* cultures were grown for 10 generations. 3 cultures of each strain were grown in 5mM nicotinamide,

the other 3 were left untreated. The untreated cells of one strain were 1:1 mixed with the nicotinamide-treated cells of the other strain, for a total of 6 competition cultures. The starting composition of the mixture was determined by plating on –Trp and +Nat plates. Each competition was 1:1,000 diluted in 0 or 0.3mM $CuSO_4$ and outgrown to saturation over 10 generations. Mixture composition of each out-growth culture was determined by plating. To ensure that the *trp1*Δ::*NatMX6* marker did not affect the result, an equal number of assays were performed with this strain as the nicotinamide-treated or untreated population.

2.10 RNA extraction and Northern blotting

25ml yeast cultures in synthetic complete 2% glucose were grown at 37°C from colony to early log phase at approximately OD 0.5. Cultures were split in two, washed with synthetic complete media lacking sugar, re-suspended in experimental condition media and grown for 4 hours at 37°C to mid-log. Cultures were then fixed with 70% ethanol at -80°C. Ethanol fixed cells were pelleted by centrifuging at 4000rpm for 1 minute and ethanol was removed. Cells were then washed with 8ml cold water and re-suspended in 1ml cold water. Cells were transferred to an RNase-free 2ml tube and pelleted by centrifugation at 14,000rpm for 10 seconds. 200µl glass beads were added to cells and 100µl lysis/binding buffer from the mirVANA kit (ThermoFisher Scientific). The cell suspension was vortexed at 4°C for 5 minutes, then proceeded with extraction of total RNA as recommended using the mirVANA kit protocol.

RNA was quantified using a NanoDrop 2000 (ThermoFisher Scientific) and 10µg RNA was mixed with a glyoxal mix (6ml DMSO; 2ml glyoxal (40% solution from sigma); 1.2ml of 10x BPTE; and 0.6ml of 80% glycerol) in a 1:5 v/v ratio. Ethidium bromide was added to each sample at 2µg ml^{-1} and incubated for 1 hour at 55°C. Samples were then loaded on a 1.2% 1xBPTE agarose gel. RNA was separated at 150V for 1 hour 50 minutes in 1xBPTE and an ethidium bromide image was acquired using a Typhoon FLA 9500 (GE). The gel was washed for 20 minutes in 75mM NaOH, followed by 20

minutes in 1.5M NaCl, 0.5M Tris pH7.5. RNA was then given at least 16 hours to transfer to a HyBond N+ membrane (GE) by capillary transfer in 6xSSC. The HyBond N+ membrane (GE) with transferred RNA was cross-linked using a Stratagene 1800 Stratalinker UV Crosslinker and pre-hybridised at 65°C for 1 hour in 8ml Ultrahyb buffer (Life Technologies). RNA probes were synthesised by *in vitro* transcription as follows. Mixed 5µl probe template PCR product; 0.6µl of 100µM rUTP; 1.2µl water; 1µl of 10mM rCTP/rGTP/rATP mix; 2µl of 100mM DTT; 0.2µl of 10mg ml^{-1} BSA, 4µl of 5x T7 transcription buffer (Promega); 5µl of 3000Ci $mmol^{-1}$ [α-^{32}P]rUTP; and 1µl T7 enzyme (Promega). Probe mix was incubated for 1 hour at 37°C, and then cleaned using a miniQuick Spin column (Roche) by centrifuging at 1000g for 4 minutes. The probe mix was then added to the membrane and hybridised at 65°C for 48 hours.

Probed membranes were washed twice with 6xSCC at room temperature, then incubated for 20 minutes at 65°C in 50ml of 6xSSC. Membranes were then incubated twice, each for 20 minutes at 65°C in 0.1xSSC 0.1% SDS. Membranes were then scrubbed in room temperature 0.1xSSC 0.1% SDS to remove background probe. Exposure, imaging and quantification of probed membranes were performed as in DNA extraction and Southern blotting above.

2.11 RNAseq

The same RNA extracted using the mirVANA kit (ThermoFisher Scientific) above for RNA extraction and Northern blotting, was used to make RNAseq libraries. Indexed mRNAseq libraries were constructed from 500ng total RNA using the NEBNext Ultra Directional RNA Library Prep Kit (NEB), with poly(A) selection using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Libraries were sequenced on an Illumina MiSeq according to manufacturer's instructions. Jon Houseley performed all bioinformatics analysis.

2.12 Microscopy for budscar counting

Unlabelled log phase and 24 hour biotin-labelled aged yeast cultures were fixed in 70% ethanol at -80°C. Aged cells were streptavidin purified using MACS LS columns (Miltenyi Biotec), eluting in approximately 1.2ml. 50µl culture was pelleted for rapid biotin staining and re-suspend in 240µl PBS, 9µl of 10% triton X100, 0.3µl of 1mg ml^{-1} streptavidin 594 for visualisation of biotinylated cells, 0.6µl of 1mg ml^{-1} Alexa 488-WGA for visualisation of budscars, and 1µl DAPI solution for identification of dead ethanol fixed cells. The staining mix was incubated for 15 minutes in the dark with rocking, then cells were pelleted, washed once with 300µl PBS containing 0.01% Triton-X100, and re-suspended in 5µl Vectashield. Re-suspended cells were mounted on a microscope slide and budscars were counted in a single-blind experiment, using an Olympus BX41 microscope.

2.13 Spotting assay

MEP 1x*CUP1* 1x*SFA1* homozygous diploid and heterozygote cells were grown to saturation over 10 generations. Cells were then diluted to 10,000 cells $µl^{-1}$ in fresh media and four 10-fold serial dilutions were made in fresh media. 10µl of each dilution was then spotted onto a complete synthetic plate containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde and grown for 7 days at 30°C in a humidified chamber. Final cell numbers plated were 100,000 cells, 10,000 cells, 1000 cells, 100 cells, and 10 cells.

2.14 Plating assay for CNV from single-copy genes

MEP 1x*CUP1* 1x*SFA1* heterozygote cells were grown to early log and then $0.25x10^7$ cells were inoculated into 125ml complete synthetic cultures containing β-estradiol, with and without 0.7mM formaldehyde. The initial number of viable cells in the cultures was determined by plating 1000 cells on –Trp plates for growth at 30°C for 3 days. Cultures were aged for 24 hours

and $1x10^6$ cells were washed with complete synthetic media to remove the β-estradiol and any formaldehyde, then plated on –Trp plates, containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde, and grown for 10 days at 30°C in a humidified chamber. 1000 washed cells were also plated on –Trp plates and grown at 30°C for 3 days, to determine the number of viable cells in the final culture. Overall cell viabilities were calculated using the initial and final cell viability plates.

3. Co-operative stimulation of CNV by transcription and chromatin modifications

3.1 Introduction

One of the prerequisites of our predicted general model for stimulated CNV is for there to be an RFS site just upstream of the candidate gene. Stalled replication forks are recognised by the cell as a type of DNA damage, and as such they become marked by γH2A, the early response marker of DNA damage. Our group performed a genome-wide ChIP-seq profile of γH2A, in wild-type BY4741 haploid yeast, and identified all the sites of DNA damage (Hull *et al.*, 2017). The study focussed on genes with γH2A peaks within 1kb of upstream sequence as candidates for having an upstream RFS site. These candidates were collectively called 'γH2A genes,' with approximately 7% of all yeast genes showing a γH2A enrichment of 2-fold or greater (Hull *et al.*, 2017).

A second prediction from our model is that genes with an upstream RFS site are likely to be generally repressed, except in response to specific environmental stimuli. A meta-analysis of published RNA sequencing datasets on steady-state mRNA levels, performed by our group, found that 'γH2A genes' are biased towards genes that are lowly expressed under optimal growth and highly expressed in suboptimal conditions (Hull *et al.*, 2017). However, this set of 'γH2A genes' did not become highly induced in response to osmotic or oxidative stresses, so cannot simply be stress responders (Hull *et al.*, 2017). The 'γH2A genes' are therefore good candidate genes that may be able to undergo stimulated CNV.

In this study I will focus primarily on one of these 'γH2A genes,' the copper-resistance *CUP1* gene. Given the known importance of *CUP1* as a CNV gene (Fogel and Welch, 1982), and the likely presence of an upstream RFS site (Hull *et al.*, 2017), *CUP1* is a prime candidate for testing the proposed model for stimulated CNV.

3.2 Induction of the bi-directional *CUP1* promoter in response to copper

The final prerequisite of our predicted general model for stimulated CNV is for the gene to be specifically expressed in response to an environmental stimulus from a bi-directional promoter. Therefore, before testing whether *CUP1* can undergo environmentally stimulated CNV, the directionality and inducibility of the *CUP1* promoter needed to be determined. Bi-directional promoters are common throughout the yeast genome, but very often the transcribed antisense RNA is a cryptic unstable transcript (CUT) (Xu *et al.*, 2009). CUTs are very hard to detect in wild-type cells because the CUT is rapidly degraded by RNA surveillance pathways (Wyers *et al.*, 2005; Houseley and Tollervey, 2009; Xu *et al.*, 2009). However, CUTs can be stabilised for detection in an *rrp6Δ* mutant, which lacks a 3'-5' exonuclease important for CUT degradation (Wyers *et al.*, 2005). Therefore, wild-type BY4741 and *rrp6Δ* yeast were grown to mid-log in YPD and then either exposed to 1mM $CuSO_4$ for 4 hours or left untreated. Total RNA was extracted from each sample and the level of induction and directionality of the *CUP1* promoter was assessed by Northern blot analysis (Fig. 3.1a).
The *CUP1* sense mRNA showed a very high level of transcript induction with copper exposure, in both wild-type and *rrp6Δ* yeast. This was unsurprising, as copper has already been shown to promote the rapid expression of *CUP1* mRNA, through Cup2-mediated transcriptional activation of *CUP1* (Peña, Koch and Thiele, 1998). As predicted, the rapid degradation of the *CUP1* antisense CUT, made detection very hard in wild-type cells, but even here a faint signal was obtained with copper induction. However, strong induction of the *CUP1* CUT was observed in the *rrp6Δ* mutant, in response to copper exposure. *ACT1* mRNA was used as a loading control for the total RNA quantities and showed little variation between the samples. Therefore, the *CUP1* promoter is bi-directional, with expression of the *CUP1* sense mRNA and antisense CUT highly induced in response to environmental copper. This meant that the *CUP1* locus fit all the predicted criteria for a stimulated CNV gene.

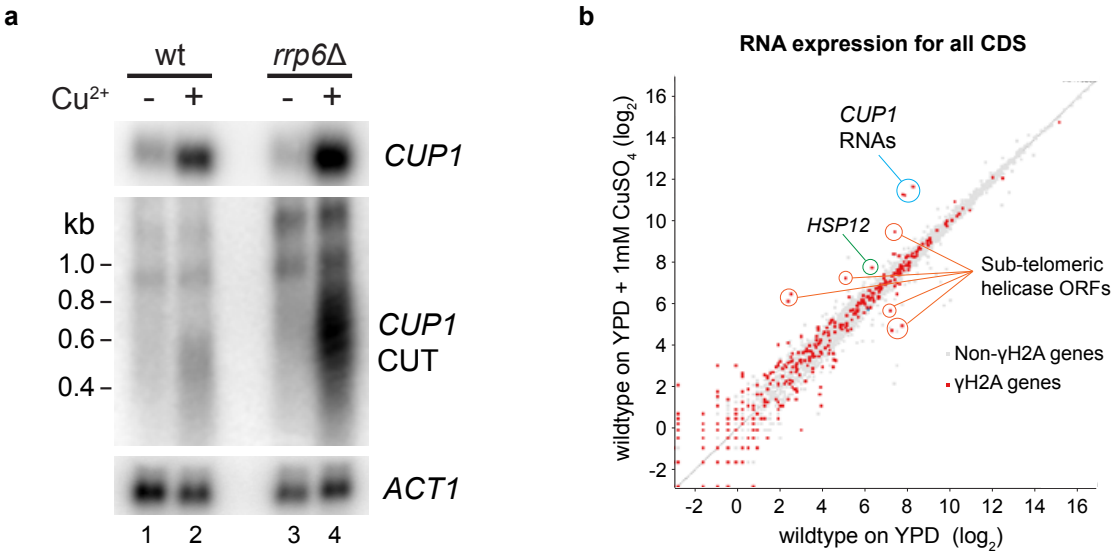


Figure 3.1: Induction of *CUP1* RNAs in response to copper. **a**) Northern analysis of *CUP1* mRNA and upstream CUT. Total RNA from BY4741 wild-type (wt) and *rrp6Δ* yeast, grown to early log-phase in YPD and then either exposed to 1mM $CuSO_4$ for 4 hours or left untreated. Probed for the *CUP1* sense mRNA (top), *CUP1* antisense CUT (middle), and *ACT1* loading control (bottom). **b**) Scatterplot of RNA levels in untreated vs. copper treated yeast. Poly(A)+ RNAseq libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB), using the same extracted RNA from the cells in **a**, and run on an Illumina MiSeq. Read counts mapping to each annotated coding sequence (CDS) were calculated and normalised for feature length. Non-γH2A genes are shown in grey and γH2A genes in red. Substantially induced or repressed genes are annotated with circles: blue circle = *CUP1* RNAs; green circle = *HSP12*; and orange circles = multi-copy sub-telomeric helicase ORFs.

To test whether the induction of the *CUP1* RNAs in response to copper was specific to the *CUP1* locus, the same total RNA used for the Northern blot analysis was also used to make poly(A)+ RNAseq libraries. Our group produced a scatterplot analysis of all yeast genes expressed in untreated vs. copper-treated cells and identified just a small selection of differentially expressed genes (Fig. 3.1b). As expected, the *CUP1* RNAs were some of the strongest differentially expressed, being up-regulated in response to copper. Most of the remaining differentially expressed genes were multi-copy sub-telomeric helicase ORFs. *HSP12* was the only other single verified differentially expressed gene that was up-regulated in copper. *HSP12* encodes a plasma membrane protein known to be important for maintaining

membrane organisation, particularly in various conditions of stress (Praekelt and Meacock, 1990). Interestingly, Hsp12 protein levels increase in response to DNA replication stress (Tkach *et al.*, 2012), which could potentially explain its up-regulation in copper. Despite the small number of differentially expressed genes in response to copper, the RNAseq data set provided evidence for the *CUP1* RNAs being the main transcripts up-regulated in copper-treated cells.

3.3 Creation of the galactose-inducible P_{GAL}-*HA* strain

Environmental exposure of yeast to high levels of copper led to the emergence of highly copper-tolerant strains with a 7-fold amplification of parental *CUP1* copy number (Giusy M Adamo *et al.*, 2012). However, proving that environmental copper actually stimulated *CUP1* CNV, instead of selecting for random pre-existing CNV mutations, required a way to measure the CNV rate in the absence of any selection pressure.

To determine whether CNV was stimulated in the genetic context of the *CUP1* locus, our group replaced the *CUP1* ORF with a 3x*HA* tag ORF that produced a non-functional protein. Then the copper-responsive *CUP1* bi-directional promoter was replaced with a galactose-responsive *GAL1* bi-directional promoter (Fig. 3.2). All other surrounding sequences, including the RFS region, were left intact. Upon transformation the construct fortuitously amplified to 17 repeats, creating the 17xP_{GAL1}-*HA* strain, referred to simply as P_{GAL}-*HA* from here on. The re-engineered *CUP1* locus could now be induced by galactose, a non-selective and non-mutagenic environmental stimulus, and also produce a product that conferred no fitness advantage in the environment. As such, P_{GAL}-*HA* yeast could be used for direct determination of environmentally-stimulated CNV in the genetic context of *CUP1*, free from any issues with selection.

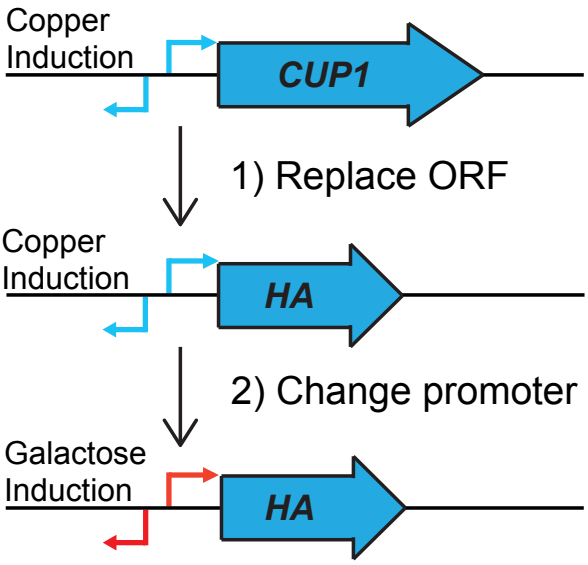


Figure 3.2: Schematic construction of *P_{GAL}-HA*. Construction of the *P_{GAL}-HA* strain first involved replacing the *CUP1* ORF with a non-functional *3xHA* tag (1), followed by replacing the copper responsive *P_{CUP1}* bi-directional promoter with a galactose responsive *P_{GAL1}* bi-directional promoter (2).

3.4 Transcriptionally stimulated CNV in the genetic context of *CUP1*

To determine whether CNV in the genetic context of the *CUP1* locus was environmentally-stimulated, wild-type BY4741 and re-engineered *P_{GAL}-HA* yeast were first grown for 10 generations to saturation in complete synthetic media containing 2% glucose. Then the saturated cultures were 1:1000 diluted in complete synthetic media containing either 2% glucose or 2% galactose and grown for another 10 generations to saturation in the new media (Fig. 3.3a). DNA was extracted from the saturated cultures for Southern blot analysis on a 0.8% agarose gel at 120V for 16.5 hours (Fig. 3.3b). The observed parental copy number for BY4741 and *P_{GAL}-HA* yeasts was 13 *CUP1* and 17 *HA* repeats respectively. The parental copy numbers were consistent with the previously reported 10 to 15 *CUP1* repeats observed in the parental S288C background, and is by no means exceptional compared to *CUP1* copy numbers in wild isolates (Zhao *et al.*, 2014; Strope *et al.*, 2015).

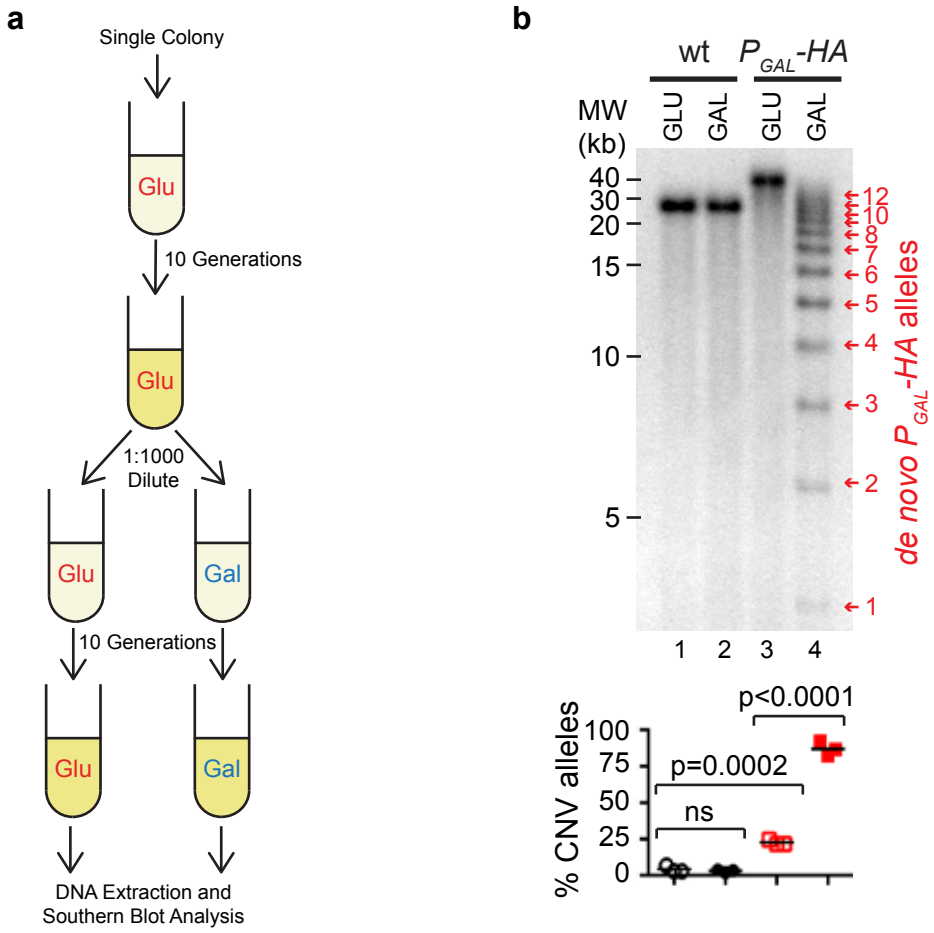


Figure 3.3: Transcriptionally-stimulated CNV. **a**) Experimental design for Southern blot analysis. 4ml complete synthetic 2% glucose cultures are inoculated with a single colony of yeast and grown for 10 generation to saturation. Saturated glucose culture was 1:1000 diluted in complete synthetic 2% glucose and 2% galactose, and grown for 10 further generations to saturation. DNA was extracted from $5x10^7$ cells from saturated cultures for Southern blot analysis. **b**) Southern blot analysis of BY4741 (wt) and P_{GAL}-*HA* yeast grown as in **a**. Copy numbers of parental alleles are 13 and 17 copies in the wt and P_{GAL}-*HA* strains respectively. Quantification shows the percentage of alleles deviating from the parental copy number, *n* = 3. *p*-values calculated by 1-way ANOVA. ns = not significant.

In wild-type yeast, neither glucose nor galactose should induce expression of the *CUP1* transcripts, and in both conditions no novel CNV alleles were detected. The only band observed by Southern blot analysis corresponded to the parental 13 *CUP1* repeats.

In P_{GAL}-*HA* yeast, only galactose should induce expression of the *HA* tag and glucose should have no effect. In glucose, very few *de novo* CNV alleles were

detected in P_{GAL}-*HA* cells, with the 17 *HA* repeat parental band being the most prominent in all three replicates. However, the glucose-grown P_{GAL}-*HA* yeast did show a significant increase in the percentage of *de novo* P_{GAL}-*HA* alleles, compared to the respective glucose-grown BY4741 wild-type cells, *p* value = 0.0002. The minor rate of CNV in glucose-grown P_{GAL}-*HA* cells is likely due to P_{GAL}-*HA* allele being more inherently unstable than the wild-type P_{CUP1}-*CUP1* allele in BY4741 cells.

Despite the P_{GAL}-*HA* allele being more inherently unstable, galactose still caused a highly significant increase in the percentage of *de novo* CNV alleles detected in P_{GAL}-*HA* cells, *p* value <0.0001 vs. the matching glucose-grown P_{GAL}-*HA* cells. The P_{GAL}-*HA* CNV rate in galactose was so high that the parental 17 repeat band was no longer detected in the population, after as little as 10 generations. Only contractions were detected in galactose-grown P_{GAL}-*HA* cells, but *de novo* CNV alleles were observed at every copy number, ranging from 12 copies down to just a single repeat.

Glucose and galactose are both non-selective and non-mutagenic environments for BY4741 wild-type and P_{GAL}-*HA* yeasts, so all observed CNV events were free from any selection pressure. The only condition that displayed a very high level of *de novo* CNV was P_{GAL}-*HA* cells in galactose and this was also the only condition where there was transcriptional induction of the promoter. Therefore promoter induction, in the genetic context of the *CUP1* locus, was sufficient to transcriptionally stimulate CNV of the expressed gene in response to the environment. However, from this experiment it cannot be concluded if promoter induction merely stimulated local CNV, or was an essential requirement for any CNV.

3.5 Genetic knockout screen for CNV mutants

As previously explained, the native *CUP1* locus cannot be tested for transcriptionally-stimulated CNV in response to copper induction of the *CUP1* promoter, because copper is also a selective pressure. So, to identify factors involved in CNV induction at the native *CUP1* locus, genetic mutants of factors predicted to be involved in CNV were grown for 10 generations to

saturation in complete synthetic glucose and then 1:1000 diluted in the same media with and without 1mM $CuSO_4$, and grown for a further 10 generations to saturation. DNA was extracted from the mutants for Southern blot analysis of *CUP1* copy number (Fig. 3.4).

The mutants were selected based upon known CNV phenotypes in the model rDNA, or for involvement in replication or recombination. The BY4741 wild-type used in Fig. 3.4a came from a stock centre and was found to be heterogeneous for *CUP1* copy number, with two dominant parental alleles being detected at 10 and 13 repeats. Based upon this first mutant screen, BY4741 cells were single colony purified to produce a wild-type population that was homogeneous for parental 13 *CUP1* repeats. This single colony purified BY4741 population with parental 13 *CUP1* repeats was used everywhere else in this study.

In Fig. 3.4a, no *CUP1* CNV was detected in the BY4741 wild-type with or without 1mM $CuSO_4$ (lanes 1 vs. 9 and 5 vs. 13), but there was a selection towards the higher 13 *CUP1* repeat parental in the copper-treated population. However, once single colony purified, there was no CNV detected in any of the copper-treated BY4741 wild-type cells (Fig. 3.4b lane 8 and Fig. 3.4c lane 8). It was hypothesised that 1mM $CuSO_4$ was not a strong enough environmental pressure, to select for CNV events with *CUP1* amplifications in BY4741 cells that already have a high *CUP1* copy number. Other work from our lab found that copper begins to precipitate out of the media at concentrations necessary to select for *CUP1* amplifications in cells with high parental *CUP1* copy number. As such, *CUP1* CNV in strains with high parental *CUP1* copy numbers cannot be detected by simply increasing the $CuSO_4$ concentration. The issue of detecting *CUP1* CNV events in cells with high parental *CUP1* copy numbers is addressed in Chapter 4. The absence of detectable CNV at *CUP1* in copper was a problem for the mutant screen, as mutants that suppress CNV could not be detected. Nonetheless copper treatment still allowed observation of amplified alleles in hyper-recombinant mutants.

At the model rDNA locus, Sir2 silenced transcription and inhibited recombination, with *sir2* mutants displaying a 10- to 15-fold increase in mitotic and meiotic intrachromosomal recombination rates (Gottlieb and Esposito,

1989; Li, Mueller and Bryk, 2006). However, no *CUP1* CNV alleles were detected in *sir2Δ* cells with or without copper (Fig. 3.4a lanes 2 and 6). Fob1 is a nucleolar protein involved in replication fork stalling at the rDNA RFB, with reduced rDNA recombination observed in *fob1Δ* cells (Kobayashi and Horiuchi, 1996; Kobayashi, Nomura and Horiuchi, 2001). As deletion of *FOB1* was predicted to have a suppressive effect on CNV, it was unsurprising that *fob1Δ* cells displayed no novel CNV events at *CUP1* with or without copper (Fig. 3.4a lanes 3 and 7). Another nucleolar protein Csm1 binds to the rDNA NTS1 region in a Tof2- and Fob1-dependent manner, where Csm1 is required for silencing of NTS1 and suppression of rDNA recombination (Huang *et al.*, 2006). However at *CUP1*, no novel CNV events were observed with or without copper in *csm1Δ* cells (Fig. 3.4a lanes 4 and 8).

Deletion of the H3K56 acetyltransferase Rtt109, the associated histone chaperone Asf1, and the opposing H3K56ac histone deacetylases Hst3 and Hst4, all caused hyper-recombination phenotypes at the rDNA (Houseley and Tollervey, 2011; Ide, Saka and Kobayashi, 2013; Jack *et al.*, 2015). However at *CUP1*, only in the *hst3Δ hst4Δ* double deletion mutant was this hyper-recombination phenotype still observed in terms of the formation of novel *CUP1* alleles, both with and without copper (Fig. 3.4a lanes 12 and 16). Surprisingly no CNV at *CUP1*, either with or without copper, was detected in *rtt109Δ* (Fig. 3.4a lanes 10 and 14) and *asf1Δ* cells (Fig. 3.4a lanes 11 and 15). All the detected CNV events in *hst3Δ hst4Δ* cells were contractions, an explanation for which is provided in the discussion. In *hst3Δ hst4Δ* mutants the rate of CNV at *CUP1* was so great that novel CNV alleles were detected at all *CUP1* copy numbers up to the parental copy number, producing a 'laddering' phenotype of CNV alleles. In copper, *hst3Δ hst4Δ* cells still observed selection towards alleles with higher *CUP1* copy numbers, but even in selective copper, the hyper-recombination phenotype was so strong that the CNV 'laddering' phenotype remained. In copper, *hst3Δ hst4Δ* cells with higher *CUP1* copy numbers may have arisen through a transcriptionally stimulated mechanism in response to copper, analogous to the galactose-stimulated CNV in the P_{GAL}-*HA* strain (Fig. 3.3b). However, since copper is a selective environment, the *hst3Δ hst4Δ* cells with higher *CUP1* copy numbers

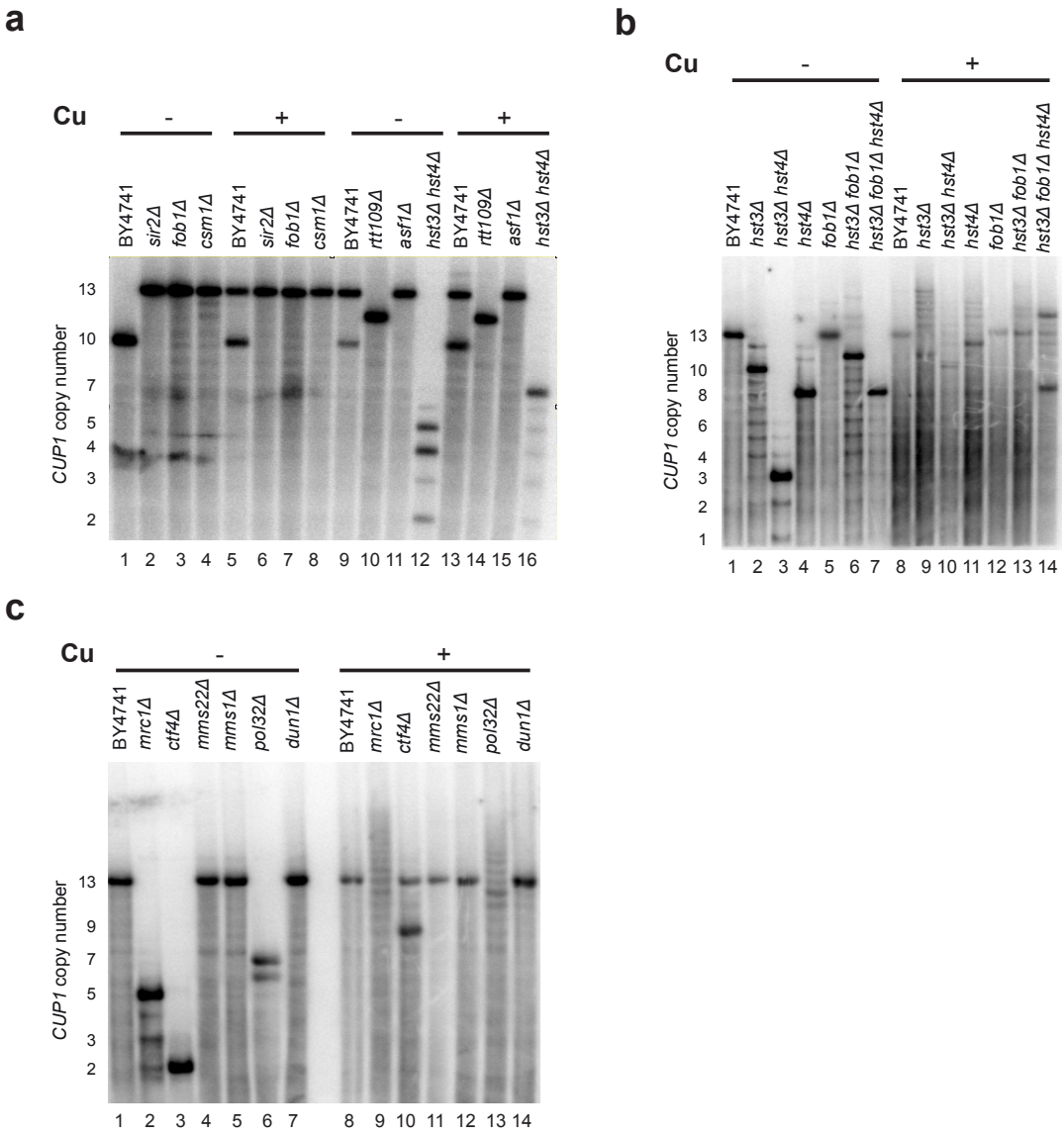


Figure 3.4: Genetic screen for CNV mutants. Southern analysis for *CUP1* copy number in BY4741 wild-type and various CNV mutants. Note parental *CUP1* copy number is 10 and 13 repeats in **a**, as it came from a mixed copy number population, prior to being single colony purified for the 13 repeat yeast used in **b** and **c**. All mutants were derived from the wild type and therefore have a parental copy number of approximately 13, even though this allele is no longer detected in highly unstable mutants. Cells were grown for 10 generations in YPD to saturation, then 1:1000 diluted in YPD with or without 1mM $CuSO_4$ and grown for a further 10 generations to saturation.

may also have pre-existed as rare cells in the starting population, which could only be detected by Southern blot analysis following selection in copper.

The hyper-recombination phenotype of *hst3Δ hst4Δ* cells on *CUP1* CNV was still present in the single *hst3Δ* (Fig. 3.4b lanes 2 and 9) and, to a lesser extent, *hst4Δ* mutants (Fig. 3.4b lanes 4 and 11), and just like the double knockout mutant, each single knockout mutant also selected towards alleles with higher *CUP1* copy numbers in copper. However, neither single mutant observed as strong a *CUP1* CNV 'laddering' phenotype as in the double mutant. In Fig. 3.4a, *fob1Δ* cells showed no CNV at *CUP1*, but deletion of *FOB1* was predicted to have a suppressive effect on recombination and as such would not have been detected in the mutant screen. To determine whether the *CUP1* CNV in the hyper-recombinant *hst3Δ* single and *hst3Δ hst4Δ* double mutants was dependent upon Fob1-meidated replication stalling at a RFB, like at the rDNA, *hst3Δ fob1Δ* and *hst3Δ fob1Δ hst4Δ* mutants were screened. Deletion of *FOB1* caused no effect on the *CUP1* CNV phenotypes of *hst3Δ* single (Fig. 3.4b lanes 6 and 13) and *hst3Δ hst4Δ* double mutants (Fig. 3.4b lanes 7 and 14), either with or without copper. Both *hst3Δ fob1Δ* and *hst3Δ fob1Δ hst4Δ* mutants also still selected towards higher *CUP1* copy number alleles in copper. Therefore *CUP1* CNV in the hyper-recombinant HDAC mutants was independent of Fob1-mediated replication fork stalling at a RFB, showing that the CNV mechanism at *CUP1* is separable from the mechanism at the model rDNA.

Fig. 3.4c screened replication protein mutants that all had previously reported CNV phenotypes at the rDNA. Mrc1 is an S-phase checkpoint protein that participates in leading-strand synthesis and activation of the DNA damage response signalling pathway upon encountering replication stress (Alcasabas *et al.*, 2001; Lou *et al.*, 2008). Mrc1 suppressed recombination at the rDNA RFB, with *mrc1Δ* cells showing de-repression of recombination initiated from the recombination-stimulating HOT1 sequence (Mohanty, Bairwa and Bastia, 2009). Deletion of *MRC1* caused rDNA CNV instability, but not hyper-amplification as observed in *rtt109Δ* or *asf1Δ* cells (Jack, 2014). The stabilising function of Mrc1 at the rDNA is likely due to the maintenance of proper checkpoint signalling, as shortening of rDNA repeats was observed in a non-phosphorylatable *mrc1-14A* checkpoint mutant in *Schizosaccharomyces pombe* (Yasuhira, 2009). Mrc1 also showed a stabilising phenotype at *CUP1*, with novel *CUP1* CNV alleles being observed

in *mrc1Δ* cells, producing another CNV 'laddering' phenotype of a hyper-recombinant strain (Fig. 3.4c lane 2). Novel CNV mutations with high *CUP1* copy numbers were still selected for in copper in *mrc1Δ* cells (Fig. 3.4c lane 9). Again the *mrc1Δ* cells in copper with high *CUP1* copy numbers may have arisen through a copper-stimulated mechanism, or pre-existed as rare cells in the initial population.

Ctf4 is a chromatin-associated protein that interacts with DNA polymerase α and couples it to the replisome progression complex through the MCM2-7 helicase (Gambus *et al.*, 2009; Tanaka *et al.*, 2009). Deletion of *CTF4* caused more frequent DSBs and end resection at arrested replication forks, resulting in hyper-amplification at the rDNA (Jack, 2014; Sasaki and Kobayashi, 2017). The parental *CUP1* copy number in *ctf4Δ* yeast was just 2 repeats. *CTF4*, like all of the mutants screened, was deleted in the BY4741 background with 13 *CUP1* repeats. As such, *ctf4Δ* cells must have undergone a major *CUP1* contraction event during transformation. In the absence of copper, *ctf4Δ* cells showed no *CUP1* CNV, but alleles with high *CUP1* copy numbers were observed in copper (Fig. 3.4c lanes 3 vs. 10). The *ctf4Δ* cells with high *CUP1* copy numbers in copper were either copper-stimulated *CUP1* amplifications or pre-existing rare cells that were selected for in copper.

Mms22 and Mms1 are subunits of the E3 ubiquitin ligase complex that is involved in stabilising the replisome at replication forks, with particular importance in prevention of fork collapse during stress (Duro *et al.*, 2008; Vaisica *et al.*, 2011). At the rDNA, deletion of *MMS22* resulted in repeat amplification, whereas deletion of *MMS1* caused a more subtle increase in rDNA instability where some, but not all cells were prone to losing a substantial amount of rDNA repeats (Jack, 2014). However at *CUP1*, neither mutation observed any effect on CNV with or without copper (Fig. 3.4c lanes 4, 5, 11 and 12).

The polymerase δ subunit Pol32, works as a processivity factor that links DNA polymerase and helicase activities, and plays an important role in mutagenic bypass repair (Gerik *et al.*, 1998; Paunesku *et al.*, 2001; Huang *et al.*, 2002). Pol32 was important for rDNA repeat expansion and *pol32Δ* cells undergo contractions at the rDNA (Houseley and Tollervey, 2011). Deletion of *POL32* also caused instability at *CUP1*, with *CUP1* contractions observed in the

absence of copper (Fig. 3.4c lane 6). This instability phenotype was maintained in copper, with novel *CUP1* CNV alleles still being detected, but at higher copy numbers than in the absence of copper (Fig. 3.4c lane 13). Again *pol32Δ* cells with high *CUP1* copy number alleles in copper may have arisen through a copper-stimulated mechanism, or were rare cells pre-existing and then selected for in copper.

Finally, Dun1 is a cell-cycle checkpoint protein kinase required for DNA damage-dependent arrest at G2/M phase, transcriptional induction of repair genes, and regulation of DNA repair pathways (Zhou and Elledge, 1993; Hammet, Pike and Heierhorst, 2002). In *dun1Δ* cells there is a reduced supply of deoxynucleotides for the replication machinery and contraction of the high-copy rDNA repeat array (Houseley and Tollervey, 2011). Also *dun1Δ* cells only allowed partial amplification of the rDNA when starting from a low-copy rDNA array (Jack, 2014). However, there was no CNV detected at *CUP1* in *dun1Δ* with or without copper (Fig. 3.4c lanes 7 vs. 14).

In summary, the mutant screen had the drawback that mutants that suppress CNV could not be detected, but despite this, CNV mutants with hyper-recombinant phenotypes were still identified. These hyper-recombinant CNV mutants were *hst3Δ* and *hst4Δ* single and double knockout cells, *mrc1Δ* and *pol32Δ* cells, and in all of these mutants there was a selection to higher *CUP1* copy numbers in copper.

3.6 Loss of the Hst3 and Hst4 histone deacetylases causes extensive *CUP1* CNV

From the screen for CNV mutants, the two HDACs Hst3 and Hst4 appeared to be the most promising genes for being involved in CNV at *CUP1*. As such, *CUP1* CNV was further explored in the Sir2-family of HDACs, which included Sir2 and the Hst3 and Hst4 proteins. BY4741 wild-type, *hst3Δ hst4Δ* double knockout, and *sir2Δ* single mutant yeast were grown in complete synthetic media for 10 generations to saturation, then 1:1000 diluted in the same media in the absence of copper and grown for a further 10 generations to saturation.

DNA was extracted for Southern blot analysis of *CUP1* copy numbers (Fig. 3.5).

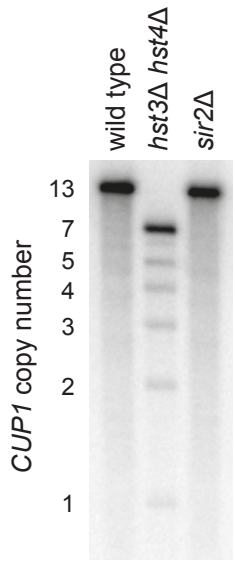


Figure 3.5: Loss of Hst3 and Hst4 destabilised the *CUP1* locus. Southern analysis for *CUP1* copy number in wild-type, *hst3*Δ *hst4*Δ, and *sir2*Δ cells. Parental *CUP1* copy number was 13 repeats for the wild-type. Note that both mutants were derived from the wild type and therefore have a parental copy number of approximately 13, even though this allele is no longer detected in the *hst3*Δ hst4Δ mutant.

Loss of *SIR2* caused no change in *CUP1* CNV, with only the parental 13 *CUP1* repeat band being detected. The absence of any CNV phenotype at *CUP1* in *sir2Δ* cells was contrary to CNV at the rDNA in *sir2Δ* cells, where mitotic and meiotic intrachromosomal recombination rates increased 10- to 15-fold (Gottlieb and Esposito, 1989). However, Sir2 primarily functions in transcriptional silencing of heterochromatin regions such as the *HML* and *HMR* mating type loci, telomeric regions and the rDNA (Bryk *et al.*, 1997; Smith and Boeke, 1997; Rusche, Kirchmaier and Rine, 2003), whereas *CUP1* is euchromatic. There is some published ChIP-seq evidence for Sir2 enrichment at ORFs of some highly expressed Pol II transcribed genes, mainly involved in fermentation, glycolysis and translation (Li *et al.*, 2013), but

nonetheless Sir2 did not have the same silencing effect at the *CUP1* locus, as at the rDNA.

In contrast, when the Sir2 homologs *HST3* and *HST4* were deleted, multiple CNV alleles were observed at the *CUP1* locus. This provided evidence for Hst3 and Hst4 protecting the stability of the *CUP1* locus, and when deleted, the locus becomes prone to recombination and CNV. Similarly, loss of *HST3* and *HST4* at the rDNA, caused a hyper-amplification phenotype (Ide, Saka and Kobayashi, 2013; Jack *et al.*, 2015), supporting a protective role for Hst3 and Hst4 in the stability of CNV loci.

3.7 Requirement for H3K56ac in stimulated CNV

The substrate for the histone deacetylases Hst3 and Hst4 is the epigenetic mark histone H3 Lysine 56 acetylation (H3K56ac); therefore an important role was hypothesised for H3K56ac in stimulated CNV at *CUP1*. Nicotinamide is an endogenous non-competitive inhibitor of the Sir2-family of HDACs (Jackson *et al.*, 2003; Sauve *et al.*, 2006, 2018; Avalos, Bever and Wolberger, 2018; Sanders *et al.*, 2018). It is also a precursor metabolite in the biosynthesis of NAD^+, an essential factor for the deacetylase activity of the sirtuins (Imai *et al.*, 2000; Ghislain, Talla and Francois, 2002). Nicotinamide can therefore regulate sirtuin activity as both a non-competitive inhibitor and as an NAD^+ precursor.

To determine whether the *CUP1* CNV phenotype of *hst3Δ hst4Δ* double knockout yeast could be pharmacologically replicated, wild-type cells were grown for 10 generations in complete synthetic media to saturation. Cells were then 1:1000 diluted in the same media with or without 5mM nicotinamide, which has no environmental selection pressure at *CUP1,* and re-grown for another 10 generations to saturation. Southern blot analysis of *CUP1* copy numbers from the saturated culture DNA showed that nicotinamide significantly stimulated CNV at *CUP1*, p value = 0.0018 vs. untreated control (Fig. 3.6a lanes 1 vs. 2). Nicotinamide-stimulated CNV also reproduced the CNV “laddering” phenotype of the hyper-recombinant *hst3Δ*

hst4Δ mutants, with novel CNV alleles being detected at every possible *CUP1* copy number, from the parental 13 repeats down to just a single copy.

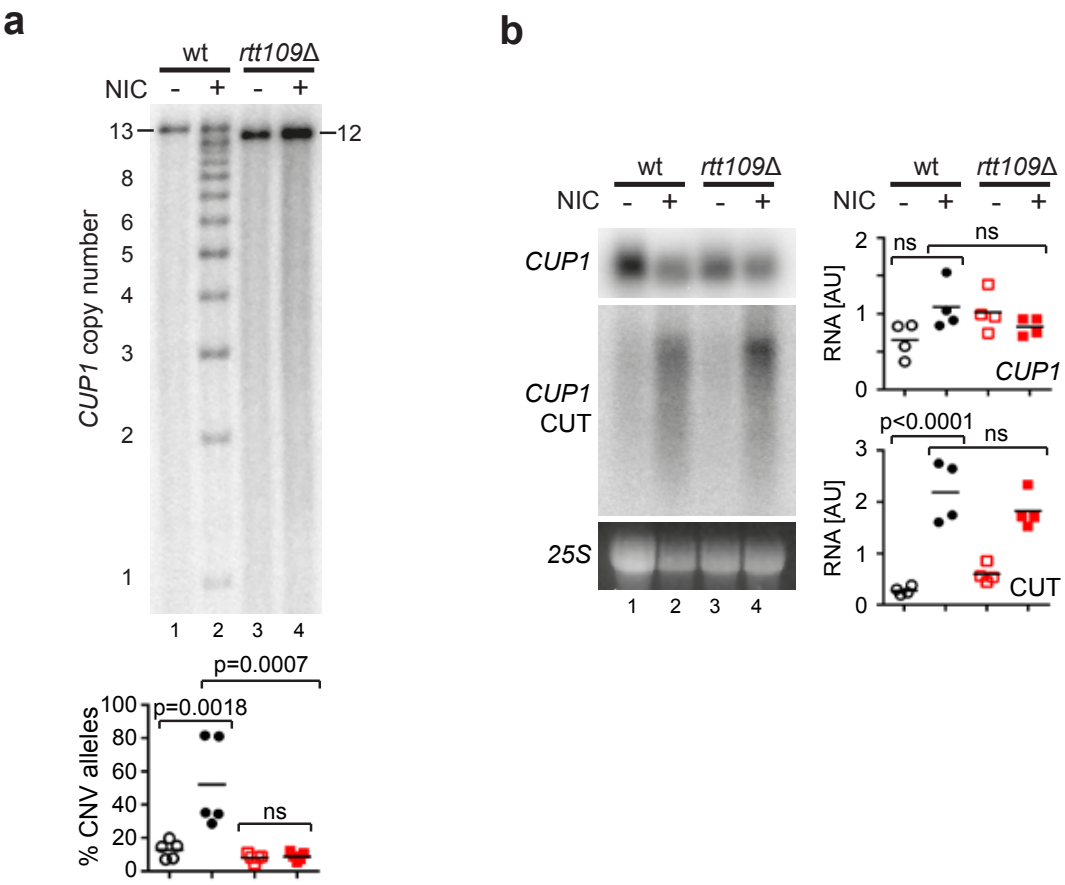


Figure 3.6: H3K56ac is required for stimulated CNV. **a**) Southern analysis of *CUP1* copy number in wild-type (wt, lanes 1 and 2) and *rtt109*Δ cells (lanes 3 and 4) treated with or without 5mM nicotinamide (NIC) for 10 generations. Parental *CUP1* copy numbers were 13 and 12 repeats for wt and *rtt109*Δ mutants respectively. Quantification shows the percentage of alleles deviating from the parental copy number, *n* = 5. *p* values calculated by 1-way ANOVA. **b**) Nicotinamide induction of the *CUP1* CUT**.** Northern analysis of *CUP1* ORF and *CUP1* cryptic unstable transcript (CUT) RNA in log-phase wild-type (wt) and *rtt109*Δ cells with or without 5mM nicotinamide (NIC). The ribosomal *25S* rRNA was included as a loading control. Quantification shows relative RNA levels in arbitrary units (AUs), *n* = 4. *p* values were calculated by 1-way ANOVA. ns = not significant.

To show that nicotinamide-stimulated CNV was acting through repression of the Hst3 and Hst4 deacetylase activity on H3K56ac, *rtt109Δ* yeast that cannot acetylate H3K56 were also grown, as above, with and without 5mM nicotinamide for Southern blot analysis (Fig. 3.6a lanes 3 vs. 4). The *rtt109Δ* cells had a parental *CUP1* copy number of 12 repeats. Nicotinamide-stimulated CNV at *CUP1* was significantly repressed in *rtt109Δ* yeast, *p* value = 0.0007, compared to nicotinamide-treated wild-type cells. Nicotinamide-stimulated CNV in the *rtt109Δ* mutant was completely suppressed to the level of CNV in untreated cells, with no *CUP1* CNV alleles being observed with or without nicotinamide treatment. Therefore, loss of Rtt109 rendered the *CUP1* locus immune to the CNV-inducing effect of nicotinamide, and provided evidence for H3K56ac being essential for stimulated CNV.

H3K56ac has been implicated in the regulation of gene expression, where it becomes preferentially enriched around active genes and promotes the recruitment of the SWI/SNF nucleosome remodelling complex and subsequent transcription (Xu, Zhang and Grunstein, 2005). To determine whether H3K56ac was essential for stimulated CNV because of its role in gene expression, our lab grew wild-type and *rtt109Δ* yeast to mid-log, with or without 5mM nicotinamide, for Northern blot analysis of the *CUP1* sense and antisense CUT transcripts (Fig. 3.6b).

In the absence of nicotinamide, the *CUP1* promoter in wild-type and *rtt109Δ* yeast showed very strong directional bias towards production of the *CUP1* sense transcript, over the anti-sense CUT (Fig. 3.6b lanes 1 and 3). This promoter directional bias was lost with nicotinamide treatment in wild-type cells (Fig. 3.6b lane 2), and would have provided supporting evidence for a regulatory role of H3K56ac in *CUP1* gene expression, had nicotinamide treatment not also caused an equivalent loss of promoter directionality in *rtt109Δ* yeast (Fig. 3.6b lane 4), which cannot acetylate H3K56. Therefore loss of promoter directionality was likely a H3K56ac-independent effect of nicotinamide treatment. As such, it was highly unlikely that H3K56ac controlled stimulated CNV through its regulation of promoter induction.

3.8 Replication Fork mutants undergo CNV without stimulation

H3K56ac is also important in replication fork stability and restart, where it perturbs replisomes, activates the DNA damage checkpoint response, and induces hyper-recombination (Han, Zhou, Li, *et al.*, 2007; Thaminy *et al.*, 2007; Celic, Verreault and Boeke, 2008). Furthermore, H3K56ac has recently been shown to inhibit extensive repair synthesis in BIR events, and to interfere with efficient fork progression, resulting in elevated recombination (Che *et al.*, 2015). To determine whether H3K56ac regulated stimulated CNV through its effects on replication fork stalling and restart, *CUP1* CNV was explored in *mrc1Δ* and *pol32Δ* cells, two replication proteins that showed promising CNV phenotypes from the mutant screen (Fig. 3.4c). Other work from our group had already revealed that there was no change in *CUP1* sense or antisense transcription in *mrc1*Δ and *pol32*Δ mutants (Hull *et al.*, 2017). The checkpoint regulatory protein Mrc1 forms part of the replisome and helps stabilise stalled replication forks, and acts as an anchor for any subsequent DNA repair events (Katou *et al.*, 2003). Pol32 encodes a non-essential subunit of polymerase δ that is needed for efficient DNA synthesis following BIR events initiated at broken replication forks (Deem *et al.*, 2008; Smith, Lam and Symington, 2009).

Wild-type, *mrc1Δ* and *pol32Δ* yeasts were grown for 10 generations to saturation in complete synthetic media for Southern blot analysis of *CUP1* copy number (Fig. 3.7). A significantly increased level of CNV at the *CUP1* locus was observed in both *mrc1*Δ, *p* value <0.0001, and *pol32*Δ cells, *p* value = 0.034. The level of CNV was so high that a CNV 'laddering' phenotype was observed in both *mrc1*Δ and *pol32*Δ strains, with novel CNV alleles detected at every *CUP1* repeat up to the parental copy number. Therefore these replication fork mutants successfully de-coupled CNV from the requirement of stimulation by copper or nicotinamide.

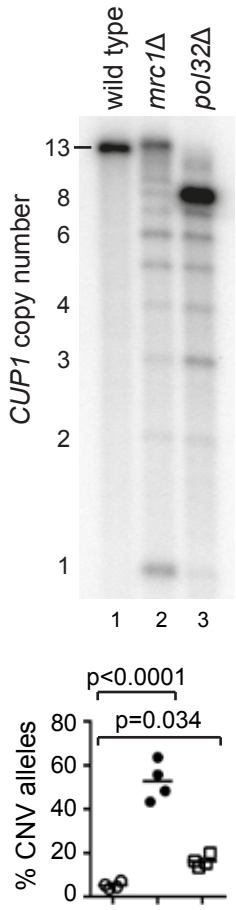


Figure 3.7: Transcriptionally uncoupled CNV in replication fork mutants. Southern analysis of wild-type (wt), *mrc1*Δ and *pol32*Δ cells grown for 10 generations in complete synthetic media. The wild-type had a parental *CUP1* copy number of 13. Note that both mutants were derived from the wild type and therefore have a parental copy number of approximately 13, even though this allele is no longer detected in the *pol32*Δ mutant. Quantification showed the percentage of alleles deviating from the parental copy number, n = 4. p values calculated by 1-way ANOVA.

For statistical analysis of the percentage CNV in *pol32*Δ cells, the conservative assumption was made that the parental copy number was the strongest band, representing 8 *CUP1* copies, with the fainter bands at higher copy numbers being amplifications. However, the true parental band in *pol32*Δ cells might have been 13 repeats, like the wild-type it was derived from. If so, the rate of CNV was so high in *pol32*Δ yeast, that the parental band was no longer detected by Southern blot analysis, and the true rate of CNV in *pol32*Δ cells would be much higher.

The ability of these replication fork stability and repair mutants to stimulate CNV in a similar manner to the H3K56ac HDAC mutants suggests that H3K56ac may affect CNV by interfering with replication fork stability or repair. Therefore raising global H3K56ac levels, by nicotinamide inhibition of the Hst3 and Hst4 HDACs, most likely interferes with stalled replication forks and/or their repair, whereby increasing the rate of recombination and CNV.

3.9 Nicotinamide cannot induce CNV at silent promoters

To support the finding that nicotinamide-stimulated CNV at the native *CUP1* locus (Fig. 3.6a), the effect of nicotinamide-stimulation was also explored in P_{GAL}-*HA* cells. P_{GAL}-*HA* yeast were grown to saturation in complete synthetic 2% glucose media, then 1:1000 diluted in complete synthetic media containing 2% glucose or 2% galactose, both with and without 5mM nicotinamide, and given a further 10 generations to grow to saturation. DNA was extracted from the saturated cultures for Southern blot analysis of *HA* copy number (Fig. 3.8).

As shown earlier in Fig. 3.3b, P_{GAL}-*HA* was again only able to transcriptionally stimulate *de novo* CNV events in galactose (Fig. 3.8 lanes 1 vs. 3), *p* value = 0.0006. However, surprisingly nicotinamide was unable to significantly stimulate CNV in glucose (Fig. 3.8 lanes 1 vs. 2), nor enhance CNV in galactose (Fig. 3.8 lanes 3 vs. 4). This was contrary to the effect of nicotinamide on the native *CUP1* locus (Fig. 3.6a), and suggested that increased global H3K56ac was not sufficient to stimulate CNV by itself.

Our lab searched for differences between the wild-type P_{CUP1}-*CUP1* and re-engineered P_{GAL}-*HA* systems, which could potentially explain the different effect of nicotinamide on the two strains. By Northern blot analysis of expressed transcripts at the *CUP1* locus in P_{GAL}-*HA* yeast, our lab found the P_{GAL} promoter to be completely repressed in glucose, independent of nicotinamide treatment (Hull *et al.*, 2017). By comparison, a high basal level of *CUP1* transcript expression in glucose was observed at the wild-type P_{CUP1} promoter, which was again independent of nicotinamide (Fig. 3.6b).

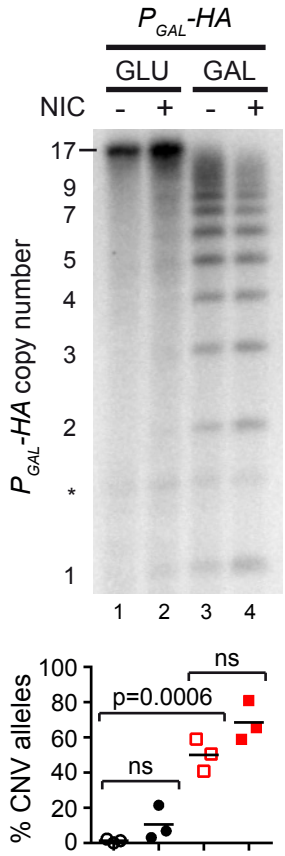


Figure 3.8: Nicotinamide does not enhance CNV in *P_{GAL}-HA* yeast. Southern analysis of *P_{GAL}-HA* copy number in *P_{GAL}-HA* cells grown for 10 generations in glucose (GLU) and galactose (GAL), with or without 5mM nicotinamide (NIC). Quantification showed the percentage of alleles deviating from the parental copy number, *n* = 3. * = non-specific band.

This result provided evidence for promoter induction being essential for stimulated CNV. However, our lab also found by Northern blot analysis that nicotinamide caused no further induction, nor any change in promoter directionality in *P_{GAL}-HA* yeast grown in 2% galactose, where the promoter was highly induced (Hull *et al.*, 2017). As H3K56ac becomes preferentially enriched at active genes (Xu, Zhang and Grunstein, 2005), it was hypothesised that growth in 2% galactose induced the P_{GAL} promoter enough to recruit a saturating local level of H3K56ac for maximum CNV at *P_{GAL}-HA*. As such, nicotinamide treatment would be unable to cause any further increase in stimulated CNV, because H3K56ac would already be saturated at the locus.

3.10 Stimulated CNV requires promoter induction and H3K56ac

To determine whether nicotinamide was unable to enhance the transcriptionally stimulated CNV rate in 2% galactose because of a saturated local level of H3K56ac, an experiment was devised whereby the P_{GAL} promoter in P_{GAL}-*HA* cells was induced to different extents by varying the concentration of galactose. If the level of promoter induction determines the local level of H3K56ac, lowering the concentration of galactose should decrease the level of P_{GAL} promoter induction and also decrease the local level of H3K56ac to a level at which nicotinamide gains an additive effect. P_{GAL}-*HA* cells were grown for 10 generations to saturation in complete synthetic 2% sucrose 2% raffinose media, then 1:1000 diluted in various galactose concentrations, along with 2% raffinose to support growth, with and without 5mM nicotinamide, for Southern blot analysis of copy number (Fig. 3.9). The galactose concentrations used for variable induction of the P_{GAL} promoter were obtained from the Tollervey group (Houseley *et al.*, 2008). When there was no galactose, and therefore complete repression of the P_{GAL} promoter, or 0.01% galactose with 0.2% glucose to repress the P_{GAL} promoter, almost no CNV events were detected with or without nicotinamide (Fig. 3.9 lanes 9-12). In the absence of glucose, when the P_{GAL} promoter was partially induced with 0.01%, 0.05%, 0.2%, and 1% galactose, *de novo* CNV events were detected with and without nicotinamide (Fig. 3.9 lanes 1-8). However, the addition of nicotinamide to all of these low-galactose concentrations significantly enhanced the stimulated CNV rate, see Fig. 3.9 for *p* values. There was also a slight increase in the percentage of *de novo* CNV events detected in the absence of nicotinamide with increasing galactose concentration, but not when nicotinamide was added. This indicated that nicotinamide-treated cells saturated the local H3K56ac level at partially induced promoters.

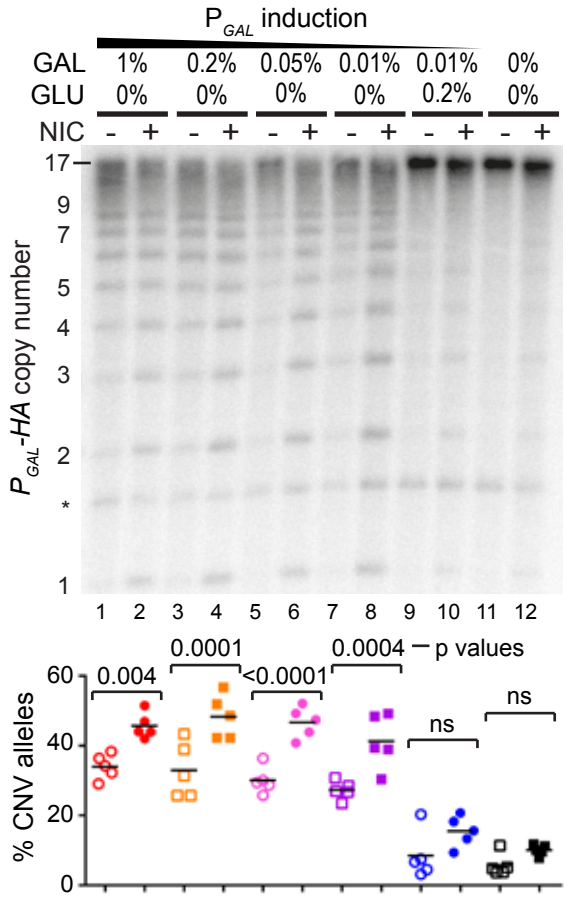


Figure 3.9: Combinatorial action of promoter activity and nicotinamide on CNV. Southern blot analysis of *P_{GAL}-HA* cells grown for 10 generations to saturation in complete synthetic 2% sucrose 2% raffinose media, then 1:1000 diluted in given concentrations of galactose (Gal) and glucose (Glu), combined with 2% raffinose, with or without 5mM nicotinamide (NIC). *p* values were calculated from pairwise comparisons of samples with or without NIC for each GLU or GAL concentration, deriving from a 1-way ANOVA of the whole data set, *n* = 5. * = non-specific band, ns = not significant.

This experiment provided evidence for the level of promoter induction controlling the local level of H3K56ac. It also provided evidence that nicotinamide can still enhance transcriptionally stimulated CNV, most likely by raising global H3K56ac levels, so long as the promoter was induced and the local level of H3K56ac was not already saturated. As such, there is a combined requirement of promoter induction and high local levels of H3K56ac for stimulated CNV in yeast.

3.11 H3K56ac is essential for transcriptionally stimulated CNV in *P_{GAL}-HA*

To determine whether H3K56ac was still essential for transcriptionally stimulated CNV in *P_{GAL}-HA* yeast, and not just important for enhancing the CNV rate, a *rtt109Δ* mutation was introduced into *P_{GAL}-HA*. Wild-type and *rtt109Δ P_{GAL}-HA* cells were then grown to saturation for 10 generations in complete synthetic 2% glucose media, then 1:1000 diluted in complete synthetic 2% glucose or 2% galactose media and grown for a further 10 generations to saturation. Culture DNA was extracted for Southern blot analysis of *P_{GAL}-HA* copy number (Fig. 3.10).

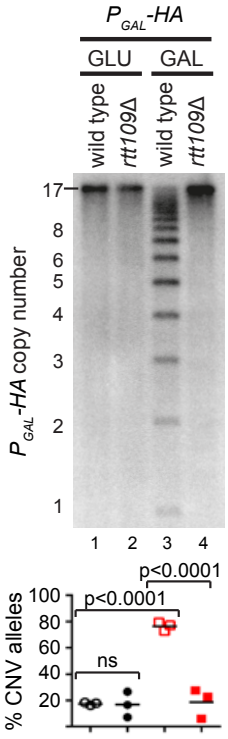


Figure 3.10: H3K56ac was required for transcriptionally stimulated CNV in *P_{GAL}-HA* cells. Southern analysis of *HA* copy number in *P_{GAL}-HA* wild-type and *rtt109*Δ cells grown to saturation for 10 generations in complete synthetic 2% glucose, then 1:1000 diluted in glucose (Glu) or galactose (Gal) and grown for another 10 generations to saturation. Parental *P_{GAL}-HA* and *P_{GAL}-HA rtt109*Δ copy numbers were 17 repeats. Quantification showed the percentage of alleles deviating from the parental copy number, *n* = 3. *p* values calculated by 1-way ANOVA. ns = not significant.

In glucose, both the P_{GAL}-*HA* wild-type and *rtt109Δ* mutant showed no CNV (Fig. 3.10 lanes 1 and 2). Therefore deletion of *RTT109* did not interfere with stimulated CNV in the absence of promoter induction. However, when the P_{GAL} promoter was maximally induced in 2% galactose, the *rtt109Δ* mutant reduced any transcriptionally stimulated CNV events to below the threshold for detection (Fig. 3.10 lanes 3 vs. 4), *p* value <0.0001. Therefore promoter induction and H3K56ac were still essential for transcriptionally stimulated CNV in P_{GAL}-*HA* cells.

3.12 Summary

Our predicted model for genes that can undergo stimulated CNV in response to the environment required an upstream RFS site, expression from a bi-directional promoter, and repressed expression under optimal growth conditions, but becoming highly induced in response to a specific environmental stimulus. The copper-resistance *CUP1* gene had all these prerequisites for it to be able to undergo environmentally-stimulated CNV. Using the genetically re-engineered P_{GAL}-*HA* strain, I first showed that promoter induction could stimulate *de novo* CNV in the genetic context of the *CUP1* locus.

Since direct determination of *CUP1* CNV in response to copper was not possible due to copper selection towards alleles with higher *CUP1* copy numbers, I instead identified a selection of gene knockout mutations that interfered with CNV at the native *CUP1* locus. Deletion of *HST3*, *HST4*, *MRC1*, and *POL32* genes caused extensive *CUP1* CNV, with hyper-recombination phenotypes, that were independent of environmental stimulation.

The H3K56ac histone deacetylases Hst3 and Hst4 were shown to be required for stabilising the *CUP1* locus against stimulated CNV, by decreasing global H3K56ac levels. I confirmed this by: 1) inhibiting Hst3 and Hst4 with the HDAC inhibitor nicotinamide, which has no environmental selection pressure at *CUP1*, but still stimulated extensive *CUP1* CNV; and 2) reducing

nicotinamide-stimulated CNV at *CUP1*, by deleting the opposing H3K56 acetyltransferase gene *RTT109*.

Using the hyper-recombinant replication protein mutants *mrc1Δ* and *pol32Δ*, I provided evidence for H3K56ac stimulating *de novo* CNV events through its function(s) in replication fork stability and/or DNA repair, and not through its regulation of promoter induction and directionality. Finally, using the P_{GAL}-*HA* strain, I provided evidence that stimulated CNV can only occur at genes with active promoters and with the combined requirement for high local levels of H3K56ac.

4. Stimulated CNV accelerates adaptation to copper in high- and low-copy *CUP1* systems

4.1 Introduction

Many human disorders have been associated with CNV of protein coding genes (Craddock *et al.*, 2010; Stankiewicz and Lupski, 2010), with some specific genetic syndromes being directly attributed to changes in copy number (van der Maarel, Tawil and Tapscott, 2011). In cancer, amplification of the driving oncogene is one mechanism cells have been shown to use in order to develop resistance to chemotherapeutic drugs (Frei *et al.*, 1984; Corcoran *et al.*, 2010; Little *et al.*, 2011). Therefore the importance of copy number in environmental adaptation leading to disease is already well established. In Chapter 3, I showed that CNV could be transcriptionally stimulated in the genetic context of *CUP1* in response to the environment, so long as the promoter was induced and there was a high local level of H3K56ac around the CNV gene. As such, environmentally-stimulated CNV could have major implications in how disorders associated with CNV and environmental adaptation are considered.

As mentioned in Chapter 3, detecting CNV at the native *CUP1* locus has two major complications. The first is that copper provides a selective pressure towards the emergence of alleles with high *CUP1* copy numbers, making the distinction between random pre-existing and adaptive mutations impossible under normal culturing conditions. The second issue is that copper begins to precipitate out of the media at concentrations necessary to select for *CUP1* amplifications in cells with high parental *CUP1* copy number. In this Chapter, I will introduce the Cell Fate Analysis strategy, which utilises the Mother Enrichment Program (MEP) devised by the Gottschling lab (Lindstrom and Gottschling, 2009), as a novel method to detect stimulated CNV events in response to the environment, by using a defined cohort of replicatively aged

yeast. Using these replicatively aged yeasts, I will explore whether cells can accelerate their adaptation to a challenging copper environment through copper-stimulated CNV at *CUP1*, and determine the importance of promoter induction and H3K56ac in the process.
I will also introduce the 3x*CUP1* system to determine whether stimulated CNV functions the same in low- and high-copy *CUP1* systems. High-copy *CUP1* systems are not uncommon for yeast from the S288C background, and are by no means exceptional compared to *CUP1* copy numbers of wild isolates (Zhao *et al.*, 2014; Strope *et al.*, 2015). Nonetheless, most genes in the yeast genome are present at low-copy, therefore establishing whether our predicted model for stimulated CNV also functions at low-copy is important for the translatability of the mechanism.

4.2 Copper stimulates *de novo* CNV at *CUP1*

It is not possible to test whether environmental copper can directly stimulate *de novo* CNV events at the native *CUP1* locus under normal culturing conditions, in the same way as the P_{GAL}-*HA* experiment (Fig. 3.3b), because of the aforementioned issues with growth in a selective copper environment. Therefore, to determine whether the native P_{CUP1}-*CUP1* locus also undergoes stimulated CNV in response to environmental copper, the issue of copper selection needed to be removed, without interfering with the natural copper responsiveness of the locus. The Mother Enrichment Program (MEP) dealt with the copper selection issue by preventing the progeny of a defined cohort of cells from replicating during growth in the presence of β-estradiol (Lindstrom and Gottschling, 2009). Therefore the MEP system allowed direct observation of the treatment on a defined cohort of cells, independent of the reproductive success of their progeny, and therefore also free from environmental selection.
The MEP, built by the Gottschling lab, uses a fusion protein, Cre-EBD78, of the bacteriophage P1 Cre recombinase and the oestrogen-binding domain of the murine oestrogen receptor that is strictly dependent upon β-estradiol for its activity. This fusion protein was put under the control of the daughter-

specific P_{SCW11} promoter, which is solely expressed in newborn cells, and integrated at the *HO* locus. P_{SCW11}-*cre-EBD78* expression is therefore restricted to the G1 phase of daughter cells. The Cre recombinase controls two *loxP* flanked target genes, *UBC9* and *CDC20*. Both *UBC9* and *CDC20* protein products are required for the degradation of mitotic cyclins and other targets vital to cell cycle progression (Sethi *et al.*, 1991; Seufert, Futcher and Jentsch, 1995; Dieckhoff *et al.*, 2004). Upon activation, the recombinase activity deletes 92% of the coding region of *UBC9* and changes the start codon of the *CDC20* gene so that it now produces a non-functional protein. In the absence of β-estradiol, daughter cells express the Cre-EBD78 fusion protein, but it is sequestered in the cytoplasm and the daughters survive. However, upon ligand binding, the fusion protein is translocated into the nucleus, where it acts on its *loxP* target sites and causes the selective permanent arrest of all daughter cells in M-phase (Lindstrom and Gottschling, 2009).

In our experiments, the MEP wild-type yeasts are diploid and maintain their *CUP1* copy number at 13 and 14 repeats on its two alleles. The MEP cells were initially pre-cultured in complete synthetic media to early log phase and then $0.25x10^7$ pre-culture cells were inoculated into two 125ml complete synthetic cultures containing β-estradiol (Fig. 4.1a). After 2 hours growth, 1mM $CuSO_4$ was added to one of the two 125ml cultures and the other one was left untreated. Both cultures were aged for 24 hours, during which time β-estradiol ensured all daughter cells produced were unable to replicate. At the end of the experimental time period, a sample of each culture condition was plated on a non-selective complete synthetic plate in the absence of β-estradiol. Progenitor cells from the initial defined cohort, which were still viable after 24 hours, gave rise to genetically identical colonies that were then grown up for DNA extraction and Southern blot analysis of *CUP1* copy number (Fig. 4.1b). The defined cohort of cells allowed for direct observation of the effect of copper treatment on *CUP1* CNV, independent of the reproductive success of the progeny and therefore without any selection bias.

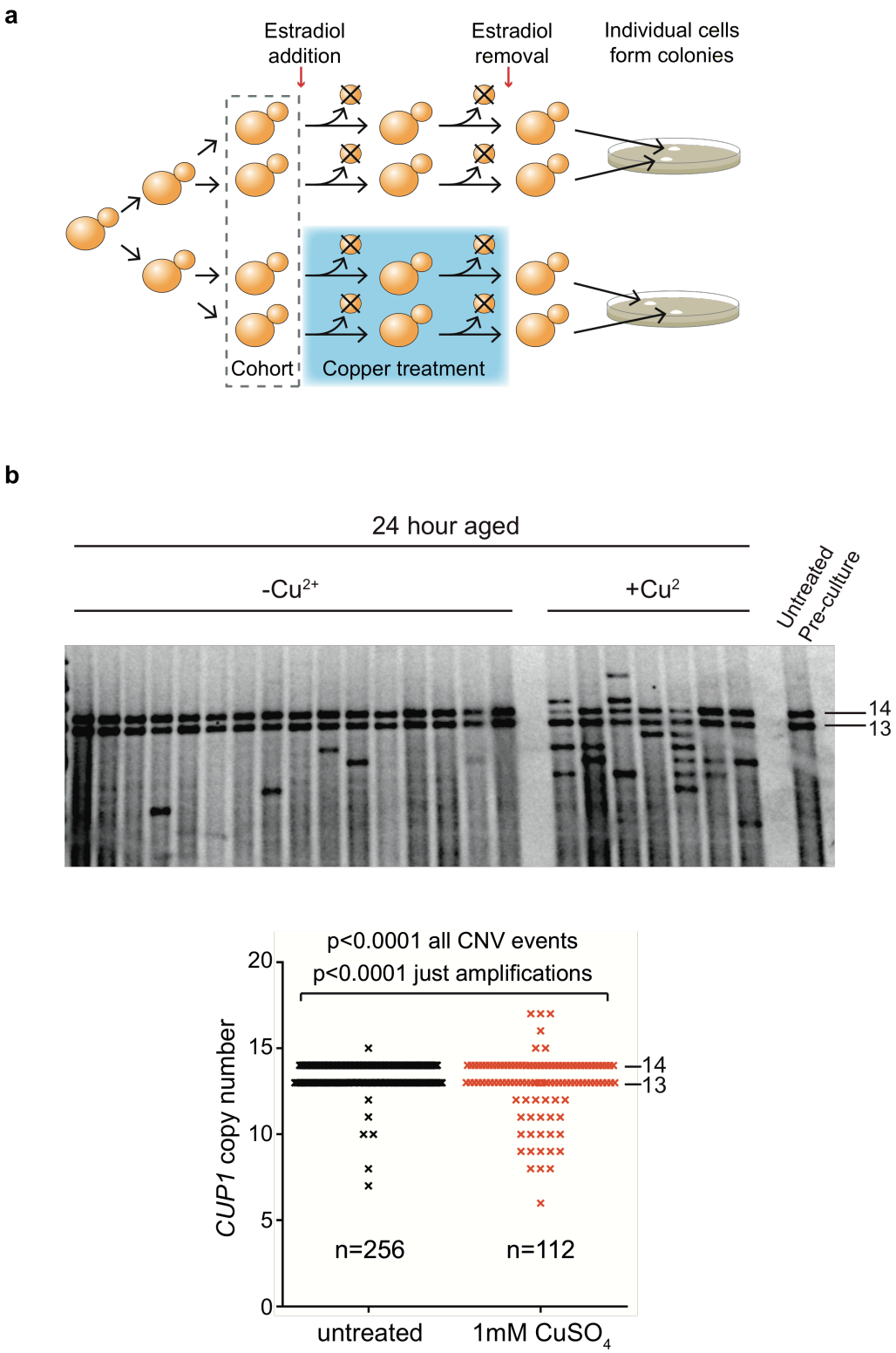


Figure 4.1: Stimulated CNV in copper treated yeast. **a**) Schematic of the Cell Fate Analysis strategy in MEP wild-type cells. An initial pre-culture was inoculated with a single colony of MEP cells and grown to early log phase. $0.25x10^7$ pre-culture cells were inoculated into 125ml cultures containing β-estradiol and grown for 2 hours. No copper (top pathway) or 1mM $CuSO_4$ (bottom pathway) was then added to the cultures and aged for 24 hours. The β-estradiol was then removed and a sample of each culture condition was plated on non-

selective media. Each viable progenitor cell gave rise to a colony of genetically identical yeast. **b**) PFGE-Southern blot and analysis of *CUP1* copy number alleles in colonies derived from 128 untreated and 56 1mM $CuSO_4$ treated diploid MEP cells after 24 hours, pooled from 2 experiments. On average, 89% of the starting cohort of cells was recovered in the untreated condition, and 40% was recovered in the 1mM $CuSO_4$ treatment. Experimentally observed mutation rates in the untreated cohort were normalised for the viability in the treated cohort, to produce the expected mutation rates for the treated condition. This included making the conservative assumption that cells lost during the experiment did not favour either yeast that do or do not undergo CNV. *p*-values were calculated by a goodness of fit χ^2 test with 1 degree of freedom between the observed and expected number of mutant to wild-type alleles across the treated cohort.

In the absence of copper, when there should be no transcriptional stimulation of the *CUP1* locus, just 7 out of the 256 alleles (128 clones) screened showed novel *CUP1* CNV that deviated from the parental 13 and 14 *CUP1* repeats on its two diploid alleles. This was a baseline observed mutation rate of 0.027 CNV events per allele. This mutation rate is the combined influence of all the CNV mutants that were pre-existing in the initial cohort of cells, plus any CNV mutants that arose *de novo* during the course of the experiment. The individual contributions to the total CNV rate of pre-existing and *de novo* CNV mutations cannot be distinguished.

In comparison, the 1mM $CuSO_4$ treated condition detected 31 alleles with novel *CUP1* CNV mutations, out of 112 screened alleles (224 clones). This was an observed mutation rate of 0.277 CNV events per allele, a 10-fold increase on the untreated condition. By random variation, a similar number of pre-existing CNV mutants would be expected in each starting cohort of cells, independent of the treatment. Therefore the increase in novel *CUP1* CNV alleles in copper treatment must be due to an increase in the number of *de novo* CNV events at *CUP1* during the 24-hour aging period.

To determine whether the copper-stimulated increase in *CUP1* CNV was statistically significant, differences in culture viability after 24 hours, across the respective treatments, needed to be accounted for. The untreated condition had a cell viability of 82-97% and the 1mM $CuSO_4$ treated condition had a viability of 36-44%. The conservative assumption was made that all non-

viable cells showed no bias between yeast that underwent *CUP1* CNV and yeast with unchanged CNV. Using the observed CNV rate in the untreated condition, and taking the different viabilities of the two conditions into account, a value was calculated for the expected number of *CUP1* CNV mutations in copper-treated cells if copper was not affecting the CNV rate. The expected number of *CUP1* CNV events in copper was then directly compared to the experimentally observed number of *CUP1* CNV events in copper, using a goodness of fit χ^2 test with 1 degree of freedom. The χ^2 test supported the conclusion that copper significantly increased the number of *de novo* CNV events at *CUP1*, well above the normal baseline rate, *p* value <0.0001.
In the same way used above for total CNV events, copper-stimulated *CUP1* amplifications were also tested for statistical significance. Just as for the total CNV rate, copper-treated cells also significantly increased the number of *de novo CUP1* amplifications, well above the normal baseline rate, *p* value <0.0001.
These calculated rates for copper-stimulated CNV are likely to even be underestimates. This is because exposure to 1mM $CuSO_4$, over the 24-hour experimental period, slowed the rate of replication. Copper-treated cells only underwent 8±3 divisions, compared to 12±2 divisions in untreated cells, as determined by budscar counting. Since CNV is currently considered to require DNA replication, copper-treated cells that have a reduced rate of replication will have had fewer occasions to undergo a CNV event at *CUP1*. As such, the Cell Fate Analysis experiment provided evidence that environmental copper stimulated the formation of *de novo* CNV and amplification events at the *CUP1* locus.

4.3 Stimulated *CUP1* CNV can accelerate adaptation to environmental copper

Having shown that *de novo* CNV events were stimulated at the *CUP1* locus in response to environmental copper, it was important to next determine whether cells could use copper-stimulated CNV to accelerate their adaptation to a challenging copper environment. Three single cell-derived lines from the MEP

Cell Fate Analysis experiment in 1mM $CuSO_4$ (Fig. 4.1b), which had: 1) gained 3 *CUP1* repeats on one allele; 2) maintained the wild-type 13 and 14 repeats on each allele; or 3) lost 7 repeats on one of its alleles, were put through a copper adaptation assay. The principle of the adaptation assay was to expose the yeast to various high concentrations of $CuSO_4$ for 3 days and then measure the optical density (OD) reading for the culture. The OD_{600nm} was a proxy readout for the ability of the culture to adapt and grow in that concentration of copper. As all three single-cell derived lines had a very high starting number of *CUP1* repeats, very high concentrations of $CuSO_4$ would be required to separate the copper-adaptive capabilities of the three lines. However, copper starts to precipitate out of the media at the concentrations necessary to separate the adaptation phenotypes of the three single-cell derived lines. To circumvent this issue, 0.5mM ascorbic acid was added to the assays. Ascorbate increases the cellular uptake of copper, thus raising the toxicity of environmental copper to the yeast. Therefore the added ascorbate enabled clear separation of the copper-adaptation profiles of the three single-cell derived lines, using $CuSO_4$ concentrations below the level at which copper begins to precipitate (Hassett and Kosman, 1995).

All three single-cell derived lines effectively grew to saturation in up to 1.5mM $CuSO_4$, and were all completely sensitive to copper concentrations above 2.25mM (Fig. 4.2). The copper-adaptation profiles of the three single-cell derived lines were only separated in the narrow concentration range of 1.5mM to 2.25mM $CuSO_4$, but nonetheless, a clear adaptive difference to copper was still observed between the three lines. The single-cell derived line that had gained 3 *CUP1* repeats (red) outperformed both of the other two lines with lower *CUP1* copy numbers. This adaptive advantage in copper was highly significant, p value = 0.005, compared to the parental *CUP1* copy number line (black). Equally, the line that lost 7 *CUP1* repeats (blue) showed a significantly decreased environmental fitness in copper, compared to the parental line, p value <0.0001. Given that these variable *CUP1* copy number alleles arose *de novo* during exposure to environmental copper (Fig. 4.1b), and provided quantifiable adaptive differences to a challenging copper environment (Fig. 4.2), the results provided evidence that cells can accelerate

their adaptation to hostile environments through environmentally-stimulated CNV.

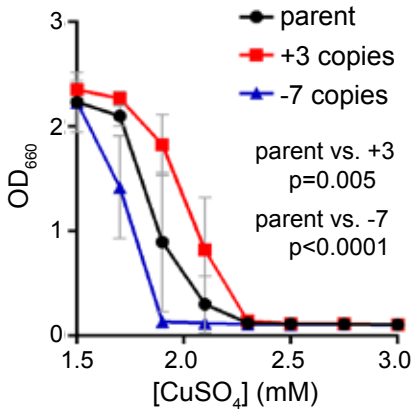


Figure 4.2: Copper adaptation of variable *CUP1* copy number strains. A copper adaptation assay of three MEP lines recovered from Fig. 4.1b with the parental 13 and 14 *CUP1* copy numbers on each allele (parent, black), +3 repeats on one allele (+3 copies, red), and –7 repeats on one allele (-7 copies, blue). $CuSO_4$ was added at indicated concentrations to media containing 0.5mM ascorbic acid to increase copper toxicity. OD_{660nm} was measured after 3 days at 30°C. Error bars represent ±1 SD; *p*-values were calculated by 1-way ANOVA of area under curves; $n = 6$ for each line.

4.4 Stimulated CNV is highly selective for the transcriptionally induced allele

So far environmental copper has been shown to stimulate *de novo* CNV events at *CUP1*, with the novel CNV alleles producing quantifiable differences in copper-fitness with the potential to accelerate a cell's adaptation to a hostile copper environment (Fig. 4.1b and 4.2). Also using the P_{GAL}-*HA* system, it was shown in Chapter 3 that *de novo* CNV was stimulated by promoter induction (Fig. 3.3b). Whether copper-stimulated CNV at *CUP1* was also a transcriptional effect still needed to be determined. One caveat of the untreated vs. copper-treated MEP experiment (Fig. 4.1b) was that copper

itself can cause genetic mutations, having been shown to induce genetic recombination and/or gene conversion events in *Saccharomyces cerevisiae* (Sakai and Takahashi, 1972; Zimmermann *et al.*, 1984). Therefore the copper-stimulated increase in *de novo* CNV could be due to induction of the *CUP1* promoter or increased mutagenesis.

In order to show that copper-stimulated *CUP1* CNV was a transcriptional response, the MEP P_{GAL}-*HA* heterozygous diploid strain was created. This strain has one wild-type copper-responsive P_{CUP1}-*CUP1* allele and one galactose-responsive P_{GAL}-*HA* allele. Using the cell fate analysis system described above (Fig. 4.1a), the heterozygote was aged for 24 hours in the same control glucose condition used before, but now also in P_{GAL} promoter-inducing galactose, instead of 1mM $CuSO_4$. Neither glucose nor galactose treatment should have any mutagenic effect on the yeast, thus enabling direct determination of the importance of transcription in stimulating CNV at *CUP1*. Extracted DNA was run on a PFGE as before and Southern blotted. The membrane was then probed first for the P_{GAL}-*HA* allele (Fig. 4.3a) and then separately for the P_{CUP1}-*CUP1* allele (Fig. 4.3b). Since the probe for the P_{GAL}-*HA* allele did not strip off the membrane very well prior to probing for the P_{CUP1}-*CUP1* allele, most of the bands detected in Fig. 4.3a were still observed in Fig. 4.3b. Therefore the true P_{CUP1}-*CUP1* alleles in Fig. 4.3b were just the novel bands observed in Fig. 4.3b that were not observed in Fig. 4.3a.

When probed for the P_{GAL}-*HA* allele, just 5 out of the 56 alleles screened showed CNV mutations in glucose (Fig. 4.3a). This was a mutation rate of 0.089 CNV events per allele and was a little higher than, but still largely comparable to, the 0.027 CNV events per allele observed in the untreated MEP diploid (Fig. 4.1b). As seen earlier in the P_{GAL}-*HA* haploid grown in glucose (Fig. 3.3b), the P_{GAL}-*HA* allele was more inherently unstable than the P_{CUP1}-*CUP1* allele, which could explain the slight increase in CNV rate observed in the MEP P_{GAL}-*HA* heterozygote in glucose. As before, this CNV rate is the sum of all novel CNV alleles that were pre-existing in the starting cohort of cells, plus any *de novo* CNV alleles formed during the 24-hour aging period.

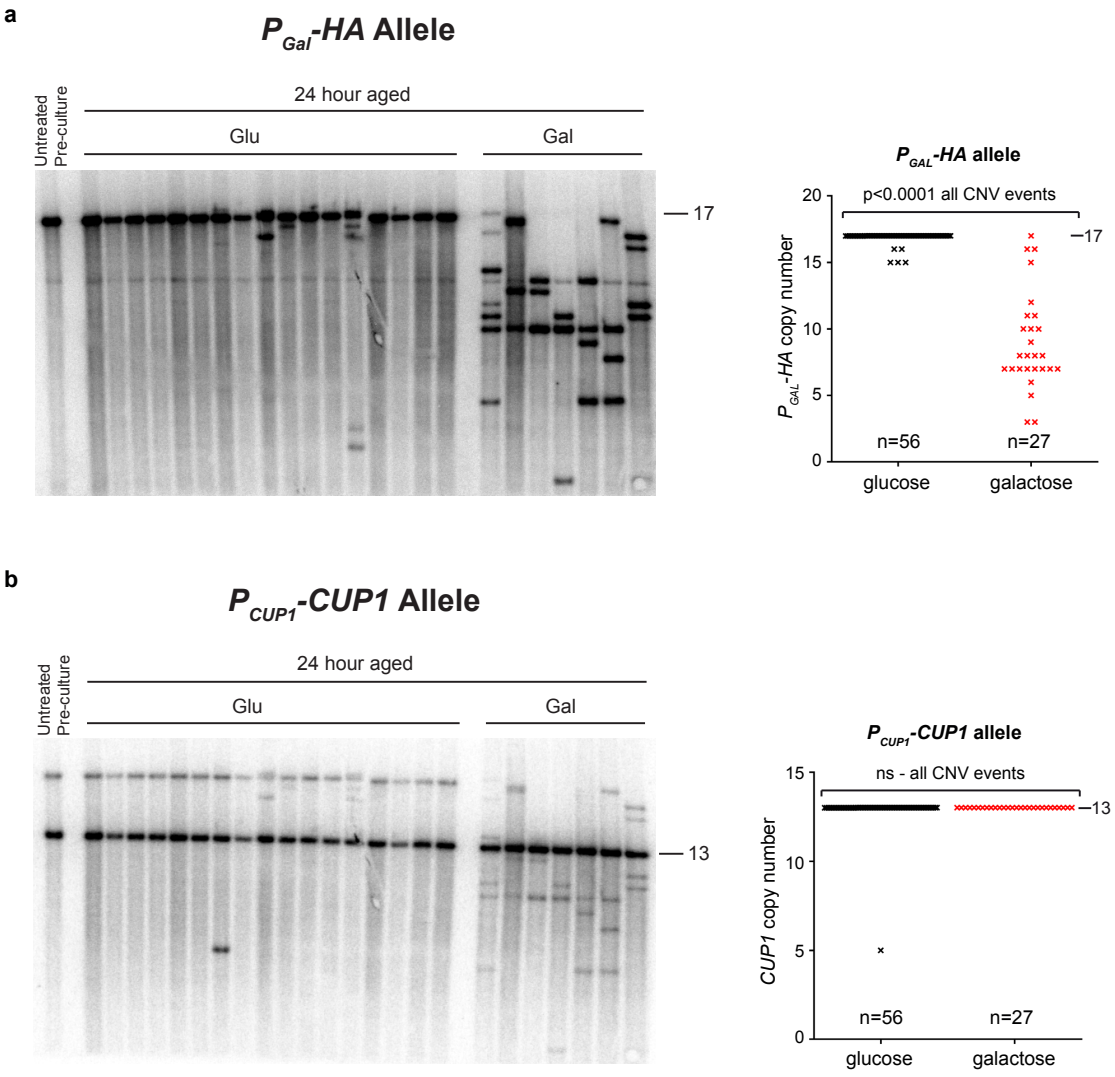


Figure 4.3: Allele specific stimulated CNV. PFGE-Southern blot and analysis of *HA* and *CUP1* copy number in 56 glucose (Glu) aged and 28 galactose (Gal) aged MEP P_{GAL}-*HA* heterozygous diploid yeast after 24 hours. Data is shown for both alleles in the same cells. Allele-specific probes covering the *GAL1* promoter and *HA* ORF in **a** and the *CUP1* promoter and ORF in **b** were used. 100% of the starting cohort was recovered in the glucose condition and 63% was recovered in the galactose treatment. Experimentally observed mutation rates in the glucose aged cohort were normalised for the viability in the galactose aged cohort, to produce the expected mutation rates for the galactose treatment. This included the conservative assumption that cells lost during the experiment did not favour either yeast that do or do not undergo CNV. *p* values were calculated by a goodness of fit χ^2 test with 1 degree of freedom between the observed and expected number of mutant to wild-type alleles across the galactose treated cohort. ns = not significant.

In comparison, 26 out of the 27 alleles screened in galactose showed CNV mutations. This was a mutation rate of 0.963 CNV events per allele, an 11-fold increase on the mutation rate in glucose. The glucose cohort recovered 100% of the starting cells and the galactose cohort recovered 63%. Taking these viabilities into account, galactose treatment caused a highly significant increase in *de novo* CNV events at the P_{GAL}-*HA* allele, *p* value <0.0001, as determined by a goodness of fit χ^2 test with 1 degree of freedom. The 11-fold increase in CNV rate at the P_{GAL}-*HA* allele with galactose, also matches the 10-fold increase in CNV rate in the MEP wild-type diploid with copper (Fig. 4.1b). Therefore promoter induction can stimulate the local CNV rate approximately 10-fold, and provided further evidence for copper-stimulated *CUP1* CNV being a transcriptional and not mutagenic effect.

To best show that the galactose-stimulated CNV was a transcriptional effect, the exact same MEP P_{GAL}-*HA* heterozygote cells used above, were then probed for the other allele, P_{CUP1}-*CUP1* (Fig. 4.3b). Neither glucose nor galactose treatment should transcriptionally induce the P_{CUP1}-*CUP1* allele. The glucose condition observed just a single CNV event out of 56 screened alleles, a mutation rate of 0.018 CNV events per allele, and the galactose treatment failed to produce any CNV events from the 27 screened alleles. There was no significant difference in *de novo* CNV events between glucose and galactose treatments at the P_{CUP1}-*CUP1* allele. This result provided evidence for CNV not being stimulated uniformly, but instead being highly specific to the transcriptionally active allele over a silent locus of similar sequence and copy number.

Also, all the CNV events detected in the MEP P_{GAL}-*HA* heterozygote were contractions. This was most likely because of the high starting copy number in these cells, combined with the absence of any selection pressure to maintain such a high copy number. For further explanations on the contraction bias see the discussion.

4.5 *RTT109* deletion reduces *de novo CUP1* CNV in response to copper

To confirm the importance of H3K56ac in copper-stimulated CNV at the *CUP1* locus, the *rtt109Δ* mutation was introduced into the MEP, and Cell Fate Analysis of MEP *rtt109Δ* diploid cells was performed as in Fig. 4.1a. Out of the 96 alleles (48 clones) screened for *CUP1* copy number, only 3 show novel CNV events in untreated cells (Fig. 4.4). This was a mutation frequency of 0.031 CNV events per allele, which was very similar to the 0.027 CNV events per allele shown earlier in the untreated MEP wild-type. Therefore the *rtt109Δ* mutation did not interfere with CNV in the absence of copper-stimulation. However, when stimulated with 1mM $CuSO_4$, the number of *de novo* CNV events only marginally increased to 5 out of 96 alleles (48 clones) screened. This was a mutation rate of 0.052 CNV events per allele and a 1.7-fold increase on the CNV rate without any stimulation. However, this slight increase in *de novo* CNV alleles with copper-stimulation in MEP *rtt109Δ* cells was not statistically significant, *p* value = 0.359, meaning that the variation in the number of observed CNV events, between the two conditions, was no greater than expected by random sampling. The copper-stimulated CNV rate in MEP *rtt109Δ* cells was also well down on the 10-fold increase in CNV seen earlier in MEP wild-type cells with copper-stimulation, which showed that deletion of the H3K56 acetyltransferase gene *RTT109* caused a 5.9-fold reduction in copper-stimulated CNV.

To check that the decreased CNV rate in the *rtt109Δ* mutant was not simply due to cells replicating less, budscar counts were determined for MEP wild-type and *rtt109Δ* cells, in both the presence and absence of 1mM $CuSO_4$. The untreated wild-type underwent 12.2±2.3 divisions and the matching *rtt109Δ* cells underwent 11.5±4.6 divisions, which was not a significant difference, *p* value = 0.9721. In copper, the wild-type underwent 8.2±3.2 divisions, with *rtt109Δ* cells undergoing 6.3±1.6 divisions, which again was not significant, *p* value = 0.2029. Therefore budscar counting confirmed that *rtt109Δ* cells replicated as often as wild-type yeast in the presence or absence of 1mM $CuSO_4$.

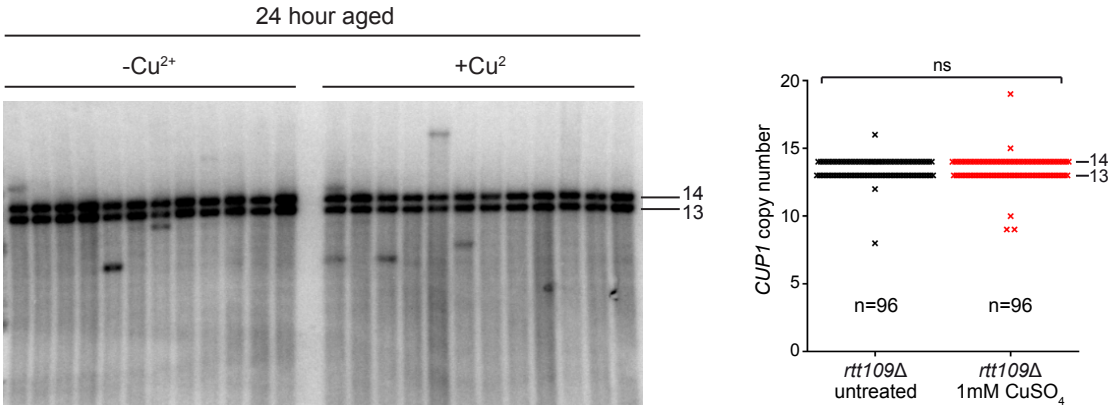


Figure 4.4: H3K56ac loss reduced *de novo* CNV at *CUP1*. PFGE-Southern blot and analysis of *CUP1* copy number alleles in colonies derived from 48 untreated and 48 1mM $CuSO_4$ treated diploid MEP *rtt109*Δ cells after 24 hours. 91% of the starting cohort of cells was recovered in the untreated condition, and 40% was recovered in the 1mM $CuSO_4$ treatment. Experimentally observed mutation rates in the untreated cohort were normalised for the viability in the treated cohort, to produce the expected mutation rates for the treated condition. This included the conservative assumption that cells lost during the experiment did not favour either yeast that do or do not undergo CNV. *p* values were calculated by a goodness of fit χ^2 test with 1 degree of freedom between the observed and expected number of mutant to wild-type alleles across the treated cohort. ns = not significant.

Copper treatment did cause a significant reduction in the number of cellular divisions in the *rtt109Δ* mutant, *p* value = 0.0010, but it did so equally in the wild-type, *p* value = 0.0012. Therefore the reduced CNV rate in the *rtt109Δ* mutant grown in copper was not due to a lower rate of replication. As such, the acetyltransferase Rtt109, and its product H3K56ac, are essential for copper-stimulated CNV at *CUP1*.

4.6 Creating the 3x*CUP1* strain

To determine whether stimulated CNV functions in low-copy systems as observed in high-copy systems, a low-copy strain with just 3 *CUP1* repeats was created (Fig. 4.5). Previous attempts in our group to manipulate this locus had encountered problems: 1) after deleting the endogenous locus, successful replacement with modified *CUP1* constructs was inefficient with

many partial constructs integrated at the wrong location; 2) the previous generation of constructs disrupted the overlapping *RSC30* gene; and 3) the absence of the *MET25* gene, in our lab standard BY4741 *MATa* haploid background, profoundly impacted copper resistance. To address these issues the native *ADE2* gene was first deleted in the *MATα* haploid BY4742 wild-type background, which carries a wild-type copy of *MET25*. *ADE2* encodes a phosphoribosylaminoimidazole carboxylase which catalyses the sixth step of the *de novo* purine biosynthetic pathway in *Saccharomyces cerevisiae* (Dorfman, 1969). *ade2Δ* yeast accumulate a red pigment as a result of adenine deprivation, which can then be used as an additional phenotypic selection marker (Zonneveld and van der Zanden, 1995; Ugolini and Bruschi, 1996). This was helpful as many of the background colonies obtained after selection on plates lacking Adenine were red, presumably due to integration of a partially functional *ADE2* marker, or integration in a repressed site.

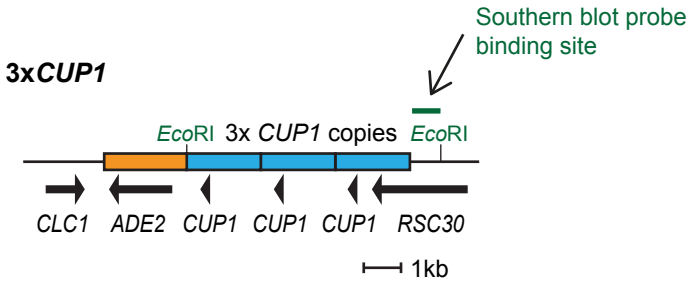


Figure 4.5: Schematic of the 3x*CUP1* locus. Blue boxes indicate *CUP1* repeats, and the orange box indicates the *ADE2* marker. The reading frame of *RSC30* is maintained across the construct boundary. Restriction enzymes for Southern analysis are shown in green along with the probe location.

The red *ade2Δ* yeasts were then transformed with a single-copy *CUP1* plasmid containing a *URA*3 marker, as *Saccharomyces cerevisiae* require a single copy of the *CUP1* gene to survive (Fogel and Welch, 1982), and then the entire native *CUP1* locus [ChrVIII: 212265...216250] was deleted using a *KanMX6* marker. A modified 3x*CUP1* plasmid, pRH9, was constructed containing 3 complete genomic *CUP1* copies [ChrVIII: 214256...216239] in

tandem, along with an *ADE2* marker and native *CUP1* flanking sequence, which was *Sac*I digested and transformed into *ade2Δ* yeast. The pRH9 construct was designed so that the reading frame of the overlapping chromatin remodelling gene *RSC30* was maintained after transformation, and the maintenance of the correct *RSC30* open reading frame was confirmed in positive clones by sequencing. White *ADE2* positive, G418 sensitive colonies were selected, genotyped by PCR and grown on FOA, to select for loss of the single-copy *CUP1* containing *URA* plasmid. This produced the YRH23/24 strains, known henceforth as 3x*CUP1*. *CUP1* copy number of the resulting strains was often variable, so the actual copy number was validated by Southern blot (data not shown). Other variants of the *CUP1* locus used later were created by a similar strategy, although it was discovered that removing the *CUP1* plasmid by plating on FOA immediately prior to transformation of the *CUP1* replacement construct substantially improved integration efficiency.

4.7 Emergence of high-copy *CUP1* alleles in 3x*CUP1* is under positive selection in copper

The results presented earlier in this Chapter using a high-copy *CUP1* system showed that environmental copper transcriptionally stimulated CNV at *CUP1*, in a replication-linked mechanism. Therefore, it is essential for stimulated CNV that cells are only ever exposed to sub-lethal copper concentrations that still allow replication to proceed. The high-copy systems replicated normally in 1mM $CuSO_4$, but this concentration was too toxic for low-copy 3x*CUP1* cells. As such, the culturing concentration of $CuSO_4$ was optimised for use with 3x*CUP1* cells, with 0.3mM $CuSO_4$ being determined to be the optimum concentration of copper to proceed with (data not shown).
To determine whether 3x*CUP1* cells could replicate in 0.3mM $CuSO_4$, the yeast was grown for 72 hours in a growth curve assay, with or without 0.3mM $CuSO_4$ (Fig. 4.6a). After about 30 hours, all 16 of the untreated replicates grew to saturation. By comparison, all 16 of the 0.3mM $CuSO_4$ treated replicates displayed clear growth retardation by the same time-point. However, fast-growing, highly copper-resistant populations were still able to

arise in late growth, from about 40 hours into the growth assay. Therefore 0.3mM $CuSO_4$ was sub-lethal to 3x*CUP1* cells, with copper-resistant cells under strong positive selection in copper.

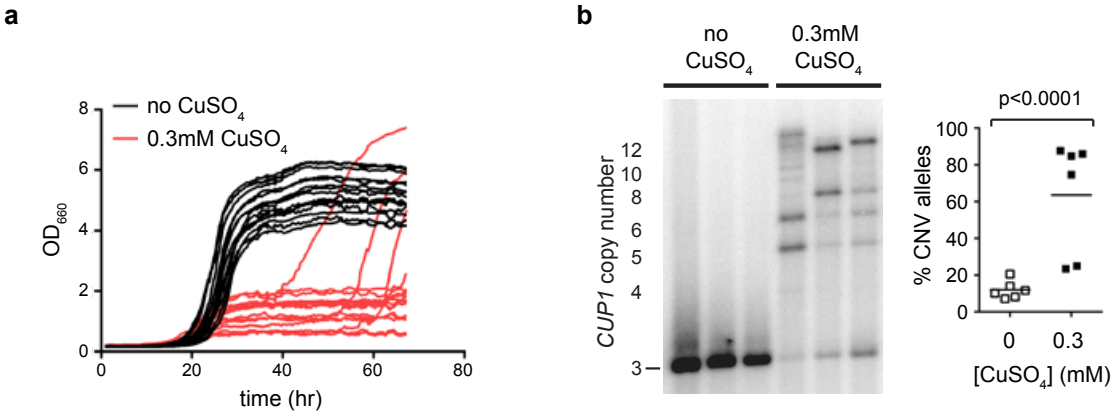


Figure 4.6: Emergence of highly copper-resistant cells with amplified *CUP1* copy number. **a)** Growth curves of 3x*CUP1* cells grown with or without 0.3mM $CuSO_4$, *n* = 16 for each condition. Note that the growth retardation caused by 0.3mM $CuSO_4$ is stronger in the 200µl 96-well plate cultures used for growth curve analysis than in the 4ml batch cultures used for Southern blot samples; although cells also grow slowly in 0.3mM $CuSO_4$ under these conditions, almost all cultures reach saturation by 72 hours. **b)** Southern blot analysis of *CUP1* copy number in 3x*CUP1* cells grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted in the same media, with or without 0.3mM $CuSO_4$, and re-grown to saturation over a further 10 generations. Quantification shows the percentage of CNV alleles, *n* = 6. *p* value calculated by a two-tailed paired *t* test.

Having shown that 3x*CUP1* cells still replicated in 0.3mM $CuSO_4$, it was important to determine whether amplified *CUP1* alleles emerge under 0.3mM $CuSO_4$ treatment. 3x*CUP1* yeast was grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted in the same media with or without 0.3mM $CuSO_4$, and re-grown to saturation over a further 10 generations. DNA was extracted for Southern blot analysis at 120V for 16.5 hours on a 1% 1xTBE agarose gel. In the absence of copper, no novel *CUP1* CNV alleles were detected by Southern blot analysis (Fig. 4.6b). Whereas novel alleles, with amplified *CUP1* copy numbers were detected in response

to 0.3mM $CuSO_4$ treatment. Copper treatment observed the emergence of significantly more novel *CUP1* alleles than the untreated condition, *p* value <0.0001. However, this result cannot determine whether the *CUP1* amplifications were stimulated by 0.3mM $CuSO_4$, or a selection towards rare pre-existing random mutations with higher *CUP1* copy number. Nonetheless, this result showed that highly copper-resistant cells, with high *CUP1* copy numbers, were under positive selection in 0.3mM $CuSO_4$, further supporting the growth curve data.

4.8 CNV and adaptation is reduced in *rtt109Δ*

In the high-copy systems, H3K56ac was essential for stimulated CNV, with total abrogation of CNV in an *rtt109Δ* mutant. So to test the importance of H3K56ac in the low-copy 3x*CUP1* system, *RTT109* was deleted in 3x*CUP1* yeast, and both 3x*CUP1* wild-type and *rtt109Δ* cells were grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted in the same media with and without of 0.3mM $CuSO_4$ and re-grown for a further 10 generations. Southern blot analysis of *CUP1* copy number showed that the emergence of *CUP1* amplifications was significantly repressed in the *rtt109Δ* mutant (Fig. 4.7 a lane 2 vs. 4), *p* value <0.0001. The *rtt109Δ* cells completely repressed all CNV in 0.3mM $CuSO_4$ to the level in untreated cells (Fig. 4.7a lane 3 vs. 4). Therefore, the *CUP1* alleles with high copy number appear to be stimulated by 0.3mM $CuSO_4$ in an H3K56ac-dependent mechanism, just as was observed in the high-copy systems.

The same cells run on the Southern blot (Fig. 4.7a), were also challenged in a copper adaptation assay (Fig. 4.7b). Since the parental *CUP1* copy number is much lower in 3x*CUP1* cells than in the MEP cells used earlier (Fig. 4.2), these adaptation assays were performed using much lower copper concentrations and without ascorbic acid.

Copper pre-treated 3x*CUP1* yeast, which underwent transcriptionally stimulated CNV, also caused a highly significant adaptive advantage in a challenging copper environment, *p* value <0.0001. This adaptive advantage was significantly reduced in *rtt109Δ* yeast, *p* value <0.0001. This reduced

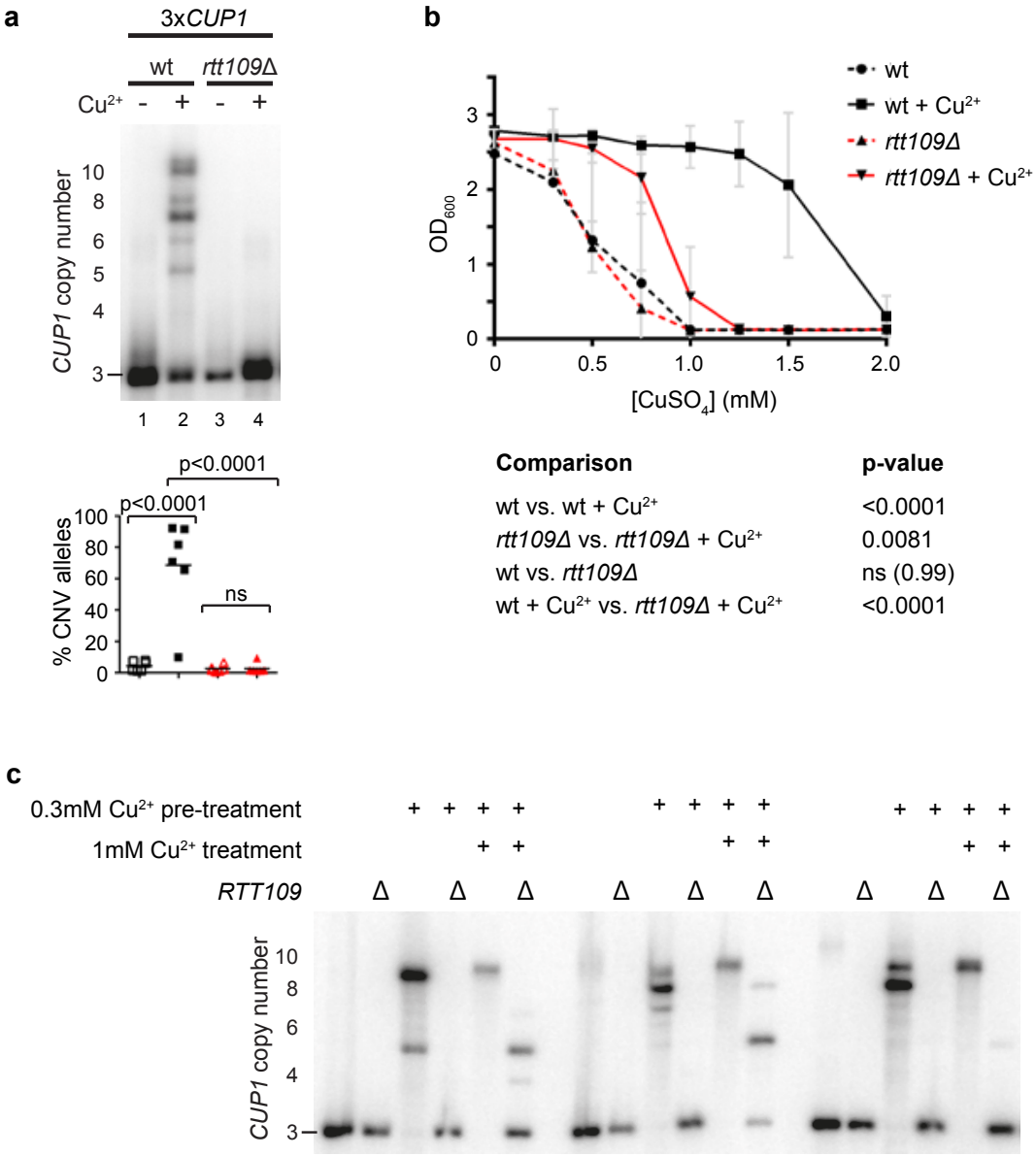


Figure 4.7: H3K56ac was required for stimulated CNV and copper adaptation in 3x*CUP1* cells. **a**) Southern analysis of *CUP1* copy number in 3x*CUP1* wild-type (wt) and *rtt109*Δ cells grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted and re-grown to saturation over 10 generations in the same media with or without 0.3mM $CuSO_4$. Quantification shows the percentage of CNV events deviating from the parental copy number, $n = 6$. p values calculated by 1-way ANOVA for repeated measurements. ns = not significant. **b**) Copper resistance of same cells from **a.** Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, and $n = 6$ cultures per condition, each tested at 8 $CuSO_4$ concentrations. **c**) Southern blot analysis of *CUP1* copy number in 3x*CUP1* wild-type and *rtt109*Δ cells from the adaptation assay in **b**, outgrown for 10 generations without copper. p values were calculated by 1-way ANOVA of area-under-curve values for each culture.

adaptive ability was not due to a general fitness disadvantage in *rtt109Δ* cells, as they showed no significant adaptive difference to the wild-type when untreated, *p* value = 0.99. Therefore loss of H3K56ac in *rtt109Δ* cells repressed stimulated CNV and impaired the ability of the cells to adapt to high copper environments.

Interestingly, copper pre-exposed 3x*CUP1 rtt109Δ* yeast still showed a significant survival advantage over its untreated counterpart, *p* value = 0.0081. This was surprising as CNV appeared to be completely repressed in *rtt109Δ*, according to Southern blot analysis (Fig. 4.7a). The partial copper-adaptation in copper pre-exposed *rtt109Δ* cells could therefore be via alternative copper-resistance mutations that do not involve *CUP1* CNV, such as point mutations or epigenetic changes. Alternatively, copper-stimulated *CUP1* CNV might still occur in r*tt109Δ* yeast, but at a greatly reduced rate that is below the threshold of detection by Southern blot. Such minor CNV would have to be independent of H3K56ac, or else cells must have another gene encoding a protein with much less efficient H3K56 acetyltransferase activity, that partially compensates in the absence of *RTT109*. Finally, as this experiment was performed without the use of the MEP, the possibility of highly copper-resistant cells, with high *CUP1* copy numbers, pre-existing extremely rarely in the starting population cannot be ruled out. As such, even with strong selection in copper, these cells may still be below the threshold of detection by Southern blot analysis after 10 generations. However, after an additional 10 generations of selective growth in the adaptation assay, these highly copper resistant yeast might finally be detected by the more sensitive copper adaptation assay.

To try and rule out some of these possibilities, 1mM $CuSO_4$ resistant 3x*CUP1 rtt109Δ* yeast from the adaptation assay (Fig. 4.7b), were outgrown for 10 generations in the absence of copper, to determine whether CNV was now detected by Southern blot analysis (Fig. 4.7c). In all 3 replicates, amplified *CUP1* alleles were now detected in *rtt109Δ* cells that survived 1mM $CuSO_4$ in the adaptation assay. Therefore the partial copper-adaptation ability of copper pre-treated *rtt109Δ* cells was still through *CUP1* CNV. Whether this CNV in *rtt109Δ* yeast was independent of H3K56ac, or dependent on H3K56ac but acetylated by an as yet unknown acetyltransferase, or simply an extremely

rare pre-existing amplification event cannot be concluded. Interestingly, the wild-type cells that could survive in 1mM $CuSO_4$ in the adaptation assay now displayed a single CNV allele at the highest observed *CUP1* copy number. This showed that in a highly selective copper environment, stimulated CNV continued until the population was dominated by highly copper-resistant cells, all with the highest possible *CUP1* copy number, for best survival in the challenging environment.

4.9 Nicotinamide-stimulated CNV in 3x*CUP1*

To determine whether CNV could be stimulated in low-copy 3x*CUP1* cells in the absence of selection, 3x*CUP1* yeast was grown for 10 generations to saturation in complete synthetic media, and then 1:1000 diluted in the same media with or without 5mM nicotinamide and re-grown for a further 10 generations to saturation. In Chapter 3, nicotinamide was shown to stimulate extensive *CUP1* CNV in high-copy BY4741 wild-type cells by inhibiting the H3K56ac HDACs, without any environmental selection pressure. DNA was extracted from the cultures for Southern blot analysis of *CUP1* copy number (Fig. 4.8), to determine whether increased global H3K56ac levels could also stimulate *CUP1* CNV in low-copy strains. Unlike in the high-copy system, where numerous novel *CUP1* CNV alleles were stimulated by nicotinamide, the only CNV allele detected in 3x*CUP1* cells was the -1 repeat band for 2 *CUP1* copies. Despite producing fewer novel CNV alleles, nicotinamide treatment did still significantly stimulate CNV in 3x*CUP1* yeast over the untreated condition, p value <0.0001. Therefore increasing global H3K56ac still stimulated CNV in low-copy *CUP1* arrays.

As there was more starting *CUP1* repeats in the high-copy system, and an RFS site in every *CUP1* repeat, the high-copy system has more opportunities for a CNV event to occur with every round of replication, compared to the low-copy system. Therefore it was unsurprising that 3x*CUP1* yeast had a greatly reduced rate of nicotinamide-stimulated CNV, compared to high-copy systems. Also nicotinamide treatment in both high- and low-copy systems primarily produced contractions, as there was no selection pressure to

maintain high-copy number alleles. See the discussion for potential explanations for the observed contraction bias.

A +1 repeat band for 4 *CUP1* copies was observed at the limit of detection in nicotinamide-treated 3x*CUP1* cells, which showed that nicotinamide also stimulated amplification events. However, the +1 repeat band was much fainter than the -1 repeat band, which showed that amplifications are more infrequent than contractions, and largely pass under the threshold of detection by Southern blot analysis.

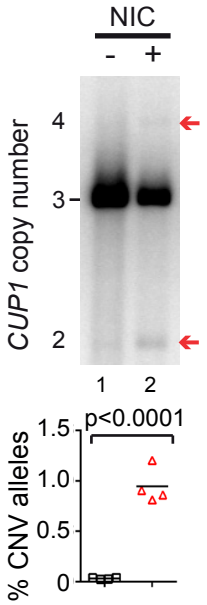


Figure 4.8: Nicotinamide stimulated CNV in 3x*CUP1* cells. Southern analysis of *CUP1* copy number in 3x*CUP1* cells grown for 10 generations with or without 5mM nicotinamide (NIC); arrows indicate -1 and +1 copy bands. Quantification shows the percentage of –1 alleles; n = 4, p value calculated by a two-tailed paired t test.

4.10 Nicotinamide enhances the emergence of high-copy *CUP1* alleles in copper

In the absence of selection, nicotinamide-stimulated CNV in 3x*CUP1* cells almost exclusively produced contractions (Fig. 4.8), whereas in the presence of selective copper only *CUP1* amplifications were observed (Fig. 4.6b). To

determine whether nicotinamide-stimulated contractions remain when a selection pressure was introduced, 3x*CUP1* cells were grown for 10 generations to stationary phase in complete synthetic media, then 1:1000 diluted in the same media in the presence or absence of 5mM nicotinamide, and with or without 0.3mM $CuSO_4$. Cultures were then re-grown for a further 10 generations to saturation, prior to Southern blot analysis of *CUP1* copy number (Fig. 4.9).

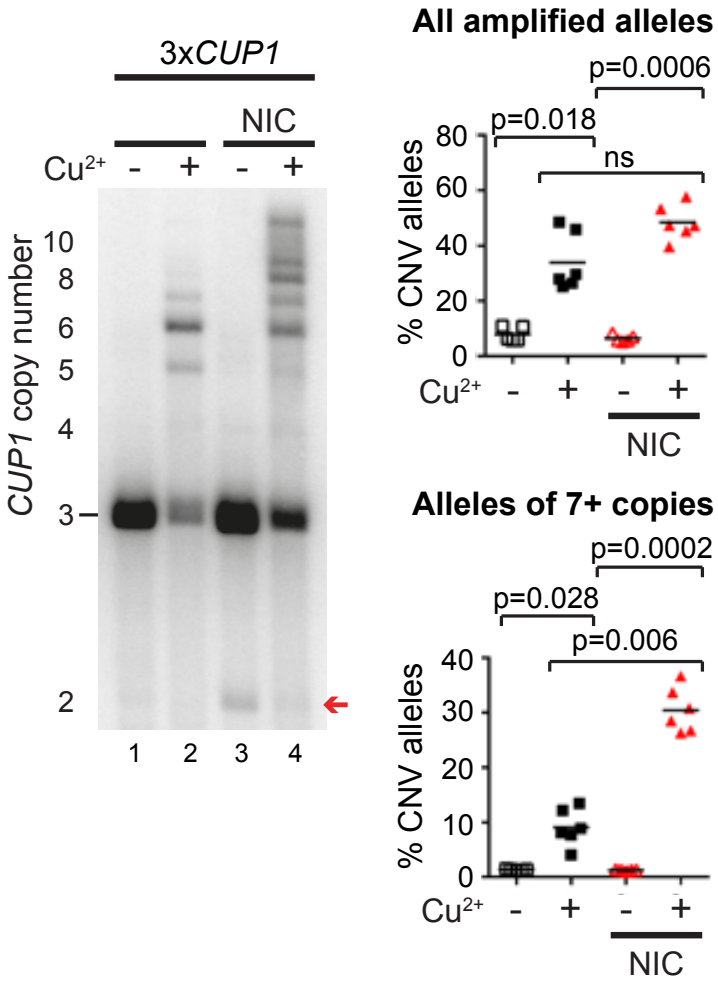


Figure 4.9: Nicotinamide enhanced the emergence of *CUP1* alleles with high copy number in copper. Southern blot analysis of *CUP1* copy number in 3x*CUP1* cells grown for 10 generations, with or without 5mM nicotinamide (NIC), and with or without 0.3mM $CuSO_4$. Upper quantification showed the percentage of alleles deviating from the parental copy number, and lower quantification showed the percentage of amplification events with greater than two times the parental copy number, $n = 6$. p values calculated by 1-way ANOVA for repeated measurements. ns = not significant.

As shown previously in Fig. 4.6b, 3x*CUP1* cells treated with copper alone again observed a significant increase in the emergence of *CUP1* alleles with higher than parental copy number, *p* value = 0.018. Also as shown in Fig. 4.8, 3x*CUP1* cells treated with nicotinamide alone primarily produced a -1 repeat contraction band. Interestingly, the treatment with nicotinamide and copper observed a reduction in the level of the -1 contraction band, compared to treatment with nicotinamide alone. Therefore copper negatively selected against nicotinamide-stimulated *CUP1* contractions. The treatment with nicotinamide and copper also observed a slight, but non-significant increase in the total number of *CUP1* CNV alleles, compared to the treatment with copper alone, *p* value = 0.1224. However, when just the novel CNV alleles with high copy number was considered, as defined by having a copy number greater than 2-fold the parental copy number, significantly more high-copy CNV alleles were detected in the treatment with copper and nicotinamide than the treatment with copper alone, *p* value = 0.006.

Individually, nicotinamide was capable of stimulating CNV events at *CUP1*, and $CuSO_4$ could select for the emergence of *CUP1* alleles with high copy number, even in a low-copy system. Alone nicotinamide-stimulated CNV primarily generated contractions, but when combined with a copper selection pressure, nicotinamide enhanced the emergence of *CUP1* alleles with high copy number. Therefore nicotinamide stimulation of *CUP1* CNV provided an additive effect to copper selection towards alleles with the highest *CUP1* copy numbers.

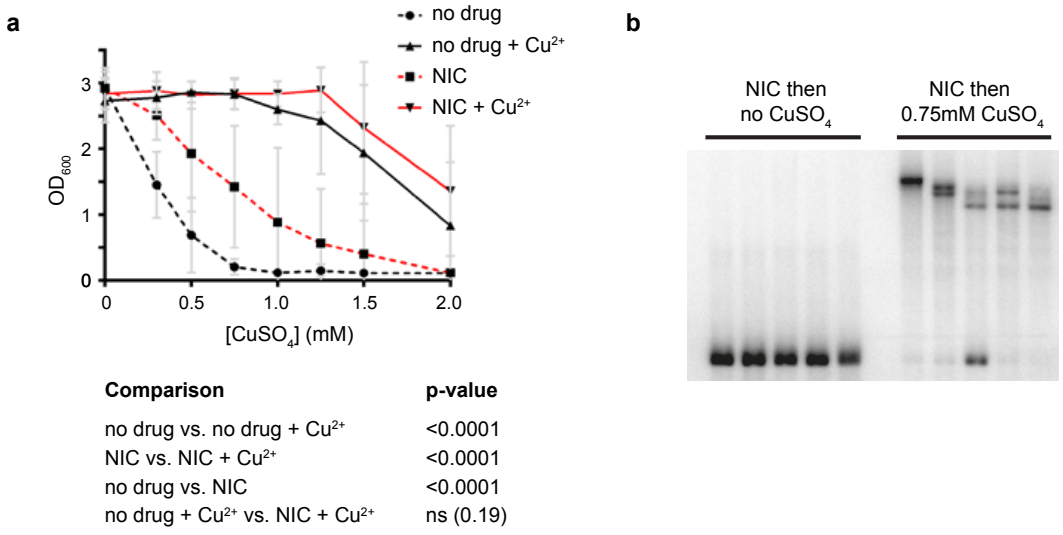
4.11 Stimulated CNV confers a fitness advantage

To determine whether nicotinamide-stimulated CNV in 3x*CUP1* cells confers an adaptive advantage in copper, the same 3x*CUP1* cells used for the Southern blot in Fig. 4.9 above, were challenged in a copper adaptation assay (Fig. 4.10a). The copper adaptation assay used the same conditions as in Fig. 4.7b.

As observed previously, copper pre-treated 3x*CUP1* yeast again showed a highly significant survival advantage over the untreated 3x*CUP1* cells, *p* value

<0.0001. Confirming that pre-exposure to copper increases the proportion of cells in a population that have alleles with higher than parental *CUP1* copy number, which provides copper pre-treated cells an adaptive advantage in a hostile copper environment.

Despite showing an increased proportion of high-copy *CUP1* alleles in 3x*CUP1* cells by Southern blot, the combined pre-treatment with 0.3mM $CuSO_4$ and 5mM nicotinamide did not cause a significant increase in adaptive ability to a hostile copper environment, over pre-treatment with copper alone, *p* value = 0.19. However, the $CuSO_4$ and nicotinamide pre-treatment did begin to outperform the treatment with copper alone at $CuSO_4$ concentrations of 1mM and above. Therefore the assay design might have used too low copper concentrations to clearly separate the high-resistance copper adaptation profiles of copper and nicotinamide pre-treated cells, from cells pre-treated with copper alone.



Comparison	p-value
no drug vs. no drug + Cu^{2+}	<0.0001
NIC vs. NIC + Cu^{2+}	<0.0001
no drug vs. NIC	<0.0001
no drug + Cu^{2+} vs. NIC + Cu^{2+}	ns (0.19)

Figure 4.10: Nicotinamide-stimulated CNV confers an adaptive advantage in copper. **a**) Copper resistance of 3x*CUP1* wild-type cells grown with or without 0.3mM $CuSO_4$ and with or without 5mM nicotinamide (NIC). Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, *n* = 12 cultures per condition, each tested at 8 $CuSO_4$ concentrations. *p* values were calculated by 1-way ANOVA of area-under-curve values for each culture. **b**) *CUP1* copy number distribution of nicotinamide pre-grown 3x*CUP1* cells from 0 and 0.75mM $CuSO_4$ in the adaptation curve in **a**, then outgrown without drug.

Surprisingly, the 5mM nicotinamide pre-treated cells, which produced mostly -1 repeat contractions by Southern blot analysis, also displayed a significant enhancement in adaptation to a hostile copper environment, over untreated 3x*CUP1* cells, *p* value <0.0001. However, the nicotinamide-stimulated adaptation to copper was still much lower than the adaptation to copper in copper pre-treated cells, *p* value <0.0001. The partial copper-adaptive phenotype of nicotinamide pre-treated cells might come from rare cells with nicotinamide-stimulated *CUP1* amplifications, which were too infrequent to observe by Southern blot in the absence of any selection (Fig. 4.9). This provided some supporting evidence for nicotinamide also stimulating *CUP1* amplifications in Fig. 4.8, where a +1 repeat band was observed at the limit of detection in nicotinamide-treated 3x*CUP1* cells.

To determine whether the nicotinamide-stimulated copper-adaptive advantage was due to *CUP1* amplifications, the nicotinamide pre-treated cells that were exposed to 0 and 0.75mM $CuSO_4$ in the adaptation assay were outgrown for 10 generations to saturation in complete synthetic media. Southern blot analysis showed that in the absence of copper selection, nicotinamide pre-treated cells still showed no observable *CUP1* amplifications (Fig. 4.10b). Whereas all 6 experimental replicates that adapted to 0.75mM $CuSO_4$ in the copper adaptation assay, could now detect *CUP1* alleles with high copy number by Southern blot analysis. This provided supporting evidence for rare nicotinamide-stimulated *CUP1* amplifications being selected for in copper and providing the partial copper-adaptive phenotype of nicotinamide pre-treated cells.

4.12 Construction of modified 3x*CUP1* strains for detection of *cis*-acting elements at *CUP1*

Cis-regulatory elements are highly conserved DNA sequences that can direct patterns of gene expression (Wittkopp, 2006). To determine whether *CUP1* had any *cis*-acting sequences required for the emergence of *CUP1* alleles with higher than parental copy number in copper, our lab created 3 modified 3x*CUP1* constructs (Fig. 4.11). The first strain, 3x[*CUP1-NodA*], had a

continuous track of 16 poly(dA) bases that occurs 260bp downstream of the 3' end of the *CUP1* ORF removed (Fig. 4.11a). A continuous track of poly(dT) bases had previously been identified as an important sequence element, where it defined the breakpoints of an adaptive mutation, a 5.1kb deletion on chromosome IV (587839–592999), during an evolution study of *Saccharomyces cerevisiae* lacking the high affinity sulphate permease Sul1, grown in sulphate-limited media (Payen *et al.*, 2016). Therefore the poly(dA) sequence was deleted in the 3x[*CUP1-NodA*] strain to determine whether poly(dA) at *CUP1* might also be involved in defining the breakpoint at *CUP1* in response to copper.

The second strain, 3x[*CUP1-Ttef*], replaced the native *CUP1* terminator sequence, starting from the 3' end of the *CUP1* ORF to the end of the repeat, with the TEF terminator from a pFA6a-*KanMX6* plasmid (Fig. 4.11b). The poly(dA) sequence was maintained at the end of the TEF terminator sequence. The 3x[*CUP1-Ttef*] strain should identify any important *cis*-acting elements at *CUP1*, downstream of the *CUP1* ORF, that are present in every repeat.

The final strain, 3x[*GFP-CUP1*], replaced the entire DNA sequence, from the start of the *CUP1* repeat to the start of the *CUP1* promoter, with a non-functional *GFP* tag (Fig. 4.11c). The replaced region upstream of *CUP1* included the predicted location of the *CUP1* RFS site. The 3x[*GFP-CUP1*] strain should identify important *cis*-acting elements upstream of the *CUP1* ORF that are present in every repeat.

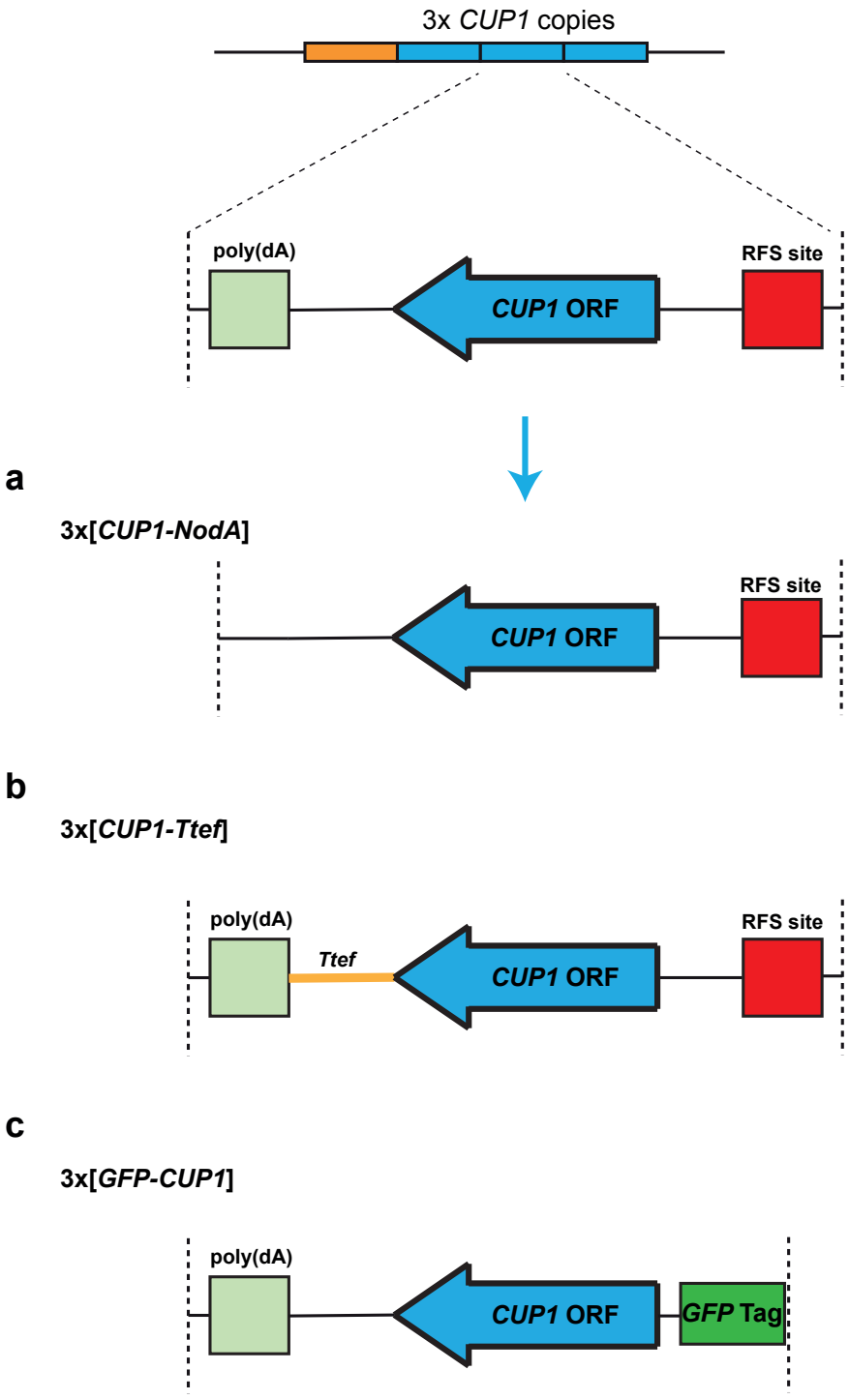
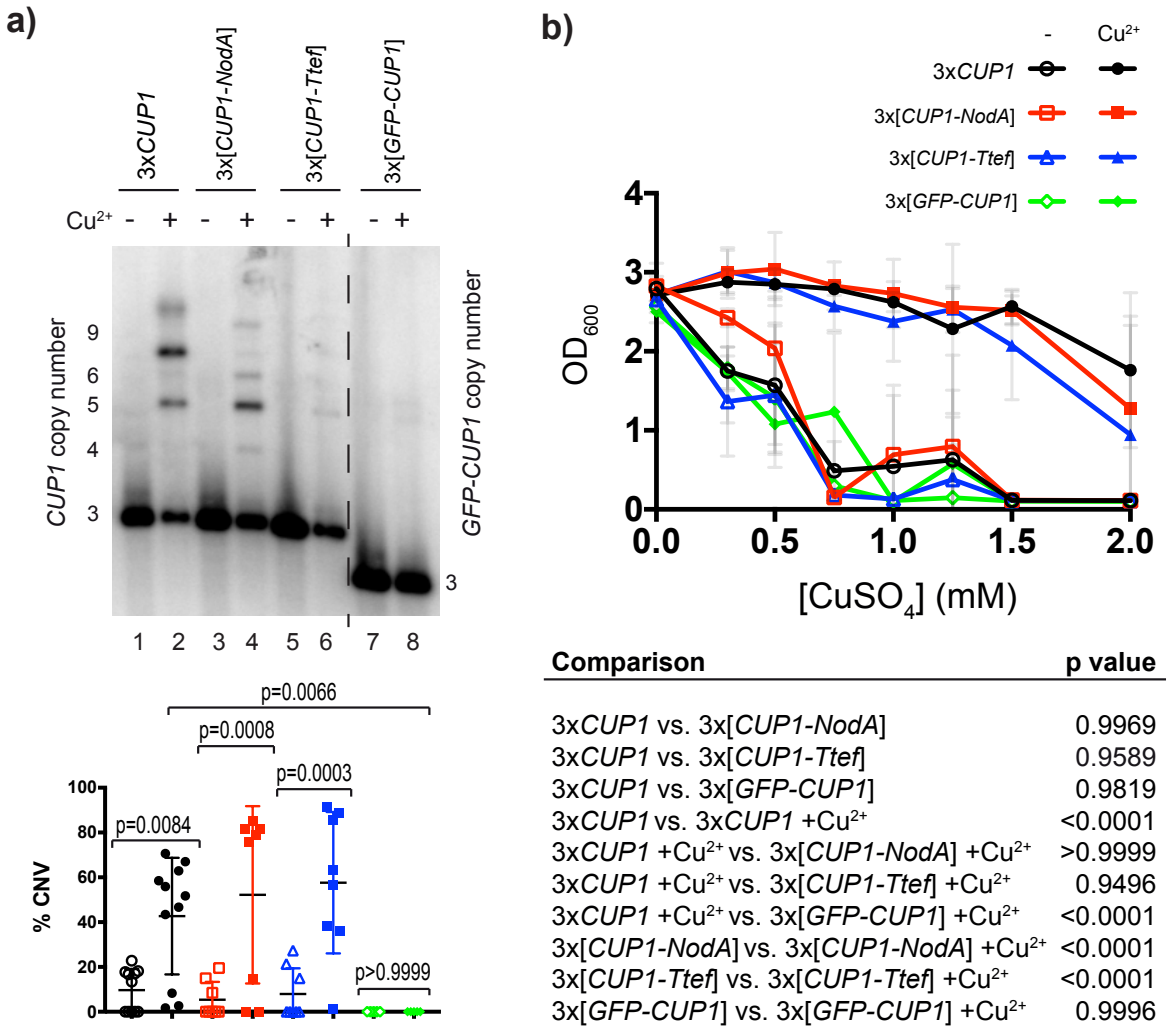


Figure 4.11: Schematics of the modified 3x*CUP1* locus in 3 strains for identification of *cis*-acting elements. Each *CUP1* repeat in 3x*CUP1* cells contains the *CUP1* ORF (blue arrow), an upstream RFS site (red box), and a continuous sequence of 16 poly(dA) downstream of the *CUP1* ORF. **a**) In 3x[*CUP1-NodA*] cells the poly(dA) region was deleted. **b**) In 3x[*CUP1-Ttef*] cells the downstream *CUP1* terminator sequence was replaced with the TEF terminator from a pFA6a-*KanMX6* plasmid, leaving the poly(dA) region intact at the end of the terminator. **c**) In 3x[*GFP-CUP1*] cells the *CUP1* upstream region, including the RFS site was replaced by a *GFP* tag.

4.13 The *CUP1* RFS site is essential for the emergence of *CUP1* alleles with higher than parental copy number in copper

To determine whether there are any *cis*-acting elements at *CUP1* that are important for the emergence of *CUP1* alleles with higher than parental copy number, the 3x*CUP1* wild-type, 3x[*CUP1-NodA*], 3x[*CUP1-Ttef*], and 3x[*GFP-CUP1*] strains were grown for 10 generations to saturation in complete synthetic media, then 1:1000 diluted in the same media with or without 0.3mM $CuSO_4$ and re-grown to saturation over an additional 10 generations. Southern blot analysis for *CUP1* copy number still detected a significant increase the number of novel *CUP1* alleles with higher than parental copy number with copper treatment in the 3x*CUP1* wild-type, *p* value = 0.0084, 3x[*CUP1-NodA*], *p* value = 0.0008, and 3x[*CUP1-Ttef*] strains, p value = 0.0003 (Fig. 4.12a). In copper, 3x[*CUP1-NodA*] and 3x[*CUP1-Ttef*] yeast also both showed no significant change in the percentage CNV detected, compared to the wild-type, *p* values = 0.9873 and 0.7996 respectively. Therefore, there did not appear to be any *CUP1 cis*-acting elements in the region from the 3' end of the *CUP1* ORF to the end of the repeat, and the poly(dA) sequence was not important for defining the breakpoint at *CUP1*. As shown in Fig. 4.11, the size of each *GFP-CUP1* repeat is smaller than the repeat size of the other 3x*CUP1* strains. As such, the Southern blot scale for *GFP-CUP1* repeats is different to the wild-type and other modified 3x*CUP1* strains. In 3x[*GFP-CUP1*] cells, the percentage of CNV alleles detected in copper was significantly reduced compared to the wild-type, *p* = 0.0066. The percentage of novel CNV alleles detected in 3x[*GFP-CUP1*] cells was completely repressed to the level of the untreated condition, *p* >0.9999. Therefore the DNA sequence from the start of the *CUP1* repeat to the start of the *CUP1* ORF contains an important *cis*-acting element essential for CNV in response to copper. Given that this *CUP1* upstream sequence contains the location of the *CUP1* RFS site (Hull *et al.*, 2017), and our predicted model for stimulated CNV proposes the requirement of an RFS site, the complete loss



Comparison	p value
3x*CUP1* vs. 3x[*CUP1-NodA*]	0.9969
3x*CUP1* vs. 3x[*CUP1-Tef*]	0.9589
3x*CUP1* vs. 3x[*GFP-CUP1*]	0.9819
3x*CUP1* vs. 3x*CUP1* +Cu^{2+}	<0.0001
3x*CUP1* +Cu^{2+} vs. 3x[*CUP1-NodA*] +Cu^{2+}	>0.9999
3x*CUP1* +Cu^{2+} vs. 3x[*CUP1-Tef*] +Cu^{2+}	0.9496
3x*CUP1* +Cu^{2+} vs. 3x[*GFP-CUP1*] +Cu^{2+}	<0.0001
3x[*CUP1-NodA*] vs. 3x[*CUP1-NodA*] +Cu^{2+}	<0.0001
3x[*CUP1-Tef*] vs. 3x[*CUP1-Tef*] +Cu^{2+}	<0.0001
3x[*GFP-CUP1*] vs. 3x[*GFP-CUP1*] +Cu^{2+}	0.9996

Figure 4.12: The *CUP1* RFS site is essential for the emergence of *CUP1* alleles with higher than parental copy number and adaptation to a challenging copper environment. Wild-type 3x*CUP1* cells, 3x*CUP1* cells with the poly(dA) region deleted (3x[*CUP1-NodA*]), 3x*CUP1* cells with the *CUP1* terminator replaced with a TEF terminator from pFA6a-*KanMX6* (3x[*CUP1-Tef*]), and 3x*CUP1* cells with the *CUP1* upstream region, including the RFS site, replaced with a *GFP* tag (3x[*GFP-CUP1*]) were grown to saturation for 10 generations in complete synthetic media, then 1:1000 diluted in the same media with or without 0.3mM $CuSO_4$ and re-grown to saturation over 10 generations. **a**) Southern blot analysis for *CUP1* copy number. Quantification shows the percentage of CNV events deviating from the parental copy number, $n = 6$ for 3x*CUP1* cells, $n = 8$ for 3x[*CUP1-NodA*] and 3x[*CUP1-Tef*] cells, $n = 5$ for 3x[*GFP-CUP1*] cells. p values calculated by 1-way ANOVA. **b**) Copper resistance of a random selection of the same cells from **a.** Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, and $n = 6$ for 3x*CUP1* cells and $n = 3$ for 3x[*CUP1-NodA*], 3x[*CUP1-Tef*], and 3x[*GFP-CUP1*] cells. Each culture was tested at 8 $CuSO_4$ concentrations.

of detectable CNV events in 3x[*GFP-CUP1*] cells support the assumption that the RFS site is essential for stimulated CNV. However, the possibility of alternative essential *CUP1 cis*-acting elements being located in the *CUP1* upstream sequence cannot be excluded.

To determine whether loss of the *CUP1* RFS site also results in an inability to adapt to a hostile copper environment, a random selection of the same cells used for the Southern blot analysis were also challenged in a copper adaptation assay (Fig. 4.12b). In the absence of copper, 3x[*CUP1-NodA*], 3x[*CUP1-Ttef*], and 3x[*GFP-CUP1*] strains showed no adaptive difference to copper, compared to wild-type cells, *p* values = 0.9969, 0.9589, and 0.9819 respectively. This showed that the modifications introduced into the 3x*CUP1* variants had no effect on the fitness of the cells in the absence of copper. Copper pre-treated 3x[*CUP1-NodA*], and 3x[*CUP1-Ttef*] cells also showed no copper-adaptive difference to wild-type cells pre-treated with copper, *p* value >0.9999, and p value = 0.9496 respectively. All 3x*CUP1* wild-type, 3x[*CUP1-NodA*], and 3x[*CUP1-Ttef*] cells showed significant increases in copper-adaptive ability in the copper pre-treated condition over the respective untreated condition, *p* values <0.0001 for all comparisons.

However, in keeping with its repression of *CUP1* CNV in copper, 3x[*GFP-CUP1*] cells pre-treated in copper significantly repressed adaptation to the hostile copper environment, *p* value <0.0001 vs. wild-type cells pre-treated in copper. 3x[*GFP-CUP1*] cells completely repressed the copper-adaptive phenotype to the level of the untreated cells, *p* value = 0.9996. The repression of adaptation to a hostile copper environment in 3x[*GFP-CUP1*] cells was even greater than the repression in 3x*CUP1 rtt109Δ* cells (Fig. 4.7b), which also repressed the emergence of *CUP1* alleles with higher than parental copy number in copper (Fig. 4.7a). As the 3x[*GFP-CUP1*] cells were not able to adapt to the high concentrations of copper used in the adaptation assay, cells could not be outgrown, as was done for the 3x*CUP1 rtt109Δ* cells (Fig. 4.7c), and so the final CNV of the 3x[*GFP-CUP1*] cells after exposure in the copper adaptation assay cannot be determined. Nonetheless, the complete inability of 3x[*GFP-CUP1*] cells to adapt to the challenging copper environment suggests that 3x[*GFP-CUP1*] cells are completely incapable of stimulating

CNV in response to environmental copper, most likely due to loss of the essential RFS site.

4.14 Summary

Adaptive mutation can be a controversial topic, not least because of limitations in current experimental approaches to conclusively separate mutations that arise after exposure to the environment, from mutations that pre-existed as rare cells in the initial population before exposure to the environment. The main challenge in separating adaptive mutations from random pre-existing mutations is selection, since mutations that confer a fitness advantage in the new environment from both origins of mutation become selected for in that environment, producing virtually indistinguishable final populations at the point of assaying. In this Chapter, I have presented the Cell Fate Analysis experiment, as a novel experimental protocol in which to study adaptive mutation events in a defined cohort of cells and free from issues of selection. The protocol utilised the MEP system in which daughter cells are unable to replicate, thereby removing the possibility for selection through propagation of cells with an adaptive advantage.
Using the Cell Fate Analysis experiment I have shown that environmental exposure to copper directly stimulates *de novo* CNV events at the wild-type *CUP1* locus. Next, I showed that these copper-stimulated CNV mutants confer quantifiable adaptive differences to high concentrations of copper. Then using the MEP P_{GAL}-*HA* heterozygote, I provided evidence for copper stimulating *de novo* CNV events at *CUP1* in a transcriptionally controlled mechanism, whereby CNV events are highly specific to the transcriptionally active allele over a silent locus of similar sequence and copy number. Finally, the acetyltransferase gene *RTT109* was deleted in the MEP system and abolished copper-stimulated CNV at *CUP1*, confirming the requirement for H3K56ac in transcriptionally stimulated CNV in response to the environment. Therefore yeast can undergo a tightly controlled transcriptionally stimulated CNV mechanism, which is induced by a specific environmental stimulus and

dependent on H3K56ac, as a novel pathway for a cell to accelerate its adaptation to a specific hostile environment.

In addition, I have provided evidence for transcriptionally stimulated CNV not being limited to the high-copy *CUP1* locus. Using a strain with 3 *CUP1* repeats, I have shown that copper treatment increases the proportion of alleles with amplified *CUP1* copy numbers detected within a population of 3x*CUP1* cells and also greatly increases the adaptive capability of the population of cells in a hostile copper environment. I have provided evidence for the emergence of *CUP1* alleles with higher than parental copy numbers in 3x*CUP1* cells being stimulated by copper, as deletion of *RTT109* reduced the observed number of *CUP1* amplifications in copper and also the ability of cells to adapt to a challenging copper environment. Furthermore, nicotinamide was able to stimulate CNV in 3x*CUP1* cells, and despite primarily stimulating *CUP1* contractions, nicotinamide pre-treatment conferred an adaptive advantage in a hostile copper environment. Also, when combined with a copper selection pressure, nicotinamide enhanced the emergence of alleles with high *CUP1* copy numbers. Therefore *CUP1* CNV and the ability to adapt to a challenging copper environment in a low-copy system were dependent upon exposure to environmental copper and H3K56ac, as was proven for high-copy repeat arrays.

Furthermore, replacement of the *CUP1* upstream sequence from the start of the repeat to the start of the *CUP1* ORF, which includes the location of the RFS site, with a non-functional *GFP* tag, completely blocked the emergence of any *CUP1* CNV alleles in copper and also completely repressed the ability of the cells to adapt to a challenging copper environment. Therefore, as our predicted model for stimulated CNV proposed, an upstream RFS site appeared to be essential for any CNV events to occur at the adjacent gene. No other potential *cis*-acting elements appeared to affect stimulated CNV at *CUP1*. As such I have provided evidence for transcriptionally stimulated CNV at *CUP1* being applicable to both high- and low-copy systems, which could have wide-ranging implications for how CNV mutations and environmental adaptation should now be considered.

5. Stimulated CNV as a general mechanism to accelerate adaptation and limitations of the system

5.1 Introduction

Using the copper-resistance *CUP1* gene, I have shown that yeast can accelerate their adaptation to a hostile environment through an environmentally stimulated CNV mechanism, which is controlled through promoter induction and local H3K56ac levels. I have also provided evidence for stimulated CNV occurring in a system with as few as 3 parental *CUP1* repeats, as well as in high-copy *CUP1* systems.

In this Chapter I will begin by exploring the fitness trade-off between the adaptive potential of stimulated CNV vs. the risk to the individual cell of mutating its genome. Then I will determine whether the TOR pathway is involved in the regulation of stimulated CNV and adaptation to a challenging copper environment, as it was at the rDNA, by treating cells with the TOR inhibitor rapamycin.

Furthermore, I will investigate potential limitations to our predicted model for stimulated CNV, by firstly determining the minimum repeat requirement for stimulated CNV using strains with 2 parental *CUP1* repeats, and even 1 genomic *CUP1* repeat, where there are no additional copies for DNA repair off a mismatched repeat to generate a CNV event. Finally, I will determine the translatability of the stimulated CNV mechanism, as defined using the model *CUP1* locus, to other potential CNV genes by using the formaldehyde-resistance gene *SFA1*, in both high- and low-copy systems.

5.2 Population benefit of stimulated CNV

In Chapter 4, nicotinamide pre-treatment conferred a survival advantage to 3x*CUP1* cells exposed to a challenging copper environment (Fig. 4.10a),

despite primarily stimulating deleterious *CUP1* contractions (Fig. 4.8 and 4.9). As nicotinamide is a non-selective environment to 3x*CUP1* cells, stimulated CNV in general has a contraction bias, but these contractions cannot be detected experimentally in selective environments like copper where amplifications confer a fitness advantage. To determine whether stimulated CNV is a beneficial mechanism for a population of cells to adapt to a hostile environment, irrespective of the contraction bias, untreated and nicotinamide pre-treated 3x*CUP1* cells were challenged in the presence and absence of copper, by a growth curve assay and a direct competition assay.

The null hypothesis of the growth curve assay was that nicotinamide pre-treated 3x*CUP1* cells would show no fitness difference, as measured by maximum growth rate, in the presence or absence of copper, compared to untreated cells of the matched condition.

3x*CUP1* cells were grown for 10 generations to saturation in complete synthetic media, then 1:1000 diluted in the same media with or without 5mM nicotinamide and re-grown to saturation over a further 10 generations. The saturated cultures were then 1:1000 diluted in complete synthetic media with and without 0.75mM $CuSO_4$, and grown for 72 hours in a 96-well plate under Growth Curve conditions. The assaying concentration of 0.75mM $CuSO_4$ was chosen from the copper adaptation assay in Fig. 4.10a, as the concentration at which growth was completely inhibited in untreated cells, but still preceded at 50% maximum growth OD in nicotinamide pre-treated cells. The maximum growth rate for untreated and nicotinamide pre-treated 3x*CUP1* cells was calculated from the OD change per hour, $\delta OD_{660nm}/\delta t(hr)$, of smoothed time-course data (Fig. 5.1a). The maximum growth rate is a measurable parameter of a culture's fitness in the surrounding environment.

In the absence of copper, nicotinamide pre-treated and untreated cells showed similar maximum growth rates in all 24 replicates. Therefore all replicates could reach their maximum growth rate in the absence of selection, and nicotinamide caused no change in growth rate. However, in the presence of 0.75mM $CuSO_4$, only a single untreated replicate (4.17%) was still able reach an equivalent, or higher maximum growth rate to that observed in the absence of copper. By comparison, 13 out of 24 (54.17%) nicotinamide pre-treated replicates could still reach an equivalent, or higher maximum growth

rate to cells in the absence of copper. Therefore nicotinamide pre-treatment caused a 13-fold increase in the number cells capable of reaching their maximum growth ability in the presence of copper. When considering the population as a whole, the adaptation advantage gained in a hostile copper environment, from nicotinamide-stimulated CNV, greatly outweighed the risk to the individual cell of nicotinamide-stimulated CNV primarily producing deleterious *CUP1* contractions.

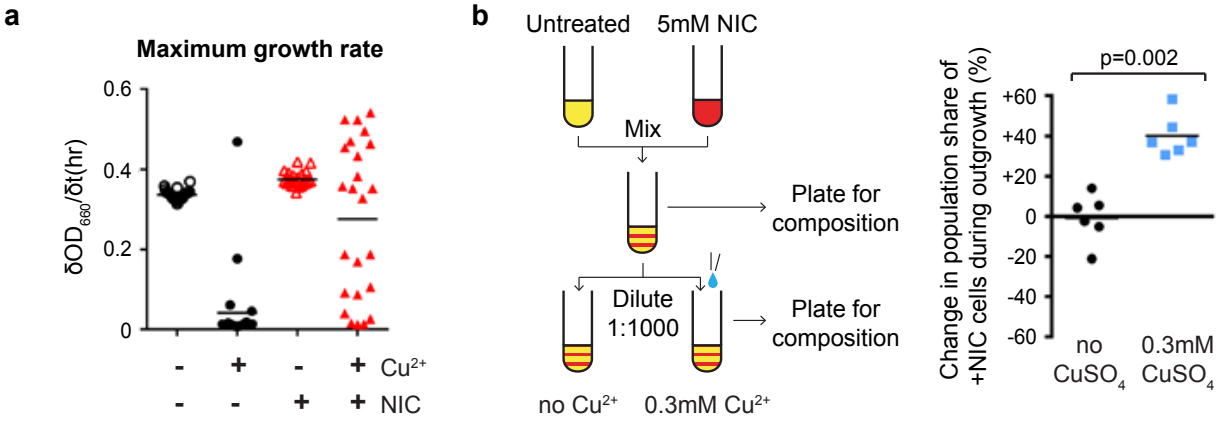


Figure 5.1: Adaptation benefit of stimulated CNV in a copper environment. **a**) Maximum growth rate in 0mM or 0.75mM $CuSO_4$ of 3x*CUP1* cells pre-treated with or without 5mM nicotinamide for 10 generations. $\delta OD_{660}/\delta t$ represents the OD change per hour. Four samples, each grown with or without 5mM nicotinamide, were each inoculated in 6 cultures for growth curve determination across 72 hours. Data was the maximum of the first derivative of smoothed OD_{600nm} time-course data for each culture. **b**) Competitive growth assay in 0mM or 0.3mM $CuSO_4$. Two populations of 3x*CUP1* cells with different selectable markers were pre-grown with or without 5mM nicotinamide, then mixed and outgrown for 10 generations in direct competition. The graph shows the change in composition of outgrowth cultures across the competition period between inoculation and saturation. *p* value was calculated by two-tailed paired *t* test, $n = 6$.

The growth curve assay used a high concentration of $CuSO_4$ to show that nicotinamide-stimulated CNV was a beneficial risk for cells to accelerate the population's adaptation to copper. However, a requirement for stimulated CNV is that cells need to still be replicating and at 0.75mM $CuSO_4$, the copper

concentration might have been too toxic for cells with non-amplified alleles to replicate. Therefore, untreated and nicotinamide-treated cells were also directly competed in 0.3mM $CuSO_4$, to test in a low copper environment where cells definitely still replicate, whether nicotinamide-stimulated CNV is a beneficial risk for cells to accelerate the population's adaptation to copper. The null hypothesis for the competition assay was that nicotinamide pre-treated cells would have no survival difference to untreated cells, in the presence or absence of copper, and so would show no change in population share of a mixed population of untreated and nicotinamide pre-treated *3xCUP1* cells.

For the competition assay, the *TRP*1 gene was replaced with the nourseothricin (Nat) antibiotic resistance cassette in 3x*CUP1* to produce the 3x*CUP1 trp*1*Δ::NatMX6* strain, which was a tryptophan (Trp) auxotroph. 3x*CUP1* wild-type and *trp*1*Δ* strains were grown for 10 generations to saturation in complete synthetic media, then 1:1000 diluted in the same media, with 3 replicates of each strain grown with and 3 replicates without 5mM nicotinamide. Cultures were grown for a further 10 generations to saturation, then cells from each wild-type treatment were 1:1 mixed with *trp*1*Δ* cells of the opposite treatment, for a total of 6 competitions (Fig. 5.1b, left panel). Since half of the competitions used nicotinamide-treated wild-type cells and the other half nicotinamide-treated *trp*1*Δ* cells, the overall result from all 6 competitions will be unbiased from any effects of the *TRP*1 deletion or Nat resistance marker.

The starting competition composition of wild-type and *trp*1*Δ* cells was determined by plating on –Trp and +Nat selection plates. Each competition was then diluted 1:1000 in 0 or 0.3mM $CuSO_4$ and grown to saturation over 10 generations. The final competition composition was then determined by plating on –Trp and +Nat selection plates. The change in the proportion of nicotinamide pre-treated cells in the final composition, from the starting composition, produced a score for the adaptation benefit or penalty of nicotinamide-stimulated CNV to a population (Fig. 5.1b, right panel).

Without copper both untreated and nicotinamide pre-treated cells performed equally well, showing that nicotinamide treatment caused no fitness difference in the absence of selection, despite stimulating CNV at *CUP1*. However,

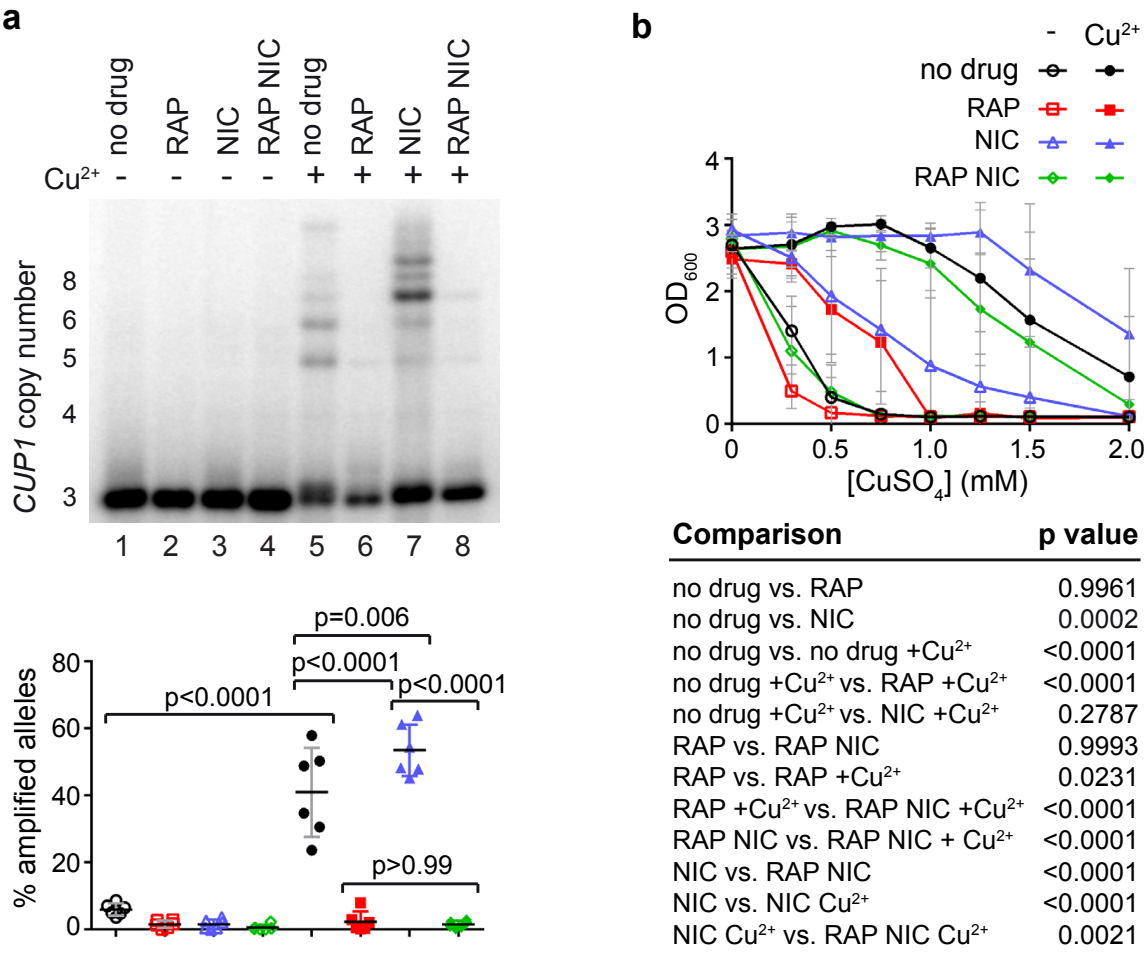
nicotinamide pre-treated cells greatly outcompeted untreated cells in all 6 competitions in 0.3mM $CuSO_4$, *p* value = 0.002. Therefore stimulated CNV was a very beneficial process for accelerating adaptation of a population of cells to a hostile environment, irrespective of the risk to the individual cell of stimulated CNV primarily producing novel alleles conferring a fitness disadvantage.

5.3 CNV and adaptation to copper is reduced by rapamycin

To determine whether stimulated CNV at *CUP1*, and by extension copper-adaptation, could be pharmacologically blocked, wild-type 3x*CUP1* yeast was grown for 10 generations with and without the TOR pathway inhibitor rapamycin. Rapamycin completely repressed copy number amplification at the rDNA in yeast with a low starting rDNA repeat number (Jack *et al.*, 2015). Rapamycin has been suggested to suppress amplification at the rDNA by inhibiting rDNA transcription (Mahajan, 1994), but also by promoting a stabilising association of Sir2 with the rDNA (Ha and Huh, 2011). Both repression of transcription, and decreased H3K56ac from greater HDAC association with CNV genes, could cause a reduction in stimulated CNV at *CUP1* and other stimulated CNV genes.

The 3x*CUP1* cells were grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted in the same media with and without 25nM rapamycin, each with or without 0.3mM $CuSO_4$ and/or 5mM nicotinamide. Cultures were grown for a further 10 generations to saturation, prior to Southern blot analysis of *CUP1* copy number (Fig. 5.2a). Rapamycin caused no change in CNV in the absence of copper (lane 1 vs. 2). However, in the presence of copper, rapamycin significantly repressed all the copper-stimulated CNV, *p* value <0.0001 (lane 5 vs. 6), to the level of rapamycin-treated cells without copper-stimulation, *p* value >0.99 (lane 2 vs. 6). Rapamycin was even able to significantly reduce all stimulated CNV in the treatment with copper and nicotinamide, where stimulated CNV was at its maximum level of induction, *p* value <0.0001 (lane 7 vs. 8). Therefore

rapamycin could significantly reduce all nicotinamide-stimulated CNV, *p* value >0.99 (lane 6 vs. 8), as well as all copper-stimulated CNV, even when there was a strong environmental pressure to stimulate CNV at *CUP1*.



Comparison	p value
no drug vs. RAP	0.9961
no drug vs. NIC	0.0002
no drug vs. no drug +Cu^{2+}	<0.0001
no drug +Cu^{2+} vs. RAP +Cu^{2+}	<0.0001
no drug +Cu^{2+} vs. NIC +Cu^{2+}	0.2787
RAP vs. RAP NIC	0.9993
RAP vs. RAP +Cu^{2+}	0.0231
RAP +Cu^{2+} vs. RAP NIC +Cu^{2+}	<0.0001
RAP NIC vs. RAP NIC + Cu^{2+}	<0.0001
NIC vs. RAP NIC	<0.0001
NIC vs. NIC Cu^{2+}	<0.0001
NIC Cu^{2+} vs. RAP NIC Cu^{2+}	0.0021

Figure 5.2: Rapamycin reduced copper- and nicotinamide-stimulated CNV and adaptation to copper. **a**) Southern analysis of *CUP1* copy number in 3x*CUP1* cells grown for 10 generations with or without 25nM rapamycin and/or 5mM nicotinamide, in the presence or absence of 0.3mM $CuSO_4$. Quantification showed the percentage of CNV events deviating from the parental copy number, *n* = 6. *p* values calculated by 1-way ANOVA for repeated measurements. **b**) Copper resistance of same cells from **a.** Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, and *n* = 6 cultures per condition, each tested at 8 $CuSO_4$ concentrations.

To determine whether rapamycin's repression of stimulated CNV could also block the population's ability to adapt to a challenging copper environment, the same cells used for Southern blot analysis were challenged in a copper adaptation assay (Fig. 5.2b). Rapamycin-treated cells showed no significant copper-adaptive difference to untreated cells, when not pre-exposed to copper, *p* value = 0.9961, showing that rapamycin treatment did not confer any adaptive difference by itself. However, when pre-exposed to copper, rapamycin significantly repressed the population's ability to adapt to a challenging copper environment, *p* value <0.0001, most likely due to rapamycin repressing stimulated CNV. However, copper pre-exposed rapamycin-treated cells did still display a significant adaptive advantage over the matched rapamycin-treated cells in the absence of copper, *p* value = 0.0231. This minor adaptation ability in copper pre-exposed rapamycin-treated cells, despite no detected stimulated CNV, was very similar to the partial adaptation observed in copper pre-treated *rtt109Δ* cells (Fig. 4.7b). As such, the results suggested that rapamycin repressed copper-stimulated CNV and adaptation to copper, through its HDAC regulating activity, reducing the global pool of H3K56ac and stabilising the *CUP1* locus.

In the absence of copper, rapamycin was also able to completely repress the nicotinamide-stimulated copper-adaptive advantage, *p* value < 0.0001 for nicotinamide vs. nicotinamide and rapamycin treated cells. Rapamycin was also able to significantly reduce the copper-adaptive phenotype of 3x*CUP1* cells pre-exposed to nicotinamide and copper, where there was the greatest adaptation advantage, *p* value = 0.0021. However, rapamycin only partially suppressed the combined nicotinamide- and copper-stimulated adaptation advantage to copper, as nicotinamide, copper and rapamycin pre-treated yeast still displayed a highly significant survival advantage over nicotinamide and rapamycin pre-treated cells in the absence of copper, *p* value <0.0001.

In summary, rapamycin could pharmacologically reduce copper- and nicotinamide-stimulated *CUP1* CNV and adaptation to hostile copper environments, similarly to an *rtt109Δ* mutant. However, copper-adaptation could still proceed in the presence of rapamycin, just at a greatly reduced rate, when a strong enough driver of stimulated CNV was present, such as the dual copper and nicotinamide treatment. If rapamycin does reduce

stimulated CNV by increasing the association of HDACs to transcriptionally active genes, it is possible that the rate of stimulated CNV is a simple competition equilibrium between rapamycin's repression of H3K56ac levels and on-going transcription's and/or nicotinamide's stimulation of H3K56ac. However, it is also possible that rapamycin could reduce copper-stimulated CNV and adaptation to a challenging copper environment by simply preventing copper from being toxic to cells, but it is currently unknown if and how rapamycin would do this.

5.4 2x*CUP1* and 1x*CUP1* constructs

This study has shown that stimulated CNV still functions in the same manner in high-copy *CUP1* systems and in a low-copy system with 3 *CUP1* repeats. To determine whether there was a limitation to stimulated CNV and the associated adaptation to a hostile copper environment in terms of a minimum requirement of starting *CUP1* repeats, strains with just a single *CUP1* repeat and with 2 genomic *CUP1* repeats were created in the same manner as the 3x*CUP1* strain was constructed in Chapter 4. It is very important to determine whether stimulated CNV, as a means to accelerate adaptation to hostile environments, still applies to 1 or 2 copy genes, as the majority of genes in the yeast genome are present at single copy or duplications.

5.5 Emergence of *CUP1* alleles with high copy number in 2x*CUP1* in copper

2x*CUP1* and 1x*CUP1* strains were grown to saturation over 10 generations in complete synthetic media, then 1:1000 diluted in the same media with and without 0.2mM $CuSO_4$, prior to Southern blot analysis (Fig. 5.3a). The 0.3mM $CuSO_4$ used for the 3x*CUP1* experiments was too toxic for the 1x*CUP1* cultures to reproducibly grow in, so the assay was optimised for the 1 and 2 *CUP1* copy strains, with 0.2mM $CuSO_4$ being the maximum copper

concentration at which all the cultures reproducibly grew to stationary phase after 3 days.

None of the 1x*CUP1* experimental cultures reached saturation in 0.2mM $CuSO_4$, but had definitely hit stationary phase after 3 days growth, since leaving the cultures for longer than 3 days did not result in any further increase in culture OD. In case 0.2mM $CuSO_4$ was still too toxic for 1x*CUP1* yeast, the cells were also cultured in 0.1mM $CuSO_4$, but displayed no improvement in culture OD after 3 days, over the 0.2mM $CuSO_4$ condition. 2x*CUP1* yeast displayed no fitness disadvantage in 0.2mM $CuSO_4$.

Southern blot analysis showed no significant increase in novel CNV alleles in response to copper in 1x*CUP1* or 2x*CUP1* cells, *p* values = 0.5529 and 0.2870 respectively. Despite the lack of significance, it was clear from the Southern blot analysis that copper still enhanced the emergence of high copy alleles in 2x*CUP1* yeast, since novel CNV alleles are detected in copper-treated 2x*CUP1* cells that are not observed in the untreated condition. The rate of CNV in 2x*CUP1* cells was greatly decreased compared to the CNV rate in 3x*CUP1* cells observed in copper, matching the earlier observation that CNV rate decreased with starting number of repeats.

Similarly, a band was observed at the limit of detection in all 6 copper-treated 1x*CUP1* replicates that was not observed in any of the untreated replicates. However, it is much harder to determine whether the band observed in copper-treated 1x*CUP1* cells was a CNV event, because 1x*CUP1* cells do not possess another *CUP1* repeat to use as a template for CNV by mismatching of DNA repeats during DSB repair. As such, the exact breakpoints in 1x*CUP1* cells, and therefore the size(s) of any copper-stimulated CNV event(s) are unknown. Assuming the band at the limit of detection in copper-treated 1x*CUP1* cells was a CNV event, the CNV rate in 1x*CUP1* yeast was even lower than the CNV rate in 2x*CUP1* yeast. The lower CNV rate in 1x*CUP1* cells is again caused by having fewer starting *CUP1* repeats, but might also be because of 1x*CUP1* cells not replicating as many times in the presence of 0.2mM $CuSO_4$, as mentioned above. Furthermore, since CNV in response to copper was not explored in 1x*CUP1* and 2x*CUP1* cells with *RTT109* deleted, it cannot be concluded whether or not the observed *CUP1* alleles with higher

than parental copy number in copper were stimulated CNV events, or selection of rare pre-existing cells with higher *CUP1* copy numbers.

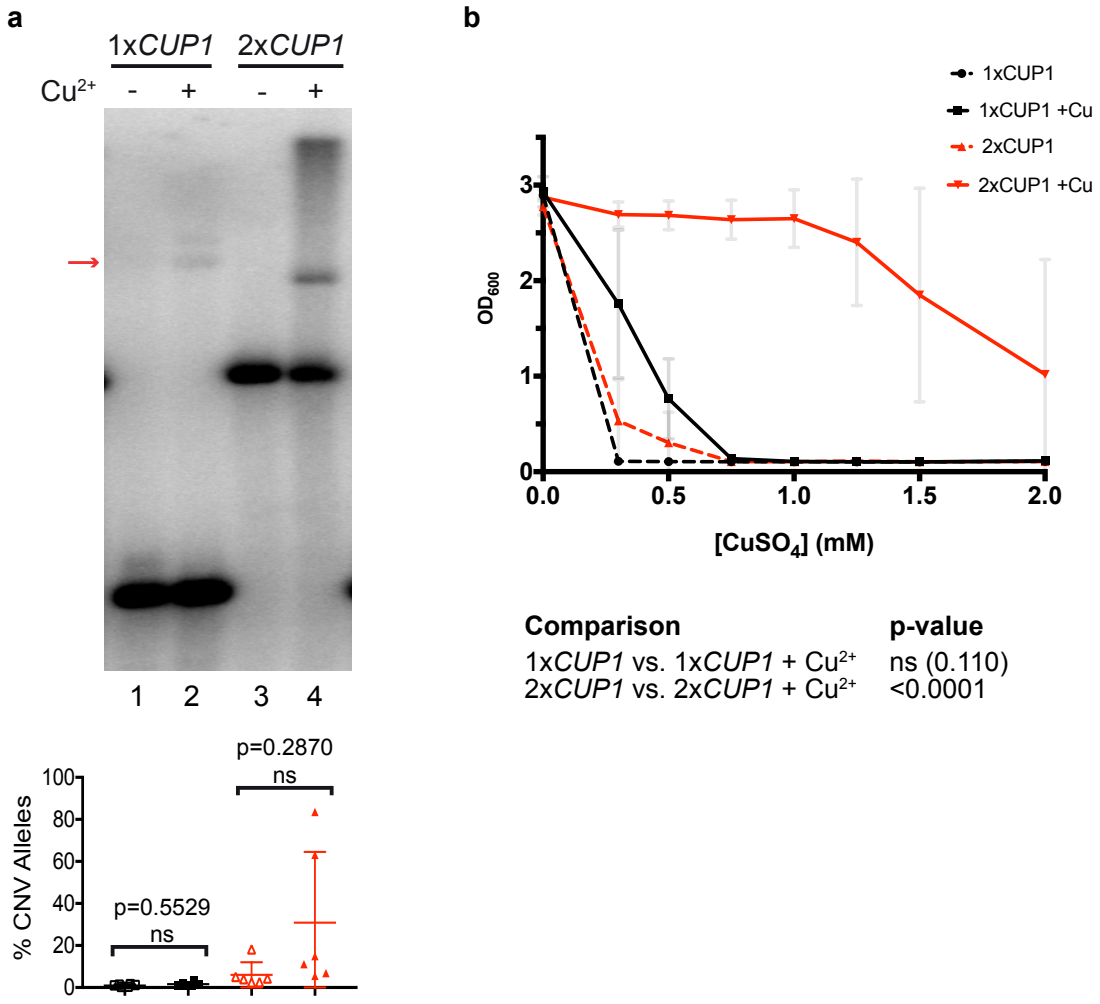


Figure 5.3: Emergence of *CUP1* amplifications and copper resistance in response to copper in 1x*CUP1* and 2x*CUP1*. **a**) Southern analysis of *CUP1* copy number in 1x*CUP1* and 2x*CUP1* cells grown for 10 generations with or without 0.2mM $CuSO_4$. Quantification showed the percentage of CNV events deviating from the parental copy number, $n = 6$. p values calculated by 1-way ANOVA for repeated measurements. **b**) Copper resistance of same cells from **a.** Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, $n = 6$ cultures per condition, each tested at 8 $CuSO_4$ concentrations.

To determine whether the decreased rate of CNV in 1x*CUP1* and 2x*CUP1* cells also caused a repression in the population's ability to adapt to a hostile

copper environment, the same cells used for the Southern blot analysis were challenged in a copper adaptation assay (Fig. 5.3b), in the same conditions used for the 3x*CUP1* cells. Copper pre-treated 2x*CUP1* cells showed a highly significant adaptation to copper over the untreated condition, *p* value <0.0001. The copper-adaptation observed in copper pre-treated 2x*CUP1* yeast was identical to that of copper pre-treated 3x*CUP1* cells, despite a decreased level of copper-stimulated CNV. Therefore, the reduced CNV rate in 2x*CUP1* caused no effect on the population's ability to adapt to a hostile copper environment. As such, the emergence of cells with high *CUP1* copy numbers in response to copper pre-treatment could effectively accelerate adaptation to a challenging copper environment, with as little as 2 starting *CUP1* repeats and irrespective of the rate of CNV, so long as CNV could occur.

Copper pre-treated 1x*CUP1* yeast showed a small, but non-significant, adaptation to copper, over the untreated condition, *p* value = 0.110. The copper-adaptive phenotype of 1x*CUP1* cells pre-treated in copper was also greatly reduced compared to 2x*CUP1* cells pre-treated in copper. This result provided evidence that the band observed at the limit of detection by Southern blot analysis of 1x*CUP1* cells in copper was a *CUP1* CNV event, which could confer partial resistance to a challenging copper environment, but was unable to reach significance because of the small number of replicates. However, the copper-adaptive phenotype of copper pre-treated 1x*CUP1* cells was weaker than the adaptation to copper observed earlier in 3x*CUP1 rtt109Δ* cells (Fig. 4.7b) and rapamycin pre-treated 3x*CUP1* cells (Fig. 5.2b), neither of which showed any copper-stimulated CNV by Southern blot analysis. This contradicts the band observed at the limit of detection in copper pre-treated 1x*CUP1* cells being a CNV event. None of the copper pre-treated 1x*CUP1* cells that partially adapted to the challenging copper environment were outgrown in the absence of selection, for Southern blot analysis, as was done for 3x*CUP1 rtt109Δ* cells (Fig. 4.7c). Therefore it cannot be concluded whether or not the partial adaptation to hostile copper phenotype in copper pre-treated 1x*CUP1* cells was indeed caused by *CUP1* CNV. In addition, in *cup1*Δ yeast, partial copper resistance was shown to be conferred by the *CRS5* gene, encoding another copper-binding metallothionein (Culotta,

Howard and Liu, 1994). Therefore the partial copper adaptation phenotype observed in 1x*CUP1* cells may also be caused by an increase in Crs5 levels, but this was not tested.

5.6 1 to 2 repeat CNV through eccDNA re-integration

There are a few possible mechanisms that 1x*CUP1* cells could use to undergo an initial duplication from 1 to 2 *CUP1* copies. This could be a replication slippage event, whereby the DNA polymerase stalls and dissociates from the replicating DNA, only to re-attach in the wrong place and re-replicate the *CUP1* region more than once. Replication slippage only requires a few bases of microhomology.

Alternatively, 1x*CUP1* yeast could undergo duplication from 1 to 2 *CUP1* repeats via a retrotransposon intermediate. For this, the retrotransposon containing the single *CUP1* repeat would be transcribed into mRNA and then reverse transcribed back to DNA and re-inserted into the genome.

A third option is for 1x*CUP1* cells to duplicate the single *CUP1* copy via re-integration of an extrachromosomal circular DNA intermediate. Extrachromosomal DNA circles (eccDNAs) are circular DNA fragments, excised from the genome during errors in DNA replication and repair, that often selectively accumulate in aged yeast cells due to their acentric nature (Windle *et al.*, 1991). eccDNAs have been suggested to be a negative sign of aging, but their role in evolution remains unknown (Sinclair and Guarente, 1997; Moller *et al.*, 2015). eccDNAs can arise in any part of the genome, even from regions with short or no repetitive sequences, as a common mutation (Moller *et al.*, 2015). However, eccDNAs arising from repetitive genomic regions, including the high-copy *CUP1* locus, become highly enriched (Moller *et al.*, 2015). Hypothetically *CUP1* eccDNAs could arise from a single genomic *CUP1* copy. Re-integration of the *CUP1* eccDNA at the native genomic locus, containing a single *CUP1* copy, would result in a duplication of *CUP1*, and be predicted to proceed by the same mechanism a plasmid re-integrates at the same genomic locus to produce a tandem array (Orr-Weaver and Szostak, 1983).

To explore *CUP1* CNV from 1 to 2 copies via a circular DNA intermediate, the *MET17* gene was deleted in 1x*CUP1* cells and then transformed with either a single-copy fake eccDNA plasmid (p*CUP1-MET25-GFP*), or an empty single-copy *MET17* plasmid (p*RS411*). The fake eccDNA contained a single genomic *CUP1* repeat with homology to the native genomic locus, as well as a *MET17* gene for selection on media lacking Methionine, and a *GFP* marker. The plasmid transformations were selected on –Met plates. The 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* and 1x*CUP1 met17Δ* p*RS411* cells were then grown for 10 generations in –Met liquid media, alongside untransformed control 1x*CUP1 met17Δ* cells in complete synthetic media.

DNA was extracted from the cultures for Southern blot analysis of *CUP1* copy number (Fig. 5.4a). The untreated controls and 1x*CUP1 met17Δ* cells transformed with the empty *MET17* plasmid showed no *CUP1* CNV. In comparison, all 4 1x*CUP1 met17Δ* replicates transformed with the fake eccDNA showed re-integration, resulting in *CUP1* duplication. This result, although not proven, provides evidence that 1 to 2-copy CNV at *CUP1* could potentially proceed via a circular DNA intermediate.

All 4 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. replicates, which showed *CUP1* CNV through re-integration of a *CUP1* circular DNA, were grown for a further 10 generations in the presence or absence of 0.3mM $CuSO_4$, to determine whether stimulated CNV could proceed, as previously shown for cells with 2 *CUP1* copies, when the second *CUP1* repeat was formed by eccDNA re-integration (Fig. 5.4b). In the presence of copper, 3 out of 4 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. replicates (75%) observed the emergence of cells with 3 *CUP1* repeats. The cells with 3 copies of *CUP1* could have arisen by copper-stimulated CNV, but again selection of rare pre-existing cells with 3 repeats cannot be excluded as the experiment was not attempted with *RTT109* deleted. Nonetheless this result suggests the possibility that stimulated CNV might be able to proceed in cells with 2 copies of *CUP1*, even when the second repeat comes from an eccDNA re-integration.

Interestingly, the only CNV event detected in response to copper was the +1 repeat amplification to 3 *CUP1* copies. This limited rate of CNV in copper was unlike that observed in 2x*CUP1* cells (Fig. 5.3a). The limited CNV in 1x*CUP1*

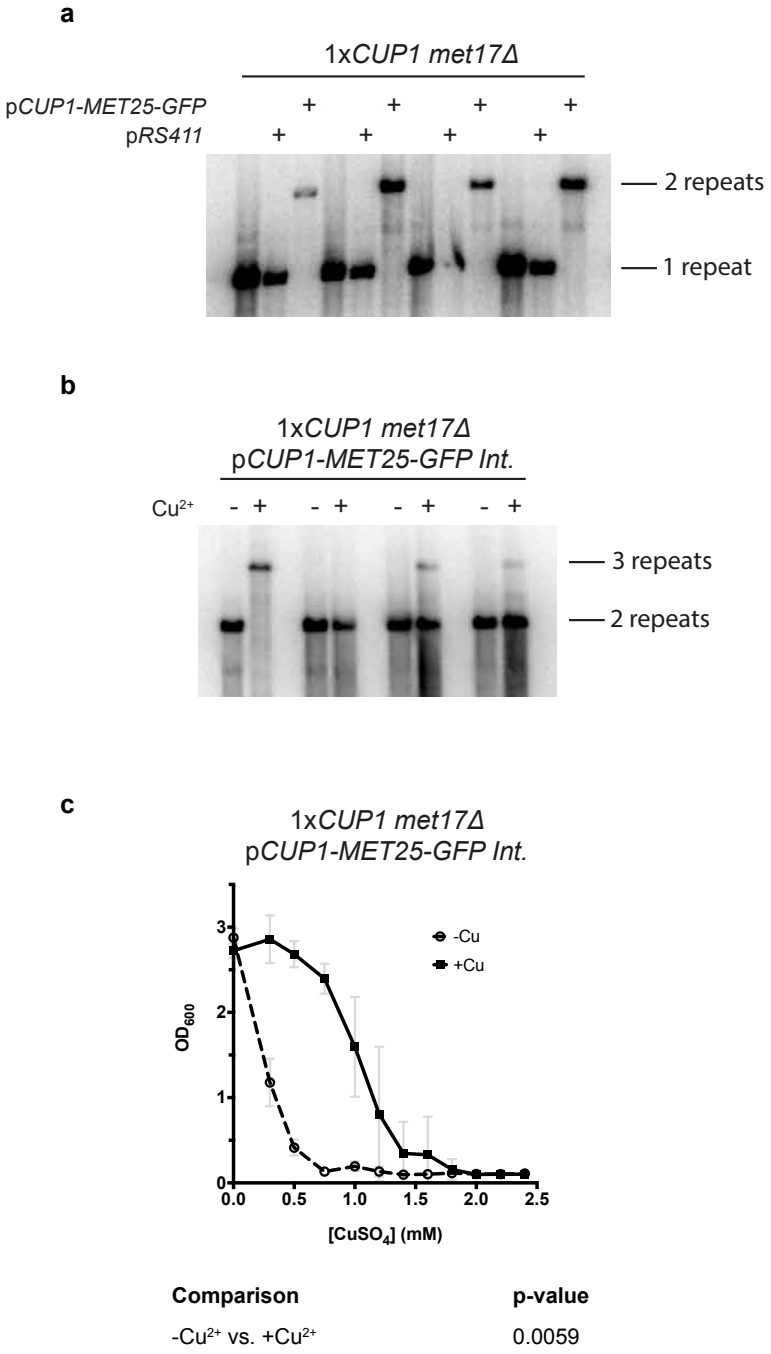


Figure 5.4: Amplification from single copy through eccDNA reintegration. **a**) Southern blot analysis of *CUP1* copy number in 1x*CUP1 met17Δ* cells, transformed with a fake eccDNA p*CUP1-MET25-GFP* or an empty *MET* p*RS411* plasmid, grown for 10 generations in Methionine drop-out media, *n* = 4 per sample. **b**) Southern blot analysis of *CUP1* copy number in 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* integrated cells from **a**, grown for 10 generations, with and without 0.3mM $CuSO_4$, *n* = 4 per sample. **c**) Copper resistance of same cells from **b**. Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{600nm} was plotted, error bars represent ±1 SD, *n* = 4 cultures per condition, each tested at 12 $CuSO_4$ concentrations.

met17Δ p*CUP1-MET25-GFP* Int. cells might be because the 2 genomic *CUP1* repeats were now separated by a greater distance, due to re-integration of the entire fake eccDNA, including marker and backbone. Therefore the *CUP1* repeats are further away in genetic distance for successful mismatched repair to occur frequently. It could also be that amplifying the entire re-integrated eccDNA sequence upon exposure to copper was inefficient, and therefore CNV was limited to the fewest events required to confer a beneficial adaptation to the environmental copper. Alternatively, the reduced rate of CNV in 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. cells in copper, might be due to a slower division rate than in 2x*CUP1* cells, and so 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. cells might have undergone fewer generations in copper for additional *CUP1* amplifications to occur, and be detected by Southern blot analysis. The division rate of 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. and 2x*CUP1* cells was not tested for.

The same 1x*CUP1 met17Δ* p*CUP1-MET25-GFP* Int. cells grown with and without 0.3mM $CuSO_4$ for Southern blot analysis (Fig. 5.4b) were diluted in fresh complete synthetic media and challenged in a copper adaptation assay (Fig. 5.4c). The copper pre-treated cells displayed a highly significant adaptive advantage over the untreated condition, *p* value = 0.0059, showing that the emergence of CNV in cells with an integrated eccDNA in copper could accelerate adaptation to a challenging copper environment.

5.7 1x*CUP1* 1x*SFA1* construct

Direct transcriptionally stimulated CNV and adaptation to copper in 1x*CUP1* cells was very hard to detect by Southern blot and copper adaptation analysis (Fig. 5.3). A potential explanation for this difficulty was that 0.2mM $CuSO_4$ could not provide a strong enough selection pressure on 1x*CUP1* cells to drive stimulated CNV to detectable levels by Southern blot analysis. However, simply treating 1x*CUP1* cells with higher $CuSO_4$ concentrations was not feasible because of copper toxicity. Therefore a more sensitive assay was required for detection of CNV from single copy genes.

Narayanan *et al*. developed a plate-based assay for detection of gross chromosomal rearrangements (GCRs), which utilised the gene dosage-dependent tolerance to copper and formaldehyde conferred by a reporter cassette containing the *CUP1* and *SFA1* genes (Narayanan *et al.*, 2006). *CUP1* copy number regulates copper resistance and *SFA1* copy number regulates resistance to formaldehyde. When combined as a *CUP1-SFA1* reporter, initial selection on media with a high concentration of copper, followed by replica plating onto media with a high concentration of formaldehyde, enabled a very tight selection for cells with amplified *CUP1* and *SFA1* genes as having undergone GCRs. The *CUP1-SFA1* system used by Narayanan *et al*. in haploid cells enabled the detection of 5- to 13-fold extrachromosomal amplifications of *CUP1* and *SFA1*, but also detected intrachromosomal amplifications, which were highly unstable and often gave rise to secondary rearrangements upon propagation (Narayanan *et al.*, 2006). Zhang *et al*. further optimised this plate-based assay for the selection of low-order gene amplification events, such as 1 to 2 copy duplications in haploid cells and 2 to 3 copy CNV in diploids (Zhang *et al.*, 2013). The study also confirmed 3 dominant/gain-of-function point mutations in the *SFA1* coding sequence (V208I, A303T, and M283I) that caused hyper-activation of the encoded formaldehyde dehydrogenase enzyme and elevated resistance to formaldehyde (Zhang *et al.*, 2013).

To improve upon the sensitivity of our current Southern blot approach for detecting whether 1 to 2-copy CNV was stimulated in response to the environment, a 1x*CUP1*-1x*SFA1* reporter system was introduced into the MEP for a plate-based screening approach, similar to that described by Zhang *et al*.. It was predicted that selection on copper and formaldehyde would provide a much stronger and tighter selection for cells that have amplified *CUP1* from a single copy.

Construction of the 1x*CUP1*-1x*SFA1* reporter system required a single copy of the *SFA1* gene to be situated adjacent to a single copy of the native *CUP1* gene, so that both genes amplify together during a stimulated CNV event. The final MEP 1x*CUP1* 1x*SFA1* heterozygote diploid was created as follows. Firstly, *CUP1* was deleted with *TRP1* in both MEP wild-type *MAT*a and α haploids. *SFA1* was then deleted with a *KanMX6* marker in both MEP

cup1Δ::TRP1 haploids. The pRH19 plasmid was created, containing a single genomic *CUP1* repeat, a *SFA1* repeat with the V208I point mutations that caused hyper-activation of the encoded formaldehyde dehydrogenase in Zhang *et al*., and a flanking *ADE2* marker gene. pRH19 was *SacI*-digested and transformed into the MEP *cup1Δ::TRP1 sfa1Δ::KanMX6 MAT*a haploid at the native *CUP1* locus, selecting for white Ade2-proficient, tryptophan auxotrophic cells. All transformation steps were confirmed by genotyping PCRs. The MEP 1x*CUP1* 1x*SFA1 sfa1Δ::KanMX6 MAT*a haploid was then mated to a MEP *cup1Δ::TRP1 sfa1Δ::KanMX6 MAT*α haploid and selected on –Ade –Trp drop-out media plates for white cells, to produce the MEP 1x*CUP1* 1x*SFA1* heterozygote diploid. A schematic of the native *CUP1* locus in MEP 1x*CUP1* 1x*SFA1* heterozygote cells is shown in Fig. 5.5.

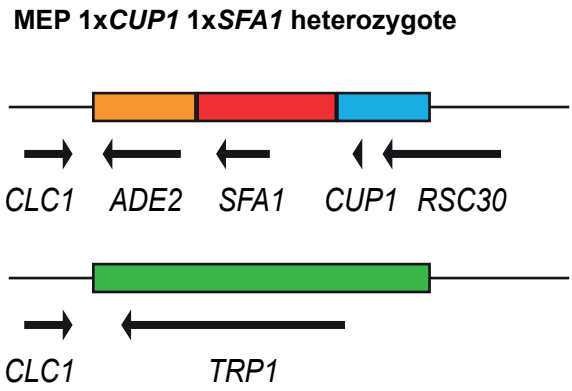


Figure 5.5: Schematic of the native *CUP1* locus in MEP 1x*CUP1* 1x*SFA1* heterozygote cells. The native *CUP1* gene was replaced with a 1x*CUP1*-1x*SFA1-ADE2* construct in MEP *MAT*a haploid cells and mated to MEP *MAT*α haploid cells where the native *CUP1* gene was replaced with a *TRP1* auxotrophic selectable marker. The native *SFA1* gene was deleted with the Kanamycin resistance cassette at both alleles.

5.8 Stimulated CNV from single copy

As seen earlier in Fig. 5.3, Southern blot analysis was not sensitive enough to detect stimulated CNV from single-copy 1x*CUP1* cells. Therefore a viability assay was used to improve the sensitivity of detection of stimulated CNV from

single-copy. Using a strain with a single copy of the *CUP1* and *SFA1* gene, and a strain with two copies of *CUP1* and *SFA1*, a spotting assay was used to determine the concentration combination of copper and formaldehyde, where the strain with 2 repeats survived but the strain with 1 repeat failed to grow. The best combination was found to be 0.2mM $CuSO_4$ and 1.6mM formaldehyde (Fig. 5.6a). Plating MEP 1x*CUP1* 1x*SFA1* heterozygote cells on 0.2mM $CuSO_4$ and 1.6mM formaldehyde could therefore distinguish between cells with parental copy number and cells that have stimulated *CUP1* and *SFA1* CNV to 2+ repeats.

MEP 1x*CUP1* 1x*SFA1* heterozygote cells were grown to early log-phase in complete synthetic media, then $0.25x10^7$ cells were inoculated into 125ml complete synthetic media containing β-estradiol, with and without 0.7mM formaldehyde. β-estradiol activated the MEP and enabled direct observation of the CNV rate on a defined cohort of cells, without selection through progeny. Cultures were grown for 24 hours and $1x10^6$ cells were plated on –Trp drop-out plates containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde. It was necessary to plate on –Trp drop-out plates, so as to only capture true, allele-specific, CNV events from single-copy and not also loss of heterozygosity (LOH) events.

The percentage of cells resistant on 0.2mM $CuSO_4$ 1.6mM formaldehyde plates were calculated for each condition, accounting for the different viabilities across the treatments. In the untreated condition, 0.053±0.031% of MEP 1x*CUP1* 1x*SFA1* heterozygote cells were resistant to 0.2mM $CuSO_4$ and 1.6mM formaldehyde, increasing to 0.143±0.052% with formaldehyde stimulation (Fig. 5.6b). This was a 2.7-fold increase in resistant cells with formaldehyde stimulation, which was highly significant, *p* value = 0.0049. This result provided preliminary evidence for transcriptionally stimulated CNV being possible even from single-copy genes.

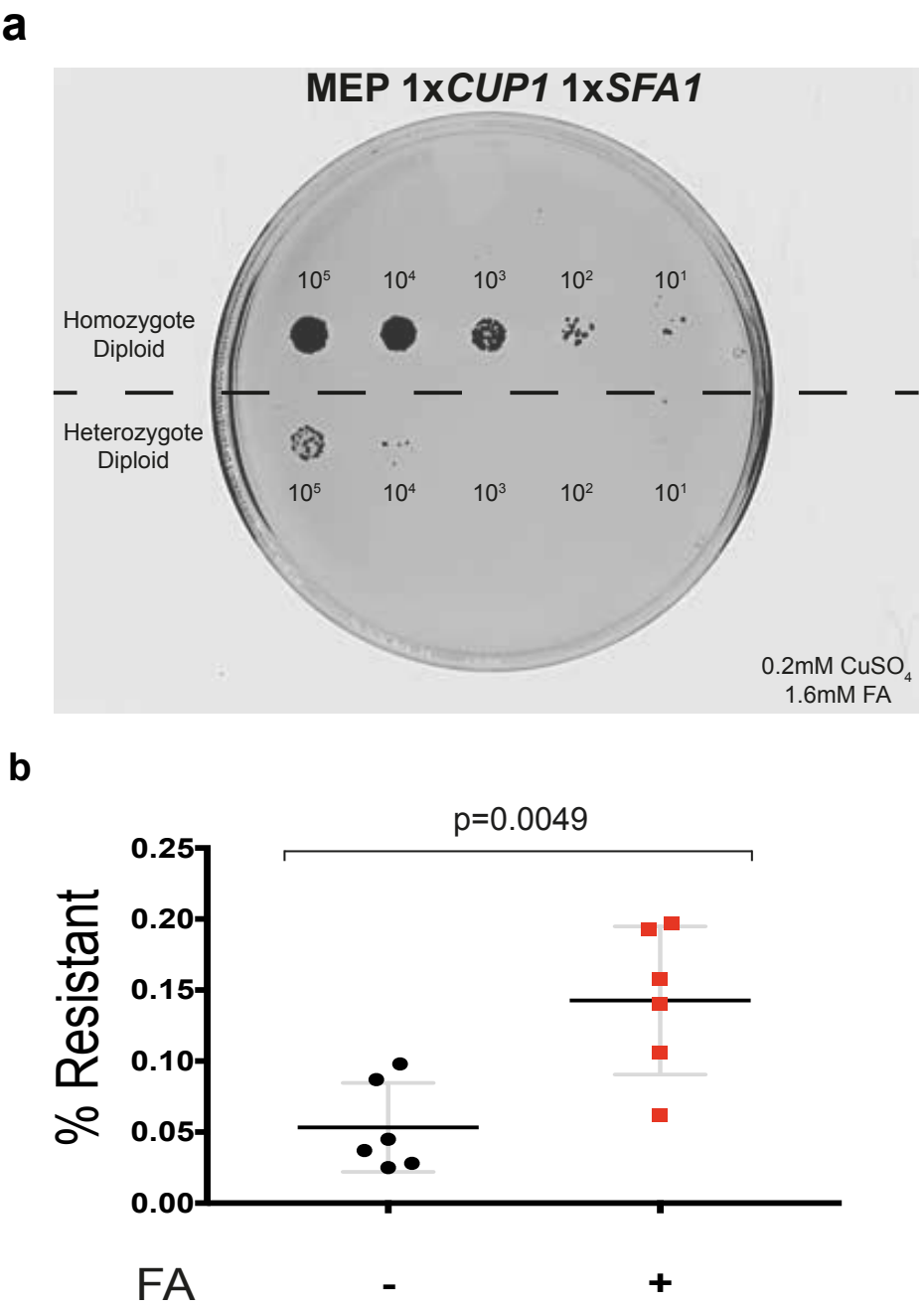


Figure 5.6: Stimulated CNV from single-copy. **a**) Spotting assay plate containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde, of MEP 1x*CUP1* 1x*SFA1* diploid (top) and heterozygote (bottom) in 10-fold serial dilutions. **b**) Percentage of 0.2mM $CuSO_4$ and 1.6mM formaldehyde resistant cells from $0.25x10^7$ early log-phase MEP 1x*CUP1* 1x*SFA1* heterozygote cells grown for 24 hours in complete synthetic media, with and without 0.7mM formaldehyde, in β-estradiol to activate the MEP. Cells were plated on –Trp drop-out plates at t=0 and t=24 hours to obtain culture viability scores. After 24 hours, the cultures were plated on –Trp drop-out plates containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde. A *p* value was calculated by unpaired two-tailed t test of percentage resistant cells, adjusted for differences in cell viability between the treatments, $n = 6$, pooled from two experiments.

5.9 Creating 3x*SFA1* strain

So far the copper-resistance *CUP1* gene has been used as a model locus for transcriptionally stimulated CNV in high-copy systems and low-copy systems starting from at least 2 genomic repeats. CNV at *CUP1* could even proceed from 1 to 2 copies via a circular intermediate, with some evidence for copper potentially stimulating further CNV in cells with an integrated eccDNA. However, to determine whether stimulated CNV could apply beyond just the *CUP1* locus, the mechanism must be tested in another model locus.
At the start of this study, the prediction was made that genes that could undergo stimulated CNV required an upstream RFS site and a bi-directional promoter, where one of the transcripts interferes with the replication machinery at the RFS site. The formaldehyde-resistance gene *SFA1* was another "γH2A gene" identified from the γH2A ChIP-seq data set, as a candidate gene for having an upstream RFS site (Hull *et al.*, 2017). *SFA1* encodes a formaldehyde dehydrogenase, which caused a hyper-resistance phenotype to formaldehyde when overexpressed on a multi-copy plasmid (Wehner, Rao and Brendel, 1993).
The *SFA1* gene is naturally present at single copy in the *Saccharomyces cerevisiae* genome, but given the experimental difficulty of detecting stimulated CNV from single copy at the *CUP1* gene, a 3 *SFA1* repeat construct was created to test for formaldehyde-stimulated CNV. To construct the 3x*SFA1* strain, first the native *SFA1* gene in YRH15 was deleted using the Kanamycin (Kan) resistance cassette. The pJH312 plasmid was then constructed, containing 3 genomic *SFA1* repeats in tandem, flanked by a single *CUP1* repeat on one end of the *SFA1* repeats, and an *ADE2* marker gene on the other end. Flanking sequence homology for the native *CUP1* locus was inserted at the ends of *CUP1*-3x*SFA1-ADE2*, so that the whole *Sac*I-digested construct could then be transformed into the native *CUP1* locus. Positive transformants were selected for as white Ade2-profficient colonies and confirmed by genotyping PCRs. Fig. 5.7 shows a schematic of the *CUP1* locus in 3x*SFA1*.

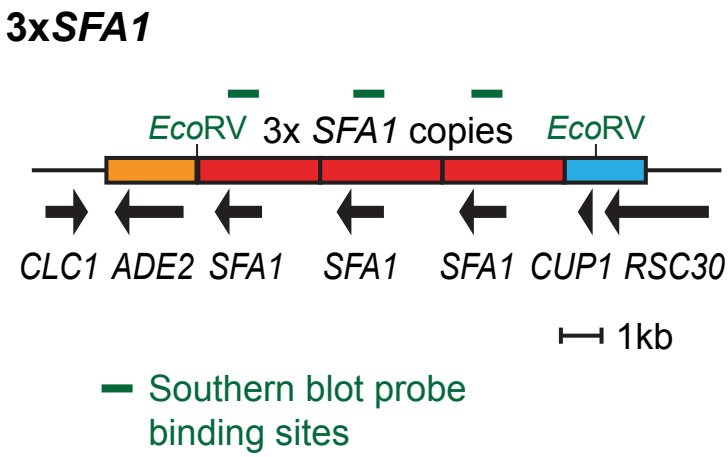


Figure 5.7: Schematic of 3x*SFA1* locus. The 3x*SFA1* construct was inserted at the endogenous *CUP1* locus. Blue boxes indicate *CUP1* repeats, red boxes indicate *SFA1* repeats, and the orange box indicates the *ADE2* marker. The reading frame of *RSC30* was maintained across the construct boundary. Restriction enzymes for Southern analysis are shown in green along with probe locations.

5.10 Formaldehyde induces *SFA1* transcripts

Before beginning to test whether formaldehyde-stimulated CNV occurs in 3x*SFA1*, it first needed to be established if the *SFA1* promoter was bi-directional and induced by formaldehyde, thereby fulfilling the predicted requirements for a stimulated CNV gene. To test transcript induction at the *SFA1* promoter, 3x*SFA1* yeasts were grown to mid-log, in the presence or absence of 0.9-1mM formaldehyde, for Northern blot analysis (Fig. 5.8). The formaldehyde concentration used in each experiment had to be slightly adjusted because of variation between batches of formaldehyde. The concentration of formaldehyde used shall simply be called ~1mM from now onwards.

Northern blot analysis identified a lowly expressed *SFA1* sense transcript and could not detect an *SFA1* upstream antisense transcript, in the absence of ~1mM formaldehyde. However, formaldehyde significantly induced expression of the *SFA1* sense transcript, *p* value <0.0001, but also stimulated the production of the *SFA1* upstream antisense transcript, *p* value <0.0001.

Therefore the *SFA1* promoter was bi-directional and highly induced in response to environmental formaldehyde. As such, *SFA1* fulfils the predicted criteria for it to be a gene capable of stimulated CNV.

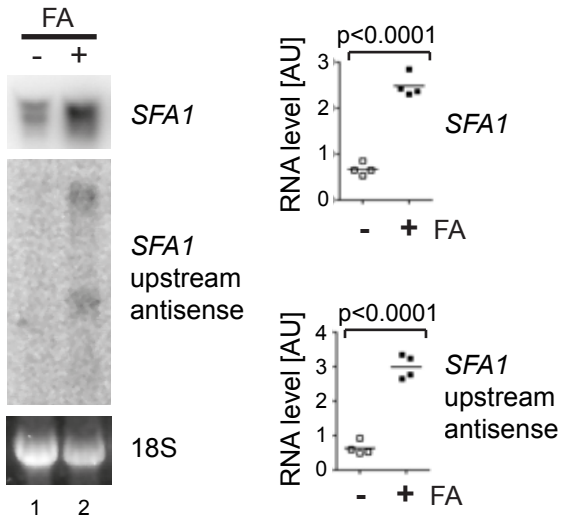


Figure 5.8: Formaldehyde induced *SFA1* transcripts. Northern analysis of *SFA1* sense and upstream antisense transcript induction, with or without a 4-hour exposure to ~1mM formaldehyde. The ribosomal *18S* rRNA was shown as a loading control. Quantification showed the levels of the indicated RNA species in arbitrary units, $n = 4$. p values were calculated by a two-tailed paired t test. Locations of probes within the *SFA1* repeat were shown in Fig. 5.7.

5.11 *rtt109Δ* reduces the emergence of *SFA1* amplifications in formaldehyde

To determine whether formaldehyde could stimulate CNV at *SFA1*, 3x*SFA1* wild-type and *rtt109Δ* cells were grown to mid-log in complete synthetic media, then 1:8000 diluted in the same media, with and without ~1mM formaldehyde and re-grown to saturation over 17 generations, prior to Southern blot analysis of *SFA1* copy numbers (Fig. 5.9).
In the absence of formaldehyde, the *SFA1* locus displayed no novel CNV alleles that deviated from the parental 3 repeat copy number. Whereas,

~1mM formaldehyde-treated 3x*SFA1* cells, showed a highly significant increase in the emergence of novel CNV alleles, *p* value = 0.002. All of the novel CNV alleles detected were amplifications, which confirmed that high-copy *SFA1* alleles were under positive selection in formaldehyde. However, the ~1mM formaldehyde selection pressure on 3x*SFA1* cells was less than 0.3mM $CuSO_4$ was on 3x*CUP1* cells, as most of the amplifications detected at *SFA1* were of lower copy number than observed at *CUP1*, despite having undergone more generations.

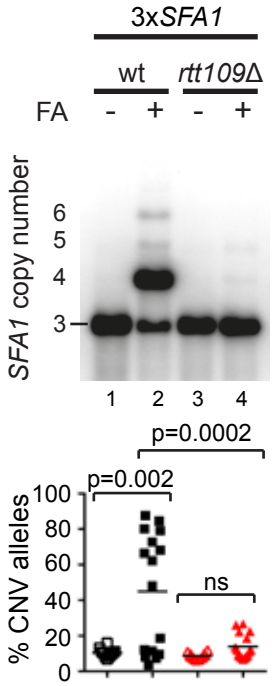


Figure 5.9: H3K56ac was required for CNV at *SFA1*. Southern analysis of *SFA1* copy number in 3x*SFA1* wild-type (wt) and *rtt109*Δ cells grown for 17 generations with or without ~1mM formaldehyde (FA). Quantification showed the percentage of alleles deviating from the parental copy number; *n* = 10 for untreated samples, *n* = 16 for FA-treated samples. *p* values calculated by 1-way ANOVA. A correction was applied to the quantification to account for the differing number of probe-binding sites in the amplified alleles.

Just as in the *CUP1* system, deletion of *RTT109* in 3x*SFA1* cells significantly reduced the emergence of *SFA1* alleles with higher than parental copy number in formaldehyde, *p* value = 0.0002, to the level in the matched

untreated cells, *p* value = 0.9152. This result provided evidence for CNV being stimulated at *SFA1* by formaldehyde, in a mechanism that was dependent upon H3K56ac, analogous to transcriptionally stimulated CNV at *CUP1*. However, +1 and +2 repeat amplifications in formaldehyde-treated 3x*SFA1 rtt109Δ* cells could still be faintly detected by Southern blot analysis, despite the highly significant reduction in stimulated CNV. These minor amplification events in formaldehyde-treated 3x*SFA1 rtt109Δ* cells must arise through a H3K56ac-independent process, or be produced by another much less efficient and as yet unidentified H3K56 acetyltransferase, or have simply pre-existed as extremely rare cells in the starting population that can only be detected after 17 generations growth in highly selective formaldehyde.

5.12 Stimulated CNV at *SFA1*

To confirm whether CNV at *SFA1* is indeed also stimulated by a H3K56ac-dependent mechanism, 3x*SFA1* cells were grown to mid-log phase in complete synthetic media, then 1:8000 diluted in the same media, with and without ~1mM formaldehyde, but also in the presence or absence of 5mM nicotinamide. Cultures were re-grown for 17 generations to saturation, prior to Southern blot analysis for *SFA1* copy number (Fig. 5.10).
As in Fig. 5.9, 3x*SFA1* cells in the absence of formaldehyde observed no novel CNV alleles that deviated from the parental 3 repeat copy number, with ~1mM formaldehyde-treated 3x*SFA1* cells showing a highly significant increase in the emergence of novel CNV alleles, *p* value = 0.001.
Nicotinamide stimulated CNV in 3x*SFA1* in the form of -1 and -2 repeat contractions, which showed that H3K56ac was still important for stimulated CNV at *SFA1*. Nicotinamide-stimulated CNV at *CUP1* only produced the -1 repeat contraction and not the additional -2 repeat contraction that was also detected at *SFA1*, suggesting that nicotinamide was a stronger driver of stimulated CNV at *SFA1* than *CUP1*. However, the combined nicotinamide and formaldehyde treatment did not select against the nicotinamide-stimulated contractions at *SFA1*, as the -1 and -2 repeat bands were still detected as strongly in the combined treatment and in the treatment with

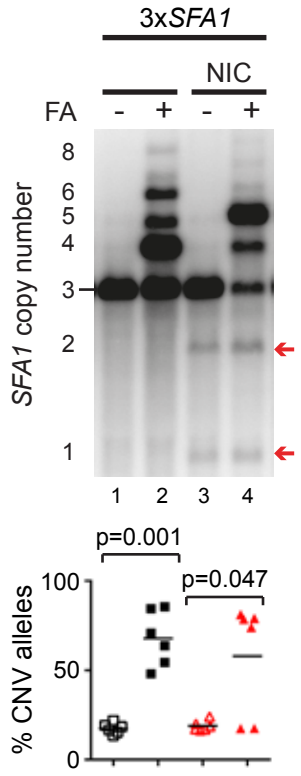


Figure 5.10: Nicotinamide-stimulated CNV at *SFA1*. Southern analysis of *SFA1* copy number in 3x*SFA1* grown for 17 generations with or without ~1mM formaldehyde (FA) and with or without 5mM nicotinamide (NIC). Red arrows identify the -1 and -2 repeat bands. Quantification showed the percentage of alleles deviating from the parental copy number, *n* = 6. *p* values calculated by 1-way ANOVA. A correction was applied to the quantification to account for the differing number of probe-binding sites in the amplified alleles.

nicotinamide alone. Also the combined nicotinamide and formaldehyde treatment did not enhance the emergence of novel amplification alleles at *SFA1*, with no significant difference in the total CNV rate or the emergence of high-copy alleles, over treatment with formaldehyde alone. Therefore unlike the *CUP1* locus, where raising global H3K56ac levels through nicotinamide treatment enhanced the copper-stimulated CNV rate, nicotinamide caused no additive effect with formaldehyde on the emergence of novel CNV alleles at *SFA1*. The P_{GAL}-*HA* system in 2% galactose in Chapter 3 (Fig. 3.8), also displayed no increase in stimulated CNV with nicotinamide treatment, when the promoter was induced strongly enough to saturate the local H3K56ac level. Therefore nicotinamide may also not be able to enhance the rate of

emergence of novel CNV alleles at *SFA1* because of strong induction of the *SFA1* promoter recruiting such a high level of H3K56ac to the *SFA1* locus, that the local H3K56ac level became saturated. However, the inability of nicotinamide to further enhance the emergence of novel *SFA1* alleles in formaldehyde could also be because the formaldehyde selection pressure was too weak to drive the production of CNV alleles with very high *SFA1* copy numbers. This scenario was supported by the nicotinamide-stimulated contractions being maintained in the combined treatment with nicotinamide and formaldehyde, where it would be expected that contractions would be selected against.

5.13 Creating the P_{GAL}-*SFA1* constructs

To confirm that formaldehyde-stimulated CNV in 3x*SFA1* was a transcriptional effect required a high-copy *SFA1* equivalent to the P_{GAL}-*HA* strain used for *CUP1*. Two galactose-driven *SFA1* constructs were designed that replaced the native promoter of *UGX2*, a divergent gene upstream of *SFA1*, with P_{GAL1}-*GFP* in all 3 *SFA1* repeats (Fig. 5.11a-b). The difference between the two galactose-driven constructs was the orientation of the inserted P_{GAL1}-*GFP*. Insertion of the *PFG-*$_{LAG}$*P* orientation caused a partial deletion of the upstream *UGX2* gene. Both constructs were *SacI*-digested and transformed into YRH12, followed by FOA selection for loss of the single-copy *CUP1* plasmid, as in the construction of the 3x*SFA1* strain (Fig. 5.7). The transformations produced the 3x*UGX2-*P_{GAL}*-GFP-SFA1* and 3x*PFG-*$_{LAG}$*P-SFA1* strains.
It was initially assumed that transcription through the *SFA1* RFS site would be enough to stimulate CNV at *SFA1*, irrespective of the where and in what orientation the bi-directional promoter was placed. However, the initial experiment only showed CNV in 3x*PFG-*$_{LAG}$*P-SFA1* cells. To determine whether the promoter orientation, or the disruption of the upstream *UGX2* gene was responsible for CNV, a third galactose-driven construct was created (Fig. 5.11c). This third strain was 3x*UGX2-*$_{LAG}$*P-GFP-SFA1* and maintained the same promoter orientation as the construct that previously showed CNV, but no longer disrupted the upstream *UGX2* ORF.

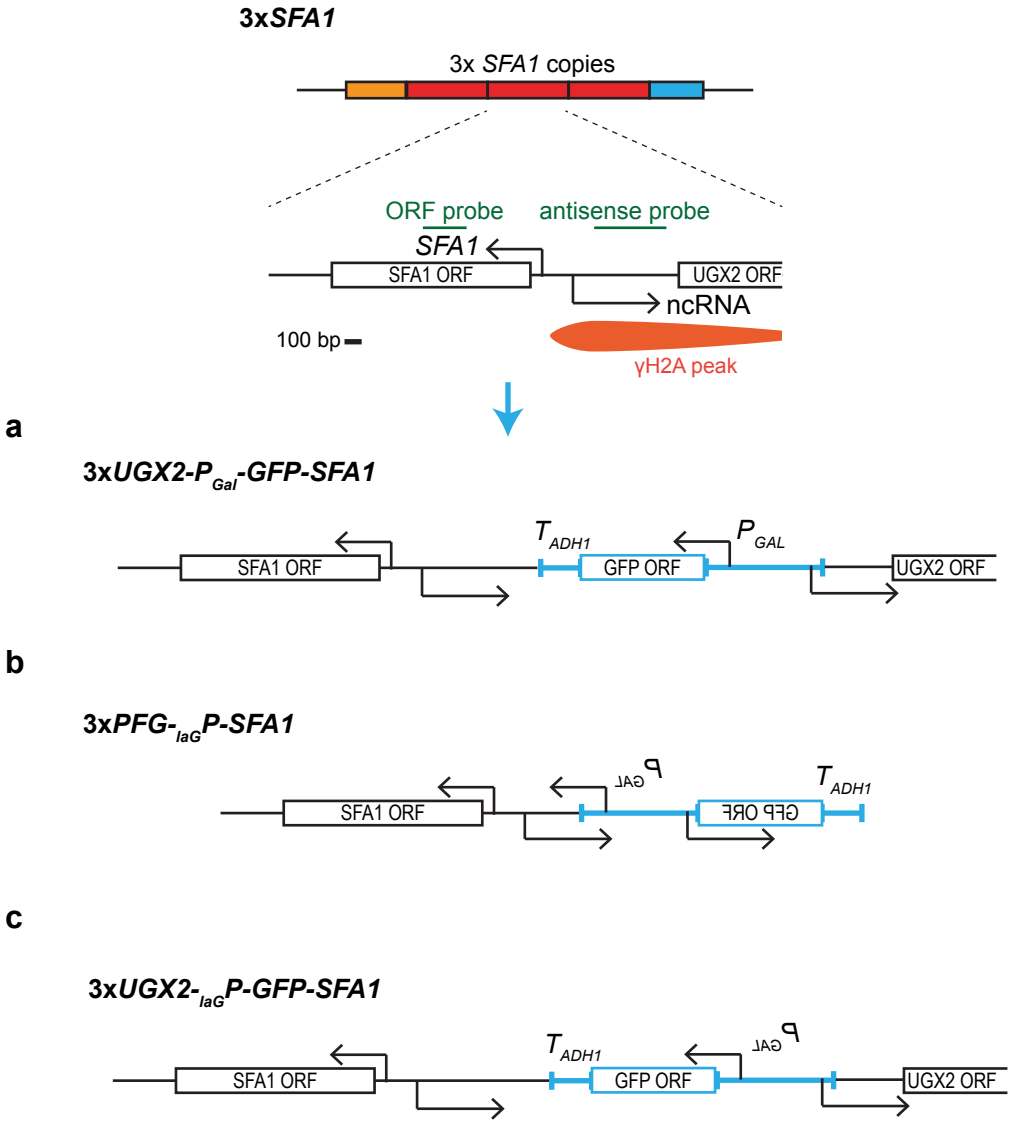


Figure 5.11: Schematic of the galactose-driven *SFA1* constructs. P_{GAL}-*GFP* was inserted into all 3 *SFA1* repeats. **a**) 3x*UGX2-*P_{GAL}*-GFP-SFA1* cells inserted the P_{GAL} promoter in the forward orientation, with the *GFP* ORF running into the *SFA1* gene. **b**) 3x*PFG-*$_{LAG}$*P-SFA1* cells inserted the P_{GAL} promoter in the reverse orientation, with the *GFP* ORF deleting the upstream *UGX2* gene. **c**) 3x*UGX2-*$_{LAG}$*P-GFP-SFA1* cells inserted the P_{GAL} promoter in the reverse orientation, with the *GFP* ORF running into the *SFA1* gene.

From the results with the P_{GAL}-*HA* strain in Chapter 3 (Fig. 3.3b), where transcriptionally stimulated CNV in the absence of selection only produced contractions, it was assumed the galactose-driven *SFA1* strains would also only show contractions when assayed by Southern blot. Earlier results also showed that the rate of stimulated CNV was greatly reduced in low-copy

systems because of fewer starting repeats. As such, it was decided best to test transcriptionally stimulated CNV at *SFA1* in a high-copy system. When making the P_{GAL}-*HA* strain, it fortuitously amplified to 17 repeats upon transformation, but this was not the case for the galactose-driven *SFA1* strains. Therefore, the galactose-driven *SFA1* strains were driven to high-copy through selective growth in ~1mM formaldehyde and isolated by re-streaking for single cells. The highest *SFA1* copy number strain isolated in the wild-type was 7x*SFA1*. The highest *SFA1* copy number strain isolated in amplified 3x*UGX2-*P_{GAL}*-GFP-SFA1* was a 10 repeat strain, but a 7 repeat strain was also isolated, for having a matching copy number to the 7x*SFA1* wild-type. The closest *SFA1* copy number strain to the 7x*SFA1* wild-type, isolated in amplified 3x*PFG-*$_{LAG}$*P-SFA1*, was a 6 repeat strain. Finally, the highest *SFA1* copy number mutant isolated from amplified 3x*UGX2-*$_{LAG}$*P-GFP-SFA1* cells was a 12 repeat strain.

5.14 Transcriptionally stimulated CNV at *SFA1*

To determine whether CNV was transcriptionally stimulated at *SFA1*, the 7x*SFA1* wild-type, 6x*PFG-*$_{LAG}$*P-SFA1*, and 7x*UGX2-*P_{GAL}*-GFP-SFA1* strains were grown for 10 generations to saturation in complete synthetic 2% glucose media, then 1:4000 diluted in complete synthetic 2% glucose and complete synthetic 2% galactose media and grown for a further 12 generations to saturation. Southern blot analysis showed that 7x*SFA1* cells did not undergo CNV in either 2% glucose or 2% galactose (Fig. 5.12a). The 6x*GFP-*$_{LAG}$*P-SFA1* strain also showed no CNV when grown in non-inducible glucose. However, when the P_{GAL1} promoter was induced in galactose, 6x*GFP-*$_{LAG}$*P-SFA1* cells displayed a significant increase in novel CNV alleles, *p* value = 0.0338. This provided evidence for CNV at *SFA1* being transcriptionally stimulated and not a mutagenic effect of formaldehyde treatment. As expected, the 7x*UGX2-*P_{GAL}*-GFP-SFA1* cells also showed no stimulated CNV alleles when grown in non-inducible glucose, but unexpectedly, no significant stimulation of CNV was observed in 7x*UGX2-*P_{GAL}*-GFP-SFA1* cells in inducible galactose. *De novo* CNV alleles were only observed in 1 of the 4

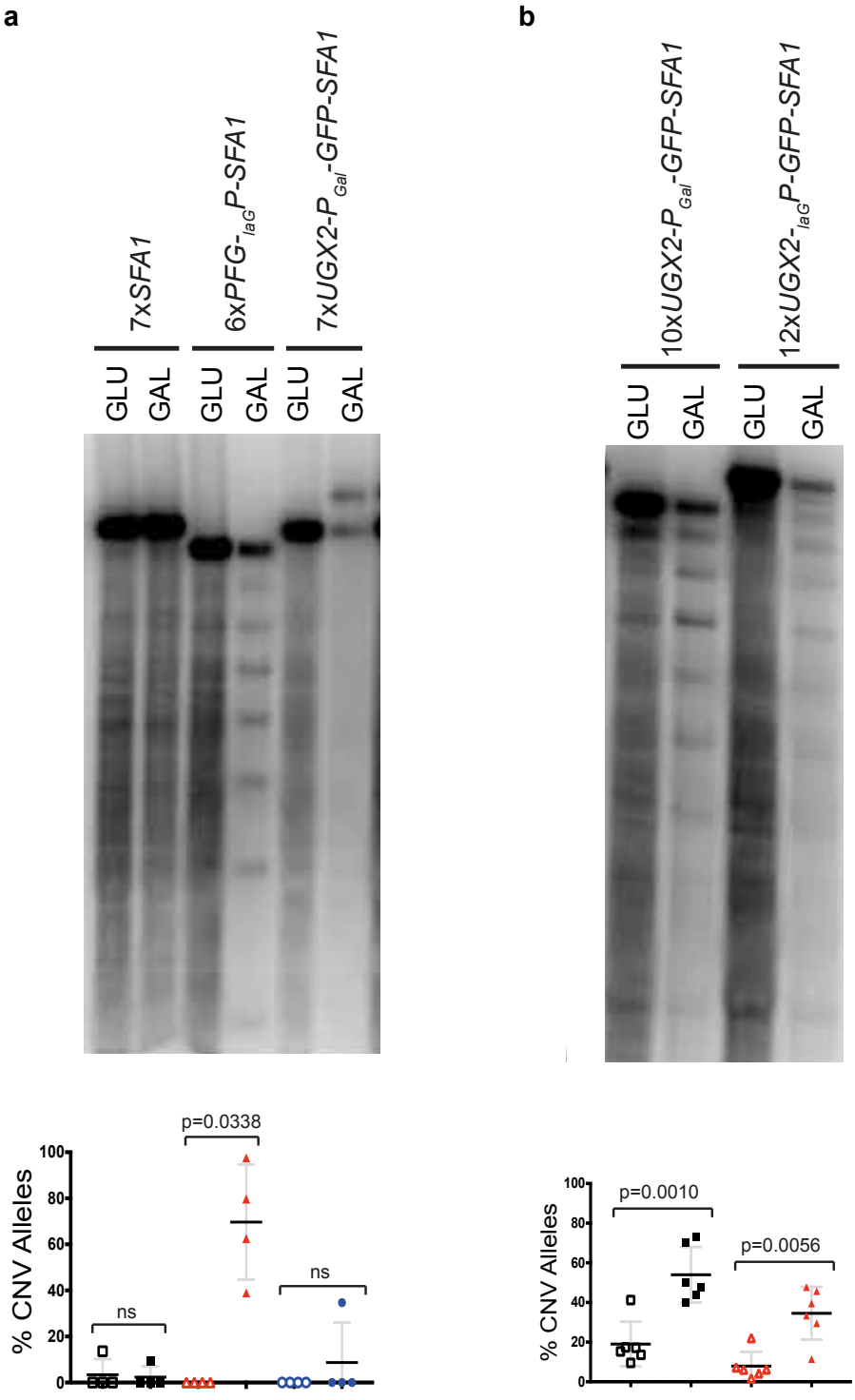


Figure 5.12: Transcriptionally stimulated CNV at *SFA1*. **a**) *SFA1* copy number analysis by Southern blot of 7x*SFA1* wild-type, 6x*PFG-$_{LAG}$P-GFP-SFA1* and 7x*UGX2-P$_{GAL}$-GFP-SFA1* cells grown for 10 generations in complete synthetic 2% glucose and 2% galactose media. *p* values were calculated by a paired 1-way ANOVA, *n* = 4. **b**) *CUP1* copy number analysis by Southern blot of 10x*UGX2-P$_{GAL}$-GFP-SFA1* and 12x*UGX2-$_{LAG}$P-GFP-SFA1* cells grown for 10 generations in complete synthetic 2% glucose and 2% galactose media. *p* values were calculated by a paired 1-way ANOVA, *n* = 6. ns = not significant.

7x*UGX2-P$_{GAL}$-GFP-SFA1* replicates, and even then, the rate of CNV was greatly reduced compared to the 6x*GFP-$_{LAG}$P-SFA1* strain. Therefore the orientation of the bi-directional promoter and/or the disruption of the upstream *UGX2* gene appeared to be important for stimulated CNV at *SFA1*. The *UGX2* gene produces a protein of unknown function, but its transcript has been shown to accumulate in response to a variety of stress conditions (Wu *et al.*, 2004), so it is not inconceivable that disruption of *UGX2* in 6x*GFP-$_{LAG}$P-SFA1* cells could cause an increase in transcriptionally stimulated CNV. To establish the importance of promoter directionality in transcriptionally stimulated CNV at *SFA1*, the 12x*UGX2-$_{LAG}$P-GFP-SFA1* strain was created, which maintained the promoter orientation of the 6x*GFP-$_{LAG}$P-SFA1* cells that showed galactose-stimulated CNV, but no longer disrupted the *UGX2* gene. 10x*UGX2-P$_{GAL}$-GFP-SFA1* and 12x*UGX2-$_{LAG}$P-GFP-SFA1* cells were grown for 10 generations to saturation in complete synthetic 2% glucose media, then 1:4000 diluted in complete synthetic 2% glucose and complete synthetic 2% galactose media. The 10x*UGX2-P$_{GAL}$-GFP-SFA1* cells only differed from the 7x*UGX2-P$_{GAL}$-GFP-SFA1* cells used in Fig. 5.12a by the addition of 3 complete genomic repeats, so as to have a closer starting copy number to the 12x*UGX2-$_{LAG}$P-GFP-SFA1* cells.

Southern blot analysis showed that both 10x*UGX2-P$_{GAL}$-GFP-SFA1* and 12x*UGX2-$_{LAG}$P-GFP-SFA1* cells underwent a low basal rate of CNV in non-inducible glucose (Fig. 5.12b). Both strains also displayed a significant induction of *de novo* CNV alleles with galactose induction, *p* values = 0.0010 and 0.0056 for 10x*UGX2-P$_{GAL}$-GFP-SFA1* and 12x*UGX2-$_{LAG}$P-GFP-SFA1* cells respectively. Galactose-stimulated CNV in 10x*UGX2-P$_{GAL}$-GFP-SFA1* cells was surprising, as there was no galactose-stimulated CNV in 7x*UGX2-P$_{GAL}$-GFP-SFA1* cells that only differed by 3 repeats. Nonetheless, this result provided evidence for transcriptionally stimulated CNV at *SFA1*, irrespective of the *P$_{GAL}$* promoter orientation and whether or not the *UGX2* gene was disrupted.

A possible explanation for galactose-stimulated CNV in 10x*UGX2-P$_{GAL}$-GFP-SFA1*, but not in 7x*UGX2-P$_{GAL}$-GFP-SFA1* cells, was that the CNV rate was simply higher in the higher-copy number strain, as observed earlier for *CUP1*. As such, galactose-stimulated CNV would still be occurring in 7x*UGX2-P$_{GAL}$-*

GFP-SFA1 cells, just at a rate that largely falls below the threshold of detection by Southern blot analysis. Since 6x*PFG-$_{LAG}$P-SFA1* cells still displayed galactose-stimulated CNV, despite having fewer starting repeats than 7x*UGX2-P$_{GAL}$-GFP-SFA1* cells, the promoter orientation and/or disruption of *UGX2* must make 6x*PFG-$_{LAG}$P-SFA1* cells more prone to transcriptionally stimulated CNV.

In summary, *SFA1* was selected as a good candidate for a stimulated CNV gene as *SFA1* had an upstream RFS site and was transcribed from a bi-directional promoter induced by formaldehyde. The emergence of *SFA1* alleles with higher than parental copy number was observed in formaldehyde and reduced in an *rtt109Δ* mutant, providing evidence that formaldehyde stimulated CNV at *SFA1* in a mechanism dependent on H3K56ac, as shown at *CUP1*. Nicotinamide could also stimulate CNV at *SFA1* in the absence of formaldehyde, confirming that CNV at *SFA1* was stimulated in a H3K56ac-dependent mechanism and not simply a selection towards rare pre-existing cells with *SFA1* CNV alleles. Finally, using a high-copy *SFA1* system, controlled by growth in a non-selective galactose environment, *SFA1* CNV was shown to be transcriptionally stimulated and largely independent of the orientation of the bi-directional promoter. Therefore *SFA1* provided evidence that our stimulated CNV model was not restricted to the *CUP1* locus, but was applicable to another 'γH2A gene'.

5.15 Summary

Already as part of this study, I had shown using the copper-resistance *CUP1* gene as a model, cells undergo transcriptionally stimulated CNV in response to a copper environment, in a H3K56ac-dependent process, as a means to accelerate their adaptation to a hostile copper environment. However, stimulated CNV is not a 'directed' mutation process, as it randomly produces copy number expansions and contractions at the stimulated gene, irrespective of the fitness outcome of the CNV mutation in the environment. Nicotinamide-stimulated CNV at *CUP1* caused a contraction bias in the absence of any selection pressure, which increased the proportion of alleles in the population

that should display a fitness disadvantage in a hostile copper environment. Using a growth curve assay and a competition assay, I have shown that despite the detrimental contraction bias, a population of nicotinamide pre-treated cells could still greatly outcompete a population of cells that have not undergone stimulated CNV, in both high- and low-copper environments. Therefore the potential reward in adaptation ability from undergoing stimulated CNV and producing a rare beneficial amplification, greatly outweighs the individual risk posed to the cell of primarily producing deleterious CNV events, and is the best option for a population of cells to accelerate adaptation to a challenging environment.

I have also provided evidence for the TOR pathway being involved in the regulation of stimulated CNV at *CUP1*, analogous to its regulation of the yeast rDNA copy number. The level of stimulated CNV and by extension copper adaptation ability, could be reduced pharmacologically with the TOR pathway inhibitor rapamycin, possibly by increasing the association of HDACs to transcriptionally active genes.

Furthermore, I provided evidence for copper-stimulated CNV and adaptation from cells with just 2 *CUP1* repeats, and preliminary data for CNV from cells with just a single repeat via an extrachromosomal circular intermediate. To improve the sensitivity of detection of CNV events from a single genomic repeat, I constructed a 1x*CUP1* 1x*SFA1* reporter construct, similar to that used in Zhang *et al*., that utilised the copper- and formaldehyde-resistance phenotypes of the respective *CUP1* and *SFA1* genes, to produce a highly sensitive plate-based detection system for genomic CNV events from single-copy in a defined cohort of cells. With its high sensitivity, the MEP 1x*CUP1* 1x*SFA1* system provided preliminary evidence for transcriptionally stimulated CNV being directly possible from single-copy genes. As most genes in the *Saccharomyces cerevisiae* genome are present as single-copy or duplications, stimulated CNV has the potential for great translatability to other 'γH2A genes' regardless of the copy number of the gene.

Finally, using the formaldehyde-resistance *SFA1* gene in another 3-copy system, I showed that environmentally stimulated CNV is not unique to *CUP1*, and using the high-copy galactose-inducible *SFA1* constructs, I confirmed that CNV at *SFA1* was also transcriptionally stimulated. Therefore stimulated CNV

is a high-risk strategy for environmental adaptation, but with a very high potential reward for the individual cell. Stimulated CNV is highly translatable to other ‘γH2A genes’ and can function at genes with as few as 2 repeats, with the possibility of even occurring at single-copy genes either directly, or indirectly through an extrachromosomal circular DNA intermediate.

6. Discussion

This thesis has provided evidence for a controlled mechanism in *Saccharomyces cerevisiae* for cells to accelerate their adaptation to challenging environments through CNV. I have shown that environmentally stimulated CNV is specific to genes with active promoters and dependent on the epigenetic mark H3K56ac. I have provided evidence for stimulated CNV being applicable to high-copy repeat arrays, right down to single copy genes. I have also shown that drugs can pharmacologically control a cell's ability to adapt to a hostile environment by interfering with the rate of stimulated CNV. The HDAC inhibitor nicotinamide can stimulate CNV events and promote adaptation, whereas the TOR pathway inhibitor rapamycin reduces CNV and adaptation. The existence of a stimulated CNV mechanism directly contradicts the notion that all CNV events are random mutations and makes us reconsider our existing understanding of adaptation. If conserved in somatic cells of higher eukaryotes, environmentally stimulated CNV could have important implications in cancer and drug resistance, where stimulated CNV could increase tumour heterogeneity and thereby accelerate the development of acquired resistance to the chemotherapeutic drug.

6.1 Proposed mechanism for stimulated CNV

From the experimental data in this thesis, we propose a mechanism for stimulated CNV at *CUP1* in *Saccharomyces cerevisiae* (Fig. 6.1), which could easily be expanded to other stimulated CNV genes. Candidate genes for stimulated CNV must first have an upstream RFS site, which accounts for ~7% all genes in *Saccharomyces cerevisiae* (Hull *et al.*, 2017). The candidate CNV gene is also likely to be lowly expressed under normal growth conditions, becoming highly expressed in response to a specific environmental stimulus. This is necessary so that cells limit stimulated CNV to genes that are transcriptionally marked as relevant in that environment.

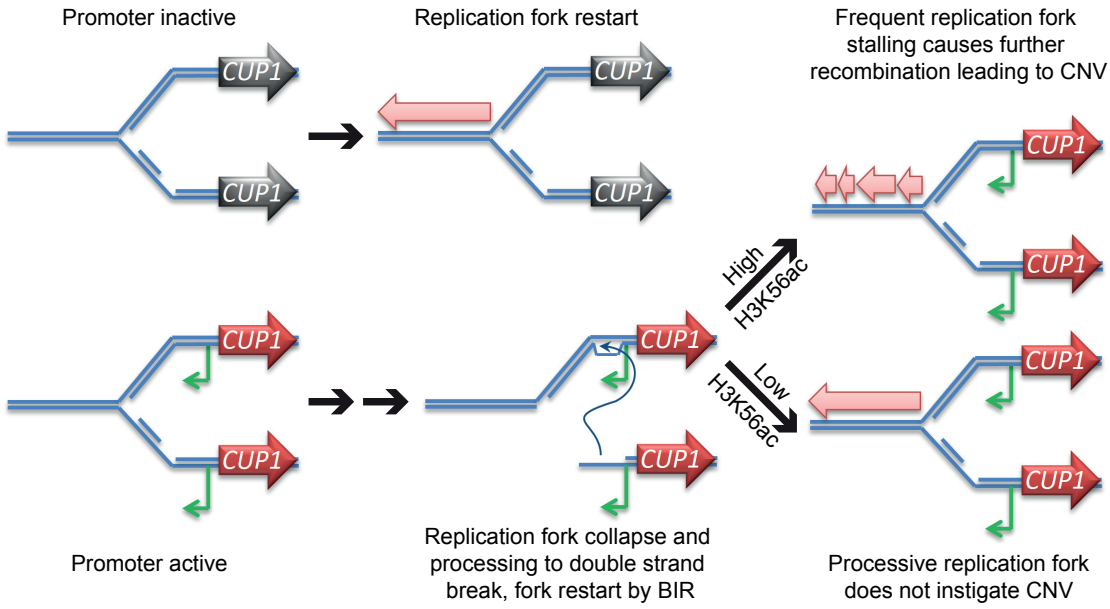


Figure 6.1: Proposed mechanism for stimulated CNV. Promoter activity and acetylated histone H3 lysine 56 (H3K56ac) contribute to stimulated CNV. DNA strands are shown in blue, the inactive *CUP1* gene is shown in black, and the induced *CUP1* gene is shown in red, with antisense CUT in green. Pink block arrows represent progression of the replication fork.

The replication machinery is predicted to stall at the RFS site upstream of the stimulated CNV gene. The stalled forks are normally resolved through error-free mechanisms that protect genome stability (reviewed in Yeeles *et al.*, 2013) and the replication fork simply restarts. However, if transcription is induced from the adjacent bi-directional promoter of a stimulated CNV gene, we propose that there are more frequent collisions between the replication and transcription machineries, or an increase in topological stress, thereby increasing the likelihood that the stalled replication fork collapses (Keszthelyi, Minchell and Baxter, 2016; Sankar *et al.*, 2016). Replication-transcription conflicts have been shown to be mutagenic, and in bacteria they are major contributors to genetic disease and evolution (Mirkin and Mirkin, 2005; Srivatsan *et al.*, 2010; Paul *et al.*, 2013; Sankar *et al.*, 2016). Therefore it is unsurprising that replication-transcription conflicts in eukaryotes could also produce mutations that accelerate adaptation.

Although essential for stimulated CNV, promoter induction is not enough to cause a CNV event by itself. After the replication fork collapses, it is

processed by introducing a double strand break (DSB), which is then repaired by a mechanism such as break induced replication (BIR). Continual collapse of replication forks and repair of the DSB by BIR reduces the processivity of the replication fork. We propose that the processivity of the replication fork predicts the level of error-prone BIR events, with frequently stalled forks more susceptible to recombination leading to CNV. We also provided evidence for H3K56ac regulating the processivity of the replication fork, where forks with low H3K56ac levels are highly processive, as in an *rtt109Δ* mutant or rapamycin-treated cells, and forks with high levels of H3K56ac frequently stall, in the case of forks near highly active promoters, in *hst3Δ hst4Δ* mutants, or in nicotinamide-treated cells.

The proposed mechanism for stimulated CNV shares a lot of features with the FoSTeS process that is thought to be responsible for complex genomic rearrangement, including a wide range of CNV events (Lee, Carvalho and Lupski, 2007; Zhang, Carvalho and Lupski, 2009). FoSTeS is proposed to initiate after the replication fork stalls, as a result of genomic instability at/near low-copy repeats, where the lagging strand serially disengages and template switches to a region of microhomology with an active replication fork, resulting in replication fork restart and DNA rearrangements (Lee, Carvalho and Lupski, 2007). The region of microhomology must be in close physical proximity, but can be separated by sizeable genomic distances, explaining the potential of FoSTeS to generate complex genomic rearrangements (Lee, Carvalho and Lupski, 2007). FoSTeS may not be limited to low-copy repeats though, as a FoSTeS/MMBIR mechanism has been proposed as a potential explanation for the observed expansion of rDNA repeats in the absence of Rad52 and Rad51 (Houseley and Tollervey, 2011). However, any contribution from transcription and H3K56ac has not been previously shown in FoSTeS. Furthermore, the data implicating H3K56ac and Pol32 in *CUP1* CNV is consistent with restart by BIR, since the replication forks newly formed through BIR are error-prone in the absence of Pol32 or in the presence of H3K56ac (Deem *et al.*, 2008; Smith, Lam and Symington, 2009; Che *et al.*, 2015). Nonetheless, other error-prone replication fork restart mechanisms are known, and may have similar dependencies (Lambert *et al.*, 2010; Mizuno *et al.*, 2013).

6.2 Proposed explanations for contraction bias in stimulated CNV

The proposed mechanism for stimulated CNV in Fig. 6.1, would predict the formation of both amplification and contraction events. However, in this study a clear contraction bias in the absence of selection was observed. Our lab provides two, not mutually exclusive mechanisms for the observed contraction bias of stimulated CNV (Fig. 6.2). The first mechanism relies on the different preferences towards template switching in high-copy repetitive regions and sites of unique sequence (Fig. 6.2a). The second mechanism relies upon the asymmetrical distribution of H3K56ac either side of the replication fork (Fig. 6.2b).

A copy number amplification event requires the fork to template switch backwards and re-replicate repeats (Fig. 6.2a, top pathway), whereas a contraction is formed when the fork template switches forward towards the end of the repeat array sequence (Fig. 6.2a, bottom pathway). As such, a replication fork that template switches backwards will spend longer replicating than a replication fork that template switches forward, thereby increasing the likelihood of additional template switching events, which could again produce an amplification or a contraction. By comparison a contraction, resulting from template switching forwards, brings the fork closer to a region of unique sequence at the end of the repeat array where template switching is disfavored. After as little as 2 successive template-switching events this would result in a theoretical 3:1 bias of contractions to amplifications.

It is highly likely that the asymmetrical distribution of H3K56ac either side of the replication fork adds another layer of complexity to the contraction bias. In wild-type yeast, grown in normal conditions, H3K56ac is primarily incorporated on newly synthesised histones in the region of DNA that has just been replicated (Masumoto *et al.*, 2005; Celic *et al.*, 2006). Therefore there is an asymmetrical distribution of H3K56ac either side of the replication fork. Stalled forks that template switch forwards encounter a previously unreplicated region of low H3K56ac, where the replication fork is highly processive. Template switching forwards results in a contraction and further

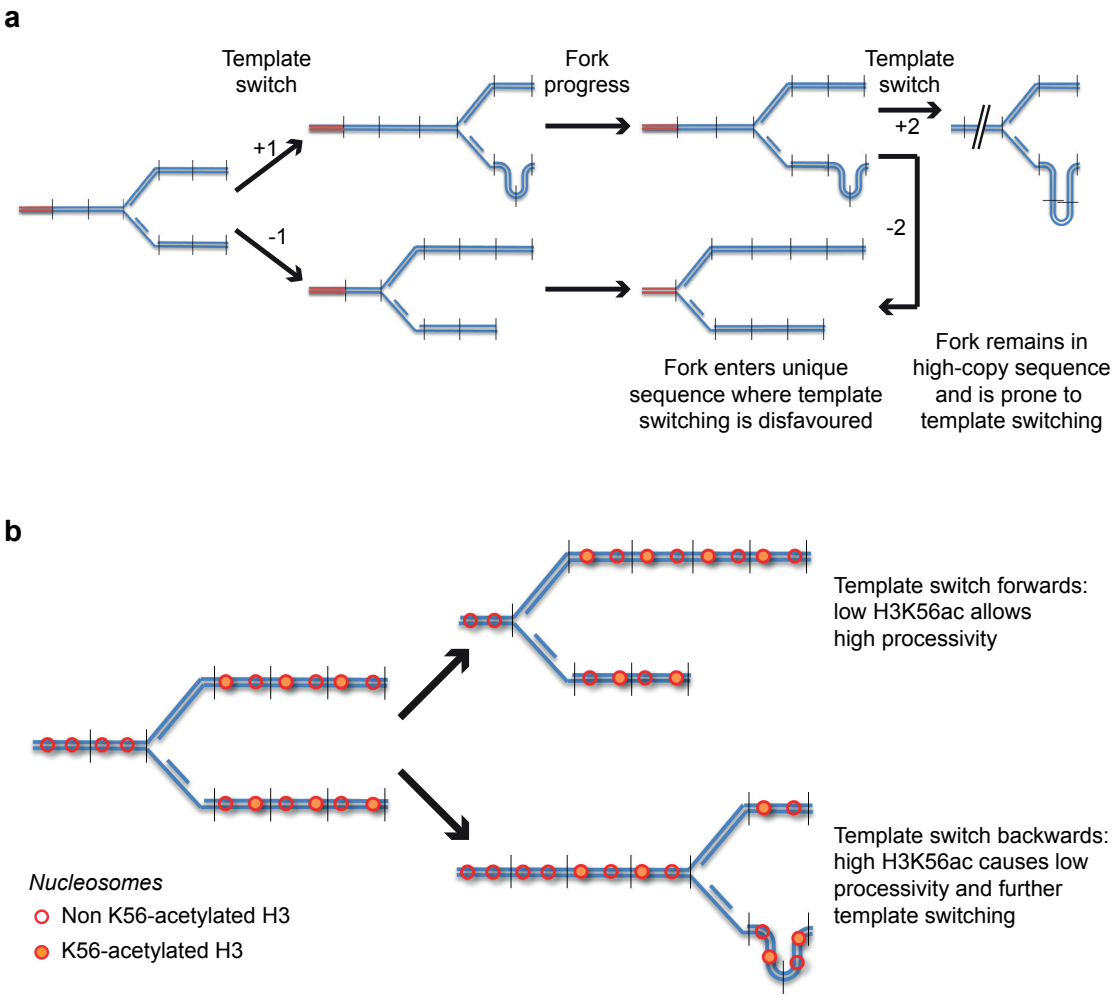


Figure 6.2: Proposed mechanisms for contraction bias in stimulated CNV. **a**) DNA strands from repeated DNA are shown in blue, unique sequences are shown in red, and vertical lines indicate repeat boundaries. Numbers indicate the change in copy number for a particular template switch. The result of 2 successive template-switching events with an intervening period of replication are shown, and result in a 3:1 ratio of contractions to amplifications. **b**) Additional asymmetry is generated by H3K56ac: nucleosomes around a replication fork (blue) are shown as red circles, either empty or shaded to represent the H3K56ac state. Template switching forward is seen to move the fork to a region of low H3K56ac, whereas template switching backwards moves the fork to a region of high H3K56ac and therefore low fork stability.

template switching is disfavoured (Fig. 6.2b, top pathway). Whereas a BIR fork that template switches backwards encounters a recently replicated region of high H3K56ac and low processivity. Template switching backwards therefore produces an amplification event, but with replication forks primed for

further template switching, which again can result in additional amplifications or contractions (Fig. 6.2b, bottom pathway). Therefore the asymmetrical distribution of H3K56ac further biases in favour of contractions, but would not prevent amplifications from occurring at a much lower frequency.

6.3 Stimulated CNV is similar but distinct from CNV at the model rDNA

At the start of this study, predictions were made about stimulated CNV based upon the specialised rDNA locus, which undergoes extensive CNV and has been commonly used as a general model for CNV. Indeed many parallels were observed between the two mechanisms, but also some important distinctions. For instance, both mechanisms have a requirement for transcription from a bi-directional promoter and CNV in the two systems could be stimulated by nicotinamide and repressed by rapamycin (Ha and Huh, 2011; Jack *et al.*, 2015). H3K56ac HDACs also silenced CNV in both systems. However, Sir2 did not show any detectable involvement in stimulated CNV at the *CUP1* locus, despite silencing transcription and inhibiting recombination at the rDNA (Gottlieb and Esposito, 1989; Li, Mueller and Bryk, 2006). This was not so surprising as Sir2 acts primarily at regions of heterochromatin, with *CUP1* being euchromatic (Loo and Rine, 1995; Moazed, 2001; Rusche, Kirchmaier and Rine, 2003). However, the converse function of the acetyltransferase Rtt109 in the two mechanisms was a surprise. In stimulated CNV, *rtt109Δ* mutants could greatly reduce CNV, whereas at the rDNA, *rtt109Δ* cells observed hyper-amplification of the ribosomal repeats (Houseley and Tollervey, 2011; Ide, Saka and Kobayashi, 2013). This discrepancy in *rtt109Δ* cells could be explained if rDNA amplification occurs via rolling circle amplification and re-integration of extrachromosomal rDNA circles (ERCs) instead of via chromosomal BIR, as has been proposed by the Kobayashi group (Ide, Saka and Kobayashi, 2013). Since *CUP1* circles are 3-4 orders of magnitude rarer than ERCs (Moller *et al.*, 2015), we reasoned that rolling circle amplification could not have such an appreciable contribution to CNV at *CUP1*. However, it remains a possible

explanation for the slow-arising, partially copper-resistant *rtt109Δ* cells with *CUP1* amplifications that were detected in cells recovered from the copper adaptation assays (Fig. 4.7c).

6.4 DNA repair by RNA

The findings of this study share similarities to reported models of DNA repair using an RNA template. RNA has long been known to be reverse transcribed from retroelements in yeast, where it can be reintroduced into the genome at novel sites (Derr, Strathern and Garfinkel, 1991). However, there is growing evidence that potentially any cellular RNA could be used as substrates for reverse transcription (Keskin *et al.*, 2014; Keskin, Meers and Storici, 2016). The reverse transcribed DNA copy (cDNA) is a potential substrate for recombination or can be captured by a DSB (Derr and Strathern, 1993; Moore and Haber, 1996; Teng, Kim and Gabriel, 1996; Meers, Keskin and Storici, 2016). Alternatively the cell can directly use RNA as a template for repair of DSBs, without needing to be first converted into cDNA (Storici *et al.*, 2007). Meers *et al*. propose 5 models for RNA-mediated DSB repair that use either the direct RNA template, or the reverse transcribed cDNA (Meers, Keskin and Storici, 2016). The eukaryotic ribonucleases (RNase) H1 and H2 cleave RNA-DNA hybrids and loss of RNase H1 and H2 increases the rate of direct transcript-RNA repair over the preferred repair by cDNA used by wild-type cells (Storici *et al.*, 2007). RNase H1 and H2 therefore control the switch between direct RNA template DSB repair and repair by cDNA.
In this thesis, promoter induction was found to be essential for CNV. Therefore it is possible that promoter induction does not stimulate CNV through replication-transcription conflicts, but instead provides an RNA template for direct or indirect repair of DSBs at collapsed replication forks. Interestingly transcript RNA is more efficient at repairing a DSB in *cis* than at a homologous ectopic locus in *trans* (Keskin, Meers and Storici, 2016). As such, RNA mediated repair is an already defined cellular process that directly links local transcription to the repair of DSBs. However, RNA mediated repair is an error-free mechanism of DSB repair and is not expected to result in

CNV, and so stimulated CNV is unlikely to utilise mechanisms of RNA-templated repair (Chakraborty *et al.*, 2016; Keskin, Meers and Storici, 2016).

6.5 Adaptive potential of stimulated CNV

Adaptive mutations have largely been assumed to be random and inevitable consequences of exposure to the cell's environment, but these adaptive mutation mechanisms have been poorly defined. The assumption that adaptive mutation is inevitable, proposes that adaptation cannot be suppressed and that loss of genome stability factors would increase the rate of adaptation. However, this study has characterised a defined mechanism that largely controlled the yeast's ability to adapt to environmental copper by amplification of the *CUP1* gene. Stimulated CNV was still regulated by the general genome stability factor Rtt109, but here loss of *RTT109* reduced CNV, challenging the assumption that all random mutation is inevitable. Therefore at least a subset of all random mutation events may be stimulated by a specific cellular mechanism, and by chance some of these random mutations may be beneficial and promote adaptation to the challenging environment.

There are instances where control of environmental adaptation in yeast could be exploited for beneficial use, such as in the prevention of antifungal resistance developing in pathogenic yeast. Copy number changes are a major mechanism in the development of antifungal resistance to the azole class of antifungal drugs (Cowen *et al.*, 2015). Types of copy number change that lead to antifungal resistance include loss of heterozygosity (LOH) events, segmental or chromosomal aneuploidies, and chromosomal CNV (Marichal *et al.*, 1997; Selmecki, Forche and Berman, 2006, 2010; Coste *et al.*, 2007; Selmecki *et al.*, 2009). This study found that the TOR inhibitor rapamycin could reduce chromosomal CNV and adaptation to a hostile environment. Therefore our proposed mechanism for stimulated CNV provides a novel pharmacologically accessible target to delay the development of acquired antifungal resistance through chromosomal CNV.

Alternatively, promoting adaptation through CNV could be exploited for biotechnology, such as in the brewing industry (Gorter de Vries, Pronk and Daran, 2017). For instance, acceleration of the adaptation process in yeast to grow on a more limited food source, or to tolerate a more toxic batch culture, would help reduce production costs and increase yields (Oud *et al.*, 2013; Voordeckers *et al.*, 2015; Venkataram *et al.*, 2016; Gorter de Vries, Pronk and Daran, 2017). In this study, nicotinamide was identified as one potential drug that can promote adaptation by stimulating CNV, showing that pharmacological manipulation of yeast's adaptation capability is already possible and has the potential to be exploited for biotechnological purposes.
This study has shown that stimulated CNV is most efficient at CNV genes present as multiple homologous repeats, decreasing in efficiency with fewer copies. This conclusion has also been observed by the Petes group, where the interhomolog and intra/sister chromatid recombination rates at the tandemly-repeated *CUP1* was almost three orders of magnitude higher than the mitotic recombination rate for the single-copy gene (Zhao *et al.*, 2017). Despite the reduced efficiency of CNV starting from low-copy, this study has shown that stimulated CNV was still detected in low-copy systems, with preliminary evidence presented for stimulated CNV even occurring from single-copy genes. Error-prone replication forks could easily initiate recombination events that utilize distal homologous sequences or even regions of microhomology as repair templates, to induce the initial *de novo* CNV event at a single-copy gene. However, breakpoints in *de novo* CNVs would be poorly defined because they initiate nearby but not at the RFS site and utilize unpredictable homologous sequences. Interestingly, the multi-copy *CUP1* repeat array has emerged many times with different breakpoints in different *Saccharomyces cerevisiae* isolates (Zhao *et al.*, 2014). This provides evidence for the initial *de novo* CNV event, which duplicated the *CUP1* gene in the wild, not initiating from a defined reproducible site.
This study provided evidence for stimulated CNV at the *CUP1* and *SFA1* genes, but this mechanism could be much more widespread throughout the genome. Approximately 7% all genes in *Saccharomyces cerevisiae* potentially have an upstream RFS site and were lowly expressed in optimum growth conditions (Hull *et al.*, 2017), and bi-directional promoters are common

throughout the yeast genome (Xu *et al.*, 2009). Our model predicts that genes with all three of these characteristics have the potential to be stimulated CNV genes, which still covers a large number of yeast genes. The tandemly repeated genes *HXT6/7* and *ENA1/2/5* are a few immediately obvious candidate genes for being able to undergo stimulated CNV, as they are known to become overexpressed in a specific challenging environment (Banuelos *et al.*, 1998; Brown, Todd and Rosenzweig, 1998), and have been detected alongside *CUP1* for their high production of eccDNAs (Møller *et al.*, 2015).

6.6 Stimulated CNV to accelerate adaptation in higher eukaryotes

It is highly likely that our stimulated CNV mechanism is conserved in higher eukaryotes, including humans, as replication fork stalling is not unique to *Saccharomyces cerevisiae* and bi-directional promoters are the norm in higher eukaryotes (Preker *et al.*, 2011; Wei *et al.*, 2011). Even the conserved histone deacetylases in mammals appear to share a similar importance in regulation of DNA repair (Das *et al.*, 2009; Yuan *et al.*, 2009). However, in somatic cells of multicellular organisms, stimulated CNV is unlikely to be desirable, as it cannot provide a heritable adaptation. Nonetheless, a stimulated CNV mechanism in somatic cells could be readily exploited for adaptation of disease states, such as tumour cells.

Interestingly, our proposed stimulated CNV mechanism shares a lot of similarities with recent reports in neurons on non-random DSBs in genes important for neuronal development (Schwer *et al.*, 2016; Wei *et al.*, 2016). Therefore stimulated CNV, or some variation on this mechanism, may well be conserved in mammals and provides a novel pathway to target pharmacologically for treatment of diseases linked to adaptation through CNV, such as chemotherapeutic drug resistance in cancer (Frei *et al.*, 1984; Corcoran *et al.*, 2010; Little *et al.*, 2011). Furthermore, there is already evidence for stimulated CNV being pharmacologically accessible. This study has provided evidence for the TOR inhibitor rapamycin being an effective

repressor of stimulated CNV and adaptation to the environment. Also there are existing drugs, already proven to be effective targets of the acetyltransferase Rtt109, which was an important protein required for adaptation through CNV in our study (Ruotolo *et al.*, 2010; Lopes da Rosa *et al.*, 2013).

6.7 Stimulated CNV is not 'directed' mutation

Adaptation through genome-wide non-random mutation is well established, particularly in bacteria and primarily in response to stress (Galhardo, Hastings and Rosenberg, 2007), but the ability of stimulated CNV to limit mutations to the relevant loci must be forcefully opposed to any 'directed' mutation systems that have been previously reported (Cairns, Overbaugh and Miller, 1988; Lenski and Mittler, 1993; Roth *et al.*, 2003; Galhardo, Hastings and Rosenberg, 2007). The major issue with 'directed' mutation systems is that they require a foreknowledge of the fitness outcome of a given mutation in response to its environment, and so cells only undergo the mutation that provides the greatest adaptive gain. This is clearly not the case for our stimulated CNV mechanism, where copper-treated MEP wild-type cells produced mostly contractions causing a detrimental effect on adaptation (Fig. 4.1b and 4.2). Despite this high risk, stimulated CNV is still worthwhile for a population of cells, as it enabled the population as a whole to gain an enhanced fitness over cells that did not stimulate the CNV (Fig. 5.1).

Most adaptation through random mutation studies ignore the existing wealth of information on gene regulatory systems, where genes that become induced in response to an environment are thereby marked as important genes for that environment. Such genes make good candidates to stimulate CNV at, as they are more likely to yield a beneficial, adaptive result than stimulation of a random gene. As such, stimulated CNV can be targeted to the most promising genes for the environment, without having to be a 'directed' mutation event.

It is also important for the cell to limit stimulated CNV to just the genes that become specifically expressed in response to the environment, and not simply at all highly expressed genes. This is because some housekeeping genes are

highly expressed and we would hypothesise that stimulating mutation in these genes would be very detrimental to cells. One way cells could distinguish highly induced genes from highly expressed genes could be through the placement of RFS sites. As such we hypothesise that the location of RFS sites throughout the yeast genome have arisen through natural selection acting on randomly located RFS sites. RFS sites positioned near genes where CNV would be detrimental, like housekeeping genes, would be rapidly lost. As such, the genome-wide location of RFS sites will have been strongly selected though evolution over many generations, to remain at just the genes where stimulated CNV is important for adaptation.

6.8 Summary

Stimulated CNV is a novel but imperfect targeted cellular mechanism that is of high-risk, but with a very high potential reward for the individual cell, as stimulated CNV offers the chance for the cell to outcompete its kin and have its individual genetic information become dominant in the population. Stimulated CNV effectively accelerates the rate of environmental adaptation to challenging environments, beyond what is achievable through random mutation and to the best of our knowledge, stimulated CNV provides the first evidence of controllable environmental adaptation in eukaryotes and is pharmacologically accessible.

7. Future Work

7.1 Further characterise proteins involved in stimulated CNV

In this study, a two-step mechanism has been proposed for stimulated CNV in response to the environment (Fig. 6.1). The importance of H3K56ac for stimulated CNV was shown using *rtt109Δ* mutants, which greatly reduced the rate of CNV at both *CUP1* and *SFA1* loci. However, the importance of efficient DNA replication following fork stalling was merely inferred using *mrc1Δ* and *pol32Δ* mutants, which shared the hyper-recombinant phenotypes of the *hst3Δ hst4Δ* HDAC mutant. To confirm that the replication protein mutants act upstream of H3K56ac, as proposed by our model, *RTT109* should be deleted in both *mrc1Δ* and *pol32Δ* hyper-recombinant mutants, with the prediction that CNV in these double mutants should be independent of H3K56ac.
Our mechanism also relies on DSB formation at the stalled replication fork site, to provide a template for BIR repair. Structure-specific endonucleases help to resolve problems at replication forks, by cleaving branched DNA substrates to leave recombination intermediates (Hanada *et al.*, 2007). The Holliday junction resolvase Yen1 (Ip *et al.*, 2008), and the structure-specific endonucleases Mus81 and Slx4 (Fricke and Brill, 2003; Hanada *et al.*, 2007) are good candidates for causing the DSB in our stimulated CNV model. If involved, deletion of *YEN1*, *MUS81*, *Slx4*, or a combination of these endonucleases in case of redundancy, should reduce the rate of stimulated CNV.

7.2 Effect of deleting the RFS site

Our stimulated CNV model predicted the importance of an upstream RFS site, as the initiation site for stimulated CNV of the downstream gene. To determine whether the RFS site was essential for stimulated CNV, the RFS

site was deleted in the 3x*GFP-CUP1* strain and blocked the emergence of *CUP1* alleles with higher than parental copy number in copper treatment, as well as completely suppressing adaptation to a challenging copper environment (Fig. 4.12). As such, it was concluded that the RFS site was essential for copper-stimulated CNV and adaptation to a hostile copper environment. However, to explore whether an RFS site is the initiation site for CNV, deletion of the RFS-containing region should be explored in the hyper-recombinant mutants *mrc1Δ*, *pol32Δ* and *hst3Δ hst4Δ*. Greatly reduced or complete loss of CNV in RFS-deleted hyper-recombinant mutants would support our hypothesis that the RFS site is essential for the initiation of CNV. Replication fork stalling sites were identified genome-wide (Hull *et al.*, 2017). Therefore it was a little surprising that stimulated CNV in 3x*GFP-CUP1* did not just switch to initiating from the next nearest RFS site and amplify a much larger sequence of DNA. Abrogation of stimulated CNV in 3x*GFP-CUP1* cells might have been due to removal of transcriptional interference. As such, collisions between the replication and transcription machineries would now be separated in genomic distance, but also the transcriptionally induced topological stress would be distanced from the RFS site, providing topoisomerases more time to correct the stress.
Alternatively, the absence of CNV in 3x*GFP-CUP1* cells may be a direct effect of genomic distance, with the rate of CNV at a gene decreasing the further away it is from the initiating RFS site. To determine whether transcriptional interference with the RFS site is the essential first step in stimulated CNV, *MRC1* and *POL32* should be deleted in P_{GAL}-*HA* yeast and grown in glucose, where the promoter is completely repressed, to explore if the hyper-recombination phenotype persists in the absence of transcriptional interference.

7.3 1 to 2-copy stimulated CNV though natural eccDNA re-integration

Fig. 5.4a provided evidence for 1 to 2-copy CNV via re-integration of a circular intermediate. However the circular intermediate used as part of this

experiment was an artificially constructed plasmid and not a naturally produced eccDNA. The rolling circle amplification model for CNV has already been proposed by the Kobayashi group to explain rDNA amplifications via re-integration of ERCs (Ide, Saka and Kobayashi, 2013). However, the current evidence for stimulated circle formation and re-integration is poor. As such, it would be very interesting to determine whether the formation of a natural *CUP1* eccDNA could be stimulated by copper and re-integrate at the genomic *CUP1* locus, causing a CNV event.

Our lab has preliminary evidence of copper stimulating the formation of *CUP1* circles (unpublished data) in aged MEP wild-type diploid cells by Southern blot analysis. However, this experiment was done in a high-copy strain, so the question remains whether copper can stimulate *CUP1* eccDNAs from a single genomic *CUP1* repeat.

In this study, we observed a reduced recombination rate with fewer genomic repeats. Therefore eccDNA formation is expected to be rare coming from genomic sites with few repeats. Our lab is working on new methodologies to improve the sensitivity and quantification of existing circle detection technologies. It is hoped that new approaches for detection and quantification of eccDNAs will increase detection of eccDNAs formed from single-copy loci, but also enable transcriptional stimulation to be determined.

In addition, formation of eccDNAs does not mean that the eccDNA can re-integrate at the genomic locus. To determine whether eccDNAs can re-integrate into the genome would require identification of mutants that enable eccDNA formation but prevent re-integration, which are currently unknown. If these eccDNA re-integration mutants were introduced into a MEP 1x*CUP1* background, copper-stimulated eccDNAs should accumulate but genomic CNV should be blocked. In addition, exploring copper-adaptation in untreated wild-type vs. copper-treated re-integration mutant MEP 1x*CUP1* cells would provide an insight into whether eccDNAs are transcriptionally active or not.

7.4 Regulation of the 1x*CUP1* 1x*SFA1* reporter

This study presented preliminary evidence for formaldehyde-stimulated CNV occurring from single-copy genes using the MEP 1x*CUP1* 1x*SFA1* plate-based reporter assay (Fig. 5.6b). However, formaldehyde is mutagenic to yeast (Chanet and von Borstel, 1979), therefore the 2.7-fold increase in the number of resistant cells detected in formaldehyde pre-treated MEP 1x*CUP1* 1x*SFA1* heterozygote cells could be the result of formaldehyde acting as a mutagen and not through transcriptionally stimulated CNV.

To determine whether resistance through CNV is transcriptionally stimulated from single-copy genes, the native *SFA1* promoter should be replaced with a galactose-responsive promoter in MEP 1x*CUP1* 1x*SFA1* heterozygote cells and then pre-treated in glucose or galactose for the plate-based assay. Galactose pre-treatment would be predicted to increase the number of resistant cells detected on –Trp drop-out plates containing 0.2mM $CuSO_4$ and 1.6mM formaldehyde, assuming that CNV is still transcriptionally stimulated from single copy genes. However, this proposed experiment has the caveat that in this study galactose treated cells only ever produced detectable contraction events. In theory amplification events should still occur, but at a greatly reduced rate that may even be too low for detection in the high sensitivity plate-based assay.

Alternatively, transcriptionally stimulated CNV from single copy can be inferred using an *rtt109Δ* mutant. In this study deletion of *RTT109* reduced stimulated CNV in high-copy systems and in cells with 3 *CUP1* or *SFA1* repeats, and H3K56ac is essential in our proposed model for stimulated CNV. *RTT109* deletion in MEP 1x*CUP1* 1x*SFA1* heterozygote cells would be predicted to reduce the number of resistant cells detected by the plate-based assay, following formaldehyde pre-treatment, thereby providing supporting evidence for transcriptionally stimulated CNV in response to the environment being translatable to all “γH2A genes,” irrespective of its parental copy number.

8. References

Adamo, G. M. *et al.* (2012) 'Amplification of the CUP1 gene is associated with evolution of copper tolerance in Saccharomyces cerevisiae.', *Microbiology (Reading, England)*. England, 158(Pt 9), pp. 2325–2335. doi: 10.1099/mic.0.058024-0.

Adamo, G. M. *et al.* (2012) 'Laboratory evolution of copper tolerant yeast strains.', *Microbial cell factories*. England, 11, p. 1. doi: 10.1186/1475-2859-11-1.

Aguilera, A. (2002) 'The connection between transcription and genomic instability.', *The EMBO journal*. England, 21(3), pp. 195–201. doi: 10.1093/emboj/21.3.195.

Aguilera, A. and Garcia-Muse, T. (2012) 'R loops: from transcription byproducts to threats to genome stability.', *Molecular cell*. United States, 46(2), pp. 115–124. doi: 10.1016/j.molcel.2012.04.009.

Aguilera, A. and Garcia-Muse, T. (2013) 'Causes of genome instability.', *Annual review of genetics*. United States, 47, pp. 1–32. doi: 10.1146/annurev-genet-111212-133232.

Aguilera, A. and Gómez-González, B. (2008) 'Genome instability: a mechanistic view of its causes and consequences', *Nature Reviews Genetics*. Nature Publishing Group, 9, p. 204.

Alani, E. *et al.* (1992) 'Characterization of DNA-binding and strand-exchange stimulation properties of y-RPA, a yeast single-strand-DNA-binding protein.', *Journal of molecular biology*. England, 227(1), pp. 54–71.

Alcasabas, A. A. *et al.* (2001) 'Mrc1 transduces signals of DNA replication stress to activate Rad53.', *Nature cell biology*. England, 3(11), pp. 958–965. doi: 10.1038/ncb1101-958.

Ames, B. N., Shigenaga, M. K. and Hagen, T. M. (1993) 'Oxidants, antioxidants, and the degenerative diseases of aging.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 90(17), pp. 7915–7922.

Anderson, R. M. *et al.* (2003) 'Nicotinamide and PNC1 govern lifespan extension by calorie restriction in Saccharomyces cerevisiae.', *Nature*, 423(6936), pp. 181–5. doi: 10.1038/nature01578.

Aplan, P. D. (2006) 'Causes of oncogenic chromosomal translocation', *Trends in Genetics*, 22(1), pp. 46–55. doi: https://doi.org/10.1016/j.tig.2005.10.002.

Argueso, J. L. *et al.* (2008) 'Double-strand breaks associated with repetitive DNA can reshape the genome.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 105(33), pp. 11845–11850. doi: 10.1073/pnas.0804529105.

Arlt, M. F., Wilson, T. E. and Glover, T. W. (2012) 'Replication stress and mechanisms of CNV formation.', *Current opinion in genetics & development*. England, 22(3), pp. 204–210. doi: 10.1016/j.gde.2012.01.009.

Avalos, J. L., Bever, K. M. and Wolberger, C. (2018) 'Mechanism of Sirtuin Inhibition by Nicotinamide: Altering the NAD^+ Cosubstrate Specificity of a Sir2 Enzyme', *Molecular Cell*. Elsevier, 17(6), pp. 855–868. doi: 10.1016/j.molcel.2005.02.022.

Bai, Z. *et al.* (2016) 'The impact and origin of copy number variations in the Oryza species', *BMC Genomics*. London: BioMed Central, 17, p. 261. doi: 10.1186/s12864-016-2589-2.

Baldacci, G., Cherif-Zahar, B. and Bernardi, G. (1984) 'The initiation of DNA replication in the mitochondrial genome of yeast.', *The EMBO journal*. England, 3(9), pp. 2115–2120.

Bando, M. *et al.* (2009) 'Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks.', *The Journal of biological chemistry*. United States, 284(49), pp. 34355–34365. doi: 10.1074/jbc.M109.065730.

Banuelos, M. A. *et al.* (1998) 'The Nha1 antiporter of Saccharomyces cerevisiae mediates sodium and potassium efflux.', *Microbiology (Reading, England)*. England, 144 (Pt 1, pp. 2749–2758. doi: 10.1099/00221287-144-10-2749.

Baranowska, H., Policinska, Z. and Jachymczyk, W. J. (1995) 'Effects of the CDC2 gene on adaptive mutation in the yeast Saccharomyces cerevisiae.', *Current genetics*. United States, 28(6), pp. 521–525.

Barbet, N. C. *et al.* (1996) 'TOR controls translation initiation and early G1 progression in yeast.', *Molecular biology of the cell*. United States, 7(1), pp. 25–42.

Basu, A. K. (2018) 'DNA Damage, Mutagenesis and Cancer.', *International journal of molecular sciences*. Switzerland, 19(4). doi: 10.3390/ijms19040970.

Bazan, J. F. (2008) 'An old HAT in human p300/CBP and yeast Rtt109.', *Cell cycle (Georgetown, Tex.)*. United States, pp. 1884–1886. doi: 10.4161/cc.7.12.6074.

Bermudez, V. P. *et al.* (2010) 'Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication.', *The Journal of biological chemistry*. United States, 285(13), pp. 9493–9505. doi: 10.1074/jbc.M109.093609.

Bernstein, C. *et al.* (2013) 'DNA Damage, DNA Repair and Cancer', in Chen, C. B. T.-N. R. D. in D. N. A. R. (ed.). Rijeka: InTech, p. Ch. 16. doi: 10.5772/53919.

Beroukhim, R. *et al.* (2010) 'The landscape of somatic copy-number alteration across human cancers.', *Nature*. England, 463(7283), pp. 899–905. doi: 10.1038/nature08822.

Bitterman, K. J. *et al.* (2002) 'Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human SIRT1.', *The Journal of biological chemistry*. United States, 277(47), pp. 45099–45107. doi: 10.1074/jbc.M205670200.

Brachmann, C. B. *et al.* (1995) 'The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability.', *Genes & development*. United States, 9(23), pp. 2888–2902.

Brachmann, C. B. *et al.* (1998) 'Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.', *Yeast (Chichester, England)*. England, 14(2), pp. 115–132. doi: 10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

Brewer, B. J. *et al.* (2011) 'Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons.', *PLoS genetics*. United States, 7(3), p. e1002016. doi: 10.1371/journal.pgen.1002016.

Brewer, B. J., Lockshon, D. and Fangman, W. L. (1992) 'The arrest of replication forks in the rDNA of yeast occurs independently of transcription.', *Cell*. United States, 71(2), pp. 267–276.

Brown, C. J., Todd, K. M. and Rosenzweig, R. F. (1998) 'Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment.', *Molecular biology and evolution*. United States, 15(8), pp. 931–942. doi: 10.1093/oxfordjournals.molbev.a026009.

Bruder, C. E. G. *et al.* (2008) 'Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles.', *American journal of human genetics*, 82(3), pp. 763–71. doi: 10.1016/j.ajhg.2007.12.011.

Bryk, M. *et al.* (1997) 'Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast.', *Genes & development*. United States, 11(2), pp. 255–269.

Buchman, C. *et al.* (1989) 'The CUP2 gene product, regulator of yeast metallothionein expression, is a copper-activated DNA-binding protein.', *Molecular and cellular biology*. United States, 9(9), pp. 4091–4095.

Butt, T. R., Sternberg, E., *et al.* (1984) 'Cloning and expression of a yeast copper metallothionein gene.', *Gene*. Netherlands, 27(1), pp. 23–33.

Butt, T. R., Sternberg, E. J., *et al.* (1984) 'Copper metallothionein of yeast, structure of the gene, and regulation of expression.', *Proceedings of the National Academy of Sciences of the United States of America*, pp. 3332–3336.

Byrne, K. P. and Wolfe, K. H. (2005) 'The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species.', *Genome research*. United States, 15(10), pp. 1456–1461. doi: 10.1101/gr.3672305.

Cafferkey, R. *et al.* (1993) 'Dominant missense mutations in a novel yeast protein related to mammalian phosphatidylinositol 3-kinase and VPS34 abrogate rapamycin cytotoxicity.', *Molecular and Cellular Biology* , 13(10), pp. 6012–6023. doi: 10.1128/MCB.13.10.6012.

Cai, J. *et al.* (1998) 'ATP hydrolysis catalyzed by human replication factor C requires participation of multiple subunits.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 95(20), pp. 11607–11612.

Cairns, J. and Foster, P. L. (1991) 'Adaptive reversion of a frameshift mutation in Escherichia coli.', *Genetics*. United States, 128(4), pp. 695–701.

Cairns, J., Overbaugh, J. and Miller, S. (1988) 'The origin of mutants.', *Nature*. England, 335(6186), pp. 142–145. doi: 10.1038/335142a0.

Caldecott, K. W. (2008) 'Single-strand break repair and genetic disease', *Nature Reviews Genetics*. Nature Publishing Group, 9, p. 619.

Cannan, W. J. and Pederson, D. S. (2016) 'Mechanisms and Consequences of Double-strand DNA Break Formation in Chromatin', *Journal of cellular physiology*, 231(1), pp. 3–14. doi: 10.1002/jcp.25048.

Castellano-Pozo, M., Garcia-Muse, T. and Aguilera, A. (2012) 'R-loops cause replication impairment and genome instability during meiosis.', *EMBO reports*. England, 13(10), pp. 923–929. doi: 10.1038/embor.2012.119.

Ceccaldi, R., Rondinelli, B. and D'Andrea, A. D. (2018) 'Repair Pathway Choices and Consequences at the Double-Strand Break', *Trends in Cell Biology*. Elsevier, 26(1), pp. 52–64. doi: 10.1016/j.tcb.2015.07.009.

Celic, I. *et al.* (2006) 'The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation.', *Current biology : CB*. England, 16(13), pp. 1280–1289. doi: 10.1016/j.cub.2006.06.023.

Celic, I., Verreault, A. and Boeke, J. D. (2008) 'Histone H3 K56

Hyperacetylation Perturbs Replisomes and Causes DNA Damage', *Genetics*, 179(4), p. 1769 LP-1784.

Cerritelli, S. M. and Crouch, R. J. (2009) 'Ribonuclease H: the enzymes in eukaryotes.', *The FEBS journal*. England, 276(6), pp. 1494–1505. doi: 10.1111/j.1742-4658.2009.06908.x.

Chakraborty, A. *et al.* (2016) 'Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes', *Nature Communications*. Nature Publishing Group, 7, p. 13049. doi: 10.1038/ncomms13049.

Chanet, R. and von Borstel, R. C. (1979) 'Genetic effects of formaldehyde in yeast. III. Nuclear and cytoplasmic mutagenic effects.', *Mutation research*. Netherlands, 62(2), pp. 239–253.

Che, J. *et al.* (2015) 'Hyper-Acetylation of Histone H3K56 Limits Break-Induced Replication by Inhibiting Extensive Repair Synthesis.', *PLoS genetics*. United States, 11(2), p. e1004990. doi: 10.1371/journal.pgen.1004990.

Chen, L. *et al.* (2001) 'Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes.', *Molecular cell*. United States, 8(5), pp. 1105–1115.

Chini, C. C. S. and Chen, J. (2004) 'Claspin, a regulator of Chk1 in DNA replication stress pathway.', *DNA repair*. Netherlands, 3(8–9), pp. 1033–1037. doi: 10.1016/j.dnarep.2004.03.001.

Chong, J. P. *et al.* (2000) 'A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 97(4), pp. 1530–1535. doi: 10.1073/pnas.030539597.

Coquelle, A. *et al.* (1998) 'A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs.', *Molecular cell*. United States, 2(2), pp. 259–265.

Corcoran, R. B. *et al.* (2010) 'BRAF gene amplification can promote acquired resistance to MEK inhibitors in cancer cells harboring the BRAF V600E mutation.', *Science signaling*. United States, 3(149), p. ra84. doi: 10.1126/scisignal.2001148.

Cost, G. J. and Boeke, J. D. (1996) 'A useful colony colour phenotype associated with the yeast selectable/counter-selectable marker MET15.', *Yeast (Chichester, England)*. England, 12(10), pp. 939–941. doi: 10.1002/(SICI)1097-0061(199608)12:10<939::AID-YEA988>3.0.CO;2-L.

Coste, A. *et al.* (2007) 'Genotypic evolution of azole resistance mechanisms in sequential Candida albicans isolates.', *Eukaryotic cell*. United States, 6(10),

pp. 1889–1904. doi: 10.1128/EC.00151-07.

Cowen, L. E. *et al.* (2015) ‘Mechanisms of Antifungal Drug Resistance’, *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 5(7), p. a019752. doi: 10.1101/cshperspect.a019752.

Cox, D., Yuncken, C. and Spriggs, A. I. (1965) ‘Minute chromatin bodies in malignant tumours of childhood.’, *Lancet (London, England)*. England, 1(7402), pp. 55–58.

Craddock, N. *et al.* (2010) ‘Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.’, *Nature*. England, 464(7289), pp. 713–720. doi: 10.1038/nature08979.

Cruz, C. and Houseley, J. (2014) ‘Endogenous RNA interference is driven by copy number.’, *eLife*. England, 3, p. e01581. doi: 10.7554/eLife.01581.

Culotta, V. C., Howard, W. R. and Liu, X. F. (1994) ‘CRS5 encodes a metallothionein-like protein in Saccharomyces cerevisiae.’, *The Journal of biological chemistry*. United States, 269(41), pp. 25295–25302.

Dancis, A. *et al.* (1992) ‘Ferric reductase of Saccharomyces cerevisiae: molecular characterization, role in iron uptake, and transcriptional control by iron.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 89(9), pp. 3869–3873.

Dancis, A., Yuan, D. S., *et al.* (1994) ‘Molecular characterization of a copper transport protein in S. cerevisiae: an unexpected role for copper in iron transport.’, *Cell*. United States, 76(2), pp. 393–402.

Dancis, A., Haile, D., *et al.* (1994) ‘The Saccharomyces cerevisiae copper transport protein (Ctr1p). Biochemical characterization, regulation by copper, and physiologic role in copper uptake.’, *The Journal of biological chemistry*. United States, 269(41), pp. 25660–25667.

Das, C. *et al.* (2009) ‘CBP/p300-mediated acetylation of histone H3 on lysine 56.’, *Nature*. England, 459(7243), pp. 113–117. doi: 10.1038/nature07861.

Datta, A. and Jinks-Robertson, S. (1995) ‘Association of increased spontaneous mutation rates with high levels of transcription in yeast.’, *Science (New York, N.Y.)*. United States, 268(5217), pp. 1616–1619.

Deem, A. *et al.* (2008) ‘Defective Break-Induced Replication Leads to Half-Crossovers in Saccharomyces cerevisiae’;, *Genetics*, 179(4), p. 1845 LP-1860.

Defossez, P. A. *et al.* (1999) ‘Elimination of replication block protein Fob1 extends the life span of yeast mother cells.’, *Molecular cell*, 3(4), pp. 447–55.

Degrassi, G. *et al.* (1999) ‘Purification and properties of an esterase from the

yeast Saccharomyces cerevisiae and identification of the encoding gene.', *Applied and environmental microbiology*. United States, 65(8), pp. 3470–3472.

Delbrück, M. and Bailey, W. T. (1946) 'Induced mutations in bacterial viruses', *Cold Spring Harbor Symposia on Quantitative Biology* , 11, pp. 33–37. doi: 10.1101/SQB.1946.011.01.006.

Deng, S. K. *et al.* (2014) 'RPA antagonizes microhomology-mediated repair of DNA double-strand breaks.', *Nature structural & molecular biology*. United States, 21(4), pp. 405–412. doi: 10.1038/nsmb.2786.

DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997) 'Exploring the metabolic and genetic control of gene expression on a genomic scale.', *Science (New York, N.Y.)*. United States, 278(5338), pp. 680–686.

Derr, L. K. and Strathern, J. N. (1993) 'A role for reverse transcripts in gene conversion.', *Nature*. England, 361(6408), pp. 170–173. doi: 10.1038/361170a0.

Derr, L. K., Strathern, J. N. and Garfinkel, D. J. (1991) 'RNA-mediated recombination in S. cerevisiae.', *Cell*. United States, 67(2), pp. 355–364.

Dieckhoff, P. *et al.* (2004) 'Smt3/SUMO and Ubc9 are required for efficient APC/C-mediated proteolysis in budding yeast.', *Molecular microbiology*. England, 51(5), pp. 1375–1387. doi: 10.1046/j.1365-2958.2003.03910.x.

Ding, J. *et al.* (2009) 'Tolerance and stress response to ethanol in the yeast Saccharomyces cerevisiae.', *Applied microbiology and biotechnology*. Germany, 85(2), pp. 253–263. doi: 10.1007/s00253-009-2223-1.

Dittwald, P. *et al.* (2013) 'NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits.', *Genome research*. United States, 23(9), pp. 1395–1409. doi: 10.1101/gr.152454.112.

Dominguez-Sanchez, M. S. *et al.* (2011) 'Genome instability and transcription elongation impairment in human cells depleted of THO/TREX.', *PLoS genetics*. United States, 7(12), p. e1002386. doi: 10.1371/journal.pgen.1002386.

Dorfman, B. Z. (1969) 'The isolation of adenylosuccinate synthetase mutants in yeast by selection for constitutive behavior in pigmented strains.', *Genetics*. United States, 61(2), pp. 377–389.

Dorsey, M. *et al.* (1992) 'Spontaneous amplification of the ADH4 gene in Saccharomyces cerevisiae.', *Genetics*. United States, 132(4), pp. 943–950.

Driscoll, R., Hudson, A. and Jackson, S. P. (2007) 'Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56.', *Science (New York,*

N.Y.). United States, 315(5812), pp. 649–652. doi: 10.1126/science.1135862.

Drolet, M. *et al.* (1995) 'Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 92(8), pp. 3526–3530.

Drolet, M., Bi, X. and Liu, L. F. (1994) 'Hypernegative supercoiling of the DNA template during transcription elongation in vitro.', *The Journal of biological chemistry*. United States, 269(3), pp. 2068–2074.

Duch, A. *et al.* (2013) 'Coordinated control of replication and transcription by a SAPK protects genomic integrity.', *Nature*. England, 493(7430), pp. 116–119. doi: 10.1038/nature11675.

Duch, A. *et al.* (2018) 'Multiple signaling kinases target Mrc1 to prevent genomic instability triggered by transcription-replication conflicts', *Nature Communications*. London: Nature Publishing Group UK, 9, p. 379. doi: 10.1038/s41467-017-02756-x.

Dunham, M. J. *et al.* (2002) 'Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 99(25), pp. 16144–16149. doi: 10.1073/pnas.242624799.

Duro, E. *et al.* (2008) 'Budding yeast Mms22 and Mms1 regulate homologous recombination induced by replisome blockage.', *DNA repair*. Netherlands, 7(5), pp. 811–818. doi: 10.1016/j.dnarep.2008.01.007.

Duvel, K. and Broach, J. R. (2004) 'The role of phosphatases in TOR signaling in yeast.', *Current topics in microbiology and immunology*. Germany, 279, pp. 19–38.

Ecker, D. J. *et al.* (1986) 'Yeast metallothionein function in metal ion detoxification.', *The Journal of biological chemistry*. United States, 261(36), pp. 16895–16900.

Ferguson, D. O. and Holloman, W. K. (1996) 'Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 93(11), pp. 5419–5424.

Fernandes, A. R. *et al.* (2005) 'Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes.', *Biochemical and biophysical research communications*. United States, 337(1), pp. 95–103. doi: 10.1016/j.bbrc.2005.09.010.

Fernandez, M. R. *et al.* (1995) 'Class III alcohol dehydrogenase from Saccharomyces cerevisiae: structural and enzymatic features differ toward

the human/mammalian forms in a manner consistent with functional needs in formaldehyde detoxication.', *FEBS letters*. England, 370(1–2), pp. 23–26.

Field, C. and Schekman, R. (1980) 'Localized secretion of acid phosphatase reflects the pattern of cell surface growth in Saccharomyces cerevisiae.', *The Journal of cell biology*. United States, 86(1), pp. 123–128.

Fillingham, J. *et al.* (2008) 'Chaperone control of the activity and specificity of the histone H3 acetyltransferase Rtt109.', *Molecular and cellular biology*. United States, 28(13), pp. 4342–4353. doi: 10.1128/MCB.00182-08.

Fogel, S. and Welch, J. W. (1982) 'Tandem gene amplification mediates copper resistance in yeast.', *Proceedings of the National Academy of Sciences of the United States of America*, 79(17), pp. 5342–6.

Fong, C. M., Arumugam, A. and Koepp, D. M. (2013) 'The Saccharomyces cerevisiae F-box protein Dia2 is a mediator of S-phase checkpoint recovery from DNA damage.', *Genetics*. United States, 193(2), pp. 483–499. doi: 10.1534/genetics.112.146373.

Force, A. *et al.* (1999) 'Preservation of duplicate genes by complementary, degenerative mutations.', *Genetics*. United States, 151(4), pp. 1531–1545.

Foster, P. L. (1997) 'Nonadaptive mutations occur on the F' episome during adaptive mutation conditions in Escherichia coli.', *Journal of bacteriology*. United States, 179(5), pp. 1550–1554.

Foster, P. L. and Trimarchi, J. M. (1994) 'Adaptive reversion of a frameshift mutation in Escherichia coli by simple base deletions in homopolymeric runs.', *Science (New York, N.Y.)*. United States, 265(5170), pp. 407–409.

Fraga, C. G. *et al.* (1990) 'Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 87(12), pp. 4533–4537.

Francke, U. (1999) 'Williams-Beuren Syndrome: Genes and Mechanisms', *Human Molecular Genetics*, 8(10), pp. 1947–1954.

Frei, E. 3rd *et al.* (1984) 'Development of methotrexate resistance in a human squamous cell carcinoma of the head and neck in culture.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 81(9), pp. 2873–2877.

French, S. L. *et al.* (2003) 'In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes.', *Molecular and cellular biology*, 23(5), pp. 1558–68.

Frenk, S. *et al.* (2017) 'Aging yeast gain a competitive advantage on non-

optimal carbon sources.', *Aging cell*. England, 16(3), pp. 602–604. doi: 10.1111/acel.12582.

Fricke, W. M. and Brill, S. J. (2003) 'Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3.', *Genes & development*. United States, 17(14), pp. 1768–1778. doi: 10.1101/gad.1105203.

Gabellini, D., Green, M. R. and Tupler, R. (2002) 'Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle.', *Cell*. United States, 110(3), pp. 339–348.

Galhardo, R. S., Hastings, P. J. and Rosenberg, S. M. (2007) 'Mutation as a stress response and the regulation of evolvability.', *Critical reviews in biochemistry and molecular biology*. England, 42(5), pp. 399–435. doi: 10.1080/10409230701648502.

Gali, V. K. *et al.* (2017) 'Translesion synthesis DNA polymerase η exhibits a specific RNA extension activity and a transcription-associated function', *Scientific Reports*, 7(1), p. 13055. doi: 10.1038/s41598-017-12915-1.

Gallo, C. M., Smith, D. L. J. and Smith, J. S. (2004) 'Nicotinamide clearance by Pnc1 directly regulates Sir2-mediated silencing and longevity.', *Molecular and cellular biology*. United States, 24(3), pp. 1301–1312.

Gambus, A. *et al.* (2009) 'A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome.', *The EMBO journal*. England, 28(19), pp. 2992–3004. doi: 10.1038/emboj.2009.226.

Ganley, A. R. D. *et al.* (2018) 'The Effect of Replication Initiation on Gene Amplification in the rDNA and Its Relationship to Aging', *Molecular Cell*. Elsevier, 35(5), pp. 683–693. doi: 10.1016/j.molcel.2009.07.012.

Gao, S. *et al.* (2016) 'The non-homologous end-joining pathway of S. cerevisiae works effectively in G1-phase cells, and religates cognate ends correctly and non-randomly', *DNA repair*, 42, pp. 1–10. doi: 10.1016/j.dnarep.2016.03.013.

Gerik, K. J. *et al.* (1998) 'Characterization of the two small subunits of Saccharomyces cerevisiae DNA polymerase delta.', *The Journal of biological chemistry*. United States, 273(31), pp. 19747–19755.

Ghislain, M., Talla, E. and Francois, J. M. (2002) 'Identification and functional analysis of the Saccharomyces cerevisiae nicotinamidase gene, PNC1.', *Yeast (Chichester, England)*. England, 19(3), pp. 215–224. doi: 10.1002/yea.810.

Giaever, G. *et al.* (2002) 'Functional profiling of the Saccharomyces cerevisiae genome.', *Nature*. England, 418(6896), pp. 387–391. doi: 10.1038/nature00935.

Gompel-Klein, P., Mack, M. and Brendel, M. (1989) 'Molecular characterization of the two genes SNQ and SFA that confer hyperresistance to 4-nitroquinoline-N-oxide and formaldehyde in Saccharomyces cerevisiae.', *Current genetics*. United States, 16(2), pp. 65–74.

Gorman, J. A. *et al.* (1986) 'Regulation of the yeast metallothionein gene.', *Gene*. Netherlands, 48(1), pp. 13–22.

Gorter de Vries, A. R., Pronk, J. T. and Daran, J.-M. G. (2017) 'Industrial Relevance of Chromosomal Copy Number Variation in Saccharomyces Yeasts', *Applied and Environmental Microbiology* , 83(11). doi: 10.1128/AEM.03206-16.

Gottlieb, S. and Esposito, R. E. (1989) 'A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA.', *Cell*, 56(5), pp. 771–6.

de Greef, J. C., Frants, R. R. and van der Maarel, S. M. (2008) 'Epigenetic mechanisms of facioscapulohumeral muscular dystrophy.', *Mutation research*. Netherlands, 647(1–2), pp. 94–102. doi: 10.1016/j.mrfmmm.2008.07.011.

Gresham, D. *et al.* (2008) 'The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast.', *PLoS genetics*. United States, 4(12), p. e1000303. doi: 10.1371/journal.pgen.1000303.

Gresham, D. *et al.* (2010) 'Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus', *Proceedings of the National Academy of Sciences*, 107(43), p. 18551 LP-18556.

Grey, M., Schmidt, M. and Brendel, M. (1996) 'Overexpression of ADH1 confers hyper-resistance to formaldehyde in Saccharomyces cerevisiae.', *Current genetics*. United States, 29(5), pp. 437–440.

Gruber, M., Wellinger, R. E. and Sogo, J. M. (2000) 'Architecture of the replication fork stalled at the 3' end of yeast ribosomal genes.', *Molecular and cellular biology*. United States, 20(15), pp. 5777–5787.

Ha, C. W. and Huh, W.-K. (2011) 'Rapamycin increases rDNA stability by enhancing association of Sir2 with rDNA in Saccharomyces cerevisiae.', *Nucleic acids research*. England, 39(4), pp. 1336–1350. doi: 10.1093/nar/gkq895.

El Hage, A. *et al.* (2010) 'Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis.', *Genes & development*. United States, 24(14), pp. 1546–1558. doi: 10.1101/gad.573310.

Hahn, M. W. (2009) ‘Distinguishing among evolutionary models for the maintenance of gene duplicates.’, *The Journal of heredity*. United States, 100(5), pp. 605–617. doi: 10.1093/jhered/esp047.

Hakem, R. (2008) ‘DNA-damage repair; the good, the bad, and the ugly’, *The EMBO Journal*. Nature Publishing Group, 27(4), pp. 589–605. doi: 10.1038/emboj.2008.15.

Hamer, D. H., Thiele, D. J. and Lemontt, J. E. (1985) ‘Function and autoregulation of yeast copperthionein.’, *Science (New York, N.Y.)*. United States, 228(4700), pp. 685–690.

Hammet, A., Pike, B. L. and Heierhorst, J. (2002) ‘Posttranscriptional regulation of the RAD5 DNA repair gene by the Dun1 kinase and the Pan2-Pan3 poly(A)-nuclease complex contributes to survival of replication blocks.’, *The Journal of biological chemistry*. United States, 277(25), pp. 22469–22474. doi: 10.1074/jbc.M202473200.

Han, J., Zhou, H., Li, Z., *et al.* (2007) ‘Acetylation of Lysine 56 of Histone H3 Catalyzed by RTT109 and Regulated by ASF1 Is Required for Replisome Integrity’, *Journal of Biological Chemistry* , 282(39), pp. 28587–28596. doi: 10.1074/jbc.M702496200.

Han, J., Zhou, H., Horazdovsky, B., *et al.* (2007) ‘Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication.’, *Science (New York, N.Y.)*. United States, 315(5812), pp. 653–655. doi: 10.1126/science.1133234.

Hanada, K. *et al.* (2007) ‘The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks.’, *Nature structural & molecular biology*. United States, 14(11), pp. 1096–1104. doi: 10.1038/nsmb1313.

Hansche, P. E., Beres, V. and Lange, P. (1978) ‘Gene duplication in Saccharomyces cerevisiae.’, *Genetics*. United States, 88(4 Pt 1), pp. 673–687.

Harper, J. W. and Elledge, S. J. (2018) ‘The DNA Damage Response: Ten Years After’, *Molecular Cell*. Elsevier, 28(5), pp. 739–745. doi: 10.1016/j.molcel.2007.11.015.

Hassett, R. and Kosman, D. J. (1995) ‘Evidence for Cu(II) reduction as a component of copper uptake by Saccharomyces cerevisiae.’, *The Journal of biological chemistry*. United States, 270(1), pp. 128–134.

Hastings, P. J. *et al.* (2000) ‘Adaptive Amplification: An Inducible Chromosomal Instability Mechanism’, *Cell*, 103(5), pp. 723–731. doi: https://doi.org/10.1016/S0092-8674(00)00176-8.

Hastings, P. J. *et al.* (2009) ‘Mechanisms of change in gene copy number’, *Nature Reviews Genetics*. Nature Publishing Group, 10, p. 551.

Hastings, P. J., Ira, G. and Lupski, J. R. (2009) 'A microhomology-mediated break-induced replication model for the origin of human copy number variation.', *PLoS genetics*. United States, 5(1), p. e1000327. doi: 10.1371/journal.pgen.1000327.

Heidenreich, E. and Wintersberger, U. (1997) 'Starvation for a specific amino acid induces high frequencies of rho- mutants in Saccharomyces cerevisiae.', *Current genetics*. United States, 31(5), pp. 408–413.

Heitman, J. *et al.* (1991) 'FK 506-binding protein proline rotamase is a target for the immunosuppressive agent FK 506 in Saccharomyces cerevisiae', *Proceedings of the National Academy of Sciences*, 88(5), p. 1948 LP-1952.

Heitman, J., Movva, N. R. and Hall, M. N. (1991) 'Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast.', *Science (New York, N.Y.)*. United States, 253(5022), pp. 905–909.

Helbock, H. J. *et al.* (1998) 'DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 95(1), pp. 288–293.

Helliwell, S. B. *et al.* (1994) 'TOR1 and TOR2 are structurally and functionally similar but not identical phosphatidylinositol kinase homologues in yeast.', *Molecular biology of the cell*. United States, 5(1), pp. 105–118.

Hoang, M. L. *et al.* (2010) 'Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination.', *PLoS genetics*. United States, 6(12), p. e1001228. doi: 10.1371/journal.pgen.1001228.

Hoeijmakers, J. H. J. (2009) 'DNA damage, aging, and cancer.', *The New England journal of medicine*. United States, 361(15), pp. 1475–1485. doi: 10.1056/NEJMra0804615.

Holmes, G. E., Bernstein, C. and Bernstein, H. (1992) 'Oxidative and other DNA damages as the basis of aging: a review.', *Mutation research*. Netherlands, 275(3–6), pp. 305–315.

Horowitz, H. and Haber, J. E. (1985) 'Identification of autonomously replicating circular subtelomeric Y' elements in Saccharomyces cerevisiae.', *Molecular and Cellular Biology* , 5(9), pp. 2369–2380. doi: 10.1128/MCB.5.9.2369.

Houseley, J. *et al.* (2007) 'Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control.', *The EMBO journal*. England, 26(24), pp. 4996–5006. doi: 10.1038/sj.emboj.7601921.

Houseley, J. *et al.* (2008) 'A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster.', *Molecular cell*. United

States, 32(5), pp. 685–695. doi: 10.1016/j.molcel.2008.09.027.

Houseley, J. and Tollervey, D. (2006) 'Yeast Trf5p is a nuclear poly(A) polymerase.', *EMBO reports*. England, 7(2), pp. 205–211. doi: 10.1038/sj.embor.7400612.

Houseley, J. and Tollervey, D. (2009) 'The Many Pathways of RNA Degradation', *Cell*, 136(4), pp. 763–776. doi: https://doi.org/10.1016/j.cell.2009.01.019.

Houseley, J. and Tollervey, D. (2011) 'Repeat expansion in the budding yeast ribosomal DNA can occur independently of the canonical homologous recombination machinery.', *Nucleic acids research*. England, 39(20), pp. 8778–8791. doi: 10.1093/nar/gkr589.

Huang, J. *et al.* (2006) 'Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer.', *Genes & development*. United States, 20(20), pp. 2887–2901. doi: 10.1101/gad.1472706.

Huang, M.-E. *et al.* (2002) 'Pol32, a subunit of Saccharomyces cerevisiae DNA polymerase delta, suppresses genomic deletions and is involved in the mutagenic bypass pathway.', *Genetics*. United States, 160(4), pp. 1409–1422.

Huang, T. and Campbell, J. L. (1995) 'Amplification of a circular episome carrying an inverted repeat of the DFR1 locus and adjacent autonomously replicating sequence element of Saccharomyces cerevisiae.', *The Journal of biological chemistry*. United States, 270(16), pp. 9607–9614.

Huertas, P. (2010) 'DNA resection in eukaryotes: deciding how to fix the break', *Nature Structural &Amp; Molecular Biology*. Nature Publishing Group, 17, p. 11.

Huertas, P. and Aguilera, A. (2003) 'Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.', *Molecular cell*. United States, 12(3), pp. 711–721.

Hughes, A. L. (1994) 'The evolution of functionally novel proteins after gene duplication.', *Proceedings. Biological sciences*. England, 256(1346), pp. 119–124. doi: 10.1098/rspb.1994.0058.

Huh, W.-K. *et al.* (2003) 'Global analysis of protein localization in budding yeast.', *Nature*. England, 425(6959), pp. 686–691. doi: 10.1038/nature02026.

Hull, R. M. *et al.* (2017) 'Environmental change drives accelerated adaptation through stimulated copy number variation', *PLoS Biology*. San Francisco, CA USA. doi: 10.1371/journal.pbio.2001333.

Iafrate, A. J. *et al.* (2004) 'Detection of large-scale variation in the human genome.', *Nature genetics*. United States, 36(9), pp. 949–951. doi:

10.1038/ng1416.

Ide, S., Saka, K. and Kobayashi, T. (2013) ‘Rtt109 prevents hyper-amplification of ribosomal RNA genes through histone modification in budding yeast.’, *PLoS genetics*. United States, 9(4), p. e1003410. doi: 10.1371/journal.pgen.1003410.

Imai, S. *et al.* (2000) ‘Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase’, *Nature*. Macmillian Magazines Ltd., 403, p. 795.

Ip, S. C. Y. *et al.* (2008) ‘Identification of Holliday junction resolvases from humans and yeast.’, *Nature*. England, 456(7220), pp. 357–361. doi: 10.1038/nature07470.

Irmisch, A. *et al.* (2009) ‘Smc5/6 maintains stalled replication forks in a recombination-competent conformation.’, *The EMBO journal*. England, 28(2), pp. 144–155. doi: 10.1038/emboj.2008.273.

Itoh, T. and Tomizawa, J. (1980) ‘Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 77(5), pp. 2450–2454.

Jacinto, E. *et al.* (2004) ‘Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive.’, *Nature cell biology*. England, 6(11), pp. 1122–1128. doi: 10.1038/ncb1183.

Jack, C. V. (2014) '*Regulated copy number changes in the rDNA locus of Saccharomyces cerevisiae*.', PhD thesis, University of Cambridge.

Jack, C. V *et al.* (2015) ‘Regulation of ribosomal DNA amplification by the TOR pathway.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 112(31), pp. 9674–9679. doi: 10.1073/pnas.1505015112.

Jackson, M. D. *et al.* (2003) ‘Mechanism of Nicotinamide Inhibition and Transglycosidation by Sir2 Histone/Protein Deacetylases’, *Journal of Biological Chemistry* , 278(51), pp. 50985–50998. doi: 10.1074/jbc.M306552200.

Jensen, L. T. *et al.* (1996) ‘Enhanced effectiveness of copper ion buffering by CUP1 metallothionein compared with CRS5 metallothionein in Saccharomyces cerevisiae.’, *The Journal of biological chemistry*. United States, 271(31), pp. 18514–18519.

Jessulat, M. *et al.* (2008) ‘Interacting proteins Rtt109 and Vps75 affect the efficiency of non-homologous end-joining in Saccharomyces cerevisiae.’, *Archives of biochemistry and biophysics*. United States, 469(2), pp. 157–164. doi: 10.1016/j.abb.2007.11.001.

Jeyaprakash, A., Welch, J. W. and Fogel, S. (1991) 'Multicopy CUP1 plasmids enhance cadmium and copper resistance levels in yeast.', *Molecular & general genetics : MGG*. Germany, 225(3), pp. 363–368.

Johansson, E., Garg, P. and Burgers, P. M. J. (2004) 'The Pol32 subunit of DNA polymerase delta contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding.', *The Journal of biological chemistry*. United States, 279(3), pp. 1907–1915. doi: 10.1074/jbc.M310362200.

Johansson, E., Majka, J. and Burgers, P. M. (2001) 'Structure of DNA polymerase delta from Saccharomyces cerevisiae.', *The Journal of biological chemistry*. United States, 276(47), pp. 43824–43828. doi: 10.1074/jbc.M108842200.

Karin, M. *et al.* (1984) 'Primary structure and transcription of an amplified genetic locus: the CUP1 locus of yeast.', *Proceedings of the National Academy of Sciences of the United States of America*, 81(2), pp. 337–41.

Katoh, M. (2008) 'Cancer genomics and genetics of FGFR2 (Review).', *International journal of oncology*. Greece, 33(2), pp. 233–237.

Katou, Y. *et al.* (2003) 'S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex', *Nature*. Macmillian Magazines Ltd., 424, p. 1078.

Keil, R. L. and Roeder, G. S. (1984) 'Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of S. cerevisiae.', *Cell*, 39(2 Pt 1), pp. 377–86.

Keller, G. *et al.* (2001) 'Haa1, a protein homologous to the copper-regulated transcription factor Ace1, is a novel transcriptional activator.', *The Journal of biological chemistry*. United States, 276(42), pp. 38697–38702. doi: 10.1074/jbc.M107131200.

Keskin, H. *et al.* (2014) 'Transcript-RNA-templated DNA recombination and repair.', *Nature*. England, 515(7527), pp. 436–439. doi: 10.1038/nature13682.

Keskin, H., Meers, C. and Storici, F. (2016) 'Transcript RNA supports precise repair of its own DNA gene.', *RNA biology*. United States, 13(2), pp. 157–165. doi: 10.1080/15476286.2015.1116676.

Keszthelyi, A., Minchell, N. E. and Baxter, J. (2016) 'The Causes and Consequences of Topological Stress during DNA Replication', *Genes*. Edited by E. Noguchi. MDPI, 7(12), p. 134. doi: 10.3390/genes7120134.

Kim, Y.-J. and Wilson, D. M. 3rd (2012) 'Overview of base excision repair biochemistry.', *Current molecular pharmacology*. United Arab Emirates, 5(1), pp. 3–13.

Knight, S. A. *et al.* (1996) ‘A widespread transposable element masks expression of a yeast copper transport gene.’, *Genes & development*. United States, 10(15), pp. 1917–1929.

Kobayashi, T. *et al.* (1998) ‘Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I.’, *Genes & development*. United States, 12(24), pp. 3821–3830.

Kobayashi, T. *et al.* (2004) ‘SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast.’, *Cell*, 117(4), pp. 441–53.

Kobayashi, T. and Ganley, A. R. D. (2005) ‘Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats.’, *Science (New York, N.Y.)*, 309(5740), pp. 1581–4. doi: 10.1126/science.1116102.

Kobayashi, T. and Horiuchi, T. (1996) ‘A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities.’, *Genes to cells : devoted to molecular & cellular mechanisms*, 1(5), pp. 465–74.

Kobayashi, T., Nomura, M. and Horiuchi, T. (2001) ‘Identification of DNA cis elements essential for expansion of ribosomal DNA repeats in Saccharomyces cerevisiae.’, *Molecular and cellular biology*. United States, 21(1), pp. 136–147. doi: 10.1128/MCB.21.1.136-147.2001.

de Kok, S. *et al.* (2012) ‘Laboratory evolution of new lactate transporter genes in a jen1Δ mutant of Saccharomyces cerevisiae and their identification as ADY2 alleles by whole-genome resequencing and transcriptome analysis’, *FEMS Yeast Research*, 12(3), pp. 359–374.

Koltin, Y. *et al.* (1991) ‘Rapamycin sensitivity in Saccharomyces cerevisiae is mediated by a peptidyl-prolyl cis-trans isomerase related to human FK506-binding protein.’, *Molecular and Cellular Biology* , 11(3), pp. 1718–1723. doi: 10.1128/MCB.11.3.1718.

Kondrashov, F. A. *et al.* (2002) ‘Selection in the evolution of gene duplications.’, *Genome biology*. England, 3(2), p. RESEARCH0008.

Kornreich, R., Bishop, D. F. and Desnick, R. J. (1990) ‘α-Galactosidase A gene rearrangements causing Fabry disease. Identification of short direct repeats at breakpoints in an Alu-rich gene’, *Journal of Biological Chemistry*, 265(16), pp. 9319–9326.

Koszul, R. *et al.* (2004) ‘Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments.’, *The EMBO journal*. England, 23(1), pp. 234–243. doi: 10.1038/sj.emboj.7600024.

Kraus, E., Leung, W. Y. and Haber, J. E. (2001) ‘Break-induced replication: a review and an example in budding yeast.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 98(15), pp. 8255–8262. doi: 10.1073/pnas.151008198.

Kreuzer, K. N. and Brister, J. R. (2010) ‘Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives.’, *Virology journal*. England, 7, p. 358. doi: 10.1186/1743-422X-7-358.

Krogh, B. O. and Symington, L. S. (2004) ‘Recombination Proteins in Yeast’, *Annual Review of Genetics*. Annual Reviews, 38(1), pp. 233–271. doi: 10.1146/annurev.genet.38.072902.091500.

Krystal, M. *et al.* (1981) ‘Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 78(9), pp. 5744–5748.

Kuruvilla, F. G. and Schreiber, S. L. (1999) ‘The PIK-related kinases intercept conventional signaling pathways’, *Chemistry & Biology*, 6(5), pp. R129–R136. doi: https://doi.org/10.1016/S1074-5521(99)80070-2.

Labib, K. and Hodgson, B. (2007) ‘Replication fork barriers: pausing for a break or stalling for time?’, *EMBO reports*. England, 8(4), pp. 346–353. doi: 10.1038/sj.embor.7400940.

Labib, K., Tercero, J. A. and Diffley, J. F. (2000) ‘Uninterrupted MCM2-7 function required for DNA replication fork progression.’, *Science (New York, N.Y.)*. United States, 288(5471), pp. 1643–1647.

Lakowski, B. and Hekimi, S. (1998) ‘The genetics of caloric restriction in Caenorhabditis elegans.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 95(22), pp. 13091–13096.

Lambert, S. *et al.* (2010) ‘Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange.’, *Molecular cell*. United States, 39(3), pp. 346–359. doi: 10.1016/j.molcel.2010.07.015.

Landry, J. *et al.* (2000) ‘The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 97(11), pp. 5807–5811. doi: 10.1073/pnas.110148297.

Landry, J., Slama, J. T. and Sternglanz, R. (2000) ‘Role of NAD(+) in the deacetylase activity of the SIR2-like proteins.’, *Biochemical and biophysical research communications*. United States, 278(3), pp. 685–690. doi: 10.1006/bbrc.2000.3854.

Lawrence, C. W., Krauss, B. R. and Christensen, R. B. (1985) 'New mutations affecting induced mutagenesis in yeast.', *Mutation research*. Netherlands, 150(1–2), pp. 211–216.

Lebenthal, E. (1987) 'Role of salivary amylase in gastric and intestinal digestion of starch.', *Digestive diseases and sciences*. United States, 32(10), pp. 1155–1157.

Lederberg, J. and Lederberg, E. M. (1952) 'REPLICA PLATING AND INDIRECT SELECTION OF BACTERIAL MUTANTS', *Journal of Bacteriology*, 63(3), pp. 399–406.

Lee, J. A., Carvalho, C. M. B. and Lupski, J. R. (2007) 'A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders', *Cell*, 131(7), pp. 1235–1247. doi: https://doi.org/10.1016/j.cell.2007.11.037.

Leman, A. R. and Noguchi, E. (2013) 'The Replication Fork: Understanding the Eukaryotic Replication Machinery and the Challenges to Genome Duplication', *Genes*. MDPI, 4(1), pp. 1–32. doi: 10.3390/genes4010001.

Lenski, R. E. and Mittler, J. E. (1993) 'The directed mutation controversy and neo-Darwinism.', *Science (New York, N.Y.)*. United States, 259(5092), pp. 188–194.

Li, C., Mueller, J. E. and Bryk, M. (2006) 'Sir2 represses endogenous polymerase II transcription units in the ribosomal DNA nontranscribed spacer.', *Molecular biology of the cell*. United States, 17(9), pp. 3848–3859. doi: 10.1091/mbc.E06-03-0205.

Li, M. *et al.* (2013) 'Genome-wide analysis of functional sirtuin chromatin targets in yeast.', *Genome biology*. England, 14(5), p. R48. doi: 10.1186/gb-2013-14-5-r48.

Li, X. and Manley, J. L. (2006) 'Cotranscriptional processes and their influence on genome stability.', *Genes & development*. United States, 20(14), pp. 1838–1847. doi: 10.1101/gad.1438306.

Libuda, D. E. and Winston, F. (2006) 'Amplification of histone genes by circular chromosome formation in Saccharomyces cerevisiae.', *Nature*. England, 443(7114), pp. 1003–1007. doi: 10.1038/nature05205.

Lieber, M. R. *et al.* (2004) 'The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination.', *DNA repair*. Netherlands, 3(8–9), pp. 817–826. doi: 10.1016/j.dnarep.2004.03.015.

Lieber, M. R. (2010) 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.', *Annual review of biochemistry*. United States, 79, pp. 181–211. doi: 10.1146/annurev.biochem.052308.093131.

Lin, F. L., Sperle, K. and Sternberg, N. (1984) 'Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process.', *Molecular and cellular biology*. United States, 4(6), pp. 1020–1034.

Lin, F. L., Sperle, K. and Sternberg, N. (1985) 'Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 82(5), pp. 1391–1395.

Lin, S. J., Defossez, P. A. and Guarente, L. (2000) 'Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae.', *Science (New York, N.Y.)*, 289(5487), pp. 2126–8.

Lindstrom, D. L. *et al.* (2011) 'Replicative age induces mitotic recombination in the ribosomal RNA gene cluster of Saccharomyces cerevisiae.', *PLoS genetics*. United States, 7(3), p. e1002015. doi: 10.1371/journal.pgen.1002015.

Lindstrom, D. L. and Gottschling, D. E. (2009) 'The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae.', *Genetics*. United States, 183(2), p. 413–22, 1SI–13SI. doi: 10.1534/genetics.109.106229.

Little, A. S. *et al.* (2011) 'Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells.', *Science signaling*, 4(166), p. ra17. doi: 10.1126/scisignal.2001752.

Liu, L. *et al.* (2000) 'Identification of a fourth subunit of mammalian DNA polymerase delta.', *The Journal of biological chemistry*. United States, 275(25), pp. 18739–18744. doi: 10.1074/jbc.M001217200.

Liu, P. *et al.* (2012) 'Mechanisms for recurrent and complex human genomic rearrangements.', *Current opinion in genetics & development*. England, 22(3), pp. 211–220. doi: 10.1016/j.gde.2012.02.012.

Liu, X. D. and Thiele, D. J. (1996) 'Oxidative stress induced heat shock factor phosphorylation and HSF-dependent activation of yeast metallothionein gene transcription.', *Genes & development*, 10(5), pp. 592–603.

Llorente, B., Smith, C. E. and Symington, L. S. (2008) 'Break-induced replication: what is it and what is it for?', *Cell cycle (Georgetown, Tex.)*. United States, 7(7), pp. 859–864. doi: 10.4161/cc.7.7.5613.

Loewith, R. *et al.* (2002) 'Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control.', *Molecular cell*. United States, 10(3), pp. 457–468.

Loewith, R. and Hall, M. N. (2011) ‘Target of Rapamycin (TOR) in Nutrient Signaling and Growth Control’, *Genetics*, 189(4), p. 1177 LP-1201.

Loo, S. and Rine, J. (1995) ‘Silencing and heritable domains of gene expression.’, *Annual review of cell and developmental biology*. United States, 11, pp. 519–548. doi: 10.1146/annurev.cb.11.110195.002511.

van Loon, N., Miller, D. and Murnane, J. P. (1994) ‘Formation of extrachromosomal circular DNA in HeLa cells by nonhomologous recombination.’, *Nucleic acids research*. England, 22(13), pp. 2447–2452.

Lopes da Rosa, J. *et al.* (2013) ‘A small molecule inhibitor of fungal histone acetyltransferase Rtt109.’, *Bioorganic & medicinal chemistry letters*. England, 23(10), pp. 2853–2859. doi: 10.1016/j.bmcl.2013.03.112.

Lorenz, M. C. and Heitman, J. (1995) ‘TOR mutations confer rapamycin resistance by preventing interaction with FKBP12-rapamycin.’, *The Journal of biological chemistry*. United States, 270(46), pp. 27531–27537.

Lou, H. *et al.* (2008) ‘Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint.’, *Molecular cell*. United States, 32(1), pp. 106–117. doi: 10.1016/j.molcel.2008.08.020.

Lunt, P. W. *et al.* (1995) ‘Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD)’, *Human molecular genetics*. England, 4(5), pp. 951–958.

Lupski, J. R. (1998) ‘Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits’, *Trends in Genetics*, 14(10), pp. 417–422. doi: https://doi.org/10.1016/S0168-9525(98)01555-8.

Luria, S. E. and Delbrück, M. (1943) ‘Mutations of Bacteria from Virus Sensitivity to Virus Resistance.’, *Genetics*, 28(6), pp. 491–511.

van der Maarel, S. M., Tawil, R. and Tapscott, S. J. (2011) ‘Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence.’, *Trends in molecular medicine*. England, 17(5), pp. 252–258. doi: 10.1016/j.molmed.2011.01.001.

Maas, N. L. *et al.* (2006) ‘Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4.’, *Molecular cell*. United States, 23(1), pp. 109–119. doi: 10.1016/j.molcel.2006.06.006.

Mager, W. H. and De Kruijff, A. J. (1995) ‘Stress-induced transcriptional activation.’, *Microbiological reviews*. United States, 59(3), pp. 506–531.

Mahajan, P. B. (1994) ‘Modulation of transcription of rRNA genes by

rapamycin.', *International journal of immunopharmacology*. England, 16(9), pp. 711–721.

Maher, R. L., Branagan, A. M. and Morrical, S. W. (2011) 'Coordination of DNA replication and recombination activities in the maintenance of genome stability.', *Journal of cellular biochemistry*. United States, 112(10), pp. 2672–2682. doi: 10.1002/jcb.23211.

Makarova, A. V, Stodola, J. L. and Burgers, P. M. (2012) 'A four-subunit DNA polymerase zeta complex containing Pol delta accessory subunits is essential for PCNA-mediated mutagenesis.', *Nucleic acids research*. England, 40(22), pp. 11618–11626. doi: 10.1093/nar/gks948.

Makino, T. and McLysaght, A. (2012) 'Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant.', *Genome research*. United States, 22(12), pp. 2427–2435. doi: 10.1101/gr.131953.111.

Mangiameli, S. M. *et al.* (2017) 'Transcription leads to pervasive replisome instability in bacteria.', *eLife*. England, 6. doi: 10.7554/eLife.19848.

Marcus, S. *et al.* (1993) 'Duplication in the hypoxanthine phosphoribosyl-transferase gene caused by Alu-Alu recombination in a patient with Lesch Nyhan syndrome', *Human Genetics*, 90(5), pp. 477–482. doi: 10.1007/BF00217444.

Marichal, P. *et al.* (1997) 'Molecular biological characterization of an azole-resistant Candida glabrata isolate.', *Antimicrobial agents and chemotherapy*. United States, 41(10), pp. 2229–2237.

Martens, M. A., Wilson, S. J. and Reutens, D. C. (2008) 'Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype.', *Journal of child psychology and psychiatry, and allied disciplines*. England, 49(6), pp. 576–608. doi: 10.1111/j.1469-7610.2008.01887.x.

Masumoto, H. *et al.* (2005) 'A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response', *Nature*. Nature Publishing Group, 436, p. 294.

Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Belknap Press.

McCulley, J. L. and Petes, T. D. (2010) 'Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 107(25), pp. 11465–11470. doi: 10.1073/pnas.1006281107.

McKenzie, G. J., Lombardo, M. J. and Rosenberg, S. M. (1998) 'Recombination-dependent mutation in Escherichia coli occurs in stationary

phase.', *Genetics*. United States, pp. 1163–1165.

McVey, M. and Lee, S. E. (2008) 'MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings', *Trends in Genetics*, 24(11), pp. 529–538. doi: https://doi.org/10.1016/j.tig.2008.08.007.

Medvedik, O. *et al.* (2007) 'MSN2 and MSN4 link calorie restriction and TOR to sirtuin-mediated lifespan extension in Saccharomyces cerevisiae.', *PLoS biology*. United States, 5(10), p. e261. doi: 10.1371/journal.pbio.0050261.

Meers, C., Keskin, H. and Storici, F. (2016) 'DNA repair by RNA: Templated, or not templated, that is the question.', *DNA repair*. Netherlands, 44, pp. 17–21. doi: 10.1016/j.dnarep.2016.05.002.

Mei, S.-C. and Brenner, C. (2015) 'Calorie Restriction-Mediated Replicative Lifespan Extension in Yeast Is Non-Cell Autonomous', *PLOS Biology*. Public Library of Science, 13(1), p. e1002048.

Metzgar, D. and Wills, C. (2000) 'Evidence for the Adaptive Evolution of Mutation Rates', *Cell*. Elsevier, 101(6), pp. 581–584. doi: 10.1016/S0092-8674(00)80869-7.

Miller, K. M., Maas, N. L. and Toczyski, D. P. (2006) 'Taking it off: regulation of H3 K56 acetylation by Hst3 and Hst4.', *Cell cycle (Georgetown, Tex.)*. United States, 5(22), pp. 2561–2565. doi: 10.4161/cc.5.22.3501.

Mirkin, E. V and Mirkin, S. M. (2005) 'Mechanisms of transcription-replication collisions in bacteria.', *Molecular and cellular biology*. United States, 25(3), pp. 888–895. doi: 10.1128/MCB.25.3.888-895.2005.

Mirkin, E. V and Mirkin, S. M. (2007) 'Replication Fork Stalling at Natural Impediments', *Microbiology and Molecular Biology Reviews* , 71(1), pp. 13–35. doi: 10.1128/MMBR.00030-06.

Mischo, H. E. *et al.* (2011) 'Yeast Sen1 helicase protects the genome from transcription-associated instability.', *Molecular cell*. United States, 41(1), pp. 21–32. doi: 10.1016/j.molcel.2010.12.007.

Mizuno, K. *et al.* (2013) 'Recombination-restarted replication makes inverted chromosome fusions at inverted repeats.', *Nature*. England, 493(7431), pp. 246–249.

Moazed, D. (2001) 'Common themes in mechanisms of gene silencing.', *Molecular cell*. United States, 8(3), pp. 489–498.

Mohanty, B. K., Bairwa, N. K. and Bastia, D. (2009) 'Contrasting roles of checkpoint proteins as recombination modulators at Fob1-Ter complexes with or without fork arrest.', *Eukaryotic cell*. United States, 8(4), pp. 487–495. doi: 10.1128/EC.00382-08.

Moller, H. D. *et al.* (2015) 'Extrachromosomal circular DNA is common in yeast.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 112(24), pp. E3114-22. doi: 10.1073/pnas.1508825112.

Moore, J. K. and Haber, J. E. (1996) 'Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks.', *Nature*. England, 383(6601), pp. 644–646. doi: 10.1038/383644a0.

Moretti, P. *et al.* (1994) 'Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1.', *Genes & development*. United States, 8(19), pp. 2257–2269.

Mortimer, R. K. *et al.* (1989) 'Genetic map of Saccharomyces cerevisiae, edition 10.', *Yeast (Chichester, England)*. England, 5(5), pp. 321–403. doi: 10.1002/yea.320050503.

Mortimer, R. K. and Schild, D. (1985) 'Genetic map of Saccharomyces cerevisiae, edition 9.', *Microbiological reviews*. United States, 49(3), pp. 181–213.

Mukherjee, K. and Storici, F. (2012) 'A mechanism of gene amplification driven by small DNA fragments.', *PLoS genetics*. United States, 8(12), p. e1003119. doi: 10.1371/journal.pgen.1003119.

Nakama, M. *et al.* (2012) 'DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation.', *Genes to cells : devoted to molecular & cellular mechanisms*. England, 17(3), pp. 218–233. doi: 10.1111/j.1365-2443.2012.01583.x.

Nam, E. A. and Cortez, D. (2011) 'ATR signalling: more than meeting at the fork.', *The Biochemical journal*. England, 436(3), pp. 527–536. doi: 10.1042/BJ20102162.

Narayanan, V. *et al.* (2006) 'The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks.', *Cell*. United States, 125(7), pp. 1283–1296. doi: 10.1016/j.cell.2006.04.042.

Naseeb, S. *et al.* (2017) 'Rapid functional and evolutionary changes follow gene duplication in yeast', *Proceedings of the Royal Society B: Biological Sciences*. The Royal Society, 284(1861), p. 20171393. doi: 10.1098/rspb.2017.1393.

Nassif, N. *et al.* (1994) 'Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair.', *Molecular and cellular biology*. United States, 14(3), pp. 1613–1625.

Nehlin, J. O., Carlberg, M. and Ronne, H. (1992) 'Yeast SKO1 gene encodes a bZIP protein that binds to the CRE motif and acts as a repressor of transcription.', *Nucleic acids research*. England, 20(20), pp. 5271–5278.

Newcombe, H. B. (1949) ‘Origin of Bacterial Variants’, *Nature*. Nature Publishing Group, 164, p. 150.

Olds, R. J. *et al.* (1993) ‘Complete Nucleotide Sequence of the Antithrombin Gene: Evidence for Homologous Recombination Causing Thrombophilia’, *Biochemistry*, 32(16), pp. 4216–4224. doi: 10.1021/bi00067a008.

Orr-Weaver, T. L. and Szostak, J. W. (1983) ‘Multiple, Tandem Plasmid Integration in Saccharomyces cerevisiae’, *Molecular and Cellular Biology*, 3(4), pp. 747–749.

Orr-Weaver, T. L., Szostak, J. W. and Rothstein, R. J. (1981) ‘Yeast transformation: a model system for the study of recombination.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 78(10), pp. 6354–6358.

Osborn, A. J. and Elledge, S. J. (2003) ‘Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53.’, *Genes & development*. United States, 17(14), pp. 1755–1767. doi: 10.1101/gad.1098303.

Oud, B. *et al.* (2013) ‘Genome duplication and mutations in ACE2 cause multicellular, fast-sedimenting phenotypes in evolved Saccharomyces cerevisiae.', *Proceedings of the National Academy of Sciences*, 110(45), p. E4223 LP-E4231.

Pacek, M. and Walter, J. C. (2004) ‘A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication.’, *The EMBO journal*. England, 23(18), pp. 3667–3676. doi: 10.1038/sj.emboj.7600369.

Pal, S. *et al.* (2018) ‘Impaired cohesion and homologous recombination during replicative aging in budding yeast’, *Science Advances*, 4(2).

Pandya, S., King, W. M. and Tawil, R. (2008) ‘Facioscapulohumeral dystrophy.’, *Physical therapy*. United States, 88(1), pp. 105–113. doi: 10.2522/ptj.20070104.

Pannunzio, N. R. *et al.* (2014) ‘Non-homologous end joining often uses microhomology: implications for alternative end joining.’, *DNA repair*. Netherlands, 17, pp. 74–80. doi: 10.1016/j.dnarep.2014.02.006.

Pâques, F. and Haber, J. E. (1999) ‘Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae’, *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 63(2), pp. 349–404.

Paul, S. *et al.* (2013) ‘Accelerated gene evolution through replication-transcription conflicts.’, *Nature*. England, 495(7442), pp. 512–515. doi: 10.1038/nature11989.

Paulsen, T. *et al.* (2018) 'Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells', *Trends in Genetics*. Elsevier. doi: 10.1016/j.tig.2017.12.010.

Paunesku, T. *et al.* (2001) 'Proliferating cell nuclear antigen (PCNA): ringmaster of the genome.', *International journal of radiation biology*. England, 77(10), pp. 1007–1021. doi: 10.1080/09553000110069335.

Payen, C. *et al.* (2008) 'Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms.', *PLoS genetics*. United States, 4(9), p. e1000175. doi: 10.1371/journal.pgen.1000175.

Payen, C. *et al.* (2016) 'High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations', *PLoS Genetics*. Edited by M. Snyder. San Francisco, CA USA: Public Library of Science, 12(10), p. e1006339. doi: 10.1371/journal.pgen.1006339.

Pellegrini, L. (2012) 'The Pol alpha-primase complex.', *Sub-cellular biochemistry*. United States, 62, pp. 157–169. doi: 10.1007/978-94-007-4572-8_9.

Peña, M. M. O., Koch, K. A. and Thiele, D. J. (1998) 'Dynamic Regulation of Copper Uptake and Detoxification Genes in Saccharomyces cerevisiae', *Molecular and Cellular Biology*. American Society for Microbiology, 18(5), pp. 2514–2523.

Perry, G. H. *et al.* (2007) 'Diet and the evolution of human amylase gene copy number variation', *Nature genetics*, 39(10), pp. 1256–1260. doi: 10.1038/ng2123.

Petermann, E., Helleday, T. and Caldecott, K. W. (2008) 'Claspin promotes normal replication fork rates in human cells.', *Molecular biology of the cell*. United States, 19(6), pp. 2373–2378. doi: 10.1091/mbc.E07-10-1035.

Pfeiffer, P., Goedecke, W. and Obe, G. (2000) 'Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations.', *Mutagenesis*. England, 15(4), pp. 289–302.

Piotrowski, A. *et al.* (2008) 'Somatic mosaicism for copy number variation in differentiated human tissues.', *Human mutation*, 29(9), pp. 1118–24. doi: 10.1002/humu.20815.

Podust, L. M. *et al.* (1995) 'Mammalian DNA polymerase auxiliary proteins: analysis of replication factor C-catalyzed proliferating cell nuclear antigen loading onto circular double-stranded DNA.', *Molecular and cellular biology*. United States, 15(6), pp. 3072–3081.

Podust, V. N. *et al.* (2002) 'Reconstitution of human DNA polymerase delta

using recombinant baculoviruses: the p12 subunit potentiates DNA polymerizing activity of the four-subunit enzyme.', *The Journal of biological chemistry*. United States, 277(6), pp. 3894–3901. doi: 10.1074/jbc.M109684200.

Pohjoismaki, J. L. O. *et al.* (2010) 'Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid.', *Journal of molecular biology*. England, 397(5), pp. 1144–1155. doi: 10.1016/j.jmb.2010.02.029.

Ponder, R. G., Fonville, N. C. and Rosenberg, S. M. (2005) 'A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation.', *Molecular cell*. United States, 19(6), pp. 791–804. doi: 10.1016/j.molcel.2005.07.025.

Pousi, B. *et al.* (1994) 'Alu-Alu recombination results in a duplication of seven exons in the lysyl hydroxylase gene in a patient with the type VI variant of Ehlers-Danlos syndrome', *American Journal of Human Genetics*, 55(5), pp. 899–906.

Prado, F. and Aguilera, A. (2005) 'Impairment of replication fork progression mediates RNA polII transcription-associated recombination', *The EMBO Journal*, 24(6), p. 1267 LP-1276.

Praekelt, U. M. and Meacock, P. A. (1990) 'HSP12, a new small heat shock gene of Saccharomyces cerevisiae: analysis of structure, regulation and function.', *Molecular & general genetics : MGG*. Germany, 223(1), pp. 97–106.

Prakash, S., Johnson, R. E. and Prakash, L. (2005) 'Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function.', *Annual review of biochemistry*. United States, 74, pp. 317–353. doi: 10.1146/annurev.biochem.74.082803.133250.

Preker, P. *et al.* (2011) 'PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters.', *Nucleic acids research*. England, 39(16), pp. 7179–7193. doi: 10.1093/nar/gkr370.

Radloff, R., Bauer, W. and Vinograd, J. (1967) 'A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 57(5), pp. 1514–1521.

Reardon, J. T. and Sancar, A. (2001) 'Excision Repair A2 - Brenner, Sydney', in Miller, J. H. B. T.-E. of G. (ed.). New York: Academic Press, pp. 673–675. doi: https://doi.org/10.1006/rwgn.2001.0437.

Redon, R. *et al.* (2006) 'Global variation in copy number in the human genome', *Nature*. Nature Publishing Group, 444, p. 444.

Rep, M. *et al.* (2001) 'The Saccharomyces cerevisiae Sko1p transcription factor mediates HOG pathway-dependent osmotic regulation of a set of genes encoding enzymes implicated in protection from oxidative damage.', *Molecular microbiology*. England, 40(5), pp. 1067–1083.

Richard, G.-F., Kerrest, A. and Dujon, B. (2008) 'Comparative genomics and molecular dynamics of DNA repeats in eukaryotes.', *Microbiology and molecular biology reviews : MMBR*. United States, 72(4), pp. 686–727. doi: 10.1128/MMBR.00011-08.

Rine, J. and Herskowitz, I. (1987) 'Four genes responsible for a position effect on expression from HML and HMR in Saccharomyces cerevisiae.', *Genetics*. United States, 116(1), pp. 9–22.

Rogina, B. and Helfand, S. L. (2004) 'Sir2 mediates longevity in the fly through a pathway related to calorie restriction.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(45), pp. 15998–16003. doi: 10.1073/pnas.0404184101.

Rosenberg, S. M. *et al.* (1994) 'Adaptive mutation by deletions in small mononucleotide repeats.', *Science (New York, N.Y.)*. United States, 265(5170), pp. 405–407.

Rosenberg, S. M. (2001) 'Evolving responsively: adaptive mutation', *Nature Reviews Genetics*. Nature Publishing Group, 2, p. 504.

Roth, G. S., Ingram, D. K. and Lane, M. A. (1999) 'Calorie restriction in primates: will it work and how will we know?', *Journal of the American Geriatrics Society*. United States, 47(7), pp. 896–903.

Roth, J. R. *et al.* (2003) 'Regulating general mutation rates: examination of the hypermutable state model for Cairnsian adaptive mutation.', *Genetics*. United States, 163(4), pp. 1483–1496.

Rüdiger, N. S. *et al.* (1991) 'DNA deletions in the low density lipoprotein (LDL) receptor gene in Danish families with familial hypercholesterolemia', *Clinical Genetics*, 39(6), pp. 451–462. doi: 10.1111/j.1399-0004.1991.tb03057.x.

Ruiz, J. F., Gomez-Gonzalez, B. and Aguilera, A. (2009) 'Chromosomal translocations caused by either pol32-dependent or pol32-independent triparental break-induced replication.', *Molecular and cellular biology*. United States, 29(20), pp. 5441–5454. doi: 10.1128/MCB.00256-09.

Ruotolo, R. *et al.* (2010) 'Chemogenomic profiling of the cellular effects associated with histone H3 acetylation impairment by a quinoline-derived compound.', *Genomics*. United States, 96(5), pp. 272–280. doi: 10.1016/j.ygeno.2010.08.005.

Rusche, L. N., Kirchmaier, A. L. and Rine, J. (2003) 'The establishment,

inheritance, and function of silenced chromatin in Saccharomyces cerevisiae.', *Annual review of biochemistry*. United States, 72, pp. 481–516. doi: 10.1146/annurev.biochem.72.121801.161547.

Rutherford, J. C. and Bird, A. J. (2004) 'Metal-responsive transcription factors that regulate iron, zinc, and copper homeostasis in eukaryotic cells.', *Eukaryotic cell*. United States, 3(1), pp. 1–13.

Sakai, K. and Takahashi, T. (1972) 'Estimation of ploidies in brewery yeasts', *Bull Brewing Sci*.

Sale, J. E. (2013) 'Translesion DNA synthesis and mutagenesis in eukaryotes.', *Cold Spring Harbor perspectives in biology*. United States, 5(3), p. a012708. doi: 10.1101/cshperspect.a012708.

Sanchez, M. R. *et al.* (2017) 'Differential paralog divergence modulates genome evolution across yeast species', *PLOS Genetics*. Public Library of Science, 13(2), p. e1006585.

Sanders, B. D. *et al.* (2018) 'Structural Basis for Nicotinamide Inhibition and Base Exchange in Sir2 Enzymes', *Molecular Cell*. Elsevier, 25(3), pp. 463–472. doi: 10.1016/j.molcel.2006.12.022.

Sandmeier, J. J. *et al.* (2002) 'Telomeric and rDNA silencing in Saccharomyces cerevisiae are dependent on a nuclear NAD(+) salvage pathway.', *Genetics*. United States, 160(3), pp. 877–889.

Sankar, T. S. *et al.* (2016) 'The nature of mutations induced by replication-transcription collisions.', *Nature*. England, 535(7610), pp. 178–181. doi: 10.1038/nature18316.

Santos-Pereira, J. M. *et al.* (2013) 'The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability.', *Genes & development*. United States, 27(22), pp. 2445–2458. doi: 10.1101/gad.229880.113.

Sasaki, M. and Kobayashi, T. (2017) 'Ctf4 Prevents Genome Rearrangements by Suppressing DNA Double-Strand Break Formation and Its End Resection at Arrested Replication Forks.', *Molecular cell*. United States, 66(4), p. 533–545.e5. doi: 10.1016/j.molcel.2017.04.020.

Sauve, A. A. *et al.* (2006) 'The Biochemistry of Sirtuins', *Annual Review of Biochemistry*. Annual Reviews, 75(1), pp. 435–465. doi: 10.1146/annurev.biochem.74.082803.133500.

Sauve, A. A. *et al.* (2018) 'Chemical Activation of Sir2-Dependent Silencing by Relief of Nicotinamide Inhibition', *Molecular Cell*. Elsevier, 17(4), pp. 595–601. doi: 10.1016/j.molcel.2004.12.032.

Scannell, D. R. *et al.* (2006) 'Multiple rounds of speciation associated with

reciprocal gene loss in polyploid yeasts.', *Nature*. England, 440(7082), pp. 341–345. doi: 10.1038/nature04562.

Schacherer, J. *et al.* (2005) 'Duplication processes in Saccharomyces cerevisiae haploid strains.', *Nucleic acids research*. England, 33(19), pp. 6319–6326. doi: 10.1093/nar/gki941.

Schacherer, J. *et al.* (2007) 'Spontaneous duplications in diploid Saccharomyces cerevisiae cells.', *DNA repair*. Netherlands, 6(10), pp. 1441–1452. doi: 10.1016/j.dnarep.2007.04.006.

Schmidt, A., Kunz, J. and Hall, M. N. (1996) 'TOR2 is required for organization of the actin cytoskeleton in yeast.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 93(24), pp. 13780–13785.

Schneider, J. *et al.* (2006) 'Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II.', *The Journal of biological chemistry*. United States, 281(49), pp. 37270–37274. doi: 10.1074/jbc.C600265200.

Schneper, L., Duvel, K. and Broach, J. R. (2004) 'Sense and sensibility: nutritional response and signal integration in yeast.', *Current opinion in microbiology*. England, 7(6), pp. 624–630. doi: 10.1016/j.mib.2004.10.002.

Scholes, D. T. *et al.* (2001) 'Multiple regulators of Ty1 transposition in Saccharomyces cerevisiae have conserved roles in genome maintenance.', *Genetics*. United States, 159(4), pp. 1449–1465.

von Schwedler, U., Jack, H. M. and Wabl, M. (1990) 'Circular DNA is a product of the immunoglobulin class switch rearrangement.', *Nature*. England, 345(6274), pp. 452–456. doi: 10.1038/345452a0.

Schwer, B. *et al.* (2016) 'Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 113(8), pp. 2258–2263. doi: 10.1073/pnas.1525564113.

Sebat, J. *et al.* (2004) 'Large-scale copy number polymorphism in the human genome.', *Science (New York, N.Y.)*. United States, 305(5683), pp. 525–528. doi: 10.1126/science.1098918.

Selmecki, A., Forche, A. and Berman, J. (2006) 'Aneuploidy and isochromosome formation in drug-resistant Candida albicans.', *Science (New York, N.Y.)*. United States, 313(5785), pp. 367–370. doi: 10.1126/science.1128242.

Selmecki, A., Forche, A. and Berman, J. (2010) 'Genomic plasticity of the human fungal pathogen Candida albicans.', *Eukaryotic cell*. United States,

9(7), pp. 991–1008. doi: 10.1128/EC.00060-10.

Selmecki, A. M. *et al.* (2009) 'Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance.', *PLoS genetics*. United States, 5(10), p. e1000705. doi: 10.1371/journal.pgen.1000705.

Selmecki, A. M. *et al.* (2015) 'Polyploidy can drive rapid adaptation in yeast.', *Nature*. England, 519(7543), pp. 349–352. doi: 10.1038/nature14187.

Sethi, N. *et al.* (1991) 'The CDC20 gene product of Saccharomyces cerevisiae, a beta-transducin homolog, is required for a subset of microtubule-dependent cellular processes.', *Molecular and cellular biology*. United States, 11(11), pp. 5592–5602.

Seufert, W., Futcher, B. and Jentsch, S. (1995) 'Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins.', *Nature*. England, 373(6509), pp. 78–81. doi: 10.1038/373078a0.

Shamji, A. F., Kuruvilla, F. G. and Schreiber, S. L. (2000) 'Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins', *Current Biology*, 10(24), pp. 1574–1581. doi: https://doi.org/10.1016/S0960-9822(00)00866-6.

Shibata, Y. *et al.* (2012) 'Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues.', *Science (New York, N.Y.)*. United States, 336(6077), pp. 82–86. doi: 10.1126/science.1213307.

Shor, E., Fox, C. A. and Broach, J. R. (2013) 'The yeast environmental stress response regulates mutagenesis induced by proteotoxic stress.', *PLoS genetics*. United States, 9(8), p. e1003680. doi: 10.1371/journal.pgen.1003680.

Sickmann, A. *et al.* (2003) 'The proteome of Saccharomyces cerevisiae mitochondria.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 100(23), pp. 13207–13212. doi: 10.1073/pnas.2135385100.

Siddiqui, K., On, K. F. and Diffley, J. F. X. (2013) 'Regulating DNA Replication in Eukarya', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 5(9), p. a012930. doi: 10.1101/cshperspect.a012930.

Sima, J. and Gilbert, D. M. (2014) 'Complex correlations: replication timing and mutational landscapes during cancer and genome evolution.', *Current opinion in genetics & development*. England, 25, pp. 93–100. doi: 10.1016/j.gde.2013.11.022.

Sinclair, D. A. and Guarente, L. (1997) 'Extrachromosomal rDNA circles--a cause of aging in yeast.', *Cell*. United States, 91(7), pp. 1033–1042.

Sinclair, D. A., Mills, K. and Guarente, L. (1997) 'Accelerated aging and

nucleolar fragmentation in yeast sgs1 mutants.', *Science (New York, N.Y.)*. United States, 277(5330), pp. 1313–1316.

Skourti-Stathaki, K. and Proudfoot, N. J. (2014) 'A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression', *Genes & Development*. Cold Spring Harbor Laboratory Press, 28(13), pp. 1384–1396. doi: 10.1101/gad.242990.114.

Slack, A. *et al.* (2006) 'On the Mechanism of Gene Amplification Induced under Stress in Escherichia coli', *PLOS Genetics*. Public Library of Science, 2(4), p. e48.

Smith, C. E., Lam, A. F. and Symington, L. S. (2009) 'Aberrant double-strand break repair resulting in half crossovers in mutants defective for Rad51 or the DNA polymerase delta complex.', *Molecular and cellular biology*. United States, 29(6), pp. 1432–1441. doi: 10.1128/MCB.01469-08.

Smith, C. E., Llorente, B. and Symington, L. S. (2007) 'Template switching during break-induced replication.', *Nature*. England, 447(7140), pp. 102–105. doi: 10.1038/nature05723.

Smith, J. S. and Boeke, J. D. (1997) 'An unusual form of transcriptional silencing in yeast ribosomal DNA.', *Genes & development*. United States, 11(2), pp. 241–254.

Solinger, J. A. and Heyer, W. D. (2001) 'Rad54 protein stimulates the postsynaptic phase of Rad51 protein-mediated DNA strand exchange.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 98(15), pp. 8447–8453. doi: 10.1073/pnas.121009898.

Solinger, J. A., Kiianitsa, K. and Heyer, W.-D. (2002) 'Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments.', *Molecular cell*. United States, 10(5), pp. 1175–1188.

Sollier, J. and Cimprich, K. A. (2015) 'R-Loops Breaking Bad', *Trends in cell biology*, 25(9), pp. 514–522. doi: 10.1016/j.tcb.2015.05.003.

Song, B. and Sung, P. (2000) 'Functional interactions among yeast Rad51 recombinase, Rad52 mediator, and replication protein A in DNA strand exchange.', *The Journal of biological chemistry*. United States, 275(21), pp. 15895–15904. doi: 10.1074/jbc.M910244199.

Srivatsan, A. *et al.* (2010) 'Co-orientation of replication and transcription preserves genome integrity.', *PLoS genetics*. United States, 6(1), p. e1000810. doi: 10.1371/journal.pgen.1000810.

Stan, R. *et al.* (1994) 'Interaction between FKBP12-rapamycin and TOR involves a conserved serine residue.', *The Journal of biological chemistry*. United States, 269(51), pp. 32027–32030.

Stankiewicz, P. and Lupski, J. R. (2010) ‘Structural variation in the human genome and its role in disease.’, *Annual review of medicine*, 61, pp. 437–55. doi: 10.1146/annurev-med-100708-204735.

Steele, D. F. and Jinks-Robertson, S. (1992) ‘An examination of adaptive reversion in Saccharomyces cerevisiae.’, *Genetics*. United States, 132(1), pp. 9–21.

Stewart, S. E. and Roeder, G. S. (1989) ‘Transcription by RNA polymerase I stimulates mitotic recombination in Saccharomyces cerevisiae.’, *Molecular and cellular biology*. United States, 9(8), pp. 3464–3472.

Storici, F. *et al.* (2007) ‘RNA-templated DNA repair.’, *Nature*. England, 447(7142), pp. 338–341. doi: 10.1038/nature05720.

Storlazzi, C. T. *et al.* (2010) ‘Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure.’, *Genome research*. United States, 20(9), pp. 1198–1206. doi: 10.1101/gr.106252.110.

Straight, A. F. *et al.* (1999) ‘Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity.’, *Cell*. United States, 97(2), pp. 245–256.

Strauss, B. S. (1992) ‘The origin of point mutations in human tumor cells.’, *Cancer research*. United States, 52(2), pp. 249–253.

Strope, P. K. *et al.* (2015) ‘The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen.’, *Genome research*. United States, 25(5), pp. 762–774. doi: 10.1101/gr.185538.114.

Sugawara, N. and Haber, J. E. (1992) ‘Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation.’, *Molecular and cellular biology*. United States, 12(2), pp. 563–575.

Sugiyama, T. and Kowalczykowski, S. C. (2002) ‘Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation.’, *The Journal of biological chemistry*. United States, 277(35), pp. 31663–31672. doi: 10.1074/jbc.M203494200.

Syeda, A. H., Hawkins, M. and McGlynn, P. (2014) ‘Recombination and Replication’, *Cold Spring Harbor Perspectives in Biology* , 6(11). doi: 10.1101/cshperspect.a016550.

Symeonidou, I.-E., Taraviras, S. and Lygerou, Z. (2012) ‘Control over DNA replication in time and space’, *FEBS Letters*, 586(18), pp. 2803–2812. doi:

https://doi.org/10.1016/j.febslet.2012.07.042.

Symington, L. S., Rothstein, R. and Lisby, M. (2014) 'Mechanisms and Regulation of Mitotic Recombination in Saccharomyces cerevisiae', *Genetics*. Genetics Society of America, 198(3), pp. 795–835. doi: 10.1534/genetics.114.166140.

Szilard, R. K. *et al.* (2010) 'Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX.', *Nature structural & molecular biology*, 17(3), pp. 299–305. doi: 10.1038/nsmb.1754.

Szostak, J. W. *et al.* (1983) 'The double-strand-break repair model for recombination.', *Cell*. United States, 33(1), pp. 25–35.

Szostak, J. W. and Wu, R. (1980) 'Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae.', *Nature*, 284(5755), pp. 426–30.

Takeuchi, Y., Horiuchi, T. and Kobayashi, T. (2003) 'Transcription-dependent recombination and the role of fork collision in yeast rDNA.', *Genes & development*. United States, 17(12), pp. 1497–1506. doi: 10.1101/gad.1085403.

Tamai, K. T. *et al.* (1993) 'Yeast and mammalian metallothioneins functionally substitute for yeast copper-zinc superoxide dismutase.', *Proceedings of the National Academy of Sciences of the United States of America*, 90(17), pp. 8013–7.

Tamai, K. T. *et al.* (1994) 'Heat shock transcription factor activates yeast metallothionein gene expression in response to heat and glucose starvation via distinct signalling pathways.', *Molecular and cellular biology*. United States, 14(12), pp. 8155–8165.

Tanaka, H. *et al.* (2009) 'Ctf4 coordinates the progression of helicase and DNA polymerase alpha.', *Genes to cells : devoted to molecular & cellular mechanisms*. England, 14(7), pp. 807–820. doi: 10.1111/j.1365-2443.2009.01310.x.

Tanida, I. *et al.* (1991) 'Yeast cyclophilin-related gene encodes a nonessential second peptidyl-prolyl cis-trans isomerase associated with the secretory pathway.', *Transplantation proceedings*. United States, 23(6), pp. 2856–2861.

Tanny, J. C. and Moazed, D. (2001) 'Coupling of histone deacetylation to NAD breakdown by the yeast silencing protein Sir2: Evidence for acetyl transfer from substrate to an NAD breakdown product.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 98(2), pp. 415–420. doi: 10.1073/pnas.031563798.

Tassabehji, M. *et al.* (1999) 'Williams syndrome: use of chromosomal microdeletions as a tool to dissect cognitive and physical phenotypes.', *American journal of human genetics*. United States, 64(1), pp. 118–125. doi:

10.1086/302214.

Taylor, J. S. and Raes, J. (2004) 'Duplication and divergence: the evolution of new genes and old ideas.', *Annual review of genetics*. United States, 38, pp. 615–643. doi: 10.1146/annurev.genet.38.072902.092831.

Tehlivets, O., Scheuringer, K. and Kohlwein, S. D. (2007) 'Fatty acid synthesis and elongation in yeast.', *Biochimica et biophysica acta*. Netherlands, 1771(3), pp. 255–270. doi: 10.1016/j.bbalip.2006.07.004.

Teng, S. C., Kim, B. and Gabriel, A. (1996) 'Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.', *Nature*. England, 383(6601), pp. 641–644. doi: 10.1038/383641a0.

Thaminy, S. *et al.* (2007) 'Hst3 Is Regulated by Mec1-dependent Proteolysis and Controls the S Phase Checkpoint and Sister Chromatid Cohesion by Deacetylating Histone H3 at Lysine 56', *Journal of Biological Chemistry* , 282(52), pp. 37805–37814. doi: 10.1074/jbc.M706384200.

Thiele, D. J. (1988) 'ACE1 regulates expression of the Saccharomyces cerevisiae metallothionein gene.', *Molecular and cellular biology*. United States, 8(7), pp. 2745–2752.

Thiele, D. J. and Hamer, D. H. (1986) 'Tandemly duplicated upstream control sequences mediate copper-induced transcription of the Saccharomyces cerevisiae copper-metallothionein gene.', *Molecular and cellular biology*, 6(4), pp. 1158–63.

Thomas, B. J. and Rothstein, R. (1989) 'Elevated recombination rates in transcriptionally active DNA.', *Cell*. United States, 56(4), pp. 619–630.

Thomas, M., White, R. L. and Davis, R. W. (1976) 'Hybridization of RNA to double-stranded DNA: formation of R-loops.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 73(7), pp. 2294–2298.

Tkach, J. M. *et al.* (2012) 'Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress.', *Nature cell biology*. England, 14(9), pp. 966–976. doi: 10.1038/ncb2549.

Torkelson, J. *et al.* (1997) 'Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation.', *The EMBO journal*. England, 16(11), pp. 3303–3311. doi: 10.1093/emboj/16.11.3303.

Truong, L. N. *et al.* (2013) 'Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 110(19),

pp. 7720–7725. doi: 10.1073/pnas.1213431110.

Tsabar, M. *et al.* (2016) 'Asf1 facilitates dephosphorylation of Rad53 after DNA double-strand break repair.', *Genes & development*. United States, 30(10), pp. 1211–1224. doi: 10.1101/gad.280685.116.

Tsang, E. and Carr, A. M. (2008) 'Replication fork arrest, recombination and the maintenance of ribosomal DNA stability.', *DNA repair*. Netherlands, 7(10), pp. 1613–1623. doi: 10.1016/j.dnarep.2008.06.010.

Tsolou, A. and Lydall, D. (2007) 'Mrc1 protects uncapped budding yeast telomeres from exonuclease EXO1.', *DNA repair*. Netherlands, 6(11), pp. 1607–1617. doi: 10.1016/j.dnarep.2007.05.010.

Tsurimoto, T. and Stillman, B. (1991) 'Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins.', *The Journal of biological chemistry*. United States, 266(3), pp. 1950–1960.

Tuduri, S. *et al.* (2009) 'Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription.', *Nature cell biology*. England, 11(11), pp. 1315–1324. doi: 10.1038/ncb1984.

Ugolini, S. and Bruschi, C. V (1996) 'The red/white colony color assay in the yeast Saccharomyces cerevisiae: epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells.', *Current genetics*. United States, 30(6), pp. 485–492.

Vaisica, J. A. *et al.* (2011) 'Mms1 and Mms22 stabilize the replisome during replication stress.', *Molecular biology of the cell*. United States, 22(13), pp. 2396–2408. doi: 10.1091/mbc.E10-10-0848.

Vandenbol, M. and Portetelle, D. (1999) 'Disruption of six ORFs on Saccharomyces cerevisiae chromosome X: the YJL069c gene of unknown function is essential to cell viability.', *Yeast (Chichester, England)*. England, 15(13), pp. 1411–1417. doi: 10.1002/(SICI)1097-0061(19990930)15:13<1411::AID-YEA463>3.0.CO;2-M.

Venkataram, S. *et al.* (2016) 'Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast', *Cell*, 166(6), p. 1585–1596.e22. doi: https://doi.org/10.1016/j.cell.2016.08.002.

Vernon, M., Lobachev, K. and Petes, T. D. (2008) 'High rates of "unselected" aneuploidy and chromosome rearrangements in tel1 mec1 haploid yeast strains.', *Genetics*. United States, 179(1), pp. 237–247. doi: 10.1534/genetics.107.086603.

Vogt, N. *et al.* (2004) 'Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in

gliomas.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(31), pp. 11368–11373. doi: 10.1073/pnas.0402979101.

Voichek, Y., Bar-Ziv, R. and Barkai, N. (2016) 'Expression homeostasis during DNA replication.', *Science (New York, N.Y.)*. United States, 351(6277), pp. 1087–1090. doi: 10.1126/science.aad1162.

Voordeckers, K. *et al.* (2012) 'Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication.', *PLoS biology*. United States, 10(12), p. e1001446. doi: 10.1371/journal.pbio.1001446.

Voordeckers, K. *et al.* (2015) 'Adaptation to High Ethanol Reveals Complex Evolutionary Pathways', *PLOS Genetics*. Public Library of Science, 11(11), p. e1005635.

Wahba, L., Gore, S. K. and Koshland, D. (2013) 'The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability.', *eLife*. England, 2, p. e00505. doi: 10.7554/eLife.00505.

Wang, H. *et al.* (2003) 'Biochemical evidence for Ku-independent backup pathways of NHEJ.', *Nucleic acids research*. England, 31(18), pp. 5377–5388.

Wang, H. and Xu, X. (2017) 'Microhomology-mediated end joining: new players join the team', *Cell & Bioscience*, 7(1), p. 6. doi: 10.1186/s13578-017-0136-8.

Wang, L. *et al.* (2008) 'Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function.', *Current opinion in structural biology*. England, 18(6), pp. 741–747. doi: 10.1016/j.sbi.2008.09.004.

Wehner, E. and Brendel, M. (1993) 'Formaldehyde lacks genotoxicity in formaldehyde-hyperresistant strains of the yeast Saccharomyces cerevisiae.', *Mutation research*. Netherlands, 289(1), pp. 91–96.

Wehner, E. P., Rao, E. and Brendel, M. (1993) 'Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in Saccharomyces cerevisiae, and characterization of its protein product.', *Molecular & general genetics : MGG*. Germany, 237(3), pp. 351–358.

Wei, P.-C. *et al.* (2016) 'Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells.', *Cell*. United States, 164(4), pp. 644–655. doi: 10.1016/j.cell.2015.12.039.

Wei, W. *et al.* (2011) 'Functional consequences of bidirectional promoters.', *Trends in genetics : TIG*. England, 27(7), pp. 267–276. doi: 10.1016/j.tig.2011.04.002.

Wei, Y. and Zheng, X. F. S. (2011) 'Nutritional control of cell growth via TOR signaling in budding yeast.', *Methods in molecular biology (Clifton, N.J.)*. United States, 759, pp. 307–319. doi: 10.1007/978-1-61779-173-4_18.

Weindruch, R. *et al.* (1986) 'The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake.', *The Journal of nutrition*. United States, 116(4), pp. 641–654.

Weischenfeldt, J. *et al.* (2013) 'Phenotypic impact of genomic structural variation: insights from and for human disease.', *Nature reviews. Genetics*. England, 14(2), pp. 125–138. doi: 10.1038/nrg3373.

Weitao, T. *et al.* (2003) 'Dna2 helicase/nuclease causes replicative fork stalling and double-strand breaks in the ribosomal DNA of Saccharomyces cerevisiae.', *The Journal of biological chemistry*. United States, 278(25), pp. 22513–22522. doi: 10.1074/jbc.M301610200.

Welch, J. *et al.* (1989) 'The CUP2 gene product regulates the expression of the CUP1 gene, coding for yeast metallothionein.', *The EMBO journal*. England, 8(1), pp. 255–260.

Wenger, J. W. *et al.* (2011) 'Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency.', *PLoS genetics*. United States, 7(8), p. e1002202. doi: 10.1371/journal.pgen.1002202.

Wiederrecht, G. *et al.* (1991) 'FKB1 encodes a nonessential FK 506-binding protein in Saccharomyces cerevisiae and contains regions suggesting homology to the cyclophilins', *Proceedings of the National Academy of Sciences*, 88(3), p. 1029 LP-1033.

Wilson, T. E. *et al.* (2015) 'Large transcription units unify copy number variants and common fragile sites arising under replication stress', *Genome Research*. Cold Spring Harbor Laboratory Press, 25(2), pp. 189–200. doi: 10.1101/gr.177121.114.

Windle, B. *et al.* (1991) 'A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration.', *Genes & Development* , 5(2), pp. 160–174. doi: 10.1101/gad.5.2.160.

Winge, D. R. *et al.* (1985) 'Yeast metallothionein. Sequence and metal-binding properties.', *The Journal of biological chemistry*. United States, 260(27), pp. 14464–14470.

Wittkopp, P. J. (2006) 'Evolution of cis-regulatory sequence and function in Diptera', *Heredity*. The Genetics Society, 97, p. 139.

Wold, M. S. and Kelly, T. (1988) 'Purification and characterization of replication protein A, a cellular protein required for in vitro replication of simian virus 40 DNA.', *Proceedings of the National Academy of Sciences of the*

United States of America. United States, 85(8), pp. 2523–2527.

Wu, J. *et al.* (2004) 'Global analysis of nutrient control of gene expression in Saccharomyces cerevisiae during growth and starvation.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(9), pp. 3148–3153. doi: 10.1073/pnas.0308321100.

Wyers, F. *et al.* (2005) 'Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.', *Cell*. United States, 121(5), pp. 725–737. doi: 10.1016/j.cell.2005.04.030.

Xu, B. and Clayton, D. A. (1996) 'RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers.', *The EMBO journal*. England, 15(12), pp. 3135–3143.

Xu, F., Zhang, K. and Grunstein, M. (2005) 'Acetylation in Histone H3 Globular Domain Regulates Gene Expression in Yeast', *Cell*, 121(3), pp. 375–385. doi: https://doi.org/10.1016/j.cell.2005.03.011.

Xu, Z. *et al.* (2009) 'Bidirectional promoters generate pervasive transcription in yeast.', *Nature*. England, 457(7232), pp. 1033–1037. doi: 10.1038/nature07728.

Yamaguchi-Iwai, Y. *et al.* (1997) 'Homeostatic Regulation of Copper Uptake in Yeast via Direct Binding of MAC1 Protein to Upstream Regulatory Sequences ofFRE1 and CTR1 ', *Journal of Biological Chemistry* , 272(28), pp. 17711–17718. doi: 10.1074/jbc.272.28.17711.

Yasuhira, S. (2009) 'Redundant roles of Srs2 helicase and replication checkpoint in survival and rDNA maintenance in Schizosaccharomyces pombe.', *Molecular genetics and genomics : MGG*. Germany, 281(5), pp. 497–509. doi: 10.1007/s00438-009-0426-x.

Yeeles, J. T. P. *et al.* (2013) 'Rescuing stalled or damaged replication forks.', *Cold Spring Harbor perspectives in biology*. United States, 5(5), p. a012815. doi: 10.1101/cshperspect.a012815.

Yu, K. *et al.* (2003) 'R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells.', *Nature immunology*. United States, 4(5), pp. 442–451. doi: 10.1038/ni919.

Yuan, J. *et al.* (2009) 'Histone H3-K56 acetylation is important for genomic stability in mammals.', *Cell cycle (Georgetown, Tex.)*. United States, 8(11), pp. 1747–1753. doi: 10.4161/cc.8.11.8620.

Zarrei, M. *et al.* (2015) 'A copy number variation map of the human genome.', *Nature reviews. Genetics*. England, pp. 172–183. doi: 10.1038/nrg3871.

Zhang, F., Carvalho, C. M. B. and Lupski, J. R. (2009) 'Complex human

chromosomal and genomic rearrangements.', *Trends in genetics : TIG*. England, 25(7), pp. 298–307. doi: 10.1016/j.tig.2009.05.005.

Zhang, H. *et al.* (2013) 'Gene Copy-Number Variation in Haploid and Diploid Strains of the Yeast Saccharomyces cerevisiae', *Genetics*. Genetics Society of America, 193(3), pp. 785–801. doi: 10.1534/genetics.112.146522.

Zhao, Y. *et al.* (2014) 'Structures of naturally evolved CUP1 tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination.', *G3 (Bethesda, Md.)*. United States, 4(11), pp. 2259–2269. doi: 10.1534/g3.114.012922.

Zhao, Y. *et al.* (2017) 'Properties of Mitotic and Meiotic Recombination in the Tandemly-Repeated CUP1 Gene Cluster in the Yeast Saccharomyces cerevisiae.', *Genetics*. United States, 206(2), pp. 785–800. doi: 10.1534/genetics.117.201285.

Zheng, D.-Q. *et al.* (2016) 'Global analysis of genomic instability caused by DNA replication stress in Saccharomyces cerevisiae', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(50), pp. E8114–E8121. doi: 10.1073/pnas.1618129113.

Zhou, Z. and Elledge, S. J. (1993) 'DUN1 encodes a protein kinase that controls the DNA damage response in yeast.', *Cell*. United States, 75(6), pp. 1119–1127.

Zimmermann, F. K. *et al.* (1984) 'Testing of chemicals for genetic activity with Saccharomyces cerevisiae: a report of the U.S. Environmental Protection Agency Gene-Tox Program.', *Mutation research*. Netherlands, 133(3), pp. 199–244.

Zonneveld, B. J. and van der Zanden, A. L. (1995) 'The red ade mutants of Kluyveromyces lactis and their classification by complementation with cloned ADE1 or ADE2 genes from Saccharomyces cerevisiae.', *Yeast (Chichester, England)*. England, 11(9), pp. 823–827. doi: 10.1002/yea.320110904.

Zuo, S. *et al.* (1997) 'DNA polymerase delta isolated from Schizosaccharomyces pombe contains five subunits.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 94(21), pp. 11244–11249.

9. Appendix

A) Hull, R. M. *et al.* (2017) 'Environmental change drives accelerated adaptation through stimulated copy number variation', *PLoS Biology*

PLOS BIOLOGY

RESEARCH ARTICLE

Environmental change drives accelerated adaptation through stimulated copy number variation

Ryan M. Hull, Cristina Cruz, Carmen V. Jack, Jonathan Houseley*

Epigenetics Programme, The Babraham Institute, Cambridge, United Kingdom

* jon.houseley@babraham.ac.uk

OPEN ACCESS

Citation: Hull RM, Cruz C, Jack CV, Houseley J (2017) Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS Biol 15(6): e2001333. https://doi.org/10.1371/journal.pbio.2001333

Academic Editor: Mark Siegal, New York University, United States of America

Received: October 12, 2016

Accepted: May 23, 2017

Published: June 27, 2017

Data Availability Statement: All sequencing data files are available from the GEO database (accession number GSE86283). Source data for all graphs are provided in Supporting Data: S1 Data (cumulative frequency gene expression data), S2 Data (gene expression data derived from RNAseq), S3 Data (Southern blot quantification), S4 Data (northern blot quantification), S5 Data (γH2A ChIPseq data for chromosomes containing CUP1 and SFA1 loci), S6 Data (mother enrichment allele copy numbers and adaptation assay), S7 Data (3xCUP1 adaptation assays and competition assay), and S8 Data (raw growth curve data). Full

Abstract

Copy number variation (CNV) is rife in eukaryotic genomes and has been implicated in many human disorders, particularly cancer, in which CNV promotes both tumorigenesis and chemotherapy resistance. CNVs are considered random mutations but often arise through replication defects; transcription can interfere with replication fork progression and stability, leading to increased mutation rates at highly transcribed loci. Here we investigate whether inducible promoters can stimulate CNV to yield reproducible, environment-specific genetic changes. We propose a general mechanism for environmentally-stimulated CNV and validate this mechanism for the emergence of copper resistance in budding yeast. By analysing a large cohort of individual cells, we directly demonstrate that CNV of the copper-resistance gene *CUP1* is stimulated by environmental copper. CNV stimulation accelerates the formation of novel alleles conferring enhanced copper resistance, such that copper exposure actively drives adaptation to copper-rich environments. Furthermore, quantification of CNV in individual cells reveals remarkable allele selectivity in the rate at which specific environments stimulate CNV. We define the key mechanistic elements underlying this selectivity, demonstrating that CNV is regulated by both promoter activity and acetylation of histone H3 lysine 56 (H3K56ac) and that H3K56ac is required for *CUP1* CNV and efficient copper adaptation. Stimulated CNV is not limited to high-copy *CUP1* repeat arrays, as we find that H3K56ac also regulates CNV in 3 copy arrays of *CUP1* or *SFA1* genes. The impact of transcription on DNA damage is well understood, but our research reveals that this apparently problematic association forms a pathway by which mutations can be directed to particular loci in particular environments and furthermore that this mutagenic process can be regulated through histone acetylation. Stimulated CNV therefore represents an unanticipated and remarkably controllable pathway facilitating organismal adaptation to new environments.

Author summary

Evolutionary theory asserts that adaptive mutations, which improve cellular fitness in challenging environments, occur at random and cannot be controlled by the cell. The mutation mechanisms involved are of widespread importance, governing diverse

images of membranes presented in the manuscript are provided in S9 Data.

Funding: MRC www.mrc.ac.uk (grant number). Doctoral Training Partnership funding for PhD students, received by RMH and CVJ. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The Wellcome Trust wellcome.ac.uk (grant number 088335, 110216). Received by JH. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. BBSRC www.bbsrc.ac.uk (grant number). Babraham Institute Epigenetics Strategic Programme Grant. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: in addition to the funding described in the Financial Disclosure, JH declares that part of this work (although not the presented data), specifically the potential of histone acetyltransferase inhibition to prevent adaptation, forms part of a patent application.

Abbreviations: γH2A, S139-phosphorylated histone H2A; ARS, autonomously replicating sequence; AU, arbitrary unit; BIR, break-induced replication; ChIP, chromatin immunoprecipitation; ChIPseq, ChIP sequencing; CNV, copy number variation; CUT, cryptic unstable transcript; ERC, extrachromosomal rDNA circle; FA, formaldehyde; GAL, galactose; GLU, glucose; H3K56ac, acetylated histone H3 lysine 56; HDAC, histone deacetylase; HR, homologous recombination; MEP, mother enrichment program; NAHR, nonallelic homologous recombination; ncRNA, noncoding RNA; ns, not significant; rDNA, ribosomal DNA; RFB, replication fork barrier; RFS, replication fork stalling; RNAseq, RNA sequencing; wt, wild type.

processes from the acquisition of resistance during chemotherapy to the emergence of nonproductive clones during industrial fermentations. Here we ask whether eukaryotic cells are in fact capable of stimulating useful, adaptive mutations at environmentally relevant loci. We show that yeast cells exposed to copper stimulate copy number amplification of the copper resistance gene *CUP1*, leading to the rapid emergence of adapted clones, and that this stimulation depends on the highly regulated acetylation of histone H3 lysine 56. Stimulated copy number variation (CNV) operates at sites of preexisting copy number variation, which are common in eukaryotic genomes, and provides cells with a remarkable and unexpected ability to alter their own genome in response to the environment.

Introduction

Copy number variation (CNV) is widespread in human populations, with 5%–10% of the human reference genome showing CNV between normal individuals [1–3]. CNV of protein-coding genes contributes to multiple disorders, and specific genetic syndromes have been directly attributed to CNV [4–6]. The pathological effects of CNV imply that gene copy number impacts gene expression, and we have recently shown that changing copy number can directly influence RNA processing [7]. However, CNV of protein coding genes is not always detrimental and can enhance cell growth, particularly in challenging environments. Evolution experiments in yeast give rise to novel CNVs that enhance growth under nutrient starvation, bestow drug resistance, and complement genetic defects [8–12]. CNVs in tumour cells also enhance proliferation, albeit at the expense of the host; for example, copy number amplification can drive tumour growth (e.g., of *FGFR2* or *CDK4* [13, 14]) or mediate drug resistance (e.g., of *DHFR*, *KRAS* or *BRAF* [15–17]). These yeast and human CNVs are examples of adaptive events in which the emergence of novel heritable alleles increases the reproductive fitness of the cell in the current environment.

The emergence of a novel allele in a population requires extensive selection such that the phenotypic observation is removed from original mutation event by many generations, and therefore causal mechanisms remain uncertain for most adaptive mutations including CNV [18]. Neo-Darwinian theory invokes natural selection of randomly occurring mutations to explain adaptation; however, random mutations need not be accidental, as the induction of genome-wide mutation under stress has been well characterised in bacteria and also reported in yeast [19–21]. Furthermore, a handful of loci in eukaryotes undergo localised and controlled genetic changes, including the mammalian immunoglobulin loci (widely reviewed, for example see [22–24]), as well as the budding yeast ribosomal DNA (rDNA) for which multiple CNV mechanisms have been described [25–27]. These loci are, however, highly specialised and their genetic changes are performed by locus-specific machinery; equivalent mechanisms acting genome-wide to induce beneficial genetic changes have not been convincingly demonstrated and present substantial theoretical difficulties [28–30].

The budding yeast rDNA has been used extensively as a model system for CNV. The rDNA consists of ~150 tandem copies of a 9.1-kb sequence encoding the ribosomal RNAs and undergoes frequent CNV [31]. rDNA recombination is initiated almost exclusively from a replication fork barrier (RFB) present in each rDNA copy (Fig 1a). A single protein, Fob1, defines the replication fork stalling site at the RFB [32, 33], and cleavage of these stalled forks is thought to initiate break-induced replication (BIR), a homologous recombination (HR) process that mediates replication fork restart using a homologous sequence on the sister chromatid [34]. Because homologous sequences are present in each rDNA copy, nonallelic homologous

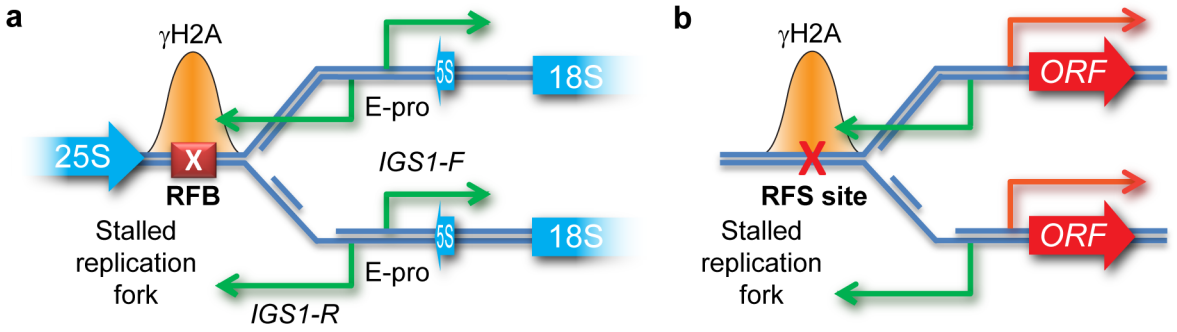


Fig 1. Systems for stimulated copy number variation (CNV) at the ribosomal DNA (rDNA) and at a model gene. a: Minimal elements implicated in control of rDNA recombination: transcription from bidirectional promoter E-pro and replication fork stalling at the Fob1-induced replication fork barrier (RFB). Green arrows represent noncoding RNAs *IGS1-R* and *IGS1-F* transcribed from the E-pro promoter; blue arrows show the rRNA genes (not to scale). **b**: Schematic representation of a general system in which a bidirectional promoter is adjacent to a replication fork stalling (RFS) site. Activation of the bidirectional promoter leads to transcription of the ORF (red arrow) and a noncoding RNA (green arrow). This system should, by analogy to the rDNA, be subject to stimulated CNV when the promoter for the indicated ORF is induced. Stalling of replication forks leads to an accumulation of S139-phosphorylated histone H2A (γH2A) (indicated by orange peaks) that can be detected by chromatin immunoprecipitation (ChIP).

https://doi.org/10.1371/journal.pbio.2001333.g001

recombination (NAHR) occurs readily during BIR, causing frequent CNV. rDNA amplification is partly controlled through transcription; recombination requires RNA Pol I transcription [35], and NAHR is enhanced by expression of 2 noncoding RNAs (ncRNAs), *IGS1-F* and *IGS1-R* [27] (Fig 1a). *IGS1-F* is a stable ncRNA, whereas *IGS1-R* is a cryptic unstable transcript (CUT), a class of noncoding RNA that is degraded instantly after transcription and is transcribed through the RFB [36, 37]. Therefore, local transcription in the context of stalled replication forks at the RFB has a profound effect on rDNA CNV and is thought to cause the environmentally regulated rDNA amplification observed in cells with insufficient rDNA copy number [34, 38, 39].

Although the rDNA recombination machinery is locus specific, replication fork stalling is not unique to the RFB, and CNV often arises from replication defects [40–43]. Collisions between replication and transcription are known to be particularly mutagenic [44], and highly transcribed loci are prone to mutation in general [44–47]. Furthermore, transcription in bacteria has been directly observed to cause replisome dissociation, and mutation rates are higher for bacterial genes oriented against the direction of replication [48, 49]. This lead us to hypothesise that environmental nutrients and toxins that invoke strong transcriptional responses may promote mutations such as CNV at induced loci, effectively focusing mutations at genes that are important for growth in the presence of those nutrients or toxins, thereby accelerating the formation of novel alleles that confer increased fitness in the current environment. Here we demonstrate that CNV of high- and low-copy repeated sequences can be directly stimulated at induced loci in response to the environment, giving rise to novel, advantageous alleles at a rate far in excess of the basal mutation rate.

Results

Promoter induction can stimulate CNV

Replication fork stalling (RFS) occurs widely in the yeast genome and is generally mutagenic [42, 50–52]. By analogy to the rDNA system, we hypothesised that CNV may be instigated from RFS sites upstream of inducible promoters when those promoters are induced (Fig 1a and 1b).

This would cause reproducible, environment-specific patterns of gene loss and gene amplification.

RFS sites are marked by S139-phosphorylated histone H2A (γH2A) [42]. We used chromatin immunoprecipitation (ChIP) sequencing (ChIPseq) for γH2A to generate a high-resolution profile of RFS in the yeast genome, producing a map that is broadly in accord with published ChIP-microarray data [42]. This experiment showed that peaks of >2-fold γH2A enrichment occur within 1 kb upstream of ~7% of *Saccharomyces cerevisiae* genes, which we classed as γH2A genes (S1 Table). We then performed a meta-analysis of published RNA sequencing (RNAseq) data, comparing steady-state mRNA levels of γH2A and non-γH2A genes; this revealed that γH2A genes are expressed at unusually low levels in yeast grown under optimal culture conditions (30˚C in YPD) (Fig 2a and S1a Fig, compare grey and blue lines). However, these γH2A genes are significantly induced under more challenging conditions, such as respiratory growth and industrial fermentation (Fig 2a and S1a Fig, red lines). γH2A genes are therefore biased towards those that are expressed primarily during growth in suboptimal conditions. However, these genes are not rapidly induced by osmotic or oxidative stress and are therefore not simply stress-response factors (S1b Fig). If the induction of RFS genes can instigate CNV, these CNV events should be more frequent at genes induced in response to specific environmental conditions.

To experimentally validate this prediction, we focused on 1 γH2A gene, *CUP1*, a well-studied gene encoding a metallothionein that sequesters excess copper [53, 54]. *CUP1* occurs in a tandem array of 2-kb repeats and has widely varying copy numbers amongst different yeast strains, with higher copy numbers conferring enhanced resistance to copper toxicity [55, 56]. The haploid strain BY4741 used here has a *CUP1* copy number of 13 in our assays, in keeping with previously reported estimates of *CUP1* copy number in the parental S288C background (10–15 copies); this copy number is high but by no means exceptional compared to wild isolates [57, 58]. As expected, most strains that we have tested from the BY4741-derived **a** mating–type deletion collection [59] also have 13 *CUP1* copies, while the S288C-derived MEP diploid has 2 *CUP1* alleles of 13 and 14 copies (see "Stimulated CNV accelerates the acquisition of copper resistance").

CUP1 is strongly induced by environmental copper, and we performed a northern blot analysis to determine whether the *CUP1* promoter is bidirectional like the rDNA ncRNA promoter, as promoter bidirectionality is important for rDNA CNV [27]. Bidirectional promoters are common in the yeast genome, but often the antisense RNA produced is an unstable CUT that is hard to detect in wild-type cells [36, 60]. We therefore analysed RNA from a wild type and from an *rrp6*Δ mutant that lacks a key exonuclease activity required for CUT degradation [36], revealing that the *CUP1* promoter is bidirectional, transcribing a CUT through the RFS site in response to copper exposure (Fig 2b and 2c). γH2A peaks upstream of the *CUP1* ORFs are readily seen in our γH2A ChIPseq data, and the *CUP1* RFS site is unaffected by growth in copper, in contrast to a previous report that ongoing transcription prevents RFS [42] (Fig 2d). This combination of an inducible bidirectional promoter adjacent to an RFS site fits our model derived from the rDNA locus (Fig 1b), making *CUP1* an excellent candidate for stimulated CNV, particularly as $CuSO_4$ induces only a handful of other γH2A genes (S2 Fig).

Copper exposure leads to the emergence of cells carrying amplified *CUP1* alleles [61], but proving that the environment actually stimulates *CUP1* CNV requires the measurement of CNV in the absence of selective pressure. To achieve this, we reengineered the *CUP1* repeat sequence, replacing the copper-responsive *CUP1* promoter and ORF with a *GAL1* promoter and a *3xHA* tag ORF while leaving surrounding sequences, including the RFS region, intact (Fig 2b). This modified construct expresses a nonfunctional protein in response to environmental galactose but not glucose, whereas the endogenous promoter is not induced by

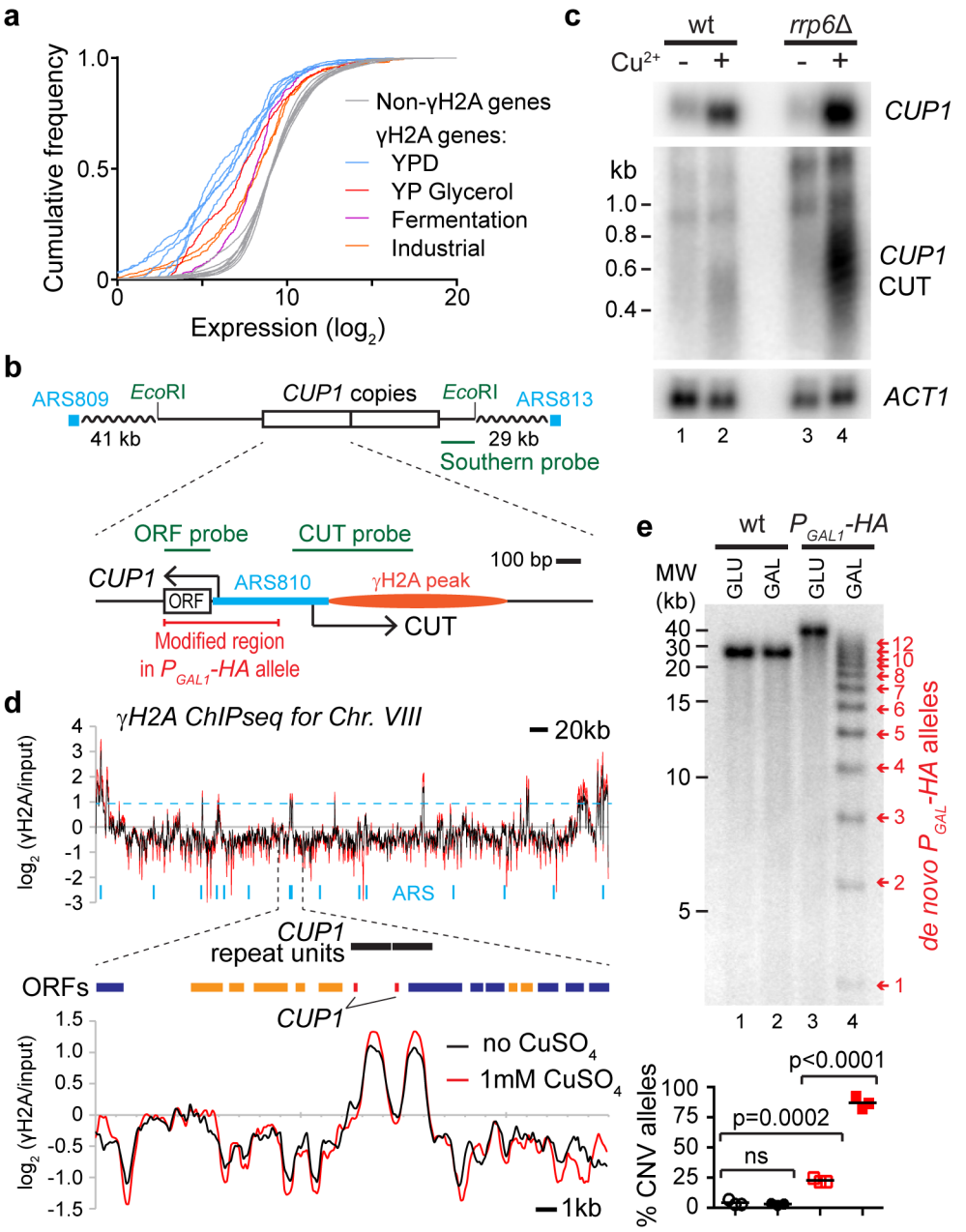


Fig 2. Candidate genes for stimulated copy number variation (CNV). a: Cumulative frequency distribution of gene expression for *S. cerevisiae* growing in various environments. Non-γH2A genes from all data sets are shown in grey, and γH2A genes are shown in blue for cells grown in YPD and in orange, red, and purple for other conditions ($p = 0.00011$, comparing γH2A genes in YPD to other conditions by nested ANOVA). **b**: Schematic of *CUP1* repeats and surrounding region of Chr. VIII, showing 2 copies of *CUP1* as annotated in the reference genome sequence (though the BY4741 wild-type [wt] used here actually has 13 copies). Close-up of a single *CUP1* copy is also shown. Probes used for northern and Southern blots are indicated in green, along with *Eco*RI sites used for Southern analysis. The nearest flanking replication origins (autonomously replicating sequences or ARS elements) are drawn in blue; each *CUP1* repeat also contains a putative ARS overlapping the *CUP1* promoter. The site of the γH2A peak in **d** is represented in orange. Arrows indicate transcription of *CUP1* mRNA and cryptic unstable transcript (CUT) from the *CUP1* promoter; the *CUP1* ORF is shown in white, and the region replaced by P_{GAL1}-*HA* in the galactose-inducible construct is highlighted in red. **c**: Northern analysis of *CUP1* mRNA and *CUP1* upstream CUT in wild-type and *rrp6*Δ cells grown in YPD and exposed to 1 mM $CuSO_4$ for 4 hours; *ACT1* is a loading control. **d**: ChIPseq data for γH2A in wild-type cells grown with or without 1 mM $CuSO_4$, showing Chr. VIII and a close-up of the region surrounding the *CUP1* genes. The dotted blue line shows the cut off for peak calling, while blue vertical marks

represent the annotated replication origins across the chromosome. **e**: Cells with *CUP1* ORF and promoter in each *CUP1* copy replaced by P_{GAL1}-*HA*, grown in glucose or galactose for 10 generations compared to wild-type cells. DNA analysis by Southern blot; arrows indicate de novo alleles formed by CNV events, with numbers indicating P_{GAL}-*HA* copy number. Copy numbers of parental alleles are 13 and 17 copies in the wild-type and the P_{GAL}-*HA* strains, respectively. Quantification shows the percentage of alleles deviating from the parental copy number, $n = 3$; *p*-values calculated by 1-way ANOVA. ns, not significant. Raw quantitation data is available in S1, S3 and S5 Data.

https://doi.org/10.1371/journal.pbio.2001333.g002

galactose. We inserted a construct containing 3 copies of this modified P_{GAL1}-*HA* repeat in place of the *CUP1* locus, which fortuitously amplified to 17 copies upon transformation. P_{GAL1}-*HA* cells were then grown for 10 generations in glucose or galactose and compared to wild-type controls grown under the same conditions. Growth of the P_{GAL1}-*HA* strain in galactose gave rise to multiple de novo CNV alleles, detected by Southern blot, whereas no change was observed in the wild-type controls (Fig 2e). This demonstrates that promoter induction in the genetic context of *CUP1* is sufficient to stimulate CNV.

Stimulated CNV accelerates the acquisition of copper resistance

Equivalent experiments, however, on the wild-type *CUP1* locus would not be informative because growth in the presence of copper selects for rare amplified alleles whether they arise through random or stimulated CNV. To determine whether copper stimulates *CUP1* CNV requires the analysis of many individual cells that are allowed to replicate with or without copper while excluding cells born during the exposure period. Quantification of de novo CNV events within this defined cohort would provide a measure of CNV rate independent of selection.

To achieve this, we employed the mother enrichment program (MEP)—a system that selectively renders new-born cells inviable in the presence of β-estradiol [62]. A cohort of MEP cells in the presence of β-estradiol can be treated for a given period with or without copper, after which time only cells in the initial cohort are viable (Fig 3a). Cells from copper-treated and control cohorts are then plated in the absence of copper and estradiol, giving rise to colonies derived from single cells that have (or have not) been previously exposed to copper; these colonies inherit the *CUP1* copy number of the progenitor cell. If copper stimulates heritable CNV at *CUP1*, then a greater number of colonies with *CUP1* alleles that deviate from the parental copy number should be detected in the copper-treated cohort.

We divided 2 populations of β-estradiol–treated MEP diploid cells and grew them for 24 hours in the presence or absence of 1 mM $CuSO_4$, then assayed 184 of the resulting colonies for *CUP1* copy number (Fig 3b). We observed 31 CNV events (including 6 amplifications) in 56 copper-treated diploid cells (112 *CUP1* alleles, 27% CNV events, 5% amplifications), compared to 7 CNV events (including 1 amplification) in 128 untreated cells (256 alleles, 3% CNV events, 0% amplifications). The difference in the number of CNV events and amplifications between copper-treated and untreated cells is significant ($p = 1.1 \times 10^{-7}$ for CNV events and $p = 0.028$ for amplifications) and represents a 9-fold stimulation of CNV by copper. Furthermore, based on bud scar counting, the untreated cells undergo more divisions than the copper-treated cells in 24 hours (12 ± 2 versus 8 ± 3 divisions), meaning that 9-fold is an underestimate of the true extent of CNV stimulation. This finding directly demonstrates that environmental copper stimulates CNV at *CUP1*.

Changes in *CUP1* copy number alter copper resistance, and we therefore measured the ability of cells arising in this experiment bearing an amplified (+3) or a contracted (–7) allele to grow at different copper concentrations. As expected, copper resistance was significantly increased in the amplified clone and decreased in the contracted clone (Fig 3c). This

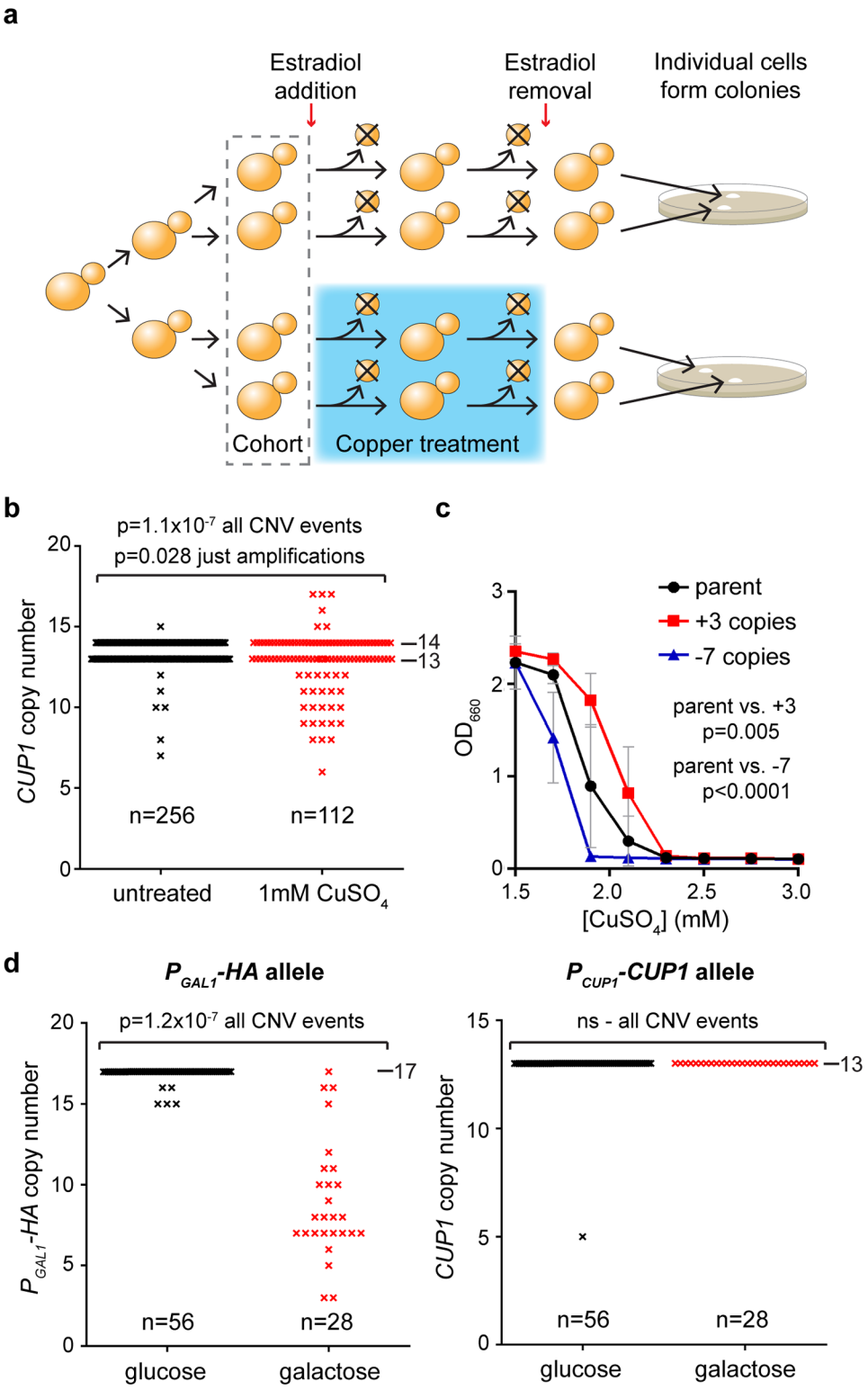


Fig 3. Stimulated copy number variation (CNV) in copper-treated cells. a: Strategy for quantifying stimulated CNV. Schematic of experimental system for measuring CNV in a defined cohort of mother enrichment program (MEP) cells. **b**: Copy number of *CUP1* alleles in colonies derived from 184 diploid MEP cells (pooled from 2 experiments), treated with or without 1 mM $CuSO_4$ for 24 hours (128 cells for–Cu, 56 cells for +Cu). 89% of starting cohort were recovered in the untreated cohort, and 40% were recovered in the

treated cohort. Observed mutation rates in the untreated cohort were normalised for the viability in the treated cohort, making the conservative assumption that cells lost during the experiment did not undergo CNV. *p*-values were calculated by a goodness of fit χ^2 test with 1 degree of freedom between the observed and expected number of mutations to wild-type alleles across the cohorts. **c**: Copper resistance of 3 colonies recovered in **b** with parental, +3, and –7 *CUP1* copy numbers on 1 allele; $CuSO_4$ was added at indicated concentrations to media containing 0.5 mM ascorbic acid to increase copper toxicity, and OD_{660} was measured after 3 days at 30˚C. Error bars represent ±1 SD; *p*-values were calculated by 1-way ANOVA of area under curves; $n = 6$ for each group. **d**: Experiment as in **b** using heterozygous diploid cells with 1 wild-type *CUP1* allele and 1 P_{GAL1}-*HA* allele; data are shown for both alleles in the same cells. Allele-specific probes covering the *CUP1* promoter and ORF or the *GAL1* promoter and *HA* ORF were used for this experiment. Copy numbers of parental alleles are indicated on each panel. ns, not significant. Raw quantitation data are available in S6 Data.

https://doi.org/10.1371/journal.pbio.2001333.g003

demonstrates that stimulated CNV gives rise to de novo alleles with quantifiable phenotypic differences, including increased copper resistance.

To ensure that stimulated CNV is a specific result of promoter induction as opposed to a mutagenic effect of copper treatment, we created diploid MEP cells heterozygous for the wild-type *CUP1* allele and the engineered galactose-responsive P_{GAL1}-*HA* allele. As predicted, galactose treatment induced extensive CNV at the P_{GAL1}-*HA* allele (96% of P_{GAL1}-*3HA* alleles underwent CNV in galactose compared to 9% in glucose) (Fig 3d, left). In contrast, the copper-responsive wild-type allele in the same cells was unaffected (0% of alleles underwent CNV in galactose and only 2% in glucose) (Fig 3d, right). This confirms that CNV is not stimulated uniformly and is highly selective for a transcriptionally induced allele over a silent locus of similar sequence and copy number.

H3K56 acetylation is required for stimulated CNV

Sir2 family histone deacetylases (HDACs) repress rDNA CNV at multiple levels, leading us to question whether HDACs also control *CUP1* CNV [26, 63]. Indeed, we observed extensive *CUP1* CNV after treating wild-type cells with the Sir2 family inhibitor nicotinamide (Fig 4a). Analysis of individual deletion mutants revealed that Sir2 itself has little impact on CNV at *CUP1*, but loss of the degenerate H3K56 HDACs Hst3 and Hst4 induces extensive CNV, suggesting a critical role for H3K56ac in regulating *CUP1* copy number (S3a Fig). Consistent with this, loss of the H3K56 acetyltransferase Rtt109 rendered the *CUP1* locus immune to nicotinamide (Fig 4a).

To determine the importance of Rtt109 for stimulated CNV, we repeated the MEP-based assay from Fig 3b in an *rtt109*Δ background. Remarkably, we found that the transcriptional stimulation of CNV in response to copper was completed abrogated by loss of Rtt109 (compare Fig 4b [showing *rtt109*Δ cells] to Fig 3b [showing wild-type cells]). This shows that stimulated CNV acts by a defined mechanism involving H3K56ac.

H3K56ac has been implicated in both *CUP1* promoter induction [64] and replication fork stability or restart [65–67]. Nicotinamide may therefore affect the stimulated CNV mechanism in 2 ways: by inducing the *CUP1* promoter or by destabilising the stalled replication fork. We observed that nicotinamide stimulates expression of the *CUP1* antisense CUT but causes little or no change in the level of the *CUP1* sense mRNA (Fig 4c). This suggests that HDAC inhibition by nicotinamide reduces promoter directionality at *CUP1* rather than inducing the promoter per se, as has been recently reported for other (as yet unidentified) HDACs [68]. This loss of directionality cannot be ascribed to H3K56 acetylation as it is also observed in *rtt109*Δ cells, and it must depend on another member of the nicotinamide-sensitive Sir2 family. Importantly, however, loss of promoter directionality cannot be solely responsible for *CUP1* CNV, as an equivalent increase in *CUP1* CUT transcript is observed in *rtt109*Δ cells in which CNV does

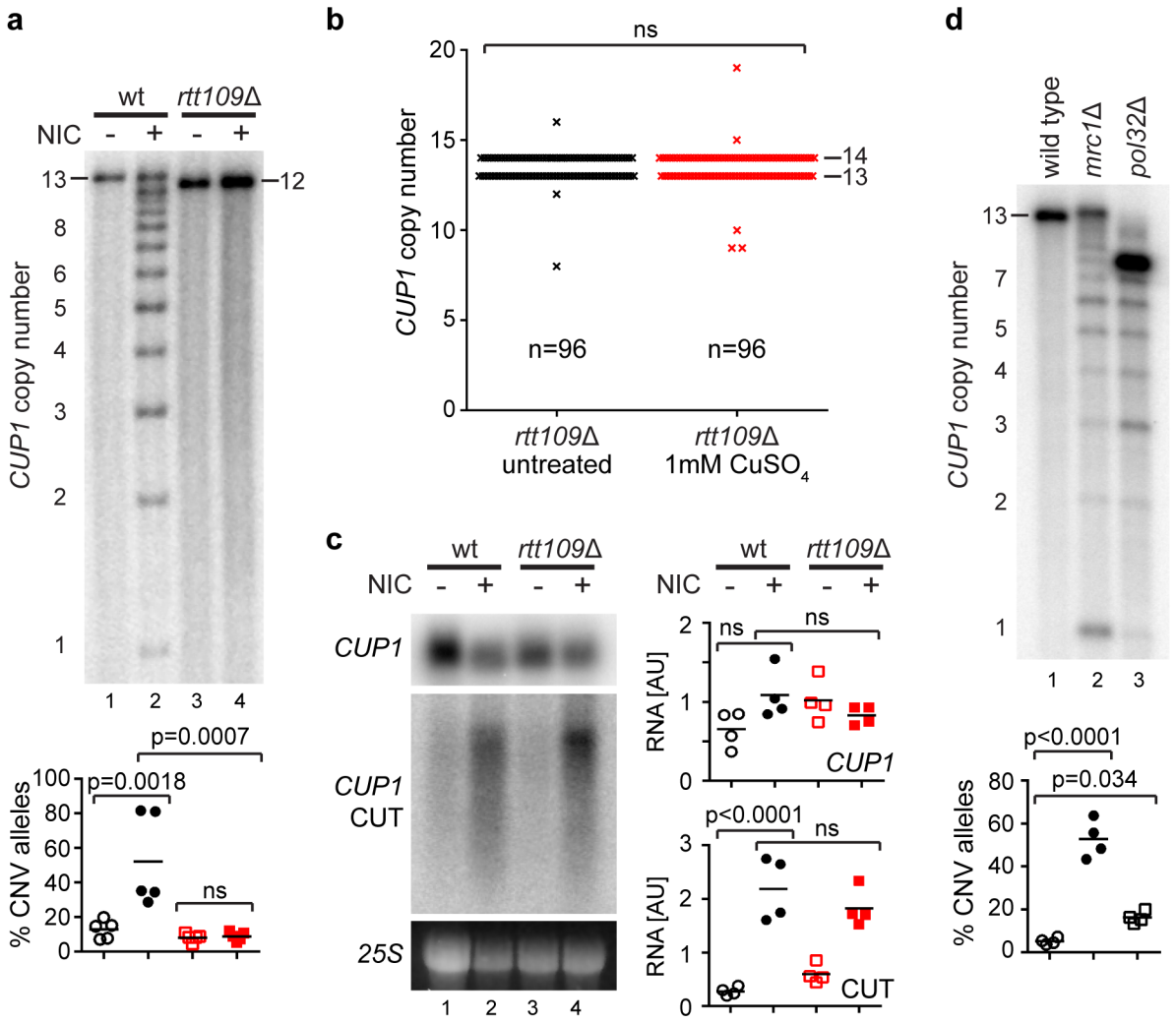


Fig 4. H3K56 acetylation has a critical role in stimulated copy number variation (CNV). a: Southern analysis of *CUP1* copy number in wild-type (wt) and *rtt109*Δ cells with 13 *CUP1* copies grown for 10 generations with or without 5 mM nicotinamide (NIC). Quantification shows the percentage of alleles deviating from the parental copy number after 10 generations; $n = 5$, *p*-values calculated by 1-way ANOVA. **b**: Measurement of CNV in a defined cohort of MEP *rtt109*Δ cells, performed exactly as Fig 3b; 48 diploid cells per condition. ns, not significant. **c**: Northern analysis of *CUP1* ORF and *CUP1* cryptic unstable transcript (CUT) RNA in log-phase wild-type and *rtt109*Δ cells with or without 5 mM nicotinamide. Quantification shows relative RNA levels in arbitrary units (AUs); $n = 4$, *p*-values calculated by 1-way ANOVA. **d**: Southern analysis of *mrc1*Δ and *pol32*Δ cells as in **a**, $n = 4$. Copy numbers of parental alleles are indicated on each panel. Note that in **d**, the mutants are derived from the wild type and therefore have a parental copy number of ~13, even though this allele is no longer detectable in the *pol32*Δ mutant. Raw quantitation data are available in S3, S4 and S6 Data.

https://doi.org/10.1371/journal.pbio.2001333.g004

not occur (compare Fig 4a to 4c). Therefore, nicotinamide treatment makes the *CUP1* promoter transcribe bidirectionally, but this effect is not Rtt109-dependent and is not the sole driver of CNV stimulation.

To assess the potential impact of H3K56 acetylation–associated replication fork defects on CNV, we asked whether mutations that destabilise or impair the processing of stalled replication forks phenocopy nicotinamide without affecting *CUP1* promoter induction or directionality. Amongst 11 deletion mutants of replication fork–associated proteins that impact rDNA stability, we observed that *mrc1*Δ and *pol32*Δ cells undergo striking *CUP1* CNV without

affecting the *CUP1* promoter (Fig 4d and S3b Fig). Mrc1 stabilises stalled replication forks [69], while Pol32 is required for efficient DNA synthesis following the BIR events that are initiated from broken replication forks [70, 71]. The high level of *CUP1* CNV observed in both mutants is consistent with abnormally frequent or inefficient BIR being a key driver of CNV. Importantly, increased H3K56ac was recently shown to impair DNA synthesis during BIR, causing frequent replication fork stalling and recombination events [72]. Such additional recombination events occurring in a repetitive region should cause extensive CNV, providing a simple explanation for the induction of *CUP1* CNV by nicotinamide, which increases H3K56ac globally through inhibition of Hst3 and Hst4.

Regulation of stimulated CNV by promoter activity and H3K56ac

To our surprise, however, nicotinamide had little effect on the P_{GAL1}-*HA* allele, showing that a global increase in H3K56 acetylation is not sufficient to drive CNV (Fig 5a). One difference between the wild-type *CUP1* allele and the re-engineered P_{GAL1}-*HA* allele is that the *GAL1* promoter is fully repressed in glucose, and therefore nicotinamide treatment does not cause the expression of an antisense transcript (S4 Fig). Given the dual effect of nicotinamide on *CUP1* promoter directionality and post-BIR replication, we suspected that both activities might be required for efficient CNV induction.

To test this, we grew P_{GAL1}-*HA* cells with or without nicotinamide in a range of galactose concentrations known to induce the bidirectional *GAL1* promoter to various extents [73]. As predicted, the effect of nicotinamide was minimal and not significant without promoter induction (Fig 5b, lanes 9–12), but with higher promoter induction, nicotinamide significantly stimulated CNV (Fig 5b, lanes 1–8). This effect was particularly striking for the formation of de novo alleles with 1–3 copies, which presumably arise through multiple sequential CNV events (S5 Fig). This experiment reveals that promoter activity and H3K56ac make additive contributions to CNV.

However, we observed that CNV becomes largely independent of nicotinamide at high galactose concentrations (Fig 5a, lanes 3–4), which is not consistent with the model proposed above whereby H3K56ac acts during BIR. We therefore tested the importance of H3K56ac in CNV induction from the *GAL1* promoter by deleting *RTT109*. Just as for the wild-type *CUP1* locus, this completely abrogated CNV induction, confirming that H3K56ac is critical for stimulated CNV in the P_{GAL1}-*HA* system (Fig 5c). These data show that stimulated CNV requires transcription and H3K56ac, but for highly induced promoters the normal physiological level of H3K56ac is sufficient to support extensive CNV such that further deregulation of H3K56 HDAC activity has little effect.

Evidence for stimulated CNV in low-copy repeats

Direct detection of stimulated CNV is facilitated by the high copy number of the *CUP1* locus. However, this raises the question of whether CNV stimulation is restricted to high-copy tandem repeat loci, which are rare amongst protein-coding genes. In contrast, copy numbers of 2–5 are very common in the yeast and human genome sequences [1, 14, 74–77], and we asked whether a 3-copy *CUP1* locus would show equivalent behaviour to the high-copy system. Individual CNV events are too rare in this 3-copy system for direct detection but, having defined the effects of modulating H3K56 acetylation on CNV, we reasoned that if stimulated CNV acts at low-copy loci then H3K56 modulation should alter the rate at which *CUP1* amplifications emerge under copper selection in a predictable manner.

To test this, we replaced the endogenous *CUP1* locus with a synthetic construct containing 3 wild-type copies of the *CUP1* repeat sequence while maintaining the reading frame of the

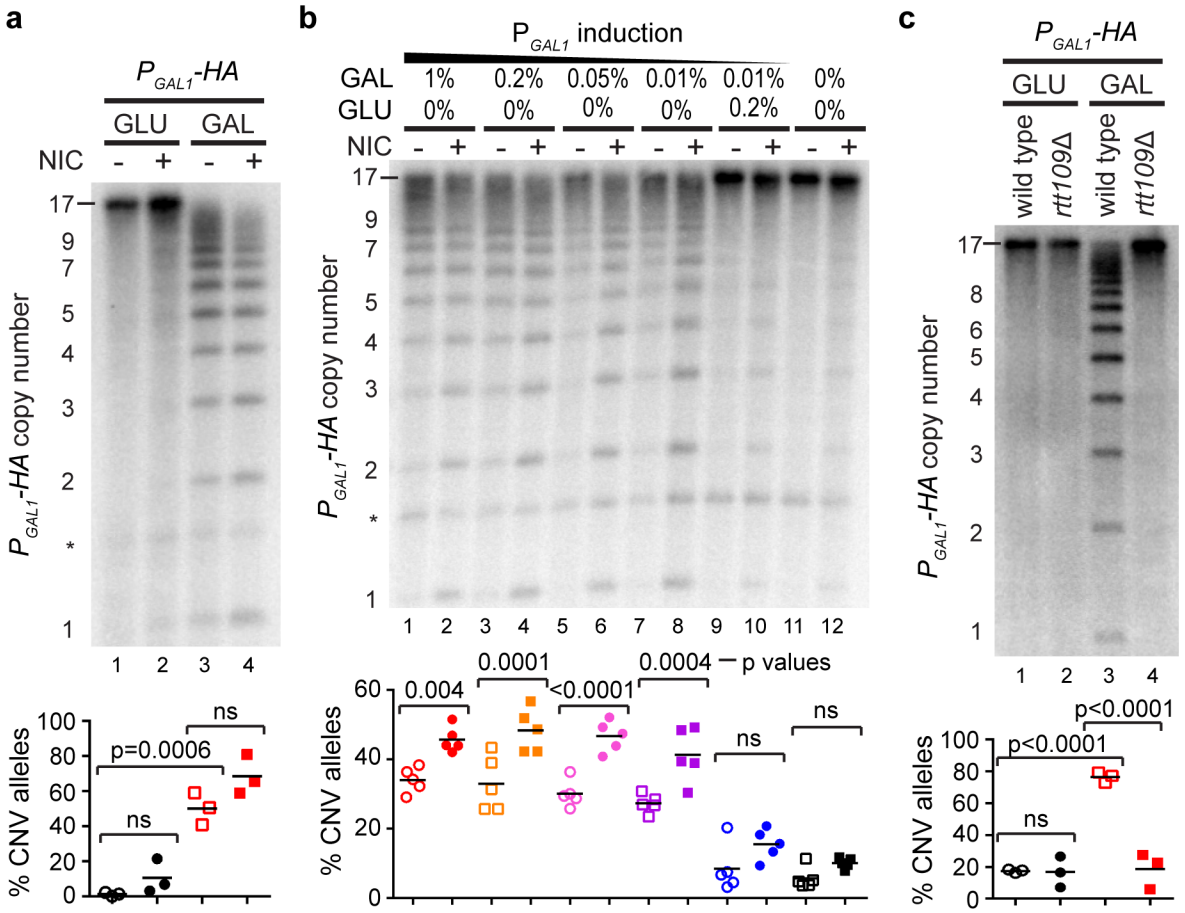


Fig 5. Combinatorial action of promoter activity and histone H3 lysine 56 acetylation (H3K56ac) on copy number variation (CNV). a: Southern analysis of copy number for P_{GAL1}-*HA* cells grown for 10 generations in glucose (GLU) and galactose (GAL) with or without 5 mM nicotinamide (NIC). Quantification shows the percentage of alleles deviating from the parental copy number after 10 generations; $n = 3$, *p*-values calculated by 1-way ANOVA. * nonspecific band. ns, not significant. **b**: Southern analysis as in **a** using given concentrations of galactose and glucose combined with 2% raffinose. $n = 5$. * nonspecific band. *p*-values were calculated from pairwise comparisons of samples with or without NIC for each GLU or GAL concentration deriving from a 1-way ANOVA of the whole data set. See also alternative analysis in S5 Fig. **c**: Southern analysis as in **a** using wild-type or *rtt109*Δ derivatives of P_{GAL1}-*HA* cells. $n = 3$. Copy numbers of parental alleles are indicated on each panel. Raw quantitation data are available in S3 Data.

https://doi.org/10.1371/journal.pbio.2001333.g005

overlapping *RSC30* gene (Fig 6a). Stimulated CNV is replication linked, effectively requiring cells to grow in a sublethal concentration of copper. We observed that 3x*CUP1* cells grow slowly in 0.3 mM $CuSO_4$, although fast-growing resistant populations often emerge late in growth, showing that resistant cells are under positive selection (S6a Fig). Importantly, when 3x*CUP1* cells were grown for 10 generations in batch culture in the presence of 0.3 mM $CuSO_4$, copy-number-amplified alleles were almost always detected in the population by Southern blot (S6b Fig), forming a quantitative assay for CNV.

We first tested whether nicotinamide treatment stimulated CNV in 3x*CUP1* cells as in the high-copy system. Nicotinamide largely stimulates contractions in high-copy *CUP1* arrays, and in 3x*CUP1* cells the only reproducible CNV event observed on nicotinamide treatment was a −1 contraction to 2x*CUP1* (Fig 6b, red arrow). In combination with 0.3mM $CuSO_4$, the 2x*CUP1* band largely disappeared, as would be expected under copper selection (Fig 6c, red

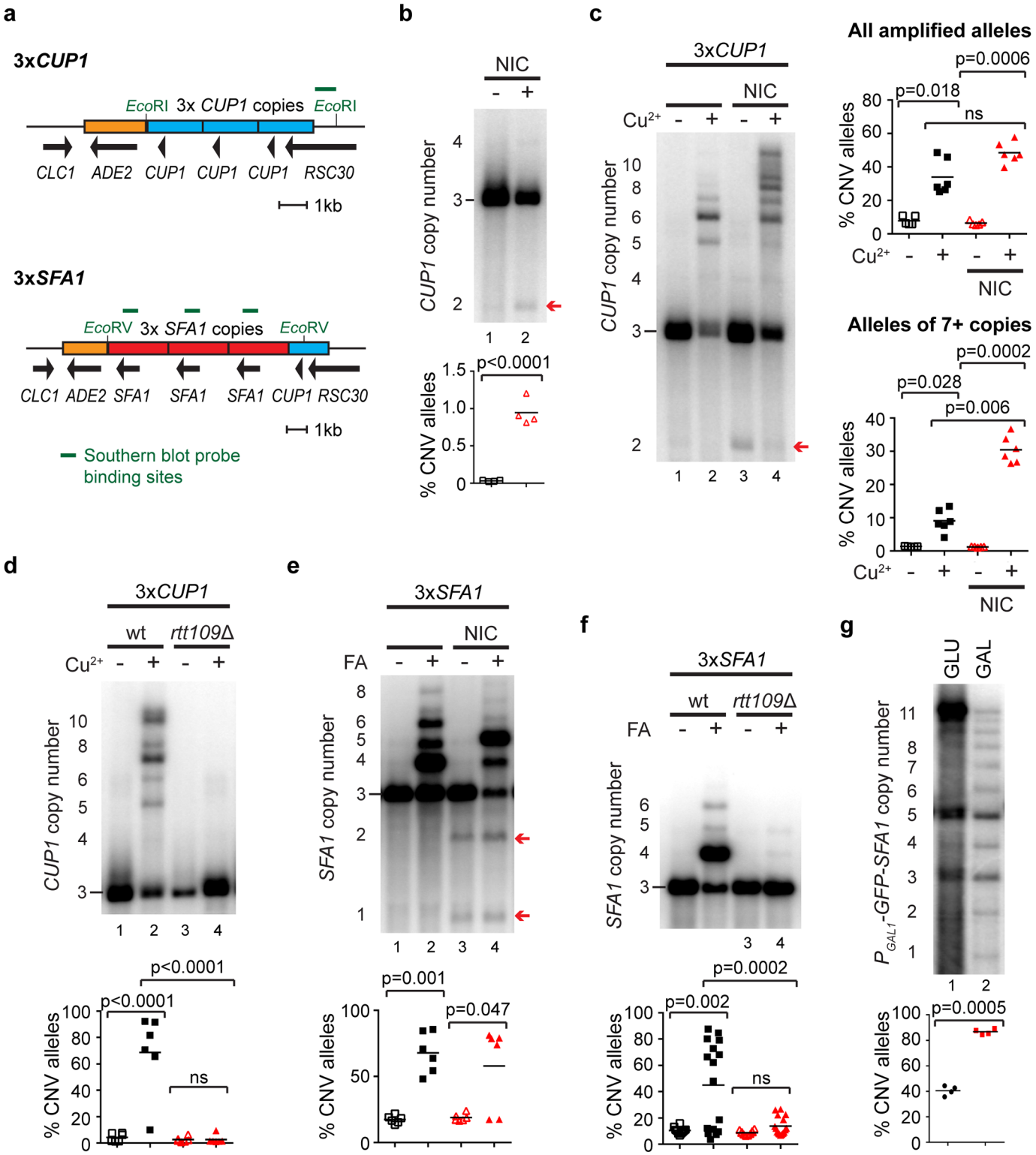


Fig 6. Stimulated copy number variation (CNV) in low-copy repeat systems. a: Schematics of the 3x*CUP1* and 3x*SFA1* constructs inserted at the endogenous *CUP1* locus. Blue boxes indicate *CUP1* repeats, red boxes indicate *SFA1* repeats, and orange boxes indicate the *ADE2* marker. The reading frame of *RSC30* is maintained across the construct boundary. Restriction enzymes for Southern analysis are shown in green along with probe locations. **b**: Southern analysis of *CUP1* copy number in 3x*CUP1* cells grown for 10 generations with or without 5 mM nicotinamide (NIC); arrow indicates –1 copy band. Quantification shows the percentage of –1 alleles; $n = 4$, p-value calculated by t test. **c**: Southern analysis of *CUP1* copy

number in 3x*CUP1* cells grown for 10 generations with or without 0.3 mM $CuSO_4$ and with or without 5 mM NIC. Upper quantification shows the percentage of alleles deviating from the parental copy number; $n = 5$, *p*-values calculated by 1-way ANOVA for repeated measurements. Lower quantification is as upper quantification, considering only alleles of 7+ copies. ns, not significant. **d**: Southern analysis of *CUP1* copy number in 3x*CUP1* wild-type (wt) and *rtt109*Δ cells after 10 generations with or without 0.3 mM $CuSO_4$ (analysis as in **c**); $n = 6$. **e**: Southern analysis of *SFA1* copy number in 3x*SFA1* grown for 17 generations with or without ~1 mM formaldehyde (FA) and with or without 5 mM NIC. Quantification shows the percentage of alleles deviating from the parental copy number; $n = 6$, *p*-values calculated by 1-way ANOVA. A correction was applied to the quantification to account for the differing number of probe-binding sites in the amplified alleles. **f**: Southern analysis of *SFA1* copy number in 3x*SFA1* wild-type and *rtt109*Δ cells grown for 17 generations with or without ~1 mM FA. Quantification as in **e**; $n = 10$ for untreated samples, $n = 16$ for FA-treated samples. **g**: Southern analysis of CNV induced in the high-copy P_{GAL1}-*GFP-SFA1* system. After selection for high copy number and outgrowth in SC media without FA, cells were grown for 10 generations in SC with 2% glucose (GLU) or galactose (GAL). Quantification shows the percentage of contracted alleles; $n = 4$, *p*-value calculated by paired *t* test. Raw quantitation data are available in S3 Data.

https://doi.org/10.1371/journal.pbio.2001333.g006

arrow), and the proportion of amplified alleles increased marginally (Fig 6c, upper quantification panel). However, the most noticeable difference was that the proportion of large alleles (more than 2-fold the progenitor allele size) increased dramatically with nicotinamide treatment (Fig 6c, lower quantification panel). These results show that copper and nicotinamide both stimulate CNV, and although CNV stimulation causes many copy number contractions, copper and nicotinamide have an additive effect in the 3x*CUP1* system that results in the formation of larger alleles that predominate in batch culture.

In contrast, since deletion of *RTT109* suppressed stimulated *CUP1* CNV in the high-copy system, we then asked if the amplifications observed in 3x*CUP1* cells are also Rtt109-dependent. Indeed, when 3x*CUP1 rtt109*Δ cells were grown for 10 generations in 0.3 mM copper, amplification was completely suppressed (Fig 6d). This demonstrates that H3K56 acetylation is required for *CUP1* amplification in the presence of copper, a very surprising result because previous studies have shown a critical role for Rtt109 in maintaining genome stability rather than promoting genome change [26, 66, 78, 79].

We then asked whether another 3-copy RFS gene would show similar behaviour. We selected to test the *SFA1* gene encoding a formaldehyde dehydrogenase, as this has a clear upstream RFS site (S6c Fig), is inducible in response to formaldehyde, and higher *SFA1* copy number increases formaldehyde resistance [80]. A tandem array of 3 *SFA1* genes with surrounding sequence was inserted at the *CUP1* locus along with a single wild-type *CUP1* copy (Fig 6a), while the endogenous *SFA1* gene was deleted. These 3x*SFA1* cells showed a sharp cut-off for growth in formaldehyde, with <0.9 mM allowing robust growth and >1 mM completely suppressing growth, while the formaldehyde concentration that gave slow but reproducible growth (which is required for these assays) varied from 0.9–1.0 mM with formaldehyde batch and had to be empirically determined. We therefore refer to the assay concentration as ~1 mM formaldehyde.

Growth of 3x*SFA1* cells with ~1 mM formaldehyde induced bidirectional transcription from the *SFA1* promoter and again gave rise to amplified *SFA1* alleles detectable by Southern blot over 17 generations (Fig 6f and S6d Fig). As in the 3x*CUP1* system, growth of 3x*SFA1* cells in nicotinamide induced copy number contractions that were readily detected in the absence of formaldehyde (Fig 6e, red arrows), although the additive effect between formaldehyde and nicotinamide was not observed. Importantly however, the copy number amplification of *SFA1* in 3x*SFA1* cells was completely suppressed in an *rtt109*Δ mutant (Fig 6f), indicating that *SFA1* amplification proceeds in the presence of formaldehyde by the same mechanism as *CUP1* amplification.

Confirming that CNV is transcriptionally stimulated at *SFA1* requires a high-copy system, but we were unable to create a direct equivalent of the high-copy P_{GAL1}-*HA* strain as this amplified fortuitously during transformation. The alternative is to select for an amplified allele using formaldehyde; however, this requires the *SFA1* gene and amplification system to be

active, which is not the case if the *SFA1* promoter is simply replaced with P_{GAL1}. We suspected that bidirectional transcription into the *SFA1* RFS would stimulate CNV irrespective of where the promoter is placed, so we created a construct in which the promoter and ORF of the upstream divergent gene *UGX2* were replaced with P_{GAL1}-*GFP* in each of the 3 *SFA1* repeats (S6e Fig). This strain was grown in 0.9 mM formaldehyde, stepped up to 2.2 mM formaldehyde and then recovered in glucose media, yielding a P_{GAL1}-*GFP-SFA1* strain with an unstable and therefore somewhat heterogeneous copy number but primarily containing 11 copies. Growth of this strain in galactose caused the disappearance of this upper band and the emergence of a prominent ladder (Fig 6g), just as in the original P_{GAL}-*HA* strain (Fig 2e), showing that transcription directly stimulates CNV at the *SFA1* RFS site.

Together, these data show that low-copy systems undergo CNV through a mechanism consistent with stimulated CNV, and that this mechanism is not restricted to *CUP1*.

Stimulated CNV enhances adaptation in a population

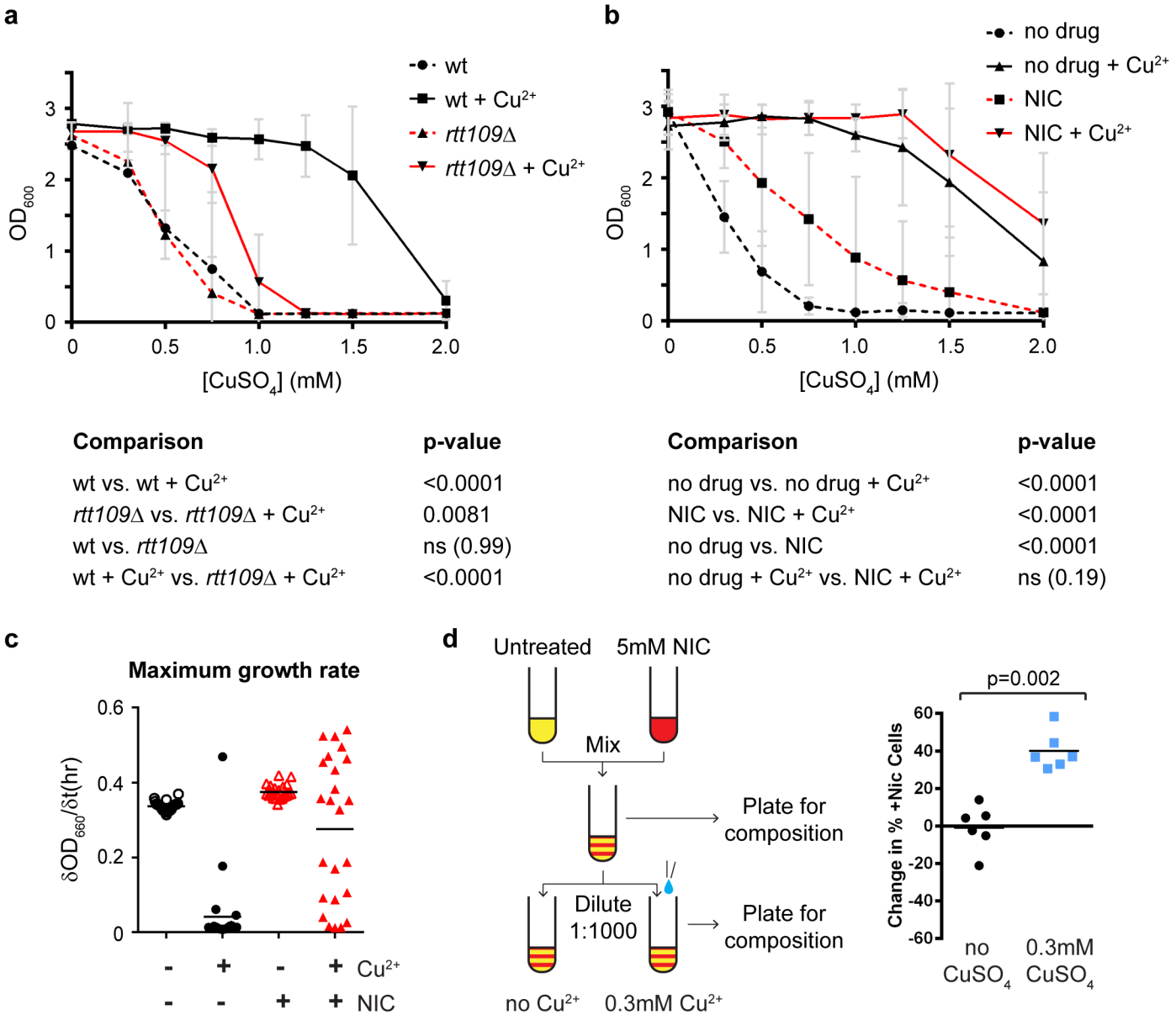
The ability of transcription to stimulate CNV and the reliance of this mechanism on H3K56 acetylation suggest that the emergence of copper adaptation, instead of being an inevitable consequence of random mutation, follows a defined mechanism that is highly sensitive to alterations in this histone mark.

To assess this, we initially tested the copper resistance of the 3x*CUP1* wild-type and *rtt109*Δ cells grown with or without 0.3 mM $CuSO_4$ shown in Fig 6d. Growth of wild-type cells in copper caused a dramatic rise in copper resistance, with GI_{50} (concentration of $CuSO_4$ causing a 50% inhibition of growth) rising >3-fold from 0.5 mM in untreated cells to ~1.7 mM in cells pregrown in 0.3 mM $CuSO_4$ (Fig 7a). This increase was clearly attributable to CNV, as cells that grew in 1 mM $CuSO_4$ carried large *CUP1* amplifications (S7a Fig). In contrast, *rtt109*Δ cells underwent a far smaller increase, rising <2-fold from 0.5 mM to ~0.8 mM after pregrowth in 0.3 mM $CuSO_4$ (Fig 7a), and of the *rtt109*Δ cells that did survive at 1 mM $CuSO_4$, albeit with much reduced growth compared to wild-type cells, only a fraction had undergone *CUP1* amplification, suggesting that most had acquired resistance through other, potentially random, mutations (S7a Fig). This shows that acquisition of copper resistance, far from being inevitable, is strongly dependent on an Rtt109-dependent amplification mechanism.

We similarly assessed the effect of nicotinamide on copper resistance using the cells grown with or without $CuSO_4$ and with or without nicotinamide shown in Fig 6c. The additive effect of nicotinamide and copper provided a small and not significant increase to the already substantial copper resistance of cells pregrown in 0.3 mM $CuSO_4$, but more surprisingly, pregrowth in nicotinamide caused a 2.5-fold increase in GI_{50} for $CuSO_4$, from 0.3 mM to 0.75mM (Fig 7b). This is in contrast to the Southern blotting data, which show that the primary effect of nicotinamide treatment is copy number loss (Fig 6b and 6c) and suggest that a substantial amount of amplifications are also generated. In support of this, the *CUP1* copy number of nicotinamide pretreated cells that grew in 0.75 mM $CuSO_4$ was substantially amplified (S7b Fig), showing that nicotinamide treatment increases copper resistance by promoting *CUP1* amplification.

These striking effects of H3K56 acetylation on copper resistance suggested that CNV stimulation should provide substantial selective advantages at the population level. As nicotinamide treatment induced constitutive stimulated CNV in the absence of copper, we used this drug to directly test the selective benefit provided by stimulated CNV.

Firstly, we examined strong copper selection based on growth in 0.75 mM $CuSO_4$, a concentration fully inhibitory to growth of 3x*CUP1* cells. We pretreated four 3x*CUP1* cultures for 10 generations with or without 5mM nicotinamide, then obtained 6 growth curves for cells



Comparison	p-value	Comparison	p-value
wt vs. wt + Cu^{2+}	<0.0001	no drug vs. no drug + Cu^{2+}	<0.0001
*rtt109*Δ vs. *rtt109*Δ + Cu^{2+}	0.0081	NIC vs. NIC + Cu^{2+}	<0.0001
wt vs. *rtt109*Δ	ns (0.99)	no drug vs. NIC	<0.0001
wt + Cu^{2+} vs. *rtt109*Δ + Cu^{2+}	<0.0001	no drug + Cu^{2+} vs. NIC + Cu^{2+}	ns (0.19)

Fig 7. Copper adaptation through stimulated copy number variation (CNV). a: Copper resistance of 3x*CUP1* wild-type (wt) and *rtt109*Δ cells grown with or without 0.3 mM $CuSO_4$ from Fig 6d. Cells were diluted in media with varying concentrations of $CuSO_4$ and grown for 3 days. Average OD_{660} is plotted, error bars represent ±1 SD, and $n = 6$ cultures per condition, each tested at 8 $CuSO_4$ concentrations. *p*-values were calculated by 1-way ANOVA of area-under-curve values for each culture. **b**: Copper resistance of 3x*CUP1* cells grown with or without 5 mM nicotinamide (NIC) and with or without 0.3 mM $CuSO_4$ from Fig 6c. Analysis as in **a**; $n = 12$. **c**: Maximum growth rate in 0 mM or 0.75 mM $CuSO_4$ of 3x*CUP1* cells pretreated with or without 5 mM nicotinamide for 10 generations. dOD_{660}/dt represents the OD change per hour. Four samples each grown with or without nicotinamide were each inoculated in 6 cultures for growth curve determination across 72 hours. Data are the maximum of the first derivative of smoothed OD_{660} time-course data (see S7c Fig) for each culture. **d**: Competitive growth assay in 0 or 0.3 mM $CuSO_4$. Two populations of 3x*CUP1* cells with different selectable markers were pregrown with or without 5 mM nicotinamide, then mixed and outgrown for 10 generations in direct competition. The graph shows the change in composition of outgrowth cultures across the competition period between inoculation and saturation (10 generations). *p*-value was calculated by paired *t* test, $n = 6$. Raw quantitation data are available in S7 and S8 Data.

https://doi.org/10.1371/journal.pbio.2001333.g007

from each culture with or without 0.75mM $CuSO_4$ in the absence of nicotinamide (S7c Fig). Maximum growth rates were derived from each growth curve to determine whether any cells had adapted sufficiently to allow as rapid growth in the presence of 0.75 mM $CuSO_4$ as in the absence of copper (Fig 7c). Nicotinamide pretreatment had no effect on growth rate in the absence of copper but dramatically increased adaptation to copper: only 1 of the 24 cultures grown in the presence of 0.75 mM Cu without nicotinamide pretreatment grew normally, whereas over half (13 of 24) cultures derived from the nicotinamide pretreated samples reached growth rates equivalent to cells growing in the absence of copper. This was not due to rare events in a few of the precultures because the distribution of growth rates obtained in samples from each pretreated culture was similar (S7d Fig). Therefore, although CNV stimulation primarily causes copy number contraction, it also dramatically enhances the ability of a subpopulation of cells to thrive in otherwise toxic concentrations of copper.

Secondly, we asked whether stimulated CNV provides a competitive advantage in low-copper environments in which nonamplified cells are still capable of growth. Under these conditions, the fact that stimulated CNV primarily causes copy number reduction (and therefore further slows growth in copper-containing environments) may put a population of cells using stimulated CNV at a disadvantage relative to a population that does not. To test this, we again made use of the nicotinamide to mimic the effect of stimulated CNV prior to growth in copper-containing media, and we directly competed 1:1 mixtures of untreated and nicotinamide pretreated populations in the same cultures, with or without 0.3 mM $CuSO_4$. Treated and untreated populations carried different selectable markers to allow the composition of the mixture to be determined by plating before and after growth (Fig 7d, left). Nicotinamide pretreatment did not alter the competitive fitness of cells in the absence of copper, but the nicotinamide-treated populations efficiently outcompeted the untreated populations in the presence of 0.3 mM $CuSO_4$, increasing their population share by 40% on average over 10 generations (Fig 7d right). These experiments clearly show that although stimulated CNV engenders many more contractions than amplifications, it still provides a major selective advantage in both low- and high-copper environments.

Together, our findings demonstrate that bidirectional promoter induction in the *CUP1* genetic context can stimulate CNV to form novel adaptive alleles and that the rate of stimulated CNV is responsive to a controllable histone modification system. Stimulated CNV provides a clear selective advantage, and amplified alleles conferring improved resistance arise in low-copy strains by a mechanism consistent with CNV stimulation.

Discussion

The assertion that adaptation occurs purely through natural selection of random mutations is deeply embedded in our understanding of evolution. However, we have demonstrated that a controllable mechanism exists in yeast for increasing the mutation rate in response to at least 1 environmental stimulus and that this mechanism shows remarkable allele selectivity. This mechanism has the potential to act widely in eukaryotic genomes, even if restricted to repeated sequences, and may therefore underlie a substantial fraction of observed CNV events.

A proposed mechanism for stimulated CNV

We propose a model for stimulated CNV in which local bidirectional promoter activity destabilises stalled replication forks, increasing the frequency of error-prone BIR events. Replication fork stalling occurs widely though nonrandomly in the yeast genome, but stalled forks are normally resolved through error-free mechanisms that protect genome stability (Fig 8a, top left) (reviewed in [81]). However, we suggest that induction of transcription from an adjacent

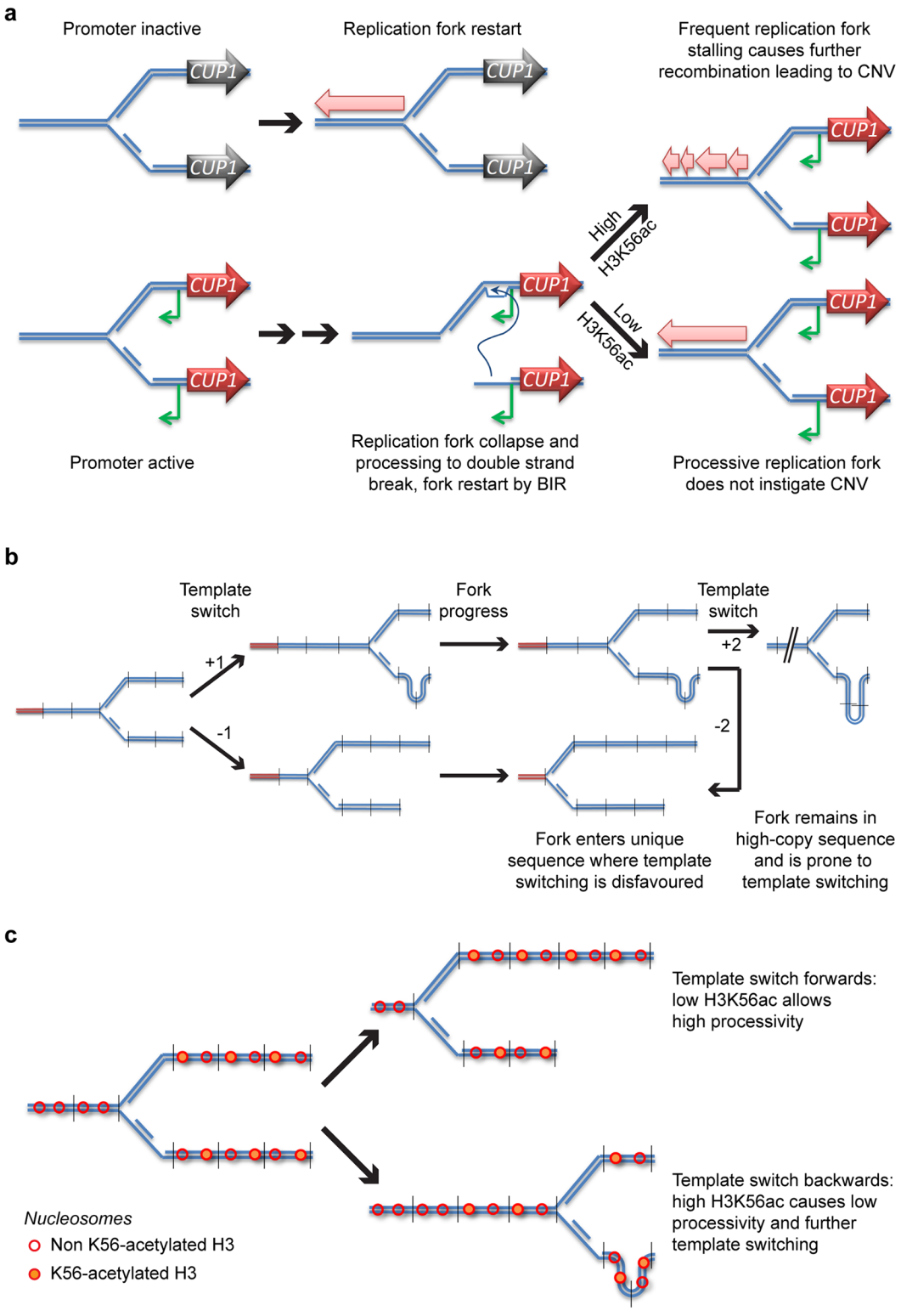


Fig 8. A scheme for stimulated copy number variation (CNV). a: Proposed mechanism by which promoter activity and acetylated histone H3 lysine 56 (H3K56ac) contribute to stimulated CNV. DNA strands are shown in blue, the inactive *CUP1* gene is shown in black, and the induced *CUP1* gene is shown in red, with antisense CUT in green. Pink block arrows represent progression of the replication fork. **b**: Proposed mechanism for CNV asymmetry. DNA strands from repeated DNA are shown in blue, unique sequences are shown in red, and vertical

lines indicate repeat boundaries. Numbers indicate the change in copy number for a particular template switch. The result of 2 successive template-switching events with an intervening period of replication are shown, resulting in a 3:1 ratio of contractions to amplifications. **c**: Additional asymmetry is generated by H3K56ac: nucleosomes around a replication fork (blue) are shown as red circles, either empty or shaded to represent the H3K56ac state. Template switching forward is seen to move the fork to a region of low H3K56ac, whereas template switching backwards moves the fork to a region of high H3K56ac and therefore low fork stability. BIR, break-induced replication.

https://doi.org/10.1371/journal.pbio.2001333.g008

bidirectional promoter increases the likelihood that stalled replication forks collapse (Fig 8a, bottom left); this may occur either through direct interference of the transcription machinery with the stalled fork or indirectly through increased topological stress. Either way, the collapsed fork must then be restarted by a mechanism such as BIR (Fig 8a, bottom middle), forming a replication fork with reduced processivity that is prone to CNV, particularly when H3K56ac is high (Fig 8a, right). This mechanism is closely related to the fork-stalling and template-switching (FoSTeS) process that is suggested to underlie a wide range of CNV events [82], but with contributions from transcription and H3K56ac. Our data implicating H3K56ac and Pol32 in *CUP1* CNV is consistent with restart by BIR, since the replication forks newly formed through BIR are error-prone in the absence of Pol32 or the presence of H3K56ac [70–72]. It is worth noting, however, that other error-prone replication fork restart mechanisms are known, and these may have similar dependencies [83, 84].

This mechanism would be expected to yield both expansions and contractions, and we suggest that 2 additional factors drive the contraction bias we actually observe. Firstly, a fork restarted by HR within a high-copy sequence has a high chance of template switching through strand invasion of a homologous sequence in a different copy, whereas a unique sequence provides only a single homologous template and so template switching is disfavoured. Copy number amplification requires the fork to template switch backwards and re-replicate multiple copies (Fig 8b, upper), whereas contraction requires a template switch forwards that moves the fork closer to the end of the high-copy sequence (Fig 8b, lower). A fork that has switched backwards will therefore spend longer replicating a high-copy sequence than a fork that has switched forwards and would have a higher chance of template switching again, with further potential to generate a contraction (Fig 8b, right). The scheme in Fig 8b shows that 2 successive template-switching events can result in a 3:1 ratio of contractions to amplifications.

This factor would prevail in nicotinamide-treated cells in which H3K56ac is uniformly high, but in normal cells, H3K56ac is primarily on the newly synthesised histones behind the fork and so H3K56ac is also asymmetrical (Fig 8c). This means that a template switch forwards would lead to BIR using a previously unreplicated template with low H3K56ac and would result in a high-processivity replication fork (Fig 8c, upper). In contrast, a template switch backwards would result in the BIR fork using a previously replicated template with high H3K56ac, and this fork would have low processivity and a higher chance of template switching again (Fig 8c, lower). Together, we believe that these 2 factors would yield a major bias towards contraction events but would not prevent amplifications occurring at a lower frequency.

We initially hypothesised that CNV at the *CUP1* locus would be mechanistically equivalent to CNV at the rDNA. Indeed, the requirement for bidirectional promoter induction and the effect of H3K56 HDACs are similar in the 2 systems. However, we see important differences: a lack of dependence on Sir2, which is not surprising as Sir2 acts primarily at heterochromatin; and the suppression of *CUP1* CNV in *rtt109*Δ cells, which conversely undergo massive rDNA amplification [25, 26]. The mechanistic analysis of rDNA amplification in *rtt109*Δ mutants by the Kobayashi group [26] provides an explanation for this discrepancy: rDNA amplification in *rtt109*Δ mutants does not proceed via chromosomal BIR; instead, rolling circle amplification of extrachromosomal rDNA circles (ERCs) forms large arrays of additional rDNA copies that

can reintegrate into the chromosomal rDNA locus [26]. *CUP1* circles have been detected but are 3–4 orders of magnitude rarer than ERCs [85], and we suspect that rolling circle amplification of these happens in too few cells to make a detectable contribution.

Adaptive potential of stimulated CNV

Stimulated CNV controls the occurrence of a subset of mutations that allow adaptation to challenging environments. It is commonly assumed that adaptive mutations occur at random, and they are largely inevitable as they occur through multiple poorly defined mechanisms. Under this assumption, loss of many genome stability factors would increase the rate of adaptation, but adaptation cannot be suppressed. In contrast, we show that the adaptation of yeast to environmental copper by amplification of *CUP1* is largely dependent on a defined pathway that requires Rtt109, despite the general role of this protein in maintaining genome stability. Therefore, adaptation occurring through apparently random mutation may in fact be stimulated by a specific cellular mechanism.

How widespread is stimulated CNV? The *CUP1* and *SFA1* model systems we have analysed are multicopy, but although CNV will be most efficient where multiple homologous sequences surround the RFS site, this mechanism should not be restricted to multicopy loci. Recombination events triggered by error-prone replication forks could easily utilise distal homologous sequences or even microhomology as templates, inducing de novo CNVs and chromosomal translocations with limited homology at the breakpoints. Breakpoints in de novo CNVs would therefore be poorly defined because they initiate nearby but not at the RFS site and utilise unpredictable homologous sequences. Interestingly, *CUP1* shows exactly this behaviour in different *S. cerevisiae* isolates, as the multicopy *CUP1* repeat has emerged many times with different breakpoints [58].

Replication fork stalling is by no means restricted to budding yeast, and bidirectional promoters are the norm in organisms, including mammals [86, 87]. Therefore, the basic machinery required for stimulated CNV is likely to be conserved. Furthermore, the histone deacetylases that regulate CNV outcome are conserved in mammals and appear to have similar functions in modulating DNA repair [88, 89]. Stimulated CNV in somatic cells of metazoans is rarely likely to be a useful organismal outcome and cannot aid heritable adaptation. However, because stimulated CNV emerges from conserved features of the replication and transcription systems, it seems likely that it would be active in mammalian cells, providing a mechanism that could be readily exploited, for example, by tumour cells. The mechanism that we have proposed is also very consistent with recent reports of nonrandom double-strand breaks formed by neurons in genes important for neuronal function [90, 91]. As such the stimulated CNV pathway provides a new set of targets by which pharmaceuticals may prevent the emergence of undesirable properties such as drug resistance in tumours or modulate natural genetic changes in particular cell types. Indeed, our observation that adaptation of yeast to copper can be effectively suppressed by removal of Rtt109, a protein for which chemical inhibitors have been described [92, 93], provides good evidence that the emergence of resistance is pharmacologically accessible.

Evidence for adaptation through genome-wide nonrandom mutation is substantial, particularly in bacteria [18], but the ability of stimulated CNV to direct mutations to relevant loci must be reconciled with forceful arguments against previously proposed 'directed mutation' systems [18, 30, 94, 95]. The primary issue is that any general mechanism that directs mutations to a particular site must 'know' in advance the fitness outcome of a particular genetic change, which is not possible except at singular, highly specialised loci such as the rDNA. However, such arguments ignore the wealth of information regarding the function of

particular loci in particular environments that is encapsulated in existing gene regulatory systems. In effect, a signalling pathway that strongly induces a gene in response to a particular environmental stimulus marks that gene as being important in that environment relative to a gene that is not expressed. CNV of such a gene is more likely to yield a useful, adaptive result than CNV of a random gene. Of course, it is also more likely to be damaging, and we see exactly this at *CUP1*: most of the CNV events we observed were contractions that reduce copper resistance (Fig 3b), but, relative to random mutation, the chance of finding an adaptive CNV remains substantial. Stimulated CNV is therefore a high-risk strategy that does not entail foreknowledge of the fitness outcome of genetic change at a particular locus, only the relative importance of that locus in the current environment.

However, many highly expressed genes are not environment specific, and such housekeeping genes are likely to be poor candidates for improving adaptation. Simply focusing CNV events at highly expressed genes would likely entail an unacceptable number of deleterious CNV events involving housekeeping genes. This problem is avoided by restricting RFS sites to the promoters of inducible genes. We suggest that the distribution of RFS sites has arisen through natural selection acting on randomly located RFS sites, as any RFS site that always engenders detrimental CNVs would have been rapidly lost.

Stimulated CNV is therefore an imperfect but useful cellular mechanism that increases the rate at which adaptive CNV events occur, particularly in suboptimal environments. Importantly, inducible gene expression systems and the placement of RFS sites are products of natural selection acting on random mutations, but these combine to yield a system that accelerates adaptation beyond what is achievable through random mutation.

Materials and methods

Yeast strains and media

Yeast strains used in this work are listed in S2 Table. Plasmids are listed in S3 Table, including construction details, and were verified by restriction digest and/or sequencing. Deletion strains were created by standard methods; oligonucleotides are listed in S4 Table. Deletion strains were verified by PCR. To create the P_{GAL1}-*HA* strain, *ADE2* was replaced with *MET25* in the *S. cerevisiae* strain BY4741 (EuroSCARF), and the resulting strain was transformed with pJH252 (1 *CUP1* repeat); then, the entire repeated region at the chromosomal *CUP1* locus [ChrVIII: 212265..216250] was replaced with a *LEU2* marker. Plasmid pJH280, containing 3 copies of the P_{GAL1}-*HA* construct and an *ADE2* marker, was digested with *Sac*I and transformed to replace the *LEU2* cassette, which fortuitously amplified on transformation to yield a 17-copy repeat tract (based on PFGE Southern blot migration). The 17xP_{GAL1}-*3HA* construct was introduced into the MEP background by mating and sporulation and was remated to a MEP wild-type haploid to form the MEP 17xP_{GAL1}-*3HA* heterozygote strain.

For construction of the 3x*CUP1* strain and its derivatives, the entire *CUP1* locus [ChrVIII: 212265..216250] was deleted in YRH12, an *ade2*Δ BY4742-derivative with a single-copy *CUP1* plasmid, to form YRH15. pRH9, which contains 3 complete *CUP1* copies [ChrVIII: 214256..216239] with an *ADE2* marker and *CUP1* flanking sequences, was digested with *Sac*I and transformed in YRH15, followed by FOA selection to yield YRH23.

Construction of the 3x*SFA1* strain and its derivatives used the same strategy as for 3x*CUP1*, except that *SFA1* was additionally deleted in YRH15 and then transformed with *Sac*I-digested pJH312, followed by FOA selection to yield YRH89.

Cells for Fig 2c and 2d were grown in YPD, and other experiments were performed in yeast nitrogen base media supplemented with CSM amino acids and 2% glucose or galactose; all cells were grown at 30˚C. YPD contains trace Cu^{2+}, and yeast nitrogen base media contains

250 nM $CuSO_4$. All media components were purchased from Formedium. Nicotinamide (Sigma I17451) was added to media at 5 mM. For Fig 6b–6d, cells were grown in SC with or without 0.3 mM $CuSO_4$ in 4-ml cultures, diluted 1:1,000 from saturated precultures. For Fig 6e and 6f, cells were grown to log phase in SC then diluted 1:8,000 into SC with or without 0.9–1 mM (batch-dependent) formaldehyde that was freshly diluted from 16% or 37% stock solution. Fig 2a represents a meta-analysis of published data; see the Bioinformatic analysis section below for accession numbers and associated culture details.

For cell-tracking analyses using the MEP system, cells were inoculated in SD from a plate for 6–8 hours, diluted and grown overnight to OD 0.2–0.5. Cultures were diluted to $2x10^4$ cells/ml, 1 μM β-estradiol (Sigma E2758) was added, and cells were grown for 2 hours prior to plating parental culture and splitting the cells for copper or galactose treatment. After 24 hours, 50 μl of each culture was plated on SD agar, and cells were grown for 2 days at 30˚C. Individual colonies were then inoculated in 200 μl of SD in a 96-well plate and grown to saturation.

Copper sensitivity and growth curve analysis

For the MEP strain, 2.5 μl saturated culture was diluted to 200 μl SC in each well of a 96-well flat-bottomed cell culture plate, with concentrations of $CuSO_4$ up to 3 mM, along with 0.5 mM ascorbic acid. Ascorbate increases the cellular uptake of copper [96], increasing the effective toxicity of copper to allow the measurement of small changes in resistance in cells with high *CUP1* copy number. This is helpful since $CuSO_4$ tends to precipitate out of media during culture at concentrations >2 mM. Plates were covered with a gas-permeable membrane and grown at 30˚C for 3 days in the dark. Cells were resuspended by pipetting, and OD_{660} was measured using a BD FLUOstar Omega plate reader. Area-under-curve measurements were calculated for each sample and compared by 1-way ANOVA. For the 3x*CUP1* strain, the assay was performed as above but with lower concentrations of $CuSO_4$ (see Fig 7a and 7b), under normal light and without ascorbic acid.

For growth curves, saturated precultures were diluted 1:1,000 into 200 μl SC per well with or without $CuSO_4$ at the required concentration. Plates were sealed as above and grown at 30˚C with shaking in a BD FLUOstar Omega plate reader; OD_{660} measurements were taken every 15 minutes. Curves were smoothed by averaging across 9 time points, and derivatives were calculated using GraphPad Prism.

Competitive growth assay

Six cultures each of 3x*CUP1* wild-type and *trp1*Δ::NatMX6 were grown for 10 generations, 3 untreated and 3 with 5 mM nicotinamide. Cultures were then mixed 1:1 pairwise to give 6 competition cultures, each containing an untreated and a nicotinamide pretreated population of the opposite genotype. The composition of the mixture was determined by plating on–Trp and +Nat plates, and each mixture was inoculated 1:1,000 in cultures containing 0 or 0.3 mM $CuSO_4$ and outgrown to saturation over 10 generations. Mixture composition of each outgrowth culture was determined by plating. To ensure that the *trp1*Δ::NatMX6 marker did not affect the result, we performed equal numbers of assays with this strain as the nicotinamide-treated or untreated population.

DNA extraction and Southern blotting

Cells from a 2 ml saturated culture were washed with 50 mM EDTA then spheroplasted with 250 μl 0.34U/ml lyticase (Sigma L4025) in 1.2 M sorbitol, 50 mM EDTA, and 10 mM DTT at 37˚C for 45 minutes. After centrifuging at 1,000g, cells were gently resuspended in 400 μl of 0.3% SDS, 50 mM EDTA, and 100μg/ml RNase A (Sigma R4875) and incubated at 37˚C for 30

minutes. 4 μl of 20 mg/ml proteinase K (Roche 3115801) was added, and samples were mixed by inversion and heated to 65˚C for 30 minutes. 160 μl 5M KOAc was added after cooling to room temperature, and samples were mixed by inversion and then chilled on ice for 30 minutes. After 10 minutes of centrifuging at 20,000g, the supernatant was poured into a new tube containing 500 μl phenol:chloroform (pH 8) and samples were mixed on a wheel for 30 minutes. Samples were centrifuged for 10 minutes at 10,000g, and the upper phase was extracted using cut tips and precipitated with 400 μl isopropanol. Pellets were washed with 70% ethanol, air-dried, and left overnight at 4˚C to dissolve in 20 μl TE. After gentle mixing, 10 μl of each sample was digested with 20 U *Eco*RI-HF (NEB) or *Eco*RV-HF (NEB) for 3 hours, phenol:chloroform extracted, ethanol precipitated, and separated on 25-cm 0.8% or 1% 1xTBE gels overnight at 120 V for *CUP1* analysis or on 1% 0.5xTBE gels in a Bio-Rad CHEF DR-III system at 6 V/cm, 15˚C, 0.5–1.5 second switch, and 120˚ included angle for 16 or 20 hours in 0.5xTBE for *SFA1* analysis. Gels were washed in 0.25 N HCl for 15 minutes, 0.5 N NaOH for 45 minutes and twice in 1.5M NaCl, 0.5M Tris (pH 7.5) for 20 minutes before being transferred to HyBond N+ membrane in 6x SSC. Membranes were probed using random primed probes (S4 Table) in UltraHyb (Life Technologies) at 42˚C and washed twice with 0.1 x SSC, 0.1% SDS at 42˚C. Bands were quantified using ImageQuant (GE) and data analysed using the GraphPad Prism v6.05 to perform 1-way ANOVA analyses comparing the means of all samples (unless otherwise noted) with Tukey correction for multiple comparisons.

Rapid DNA extraction and PFGE for cell fate tracking analysis

Colonies were analysed in pools of 4. Cells obtained from 50 μl from each of the 4 saturated cultures were resuspended in 50 μl 50 mM EDTA containing 17 U lyticase (Sigma L2524) and incubated at 37˚C for 45 minutes. 1.6 μl 10% SDS and 1 μl 20 mg/ml proteinase K were added and samples were incubated at 65˚C for 30 minutes. After addition of 32 μl 5 M KOAc and 30 minutes on ice, samples were centrifuged for 10 minutes at 20,000g at room temperature, and the supernatant was decanted to a new tube containing 100 μl isopropanol and 1 μl glycogen. Samples were centrifuged for 15 minutes at 20,000g at 4˚C, and the pellet was washed with 70% ethanol before overnight elution in 20 μl 1x NEB CutSmart buffer with 20 U *Eco*RI-HF (NEB) at 37˚C. DNA was quantified using PicoGreen (Thermo Fisher Scientific) and separated on PFGE gels using a Bio-Rad CHEF DR-III (1% 0.5xTBE gel, 6 V/cm, 15˚C, 0.5–1.5 second switch, 120˚ included angle for 20 hours in 0.5xTBE), then blotted and probed as above. Copy numbers of individual alleles were plotted using GraphPad Prism v6.05.

To calculate *p*-values, we first estimated the background CNV or amplification mutation rate in the population based on the number of CNV or amplification events observed by PFGE for the unstimulated condition, including the viability of this population after 24 hours of aging. Using this estimate, we then calculated the number of CNV or amplification events in the stimulated condition that would be expected to arise through unstimulated CNV given the viability of this population. We then compared the number of events observed by PFGE to the expected number of CNV or amplification events for the stimulated condition using a goodness of fit χ^2 test with 1 degree of freedom. This provides a *p*-value based on the null hypothesis that all observed CNV or amplification events arose through random mutation. This estimate includes the conservative assumption that any cells that lost viability during the experiment did not undergo CNV.

RNA extraction and northern analysis

Total RNA was extracted using a mirVANA kit (Thermo Fisher Scientific) according to manufacturer's instructions (Fig 2) or using GTC-phenol as described [7] (Figs 4, 5 and 6), and

analysed as previously described [7] using probes listed in S4 Table. RNA probes were hybridised at 65˚C, DNA probes at 42˚C. Indexed mRNAseq libraries were constructed from 500 ng total RNA using the NEBNext Ultra Directional RNA Library Prep Kit (NEB), with poly(A) selection using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB), and sequenced on an Illumina MiSeq.

Chromatin immunoprecipitation

$0.5x10^9$ cells grown in YPD, with or without 4 hours of 1 mM $CuSO_4$ treatment, were fixed for 15 minutes in 1% formaldehyde and quenched with 150 mM glycine. Cells were washed 2 times with cold PBS, then resuspended in 600 μl lysis buffer (50 mM HEPES [pH 7.5], 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 1x Roche Complete Protease Inhibitors), broken with 500 μl 0.5mm zirconium beads (BioSpec) in an MP Biomedical Fast Prep (6 cycles, 30 seconds each), then the lysate was separated from the beads and diluted to 1 ml final volume in lysis buffer. Samples were sonicated 19 times, 30 seconds each in a Diagenode Bioruptor on High and cleared by centrifugation at 20,000g at 4˚C for 5 minutes. 1 μl anti-γH2A (Millipore 07-745-I) was added to 100 μl lysate and incubated overnight at 4˚C before addition of 15 μl Gammabind beads (GE) in 25 μl lysis buffer (preblocked with 1% BSA) and incubation for 2 hours at 4˚C. Beads were washed 5 minutes each with lysis buffer, 0.5 M salt lysis buffer, wash buffer (10 mM Tris [pH 8.0], 0.25 M LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA), and TE, then DNA was eluted overnight at 65˚C in 200 μl 50 mM Tris [pH 8.0], 10 mM EDTA, 1% SDS. DNA was purified by phenol:chloroform extraction and then ethanol precipitated and eluted in 50 μl TE. Sequencing libraries were prepared from 5 ng of immunoprecipitated material using a NEBNext DNA Ultra kit (NEB) and sequenced on an Illumina HiSeq.

Bioinformatic analysis

γH2A ChIP: Reads were mapped to the *S. cerevisiae* reference genome R64-2-1 using Bowtie 2 v2.2.5 (default parameters). Peaks were called using MACS2 v2.1.0 (-g 12e6—nomodel—extsize 250—keep-dup all), artifactual peaks containing a single mismapped read were manually removed, and only peaks with a 2-fold or greater enrichment were considered. Coding sequences (CDS) were categorised as upstream-RFS if a γH2A peak was present in 1 kb upstream of the annotated start site using the R script in S1 Text.

RNAseq: Read data (including accessions GSE61783 [97], GSE74642, GSE70835, GSE54831 [98], GSE54825 [99], GSE43002 [100], and GSE41834 [101] deposited at GEO) were mapped to genome R64-2-1 using HISAT2 v2.0.5 (—sp 1000,1000). Log_2 read counts were performed for each annotated CDS using Seqmonk v0.34.1 and normalised for CDS length, then the whole data set was normalised for a median expression of 9 (an arbitrary value that maintained most expression data as positive and required minimal scaling for most data sets). Genes were categorised as γ-H2A or non-γH2A using the R script in S1 Text. Cumulative frequency distributions for the data sets were calculated using GraphPad Prism v6.05. Frequency distributions were compared by nested ANOVA, and assumptions for using parametric tests were checked prior to run the analyses. Values for skewness, kurtosis, and variance were consistent with normality and homoscedasticity.

Image processing and data availability

Sequencing data has been deposited at GEO (GSE86283). Source data for all graphs is provided in S1 Data (cumulative frequency gene expression data), S2 Data (RNAseq data for cells grown with or without Cu), S3 Data (Southern blot image quantification), S4 Data (northern blot

image quantification), S5 Data (γH2A ChIPseq data for chromosomes containing *CUP1* and *SFA1* loci), S6 Data (mother enrichment allele copy numbers and adaptation assay), S7 Data (3x*CUP1* adaptation assays and competition assay), and S8 Data (Raw growth curve data).

Gel images were processed with ImageJ 1.50i—processing involved rotating and cropping, denoising if required (Despeckle filter), and altering window-level settings to improve contrast of relevant bands. Full images of membranes presented in the manuscript are provided in S9 Data —these have been cropped to the borders of the membrane and have undergone minimal window-level adjustments if required to make the bands shown in the presentation figure visible.

Supporting information

S1 Fig. Cumulative frequency distributions showing expression of upstream-RFS genes. a: Data from matched set of cells grown in YPD and YPGlycerol, GSE74642, showing cumulative frequency distributions for the expression of genes either with (γH2A) or without (non-γH2A) an upstream stalled replication fork site. The YPGlycerol data set overlaps with some YPD data in Fig 2a but is clearly separable from the matched YPD control. **b**: Data sets as in **a** for cells subjected to oxidative (0.4 mM H_2O_2 for 30 minutes) or osmotic stress (0.4 M KCl for 10 minutes). Data were reanalysed from GSE42983 [100] and GSE61783 [97] using the R script provided in S1 Text; raw quantitation data are available in S1 Data.
(TIF)

S2 Fig. RNAseq analysis of copper-treated cells. Wild-type cells were grown in YPD with or without 1 mM $CuSO_4$ for 4 hours, analysed by poly(A)+ RNAseq, and read counts mapping to each annotated coding sequence (CDS) were calculated and normalised for feature length. **a**: Cumulative frequency distribution showing expression of genes with an upstream γH2A peak relative to control genes under each condition. **b**: Scatterplot of RNA levels, with γH2A genes highlighted in red. CDS for γH2A genes that are substantially induced or repressed by copper are annotated: the blue circle shows *CUP1* locus genes, the green circle shows the single other verified coding sequence induced by copper (*HSP12*), and orange circles are CDS representing the multi-copy, subtelomeric helicase ORFs. Raw quantitation data are available in S1 and S2 Data.
(TIF)

S3 Fig. Supplement to role of H3K56 acetylation in stimulated CNV. a: Southern analysis of *CUP1* copy number in wild-type, *sir2*Δ, and *hst3*Δ *hst4*Δ cells. **b**: Northern analysis of *CUP1* ORF and *CUP1* CUT RNA in log-phase wild-type, *mrc1*Δ, and *pol32*Δ cells. *p*-values are non-significant by 1-way ANOVA, $n = 3$. Raw quantitation data are available in S4 Data.
(TIF)

S4 Fig. *GAL1* promoter does not respond to nicotinamide. Northern analysis of *HA* ORF and *CUP1* CUT RNA in log-phase P_{GAL1}-*HA* cells grown on glucose or galactose with or without 5 mM nicotinamide. *p*-values were calculated by 1-way ANOVA, $n = 4$. Raw quantitation data are available in S4 Data.
(TIF)

S5 Fig. Nicotinamide stimulation of CNV to low copy number. Reanalysis of Southern data from Fig 5b, quantifying *CUP1* alleles with 1–3 copies compared to the total of all alleles. This shows that nicotinamide stimulation is particularly potent in the production of small alleles that presumably arise from multiple CNV events. *p*-values were calculated from pairwise comparisons of negative and positive NIC samples for each GLU or GAL concentration, derived from a 1-way ANOVA of the whole data set. Raw quantitation data are available in S3 Data.
(TIF)

S6 Fig. Characteristics of 3x*CUP1* and 3x*SFA1* systems. a: Growth curves of 3x*CUP1* cells growing with or without 0.3 mM $CuSO_4$. Note that the growth retardation caused by 0.3 mM $CuSO_4$ is stronger in the 200 μl 96-well plate cultures used for growth curve analysis than in the 4 ml batch cultures used for Southern blot samples; although cells also grow slowly in 0.3 mM $CuSO_4$ under these conditions, almost all cultures reach saturation by 72 hours. **b**: Southern analysis of *CUP1* copy number in 3x*CUP1* cells grown for 10 generations with or without 0.3 mM $CuSO_4$. Quantification shows the percentage of amplified alleles; $n = 6$, p-value calculated by t test. **c**: γH2A signal in the region surrounding *SFA1*; analysis performed as in Fig 2d. **d**: Induction of *SFA1* and upstream antisense transcripts after a 4-hour exposure to 1 mM formaldehyde, assayed by northern blot. 18S rRNA is shown as a loading control. Quantification shows the levels of the indicated RNA species in arbitrary units; p-values were calculated by t test, $n = 4$. Locations of probes within the *SFA1* repeat are shown in **e**. **e**: Schematic of the wild-type *SFA1* region in the 3x*SFA1* construct, along with the modified P_{GAL1}-*GFP-SFA1* construct. All *SFA1* copies carry this construct in the P_{GAL1}-*GFP-SFA1* strain. Transcriptional start sites indicated by black arrows are approximate. Raw quantitation data are available in S3, S4, S5 and S8 Data.
(TIF)

S7 Fig. Supplement to copper adaptation through stimulated CNV. a: *CUP1* copy number distribution of 3x*CUP1* wild-type and *rtt109*Δ cells after growth for 10 generations with with or without 0.3 mM $CuSO_4$, followed by growth with or without 1 mM $CuSO_4$ under adaptation curve conditions, then outgrown without drug. **b**: *CUP1* copy number distribution of 3x*CUP1* cells after growth for 10 generations with 5 mM nicotinamide, followed by growth with or without 0.75 mM $CuSO_4$ under adaptation curve conditions, then outgrown without drug. **c**: Individual growth curves of 3x*CUP1* cells preexposed to 5 mM nicotinamide before growth with or without 0.75 mM $CuSO_4$. **d**: Maximum growth rate data for +NIC +$CuSO_4$ samples from Fig 7d, separated to show distributions of data points derived from the 4 different precultures. Raw quantitation data are available in S8 Data.
(TIF)

S1 Table. List of *S. cerevisiae* genes with upstream γH2A peaks. Peaks were identified by MACS analysis of γH2A data on unstressed cells growing at log phase in YPD (see Materials and methods). Peaks were matched to annotated coding sequences (CDS) using an R script (provided as S1 Text).
(XLSX)

S2 Table. *S. cerevisiae* yeast strains used in this study.
(DOCX)

S3 Table. Newly derived plasmids used in this study. See S4 Table for oligonucleotides used to amplify cloning fragments.
(DOCX)

S4 Table. Oligonucleotide pairs used in this study.
(DOCX)

S1 Text. R script for meta-analysis of gene expression data sets.
(R)

S1 Data. Cumulative frequency gene-expression data.
(XLSX)

S2 Data. Gene-expression values derived from RNAseq.
(XLSX)

S3 Data. Southern blot quantification.
(XLSX)

S4 Data. Northern blot quantification.
(XLSX)

S5 Data. γH2A ChIPseq data for chromosomes containing *CUP1* and *SFA1* loci.
(XLSX)

S6 Data. Mother enrichment allele copy numbers and adaptation assay.
(XLSX)

S7 Data. 3x*CUP1* adaptation assays and competition assay.
(XLSX)

S8 Data. Raw data for growth curves.
(XLSX)

S9 Data. PDF file containing uncropped images of Southern and northern blots.
(PDF)

Acknowledgments

We thank Felix Krueger and Anne Segonds-Pichon for help with bioinformatic and statistical analysis, Dan Gottschling for the MEP strain, Wolf Reik and Jane Alfred for critical reading, and Simon Cook for insightful comments.

Author Contributions

Conceptualization: Ryan M. Hull, Jonathan Houseley.

Funding acquisition: Jonathan Houseley.

Investigation: Ryan M. Hull, Cristina Cruz, Carmen V. Jack, Jonathan Houseley.

Methodology: Ryan M. Hull, Jonathan Houseley.

Project administration: Jonathan Houseley.

Supervision: Jonathan Houseley.

Visualization: Ryan M. Hull, Jonathan Houseley.

Writing – original draft: Jonathan Houseley.

Writing – review & editing: Ryan M. Hull, Jonathan Houseley.

References

1. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet. 2015; 16(3):172–83. https://doi.org/10.1038/nrg3871 PMID: 25645873.
2. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004; 305(5683):525–8. https://doi.org/10.1126/science.1098918 PMID: 15273396.
3. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36(9):949–51. https://doi.org/10.1038/ng1416 PMID: 15286789.

4. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010; 464(7289):713–20. https://doi.org/10.1038/nature08979 PMID: 20360734;
5. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annual review of medicine. 2010; 61:437–55. https://doi.org/10.1146/annurev-med-100708-204735 PMID: 20059347.
6. van der Maarel SM, Tawil R, Tapscott SJ. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends in molecular medicine. 2011; 17(5):252–8. https://doi.org/10.1016/j.molmed.2011.01.001 PMID: 21288772;
7. Cruz C, Houseley J. Endogenous RNA interference is driven by copy number. eLife. 2014; 3:e01581. https://doi.org/10.7554/eLife.01581 PMID: 24520161;
8. Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, et al. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. PNAS. 2002; 99(25):16144–9. https://doi.org/10.1073/pnas.242624799 PMID: 12446845;
9. Gresham D, Usaite R, Germann SM, Lisby M, Botstein D, Regenberg B. Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. PNAS. 2010; 107(43):18551–6. https://doi.org/10.1073/pnas.1014023107 PMID: 20937885;
10. Huang T, Campbell JL. Amplification of a circular episome carrying an inverted repeat of the DFR1 locus and adjacent autonomously replicating sequence element of Saccharomyces cerevisiae. J Biol Chem. 1995; 270(16):9607–14. PMID: 7721892.
11. Libuda DE, Winston F. Amplification of histone genes by circular chromosome formation in Saccharomyces cerevisiae. Nature. 2006; 443(7114):1003–7. https://doi.org/10.1038/nature05205 PMID: 17066037.
12. Payen C, Koszul R, Dujon B, Fischer G. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet. 2008; 4(9): e1000175. PMID: 18773114. https://doi.org/10.1371/journal.pgen.1000175
13. Katoh M. Cancer genomics and genetics of FGFR2 (Review). International journal of oncology. 2008; 33(2):233–7. PMID: 18636142.
14. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463(7283):899–905. https://doi.org/10.1038/nature08822 PMID: 20164920;
15. Frei E 3rd, Rosowsky A, Wright JE, Cucchi CA, Lippke JA, Ervin TJ, et al. Development of methotrexate resistance in a human squamous cell carcinoma of the head and neck in culture. PNAS. 1984; 81 (9):2873–7. PMID: 6201865.
16. Corcoran RB, Dias-Santagata D, Bergethon K, Iafrate AJ, Settleman J, Engelman JA. BRAF gene amplification can promote acquired resistance to MEK inhibitors in cancer cells harboring the BRAF V600E mutation. Sci Signal. 2010; 3(149):ra84. https://doi.org/10.1126/scisignal.2001148 PMID: 21098728;
17. Little AS, Balmanno K, Sale MJ, Newman S, Dry JR, Hampson M, et al. Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. Sci Signal. 2011; 4(166):ra17. https://doi.org/10.1126/scisignal.2001752 PMID: 21447798.
18. Galhardo RS, Hastings PJ, Rosenberg SM. Mutation as a stress response and the regulation of evolvability. Critical reviews in biochemistry and molecular biology. 2007; 42(5):399–435. https://doi.org/10.1080/10409230701648502 PMID: 17917874;
19. Hastings PJ, Bull HJ, Klump JR, Rosenberg SM. Adaptive amplification: an inducible chromosomal instability mechanism. Cell. 2000; 103(5):723–31. PMID: 11114329.
20. Radman M. SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. Basic Life Sci. 1975; 5A:355–67. PMID: 1103845.
21. Shor E, Fox CA, Broach JR. The yeast environmental stress response regulates mutagenesis induced by proteotoxic stress. PLoS Genet. 2013; 9(8):e1003680. https://doi.org/10.1371/journal.pgen.1003680 PMID: 23935537;
22. Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. Annual review of immunology. 2008; 26:261–92. https://doi.org/10.1146/annurev.immunol.26.021607.090248 PMID: 18370922;
23. Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, et al. The biochemistry of somatic hypermutation. Annual review of immunology. 2008; 26:481–511. https://doi.org/10.1146/annurev.immunol.26.021607.090236 PMID: 18304001.

24. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annual review of immunology. 2006; 24:541–70. https://doi.org/10.1146/annurev.immunol.23.021704.115830 PMID: 16551259.

25. Houseley J, Tollervey D. Repeat expansion in the budding yeast ribosomal DNA can occur independently of the canonical homologous recombination machinery. Nucleic Acids Res. 2011; 39(20):8778–91. https://doi.org/10.1093/nar/gkr589 PMID: 21768125;

26. Ide S, Saka K, Kobayashi T. Rtt109 Prevents Hyper-Amplification of Ribosomal RNA Genes through Histone Modification in Budding Yeast. PLoS Genet. 2013; 9(4):e1003410. https://doi.org/10.1371/journal.pgen.1003410 PMID: 23593017;

27. Kobayashi T, Ganley AR. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science. 2005; 309(5740):1581–4. PMID: 16141077. https://doi.org/10.1126/science.1116102

28. Cairns J, Foster PL. Adaptive reversion of a frameshift mutation in Escherichia coli. Genetics. 1991; 128(4):695–701. PMID: 1916241;

29. Torkelson J, Harris RS, Lombardo MJ, Nagendran J, Thulin C, Rosenberg SM. Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation. EMBO J. 1997; 16(11):3303–11. https://doi.org/10.1093/emboj/16.11.3303 PMID: 9214645;

30. Lenski RE, Mittler JE. The directed mutation controversy and neo-Darwinism. Science. 1993; 259 (5092):188–94. PMID: 7678468.

31. Szostak JW, Wu R. Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae. Nature. 1980; 284(5755):426–30. PMID: 6987539.

32. Keil RL, Roeder GS. Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of S. cerevisiae. Cell. 1984; 39(2 Pt 1):377–86. PMID: 6094015.

33. Kobayashi T, Horiuchi T. A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. Genes Cells. 1996; 1(5):465–74. PMID: 9078378.

34. Kobayashi T, Heck DJ, Nomura M, Horiuchi T. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. Genes Dev. 1998; 12(24):3821–30. PMID: 9869636.

35. Stewart SE, Roeder GS. Transcription by RNA polymerase I stimulates mitotic recombination in Saccharomyces cerevisiae. Mol Cell Biol. 1989; 9(8):3464–72. PMID: 2677675;

36. Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell. 2005; 121(5):725–37. PMID: 15935759. https://doi.org/10.1016/j.cell.2005.04.030

37. Houseley J, Kotovic K, El Hage A, Tollervey D. Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. EMBO J. 2007; 26(24):4996–5006. https://doi.org/10.1038/sj.emboj.7601921 PMID: 18007593;

38. French SL, Osheim YN, Cioci F, Nomura M, Beyer AL. In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol Cell Biol. 2003; 23(5):1558–68. PMID: 12588976; https://doi.org/10.1128/MCB.23.5.1558-1568.2003

39. Jack CV, Cruz C, Hull RM, Keller MA, Ralser M, Houseley J. Regulation of ribosomal DNA amplification by the TOR pathway. PNAS. 2015; 112(31):9674–9. https://doi.org/10.1073/pnas.1505015112 PMID: 26195783;

40. Sima J, Gilbert DM. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. Current opinion in genetics & development. 2014; 25C:93–100. https://doi.org/10.1016/j.gde.2013.11.022 PMID: 24598232.

41. Liu P, Carvalho CM, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Current opinion in genetics & development. 2012; 22(3):211–20. https://doi.org/10.1016/j.gde.2012.02.012 PMID: 22440479;

42. Szilard RK, Jacques PE, Laramee L, Cheng B, Galicia S, Bataille AR, et al. Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. Nat Struct Mol Biol. 2010; 17 (3):299–305. https://doi.org/10.1038/nsmb.1754 PMID: 20139982;

43. Slack A, Thornton PC, Magner DB, Rosenberg SM, Hastings PJ. On the mechanism of gene amplification induced under stress in Escherichia coli. PLoS Genet. 2006; 2(4):e48. PMID: 16604155. https://doi.org/10.1371/journal.pgen.0020048

44. Aguilera A, Garcia-Muse T. Causes of genome instability. Annu Rev Genet. 2013; 47:1–32. https://doi.org/10.1146/annurev-genet-111212-133232 PMID: 23909437.

45. Datta A, Jinks-Robertson S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science. 1995; 268(5217):1616–9. PMID: 7777859.

46. Thomas BJ, Rothstein R. Elevated recombination rates in transcriptionally active DNA. Cell. 1989; 56 (4):619–30. PMID: 2645056.

47. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. The nature of mutations induced by replication-transcription collisions. Nature. 2016; 535(7610):178–81. https://doi.org/10.1038/nature18316 PMID: 27362223;

48. Mangiameli S, Merrikh CN, Wiggins PA, Merrikh H. Transcription leads to pervasive replisome instability in bacteria. eLife. 2017; 6:e19848. https://doi.org/10.7554/eLife.19848 PMID: 28092263.

49. Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. Accelerated gene evolution through replication-transcription conflicts. Nature. 2013; 495(7442):512–5. https://doi.org/10.1038/nature11989 PMID: 23538833;

50. Carr AM, Lambert S. Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination. J Mol Biol. 2013; 425(23):4733–44. https://doi.org/10.1016/j.jmb.2013.04.023 PMID: 23643490.

51. Raveendranathan M, Chattopadhyay S, Bolon YT, Haworth J, Clarke DJ, Bielinsky AK. Genome-wide replication profiles of S-phase checkpoint mutants reveal fragile sites in yeast. EMBO J. 2006; 25 (15):3627–39. https://doi.org/10.1038/sj.emboj.7601251 PMID: 16888628.

52. Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. Mol Cell. 2009; 34(6):722–34. https://doi.org/10.1016/j.molcel.2009.05.022 PMID: 19560424.

53. Karin M, Najarian R, Haslinger A, Valenzuela P, Welch J, Fogel S. Primary structure and transcription of an amplified genetic locus: the CUP1 locus of yeast. PNAS. 1984; 81(2):337–41. PMID: 6364141;

54. Prinz R, Weser U. A naturally occurring Cu-thionein in Saccharomyces cerevisiae. Hoppe Seylers Z Physiol Chem. 1975; 356(6):767–76. PMID: 1102411.

55. Fogel S, Welch JW. Tandem gene amplification mediates copper resistance in yeast. PNAS. 1982; 79 (17):5342–6. PMID: 6291039;

56. Fogel S, Welch JW, Cathala G, Karin M. Gene amplification in yeast: CUP1 copy number regulates copper resistance. Curr Genet. 1983; 7(5):347–55. https://doi.org/10.1007/BF00445874 PMID: 24173415.

57. Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, et al. The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. Genome research. 2015; 25(5):762–74. https://doi.org/10.1101/gr.185538.114 PMID: 25840857;

58. Zhao Y, Strope PK, Kozmin SG, McCusker JH, Dietrich FS, Kokoska RJ, et al. Structures of naturally evolved CUP1 tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. G3 (Bethesda). 2014; 4(11):2259–69. https://doi.org/10.1534/g3.114.012922 PMID: 25236733;

59. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002; 418(6896):387–91. PMID: 12140549. https://doi.org/10.1038/nature00935

60. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009; 457(7232):1033–7. https://doi.org/10.1038/nature07728 PMID: 19169243;

61. Adamo GM, Lotti M, Tamas MJ, Brocca S. Amplification of the CUP1 gene is associated with evolution of copper tolerance in Saccharomyces cerevisiae. Microbiology. 2012; 158(Pt 9):2325–35. https://doi.org/10.1099/mic.0.058024-0 PMID: 22790396.

62. Lindstrom DL, Gottschling DE. The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. Genetics. 2009; 183(2):413–22. https://doi.org/10.1534/genetics.109.106229 PMID: 19652178;

63. Gottlieb S, Esposito RE. A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. Cell. 1989; 56(5):771–6. PMID: 2647300.

64. Xu F, Zhang K, Grunstein M. Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell. 2005; 121(3):375–85. https://doi.org/10.1016/j.cell.2005.03.011 PMID: 15882620.

65. Celic I, Verreault A, Boeke JD. Histone H3 K56 hyperacetylation perturbs replisomes and causes DNA damage. Genetics. 2008; 179(4):1769–84. https://doi.org/10.1534/genetics.108.088914 PMID: 18579506;

66. Han J, Zhou H, Li Z, Xu RM, Zhang Z. Acetylation of lysine 56 of histone H3 catalyzed by RTT109 and regulated by ASF1 is required for replisome integrity. J Biol Chem. 2007; 282(39):28587–96. https://doi.org/10.1074/jbc.M702496200 PMID: 17690098.

67. Thaminy S, Newcomb B, Kim J, Gatbonton T, Foss E, Simon J, et al. Hst3 is regulated by Mec1-dependent proteolysis and controls the S phase checkpoint and sister chromatid cohesion by deacetylating histone H3 at lysine 56. J Biol Chem. 2007; 282(52):37805–14. https://doi.org/10.1074/jbc.M706384200 PMID: 17977840.

68. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, et al. Gene loops enhance transcriptional directionality. Science. 2012; 338(6107):671–5. https://doi.org/10.1126/science.1224350 PMID: 23019609;

69. Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, et al. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature. 2003; 424(6952):1078–83. https://doi.org/10.1038/nature01900 PMID: 12944972.

70. Deem A, Barker K, Vanhulle K, Downing B, Vayl A, Malkova A. Defective break-induced replication leads to half-crossovers in Saccharomyces cerevisiae. Genetics. 2008; 179(4):1845–60. https://doi.org/10.1534/genetics.108.087940 PMID: 18689895;

71. Smith CE, Lam AF, Symington LS. Aberrant double-strand break repair resulting in half crossovers in mutants defective for Rad51 or the DNA polymerase delta complex. Mol Cell Biol. 2009; 29(6):1432–41 PMID: 19139272. https://doi.org/10.1128/MCB.01469-08

72. Che J, Smith S, Kim YJ, Shim EY, Myung K, Lee SE. Hyper-Acetylation of Histone H3K56 Limits Break-Induced Replication by Inhibiting Extensive Repair Synthesis. PLoS Genet. 2015; 11(2): e1004990. https://doi.org/10.1371/journal.pgen.1004990 PMID: 25705897;

73. Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. A ncRNA Modulates Histone Modification and mRNA Induction in the Yeast GAL Gene Cluster. Mol Cell. 2008; 32(5):685–95. PMID: 19061643. https://doi.org/10.1016/j.molcel.2008.09.027

74. Adamczyk J, Deregowska A, Skoneczny M, Skoneczna A, Natkanska U, Kwiatkowska A, et al. Copy number variations of genes involved in stress responses reflect the redox state and DNA damage in brewing yeasts. Cell stress & chaperones. 2016; 21(5):849–864. https://doi.org/10.1007/s12192-016-0710-8 PMID: 27299603.

75. Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome research. 2012; 22(5):908–24. https://doi.org/10.1101/gr.130310.111 PMID: 22369888;

76. Steenwyk J, Rokas A. Extensive Copy Number Variation in Fermentation-Related Genes Among Saccharomyces cerevisiae Wine Strains. G3 (Bethesda). 2017; 7(5):1475–1485. https://doi.org/10.1534/g3.117.040105 PMID: 28292787.

77. Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, Abrusan G. Analysis of the largest tandemly repeated DNA families in the human genome. BMC Genomics. 2008; 9:533. https://doi.org/10.1186/1471-2164-9-533 PMID: 18992157;

78. Driscoll R, Hudson A, Jackson SP. Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. Science. 2007; 315(5812):649–52. https://doi.org/10.1126/science.1135862 PMID: 17272722;

79. Yang JH, Freudenreich CH. The Rtt109 histone acetyltransferase facilitates error-free replication to prevent CAG/CTG repeat contractions. DNA Repair (Amst). 2010; 9(4):414–20. https://doi.org/10.1016/j.dnarep.2009.12.022 PMID: 20083442;

80. Wehner EP, Rao E, Brendel M. Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in Saccharomyces cerevisiae, and characterization of its protein product. Molecular & general genetics:MGG. 1993; 237(3):351–8. PMID: 8483449.

81. Yeeles JT, Poli J, Marians KJ, Pasero P. Rescuing stalled or damaged replication forks. Cold Spring Harbor perspectives in biology. 2013; 5(5):a012815. https://doi.org/10.1101/cshperspect.a012815 PMID: 23637285;

82. Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell. 2007; 131(7):1235–47. https://doi.org/10.1016/j.cell.2007.11.037 PMID: 18160035

83. Lambert S, Mizuno K, Blaisonneau J, Martineau S, Chanet R, Freon K, et al. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. Mol Cell. 2010; 39(3):346–59. https://doi.org/10.1016/j.molcel.2010.07.015 PMID: 20705238.

84. Mizuno K, Miyabe I, Schalbetter SA, Carr AM, Murray JM. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. Nature. 2013; 493(7431):246–9. https://doi.org/10.1038/nature11676 PMID: 23178809;

85. Moller HD, Parsons L, Jorgensen TS, Botstein D, Regenberg B. Extrachromosomal circular DNA is common in yeast. PNAS. 2015; 112(24):E3114–22. https://doi.org/10.1073/pnas.1508825112 PMID: 26038577.

86. Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, et al. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. Nucleic Acids Res. 2011; 39(16):7179–93. https://doi.org/10.1093/nar/gkr370 PMID: 21596787;

87. Wei W, Pelechano V, Jarvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. Trends Genet. 2011; 27(7):267–76. https://doi.org/10.1016/j.tig.2011.04.002 PMID: 21601935;

88. Yuan J, Pu M, Zhang Z, Lou Z. Histone H3-K56 acetylation is important for genomic stability in mammals. Cell Cycle. 2009; 8(11):1747–53. https://doi.org/10.4161/cc.8.11.8620 PMID: 19411844;

89. Das C, Lucia MS, Hansen KC, Tyler JK. CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature. 2009; 459(7243):113–7. https://doi.org/10.1038/nature07861 PMID: 19270680;

90. Schwer B, Wei PC, Chang AN, Kao J, Du Z, Meyers RM, et al. Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. PNAS. 2016; 113 (8):2258–63. https://doi.org/10.1073/pnas.1525564113 PMID: 26873106;

91. Wei PC, Chang AN, Kao J, Du Z, Meyers RM, Alt FW, et al. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. Cell. 2016; 164(4):644–55. https://doi.org/10.1016/j.cell.2015.12.039 PMID: 26871630;

92. Ruotolo R, Tosi F, Vernarecci S, Ballario P, Mai A, Filetici P, et al. Chemogenomic profiling of the cellular effects associated with histone H3 acetylation impairment by a quinoline-derived compound. Genomics. 2010; 96(5):272–80. https://doi.org/10.1016/j.ygeno.2010.08.005 PMID: 20732410.

93. Lopes da Rosa J, Bajaj V, Spoonamore J, Kaufman PD. A small molecule inhibitor of fungal histone acetyltransferase Rtt109. Bioorg Med Chem Lett. 2013; 23(10):2853–9. https://doi.org/10.1016/j.bmcl.2013.03.112 PMID: 23587423;

94. Cairns J, Overbaugh J, Miller S. The origin of mutants. Nature. 1988; 335(6186):142–5. https://doi.org/10.1038/335142a0 PMID: 3045565.

95. Roth JR, Kofoid E, Roth FP, Berg OG, Seger J, Andersson DI. Regulating general mutation rates: examination of the hypermutable state model for Cairnsian adaptive mutation. Genetics. 2003; 163 (4):1483–96. PMID: 12702691;

96. Hassett R, Kosman DJ. Evidence for Cu(II) reduction as a component of copper uptake by Saccharomyces cerevisiae. J Biol Chem. 1995; 270(1):128–34. PMID: 7814363.

97. Adhikari H, Cullen PJ. Metabolic respiration induces AMPK- and Ire1p-dependent activation of the p38-Type HOG MAPK pathway. PLoS Genet. 2014; 10(10):e1004734. https://doi.org/10.1371/journal.pgen.1004734 PMID: 25356552;

98. Mayhew D, Mitra RD. Transcription factor regulation and chromosome dynamics during pseudohyphal growth. Molecular biology of the cell. 2014; 25(17):2669–76. https://doi.org/10.1091/mbc.E14-04-0871 PMID: 25009286;

99. Lin Y, Chomvong K, Acosta-Sampson L, Estrela R, Galazka JM, Kim SR, et al. Leveraging transcription factors to speed cellobiose fermentation by Saccharomyces cerevisiae. Biotechnology for biofuels. 2014; 7(1):126. https://doi.org/10.1186/s13068-014-0126-6 PMID: 25435910;

100. Baker LA, Ueberheide BM, Dewell S, Chait BT, Zheng D, Allis CD. The yeast Snt2 protein coordinates the transcriptional response to hydrogen peroxide-mediated oxidative stress. Mol Cell Biol. 2013; 33 (19):3735–48. https://doi.org/10.1128/MCB.00025-13 PMID: 23878396;

101. Carvalho-Netto OV, Carazzolle MF, Mofatto LS, Teixeira PJ, Noronha MF, Calderon LA, et al. Saccharomyces cerevisiae transcriptional reprograming due to bacterial contamination during industrial scale bioethanol production. Microb Cell Fact. 2015; 14:13. https://doi.org/10.1186/s12934-015-0196-6 PMID: 25633848;