LOSS OF HETEROZYGOSITY DUE TO SHORT-TRACT AND LONG-TRACT GENE CONVERSIONS AT FRAGILE SITE FS2 IN *SACCHAROMYCES CEREVISIAE*

Thomas Coates

Faculty Sponsor: Anne Casper, Ph.D

A Senior Thesis Submitted to the Eastern Michigan University

Honors College

In Partial Fulfilment of the Requirement for Graduation

With Honors in Biology

Approved at Ypsilanti, Michigan, on this date _____

Supervising Instructor

Acting Honors Advisor

Department Head

Honors Director

Table of Contents

List of Tables

List of Figures

Abstract

The repair of DNA breaks at common fragile sites can lead to a loss of heterozygosity at genes surrounding these events. If loss of heterozygosity leads to deactivation of tumour suppressor genes, it can accelerate or initiate tumorigenesis. Homologous recombination mechanisms are often used to repair DNA breaks, such as those at common fragile sites. Of the recombination mechanisms used to repair fragile site breaks, the least well understood are mechanisms leading to gene conversions. Here, using the model yeast organism *Saccharomyces cerevisiae* containing the fragile site FS2, we find evidence for two types of gene conversion following breaks in DNA at the fragile site: short-tract gene conversion and long-tract gene conversion. We find the frequency of gene conversions over a gene marker adjacent to the fragile site FS2 is $6.23*10^{-4}$ gene conversions per cell, and find evidence for two possible classification methods of short-tract and long-tract gene conversions. We believe these different types of gene conversions may have implications for the possible extent of loss of heterozygosity after repair at fragile sites, with long-tract gene conversions contributing to greater tumorigenesis.

Introduction

The development of medical sciences and our understanding of the biochemical nature of many disorders and diseases over the last fifty years has led to giant leaps in our ability to combat and prevent mortality and fatality from genetic and pathogenic causes. Vaccinations and the development of antibiotics alone have allowed us to statistically decimate communicable disease. However as old problems are overcome, new ones arise. As an indirect result of the progress made in combatting communicable diseases during the twentieth century, the burden of cancer on society has grown to the point where it is now the second leading cause of death in developing nations, and the leading cause in developed nations (Jemal *et al*., 2011). Unfortunately with our current understanding of cancer, we are limited in our influence over the disease While we understand some behaviours increase our risk of cancer, we can do very little to proactively prevent the disease. The economic impact of cancer is also astounding. The 2008 estimate for the global cost of cancer due to premature death and disability was $895 Billion. In light of this, it is clearly vital that we continue to increase our understanding of this disease if we wish to reduce its impact in society.

Cancer is a genetic disease, caused by the ill effects of genomic damage in the DNA of otherwise non-cancerous cells (Stratton *et al*., 2009). While the disease is given a simple umbrella term, all cancers are unique due to the complex interplay between the genetic components of a cell that control its replication cycle. In order to become tumorous a cell must obtain sufficient mutations to be able to reproduce rapidly, without being identified as tumorous by the immune system, and obtain and maintain a supply of nutrients sufficient to feed this unnatural growth. In addition to this, cancerous cells may

also metastasize, leaving their native tissue and entering other regions of the body. Tumours that are able to metastasize are termed malignant tumours, or cancers. Tumours that are not able metastasize are termed benign tumours. Benign tumours may be cause for threat due to their location (*i.e*. the brain), but it is malignant cancers that are more often the cause of disability and death in humans (Stratton *et al*., 2009).

The genetic damage that allows a cell to become cancerous is widely varied from one case to the next, but there are some generalizable changes that have been identified as important for cancer initiation. Cancer cells generally undergo a series of genetic alterations, leading to the development of a cancerous cell, this process is known as tumorigenesis (Bignell *et al*., 2010). Damage in genes that encode proteins that control the cell cycle is extremely common in cancer cells. Genes encode proteins that promote cell division are termed proto-oncogenes (when mutated, these are referred to as oncogenes). Genes encode proteins thattypically control cell cycle checkpoints, allowing for detection of incorrect growth and for abortion of replication in the instance of damaged DNA are termed tumour suppressor genes. Mutations in either one of these classes of genes are very common for cancerous cells, with oncogenes typically mutated into becoming over active, and tumour suppressor genes typically being deactivated (Stratton *et al*., 2009).

Loss of heterozygosity (LOH) is one method of losing functionality in a tumour suppressor gene. LOH involves the loss of function of a particular, functional, copy of a gene and subsequent conversion of the host cell to homozygous or hemizygous. In cells containing only one functional copy of a tumour suppressor gene LOH can result in deactivation of a tumour suppressor gene and subsequent tumorigenesis.

Fragile sites are regions within the genome prone to breaking when a cell is undergoing stressed replication (Le Tallec et al., 2014). The study of fragile sites began with the study of the FRAXA rare fragile site, and this field of investigation grew when breaks at other regions in the human genome were identified in all individuals under the conditions (low folate) being used to induce breaks at FRAXA (Durkin and Glover 2007; Arlt *et al*., 2003; Glover *et al.,* 1984). There are currently two consolidated classes of fragile sites, each with its own unique structural components and implications for clinical pathology. Rare fragile sites (RFS) are heritable regions in the genome that are the result of a nucleotide repeat expansion (generally tri-nucleotide repeats), these are only found in ~5% of humans (Durkin and Glover 2007). Common fragile sites (CFS) are highly conserved regions of the genome found in all humans that are a hotspot for DNA breaks when the cell is under replication stress (Palumbo *et al*., 2010). A possible third new class of fragile site was recently reported, termed "early replicating fragile sites"; however, little is known or understood of them (Glover and Wilson, 2013).

Rare fragile sites (RFS) are uncommon in the human genome, and are inherited in a Mendelian manner. They are the result of expansions in tri- and dinucleotide repeats (Durkin and Glover, 2007). This increase in tandem repeats allows for the formation of secondary structures, leading to fragility in rare fragile sites during replication. In the clinical context, rare fragile sites are often linked with conditions causing mental retardation. Breaks at RFS are implicated in conditions such as Fragile X syndrome (RFS, FRAXA) and Jacobsen syndrome (RFS, FRA11B) as well as being associated with many forms of uncategorized mental retardation (Debacker and Kooy, 2007; Matuszek *et al*., 2009; Winneonnenckx *et al*., 2007).

Common Fragile Sites in Human Disease

Hotspots for genetic alterations are of particular importance in the development of cancers. Common fragile sites are a highly conserved part of the human genome, present in all individuals and prone to breakage when a cell is under replication stress. They are regions prone to breaking under replication stress and the result of this damage has been implicated in cancer development (Bignell *et al.,* 2010). There are 87 common fragile sites currently recognized in the human genome(Ozeri-Galai *et al*., 2011).

Of great importance is the identification of the genetic consequences of breaks at fragile sites, as well as the discovery of the mechanisms that cause and repair these breaks at the molecular level. CFS have thus far been implicated in a number of human disorders.

Cancer. Alterations in the genome leading to tumorigenesis can be the result of many things including (but not limited to) amplification, deletion, point mutations, translocations or insertions of specific genes or genomic regions. Tumour suppressor genes are a class of genes whose role is vital to ensuring a cell does not become neoplastic (Stratton *et al*., 2009). Recently the loss of function of tumour suppressor genes in breast, lung, ovarian, prostate, oesophageal and other cancers have been correlated with the presence of CFSs in those genes (Arlt *et al*., 2002; Ingvarsson, 2001; Aqeilan *et al*., 2007; Huebner and Croce, 2001; Fang *et al*., 2001).

The fragile histidine triad (*FHIT*) gene is a, large (~1 Mb), well-characterised tumour suppressor gene found on the human chromosome III. *FHIT* consists of 10 small exon regions, and when expressed produces a processed product of only 1.1kb (Smith *et*

al., 2006). Loss of *FHIT* can be indicative of the early stages of tumorigenesis (Ingvarsson, 2001). Overlapping in sequence with the *FHIT* gene is the large (~4Mb) CFS FRA3B, notable as being the most commonly broken in lymphocytes (the cell type traditionally used in CFS research) and first characterised CFS (Durkin and Glover 2007; Smith *et al*., 2006).

Loss of the functionality of the *FHIT* gene due to loss of heterozygosity, deletion or translocation appears to be a significant component of the development of some of breast, renal, oesophageal and many lung cancers (Arlt *et al*., 2002; Ingvarsson, 2001; Aqeilan *et al*., 2007; Huebner and Croce, 2001; Fang *et al*., 2001). While the results of the studies above showed evidence for breakages as FHIT to be a driven in tumour development, it is interesting to note that deletions and other mutations found in tumour tissue associated with the FRA3B region in *FHIT* are frequently within intron regions, and most commonly only found in introns later in tumour development (Letessier *et al*., 2011; Smith *et al*., 2006). This indicates that mutations in intron regions may affect *FHIT* function through loss of regulatory regions. In addition to this, in some breast and oesophageal cancers the loss of function of FHIT is associated with translocation in the FRA3B region. In some of these translocation instances the homologous chromosome had then lost its functional copy of *FHIT* via some other variable mutation (Popovici *et al*., 2002; Fang *et al*., 2001). Ultimately the sequence of events that result in loss of function of the *FHIT* gene are complex, and involve interplay between the CFS FRA3Bsbreakage and repair, and other possible mutations and translocations that can lead to loss of the *FHIT* gene and its function.

WW domain-containing oxidoreductase (*WWOX)* (also identified as *FOR*1 [Ried *et al*., 2000] and *WOX*1 [Chang *et a*l., 2001] is another tumour suppressor gene associated with a CFS. Like *FHIT*, *WWOX* is a large (~1 MB) gene which transcribes to a relatively small 1.2kb product (Durkin and Glover 2007; Sozzi *et al*., 1996) and like *FHIT* it is associated with a large and frequently broken CFS FRA16D. LOH due to breakage or translocation at FRA16D has been associated with breast, prostate, hepatocellular and ovarian cancer (Krummel *et al*., 2000; Ried *et al*., 2000; Finnis *et al*., 2005). The tissues which are most often associated with *WWOX* expression (mostly tissues in sexual and endocrine organs) appear to be the tissues in which loss of *WWOX* is associated with tumorigenesis (Nunez and Ludes-Meyers, 2006), highlighting the important role epigenetics plays in determining the path for tumorigenesis, and the risk factors associated with it. In addition to this, emerging research on the role the *WWOX* gene plays *in vivo* indicate it may have links to the tumour supressorp53, and possibly even to tau phosphorylation best known for its role in Alzheimer's disease (Avila *et al.,* 2003). A study of cancer cell lines found that regions that were broken in these lines *in vivo* were the same as regions that are frequently broken in *in vitro* studies of FRA16D, indicating a particularly strong link between this fragile site and tumorigenesis (Ried *et al*., 2000).

The evolving 'life-history' of the CFS genomic region during tumorigenesis is not identical between all CFS. For example, the continual fragility of FRA3B throughout cancer development appears to be greater than that of FRA16D, which seems to mutate to loss of function early and then remain relatively unaltered throughout cancer development. Cancer cell lines with FRA3B deletions have an 80% chance of having

FRA16D deletions, while cancer cell lines with FRA16D deletions only have a 40% chance of having FRA3B deletions (Finnis *et al*., 2005). This indicated that there is still much to be understood about the subsequent increase in fragility caused by loss of function of the tumour suppressor genes at CFSs and the complex interplay between fragility, the specific functional expression of different tumour suppressor genes and CFS associated with other tumour suppressor genes.

Like *FHIT* and *WWOX* the *PARK2* gene is extremely large (~1.53Mb), and like *FHIT* and *WWOX*, *PARK2* has been implicated in possible development of cancer (Denison *et al*., 2003). *PARK*2 contains the CFS FRA6E and breaks at this fragile sites have been linked to both Parkinson's disease and ovarian cancer (Denison *et al*., 2003; Coelln *et al.,* 2004). Parkinson's disease is a progressive neural disorder characterized by tremors, impaired balance, loss of automatic movements and other compromised motor control issues (mayoclinic.org/diseases-conditions/parkinsons-disease/basics/definitions/con-200028488). Autosomal recessive juvenile Parkinson's is an early onset form of the disease which has been linked to the gene *PARK2* on chromosome IV. This elucidates further the importance of fragile sites in development of cancer and other human disorders.

While FRA3B and FRA16D are the most studied and best characterized sites of tumour suppressor gene/CFS interaction, they are not the only such sites. FRA7G has been associated with the *MET* oncogene and together they have been correlated with oesophageal and ovarian cancer (Miller *et al*., 2006; Huang *et al.,* 1999). FRA9E and FRA6E, too, have been linked to ovarian cancer (Callahan *et al*., 2003). Further studies

may yet unearth more CFS associated with genes that have tumour suppressor or oncogene roles.

There has also been the identification of a possible mechanism for oncogene amplification dubbed the breakage-fusion-bridge cycle. This mechanism usually result from double strand breaks, such as those caused by CFS, at the border of the to-be-amplified oncogene (Ozeri-Galai *et al*., 2012; Conquelle *et al*., 1997). Recently *in viv*o evidence has surfaced implicating CFS sequences as hotspots for the breakage-fusion-bridge cycle (Ozeri-Galai *et al*., 2012).

The connection between CFS and cancer is an unfortunately convoluted affair. Simple correlations between CFS and tumour suppressor genes are insufficient to proclaim CFS as the cause of cancer and a recently described potential epigenetic component contributing to CFS breaks has added an additional layer of complexity to this field of research (Debatisse *et al*., 2011). The finding that different CFS sequences break at different rates in different cells does, however, help explain some otherwise findings, such as the facts that FRA3B and mutations at *FHIT* are not involved in renal cancer development but are in breast cancer (Ingvarsson, 2001; Bugert *et al*., 1997)

It is important to note that the correlation between CFS and cancer could be a statistical anomaly. There is a known shared genomic region in many CFS and tumour suppressor genes, and as the evolution of cancer favours breaks in these tumour suppressor genes the correlation with CFS at the same sites may simply be one of happenstance. Research to eliminate this concern has indicated that breaks at CFS are drivers, not passengers, in the development of cancerous cells (Bignell *et al*., 2010). The fact that CFS break even when not associated with tumour suppressor genes also

indicated that CFS instability drives cancer development (Arlt *et al*., 2002; Bignell *et al*., 2010; Ingvarsson, 2001).

Finally, human papilloma virus (HPV) is a DNA virus which is a causative agent of 70 and 80% of cervical and anal cancer cases, respectively. Of the different viruses that make up the HPV family, HPV16 is a particularly dangerous form, extremely prevalent in its contribution to the statistics above (cancer.gov/cancertopics/factsheet/Risk/HPV). Integration of HPV16 is preferential in CFS including FRA3B and FRA16C (Wilke *et al*., 1996). In addition to these fragile sites providing places for the HPV virus to integrate, the HPV viral integration appears to actively induce cellular replication stress. The result of this is increased breaks at CFS in cells where the HPV virus has integrated into the genome (Bester *et al.,* 2011).

Copy number variants. Copy number Variants (CNVs) are relatively large (>50bp) duplications or deletions of regions within the human genome. Despite their size relative to other smaller repeateding genomic seuqnces (ie: small tandem repeats) they are often a natural component of the human genome, contributing to genetic and phenotypic diversity (Arlt *et al*., 2011; Girirajan *et al*., 2010). CNVs are correlated with a number of disorders, including autism, schizophrenia and childhood cognitive development issues (Eddwin and Scherer 2008; Morrow 2010). CNVs are not well understood, but it has been speculated they may provide a novel mechanism for explaining the broad spectrum encompassed within their correlated disorders, because variability in copy number and the differences in precise location of endpoints of the variable region enables explanation for the variable expressivity of these genomic disorders (Girirajan *et al*., 2010).

CNVs have been recently broken into two broad classes, recurrent and non-recurrent. Recurrent CNVs are characterized by low copy repeats (between two and five) and duplications at large segmental duplication regions. They are meiotic in origin and most of the mechanisms proposed to result in recurrent CNV changes are associated with non-homologous meiotic recombination and misalignment of repeated/ duplicated sequences (Arlt *et al.,* 2010). Non-recurrent CNVs on the other hand appear to be mitotic in origin, they are characterized by unique and inconsistent breakpoints, complex rearrangements similar to chromothripsis a massive chromosome rearrangement also known as chromosome shattering, and are also associated with neurological disorders. The proposed mechanisms behind non-recurrent CNV alteration includes homologous repair, mitotic recombination and breakage-fusion-bridge cycles (Arlt *et al*., 2010).

The relationship between CNVs and CFS became apparent with the discovery that treatment of cells with aphidicolin resulted in an increase in altered CNVs (Arlt *et al*., 2009). Further studies indicated that hydroxyurea (another compound used to induce breaks at fragile sites) also induced alterations in CNVs and as such provided further evidence for a link between CNVs and CFS. Aphidicolin and Hydroxyurea are both inhibitors of DNA replication which slow replication rates. CNVs were found to be more likely to be altered in the FRA16D CFS than in the rest of the genome(Arlt *et al*., 2011). The similarity between *de novo* non-recurrent CNVs and those induced by hydroxyurea and aphidicolin indicate that CFS play a role in CNV duplications and deletions. The study of the connection between CNVs and CFSs is in its infancy, but holds promise for increasing our understanding of how CFSs play yet another role in genetic variability and pathogenesis.

Table 1: Common Fragile Sites and Associations

CFS	Associated with
FRA3B	*FHIT* and HPV Integration
FRA16C	HPV Integration
FRA16D	*WWOX*
FRA6E	*PARK2*
FRA7G	*MET*

The Mechanisms of Common Fragile Site Instability

The mechanisms that are associated with the breakage and repair of CFS sequences largely decide the extent and severity of DNA damage done by the fragile nature of these regions. Characterising these breakage and repair mechanisms may allows us to understand the cause behind the loss of heterozygosity and other genetic mutations that lead to many of the disorders explained above. There are five main proposed explanations for why CFS break, but recent consensus is on a molecular cause that combines multiple of these. Common to each of these hypotheses is the understanding that CFS are late replicating, often carrying on their replication into G2 (Handt *et al*., 2000; Le Beau *et al*., 1998).

Formation of secondary structures hypothesis. The formation of secondary structures has long been a hypothesis for explaining why CFS are prone to breaking under conditions of replication stress. This hypothesis contains four main components

which theoretically interact to culminate in a double strand break. First is a slow replication rate, resulting in a lagging polymerase complex. As a result of this a 'run away' helicase is able to proceed a disproportionate distance ahead of the polymerase complex, creating a large region of single stranded DNA (ssDNA) on the leading strand. Next, the DNA in this single-stranded area bends back on itself and a region of sequence that can self-pair (such as a series of AT repeats) creates a secondary structure, such as a hairpin (Arlt and Glover, 2010). Finally, DNA breakage occurs either directly as a result

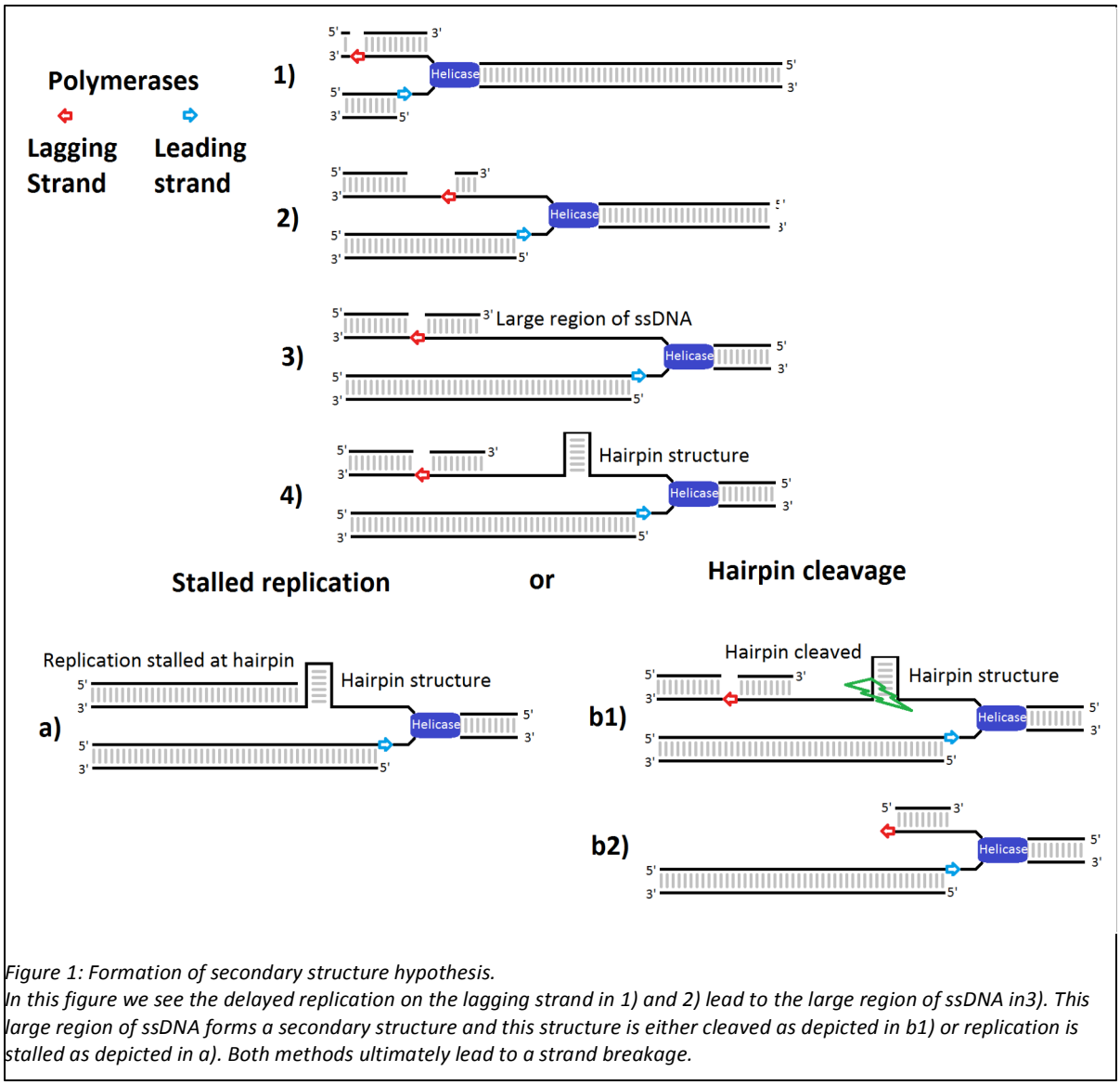

Figure 1: Formation of secondary structure hypothesis.
In this figure we see the delayed replication on the lagging strand in 1) and 2) lead to the large region of ssDNA in3). This large region of ssDNA forms a secondary structure and this structure is either cleaved as depicted in b1) or replication is stalled as depicted in a). Both methods ultimately lead to a strand breakage.

of cleavage at the site of the secondary structure (Durkin and Glover 2007), or as a result of the secondary structure stalling replication sufficiently to ensure it is not completed before cell division, and anaphase bridge breakage may occur (Chan *et al*., 2009). This process is depicted in figure 1.

As an extension of this hypothesis, the characterization of "flexibility peaks" as regions with abnormal flexibility (4.5 standard deviations above the mean flexibility when calculated by the flexstab program) in DNA twisting has opened up a new possible explanation for CFS breaks (Mishmar *et al*., 1998). The prevalence of AT stacked nucleotides in regions of DNA increases the twist flexibility of DNA at that region, and that flexibility coupled with the potential for self-pairing in AT dinucleotide runs could be correlated with hotspots for double strand breaks in CFS (Durkin and Glover 2007, Glover *et al.,* 2005). However recent testing has indicated that the presence of breakage 'hot spots' is not at flexibility peaks (Casper *et al*., 2012). Ultimately the overall evidence for flexibility peaks as the cause of CFS breaks is weak, but the formation of secondary structures is an important hypothesis which is probably at least partially responsible for CFS breaks.

Replication origin paucity hypothesis. The discovery that origins are scarce in fragile sites led to the hypothesis that the paucity of origins is a causative agent in why CFSs break (Letessier *et al*., 2011). Further evidence for this hypothesis has been provided by studies indicating that additional origins of replication are activated in normal genomic regions when a cell is under stress. However at CFS all origins are activated, even under unstressed condition. As a result of the inability of CFS to activate more origins of replication during periods of replication stress, these regions of DNA

may not be fully replicated before the cell enters G2. This hypothesis proposes that this is the reason that CFS are prone to breakage, specifically under conditions of replication stress (Ozeri-Galai *et al.,* 2011)

As the activation of origins of replication is cell specific. As such it follows that any genetic instability caused by replication origin paucity would be different in different cell types. Recent studies have shown that this is, indeed, the case as cells which utalize more origins of replication in specific regions experience fewer breaks in that region(Metsu *et a*l., 2014; Letessier *et al*., 2011; Ozeri-Galai *et al*., 2012). This also helps explain the epigenetic nature of CFS.

Replication and transcription machinery collision hypothesis. The fact that CFSs are found preferentially at very large genes (>600kb) and an even stronger connection to genes of at least 300kb (only 3.4% of genes are this long) led to the formulation of the hypothesis that breaks at CFS result from collisions between transcription and replication machinery (Le Tallec *et al*., 2013). This is particularly relevant as it has also been found that all 7 large genes which contain recurrent cancer deletions are also associated with CFS (Le Tallec *et al*., 2013). This hypothesis has been supported by evidence showing that CFSs that are associated with large genes break more frequently when that gene is being transcribed (Helmrich *et al*., 2011). Direct evidence of this hypothesis however is scarce, with the only study looking specifically at CFS breakages while transcription of these large, CFS associated, genes was occurring finding that CFS were not less prone to breakage while their associated gene was undergoing transcription (Le Tallec *et al.,* 2013).

Chromatin structure hypothesis. The condensation of chromatin plays an important role in epigenetic control and recent research indicates that it may also be linked to the breakage of DNA at CFS. Histone acetylation is noticeably less at CFS than at surrounding DNA regions, indicating that the DNAt is more compact and perhaps harder to reach for replication machinery and therefore harder and slower to replicate (Jiang *et al*., 2009; Ozeri-Galai *et al*., 2012).

The clustering of nucleosomes at the fragile site FRA3B also supports the hypothesis that chromatin is linked to the mechanism behind CFS breakage, however results *in vivo* and *in vitro* are inconsistent and indicate that further factors are influential in this relationship (Ozeri-Galai *et al*., 2012).

Replication border hypothesis. The fragile site FRA10B was first found to be located at a boundary between early and late replication zones (Handt *et al*., 2000). Detailed analysis of replication timing has led to the discovery that the CFS FRA6E is also at the border of an early and late replicating region (Palumbo *et al*., 2010). This hypothesis states that the early replication region completes its replication at the border of this zone, with the replication fork waiting for the adjacent late replication region collapsing before replication in the adjacent region is complete. Alternatively the adjacent region simply does not complete replication in time and the break of the fragile site occurs during condensation of the DNA.

It is important to note that none of the above mechanisms are mutually exclusive, and that current research indicated overlap between causative mechanisms (Durkin *et al.* 2007). Research to determine the mechanistic cause of CFS breakage has also been conducted most commonly in two very different systems, human cell cultures

(traditionally lymphocyte but more recently other cell types such as fibroblasts), and in the yeast *S. cerevisiae*.

Research on CFS are most commonly completed using two distinct approaches. The study of CFS in their natural context within the human cells allows for the most 'relevant' method of research, with results that are more applicable than research in model organisms. In addition to this human cell research allows for researchers to study the CFS 'in context' with all adjacent and influencing promoters, origins of replication and other surrounding genetic and molecular components in position. While there are obvious disadvantages of taking the study of CFS out of their natural human context, there are also some advantages. Research in *S. cerevisiae* allows for CFS regions to be taken out of context, and to fundamentally separate structural and context based mechanisms of CFS breaks. This is able to be achieved to ways. First the insertion of a human CFS into yeast cells allows for research to be conducted on these regions out of context. This means we can test hypotheses such as the secondary structure, or flexibility peak hypothesis, without having to account for origin paucity or spread. Second the study of yeast fragile sites, which are naturally occurring sites similar in nature but not sequence to human fragile sites, allows for research to be conducted on the yeast native repair mechanisms such as homologous recombination, and the subsequent consequences for genetic mutation due to these repair pathways.

Repair of Common Fragile Sites after Breakage

In order to avoid the loss of tumour suppressor genes or conversion of proto-oncogenes to oncogenes events discussed above, the cell has a number of ways to deal with breakages at CFS. These repair mechanisms help ensure the cell does not

undergo tumorigenesis and ultimately conserve genetic integrity. Unfortunately the repair mechanisms themselves are not perfect, and can often lead to loss of heterozygosity (LOH) which can be highly detrimental if it occurs in genes such as *FHIT* and *WWOX*. Understanding the repair mechanisms used by cells allows us to gain a better insight into the genetic alterations that are occurring due to CFS breakages and how these can lead to the loss of essential genetic information via LOH. LOH occurs to different degrees for each of the different repair mechanisms, and as such an individual understanding of the details of each repair pathway is desirable. Some mechanisms (such as break induced repair and reciprocal crossover) are better researched, characterised and understood than others, so it is vital we continue to research alternate repair pathways such as gene conversion in order to fully understand the events that lead to tumorigenesis.

Homologous recombination. Homologous recombination is a mechanism for repair of lost or damaged genomic regions using the homologous chromosome. There are a number of forms of homologous recombination used in these circumstances. While homologous recombination is a mechanism capable of ensuring loss of important genetic information does not occur, it does have its drawbacks. As the sequence being used to replicate off during this method of repair is on the homologous chromosome, the alleles and genes on this chromosome are duplicated. Depending on the method of homologous recombination used, this can lead to LOH at specific loci repaired by homologous recombination in the region surrounding the initial break (Rosen *et al.* 2013).

The various forms of homologous recombination will be described in detail below, and figure 2 gives a quick guide to reading these homologous recombination figures. All homologous recombination methods begin with the initial

break and a partial resection of DNA in the surrounding region to expose a ssDNA segment with a free 3' end that is capable of searching for homology on a homologous chromosome (Figure 3). It's important to note that resection as depicted in figure 3 can be quite extensive and makes prediction of the initial point of breakage quite difficult to ascertain.

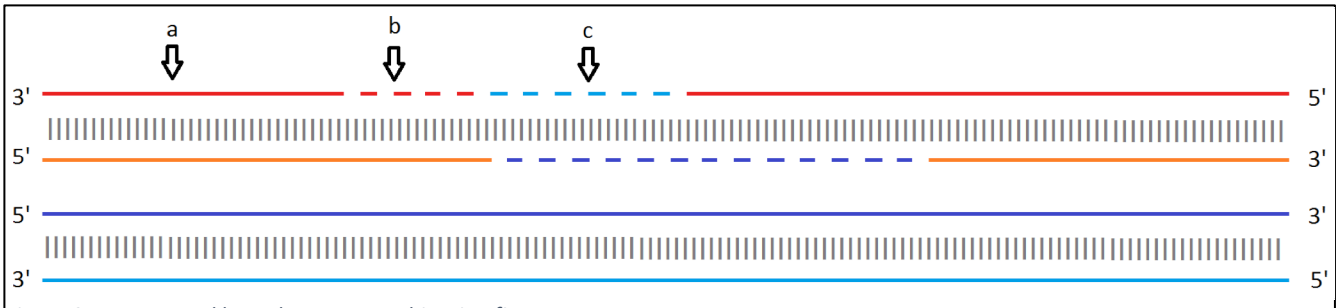

Figure 2: How to read homologous recombination figures
A pair of homologous chromosomes is shown. The homolog depicted with red and orange strands is the homolog that experiences a break. The homolog depicted with blue and purple strands is the homolog that does not experience damage and acts as a template for repair by homologous recombination. Hydrogen bonds are depicted by short grey vertical lines. a) A 'natural' segment of DNA; b) A segment of DNA synthesised after DNA was damaged, but still containing its 'natural' (red) sequence; c) A segment of DNA synthesised on the strand equal to that of 'a)' but repaired off a homolog and now containing the sequence indicated by the colour of the dashed line (light blue).

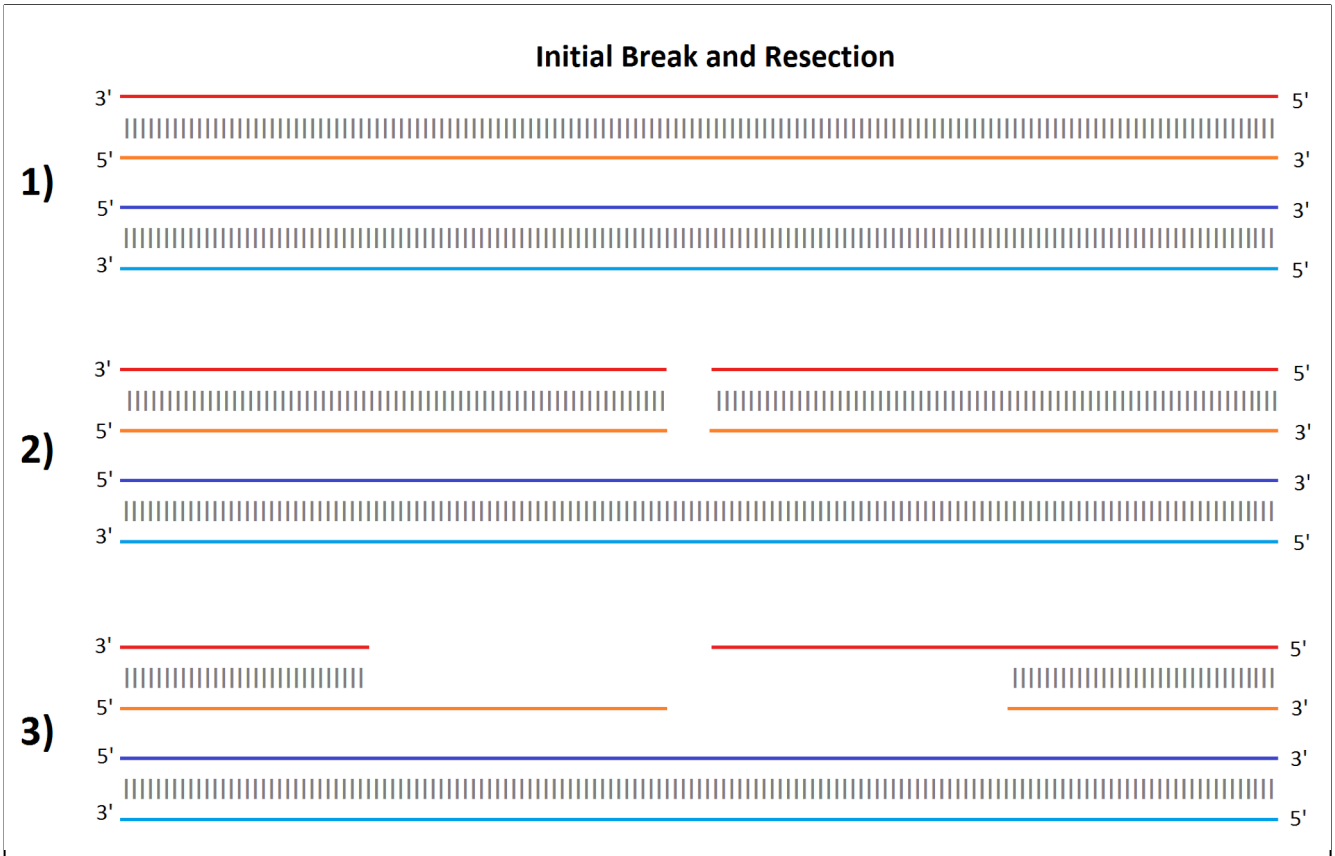

Figure 3: Initial stages common to all homologous recombination methods of repair
This figure shows the three stages common to all homologous recombination events. In 1) we see two homologous chromosomes aligned adjacent each other. In 2) we see a double strand break in the upper chromosome, while the lower chromosome remains intact. In 3) we see resection of DNA around the initial break site to expose a single stranded region of DNA with a 3' end capable of invading the homologous chromosome.

Break induced repair (BIR). Like all mechanisms of homologous recombination, the BIR mechanism begins with resection of DNA from the breakpoint to create single stranded overhangs of DNA, with free 3' ends. The 3' end itself is also frequently resected to some degree. As shown in figure 4 the 3' end of one of these ssDNA regions invades the homologous chromosome and finds a region of homology. DNA replication is initiated at the invading 3' end and replication continues in the 5' to 3' direction. Lagging strand DNA synthesis is initiated on the newly formed strand to complete the process of replication. Thus, BIR results in loss of heterozygosity at all points centromere distal to point of resection from the break. BIR is the only method of homologous

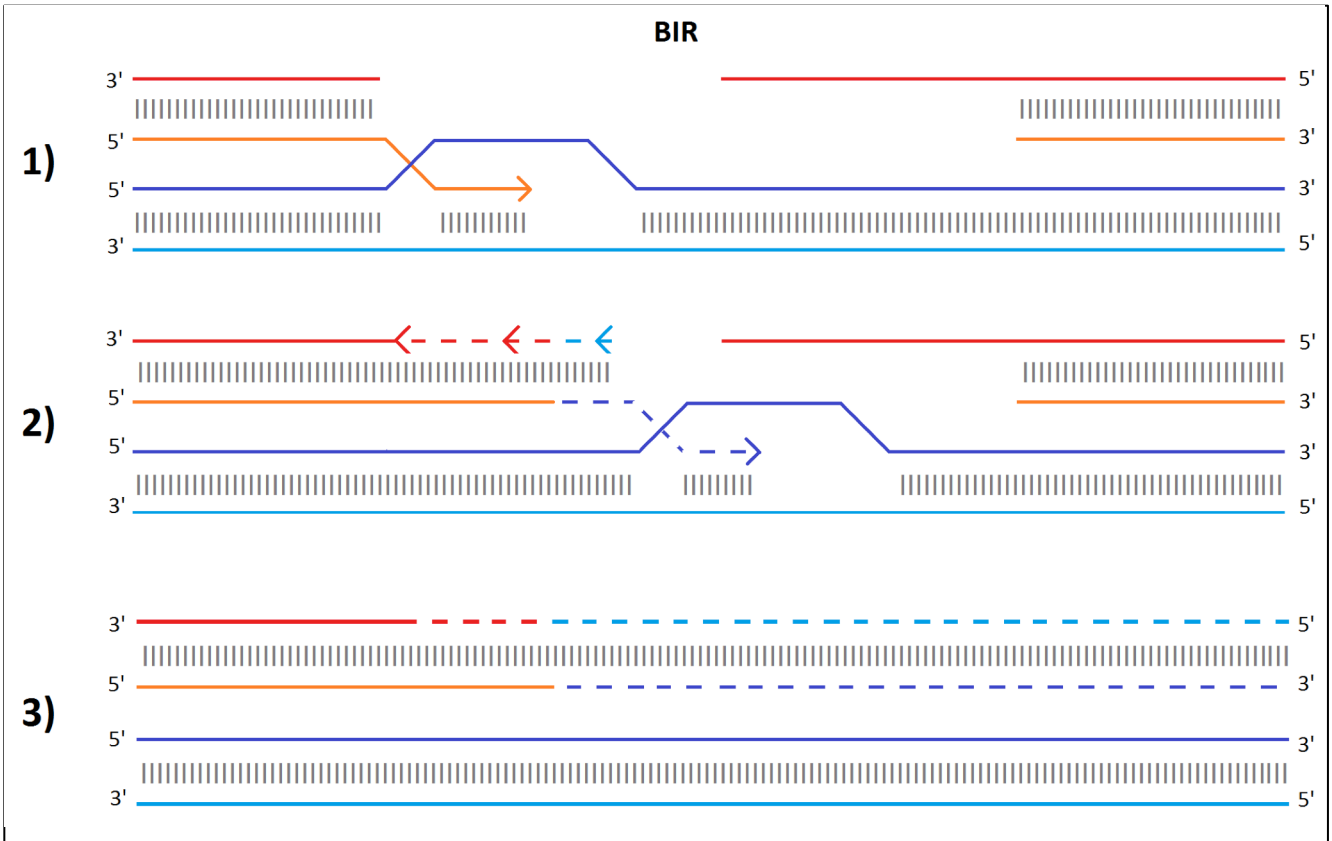

Figure 4: Break induced repair mechanism of homologous recombination
This mechanism involves an initial invasion of the resected 3' end of the broken strain into its homolog, creating a D-loop structure as seen in 1). The invading strand proceeds to replicate, progressing with its D-loop as seen in 2). The lagging strand repairs using the recently repaired (blue) leading strand as a template as seen in 2). Completion of replication involves loss of the region of DNA centromere distal to the initial break point as seen in 3).This method of repair ultimately results in a LOH for all points centromere distal to the initial break point.

recombination that does not require "capture" of the other broken end (the centromere-distal portion of the break) (Mitchel *et al*., 2010; Lee *et al*., 2009)

Double break induced repair (dBIR). The recent observation of a class of extremely large gene conversions (long-tract gene conversions) has led to the hypothesis of a double BIR mechanism, in which the initial invading strand stops replicating the homologous chromosome and then proceeds to invade its initial chromosome and initiate replication of this chromosome as a template. This mechanism does require capture of the centromere distal portion of the chromosome (Llorente *et al*., 2008) and is depicted in figure 5.

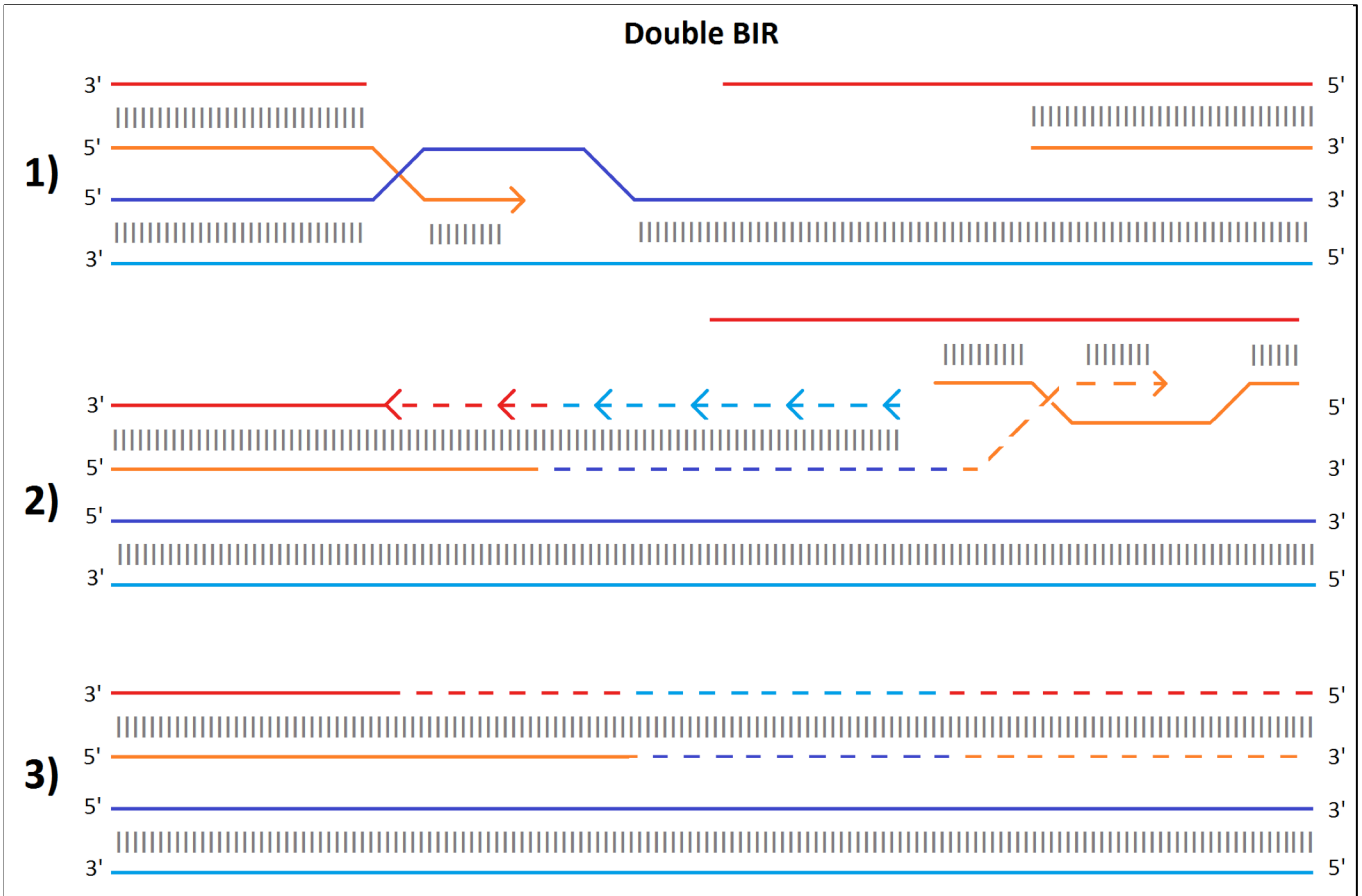

Figure 5: Double break induced repair mechanism of homologous recombination
Repair by double BIR involves and initial segment of replication identical to the method used for a regular BIR as seen in 1). The template is then switched however and the invading strand exits its homologous chromosome to replicate off its original template as seen in 2). After replication is complete the ultimate LOH is isolated to the region where replication was occurring off the homologous chromosome as in 1) before switching back to the initially broken chromosome in 2).

Double Holliday junction (dHJ). The formation of a dHJ results from the invasion of a strand, as in BIR, followed by the annealing of the ssDNA from the resection of the centromere distal portion of the chromosome break to the other strand of the homologous chromosome. In this method (Figure 6) two Holliday junctions are formed, which can be resolved in a number of ways, by either cleavage or dissolution of the Holliday junctions. This method of repair is more likely to be utilized if a long initial tract of heteroduplex DNA is formed upon strand invasion, creating large D-loop structures that are more stable and less likely to collapse, ejecting the invading strand

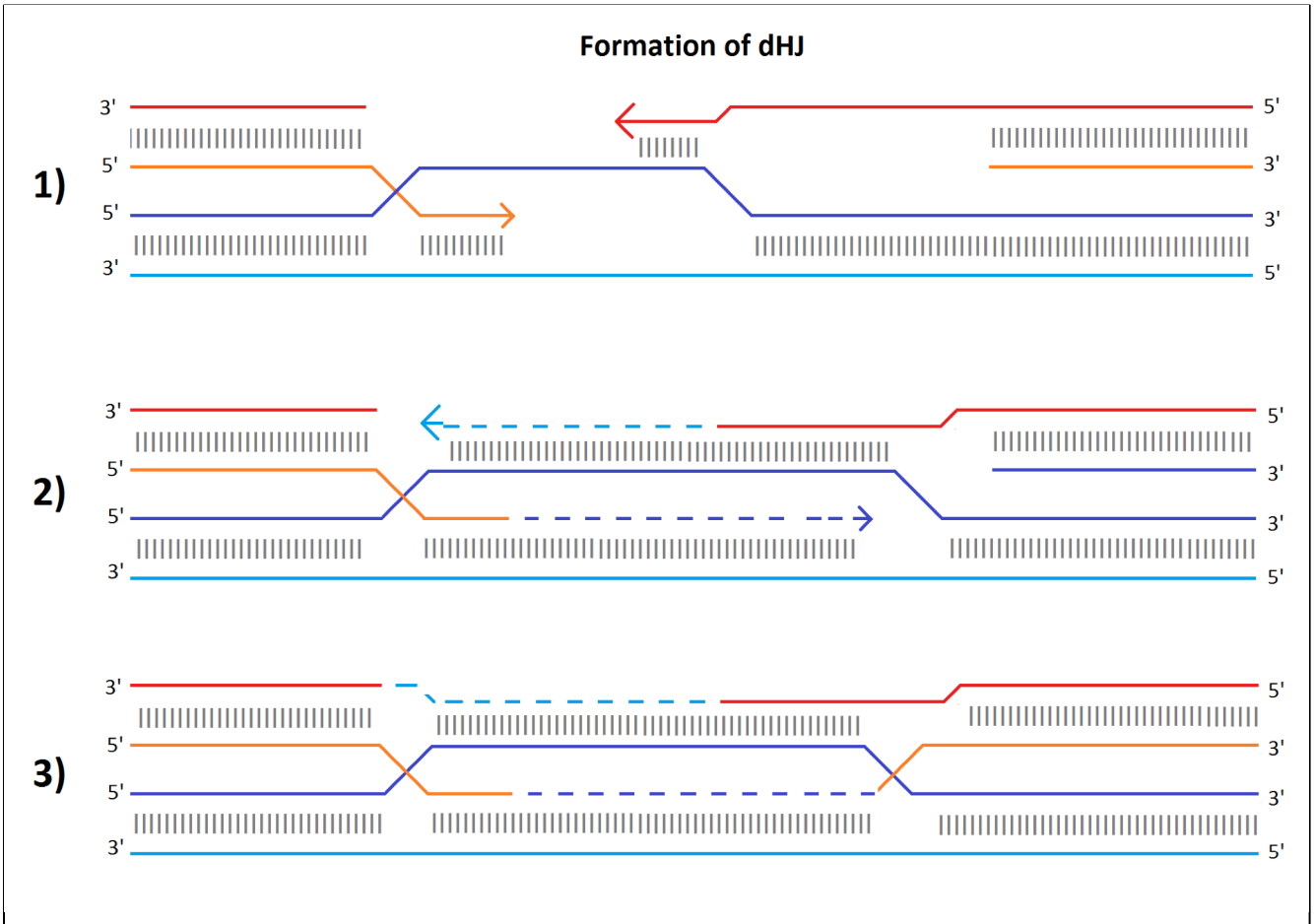

Figure 6: Formation of the double Holliday junction prior to resolution or dissolution
In this method of homologous recombination the invasion of both the centromere proximal and the centromere distal resection portions of DNA occur as can we seen in 1)s. A section of DNA is replicate by both sidesd, repairing the initial damage as can be seen in 2). The invading strand is ejected from the homologous chromosome, re-joining with its initial chromosome segment and this forms a double Holiday Junction as seen in 3), which can be resolved in the methods shown in figure 6.

(Mitchel *et al*., 2010).

As shown in the top panel of figure 7, the resolution of both Holliday Junctions by nuclease cleavage in the same plane results in a non-crossover, with a region of gene conversion in the sequence between the two resolved Holliday junctions. Resolution of the two junctions by nuclease cleavage in different planes (Figure 7, lower panels) results

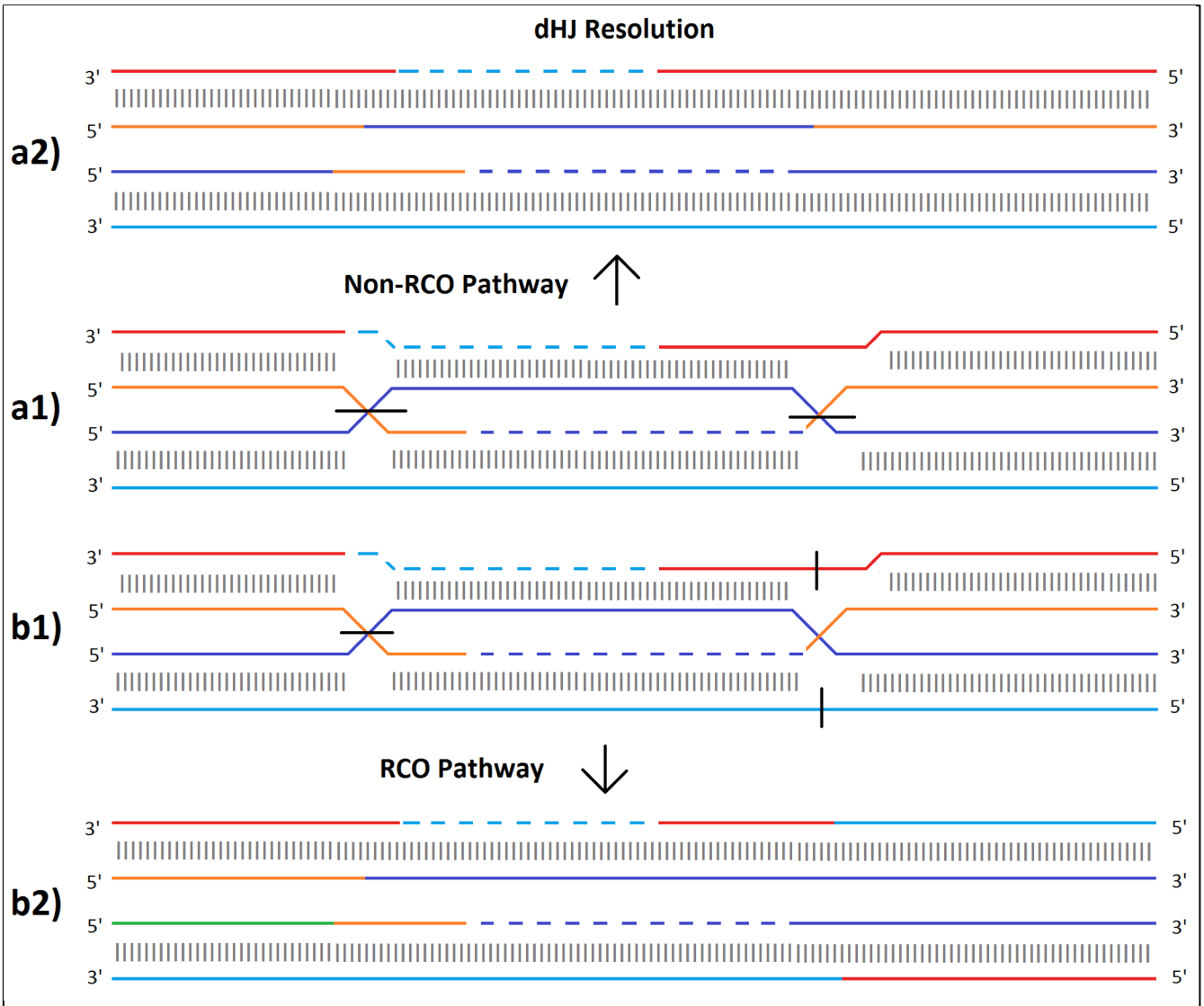

Figure 7: Resolution of a double Holliday junction can follow RCO and a Non-RCO pathways
The two possible pathways for resolution of a dHJ result from different cleavage pathways. The Non-RCO Pathway results from cleavage of each Holiday Junction in the same direction as can be seen in a1).The subsequent rearrangement of DNA sequence can be seen in a2) and ultimately results in a patch gene conversion. The RCO Pathway results from the cleavage of each Holiday Junction in a different direction as can be seen in b1). This results in the 'switching' of homolog sequences at all points centromere distal to the repair site leading to the chromosomal rearrangement show in b2).

in a reciprocal crossover. This crossover is also usually associated with a region of gene conversion. The size of gene conversion regions associated with double Holliday junctions can vary greatly.

Finally, dissolution of the dHJ by helicase and topoisomerase results in gene conversion between the two junctions, and does not result in reciprocal crossover (Mitchel *et al*., 2010; Yim *et al*., 2014; Lee P.S., 2009).

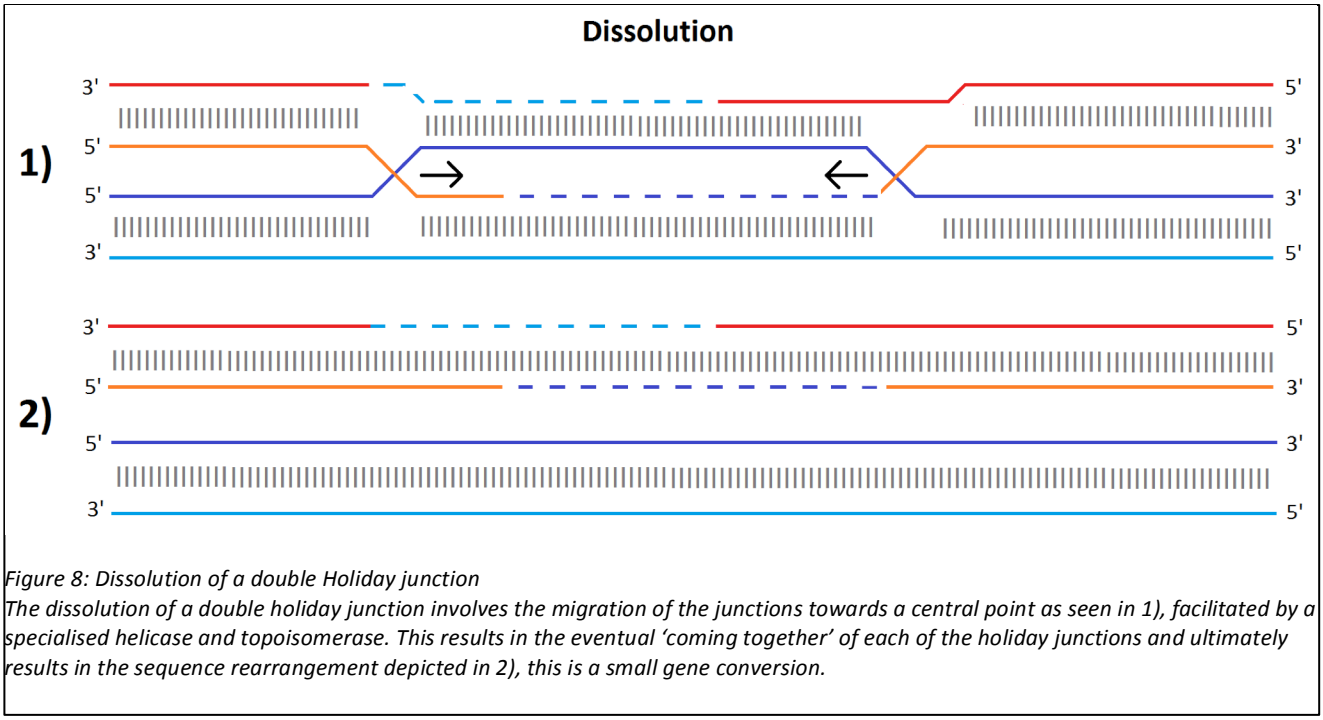

Figure 8: Dissolution of a double Holiday junction
The dissolution of a double holiday junction involves the migration of the junctions towards a central point as seen in 1), facilitated by a specialised helicase and topoisomerase. This results in the eventual 'coming together' of each of the holiday junctions and ultimately results in the sequence rearrangement depicted in 2), this is a small gene conversion.

Synthesis dependent strand annealing (SDSA). This is the typical cause of short-tract gene conversions and is the result of the invasion of a single strand of resected DNA from the break site into the homologous chromosome, followed by a period of replication and finally exit of the strand and re-joining with the original centromere-distal portion of DNA from the parent chromosome. This method, depicted in figure 9, is favoured when

the invading region of DNA is short and when the second end of the broken parent chromosome is readily available (Mitchel *et al*., 2010).

In instances where mismatched DNA is experienced as a result of different sequences in regions where an invading strand has bound to a region of homology repair on its homolog, mismatch repair consolidates that sequence to that of the chromosome

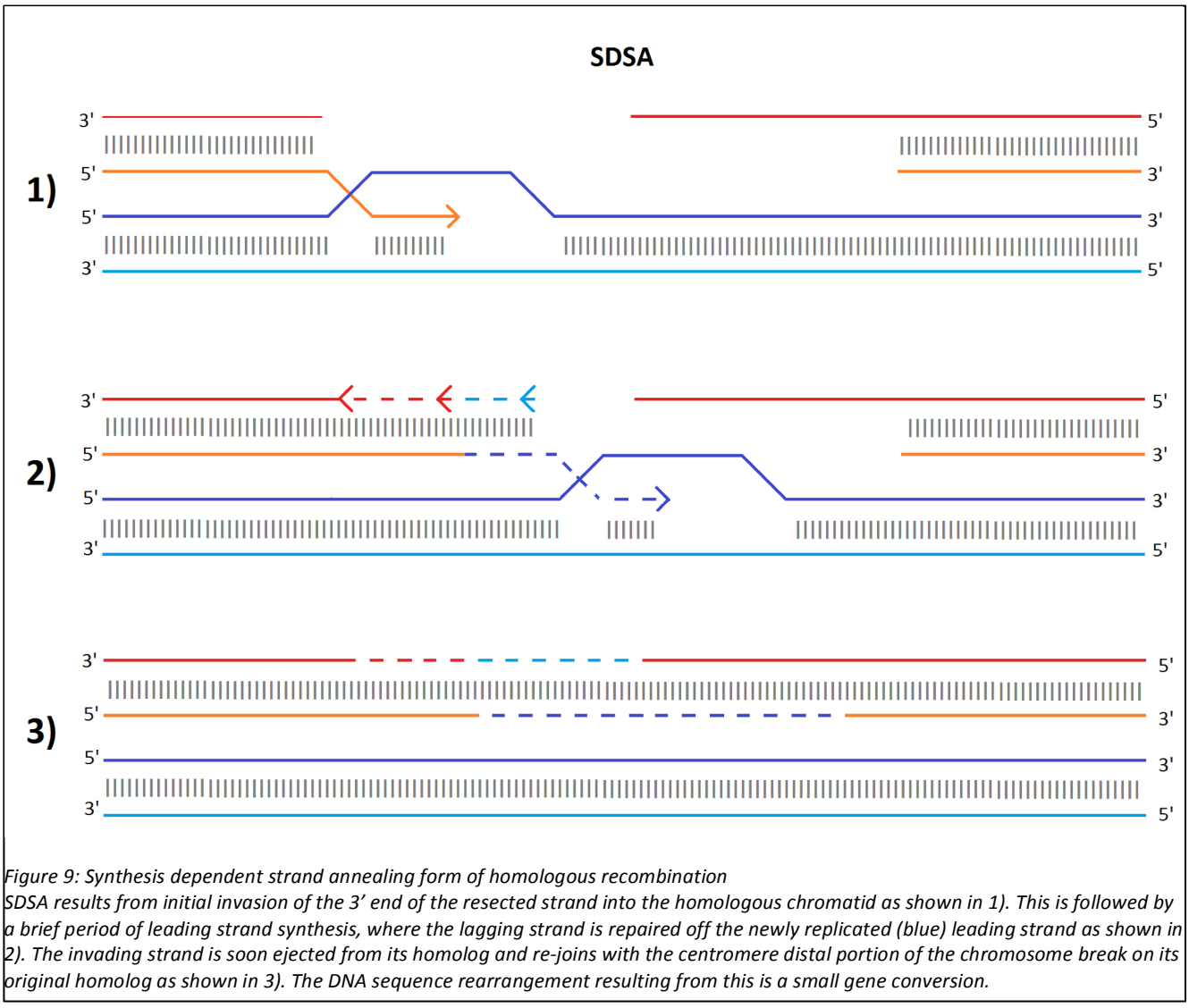

Figure 9: Synthesis dependent strand annealing form of homologous recombination
SDSA results from initial invasion of the 3' end of the resected strand into the homologous chromatid as shown in 1). This is followed by a brief period of leading strand synthesis, where the lagging strand is repaired off the newly replicated (blue) leading strand as shown in 2). The invading strand is soon ejected from its homolog and re-joins with the centromere distal portion of the chromosome break on its original homolog as shown in 3). The DNA sequence rearrangement resulting from this is a small gene conversion.

that was used, or is going to be used, as the template for repair (Lee P.S., 2009). In all of the mechanisms described above, it is important to note that the process of homologous recombination is able to proceed in either direction along the chromosome, however, it

will not cross the centromere. While the biochemical control of these repair pathways is fairly well researched it is beyond the scope of this paper, the interested reader is directed to Filipp *et al*., 2008.

Chromosome loss and telomere capping. In instances when the cell does not utilize homologous recombination as a method of repair, it may lose the damaged chromosome altogether, or simply cap the chromosome with a telomere at a point after resection from the break site (Lemoine *et al*., 2005). Unlike the homologous recombination events described above, chromosome loss and telomere capping result in LOH due to a conversion to hemizygosity, not homozygosity.

The LOH that results from the mechanisms described above is an event of particular importance in the initiation and progression of cancer. As such, it is vital to understand how these different pathways can result in LOH in daughter cells after mitosis is complete (Figure 10). BIR and telomere capping will cause LOH of any gene centromere-distal to their respective initiation points, while SDSA will only cause LOH for genes in the short gene conversion patch that results. A reciprocal crossover may or may not be associated with a gene conversion, and that gene conversion may be of varying size. RCOs are also unique in that the possible LOH depends on the independent assortment of chromosomes after repair. The ultimate result will either be two daughter cells that both have experienced LOH (one homozygous for allele of interest and one homozygous for another allele) or two daughter cells that have recombined homologous chromosomes but no LOH. This is also depicted in figure 10.

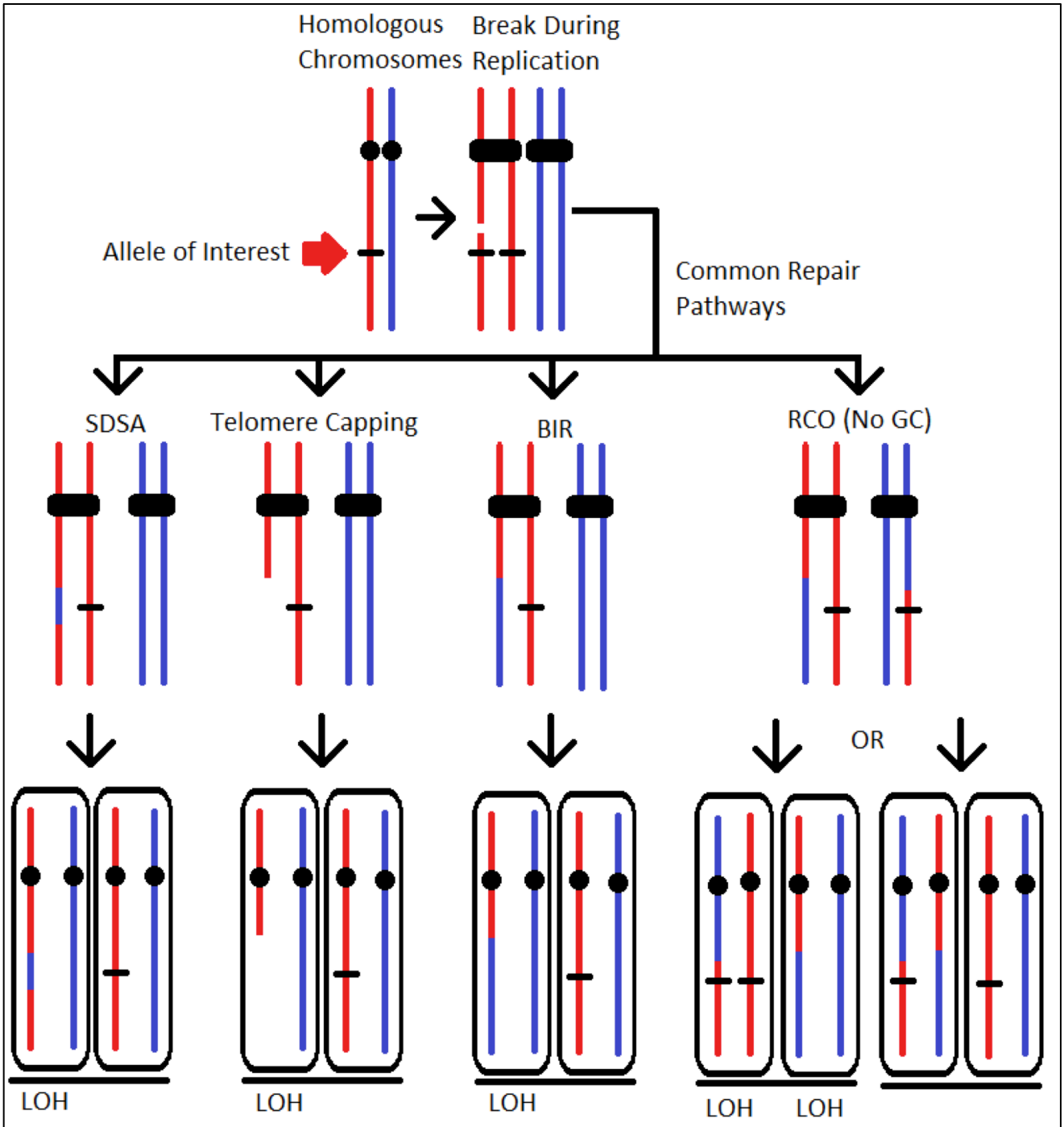

Figure 10: Methods of loss of heterozygosity
This figure represents common methods of LOH resulting from a break, and subsequent repair, at a gene of interest during mitosis. This figure depicts a single pair of homologous chromosomes going through the initial stage of breakage, followed by one of four common repair pathways, and ultimately segregating into daughter cells which contain varying degrees of LOH. SDSA results in a small patch of LOH known as a gene conversions, while the telomere capping and BIR pathway result in large regions of LOH centromere distal to the break site. Note RCO depicted has no associated gene conversion and the LOH that results from this repair pathway depends largely on the subsequent independent assortment of its chromosomes during replication.

The characterization of these repair mechanisms is an ongoing process, with a limited amount known about the specifics of a break that result in a cell utilizing one repair mechanism over another, or how the type of initiating lesion in DNA affects the frequency of each type of repair mechanism. The goal of our experiment is to observe the frequency of the different forms of repair, specifically those that result in LOH. Specifically we hope to find and analyse gene conversions that are not associated with reciprocal crossover, as these are the least studied and least well understood of the repair pathways above.

The Fragile Site FS2 provides a unique system for the study of repair systems in the yeast strain *S. cerevisiae* (Lemoine *et al*., 2005). Previous research by the Casper Lab utilized *S. cerevisiae* to study repair mechanisms in response to lesions at the naturally-occurring fragile site FS2 on yeast chromosome III. This research was able to identify hotspots for the initiation of BIR events and RCO events at this fragile site, but was unable to detect gene conversions at the fragile site (Rosen *et al.* 2013). As an extension of this research we have altered the yeast strain to enable detection of small gene conversion events at fragile site FS2 and ultimately gain greater insight in their role in mitotic recombination and LOH. We hypothesised that gene conversion events will be present at a higher frequency at FS2 than surrounding regions of chromosome III and that such events will result in the LOH of our FS2 adjacent marker gene. a

Materials and Methods

Method Overview

Identification of gene conversion, BIR and RCO LOH events required engineering of a novel yeast strain. The strain we designed is a diploid and contains five key elements that are vital to our experiment. It is related to the system first used by Lee *et al.* (2009). This system was later adapted to chromosome III and utilized in a study of homologous repair mechanisms by Rosen *et al.* (2013). Our strain has a direct linkage to the Rosen *et al.* strain. The Rosen *et al.* strain contained a *SUP4-o* marker gene highly centromere distal to its FS2 fragile site, as is shown in figure 11. In an attempt to capture and analyse localised gene conversion evens we have moved the SUP4-o position up from 159kb distal from the centromere to only 55kb from the centromere. Our *SUP4-o* marker is now only 2kb from FS2, as opposed to 106kb in Rosen *et al* (2013). \

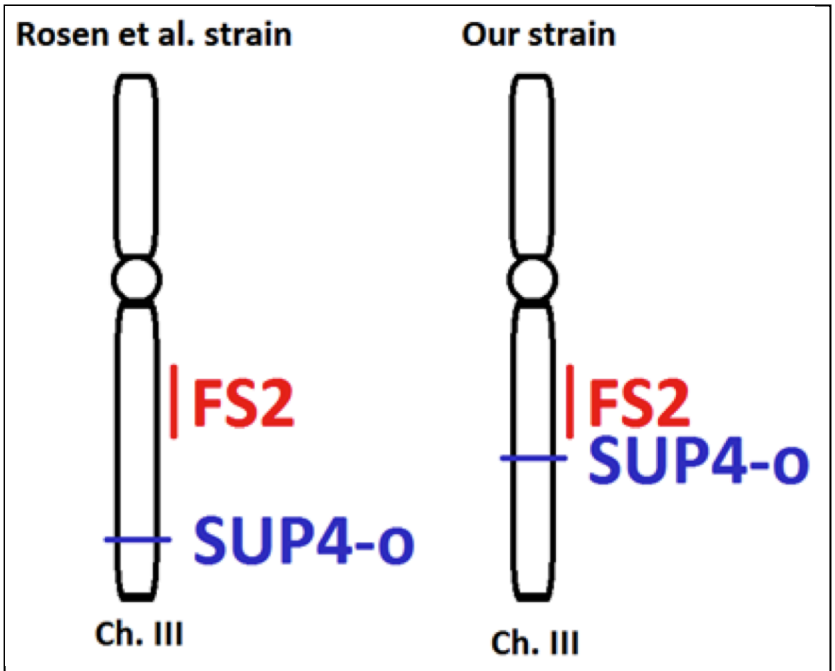

Figure 11: Comparison of our strain to that used by Rosen et al.
As can be seen, on the depicted homolog of chromosome III the position of the SUP4-o gene has been moved to be much closer to the FS2 fragile site region. This allows for the loss of SUP4-o by localized gene conversions to be detected.

First our strain is homozygous for a mutated *ade2-1* gene. This mutation results in a dysfunctional enzyme in the inosine monophosphate (IMP) synthesis pathway, resulting in the build-up of the red intermediate pigmented molecule 5'-phosphoribosyl-5-aminoimidazole (Escobar-Henriques and Diagnan-Fornier, 2001). The use of the pathway is controlled through altering concentrations of adenine; the less adenine present in the medium, the more a cell will employ the pathway (yeastgenome.org/cgi-bin/locus.fpl?dbid=S000005654). The specific mutation in the *ade2-1* gene is the presence of an ochre stop codon in codon 64, this results in a nonsense mutation, which in turn results in a truncated protein (wiki.yeastgenome.org/index.php/commonly_used_auxotrophic_markers). This relationship is depicted in figure 12.

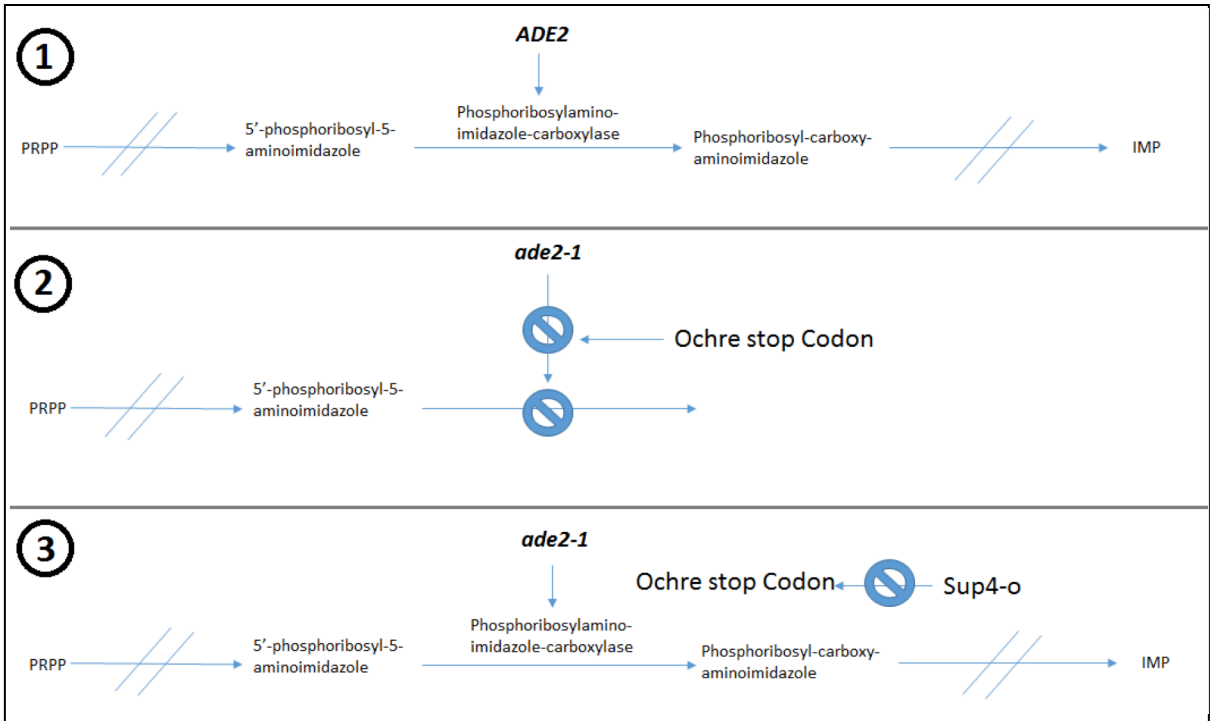

Figure 12: SUP4-o, ade2-1 and the IMP synthesis pathway
This figure shows the interplay between the IMP synthesis pathway, the ade2-1 mutated gene our strain contains and the SUP4-o mutation found on one of our homologs. Diagram 1 depicts the wild type process, with ADE2 synthesizing Phosphoribosylamino-imidazole-carboxylase, the enzyme used to catalyze 5'-phosphoribosyl-5-aminoimidazole to Phosphoribosyl-carboxy-aminoimidazole in the pathway to IMP. Diagram 2 depicts the ochre stop codon in ade2-1 preventing the production of the Phosphoribosylamino-imidazole-carboxylase required, and subsequently leading to a build-up of the red intermediate 5'-phosphoribosyl-5-aminoimidazole. Diagram 3 shows the SUP4-o mutation enabling read-through of the Ochre stop codon in ade2-1, functionally reversing the mutation and re-enabling the production of Phosphoribosylamino-imidazole-carboxylase and subsequent restoration of a functional IMP synthesis pathway.

Second, our strain contains the fragile site FS2 on a single homolog of chromosome III; this allowed us to study events occurring specifically at this fragile site (Lemoine *et al.,* 2005). Third, our strain contains one copy of the *SUP4*-o gene inserted on a single homolog of chromosome III at a location close, but centromere distal by 2kb, to the FS2 fragile site on chromosome III. *SUP4*-o is a tRNA gene that allows for "reading-through" ochre stop codons, functionally reversing the mutation caused by the homozygous ade2-1 (Figure 12) and allowing cells to grow up a pale white colour (Lee *et al.,* 2009). Loss of this *SUP4*-o gene results in a loss of this functional mutation reversal and a subsequent accumulation of red pigment and coloration in the yeast colony containing the *ade2-1*

mutation. Yeast colonies with cells homozygous for the *SUP4-o* gene will grow up white, those containing cells that are heterozygous will grow up pink, and those containing cells that lack the *SUP4*-o gene will grow up red. Fourth, our strain contains the *GAL1/10* promoter in front of the *POL-1* gene. This allows us to control the levels of cellular polymerase α, thus inducing replication stress by growing our strain in media with low levels of galactose (Lemoine *et al.,* 2009).

Our fifth element is a 0.5% divergence in the sequence of homologous chromosomes, obtained from mating two haploid yeast strains from different genetic backgrounds. Our two haploid strains are known as YJM and SGD, and their sequence divergence is due to single nucleotide polymorphisms (SNPs) along the lengths of each of their chromosomes (Lee *et al.,* 2009; Wei *et al.,* 2007). These SNPs are what enable us to characterise the genetic composition at particular loci once our diploid strain has been created.

It is important to note that at the point I entered this experiment components one through three of our strain were already in place. My contribution to strain construction involved only components four and five.

Our YJM haploid contains the *SUP4-o* gene, while our SGD haploid does not. Mating of these two strains creates a diploid that is heterozygous for *SUP4*-o. This means that breakages at FS2 repaired by homologous recombination and resulting in conversion of chromosome sequence from YJM to SGD sequence in the region of the *SUP4*-o gene will result in a LOH event, and subsequent production of red colonies as per the process described above (Figure 12). We are then able to utilize the SNPs to determine the identity of the sequence at different positions along chromosome III enabling us to

ultimately determine the method and extent of homologous recombination or other repair mechanism that resulted in LOH.

The result of this strain engineering is that we now have a strain that we can control stress levels in. Growing under stress and the subsequent expected breaks at FS2 and repair by methods of homologous recombination and telomere capping that result in LOH of *SUP4-o*. This loss of *SUP4-o* will result in pigment accumulation due to the pathway described in figure 12. After the growth of a colony of cells there is a clear sectoring within the colony of cells that are the progeny of the daughter cells of an LOH event. The colony will be split into a portion of red cells having no copies of *SUP4-o* and

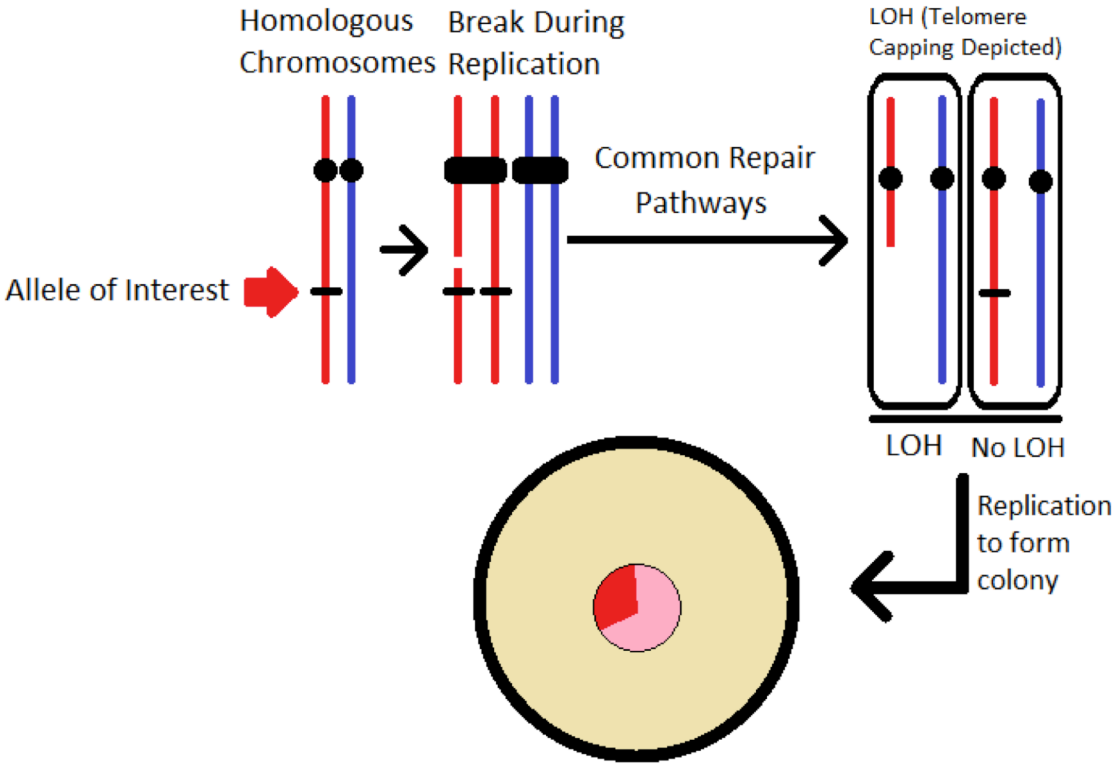

Figure 13: The experimental pathway and identification of LOH
The growth of cells under stress results in breakage at FS2 which in turn is repaired by a pathway that leads to LOH. After multiple rounds of replication a colony is formed as depicted, with a segment of red cells and a segment of white cells.

now being red, and a portion of pink (or white in the case of RCO) cells that contain one (or two for RCO) copies of *SUP4-o.* This is depicted in figure 13.

Our experiment began with the creation of the strain. This involved the Transformation Protocol and required the insertion of the *GAL1/10* promoter and mating of the SGD and YJM haploid strains to create our diploid experimental strain. We then confirmed the position of our *GAL1/10* promoter using a DNA Harvest PCR Protocol. Next we used the no Gal inoculation to induce breaks at our fragile site. We identified and isolated the daughter cells of a cell that had experienced LOH using the isolation and freeze down protocol. After this we conducted preliminary tests of SNP 3 (the SNP closest to the fragile site FS2) to determine if analysable LOH had occurred using SNP Analysis. For strains that did not test heterozygous in both the red and white colonies, we conducted further SNP analysis at eight more SNPs along chromosome III (again using SNP Analysis). For strains that appeared to be gene conversions we conducted addition testing of SNPs in order to increase the resolution of the borders of the gene conversion event, some of these further SNPs required digestion by enzymes that did not work in our Taq Polymerase buffer and as such a DNA purification protocol was used alongside the SNP Analysis Protocol. For strains that appeared to have either lost a chromosome or a chromosome arm, we did SNP testing at regions on the left arm of the chromosome utilizing the SNP Analysis protocol again.

Strain Creation. At the point where I entered this project, much of the necessary strain construction had been completed by Dr. Casper. The only step left was the addition of the *GAL1/10* promoter to the haploid strain containing SUP4-o, and mating of the SGD and YJM strains. We started with the strain Y651 and, upon successful addition of

the promoter re-named it Y655 then checked using DNA Harvest PCR, which allows us to amplify larger segments of DNA than colony PCR, which our promoter was in the correct position. Following this Dr Casper mated the SGD strain Y655 with the YJM strain Y325 and createdd the diploid strain Y657 that was to be used for our experiment.

Addition of GAL1/10 promoter protocol. In order to do this we inoculated our strain in 5mL liquid culture in YPD then incubated at 30°C shaking overnight. We used 1mL of the overnight culture to inoculate 50 mL of liquid media and grew this at 30°C for 3-4 hours. Checking OD_{600} till it measured between 0.4 and 0.6. Wetransfered culture to a 50 mL Falcon tube and spin at 5000 g for 5 minutes to pellet. We discarded the supernatant and resuspended the pellet in 1mL of sterile water in an Eppendorf tube. We re-centrifuged the cells at full speed for 30 seconds, dumped the supernatant again and re-suspended the pellet in 125µL each of 0.1M Lithium Acetate and 1X TE Buffer (note 1X TE Buffer consists of 0.1M Tris and 0.05M EDTA). We then mixed 100µL of salmon sperm carrier DNA (containing 100µL of herring testes carrier DNA containing the Gal Pol Promoter and a *G418* Resistance gene) and 10µL of transformation DNA in an Eppendorf tube before adding 100µL of yeast competent Y651 cells. We then added 600 µL of sterile 40% PEG/Lithium Acetate to each tube and vortex to mix before incubating at 30°C for 30 minutes, shaking at 200RPM. The cells were removed from the incubator and weadded 70 µL of 100% DMSO and mix by inversion before heat shocking the cells for 30 minutes in a water bath at 65°C. We then chilled the cells on ice for 2 minutes before centrifuging at 16,000gand discarding the supernatant. We resuspended the pellet in 150 µL of 1XTE and plated onto YPR+HG allowing the cells to grow in a 30°C incubator overnight. The next day we replica plated onto YPR+HG+G418 and allowed to

grow for 4 more days in an incubator at 30°C to identify transformed cells. (Gietz and Schiestl, 2007)

DNA harvest PCR protocol. We inoculated cell in 5mL of YPR+HG overnight, shaking at 32°C before spin down to pellet cells and resuspending the pellet in 500 µl of lysis buffer. We then add 6µL of 25 mg/mL 20T zymolases and incubate for 1 hour at 30°C. We centrifuge at 5200gfor 1 minute to pellet cells, discard the supernatant and resuspend in 500 µL of 50 mM Tris, 20mM EDTA before adding 50µL of 10% SDS and vortexing. We incubated the cells at 65°C for 30 minutes before adding 200µL KOAc and vortexing to mix then transferring to a refrigerator at 4°C for one hour. After removing the cells from the refrigerator we centrifuge at 16,000g for 15 minutes and transfer the supernatant to a new tube, adding 700µL of 2-propanol and mixing by inversion. We then rinse the pellet with 70% EtOH at 0°C and allow to air dry for 30 minutes. We resuspend the rinsed pellet in 300 µL 1XTE and then add 2 µL RNAse A at 10 mg/mL for a final concentration of 66.7 µg/mL before incubating for 15 minutes at 65°C, this breaks down any RNA which may interfere later during PCR. After this we add 30µL of 3 M NaOAc and mix by inversion, then add 500µL of room temperature 2-propanol and mix by inversion. We leave at room temperature for 5 minutes then centrifuge 5 minutes at 16,000g. We re-rinse the pellet with 0°C 70% ethanol then air dry for 30 minutes and then resuspend in 40 µL of TE. We create a master mix for the PCR containing 5 µL of Go Taq and 1µL of the respective primer mix per reaction, giving a final concentration of 0.5 µL Go Taq / 1 µL reaction and 0.1 µL primer mix / 1 µL reaction. We then add 4µL of supernatant from the boil freeze DNA to 6uL of master mix and place in a PCR machine, running the DNA Harvest cycle. 1*(94°C for 2 minutes),

3*(94°C for 30 seconds, 50°C for 30 seconds, 72°C for 2 minutes), 1*(72°C for 7 minutes).

Growth of cells under Stress. In order to determine the frequency of LOH events we needed a method that would allow us to quantify the number of LOH events experienced during replication in a defined number of cells. To do this we grew cells in a rich liquid media, then switch them to a liquid low galactose media to induce replication stress. We then determined our cell concentration. We used this information to spread approximately 500 cells over a plate of adenine deficient media (to promote the adenine synthesis pathway and hence build-up of intermediate) and grow them to colonies. We then determined LOH events by screening for red/white sectored colonies on these plates.

No Gal inoculation protocol. We inoculated Y657 cells in 5mL of YPR+HG and grew them overnight at 30°C. We then centrifuge cells and pour off supernatant before adding 10 mL of sterile water and re-centrifuge the cells, in order to wash off any residual galactose. We re-suspended cells in 10 mL of YPR+ No Gal and grew at 30°C in incubator for 6 hours, allowing enough time for approximately one cell division. We measured OD_{600} of the media after growth and converted the reading to cells/mL by multiplying by $3*10^7$. We then make serial dilutions to reduce the concentration of cells in a 10mL sample to 5000cells/mL. We then plate 100 µL of the diluted cell solution to R/W Analysis plates. These plates are incubated at 30°C to allow colony growth for 3 days. The plates are then removed from the incubator and put in a 6°C refrigerator overnight to allow for development of red coloration in *SUP4-o* deficient cells. The plates are removed from the incubator and colonies are counted by hand, with sectored colonies circled and noted.

Separation and purification of red/white sectored colonies. In order to analyse the cause of LOH on red/white (R/W) sectored colonies we must obtain a pure sample of each colony that can be drawn upon repetitively without losing genetic integrity. In order to do this we separate out each individual sectored strain and grow up a large number of cells which are then frozen for later use in glycerol.

Isolation and freeze down protocol. After identification of red white colonies each colonies' red portion and white portion are separated and grown up individually on YPR+HG plates for three days at 30°C.If complete separation is not successful on the first attempt we creating a streak of cells across a plate will allow for selection of individual progeny colonies of only one cell type (red or white). A pure sample is obtained it is plated onto YPR+HG and grown for 3 days at 30°C. In order to create a stock of each of the isolated strains we obtain an Eppendorf tube and add 250 µL of 20% Glycerol with a large number (about 100µL equivalent volume blob) of cells. The strain in each tube is assigned a name SC#R for red strains and SC#W for white strains originating from the same colony. We repeat the creation of glycerol stocks twice per strain and with each tube labelled with the strain name, we freeze down a working stock and a backup stock in separate -80°C freezers.

Single-nucleotide polymorphism (SNP) analysis. Our analysis of SNPs at each of the sites for each strain involves three steps. The first step involves isolation and amplification of DNA in the region surrounding the SNP, this is done by colony PCR. Second we take our PCR product and digest it with a sequence-specific endonuclease enzyme, which will cut at one of two possible sequences at the SNP site. Finally we view

the result of our PCR and digest using agarose gel electrophoresis and UV imaging. This allows us to discern bands of specific fragment lengths which can be interpreted as indicative of different sequences being present at our SNP sites.

Colony PCR protocol. We create a primer mix by mixing the two primers required to obtain a PCR product at each SNP, to an overall 1X concentration (using sterile water to dilute). The volume of primer mix created can vary depending on the amount we will need in upcoming experiments. We grew a patch of cells from the strain of interest (eg. SC998R) from glycerol stocks on YPR+HG plates for 2 days. We mixed 40μL of distilled H_2O and a blob of cells (about 20μL equivalent volume) in a PCR strip tube. We vortexed to mix cells and prevent sedimentation, before boiling the PCR strip tube in a PCR machine for 6 minutes, and transferring to a -80°C freezer for 10 minutes. This ruptures the cell membranes, releasing fragments of DNA to which our primers can later anneal. While the cells are in the freezer we created a PCR master mix, containing 5 μL of Go Taq and 1μL of 10 μM primer mix perer PCR reaction. Cells are removed from the freezer and thawed, before being spun down to pellet the cell debris and enable DNA to be collected from the supernatant with a minimum of other cell components. We added 4μL of supernatant from spin down to a new PCR strip tube being careful not to collect any cells in the supernatant then add 6μL of PCR master mix to each tube. We placed the tubes in a PCR machine and run the colony PCR cycle. 1*(94°C for 2 minutes), 3*(94°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 minute), 1*(72°C for 7 minutes).

Digest protocol. A digest master mix was made containing 1.25 μL of deionized H_2O for each sample, 1.25 μL of enzyme specific buffer for each sample and 0.5 μL of respective endonuclease for each sample. We mixed 3 μL of digest master mix to the

product of a colony PCR protocol and incubate each overnight in a water bath at the temperature stipulated by the enzyme being used.

Gel electrophoresis protocol. To create a 2% agarose gel we first mix 2 g of agarose in 100mL of deionized H_2O. We microwaved the mixture for 1 minute, removed and stired, then microwave till mixture is clear and liquid. We add 1.5μL of Gel Red to obtain a final concentration of $1.5*10^{-3}$ v/v to later enable us to view our DNA bands under UV light, then poured the gel into a cast and add combs to create wells for DNA samples. We allowed the gel to dry at room temperature for ~25 minutes and once the gel is dry we removed the combs and place the gel in a gel station filled just over the gel top with 0.5X TBE solution. We then loaded 3-5 μL (depending on well size) of digest protocol product into each well and run the gel at ~90V for ~40 minutes or until sufficient separation of band has occurred to allow for analysis of SNPs. We imaged the gel using the BIO RAD Molecular Imager and saved a natural image, as well as an image that was inverted and optimized for band analysis.

Analysis of Gels. Each SNP had a unique band size and cut site which results in the creation of two smaller bands. Table 2 is a list of expected band sizes depending on the sequence of the DNA at that chromosome region. According to these fragment sizes we are able to ascertain the identity of sequence at each SNP by analysing our gels. Figure 14 is an example of a SNP analysis, showing how the difference in product sizes is interpreted.

Table 2: List of SNPs used and expected band sizes

Cut Site on Chromosome III	SNP Name	Cutting Restriction Enzyme	PCR Product Size	YJM Product(s) Size	SGD Product(s) Size
112643	"-2"	*Bst*BI	402	285 & 117	402
113209	"-1"	*Mnl*I	462	462	335 & 127
114919	"0"	*Hinf*I	176	131 & 44	176
120340	"1"	*Hinf*I	452	452	280 & 170
130370	-	*Msp*I	422	422	236 & 186
148133	-	*Rsa*I	370	370	233 & 147
152544	-	*Alu*I	352	238 & 114	352
159890	-	*Ban*I	466	466	330 & 135
164273	"2"	*Msp*I	442	288 & 154	442
167720	-	*Hpy*CH4III	390	390	311 & 79
175324	"3"	*Ban*I	298	192 & 106	298
181520	-	*Bts*I	438	438	257 & 181
189048	-	*Alu*I	398	398	229 & 169
193671	"4"	*Hinf*I	351	229 & 122	351
195583	-	*Alu*I	364	364	215 & 149
201157	"5"	*Hinf*I	428	428	250 & 179
223672	"A"	*Dpn*II	369	202 & 167	369
233758	-	*Hpy*CH4III	369	369	274 & 95
247475	"6"	*Hinf*I	355	255 & 100	355
289633	"7"	*Msp*I	466	284 & 182	466
298875	"8"	*Rsa*I	353	226 & 127	353

Table 3: Activity of Restriction Enzymes Used

Restriction Enzyme	Units of Activity (in units/mL)
*Ban*I, *Hpy*CH4III, *Mnl*I	5,000
*Alu*I, *Hinf*I, *Rsa*I, *Dpn*II	10,000
*Bst*BI, *Msp*I	20,000

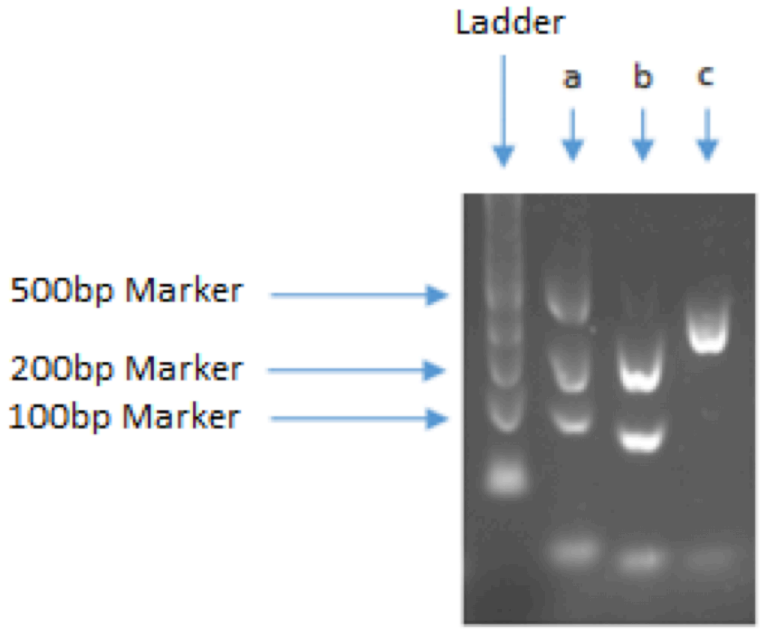

Figure 14: Example gel image of possible outcome of SNP analysis
a) represents a heterozygous test, containing all three bands means the two sequence amplified off two separate homologs in PCR were both differed, with one containing the sequence cut by the digestive enzyme used, and one not containing that sequence. b) Indicates a homozygous sequence. These two bands are both the result of PCR products off each homolog that was cut by the digestive enzyme to create two smaller bands (ignore the very faint third band, this is the result of a partial digest). c) Also indicates a homologous sequence but for the opposite identity to b. The PCR product of both homologs was not cut in c.

Purification of PCR product. Some of the digestive enzymes used (specifically the *Hpy*Ch4III enzyme) will not work in the presence of a GoTaq, and as such the purification of amplified DNA products after PCR is required. In order to purify DNA we used a QIAquick PCR Purification Kit (50).

DNA Purification protocol. We first added 50µL of PB buffer to the 10µL PCR product then place a provided QIAquick column in a 2mL collection tube. We applied the sample to the sample mixture into the QIAquick column and centrifuge at 17,900 x g for 60 seconds to allow the DNA to bind to the QIAquick column. We discarded the flow-through and added 600µL of Buffer PE to the QIAuick column and centrifuge s at 17,900 x g for 60 second to wash away residual contaminants. After discarding the flowthrough we centrifuge on last time for 1 minutes at 17,900 x g. We then placed the QIAquick column in an Eppendorf tube and add 30µl of water to the QIAquick column, allowing it to stand for one minute dissolving the annealed DNA and then centrifuge for one minutes at 17,900 x g to wash the DNA into our Eppendorf tube. We added one volume of provided loading dye to five volumes of DNA to allow for visualisation of DNA bands later during Gel Electrophoresis.

Media Formulas

Table 4 represents the formulas of the different kinds of media that were utilized throughout our experiment. We used two main types of media, and two main types of plates. Yeast peptide raffinose media with no galactose (YPR+ No Gal), yeast peptide raffinose media with high galactose (YPR + HG), yeast peptide raffinose plates with high galactose (YPR + HG) and red white sectored colony plates (R/W SC). For all media and plates components are mixed in water and sterile conditions are maintained.

Table 4: Media Formulas

	YPR + No Gal (Liquid Media)	YPR + HG (Liquid Media)	YPR + HG (Plates)	R/W SC (Plates)
Yeast Extract	1%w/v	1%w/v	1%w/v	-
Peptone	2%w/v	2%w/v	2%w/v	-
Raffinose	3%w/v	3%w/v	3%w/v	3%w/v
Agar	-	-	3%w/v	3%w/v
Galactose	-	$2.78*10^{-3}$M	$2.78*10^{-3}$M	$2.78*10^{-3}$M
Amino Acid Mix (No Adenine)	-	-	-	0.14w/v
Yeast Nitrogen Base (No aa's, No $(NH_4)_2SO_4$)	-	-	-	0.17%w/v
$(NH_4)_2SO_4$	-	-	-	0.5%w/v

Results

A total of 14,435 colonies were screened for LOH at the *SUP4*-o locus following induction of replication stress. The number of sectored colonies identified was 58, which is an overall LOH frequency of $4.01 \times 10^{-3,}$ or one per 189.9 colonies. Each sectored colony was first analysed at the SNP closest to the *SUP4*-o locus. Results of this initial analysis indicated heterozygosity at this SNP in both the red and white sides of the sectored colony in 18 strains, and these were subsequently excluded from analysis (they do not make up any of our 58 LOH events). The remaining 58 sectored colonies were analysed for a panel of eight additional SNPs on the right arm of yeast chromosome III. A SNP diagram was created for each of these sectored colonies after completion of this panel of additional SNPs. These diagrams facilitated categorization of the event responsible for LOH in each sectored colony. Examples of a typical SNP diagram for whole chromosome loss, chromosome right arm loss, RCO, BIR and GC are given in figure 15.

A typical gene conversion is indicated by a tract of conversion to YJM sequence on the SGD homolog in the red side of the sectored colony. A typical reciprocal crossover is indicated by a region of homozygous YJM sequence in the red side, correlating with a corresponding region of homozygous SGD sequence in the white side. A typical BIR is indicated by a region of homozygous YJM sequence in the red side, extending to the end of the chromosome. It is important to note that our method does not allow for distinguishing between BIR and telomere capping events; however, previous research has indicated telomere capping is quite rare and consequentially we are assuming all SNP diagrams which appear as formerly described are the result of a BIR

(Lemoine *et al.,* 2005). A typical whole chromosome loss is indicated by a complete conversion of all SNPs from -2, to 8 in the red strain to YJM sequence. A chromosome right arm loss is indicated by Conversion of all SNPs from 0 to 8 to YJM sequence on the SGD homolog in the red strain, but no maintenance of heterozygosity at SNPs -1 and -2.

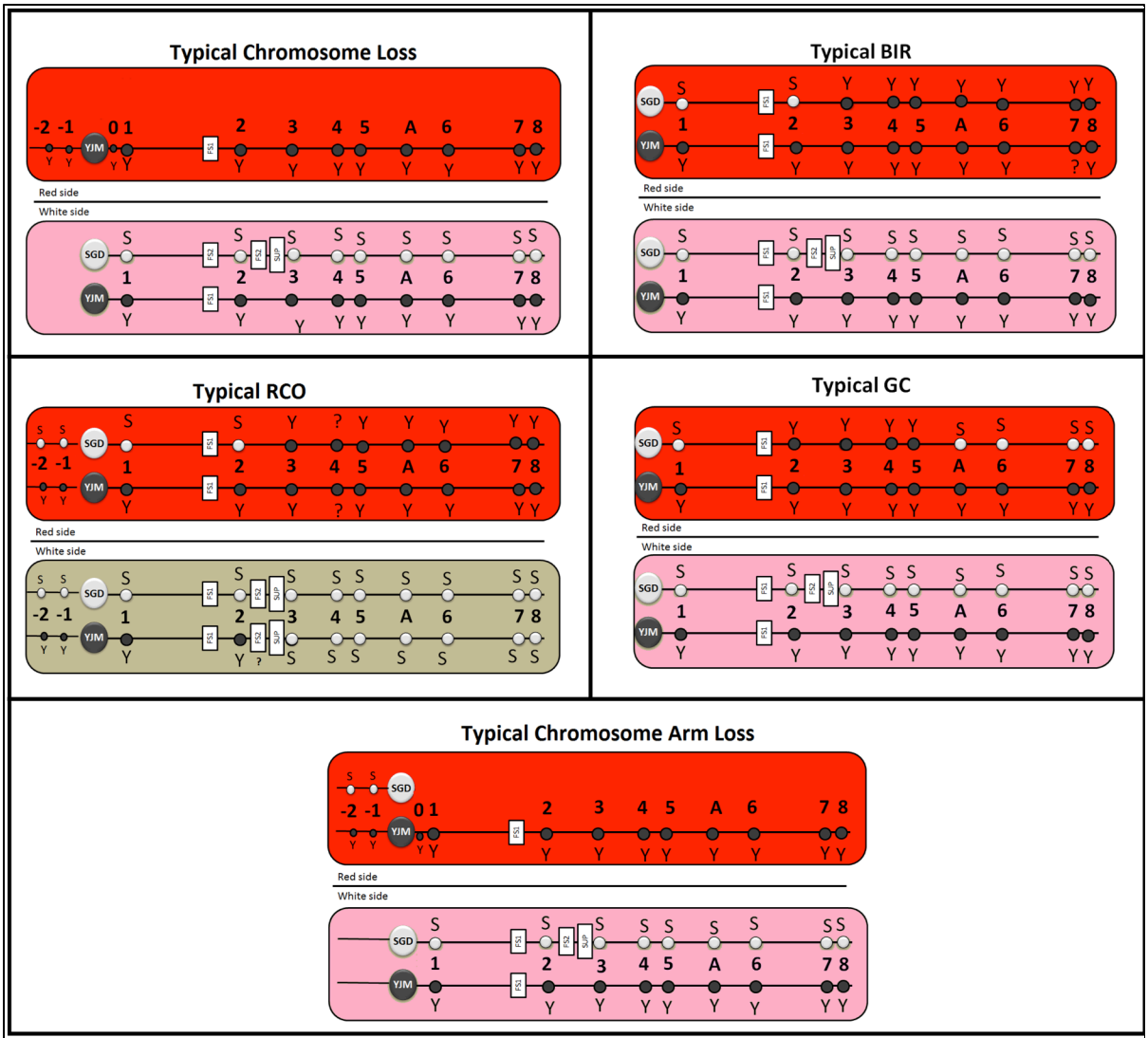

Figure15: Typical SNP Diagrams
This figure represents the typical SNP diagram created after SNP testing, and the events we'd interpret that diagram as. This figure shows only chromosome III of each cell, and cells are coloured to represent expected colouration of the cell due to presence or absence of SUP4-o. The homologous chromosomes are represented as white for SGD and black for YJM, with small circles of each colour (as well as the letter Y or S, for YJM and SGD respectively) at each SNP showing the sequence indicated by SNP testing. The single large circle on each homolog represents the centromere while smaller circles are shown for each of the 9 main SNPs and very small circles represent addition SNPs that were tested beyond the original 9 (SNPs -2,-1, and 0) colour of these circles indicates sequence at that region. Spacing of the SNPs shown is indicative but not a to scale precise representation of their positions on Ch. III of S. Cerevisiae. The question mark under FS2 on the SGD homolog of the white strain in the Typical RCO event indicates uncertainty of the presence of this structure due to a lack of resolution in our SNP testing. FS1 is a yeast fragile site consisting of two Ty elements, similar to FS2. However FS1 is less prone to breakage and is not a focus of our study (Casper et al., 2009)

Of the 58 sectored colonies analysed using the panel of eight additional SNPs, 55 could be classified into one of the typical categories shown in figure 15. Figure 16 gives diagrams for three sectored colonies that were not easily classified into one of the typical patterns.

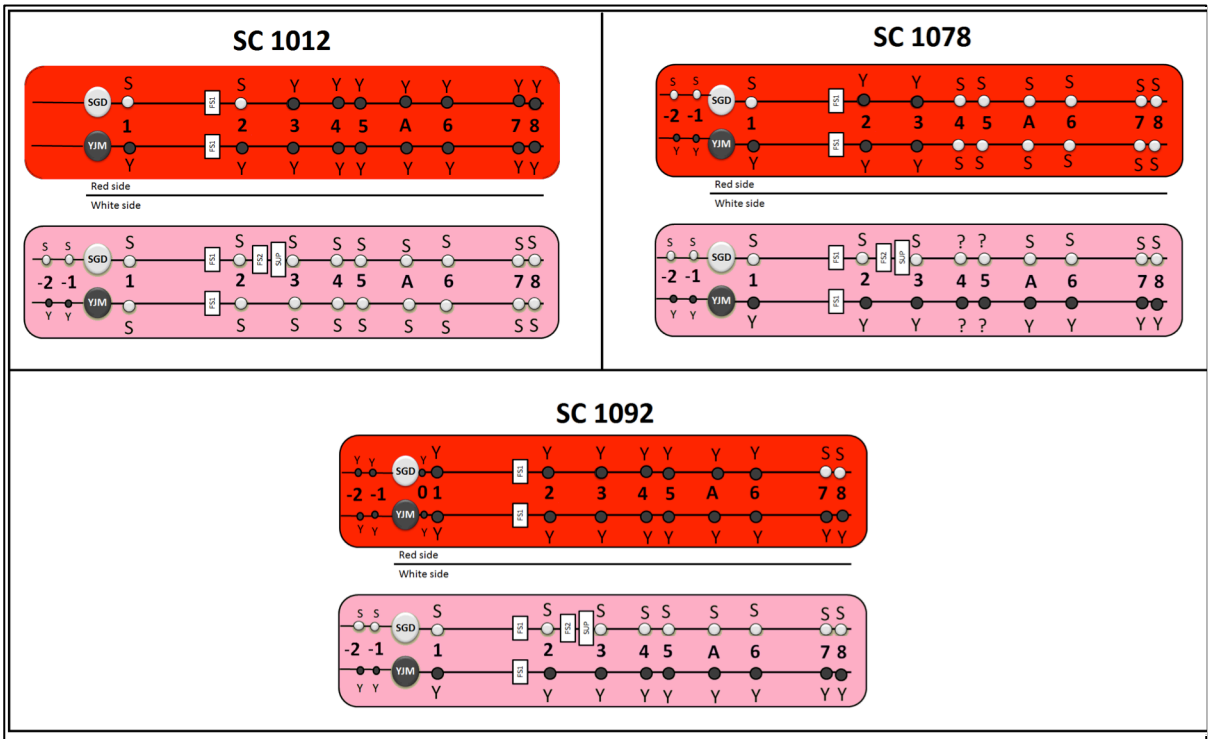

Figure 16: Unique SNP Diagrams
This figure represents the SNP diagrams of the three unique events we encountered during our experiment. See caption in figure 13 for details on how to read diagrams. SC1012 is best explained by a simultaneous BIR on the red strain, with a chromosome arm loss on the white. SC1078 is best described by simultaneous BIR events on each of the homologs in the red strain leading to two regions of LOH. SC1092 is not able to be explained by any repair mechanism.

These three sectored colonies are SC1012, SC1078 and SC1092. The pattern in SC1012 could be explained by a BIR on the red strain converting SNPs 3 to 8 to YJM sequence on the SGD homolog in the red strain, while a chromosome arm loss on the YJM homolog in the white strain converted SNPs 0 to 8, to SGD sequence. The patterns in SC1078 could be explained by simultaneous BIRs, with the SGD homolog beginning

at SNP 2 and the YJM homolog beginning at SNP 4, in the red strain. In this scenario by the time the invading SGD strand reached and was replicating off the SNP 4 region on the YJM homolog, the YJM homolog would have completed its repair of this region and converted that sequence to SGD sequence. Alternatively a gene conversion on the SGD homolog to the YJM sequence between SNPs 2 and 3, while a simultaneous telomere capping or BIR on the SGD homolog from SNP3, in the red strain, could also explain this diagram. Due to the uncertainty surrounding the repair of this strain we have not included it in our analysis as a gene conversion. The pattern in SC1092 cannot be explained by any known repair mechanism. This sectored colony exhibits a pattern similar to a long-tract gene conversion but is not consistent with this type of event, due to the fact that the proposed region of gene conversion crosses the centromere to include SNPs -1 and -2.

Classification of the 58 sectored colonies analysed by the panel of eight additional SNPs indicated that BIR was the most common cause of LOH, being present in 22 of 58 sectored colonies (38%). Chromosome loss was the next most common event, present in 20 of 58 strains (34%). Gene conversion was the least common of the major events, representing 9 of 58 (16%) of LOH events. Of the less common events we saw a single RCO (1.7%), 2 instances of multiple events (3.4%), 3 instances of chromosome arm loss (5.2%) and 1 unknown event (1.7%). Note that due to the two possible methods of independent assortment after RCO (Figure 10) the number of events we find is only half the number of events that have occurred, as such frequency of RCO events is calculated by doubling the number found then dividing by the total colony count. Table 5 indicates the frequency of each type of event, as well as the number we found and the relative percentage of LOH that was ascribed to this event.

Table 5: Frequency of LOH causes

	BIR	Ch Loss	GC	Unknown	Multiple Events	Ch Arm Loss	RCO
Number Found	22	20	9	1	2	3	1
Responsible for LOH	38%	34%	16%	1.70%	3.40%	5.20%	1.70%
Frequency	$1.62*10^{-3}$	$1.38*10^{-3}$	$6.23*10^{-4}$	$6.93*10^{-5}$	$1.39*10^{-4}$	$5.17*10^{-4}$	$1.39*10^{-4}$

Our BIR events were analysed and the section of the chromosome where the sequence changed from that of the original SGD homolog to that of the YJM homolog was identified. This 'border' for the BIR event gives us an idea of where the original double strand break occurred as it is in this region that strand invasion and homologous repair begins. Because the resolution provided by our SNP testing is not very high in this region, we are uncertain about the exact beginning point of BIR, as such we have indicated this 'unsure' region as well as the pre-repair region containing the original SGD sequence and the post-repair region containing the new YJM sequence as a result of BIR. We found 5 events which initiated between SNPs 0 and 1, 9 events which initiated between SNPs 1 and 2, 7 events which initiated between SNPs 2 and 3 and 1 event which initiated between SNPs 3 and 4 (Figure 17).

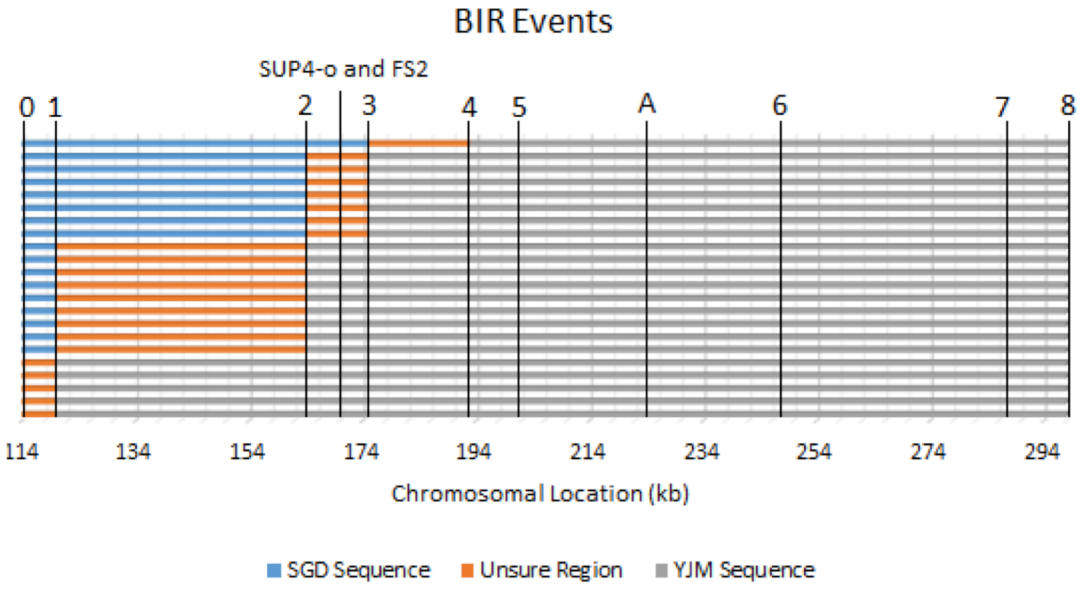

Figure 17: Break induced repair events
This figure depicts the location of the conversion of SGD to YJM sequence in our BIR events. The blue region indicates SGD sequence, the orange region is the 'unsure' region within which SGD sequence ends and YJM sequence begins, and the grey region indicates YJM sequence. The position of our SNPs 0 through 8 are indicated by black lines, as is the position of SUP4-o and FS2. At this resolution both SUP4-o and FS2 are indicated by a single line.

To determine the length of each GC we took the middle of each unsure region as the beginning/end of the GC tract, these lengths are reported in Table 6. We the used this data as well as the information in our SNP analysis to determine the position and length of each gene conversion. For typical gene conversions we determined the centromere proximal and centromere distal borders of each gene conversion to the highest resolution possible using all available SNPs, while we determined resolution for our larger gene conversions (those reaching to SNP 0) using only the initial 9 SNPs. Figure 18 represents the result of mapping our gene conversion tracts. We found that the image in figure 18 and the lengths calculated for table 6 suggested there may be two different classes of gene conversions, short-tract conversions and long-tract conversions, as has been recently reported in another yeast mitotic recombination system (Yim *et al.,* 2014). There are two

different possible separation points for the classification of short and long gene conversions in our data. In the first classification method (Classification Method 1) short-tract gene conversions are identified as those with conversion tracts under 25 kb, with long-tract gene conversions classified as anything larger than this. In the second classification method (Classification Method 2) short-tract gene conversions are classified as anything with a conversion tract under 100 kb, with long-tract gene conversions being anything greater than this.

Table 6: Estimated Length of Conversion Tracts and Possible Classification Methods

Gene Conversion Number	Classification Method 1	Classification Method 2	Length of Conversion Tract
1	Long-tract	Long-tract	153.5kb
2	Long-tract	Long-tract	153.5kb
3	Long-tract	Long-tract	125.5kb
4	Long-tract	Short-tract	85kb
5	Long-tract	Short-tract	79kb
6	Long-tract	Short-tract	69kb
7	Short-tract	Short-tract	23kb
8	Short-tract	Short-tract	19.5kb
9	Short-tract	Short-tract	14kb

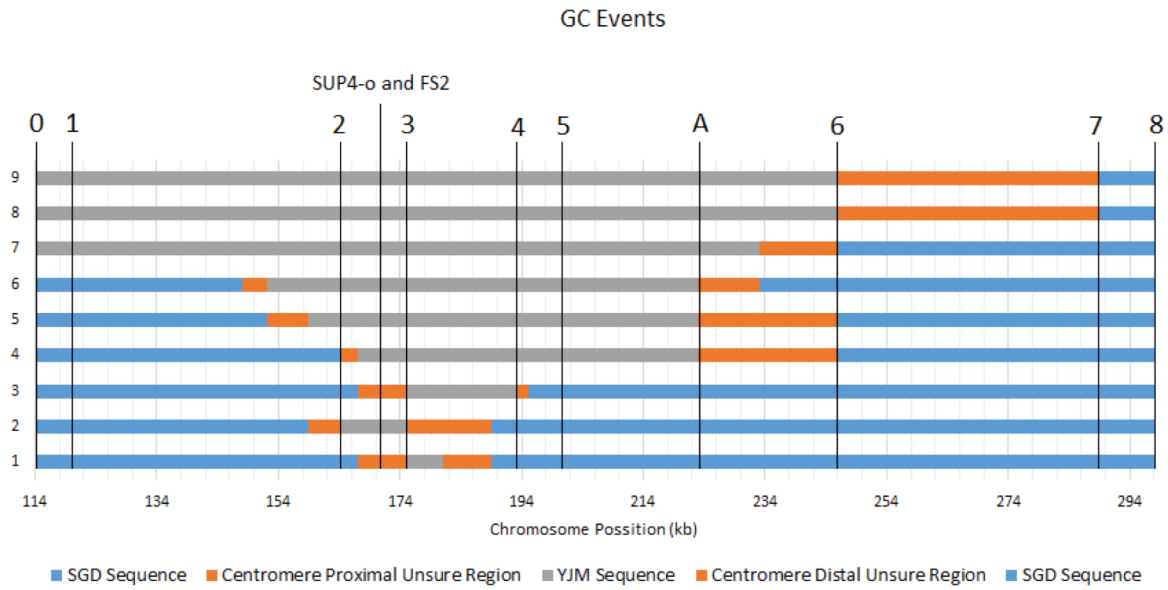

Figure 18: Gene conversion events
This figure depicts the conversion tracts of each of our gene conversion events. Shown in blue is the original SGD sequence of each chromosome, shown in orange are the centromere proximal (left) and centromere distal (right) regions of unknown sequence, shown in gray are the regions of YJM sequence. Also shown are the positions of SNPs 0-8 and SUP4-o as well as FS2, at this resolution FS2 and SUP4-o are depicted by a single line.

Utilizing the results in table 6 and classification method 1, we calculated the average length of each class of gene conversions. By this classification method, the average length for short-tract gene conversion events in our system is 18.8±3.7 kb, and 110.9±34.8 kb for long-tract gene conversion (Figure 19). In order to calculate if there was a statistically significant difference in average tract length between the two classes of gene conversion we used a one-tailed Student's t-test. The t-test using the values for gene conversion length in table 6 and classification method 1 gave a p value of 0.00251, indicating a significant difference between the two classes of gene conversion using this classification method.

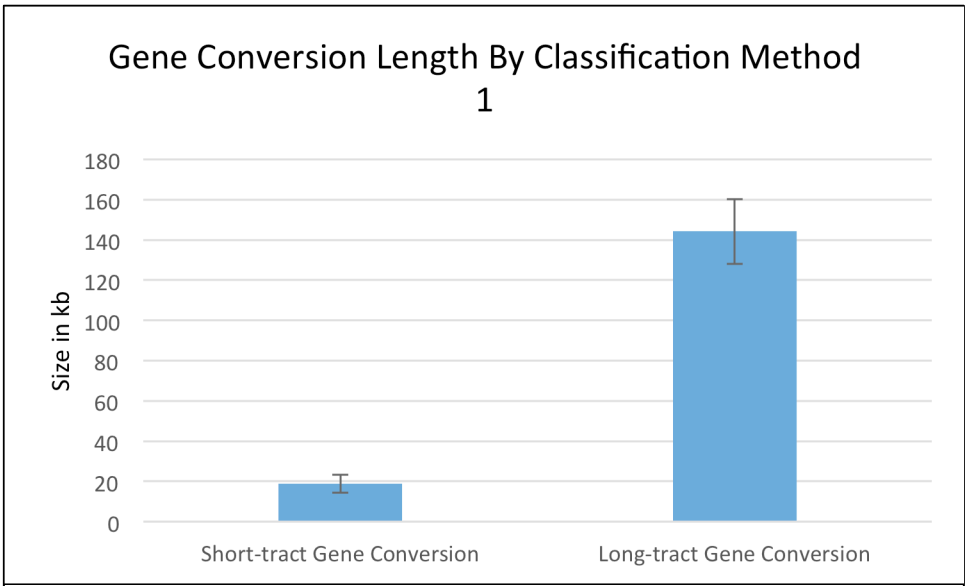

Figure 19: Mean conversion tract length for classification method 1
This figure represents the average tract lengths for short-tract and long-tract gene conversions using Classification Method 1. Short-tract gene conversions have an average length of 18.8±3.7 kb while long-tract gene conversions have an average length of 144.17±16.2 kb. A t-test between these two classes using this classification method gave a p value of .00251.

Utilizing the results in table 6 and classification method 2, we calculated the average length of each class of Gene Conversions. By this classification method, the average length for short-tract gene conversion events in our system is 48.3±29.9 kb, and 110.8±34.8 kb for long-tract gene conversion (Figure 20). In order to calculate if there was a statistically significant difference in average tract length between the two classes of GC we used a one-tailed Student's t-test. The t-test using the values for gene conversion length in table 6 and Classification Method 2 gave a p value of 0.00113, indicating a significant difference between the two classes of gene conversion using this classification method.

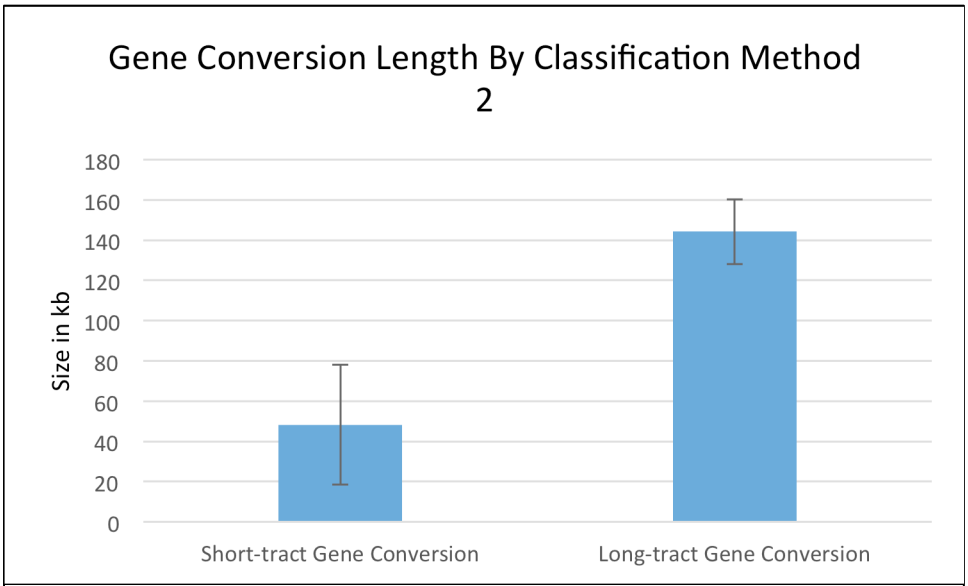

Figure 20: Mean conversion tract length for classification method 2
This figure represents the average tract lengths for short-tract and long-tract gene conversions using classification method 2. Short-tract gene conversions have an average length of 48.3±29.9 kb while long-tract gene conversions have an average length of 110.9±34.8 kb. A t-test between these two classes using this classification method gave a p value of .00113.

Discussion

The aim of this experiment was to analyse the use of homologous recombination resulting in gene conversion to repair breaks at the fragile site FS2 in *Saccharomyces cerevisiae*. We began by modifying a strain developed by Rosen *et al*. (2013) for analysis of mitotic recombination on yeast chromosome III where FS2 is located (Lemoin *et al*., 2005). By moving the *SUP4-o* marker gene up closer to the fragile site FS2 we were able to screen for LOH methods including localized gene conversions. This allowed us to identify two distinct classes of gene conversions and obtain an indication of the frequency of their use as a mechanism to repair breakage at FS2. Our results indicated that gene conversions are a significant method of repair by homologous recombination, being utilized in 12% of LOH events. They are not, however, as common as other methods such as break induced repair possibly due to the fact that a BIR event does not require the centromere distal arm of the chromosome to be caught after the initial strand break. As our analysis did not directly detect the repair pathway utilized it is important to discuss the possible mechanistic causes for the observed gene conversions at FS2.

The rate of overall spontaneous LOH rate in our system was $4.01*10^{-3}$ which is similar to what has previously been found in studies in our lab with the FS2 fragile site on chromosomes III which detected an event of $7.47*10^{-3}$ and $4.17*10^{-3}$ in their two respective experimental strains (Rosen *et al.,* 2013). This same previous study found rates of BIR to be $2.92*10^{-3}$ and $1.88*10^{-3,}$ very similar to our value of $1.52*10^{-3.}$ Finally our frequency of Crossover events was $1.39*10^{-4}$ while Rosen *et al.* (2013) found much greater frequencies of $2.56*10\text{-}^{3}$ and $2.02*10^{-3}$. These discrepancies could be the result of chance, as our small sample size means it is possible that with a larger total number of

colonies our detection of RCO events would be greater and this difference would become a statistical anomaly.

The overall frequency of gene conversion was $1.39*10^{-4}$ which is approximately ten times greater than has been previously found in studies lacking the FS2 fragile site where the frequency was $1.3*10^{-6}$ (Yim *et alk.,* 2014). As far as we know there have been no other characterizations of gene conversion rates in systems containing a fragile site, it will be interesting to see if this frequency is supported in future studies.

Previous research has reported evidence for two classes of gene conversions, short-tract and long-tract gene conversion, and it seems likely that these two classes arise from different recombination mechanisms. Yim *et al*. (2014) used a *S. cerevisiae* yeast system with a gene marker on chromosome V, they had a single Ty element near their gene marker, but no fragile site present. After finding two different sizes of gene conversions they proposed that the class of gene conversions identified as long-tract gene conversions is the product of double BIR events, while the class of smaller short-tract gene conversions is the product of the SDSA repair mechanism. Although the mechanism of SDSA leading to short-tract gene conversion is well established, a method of direct observation of a double BIR event will be needed to confirm its usage is responsible for the large-tract class of gene conversions. Our research also indicated that there may be two classes of gene conversions, however given the small number of gene conversions in of our data set, it is unclear which conversions should be included in each class. Yim *et al* (2013) classified their short-tract and long-tract gene conversions using a classification method that defined conversions under 25kb as short-tract and those over 25kb as long-

tract gene conversion. Their short tract gene conversions had a median size of 6.4kb, while their long tract had a median size of 54kb.

The results of our study also indicated two distinct classes of gene conversion. We found evidence for two different possible classification methods, but our median sizes using either method were not similar to what was found by Yim *et al.* (2013). Our classification method 1 has a mean short-tract gene conversion size of 18.8kb and mean long-tract gene conversion 110kb while our classification method two has a mean short-tract gene conversion size of 48.3kb and mean long-tract gene conversion 144.2kb. Classification by Method 1 contains a slightly smaller standard deviation for the average short-tract gene conversion class but a slightly larger standard deviation in the long-tract gene conversion. The classification method 1 also has a slightly larger p value (p=.00251) also and does not contain the qualitative property of the universal tract extension to the centromere found in classification method 2. Classification method 1 described in this paper is, however, the same as the classification method used by Yim *et al.* (2013). Higher numbers of colonies and better resolution by analysis by using more intermediate SNPs or DNA sequencing would allow for superior data, and enable better distinction between these classes of gene conversion. It may also be prudent, once more data is collected, to compare the tract-length classes to a standard and normal distribution to determine if classification is indeed statistically justified, and if so where bins of classification should be divided. It would also allow for better comparison to the Yim *et al.* (2013) and perhaps enable us to understand the discrepancies between our results.

Some of the red/white sectoring events we obtained in the screening process would require additional analysis to fully evaluate. Our sectored colonies which were

initially retained in the screening process because they had a clear red/white sectoring pattern that suggests LOH at the *SUP4-o* locus, but were later tested was determined to be heterozygous at SNP3 in both the red and white side of the colony, can be explained in a number of ways. First it is possible that there was a point mutation in SUP4-o resulting in a loss of function and subsequently colouration of the progeny colony. Such a point mutation would not be detected by our SNP analysis system. Second a very small Gene Conversion could be present spanning over *SUP4-o* but not the adjacent SNP 3. Finally it is possible that after homologous recombination a point mutation at the point of restriction for our SNP enzyme on one of the stains homologs could result in a false unrestricted band showing up on our gel. These possibilities are indistinguishable and would require higher resolution analysis or DNA sequencing to differentiate. The instances of multiple events observed also call for further analysis. The ability of a cell to simultaneously utilise multiple pathways may hold answers to some of the more unusual SNP diagrams we saw (Figure 16).

It is also important to note that our results are indicative of LOH events that influenced the *SUP4-o* marker only on the SGD homolog of chromosome III. Our results do not represent the universal use of the homologous recombination repair mechanism in yeast, nor even its universal use for repair on yeast chromosome III, because we can only detect events that result in LOH at the *SUP4*-o marker, and events elsewhere will be missed in our system. While the future study of CFS and their associated repair mechanisms and LOH should delve deeper into the mechanisms leading to LOH, there is much research to be conducted in different areas of this topic as well. In particular none of the hypotheses for why the conservation of CFS is so well-conserved in mammals are well studied or

supported. Perhaps the opportunities for chromosomal rearrangement after breakage at these sites allows for a novel method of gene duplication and chromosome evolution. Adding a second copy of a gene permits mutations to accumulate in the duplicate while the original gene form can be maintained, eventually producing new genes which encode for proteins with new, and useful, functions. This would provide organisms with the advantage of being able to more rapidly adapt to their environment, and the advantage of adaptation by gene duplication is already a supported hypothesis (Qian and Zahng, 2014). Research to answer hypotheses such as this and others requires the identification and analysis of common fragile sites and their surrounding loci in a broad range of organisms. There is also a lack of research on the behaviour of CFS *in vivo*. There is limited research on the causes of replication stress *in vivo* in cancer cells, and in other tissues of the body. Without the information this research would provide, we are lacking a vital segment of the overall picture of common fragile sites and their importance to our health.

Ultimately this study has achieved its intended purpose. We successfully identified LOH due to gene conversion, and unexpectedly added to the evidence of the very recently proposed separation of gene conversions into two distinct classes depending on the size of their conversion tract. Further research will be needed in order to fully understand the interplay between mechanisms of homologous recombination and their associated effect on LOH.

References

Aqeilan R.I., Trapasso F., Hussain S., Constinean S., Marshal D., Pekarsky Y., Hagan J.P., Zanesi N., Kaou M., Stein G.S., Lian J.B., and Croce C.M. 2007. Targeted deletion of Wwox reveals a tumor suppressor function. *The National Academy of Science of the USA.* 104(10), 3949-3954

Arlt M.F., Miller D.E., Beer D.G., and Glover T.W. 2002. Molecular Characterization of FRAXB and Comparative Common Fragile Site Instability in Cancer Cells. *Genes, Chromosomes and Cancer.* 33, 82-92.

Arlt M.F., Casper A.M., and Glover T.W. (2003). Common fragile sites. *Cytogenetic and Genome Research.* 100 (92-100).

Arlt M.F., Mulle J.G., Schaibley V.M., Ragland R.L., Durkin S.G., Warren S.T., and Glover T.W. 2009. Replicataion Stress Induces Genome-wide Copy Number Changes in Human Cells that Resemble Polymorphic and Pathogenic Variants. *The American Journal of Human Genetics.* 84, 339-350

Arlt M.F., Glover T.W. 2010. Inhibition of topoisomerase I prevents chromosome breakage at common fragile sites. *DNA Repair.* Doi: 10.1016/j.dnarep.2010.03.005

Arlt M.F., Wilson T.E., and Glover T.W. 2012. Replication Stress and Mechanisms of CNV Formation. *Current Opinion in Genetics and Development.* 22(3), 204-210.

Arlt M.F., Ozdemir A.C., Birkeland S.R., Wilson T.E., and Glover T.W. 2011. Hydroxyurea induces de novo copy number variants in human cells. *Proceedings of the National Academy of Science.* 108(42) 17360-17365.

Avila J., Lucas J.J., Perez M., and Hernandes F. 2003. Role of Tau Protein in both Physiological and Pathological Conditions. *Physical Review.* 84, 361-399

Bester A.C., Roniger M., Oren Y., Im M.M., Sarni D., Chaoat M., Bensimon A., Zamir G., Schewach D.S., and Kerem B. 2011. Nucleotide Deficiency Promotes Genomic Instability in early Stages of Cancer Development. *Cell.* 145, 435-446

Bignell G.R., Greenman C.D., Davies H., Butler A.P., Edkins S., Andrews J.M., Buck G., Chen L., Beare D., Latimer C., Widaa S., Hinton J., Fahey C., Fu B., Swamy S., Dagliesh G.L., Teg B.T., Deloukas P., Yang F., Campbell P.J., Futreal P.A., and Stratton M.R. 2010 Signatures of mutation and selection in the cancer genome. *Nature.* 463, 893-898

Bugert P., Wilhelm M., and Kovacs G. 1997. FHIT Gene and the FRA3B Region Are Not Involved in the Genetics of Renal Cell Carcinomas. *Genes, Chromosomes and Cancer.* 20, 9-15.

Casper A.M., Rosen D.M., Rajula K.D. 2012. Sites of genetic instability in mitosis and cancer. *Annals of the New York Academy of Science.* 1267. 24-30

Callahan G., Denison S.R., Phillips L.A., Shridhar V., and Smith D.I 2003. Characterization of the common fragile site FRA9E and its potential role in ovarian cancer. *Oncogene.* 22, 590-601

Chan K.L., Palmai-Pallag T., Ying S., and Hickson I.D. 2009. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nature Cell Biology.* 11, 753-761.

Chang N.S., Pratt N., Heath J., Schultz L., Sleve D., Carey G.B., and Zevotek N. 2001. Hyaluronidase inducation of a WW domain-containing oxireductase that enhances tumor necrosis factor cytotoxicity. *Journal of Biological Chemistry.* 276(5), 3361-3370

Coelln R.V., Dawson V.L., Dawson T.M. 2004. Parkin-associated Parkinson's disease. *Cell Tissue Review.* 318, 175-184.

Conquelle A., Pipiras E., Toledo F., Buttin G., and Debatisse M. (1997) Expression of Fragile Sites Triggers Intrachromosomal Mammalian Gene Amplification and Sets Boundaries to Early Amplicons. *Cell.* 89, 215-225.

Debacker K., Kooy R.G. (2007) Fragile sites and human disease. *Human Molecular Genetics.* 16(2), 150-158.

Debatisse M., Le Tallec., B., Letessier A., Dutrillaux B., and Brison O. 2011. Common fragile sites: mechanisms of instability revisited. *Trends in Genetics.* Xx, 1-11

Denison S.R., Callahan G., Becker N.A., Phillips L.A., and Smith D.I. 2003. Characterization of FRA6E and Its Potential Role in Autosomal Recessive Juvenile Parkinsonism and Ovarian Cancer. *Genes, Chromosomes and Cancer.* 38, 40-52

Denison S.R., Wang F., Becker N.A., Schule B., Kock N., Phillips L.A., Klein C., and Smith D.I. 2003 Alterations in the common fragile site gene *Parkin* in ovarian and other cancers. *Oncogene* 22, 8370-8379

Durkin S.G., and Glover T.W. 2007. Chromosome Fragile Sites. *Annual Review of Genetics.* 41, 169-92

Edwin Jr H. and Scherer S. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature.* 455(16), 919-923.

Escobar-Henriques M., and Daidnan-Fornier B. 2001. Transcriptional Regulation of the Yeast GMP Synthesis Pathway by Its End Products. *The Journal of Biological Chemistry.* 276(2), 1523-1530

Fang J.M., Arlt M.F., Burgess A.C., Dagenais S.L., Beer D.G., and Glover T.W. 2001. Translocation Breakpoints in FHIT and FRA3B in Both Homologs of Chromosome 3 in an Esophageal Adenocarcinoma. *Genes, Chromosomes and Cancer.* 30, 292-298

Filipp J.S., Sung P., and Klein H. 2008. Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry.* 77, 229-57

Finnis M., Dayan S., Hobson L., Chenevix-Trench G., Friend K., Ried K., Venter D., Woollatt E., Baker E., and Richards R.I. Common chromosomal fragile site FRA16D mutation in cancer cells. *Human Molecular Genetics.* 14(10) 1341-1349

Girirajan S., and Eicher E.E. 2010. Phenotypic variability and genetic susceptibility to genomic disorders. *Human Molecular Genetics.* 19(2), R176-187

Glover T.W., Berger C., Coyle G., and Echo B. 1984. DNA polymerase a inhibition by aphidicolin induces gaps and breaks in common fragile sites in human chromosomes. *Hman Genetics.* 67, 136-142

Glover T.W., Arlt M.F., and Casper A.M., Durkin S.G. (2005). Mechanisms of common fragile site instability. *Human Molecular Genetics.* 14(2), 197-205.

Glover T.W., and Wilson T.E. (2013) Breaking News on Fragile Sites in Cancer. *Cancer Cell.* 23, 137-139

Gostout B.S., Myers S.L., Smith D.I., and Thorland E.C. 2003. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene.* 22(8), 1225-1238

Handt O., Baker E., Dayan S., Gartler S.M., Woollatt E., Richards R.I., and Hansen R.S. 2000. Analysis of replication timing at the FRA10B and FRA16B fragile site loci. *Chromosome Research.* 8. 677-688.

Huang H., Reed C.P., Mordi A., Lomberk G., Wang L., Shridhar V., Hartmann L., Jenkins R., and Smith D.I. (1999) Frequent Deletions Within FRA7G at 7q31.2 in Invasive Epithelial Ovarian Cancer. *Genes, Chromosomes and Cancer.* 24, 48-55.

Huebner K., Croce C.M. (2001) FRA3B and Other Common Fragile Sites: The Weakest Link. *Nature Reviews/Cancer.* 1, 214-221.

Helmrich A., Ballarino M., and Tora L. 2011. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell.* 44, 966-977

Ingvarsson S. (2001). FHIT alterations in breast cancer. *Cancer Biology.* 11, 361-366

Jemal A., Bray F., Center M.M., Ferlay M.E., Ward E., and Forman D. 2011. Global Cancer Statistics. *California Cancer Journal for Clinicians.* 61, 69-90

Jiang Y., Lucas I., Young D., Elizabeth D., Karrison T., Rest J., and Le Beau M. (2009). Common fragile sites are characterized by histone hypoacetylation. *Human Molecular Genetics*. 18(23), 4501-4512.

Krummel K.A., Roberts L.R., Kawakami M., Glover T.W., and Smith D.I. 2000. The Characterization of the Common Fragile Site FRA16D and Its Involvement in Multiple Myeloma Translocations. *Genomics.* 69, 37-46

Le Beau M., M., Rassool F.V., Neilly M.E., Ill R.E., Glover T.W., Smith D.I., McKeithan T.W. 1998. Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Human Molecular Genetics.* 7(4), 755-761

Le Tallec B., Millot G.A., Blin M.E., Brison O., Dutrillaux B., and Debatisse M. 2013. Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals that Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes. *Cell Reports.* 4, 420-428

Le Tallec B., Koundroukoff S., Wilheim T., Letessier A., Brison O., and Debatisse M. 2014. Updating the mechanisms of common fragile site instability: how to reconcile the different views?. *Cell Mol Life Sci. Unpublished.*

Lee P.S., Greenwell P.W., Dominska M., Gawel M., Hamilton M., and Petes T.D. 2009. A Fine-Structure Map of Spontaneous Mitotic Crossovers in the Yeast *Saccharomyces cerevisiae.* 5,. *Public Library of Science: Genetics.* 5(3), e1000410

Letessier A., Millot G.A., Koundrioukoff S., Lachages A.M., Vogt N., Hansent R.S., Malfoy B., Brison O., Debatisse. 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature.* 470. 120-124

Lemoine F.J., Dagtyareva N.P., Lobachev K., and Petes T.D. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase A model for chromosome fragile sites. *Cell.* 120(5), 587-598

Llorente B., Smith C.E., and Symington L.S. 2008. Break-induced replicataion: what is it and what is it for? *Cell Cycle.* 7, 859-864

Matuszek G., and Talebizadeg Z. 2009. Autism genetic database (AGD): a complrehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. *BMC Medical Genetics.* 10:102, 10.1186/1471-2350-10-102.

Metsu S., Rooms L., Rainger J., Taylor M.S., Bengani H., Wilson D.I., Chilamakuri C.S.R., Morrison H., Vandeweyer G., Reyniers E., Douglas E., Thompson G., Haan E., Gecz J., FitzPatrick D., and Kooy R.F. (2014) FRA2A Is a CGG Repeat Expansion Associated with Silencing of *AFF*3. *PLOS Genetics.* 10(4), e1004242

Miller C.T., Lin L., Casper A.M., Lim J., Thomas D.G., Orringer M.B., Chang A.C., Chambers A.F., Giordano T.J., Glover T.W., and Beer D.G. 2006. Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in oesophageal adenocarcinoma. *Onceogne.* 25, 409-418

Mishmar D., Rahat A., Scherer S., Nyakatura G., Hinzmann B., Kohwi Y., Mandel-Gutfroind Y., Lee J., Drescher B., Sas D., Margalit H., Platzer M., Weiss A.,

Tsui L., Rosenthal A., and Kerem B. (1998) Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *The National Academy of Sciences.* 95, 8141-8146.

Mitchel K., Hengshan Zhang., Welz-Voegele C., and Jinks-Robertson S. (2010). Molecular Structures of Crossover and Noncrossover Intermediates during Gap Repair in Yeast: Implications for Recombination. *Molecular Cell*. 38, 211-222.

Morrow E. (2010) Genomic Copy Number Variation in Disorders of Cognitive Development. *Journal of American Child and Adolescent Psychiatry.* 49(11), 1091-1104.

Nunez M.I., and Ludes-Meyers J. 2006. WWOX protein expression in normal human tissues. *Journal of Molecular Histology.* 37, 115-125.

Ozeri-Galai E., Lebofsky R., Rahat A., Bester A.C., Bensimon A., and Karem B. 2011. Failure of Origin Activation in Response to Fork Stalling Leads to Chromosomal Instability at Fragile Sites. *Molecular Cell.* 43, 122-131

Ozeri-Galai E., Bester A.C., and Batsheva K. 2012. The complex basis underlying common fragile site instability in cancer. *Trends in Genetics.* 28(6), 295-302

Palumbo E., Matricardi L., Tosoni E., Bensimon A., and Russo A. 2010. Replication dynamics at common fragile site FRA6E. *Chromosoma.* 119, 575-587.

Popovici C., Basset C., Bertucci B., Orsetti B., Adelaide J., Mozziconacci M., Conte N., Murati A., Ginestier C., Charafe-Jauffret E., Ethier S.P., Lafage-Pochitaloff M., Theillet C., Birnbaum D., and Chaffanet M. (2002) Reciprocal Translocations in

Breast Tumor Cell Lines: Cloning of a t(3;20) That Targets the *FHIT* Gene. *Genes, Chromosomes and Cancer.* 35, 204-218

Qian W., and Zhand J.G. 2014. Genomic evidence for adaptation by gene duplication. *Genomic Research. Early Online.* gr 172098

Ried K., Finnis M., Hobson L., Mangelsdorf M., Dayan S., Nancarrow J.K., Woollatt E., Kremmidiotis G., Gardner A., Venter D., Baker E., and Richards R.I. 2000. Common chromosomal fragile site FRA16D sequence: identification of the *FOR* gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Human Molecular Genetics.* 9(11), 1651-1663

Rosen D.M., Younkin E.M., Miller S.D., Casper A.M. 2013. Fragile site Instability in *Saccharomyces cerevisiae* Cases Loss of Heterozygosity by Mitotic Crossover and Break-Induced Replication.

Smith D.I., Zhu Y., McAvoy S., and Kuhn R. 2006. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Letters.* 232, 48-57.

Sozzi G., Veronese M.L., NegriniM., Baffa R., Cotticelli M.G., Inoue H., Tornielli S., Pilotti S., Gregorio L.D., Pastorino U., Pierotti M.A., Ohta M., Huebner K., and Croce C.M. 1996. The FHIT Gene at 3p14.2 Is Abnormal in Lung Cancer. *Cell Press.* 85, 17-26

Stratton M.R., Campbell P.J., and Futreal P.A. 2009. The cancer genome. *Nature.* 458(9), 719-724

Wei W., McCasher J.H., Hyman R.W., Jones R., and Ning Y. 2007 Genome Sequencing Comparative Analysis of *Saccharomyces cerevisiae* strain Y789. *Proceedings of the National Achademy of Science.* 104: 12825-12830.

Wilke C.M., Hall B.K., Hoge A., Paradee W., Smith D.I., and Glover T.W. 1996. FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Human Molecular Genetics.* 5(2) 187-195

Winneoenninckx B., Debacker K., Ramsay J., Smeets., Smits., FitzPatrick D.R., and Kooy R.F. (2007) CGG-Repeat Expansion in the DOP2B Gene Is Associated with the Fragile Site FRA12A on Chromosome12q13.1. *The American Journal of Human Genetics*. 80, 221-231.

Yim E., O'Connell K.E., St Charles J., and Petes T.D. 2014. High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae. Genetics: Early Online* 10.1534/genetics.114.167395