UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVI Ciclo

Exploring genome integrity pathways in *Saccharomyces cerevisiae*

Muthukumar Shanmugam

PhD Thesis

Scientific tutor: Prof. Paolo Plevani

Academic year: 2013-2014

SSD: BIO/11

Thesis performed at Dipartimento di Bioscienze

Contents

Part 1

Abstract:

Genomic DNA is under constant attack from both endogenous and exogenous DNA damaging agents like reactive oxygen species which include O_2, H_2O_2, OH, reactive carbonyl species, alkylating agents such as estrogen and cholesterol metabolites, radiations (like UV, x-rays and gamma rays) and mutagenic chemicals. Moreover, threats to DNA integrity can also come from DNA metabolism such as replication, transcription and recombination. In order to survive and faithfully transmit the genetic material to the progeny, cells must detect the damage and activate repair mechanisms and, if the damage cannot be repaired, trigger the apoptotic program. All these processes, which are collectively known as DNA damage response (DDR), are coordinated by surveillance mechanisms often called DNA damage checkpoint, which temporarily halt or slow down cell cycle progression to provide enough time for DNA repair. The failure of the DNA damage response and other mechanisms deputed to the maintenance of genome integrity leads to a condition called “Genome Instability”, consisting in the accumulation of damage, genomic aberrations, such as mutations, gross chromosomal rearrangements and chromosome loss. Genome instability is a hallmark of cancer and a driving force in tumorigenesis.

We exploit budding yeast *Saccharomyces cerevisiae* as a model system for studies on genome maintenance pathways which are highly conserved throughout evolution from yeast to human. Despite recent advances in the

field, genome integrity pathways are not yet fully understood and not all the genes involved have been identified. We developed a screening strategy, based on the overexpression of *DDC2*, a critical DNA damage checkpoint gene in the contest of a yeast deletion collection, in order to identify genes controlling genome integrity on the basis of spontaneous accumulation of endogenous DNA damage. We identified several genes and pathways associated with genome integrity maintenance, among which are many genes induced in peroxisome biogenesis and mitochondria structure and function, as well as several uncharacterized ORFs.

State of the art

1. Introduction to DNA damage and genome maintenance

DNA, the blueprint of life, encodes the genetic information required for the development, functioning and survival of all known living organisms and many viruses. DNA is the carrier of genetic material and has to be replicated and passed through generations without or with few errors. Genomic DNA is under constant attack from both endogenous and exogenous DNA damaging agents (See Fig 1). Many of the DNA lesions caused represent structural impediments to DNA replication or transcription.

1.1 Endogenous DNA damage

The majority of DNA modifications are endogenous in origin (De Bont, R., et al., 2004). Spontaneous hydrolysis is the simplest form of endogenous DNA damage leading to depurination (Lindahl T, 1993). The N-glycosidic bond between the DNA base and the deoxyribose is particularly prone to acid-catalyzed hydrolysis which leads to abasic or apurinic/apyrimidinic (AP) sites, and the loss is estimated to occur at the rate of approximately 10,000 per cell per day (Lindahl T, 1993; Lindahl T, 1972). The AP sites are genotoxic as they stall DNA synthesis and may lead to collapse of the replication forks, causing the formation of double strand breaks

(DSBs). AP sites are also mutagenic, as the replicative DNA polymerases lack a template, and thus a random base is inserted, although in 54% of the cases is an adenine (De Bont and van Larebeke, 2004; Lawrence et al., 1990). Deamination is another common reaction of hydrolysis where DNA bases carry the exocyclic amino group (Lindahl T, 1993; Yonekura S. et al., 2009). The formation of uracil from cytosine is the most common lesion that occurs at an estimated 100-500 times per cell per day (McKinnon PJ, 2009; De Bont R, 2004). Similarly, adenine and guanine deaminate to form hypoxanthine and xanthine, although at a much lower rate (Lindahl T, 1993).

Moreover, normal cellular metabolism acts as a source of endogenous reactive oxygen species (ROS) and reactive nitrogen species. The ROS, which include O_2, H_2O_2, OH (Lindahl T 1972; Sugiyama H, 1994) generate more than one hundred different oxidative DNA lesions, such as base modification, deoxyribose oxidation, single or double strand breaks and DNA-protein cross links (Cadet J, 1997). Similarly, endogenous reactive nitrogen species can produce oxidative adducts from primarily nitric oxide and its by-products (Burney S, et al., 1999). One of the most extensively studied oxidative DNA lesions is 8-oxoguanine, which is routinely used as an analytical measure of oxidative DNA damage in biological systems (Ravanat J-L, 2005). The reactive carbonyl species (RCS) are potent mediators of cellular carbonyl stress originating from endogenous chemical processes such as lipid peroxidation and glycation (Roberts et al., 2003). The oxidized lipid products react with DNA, which can result in the severe crosslinking between opposite DNA strands, (interstrand crosslinks, ICLs) (Friedberg, 2006). Moreover, alkylation to DNA damage may arise from endogenous (i.e., S-adenosylmethionine pool, lipid peroxidation products) or exogenous sources. However, due to endogenous agents reactivity, mutagenic and cytotoxic adducts can also be formed (De Bont and van Larebeke, 2004).

1.2 Exogenous DNA damage

Overall, exogenous DNA damage is more bulky compared to endogenous DNA damage, and is the main source for DSBs (De Bont and van Larebeke, 2004). Exogenous DNA damage can be produced by physical or chemical sources. For example, physical genotoxic agents from sunlight are ionizing radiation (IR) and ultraviolet (UV) light, the latter is estimated to induce 10^5 DNA lesions (pyrimidine dimer and 6-4 photoproducts) per cell per day (Hoeijmaker, 2009). IR (from, e.g., Cosmic radiation and X-rays or radiotherapy) can induce oxidation of DNA bases and generate single-strand DNA break (SSBs) and double-strand DNA break (DSBs).

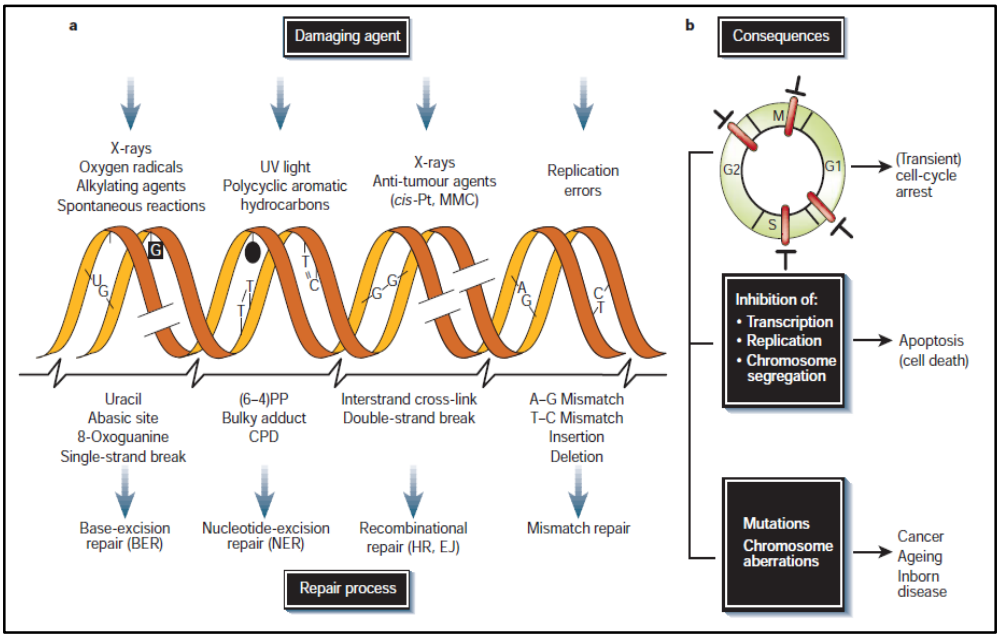


Figure 1. *Schematic representation of DNA damage, repair and its consequences* (taken from Jan H. J**.** Hoeijmakers, 2001)

Besides the physical insults, cells must also cope with several chemical sources of DNA damage (Wogan GN, 2004; Irigaray P, 2010). For

example, different types of chemical agents have been developed and used to target DNA as a means to treat cancer, and they cause a variety of DNA lesions. Among these agents we can mention alkylating agents such as methyl methanesulfonate (MMS) and temozolomide, as well as bifunctional alkylating agents, such as nitrogen mustards, platinum compounds and the natural product mitomycin C (MMC) that causes DNA damage in the form of intrastrand and interstrand cross-links (Noll DM, et al., 2006).

Chemotherapeutic drugs, such as topoisomerse I or II inhibitors (e.g., camptothecin or etoposide, respectively), generate SSBs or DSBs by trapping topoisomerase–DNA covalent complexes, respectively (Sinha BK, 1995). Widely used DNA-damaging chemicals include N nitrosoamines, heterocyclic amines, and polycyclicaromatic hydrocarbons (e.g., benzo[a]pyrene), which are commonly found in the diet, with the latter also being produced in air emissions, such as cigarette smoke and vehicle exhaust. DNA damage can be both mutagenic and cytotoxic to the cell. If left unrepaired upon DNA replication, damage to DNA can give rise to mutations, and accumulation of mutations in genes coding for proteins involved in the cell's regulation of growth and death may, in the worst case, give rise to immortal cancer cells. DNA damage may also impair protein synthesis, arrest the cell cycle, ultimately leading to cell death - so the cytotoxicity of the DNA damage.

2. Genome instability and its consequences

Genome instability is a broad term encompassing many forms of genome aberrations, going from point mutations to chromosomal rearrangements or loss (Aguilera and Gomez-Gonzalez, 2008). Although detrimental to the cell in most instances, these rearrangements and mutations can be of beneficial as they drive evolution at the molecular level, generating genetic variation

(Aguilera and Gomez-Gonzalez, 2008). Depending on the mechanisms involved, genome instability that can result in: a) mutations, including point mutations; b) microsatellite instability due to contraction or expansion; c) variation in the chromosome number caused by failures in the chromosome segregation apparatus or the mitotic checkpoint, termed chromosome instability (CIN), (Aguilera and Gomez-Gonzalez, 2008); d) gross chromosomal rearrangement (GCR) such as duplications, deletions, translocations and inversions that involve changes in genetic linkage between at least two DNA fragments (Aguilera and Gomez-Gonzalez, 2008). In addition other types of genetic alternations may occur, including copy number variation (CNV), hyper-recombination, and loss of heterozygosity (LOH) (Aguilera and Garcia-Muse, 2013). Genomic instability is a major driving force for tumorigenesis, and it is a feature of almost all types of human cancers (Negrini et al, 2010). During cell division, genomic instability is associated with the failure of parental cells to duplicate accurately the genome and precisely distribute the genomic material to the daughter cells, which result in various forms of genome alterations in the daughter cells. Accumulation of these genomic alterations may cause dysregulation of cell division, imbalance in cell growth and cancer. Most tumors are genetically unstable, providing the genetic plasticity to drive the stepwise progression of genetic changes required for the development of malignancy. There are two main models, which could explain the cancer development, namely; a) mutator phenotype, and b) oncogene induced DNA damage.

a) Mutator phenotype: The mutator phenotype hypothesis describes cancer as a process, where cells subsequently undergo multiple rounds of mutation and selection (Loeb LA., 2011). Misregulation or mutation in gene responsible for genome stability, could increase the possibilities that a

subsequent mutation will occur in an oncogene resulting in driver mutations, which confer a growth advantage (See Fig. 2).

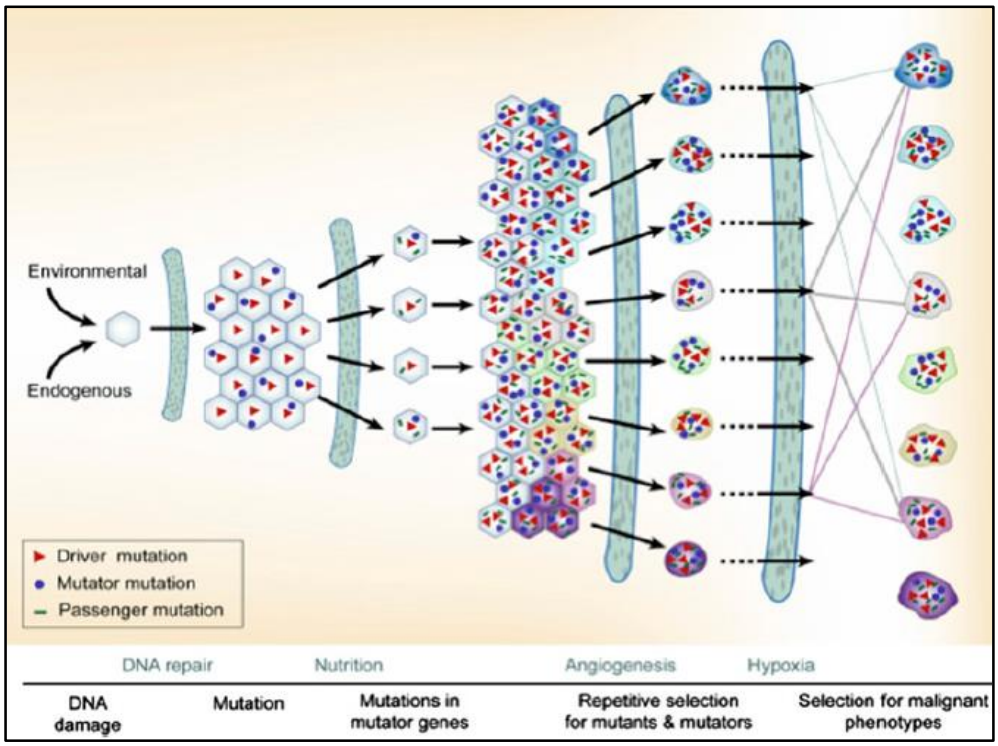


Figure 2: *Mutator hypothesis shows that cancer cells exhibit a mutator phenotype* (taken from Edward. J. Fox, 2013)

Accordingly, studies in yeast (Kolodner, R. D, et al., 2002; Herr, A. J, et al., 2011) and bacteria (Millerhe J. H., et al., 1999), indicated that many mutator mutations confer initial growth advantage compared to the wild type. Indeed, cancers are known to exhibit genome instability and mutations (Ellis NA, et al., 1995) in DNA repair and DNA maintenance genes are associated with hereditary cancers or with a mutator phenotype (Schmitt MW, et al., 2012). The concept of mutator phenotype in human cancer has been discussed for many years (Loeb, L. A., et al., 1974; Cleaver JE, 1968; Ellis NA, et al., 1995 and cancers are known to exhibit genome instability (Schmitt MW, et al.,

2012). The marked heterogeneity of cancer and the Darwinian evolution of a tumor suggest that different regions within the tumor may have different mutations (See Fig. 3).

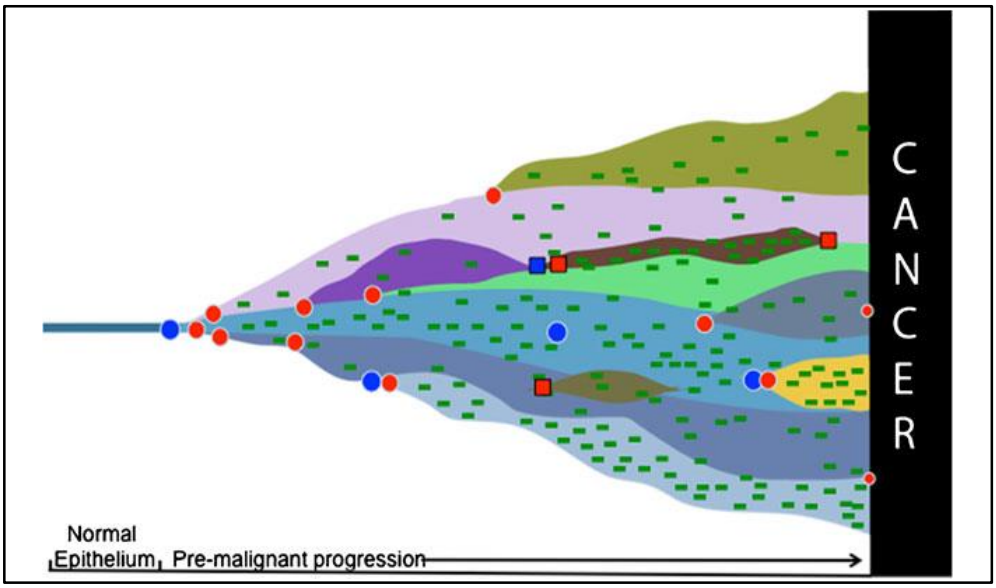


Figure 3. *Geographically delineated different mutations in different regions of the tumor* Blue circle represent mutator mutations that result in an elevated mutation rate. Red circles indicate driver mutations that confer a phenotype that allows a given lineage to overcome a particular barrier to progression. Green bars represent passenger mutations accumulated as tumor progression proceeds which do not in themselves confer selected phenotypes, however do contribute significant genetic heterogeneity. Blue and red squares represent mutator and driver mutations that result in outcompeted lineages that do not progress to malignancy. (Adapted from Edward J. Fox et al., 2013).

The mutator phenotype model suggests different approaches for therapeutic intervention: as the rate of mutation fuels the disease progression, strategies that attenuate the accumulation of mutation could dramatically slow the rate of development of early stage cancer. On the other hand, by inducing

excessive levels of mutation, it might be possible to treat certain tumors by lethal mutagenesis.

b) Oncogene induced DNA damaged model: This model developed initially from the observation that both precancerous and cancerous lesions exhibit a persistent DNA damage response indicating the presence of DSBs (Bartkova et al., 2005; Gorgoulis et al., 2005). This model proposes that oncogene-induced genome instability in the early stage of cancer development is due to DNA replication stress, resulting in DNA double strand breaks (Halazonetis et al., 2008). Common fragile sites (CFSs) are specific genomic sites that are particularly sensitive to DNA replication stress, suggesting that genome instability preferentially affect CFS in precancerous lesions and also in several models in which oncogenes have been activated. It is noteworthy that, both models (i.e., mutator hypothesis and oncogene induced DNA damage), stress genome instability as a vital factor for tumor development.

3. Genome instability and its causes

Accurate and complete replication of DNA in every cell cycle and repair of DNA lesion are critical for maintenance of genomic stability in the cell (Aguilera and Gomez-Gonzalez, 2008; Branzei and Foiani, 2008). Under physiological conditions, the genome has a natural tendency to undergo genome alternation by various types of endogenous and exogenous sources. Damage to DNA linked to various DNA metabolic pathways is another serious threat to genome integrity, being transcription and replication major sources of chromosome breakage. Replication dysfunction due to replication stress or replication errors seems to be the main cause of genome instability (Gorgoulis et al., 2005; Kunkel 2004; Aguilera and Gomez-Gonzalez, 2008; Halazonetis

et al., 2008). Mostly fragile site and highly transcribed regions are accountable for dysfunction of replication (Durkin and Glover, 2007; Aguilera and Garcia-Muse 2012).

3.1 Replication dysfunction: a major culprit of genome instability

Although highly coordinated during the cell cycle, DNA replication is a particularly dangerous process, as it is very susceptible to endogenous and exogenous events that can interfere with the progression, stability and restart of replication forks (Branzei and Foiani, 2005). The replication fork must frequently overcome various structurally unrelated hurdles like DNA lesions, non-histone proteins tightly bound to DNA, peculiar DNA sequences causing secondary structures such as cruciform structures and possibly G-quadruplexes, nucleotide pool imbalance and conflicts with the transcription machinery (Mirkin EV and Mirkin SM, 2007; Lambert S, Froget B, Carr AM, 2007). If the fork is not able to bypass the lesion and restart the replication, the cell must depend on an incoming replication fork from the opposite side to complete replication. Replication difficulties or obstacles may have stronger consequences in DNA regions with a paucity of replication origins or forks that replicate at the end of S phase (Durkin and Glover 2007; Letessier et al., 2011). Each replication fork is associated with a replisome, which consists of the replicative helicase and polymerases, primases and other accessory factors (Labib and Hodgson, 2007). All the above mentioned obstacles can impede replication by uncoupling the replicative polymerases, and helicases resulting in the production of large segments of ssDNA (“replication fork stalling”). In addition, replisome disassembly can occur with the consequent generation of SSBs and DSBs (Sogo et al., 2002; Lopes et al., 2001; Cotta-Ramusino et al. 2005). Processing of this aberrant structure is required before replication can

and this processing may lead to undesired recombination events, eventually leading to increased genome instability (Cotta-Ramusino et al., 2005; Branzei and Foiani, 2007 and Tourriere and Pasero, 2007).

Alternatively, a template strand switch mechanism may allow the bypass the replication-blocking lesion. This mechanism includes conversion of the fork to a Holliday junction (known as the 'chicken foot' structure) by branch migration (Postow et al., 2001; Sogo et al., 2002). In the chicken foot structure, an elongated lagging strand provides a template for extension of the leading strand beyond the point to the replication block of the parent template. After extension of the leading strand by DNA polymerase, the chicken foot structure can be reset by reverse branch migration into the fork, so that the lesion is bypassed. Alternatively this structure can be cleaved resulting in one-end double strand breaks (Jaktaji and Lloyd, 2003; Heller and Marians, 2006) and this one ended DSBs can restore replication and may result in a second Holliday junction.

There are several different pathways and processing mechanism of DSBs (Aguilera and Gomez-Gonzalez, 2008). DSBs can be repaired either by homologous recombination (HR) in a process which requires a homologous partner, or by non-homologous end-joining (NHEJ) in situations where no sister chromatid is available with a homologous DNA sequence (For details about the repair pathways, please see section: DNA repair pathways). The initially formed DSBs can be resected by nucleases followed by HR or synthesis-dependent strand annealing (SDSA), yielding interstitial deletions, duplications, reciprocal translocations and inversions; break-induced replication (BIR) results mainly in non-reciprocal translocations, but also interstitial deletions and inversions; repair by single-strand annealing (SSA) causes interstitial deletions. Conversely, direct repair by non-homologous end

joining (NHEJ), may lead to interstitial deletions, insertions, inversions and translocations, while de-novo telomere addition results in a terminal deletion.

3.2 Fidelity of DNA replication

Watson and Crick showed that each strand of a parental DNA helix acts as template for daughter strands DNA synthesis following the rules of base complementarity (Watson and Crick, 1953). In eukaryotic cells, DNA synthesis is catalyzed by replicative polymerases (B-family polymerases α, δ, and ε for nuclear DNA and A-family polymerase γ for mitochondrial DNA synthesis). High nucleotide selectivity is achieved by hydrogen bonding

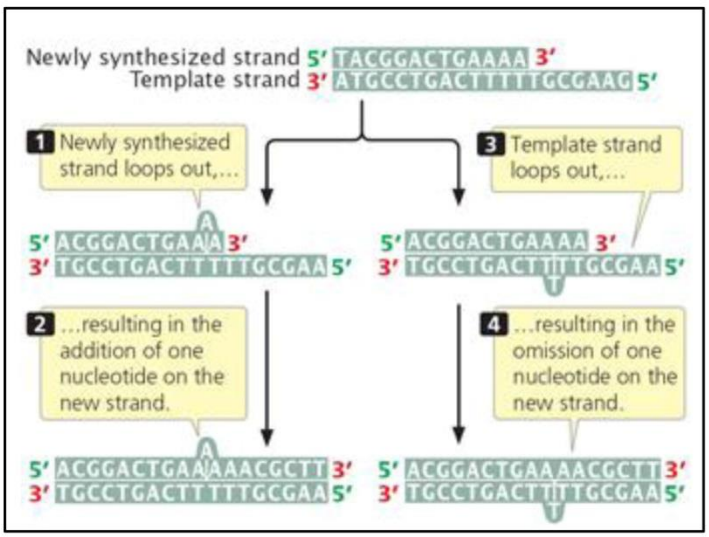


Figure 4. *Strand slippage may result in Insertion and deletion* (taken from W. H. Freeman Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed., 2005)

between the template bases, solvent exclusion from the active site, and geometric selection of shape and size of the nascent base pair in the catalytic site (McCulloch and Kunket, 2008). All the four above mentioned

polymerases have high accuracy for the insertion of the correct nucleotide, generating ~ one error every 10,000 correct incorporation events (Scott D. McCulloch et al., 2008). Moreover, insertions and deletions may result from strand slippage (See Fig.4). Remarkably, many mammalian DNA polymerases (pols δ, ε, and γ) have intrinsic 3′ exonucleolytic proofreading activity by extending mismatched primer termini less efficiently than the matched primer (Kunkel, 2004). The intrinsic error rate for any given DNA polymerase is an important feature of DNA replication because uncorrected errors during DNA synthesis lead to the generation of mutations. The estimated base substitution error rate of the replicative polymerase in vivo is in the range of 10^{-7} to 10^{-8} (Schaaper, R. M, 1993; Loeb, L. A, 1991).

3.2.1 Ribonucleotide misincorporation in DNA

Modification of nucleotides in DNA poses a threat to genome integrity of cell, often resulting in mutation or cell death. The presence of ribonucleotides in the DNA backbone appears to be one of the most common threats to genomic stability. Ribonucleotides are incorporated in DNA by DNA polymerase (Pol α, δ, and ε): Although DNA polymerases can discriminate ribonucleoside triphosphates (rNTPs) efficiently, they are incorporate into DNA likely because the cellular rNTPs concentration is 10 to 100 fold higher compared to dNTPs concentration (Nick McElhinny SA et al., 2010a). Studies of the yeast replicative polymerases estimate that more than 10^4 rNTPs may be stably incorporated into the genome during one round of replication (Nick McElhinny SA et al., 2010a; Nick McElhinny SA et al., 2010b). Ribonucleotide monophosphates (rNMPs) incorporated in the nuclear genome may promote genome instability in several ways. Primarily, compared to RNA, DNA is inherently more resistant to strand cleavage due to the absence of a reactive 2′ hydroxyl on the ribose ring. Therefore, RNA is 10^5

fold or more prone to hydrolysis than DNA (Li Y. and Breaker, 1999). Furthermore, the presence of rNMPs can cause a distortion of the DNA helix parameters (Jaishree et al., 1993; DeRose et al., 2012).

3.3 Defective nucleosome assembly and remodeling

Chromatin is the complex of DNA and cellular proteins, which form eukaryotic chromosomes. It is composed of an elementary repeating unit called the nucleosome, which is the major factor of DNA packaging in eukaryotic genomes. Nucleosomes are DNA-protein complexes, which are comprised of a core particle of 1.6 left-handed turns of DNA (roughly 146 bp) wound around a protein complex called the histone octamer. De novo nucleosome assembly strictly depends on DNA replication progression. Deregulation of nucleosome assembly causes replication and checkpoint defects, resulting in recombinogenic DSBs and ssDNA gaps (Ye X, 2003) as well as hyper-recombination (Prado F, 2005). The size and distribution of Okazaki fragments due to defects in chromatin assembly suggest that instability might be caused by dysfunctional lagging-strand processing (Smith DJ, 2012). Finally, we cannot exclude the fact that defects in DSB repair drive to nucleosome assembly dysfunctions, resulting in genome instability (Alabert C, 2012).

3.4 Secondary Structures

Right-handed double helical conformations are adopted by DNA in most cases, but specific sequences can also allow the formation of alternative DNA structures (Gacy A.M. et al., 1995; Mirkin and Mirkin, 2007; Mitas et al., 1995; Moore et al., 1999; Wells, 1996). Several factors can influence the formation of alternative DNA structures. Inverted repeats (IR) have the tendency to form ssDNA hairpins and dsDNA cruciform structures. Triple-

helical DNA, called H-DNA, conformation can be adopted by mirror repeats (MR), whereas G-quadruplexes, the left-handed Z-DNA conformation and the slip stranded DNA conformation of S-DNA is the outcome of direct tandem repeats (DTRs) (Mirkin and Mirkin, 2007). Early studies show that trinucleotide repeats (TNR) are conserved in both mammals (Shiraishi et al., 2001) and yeast (Cha and Kleckner, 2002; Lemoine et al., 2005; Raveendranathan et al., 2006).

Instability in TNRs is directly related to secondary structure formation. These sequences can form stem-loops, hairpins and triplexes on the leading strand representing physical barriers that can perturb DNA synthesis by either causing slippage or fork stalling, which is enhanced under replication stress (Aguilera and Gomez-Gonzalez, 2008). Additionally, secondary structures on the lagging strand, have been shown to promote expansions (Aguilera and Gomez-Gonzalez, 2008). Regardless of the specific type of DNA structure present, these regions represent obstacles for normal fork progression, which ultimately may result in genomic instability.

3.5 Challenges to replication machinery

Replication fork encounters various numbers of challenges as it progresses along the chromosomes. Fragile sites are associated with chromosome breakage and genomic rearrangement, as they can induce fork stalling (Cha and Kleckner, 2002). These sites include tRNA genes (Deshpande and Newlon, 1996), slow replication zones (Cha and Kleckner, 2002), inverted repeats (Lemoine et al., 2005) and specialized protein-mediated replication fork barriers (Branzei and Foiani, 2007; Takeuchi et al., 2003).

3.5.1 Fragile sites

Fragile sites were defined originally by Magenis and colleagues in 1970, describing the recurrent chromosome breakage on the long arm of human chromosome 16 (Magenis et al., 1970). Human chromosomes have nearly 120 fragile sites that were observed and named according to the band chromosome (Debacker et al., 2007; Lukusa et al., 2008). Chromosomal fragile sites are heritable specific loci that preferentially show instability, visible as non-random gaps and breaks on metaphase chromosomes. They are mainly associated with rearrangement (like translocations, integration of exogenous DNA and gene amplifications). Fragile sites are conserved among mammals (Arlt et al., 2003) and were also found in lower eukaryotes like the yeast *S. cerevisiae* (Cha and Kleckner, 2002; Ivessa et al., 2003; Lemoine et al., 2005; Anne Helmrick, 2008). Fragile sites are regions likely vulnerable to breakage after low replication stress and might be one of the driving force in cancer progression (Glover TW et al, 2013). Inhibition of DNA polymerase and decrease in the dNTP pools lead to replication stress and interference in the completion of replication (Ikegami et al., 1978; Wist and Prydz, 1979). Based on the population frequency and mode of induction, the fragile sites are classified into “common fragile site” and “rare fragile sites”.

3.5.1.1 Rare fragile sites

Only less than 5% of the fragile sites in the human genome are rare (Kremer et al., 1991; Sutherland et al., 1998). According to normal Mendelian patterns of inheritance, the rare fragile sites are passed from parents to offspring. Fragile sites are usually associated with trinucleotide repeats (TNRs) or with long AT-rich repeats, and these repeats are mostly associated

with genetic disease (Durkin and Glover, 2007; Lopez Castel et al., 2010). Their fragility is linked to repeats expansion or contraction.

3.5.1.2 Common fragile sites

In contrast with rare fragile sites, common fragile sites account for 95% of all know fragile sites and are present in all individuals as a peculiarity of normal human chromosomes (Glover et al., 1984). Yeast also contains fragile site-like sequences, suggesting an evolutionary significance (Durkin and Glover, 2007). Common fragile sites (CFSs) are AT-rich sequences (*See section: Secondary structure & fragile sites*), but do not contain expansions of the specific repeated sequences seen in rare fragile sites (Debacker and Kooy, 2007). Genomic instability, a hallmark of cancer, occurs preferentially at CFSs. Common fragile site breakage has been identified after treatment with various replication inhibitors (Glover, T.W et al, 2007). Several factors contribute to CFSs instability; these factors are both intrinsic characteristics of fragile regions and events that interfere with the replication process (See Fig 5). Recent studies suggest that transcription might contribute to the fragility of CFS. Mapping of the majority of CFS in the coding regions of a large number of genes and the finding that transcription of such genes requires a long time to be completed, suggest that in these regions transcription and replication may occur at the same time. In this case, the transcription machinery and replication forks may collide, resulting in replication fork impairment (A. Aguilera and T. Garcıa-Muse, 2012; A. Helmrich, M. Ballarino, and L. Tora, 2011). Additionally, CFSs are also preferably involved in sister chromatid exchange (SCE), deletions and translocations (Glover, T.W et al., 1987; Glover, T.W, 1988; Wang, N.D, Testa, J.R., Smith, D.I, 1993; Chan, K.L et al., 2009).

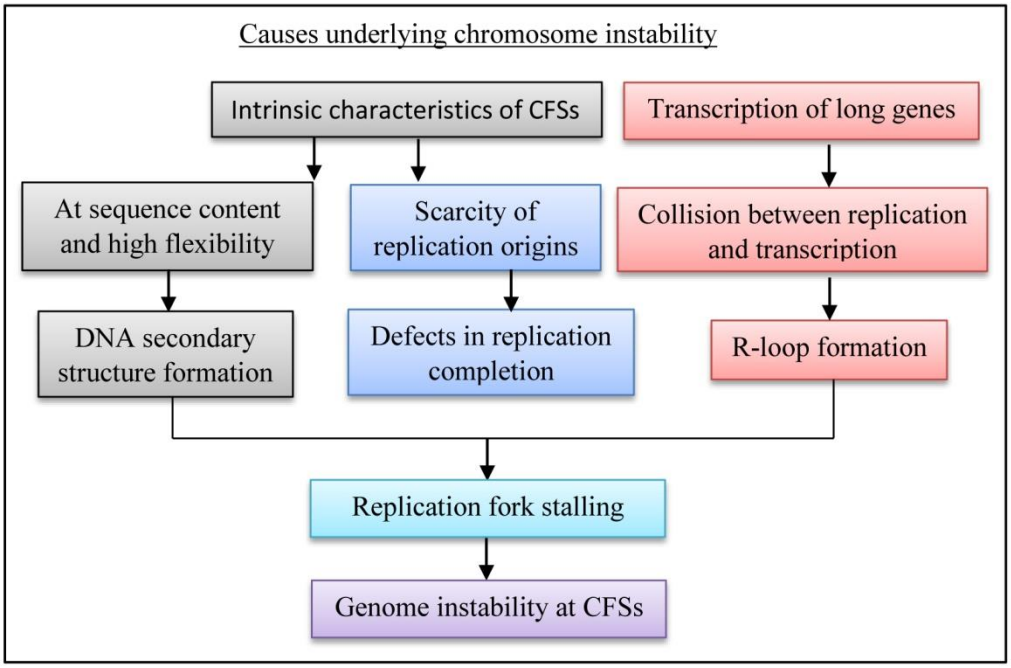


Figure 5. *Potential source underlying CFS fragility and the final impact on genome stability*

3.5.2 Replication fork barriers

During normal chromosome replication, replication fork can pause naturally where particular proteins are tightly bound to DNA (Labib et al., 2007). These specific replication fork arrest sites are called "Replication Fork Barrier (RFB)", and include DNA bases and intrinsic RFBs. Different forms of damaged DNA bases (See section endogenous and exogenous DNA damage) and bulky adducts are obstacles to replicative polymerase. In the presence of these obstacles replicative polymerases are unable to incorporate nucleotides. Inter-strand cross links (ICLs), are obstacles to replication fork progression, as they prevent the DNA duplex to unwind ahead of the work (Dronkert and Kanaar, 2001). Early studies showed that budding yeast cells have to face more than 1400 natural RFBs caused by DNA-proteins per

replication cycle (Ivessa et al., 2003). Replication fork arrest has been shown in yeast and bacteria system by introducing the exogenous protein binding sequences into the genome combined with the expression of the corresponding binding protein (Sofueva et al., 2011; Possoz et al., 2006). These experiments show that the frequency of replication fork arrest is increased when the ancillary replicative helicase is lost (Sofueva et al., 2011) and the arrest is dependent on the strength of DNA-protein association (Dubarry et al., 2011).

3.5.3 Transposable elements

Repetitive elements are one of the substrates frequently involved in genomic rearrangement. As nonallelic homologous sequences, transposable elements (TEs) have the ability to intrfer with essential DNA repair processes often leading to genome alterations. There are varieties of rearrangements involving TE elements, ranging from mutations to inter-or intra-chromosomal alteration. There are approximately 25 distinct human genetic diseases that are related to TE element rearrangements resulting in genome instability. The role of TE elements in the fragile site related genome rearrangements is still obscure, but recombination is a source of genomic rearrangement between TE elements (Lemoine et al., 2005). Retrotransposons are eukaryotic mobile elements that transpose through RNA intermediates and retrotransposon (RT) integration is a potential source of mutagenesis (Scholes et al., 2001).

It has been shown that reduced levels of replicative DNA polymerase α in yeast causes chromosome translocations (Lemoine et al., 2005). Interestingly, the breakpoints of these rearrangement events were mapped within Ty elements, especially to those elements in a head-to-head conformation (Lemoine et al., 2005). Therefore, the elevated rate of breakage under replication stress at retrotransposons suggests that these sites may, in fact, be another type of fragile site. Recent studies in budding yeast demonstrated high

rates of chromosome aberrations with breakpoints corresponding to Ty or LTRs (Vernon et al., 2008). While the exact mechanism that generates chromosomal rearrangements at these sites remains unclear, two different models are likely to play a role.

First, Ty elements transpose very near a second Ty element, producing an inverted repeat, which can form a secondary structure interfering with fork progression and generating a high rate of chromosome rearrangements (Lemoine et al., 2005). Second, retrotransposons are commonly dispersed repetitive sequences throughout the genome that, through ectopic recombination may give rise to GCR events (Lemoine et al., 2005; Umezu et al., 2002). Additionally, LTR sequences have been shown to insert at DSBs as a repair mechanism (Moore and Haber, 1996). Therefore, in addition to facilitating genomic rearrangements, they may also act as a marker for genomic sites that are prone to breakage (Admire et al., 2006).

4. Transcription linked genome instability

"Transcription" is the synthesis of RNA from DNA. Increasing evidence in the last three decades has shown that transcription is an important source of genome instability. Studies in yeast (specifically in budding yeast), shows that transcription stimulates spontaneous mutation in eukaryotes (Datta, A and Jinks-Robertson, 1995), a phenomenon that is known as transcription associated mutation (TAM) (Datta, A and Jinks-Robertson, 1995; Beletski and Bhagwat, 1996), while the increased recombination induced by transcription is called transcription association recombination (TAR) (Thomas and Rothstein, 1989; Nickoloff, 1992). TAR and TAM together are termed as Transcription-associated genome instability (TAGIN) (Gaillard H. et al., 2013). Recent studies demonstrate that TAM primarily reflects damage to the non-

transcribed strand (NTS) of the DNA template, whereas TAR is largely due to transcription-replication collision.

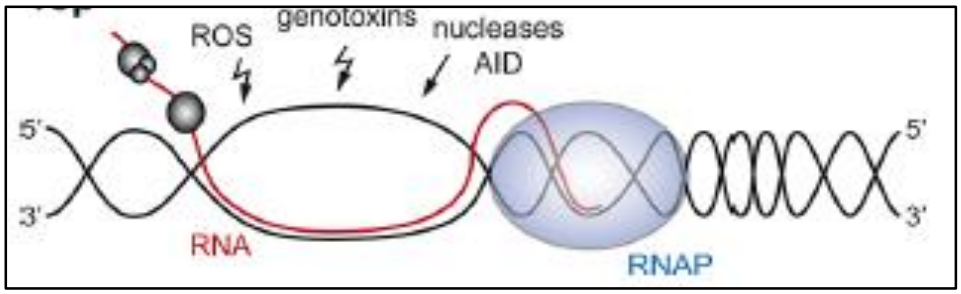


Figure 6. *Model to explain the R loop formation facilitate the exposure of NTS to genotoxic agents* (adapted from **Gaillard**, H. 2013)

TAM mechanism: TAM depends strictly on the transcription process itself and not on the potentially beneficial changes in the polypeptide sequence (Francino, M. P et al., 2001; Green, P et al., 2003; Mugal, C.F, 2009).

During transcription, the newly synthesized RNA molecule remains transiently paired with the transcribed DNA strand (TS) forming a 9-12 nucleotides long RNA: DNA hybrid. Hence, the short complementary non-transcribing strand (NTS) remains unpaired and, therefore, single stranded increasing its vulnerability towards nucleolytic attack and DNA damaging agents. Moreover, the process of transcription triggers topological changes, including negative supercoiling behind the RNAP (RNA polymerase) (Liu, L. F. et al., 1987), which favors the formation of ssDNA containing vulnerable unpaired bases, and secondary structures such as stem-loops or other forms of DNA structure (See Fig. 6). Alternative mechanisms supporting TAM may be linked to the observation that some DNA repair machineries do not workproperly on non-B DNA structures. For example, high levels of transcription reduce the efficiency of MMR on plasmid-based microsatellite and increase the rate of DNA synthesis errors (Wierdl, M et al., 1996).

TAR mechanism: TAR results from the collision between the transcription and replication machineries. Transcription interferes with replication fork progression, thus promoting recombination. Impaired replication fork progression which can be due to the formation of R-loops, or the collision between replication and transcription complexes can cause replication fork impairment, DSBs and TAR (Prado and Aguilera, 2005; Gottipati et al., 2008; Azvolinsky et al., 2009), leading to genome instability. Previous research showed in yeast that cotranscriptional R-loops are an important mediator of transcription-associated instability (Huertas P, Aguilera A. 2003). Transcription and co-transcription R-loops may determine some hotspot for genome instability for e.g., trinucleotide repeats (TNRs) (Grabczyk E et al., 2007; Lin Y et al., 2010) and fragile sites as well as sites of programmed instability, such as class switch recombination (CSR) in the immune system. Therefore, R-loops can be a principal cause of genome instability (Aguilera, A & García-Muse T, 2012).

5. Telomeres and genome instability

Telomeres are the nucleoprotein complexes that stabilize the chromosomal ends preventing them from aberrant recombination and from being recognized as a double-strand breaks (Maser, R. S. and DePinho, 2004). Mammalian telomeres contain up to two thousand repeats of the 5'-TTAGGG-3' sequence (Moyzis R. K, et al., 1988). Telomeres have a unique structure whereby the terminal end of the DNA loops back and inserts it into the terminal telomeric repeat sequence, known as the "T-loop" (Griffith, J. D et al., 1999) via a short 3' overhang with G rich tails invading the duplex and

forming a D-loop (displacement loop). Telomeric proteins stabilize the D-loop structure and seem to function in protecting the telomere structure from nuclease degradation and recombination. Loss of telomeric DNA or telomere protection leads to telomere dysfunction and activation of the DNA damage response pathways (Karlseder J., Smogorzewska., Dai Y., Hardy S, and de Lange, T. 1999; Celli G. B, and de Lange T, 2005; Karlseder J, Smogorzewska A, and de Lange, T. 2002).

Several factors contribute to telomere shortening and telomere dysfunction, for example, oxidative stress (Kawanishi, S. and Oikawa, S. 2004; Richter, T. and Proctor, C., 2007), stochastic deletion (Baird D. M et al., 2003), or the "end-replication" problem of chromosomal ends (Ohki R, 2001). Since the DNA polymerase replicates the DNA only in 5' to 3' direction, normal lagging strand DNA replication fails to copy the 5' end of the chromosome, thus leaving a gap between the final RNA priming event and the terminus (Harley C. B, 1990; Lindsey J et al., 1991): this leads to progressive telomere shortening of about 50-200 bps at each cell division (Harley C. B, 1990), finally resulting in loss of DNA function.

Genome instability may arise due to the loss of telomere function, which further leads to dysfunction in genes responsible for genome stability. Due to the nature of chromosome ends which are sticky, end-to-end fusions may occur (with the formation of dicentric chromosomes and anaphase bridges. Anaphase bridges facilitate chromosome instability with fusion and rearrangements through "break fusion-bridge cycles" (BFB) (McClintock B, 1941; Mathieu N et al., 2004). Several studies showed that genetic intra tumor heterogeneity is caused by break fusion-bridge cycles (Gisselsson D et al., 2000; Gisselsson D et al., 2001).

6. The DNA damage response

To counteract the effects of genomic DNA damage and to ensure a faithful duplication and inheritance of the genetic material, eukaryotic cells have evolved a complex network of cellular responses, collectively known as "DNA damage response" (DDR), which impacts on cell cycle progression, DNA repair, senescence and apoptosis (See Fig 7). I will discuss the DNA damage checkpoints and DNA repair pathways separately, although they share many components and functions.

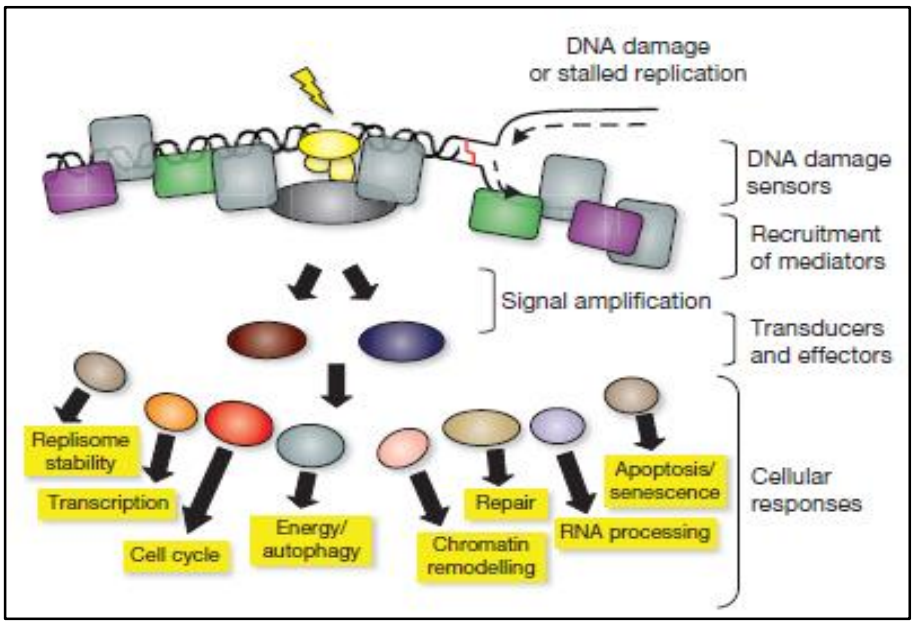


Figure 7. *DNA damage response* (taken from Stephen P. Jackson & Jiri Bartek, 2009)

7. DNA damage checkpoints

Whenever there is a DNA break or a DNA lesion, DNA damage response (DDR) proteins sense the alterations and transmit a signal to activatesignaling pathways often called DNA damage checkpoints. A

checkpoint can be considered as a surveillance mechanism, which not only plays a regulatory role in the transition from on cell cycle phase to the next one, but also coordinate the crosstalks among various cellular processes, like DNA repair and apoptosis. Although the DNA damage checkpoint pathways may partially differ in various cell cycle phases, the general scheme of the DDR pathway is a highly conserved process (See Fig 8): the majority of its components shows a high degree of functional homology among eukaryotic organisms from yeast to man. *(See fig. 8 for a comparison between yeast and human proteins)*

	Budding yeast	Fission yeast	Human
PIKK	Mec1	Rad3	ATR
PIKK	Tel1	Tel1	ATM
Adaptor	Rad9	Crb2	53BP1, MDC1, BRCA1?
Rfc1 homolog	Rad24	Rad17	Rad17
9-1-1 clamp	Rad17	Rad9	Rad9
	Mec3	Hus1	Hus1
	Ddc1	Rad1	Rad1
MRX complex	Mre11	Mre11	Mre11
	Rad50	Rad50	Rad50
	Xrs2	Nbs1	Nbs1
BRCT domain adaptor?	Dpb11	Rad4/Cut5	TopBP1
Signaling kinase	Rad53	Cds1	Chk2
Signaling kinase	Chk1	Chk1	Chk1
Polo kinase	Cdc5	Plo1	Plk1
Securin	Pds1	Cut2	Securin
Separase	Esp1	Cut1	Separase
APC-targeting subunit	Cdc20	Slp1	$p55^{CDC}$/CDC20

Figure 8. *Schematic representation of DNA damage checkpoint proteins conserved from yeast to human* (taken from Harrison, 2006)

The DNA damage checkpoint works throughout the cell cycle and responds to almost all kinds of internal and external threats to the genome. There are three major DNA damage checkpoints in cells, which respectively delay the G1/S transition, arrest cells at G2-M phase or slow down S phase progression upon induction of DNA damage to allow time for repair. Early evidence shows that

dysfunction of all these checkpoint pathways leads to genomic instability. Within the DNA damage checkpoint signaling pathway, we can classify four groups of proteins that are often classified as sensors, mediator, transducers and effectors (See Fig. 9).

7.1 Sensors of DNA damage

Ataxia-telangiectasia mutated (ATM or Mec1 in *S. cerevisiae*/yeast), and Ataxia-telangiectasia and Rad3-related (ATR or Tel1 in yeast) are the two apical kinases of the DDR pathways and they belong to the phosphoinositide-3 kinase-related protein kinase (PIKK) family.

Function	Class	Gene
Sensors	RFC-like	*RAD17, RFC2-5*
	PCNA-like	*RAD9, RAD1, HUS1 (9-1-1)*
Mediators	BRCT-containing	*BRCA1, 53BP1,TopBP1, MDC1, Claspin*
	DSB recognition/repair (MRN complex)	*Mre11, RAD50, NBS1*
Transducers	PI3 kinase-like protein	*ATR, ATM*
	PIKK binding protein	*ATRIP*
	Protein kinase	*CHK1, CHK2*
Effectors	Transcription factor	*p53*
	Phosphatase	*CDC25A, B, C*
	Protein kinase	*CDKs, CDC7*

Figure 9. *Classification of human genes involved in DNA damage checkpoint*

ATR/Mec1 is crucial for signaling ssDNA at DNA lesion and stalling replication fork, while ATM/Tel1 signals DSBs. After DNA damage, the first step is the recognition of the lesion and activation of the signaling cascade. The 9-1-1 and RFC-like complexes are responsible for the activation of the

checkpoint signal both in yeast and mammals (Parrilla-Castellar, ER, 2004; Melo J. and Toczyski D, 2002; Paulovich AG, 1998 and Longhese M.P, 1998). The Rad9, Rad1 and Hus1 proteins form a heterotrimeric complex, the 9-1-1 complex, whose structure resembles that of the proliferating cell nuclear antigen (PCNA)-sliding clamp (Shiomi Y, 2002). Additionally, the Rad17 protein substitutes RFC1 in the interaction with the other RFC subunits thus forming the RFC-like complex, which acts as a clamp loader complex to bring the PCNA-like clamp near the DNA lesion on damaged DNA (Griffith JD, 2002; Kondo T, 1999; Green CM, 2000; Naiki T, 2000). Therefore, when DNA damage is occurring, the 9-1-1 complex is recruited to the damage site with the help of the Rad17 complex. Then, the chromatin bound 9-1-1 complex is phosphorylated by the apical kinases ATR or ATM. ATR (in yeast, Mec1), and its interacting partner ATRIP (in yeast Ddc2) is recruited to the site of DNA damage independently on Rad17 and 9-1-1 complexes. Ddc2/ATRIP binds RPA-coated ssDNA and recruits Mec1/ATR to the site of DNA damage (Kondo T, 2001; Melo J. A, 2001). Alternativily, in response to DSBs, the MRN complex (Mre11, Rad50, Nbs1) acts as the main sensor recruiting the ATM apical kinase through interaction with its Nbs1 subunit.

7.2 Mediators:

The mediator proteins are, Breast cancer 1, early onset (BRCA1) and p53 binding protein 1 (53BP1), which are the homologs of Rad9 in the yeast S. *cerevisiae* (Schultz L.B, 2000; Wang B, 2002; DiTullio R.A, 2002; Ward I.M, 2003 and Saka Y, 1997), the topoisomerase binding protein 1 (TopBP1) (Yamane K, 2002), the homolog of yeast Dpb11, and the MRN complex (Goldberg M, 2003; Lou Z, 2003 and Stewart G.S, 2003), (MRX in yeast). BRCT domain mediator proteins help in targeting activated ATM to sites of DNA damage and multiprotein interactions controlled by these mediators

facilitate ATM signaling (Kitagawa R, 2004; Uziel T, 2003; Carson C.T, 2003; Horejsi Z., 2004; D'Amours D, and Jackson S.P, 2002; Petrini J.H, and Stracker T.H, 2003).

7.3 Transducers

ATM and *ATR*

In mammalian cells the ATR and ATM apical kinases, activate the downstream effector kinases Chk1 and Chk2 and many other protein factors that modulate processes such as cell cycle, DNA replication, DNA repair and apoptosis, among which phosphatases of the Cdc25 family, activators of cyclin /Cdk complexes, the p53 transcription factor and others (Langerak and Russell, 2011). Yeast cells lack p53 and do not have a robust apoptotic pathway that can eliminate damaged cells. In yeast, Tel1 (the homolog of ATM) plays a minor role in DSB repair, but it is primarily involved in telomere maintenance. ATM is present as an inactive homo dimmer form and, upon DSBs formation, it undergoes a conformational change, which leads to intermolecular phosphorylation at serine1981, causing dimer dissociation. The activated monomer is now ready to act on its numerous downstream substrates, like p53, Nbs1, Brca1 and Smc1 (Bakkenist C J, 2003).

While ATM is mostly activated in response to DSBs, the ATR kinase plays an essential role in response to damage caused by UV irradiation and in replication stress. Moreover, ATR-knockout mice are embryonic lethal suggesting a role for ATR in normal cellular function. Known ATR targets are numerous, including the effector kinase Chk1, Rad17, TopBP1, RPA, ATRIP, 9-1-1 and Claspin.

7.4 Effectors

The apical ATM/ATR kinases and their effector kinases Chk2/Chk1 through phosphorylation of multiple targets transiently delay cell cycle progression thus ensuring the accuracy of replication and transmission of DNA after DNA damage.

7.4.1 G1 and G1/S cell cycle checkpoint

In response to DNA damage, G1 cells delay entry into S phase, to repair DNA lesions, thus preventing the replication of damaged DNA, which would lead to mutations or replication fork stalling or collapse (Liu et al., 2012). There are two pathways that activate the G1/S checkpoint. The first pathway acting through Chk2 and Cdc25 blocks the loading of Cdc45 into the pre-replication complexes and, consequently, prevent the firing of replication origins. However, the arrest induced by Chk2-Cdc25 is only transient (Deckbar et al., 2011) and cells eventually enter S-phase. However, a second control pathway acts through the phosphorylation of p53 at serine 15 and 20 within its amino-terminal transactivation domain. Meanwhile, the ubiquitin ligase for p53, Mdm2, is targeted by ATM (ATR) /Chk2 (Chk1) for phosphorylation, leading to a loss of its ability to ubiquitylate and degrade p53. All these modifications contribute to both the stability and activity of p53 as a transcription factor (Bashkirov V. I et al., 2003). Activated p53 induces the transcription of p21, which is an inhibitor of cyclin E/Cdk2 complex. In addition, transcription of genes essential for DNA replication is suppressed as p21 inhibits phosphorylation of Rb protein, thus preventing the release and activation of the E2F transcription factor (Falck et al., 2001).

7.4.2 Intra S-phase checkpoint

The intra S-phase checkpoint decreases the rate of DNA synthesis following DNA damage during replication and it monitors replication fork stalling or collapse. The RPA bound to the ssDNA accumulated in conditions of replication stress recruits the sensor kinase ATR through the RPA-ATRIP interaction (Petermann et al., 2010). Fully activated ATR phosphorylates Chk1 that in turn phosphorylates effector proteins required to stabilize stalled forks, repair collapsed forks, and inhibit late origins firing to prevent further encounter of replication forks with DNA lesions (Ciccia and Elledge, 2010). Once DNA lesions are repaired by NHEJ or HR, forks may restart with the help of HR proteins (Budzowska and Kanaar, 2009). Alternatively, TLS polymerases are recruited to the lesions through the action of monoubiquitinylated PCNA and are able to bypass the lesions in DNA (Moldovan et al., 2007). If stalled forks cannot restart in time, DSBS may be generated, which are sensed and repaired as a canonical DSBs (Chanoux R.A et al., 2009).

7.4.3 G2/M checkpoint

The G2 checkpoint acts mainly to prevent cells from entering M phase with damaged DNA and this is the last opportunity for the cells to repair the lesions before passing the genome to the daughter cells. In late S and G2, the sister chromatid is available for recombinational repair, providing the highest efficiency for lesion removal. Both ATR and ATM pathways are triggered in G2, and target Cdc25C blocking mitotic entry (Furnari et al., 1997). In addition, Chk1 facilitates G2 checkpoint by inhibiting Wee1 kinase (O'Connell et al., 1997). Meanwhile, many other inhibitors of cell cycle progression such as p21, Gadd45, and 14-3-3δ are upregulated by p53 and

Brca1, leading to the reinforcement and maintenance of the G2 checkpoint (Kastan M.B., et al., 2004).

8. DNA damage repair pathways

DNA is the target of several endogenous and exogenous damaging agents causing various kinds of lesions. Cells can either repair the damage and restore the DNA structure to a normal state, or activate ceratin pathways capable to tolerate the damage. Several DNA repair mechanisms exist capable to take care of a subset of lesions. Direct repair systems such as photoreactivation and demethylation, a base excision repair system (BER) mainly involved in repairing nitrogen base lesions and single strand breaks (SSB), nucleotide excision repair (NER), which repairs lesions causing large DNA distortions such as those caused by UV irradiation, mismatch repair (MMR) which repairs mismatched bases in the DNA double helix. Double-strand break repair systems, such as NHEJ and HR which repairs DSBs.

When a replicative polymerase is blocked by unrepaired DNA lesions, DNA damage tolerance mechanisms, also known as post replication repair (PRR), can take place. In fact, DNA repair mechanisms during replication are risky. PRR can be divided in two sub-pathways: translesion DNA synthesis (TLS) involving a variety of TLS polymerases and template switching (TS), a sort of error-free recombination pathway involving the temporary annealing of the two newly synthesized DNA strands. During PRR, DNA is synthesized past the damaged bases, and eventually the lesions can be repaired after the passage of the replication forks (Sale et al., 2012). During TLS, the high fidelity polymerases are replaced by specialized TLS polymerases that can

proceed through the damaged site. TLS polymerases are error-prone because they lack exonuclease proofreading activities (Sale et al., 2012).

8.1 Repair of single-strand DNA lesions

8.1.1 Photoreactivation

Photoreactivation is one of the pathways able to remove UV-induced DNA lesions from the genome. It is also called light repair since the mechanism is dependent on a light source. Photoreactivation is found in many prokaryotic and eukaryotic organisms, but not in mammals. Cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts, two lesions caused by UV irradiation, have the ability to distort the DNA helix. CPD photolyase, specifically binds and repairs the pyrimidine dimers, while 6-4 photolyase, can bind and cleave the 6-4 photoproducts (Thoma, 1999).

8.1.2 Demethylation

The O^6-methylguanine and O^4-methylthymine are DNA lesions resulting from methylations events. The repair enzyme O^6-methylguanine-DNA-methyltransferase can directly reverse the lesion by removing the methyl group. In this process, a methyl group is transferred to a cysteine residue of the enzyme, resulting in enzyme inactivation (Sedgwick 2004).

8.1.3 Base excision repair (BER)

Base excision repair (BER) is an error-free mechanism, which repair base modifications and oxidative damage (Fortini P et al., 2003). Specialized DNA glycosylases recognize the damaged DNA base and cleave the N-glycosyidic bond between the base and the sugar component of the nucleotide (Lindahl et al., 1997), causing the formation of an AP (apurinic/apyrimidinic

site) or abasic site (See Fig 10). The AP site is cleaved by the AP endonuclease 1 (APE1) in the 5' region, creating a single-strand break (SSB), flanked by 3'-OH and 5'-deoxyribose termini. In order to complete the repair process, the blocked termini must be restored to conventional 3'-OH and 5'-phosphate ends, which are essential for DNA polymerase and subsequent DNA ligase reaction. BER can be divided in 2 sub-pathways: long patch BER in which DNA Polymerase ε can using the 3'OH terminus for extension and its strand displacement activity generates a 5' single strand overhang which is then removed by the flap endonuclease Rad27. The final step is ligation by DNA ligase (Boiteux S. and M. Guillet, 2004). The second BER sub-pathway is a short-patch repair mechanism: in this case, DNA polymerase β is inserting one nucleotide and remove the 5' extremity through its lyase activity, followed by ligation which is performed by XRCC1-Lig3 complex (Hoeijmakers J.H, 2001).

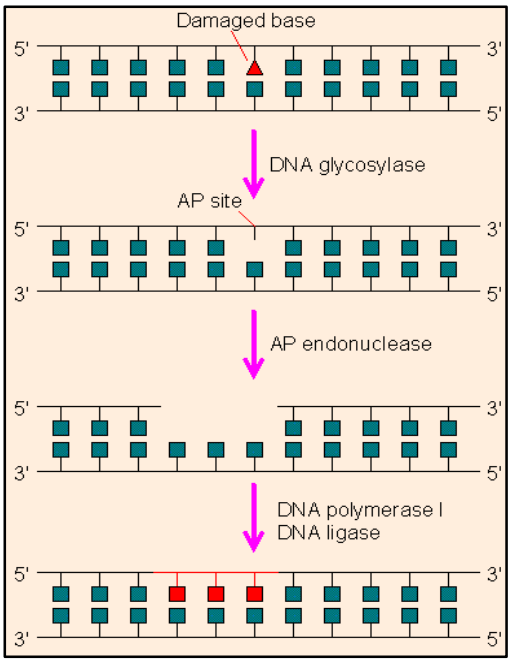


Figure 10. *Simplified version of base excision repair* (BER) (taken from http://www.web-books.com)

8.1.4 Nucleotide excision repair (NER)

NER recognizes a variety of bulky helix distorting adducts caused by chemical mutagens (e.g. anticancer compounds), UV-induced dimers (6-4 photoproducts (6-4 PP) and cyclobutane pyrimidine dimers (CPDs), alkylated nucleotides, including O^6-methyl-/or ethylguanine, N^6-methyladenine etc. The NER pathway (See Fig 11) is mediated by the sequential assembly of repair proteins at the site of DNA damage and is more complex compared to BER.

The NER system consists of two related sub-pathways: a) Transcription Coupled Repair (TCR), which removes lesions in the transcribed strand of the active genes (Hanawalt et al., 2003) and is activated by stalling of RNA polymerases and b) Global Genome NER (GG-NER), which is responsible for removing UV-induced lesions from the rest of the genome. Both the pathways share common mechanisms, but differ in the initial recognition steps. After detection of a photolesion, both the TC-NER and GC-NER follow a three step mechanism, allowing excision of an oligonucleotide containing the lesion and completion of repair by a gap filling step which allows the recovery of the lost information (de Laat, et al., 1999). Factors that detect helix-distorting lesions during GG-NER and TC-NER are in yeast Rad4/Rad23 (XPC/h Rad23b in humans), Rad7/Rad16 (functional equivalent of mammalian UVDDB1/2) and Rad26-RNA Pol II (CSB/RNA pol II). The complex Rad14-RPA (XPA/RPA) and Rad4/Rad23 work in common for both GG-NER and TC-NER, while Rad7/Rad16 is specific to GG-NER. Rad3 and Rad25, subunits of the TFIIH helicase, help to unwind the DNA before the incision step, which is carried out by the two structure- dependent endonucleases Rad1-Rad10 and Rad2 at the 5' and 3' side, respectively, of the damage. Rad14-RPA and Rad4-Rad23 complexes are also essential for

incision of the damage. In the final step, the replication machinery fills the gap and completes the repair (Prakash, S. and L. Prakash, 2000).

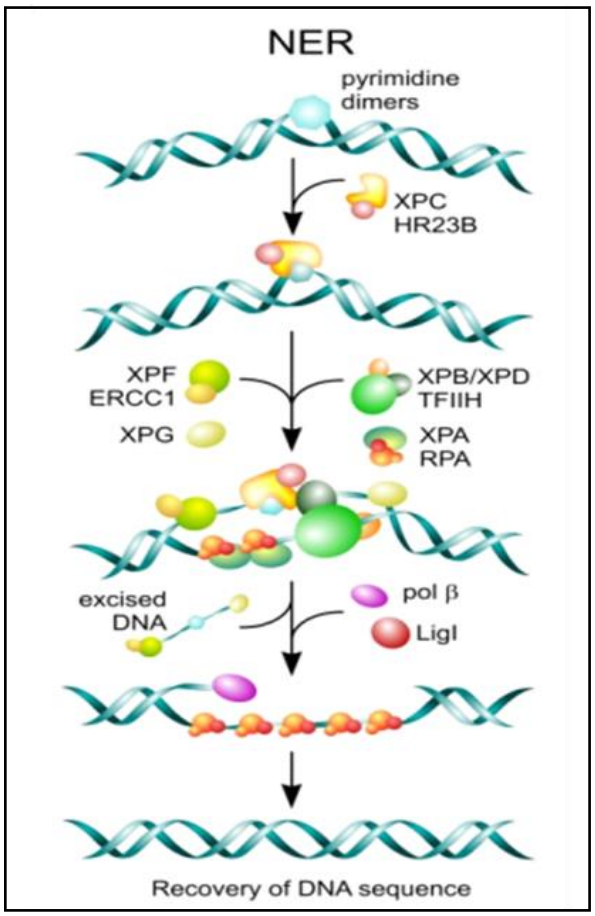


Figure 11. *Schematic representation of normal pathway of nucleotide excision repair (NER)* (taken from the book," The pathway of double stranded break" by Emil Mladenov and George Ilakis, 2011)

8.1.5 Mismatch repair (MMR)

The MMR pathway plays an important role in repairing misincorporated base (base-base mismatches) during DNA replication (See Fig 12) that have

escaped from the proofreading activity of replication polymerases and in the repair of insertion and deletion loops (IDLs) resulting from polymerase slippage during replication of repetitive DNA sequences. Initial mismatch recognition is fulfilled by two MutS activities that function as heterodimers and recruit the MutL complex. The heterodimer Msh2 and Msh6, also known as MutSα, recognizes base mismatches and small IDLs (1-2 nucleotides), while the Msh2 and Msh3 heterodimer also knows as MutSβ, detects large IDLs.

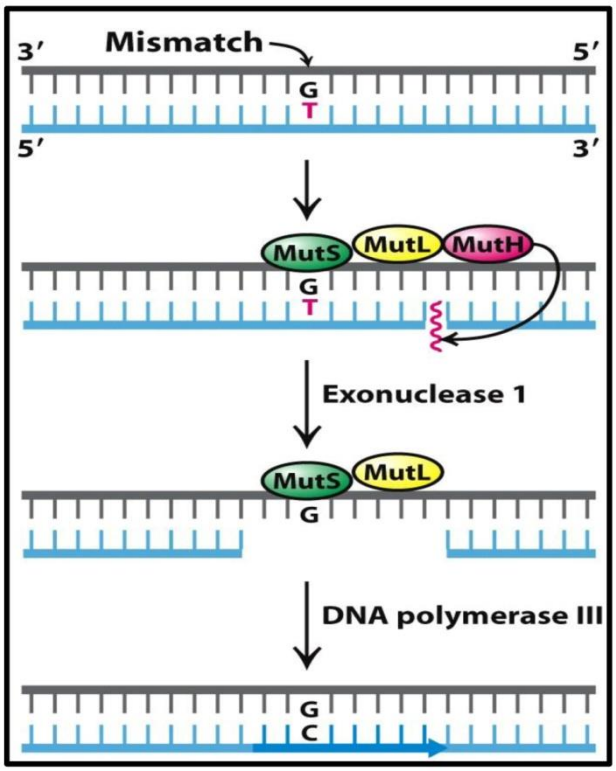


Figure 12. *Schematic representation of mismatch repair (MMR)* (taken from Kevin Aherns Biochemistry, 7th edition, 2012)

The binding on either of the two complexes (Mutα or MutSβ) induces a conformational change that attracts the Mlh1-Pms1 complex (MutLα) to the lesion. MutS/MutL complexes guide exonuclease 1 (EXO1) to the side of the damage. A nick on either side of the mismatch allows further processing by exonuclease such as Exo1 (Tran PT et al., 2004). On the lagging strand, nicks are likely due to the removal of Okazaki fragments. Then, the polymerases are thought to complete the repair process by performing the DNA synthesis and ligation steps (Harfe, B.D. and S. Jinks-Robertson, 2000).

8.2 Repair of DNA DSBs

DSBs are among the most harmful types of DNA damage. Persistent or incorrectly repaired DSBs result in GCRs, which can lead to carcinogenesis through activation of oncogenes or inactivation of tumor-suppressor genes. Thus, the repair of DSBs is critical for cell survival and maintenance of genome stability (van Gent DC et al., 2001; Khanna KK, Jackson SP, 2001). There are two main mechanisms by which mammalian cells repair DSBs: homologous recombination (HR) and non-homologous end-joining (NHEJ). Homologous recombination is an error-free repair mechanism which utilizes the genetic information contained in the undamaged sister chromatid as a template (Li X et al., 2008); in contrast, NHEJ is often error-prone and involves elimination of DSBs by direct ligation of the broken ends (Lieber MR et al., 2010).

8.2.1 Homologous repair (HR)

The HR mainly occurs in the S and G2 phase of the cell cycle (Durant and Nickoloff, 2005), since it uses the sister chromatid to copy the information required to repair the break. HR starts with DSB processing through 5’ to 3’ end resection (See Fig 13) which is initiated by the MRX complex (Mre11-

Rad50-Xrs2) and its associated partner Sae2 (Nicolette ML et al., 2010). Then 5' to 3' resection continues through the action of Exo1 exonuclease or Sgs1 helicase and helicase/nuclease Dna2, in a second resection step called long range resection. After the resection step, 3'-single stranded overhangs (ssDNA) are rapidly coated with RPA protein to remove secondary structures that would interfer with the recruitment of the Rad51 recombinase.

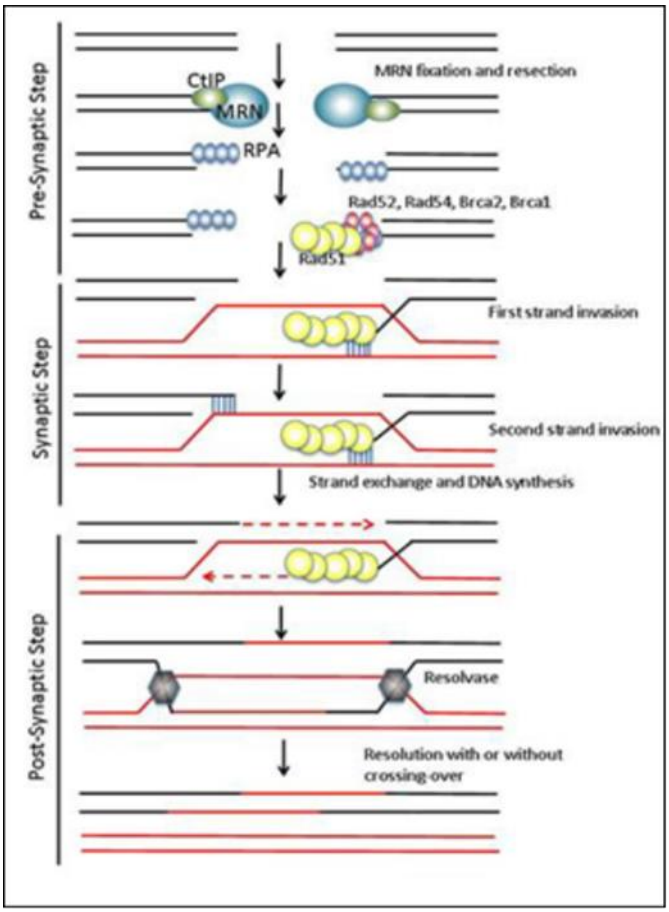


Figure 13. *Schematic representation of DSB repair by homologous recombination (HR)* (taken from Rass, E, 2009)

In fact, in the next step RPA is replaced by the Rad51 protein in a Rad52-, Rad55-and Rad57-dependent manner, to form a Rad51 nucleoprotein filament. The filament, in concert with the Swi/Snf complex and Rad54,

directs the search for homology sequences and once the homology has been identified, it further proceeds with DNA strand invasion, where the damaged DNA strand invades the template DNA duplex. Annealing of the filament with the homologous template initiates DNA synthesis from the 3'end of the invading strand, which is carried out by DNA polymerase η, followed by successive ligation by DNA ligase I to yield a four-way junction intermediate structure, known as a Holliday junction (McIlwraith MJ et al., 2005). Finally, the Sgs1-Top3-Rmi complex resolves the join molecules (Heyer, W.D et al., 2006; San Filippo J, 2008).

8.2.2 Non-homologous DNA end-joining (NHEJ)

NHEJ is active during the whole cell cycle, but it is predominant in the G1 phase (Krokan HE et al., 2004). In mammalian cells, NHEJ starts with a

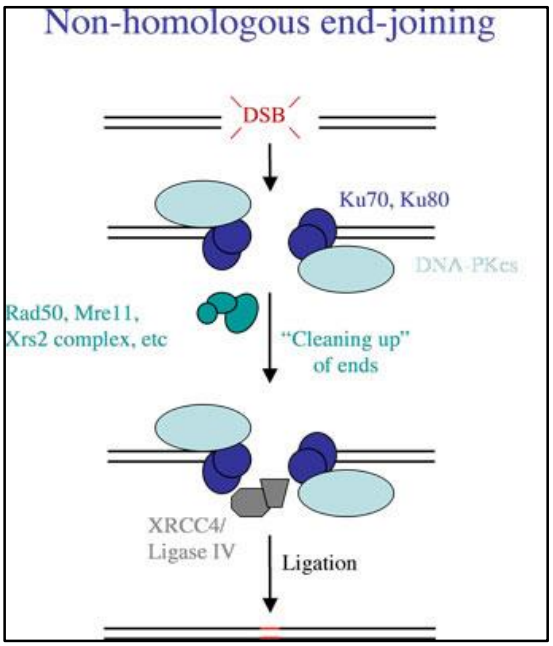


Figure 14. *Schematic representation of the DSB repair pathway: non-homologous end joining (NHEJ)* (taken from Doherty and Jackson, 2001)

limited processing of DNA ends by the MRN complex (Mre11/Rad50/Nbs1). Then, the proteins Ku-70 and Ku-80 bind the DNA ends and recruit the DNA dependent-protein kinase (DNA-PK). Once bound to broken ends, DNA-PK activates itself and phosphorylates its targets, including RPA and Artemis. DNA ligase IV will then seal the break (See Fig 14).

8.3 DNA damage tolerance pathway

The process of coping with DNA damage during replication (PRR) is referred to as a DNA damage tolerance pathway. This process is biologically important as the DNA repair pathways. Earlier studies in both yeast and mammalian cells suggested two major pathways for PRR: translesion synthesis (TLS) and a damage avoidance mechanism acting through template switching (TS) (Chang DJ et al., 2009; Branzei D and Foiani M, 2010; Lee KY et al., 2008; Klarer AC et al., 2011) (See Fig 15). During PRR the lesion is bypassed and left unrepaired but offering the possibility of being fixed in subsequent stages of the cell cycle by the DNA repair mechanisms described above (Budzowska and Kanaar, 2009). This temporary bypass and tolerance of a DNA lesion often come at a cost. There is an increased mutation rate at the lesion site due to the error-prone nature of this process (McCulloch and Kunkel, 2008). The predominant mechanism of DNA damage tolerance is translesion synthesis. Translesion synthesis is the replicative bypass of DNA damage by non-classical DNA polymerases. This process involves the incorporation of nucleotides directly across a DNA lesion which blocks DNA replication because by classical polymerases are unable to accommodate the lesion in their active site. This process is error-prone because the polymerases responsible for translesion synthesis have a reduced fidelity of nucleotide incorporation, a property that allows them to accommodate the structural distortions caused by various types of DNA lesions (Prakash S. et al 2005). In

fact, replication errors associated with translesion synthesis are believed to be responsible for almost all DNA damage-induced mutations (Guo C. et al., 2009). Each of the TLS polymerases has different substrate specificities for different types of DNA damage (Sharma S, 2013; Lange SS et al., 2011; Makridakis NM et al., 2012). The non-classical polymerases involved in the translesion synthesis in eukaryotes are polymerase η, polymerase ι, polymerase ζ, polymerase κ, and the Rev1 protein.

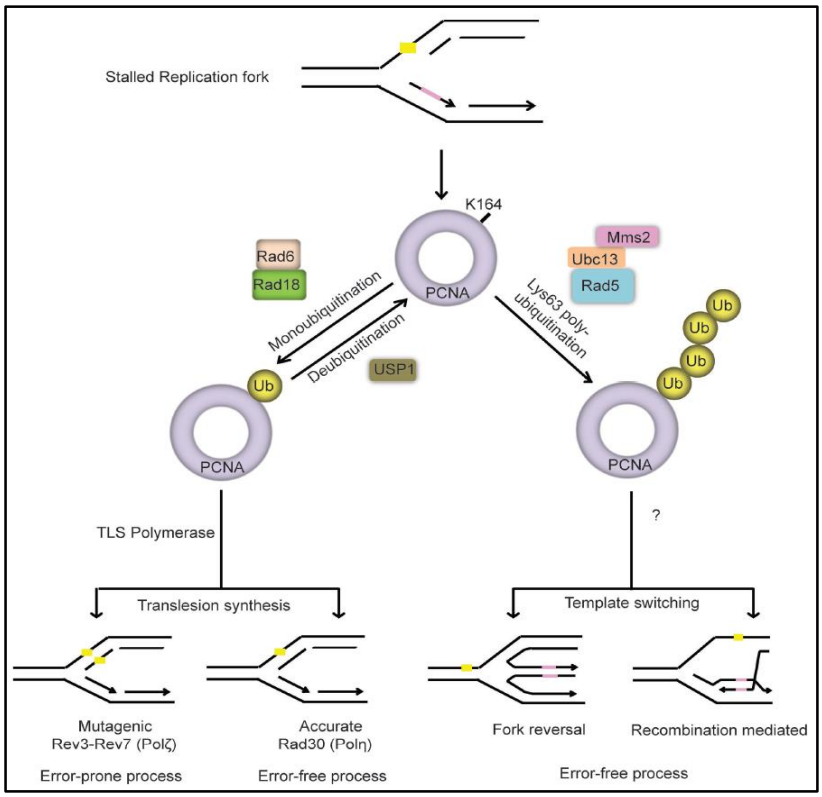


Figure 15.*Schematic representation of DNA damage tolerance pathway.* (taken from Goshal G. et al., 2013): DNA damage tolerance pathway (DDT): lesions (yellow Square) in the DNA template blocks progression of high-fidelity replicative polymerase resulting in stalled replication forks. DNA damage tolerance mechanism mediates bypass of lesions by replicating over damaged DNA by low-fidelity DNA polymerases (translesion synthesis) or using the undamaged sister chromatid as a template (template switching). ***Template switching*** is mediated by structural rearrangement of the replication

fork either by recombination or fork reversal. The key regulator of DDT pathway is the modification of PCNA. Under undamaged conditions replicative polymerase binds to unmodified PCNA during DNA replication. Upon genotoxic stress, PCNA is ubiquitinated at K164 to initiate DNA damage tolerance pathways. Monoubiquitination of PCNA promotes translesion synthesis, while polyubiquitination facilitates template switching. PCNA is monoubiquitinated by RAD18-RAD6 E3-ligase and polyubiquitinated by Rad5 (human homologue, SHPRH or HLTF). Following lesion bypass Usp1 deubiquitinates PCNA, thereby facilitating loading of the replicative polymerase to resume DNA synthesis.

To employ these non-classical polymerases, the stalled classical polymerase at the site of DNA damage must be exchanged for a TLS polymerase. The non-classical polymerase will then bypass the damage, and a second exchange will occur between the TLS and the classical polymerase. This switching event is mediated by replication factors at the replication fork, mainly by PCNA (de Saro 2009; Lehmann et al., 2007). The concerted actions of Rad6, the E2 ubiquitinating conjugating enzyme and the E3 ubiquitinating ligase Rad18 is required for the monoubiquitination of Lys164 of PCNA which recruits TLS polymerases in response to stalled replication caused by DNA damage (Friedberg, Lehmann et al., 2005).

9. *S. cerevisiae* as a model to study genome stability maintenance

9.1 *S. cerevisiae* as a lab model

The yeast *Saccharomyces cerevisiae,* (budding yeast) is a unicellular eukaryotic organism and it belongs to the fungi kingdom (Kurtzman CP, Fell JW, 2005). It has a haploid genome composed of a total of 13 Megabases (Mb), organized in 16 chromosomes. After its complete genome sequence, we know that SC DNA contains approximately 6000 open reading frames (ORFs) most of which encode specific proteins. Yeast is non-pathogenic and serves as

an ideal model organism in many aspects of eukaryotic biology, from gene structure to protein function (Botstein and Fink, 1988). Many pathways in eukaryotes are evolutionary conserved from yeast to multicellular organisms, including humans. Therefore, data obtained from yeast can in many cases be transferred and applied to human cells. Yeast has a short generation time, and it is relatively easy and cheap to use and maintain. In addition, budding yeast has a highly efficient DNA recombination system, making gene deletions, gene modifications and epitope tagging experiments relatively straightforward and efficient.

The yeast has become the 'test bed' for developing many new technologies, for example, synthetic genetic array (SGA) technology. The creation of a complete deletion collection of non-essential yeast genes allows genome-wide screenings to dissect cellular pathways, as well as to facilitate the structure-function analysis of genes and proteins. Interfering with the function of proteins in molecular complexes can be uncovered by synthetic genetic interactions, usually identified when a specific mutant is screened for second-site mutations or overexpression effects that either suppress or enhance the original phenotype.

9.2 SGA technology

Synthetic genetic array (SGA) is a high-throughput technology, which is now quite routinely applied using yeast cells. The SGA analysis allows the systematic construction of double mutant strains (Tong et al. 2001), and the subsequent large scale analysis of synthetic genetic interactions (Tong et al., 2004; Baryshnikova A, 2010). The query mutant strain is crossed with the array of approximately 5000 viable deletion mutants and the resulting diploids are transferred to a reduced nitrogen medium for sporulation. The haploid progeny is then put through a series of selection platings and incubations in

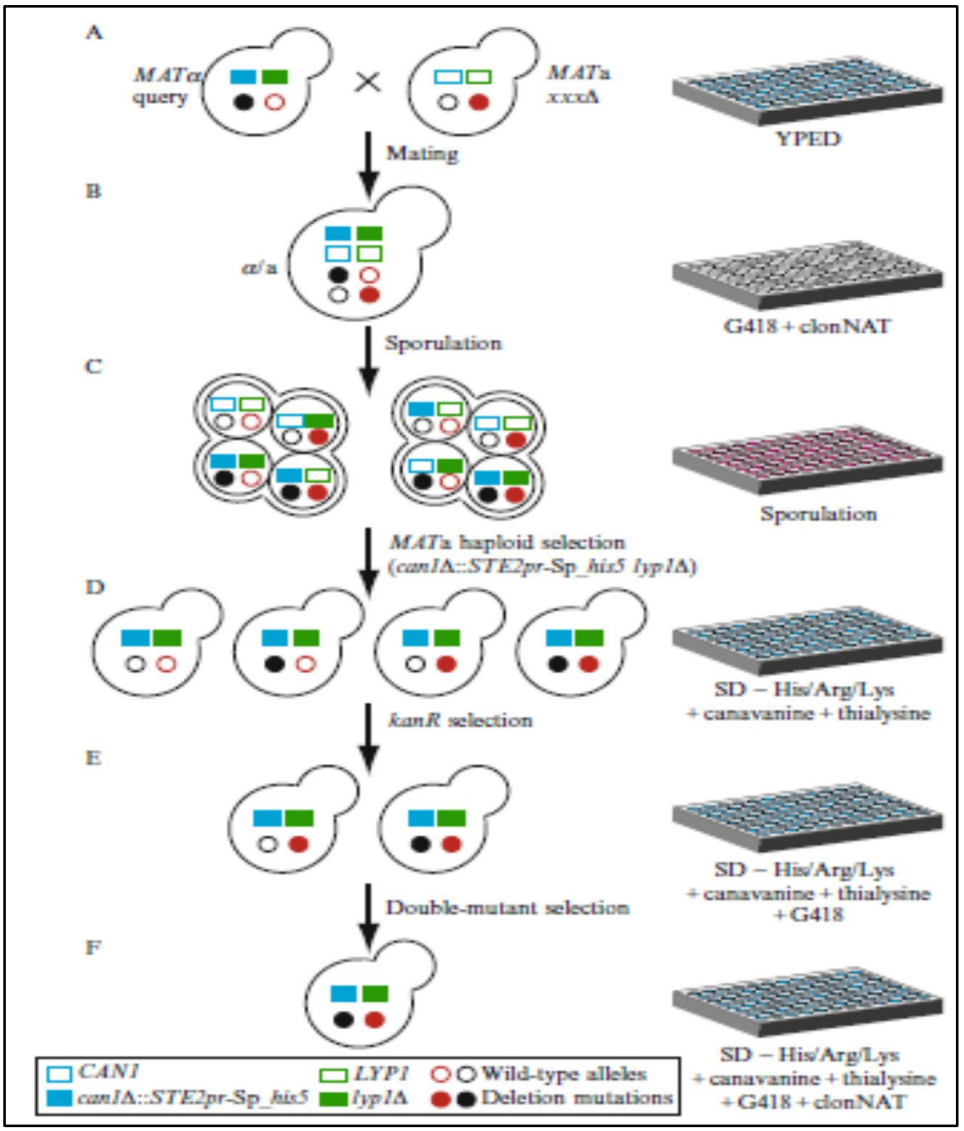


Figure 16. *The Schematic representation of SGA technology for the selection of double mutant* (taken from Baryshnikova A, 2010)

order to select for the double mutant (See Fig 16); finally, the colonies are scored for growth defects by a computer software analysis. The SGA technology allows to identify genetic interaction networks, thus providing functional informations associated with the position and connectivity of a gene in the network. The methodology can be adapted to many different functions

because any genetic element marked by a selectable marker can be analyzed. In addition, using the SGA methodology, strains containing specific alleles, including temperature-sensitive alleles (ts-mutants), point mutants, or plasmid can be crossed with any ordered array of mutant strains, providing a systematic tool for genetic suppression analysis, dosage lethality and dosage suppression screens or plasmid shuffling.

9.3 *S. cerevisiae*, the tool for studying genome instability

Saccharomyces cerevisiae proved to be a great model for several decades to study the function of genes and pathways (epistasis studies) involved in genome stability. For example, DNA damage repair and DNA damage checkpoint pathways are well conserved from yeast to human, and their analysis in yeast enable us to understand better the human counterparts. Multiple assays have been developed to study genome instability in yeast, like GCR assay (Schmidt KH et al., 2006; Christopher D. Putnam and Richard D. Kolodner, 2010), chromosomal loss assay (Klein HL, 2001), point mutation assay (Foster PL. 2006), spontaneous recombination assay (Spell RM, Jinks-Robertson S, 2004) and repetitive sequences instability assay (Kenneth Larkin and Michael Schweizer, 1999). For example, the GCR assay allows to measure the rate of accumulation of spontaneous GCRs in *S. cerevisiae* and allowed to understand the involvement of many DNA replication, repair, recombination, checkpoints, telomere maintenance and chromosome remodelling proteins in genome integrity maintenance. Likewise, many other above-mentioned assays helped to identify and characterize a wide range of genome aberrations in yeast that proved to help our understanding of the role of specific genes in genome maintenance.

Aim of the project

The integrity of the genome is crucial for normal cell cycle progression, and for the propagation of genetic information to subsequent generations and genome integrity can be threatened by several endogenous and exogenous factors. Therefore, different mechanisms must exist in the cell to overcome this challenges. In the last 20 years several studies have been successfully carried out to improve our understanding of the molecular mechanisms preserving genome integrity. Nevertheless, it is likely that not all the genes and pathways involved in genome integrity maintenance have been identified and fully characterized.

In the attempt to find out new genome stability genes, we developed a screening strategy based on the spontaneous accumulation of endogenous DNA damage. The screening relies on the overexpression of the *DDC2* DNA damage checkpoint gene in the yeast deletion mutant collection: indeed, high Ddc2 levels affect the viability of strains experiencing endogenous DNA damage, and may allow the identification of new genome integrity mutants.

Main results

1. The strategy behind the screening

The idea of the screening came from previous observations. Yeast strains suffering genome instability undergo spontaneous accumulation of endogenous DNA damage which can be detected by phosphorylation of the Rad53 checkpoint kinase, indicating a chronically activated DNA damage checkpoint response (Zhang et al., 2006a; Driscoll et al., 2007; Duro et al., 2008; figure 17).

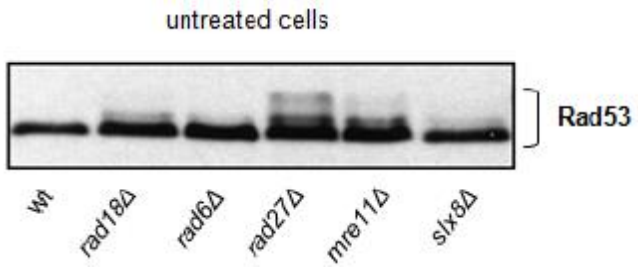


Figure 17. ***Yeast strains deleted for genes which are known to be involved in the control of genome stability display chronic Rad53 phosphorylation*****.** Protein extracts were prepared from exponentially growing cultures of the indicated strains and the phosphorylation status of Rad53 was analysed by SDS-PAGE and western blotting with anti-Rad53 antibodies.

We designed a strategy to select this phenotype, in order to screen the yeast deletion collection for genome integrity genes, based on the accumulation of endogenous DNA damage.

The rationale of the screening is based on the overexpression of the DDC2 gene. Ddc2 is the binding partner of the apical kinase Mec1, which is the main activator of the DNA damage checkpoint. It has been demonstrated that *DDC2* overexpression in the presence of DNA damage leads to checkpoint hyper-activation, resulting in prolonged cell cycle arrest and cell death, while it has no effect on undamaged cells (Clerici et al., 2001). We confirmed the observation that *DDC2* overexpression increases the sensitivity of yeast cells to low doses of DNA damaging agents, and we speculated that a similar effect might be observed if some cell mutants will accumulate endogenous DNA damage (i.e., in the absence of any treatment with external DNA damaging agents). This hypothesis was confirmed by testing the effect of *DDC2* overexpression in yeast strains bearing deletion of genes known to play a role in genome integrity maintenance: indeed, all the strains tested showed severe growth defects in response to *DDC2* overexpression. Thus, we decided to screen the yeast deletion collection to identify strains which are unable to grow in *DDC2* overexpression conditions, due to accumulation of endogenous DNA damage. We chose to overexpress *DDC2* under the control of the *GAL1* promoter, which is induced in the presence of galactose on a multicopy plasmid in order to increase the sensitivity of the screen in the presence of moderate levels of DNA damage. We exploited for our screening the power of the SGA (Synthetic Genetic Array) technology (Baryshnikova A, et al., 2010a)**,** by introducing the *GAL1-DDC2* overexpressing plasmid in a yeast deletion collection.

The procedure can be summarized as follows: mating of a query strain containing the plasmid with the deletion collection of the opposite mating type and selection of the diploids. After sporulation, selection of yeast haploid mutants with both the gene deletion and the plasmid (figure 18).

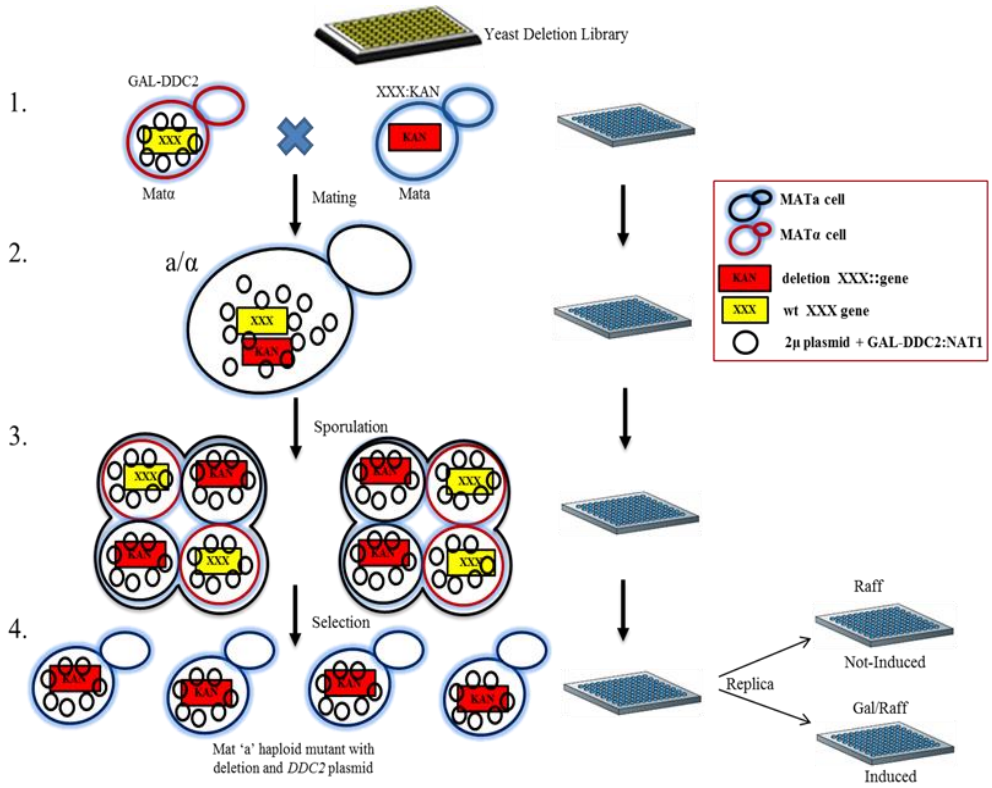


Figure 18. ***Overview of the screen for genome stability genes based on DDC2 overexpression***

The resulting colonies are then replicated in galactose containing medium to induce *DDC2* overexpression and cell proliferation is assessed by checking the colony formation ability. Each mutant's fitness is scored by measuring the size of the colony on galactose medium compared to the control experiment with the empty vector. The screening can be carried out only with a well equipped SGA platform and it was performed in Toronto in collaboration with G. Brown and C. Boone's labs, recognized leaders in this field.

Processing of SGA data was done (figure 19) as follows: i) the galactose plates were photographed with a high-resolution digital camera; ii) the digital images of the plates were processed by image processing software

that identifies the colonies and measures their areas in terms of pixels; iii) the quantified colony size was converted in a 'score' number through further Bioinformatics analysis performed by Anastasia Baryshnikova (Baryshnikova et al., 2010b).

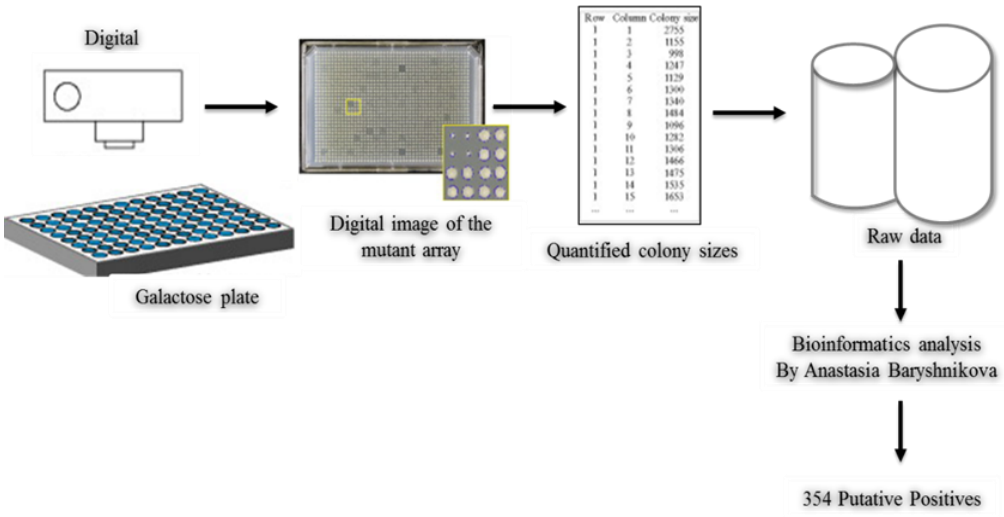


Figure 19. ***Processing the SGA data by computational analysis and selection of putative positives***

2. Identification of new genes controlling genome stability by *DDC2* overexpression

The SGA screening was performed twice. By merging the results obtained from both screenings, we obtained 354 putative positives which can be gathered in various cellular pathways represented in figure 20. Apart from known DNA damage response or genome stability genes, we identified genes involved in many pathways such as mitochondrial structure and function, peroxisomes biogenesis, ribosome biogenesis, protein sorting, cell wall, oxidative stress response, cytoskeleton, chromatin remodeling, autophagy, protein degradation, transcription regulation, RNA processing, plasma membrane transport, vacuolar function, meiosis and amino acid biosynthesis.

Interestingly, we found many mitochondria, peroxisomes and oxidative stress genes, which together form the most represented category and which were not identified in previous screenings for genome integrity genes (Bonekamp NA. et al., 2009; Moldovan L. et al., 2004; Ouspenski II. et al., 1999; Myung K. et al., 2001; Huang ME. et al., 2003; Measday V, et al., 2005; Pan X, et al., 2006; Storchov_a Z, et al., 2006; Andersen MP. et al., 2008; Smith S, et al., 2004; Yuen KWY, et al., 2007**;** Strome ED. et al., 2008; Ungar L, et al., 2009; Alabrudzinska M. et al., 2011; Stirling PC, et al., 2011; Zhang Y, et al., 2012). Surprisingly, we also obtained a set of genes growing better in *GAL-DDC2* overexpression conditions, but this class of genes needs further investigation. Moreover, many of the identified genes are still uncharacterized (20%), which paves the way for future studies.

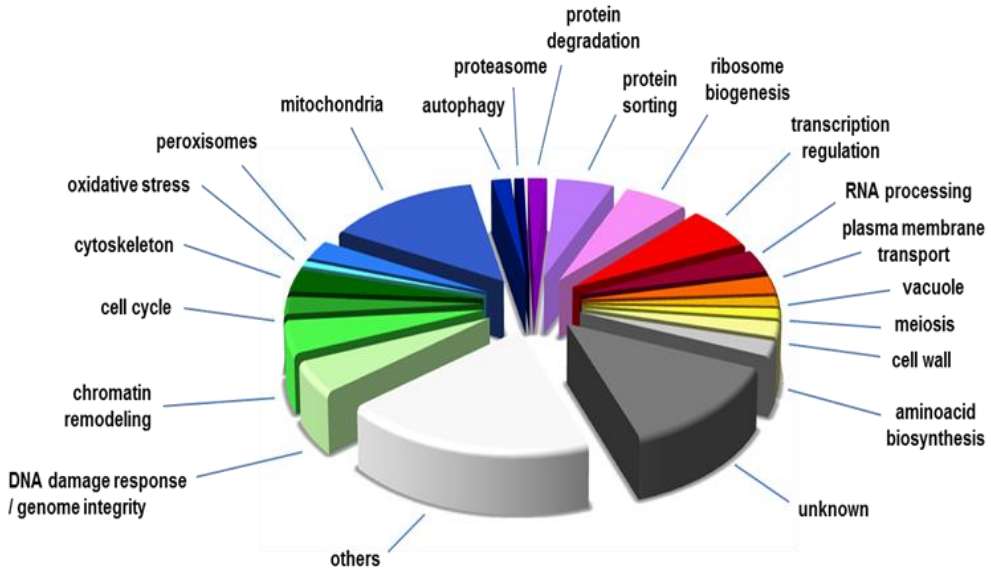


Figure 20. ***Schematic representation showing the putative roles of the positives genes identified in the SGA screening***

3. **Direct validation of putative positives**

The first 96 putative positive genes were directly validated by transformation and drop assays. Selected 96 candidate strains were retrieved from the Euroscarf deletion collection (Background: BY4741), and transformed with the *GAL-DDC2* plasmid or with the empty plasmid, separately. Drop assays were performed with all the transformed strains on galactose containing plates to check the sensitivity to *DDC2* overexpression, and on glucose containing plates as a control. We obtained a list of confirmed true positives: out of the 96 tested candidates, we were able to confirm 67% of them, while the remaining 33% were rejected as false positives. We conclude that the overall quality of the screening was good compared to previous screenings done using the SGA technology **(**Amy Hin Yan Tong, et al., 2001**).** In the case of the confirmed positives, the strains were appreciably sensitive to *GAL-DDC2* overexpression in the presence of galactose, compared to the wild type control strains. Conversely, when the mutant strains showed a strongly reduced fitness in galactose both with the *GAL-DDC2* plasmid and the empty vector or when the same growth was observed in the wild type control and in the mutant after *DDC2* overexpression the candidate gene was discarded as a false positive (figure 21).

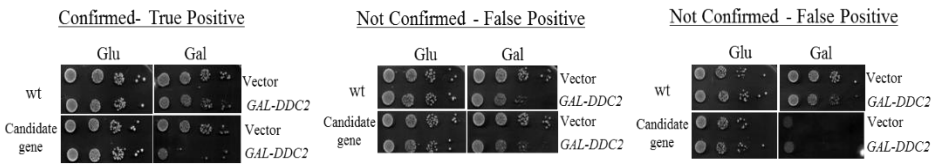


Figure 21. ***LEFT: Example of a confirmed true positive candidate gene, showing the sensitivity to GAL-DDC2 overexpression compared to the wild type and empty vector controls. MIDDLE: a non-confirmed false positive candidate gene shows the growth pattern of the wild type and empty vector control. RIGHT: a non-confirmed false positive candidate gene shows lethality also with an empty vector.***

4. The GAL-*DDC2* screen with essential genes: Preliminary results

Additionally, we decided to extend our search for genome integrity genes by including ~1000 genes which are essential for cell viability in yeast. To do this, we applied the same experimental approach except that we used a temperature sensitive (ts)-mutant collection of essential genes. The screening was performed at two different temperature 26ºC and 30ºC. At present we only have preliminary results because the screening was performed only once. From the analysis at 26ºC, we obtained 57 putative positives, which plays roles in different cellular pathways, such as genome maintenance, cell cycle, RNA processing, cell wall function, protein sorting and ribosome biogenesis (figure 22a). From the analysis at 30ºC, we obtained 125 putative positives involved in various process, such as, DNA repair, protein degradation, transcription regulation, mitochondrial function, chromatin remodelling, plasma membrane transport, RNA processing, protein sorting, cell wall metabolism and cell cycle control (figure 22b). Currently, we are repeating the screening to confirm these results.

ts-mutant 26ºC

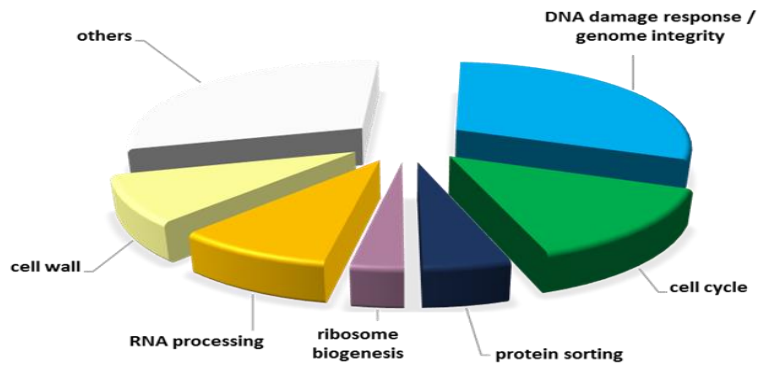


Figure 22a. ***Schematic representation of putative positive genes identified with the ts-mutant collection at 26ºC***

ts-mutant 30ºC

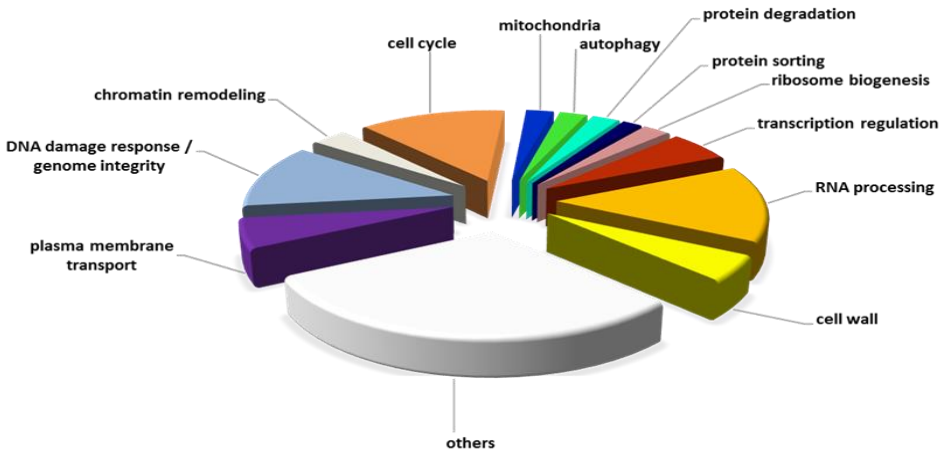


Figure 22b. ***Schematic representation of putative positive genes identified with the ts-mutant collection at 30 ºC***

Conclusions and future perspectives

Several pathways in cell are responsible for the maintenance of genome stability and if these pathways are unable to function properly or provide support to the normal cell function leads to “genome instability” a hallmark of cancer and various disease. In order to conclude the work in the thesis as summarize as follow:

1. We identified the several novel genes which might play a vital role in the maintenance of genome stability.
2. We identified several genes from ts-mutant screening which might play an important role in genome stability and maintenance.
3. The *DDC2* overexpression in yeast strains acts as a new tool for unveiling the new genes and pathways involved in the maintenance of genome stability.
4. Our screening strategy is unique from many other genome instability screens because we selected the genes based on the endogenous DNA damage (i.e., without external genotoxic agents).

From the above conclusion, interestingly, it results in paved a way for future studies, where we are particularly interested in unveiling the biological and molecular function of ORF (i.e., from the list of confirmed true positives)

in genome stability using classical genome instability assays and test the sensitivity to different DNA damaging agents. At the same time it is also worth to investigate the contribution of series of peroxisome and mitochondria genes in the role of genome integrity maintenance by performing specific assays to detect the reactive oxygen species level and also to perform comet assay for detecting the DNA breaks and its frequency. Moreover, unexpectedly we identified some mutants which were growing better in *DDC2* overexpression condition, but this class of genes needs further investigation.

Bibliography

Admire, A., L. Shanks, N. Danzl, M. Wang, U. Weier, W. Stevens, E. Hunt, and T. Weinert. (2006). Cycles of chromosome instability associate with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev*. *20*:159-73.

Amy Hin Yan Tong, Marie Evangelista, Ainslie B. Parsons, Hong Xu, D. Bader, Nicholas Mark Robinson, Sasan Raghibizadeh, Christopher W. V. Hogue, Howard Bussey, Brenda Andrews, Mike Tyers, Charles Boone. (2001). Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion. *Science.* 5550 pp. 2364-2368

Andersen MP, Nelson ZW, Hetrick ED, Gottschling DE (2008). A genetic screen for increased loss of heterozygosity in Saccharomyces cerevisiae. Genetics 179:1179{1195.

Anne Helmrick, (2008). Evolution of Common Fragile Sites. http://www.els.net/WileyCDA/ElsArticle/refId-a0020747.html

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

Aguilera A, García-Muse T. (2013). Causes of genome instability. *Annu Rev Genet. 2013* Jul 31.

Aguilera, A., & García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell , 46* (2), 115-124.

Aguilera, A., & Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet , 9 (3)*, 204-217.

Alabert C, Groth A., (2012). Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol*., *13(3):*153-67.

Alabrudzinska M, Skoneczny M, Skoneczna A (2011). Diploid-specific genome stability genes of S. cerevisiae: genomic screen reveals haploidization as an escape from persisting DNA rearrangement stress. PLoS One 6:e21124.

Azvolinsky, A., Giresi, P. G., Lieb, J. D., & Zakian, V. A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol Cell , 34* (6), 722-734.

Baird, D. M., Rowson, J., Wynford-Thomas, D., and Kipling, D. (2003). Extensive allelic variation and ultra short telomeres in senescent human cells. *Nat Genet, 33*: 203-207

Bakkenist C J, Kastan M B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, *421(6922):* 499-506.

Baryshnikova A, et al. (2010)a. Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Methods Enzymol 470:145{179.

Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn JY, Ou J, San Luis BJ, Bandyopadhyay S, Hibbs M, Hess D, Gingras AC, Bader GD,Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL. (2010)b. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods, 7(12):1017-24.

Bartkova, J., Horejsí, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., et al. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature , 434* (7035), 864-870.

Bashkirov, V.I. (2003). Direct kinase to-kinase signaling mediated by the FHA phosphoprotein recognition domain of the Dun1 DNA damage checkpoint kinase. *Mol Cell Biol, 23(4)*:1441-52

Beletskii, A., & Bhagwat, A. S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proc Natl Acad Sci U S A , 93* (24), 13919-13924.

Bonekamp NA, Volkl A, Fahimi HD, Schrader M (2009) Reactive oxygen species and peroxisomes: struggling for balance. Biofactors 35:346-355.

Boiteux, S. and M. Guillet, 2004. Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair (Amst),* 3(1): p. 1-12.

Botstein, D. and Fink, G. R. (1988). Yeast: an experimental organism for modern biology. *Science 240*, 1439–1443.

Budzowska M, Kanaar R. (2009). Mechanisms of dealing with DNA damage-induced replication problems. *Cell Biochem Biophys*. *53(1*):17-31

Burney S, Caulfield JL, Niles JC, Wishnok JS, Tannenbaum SR (1999). The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat Res 424(1–2):*37–49

Branzei D, Foiani M. (2005). The DNA damage response during DNA replication. *Curr Opin Cell Biol. 17(6):*568-75.

Branzei D., Foiani M. (2007). Template switching: from replication fork repair to genome rearrangements. *Cell. 131*, 1228-30

Branzei D, Foiani M., (2008). Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol. Apr; 9(4)*:297-308.

Branzei D, Foiani M. Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol; 11*:208-19

Brown E J, Baltimore D. (2003). Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes Dev, 17(5)*: 615-628.

Cadet J, Berger M, Douki T, Ravanat JL (1997). Oxidative damage to DNA: formation, measurement, and biological significance. *Rev Physiol Biochem Pharmacol 131*:1–87

Carson, C.T., Schwartz, R.A., Stracker, T.H., Lilley, C.E., Lee, D.V. and Weitzman, M.D. (2003). The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J., 22*, 6610–6620.

Celli, G. B., and de Lange, T. (2005). DNA processing is not required for ATM mediated telomere damage response after TRF2 deletion. *Nat. Cell Biol. 7*, 712-718

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol.Cell 40*, 179–204.

Clerici M, et al. (2001) Hyperactivation of the yeast DNA damage checkpoint by TEL1 and DDC2 overexpression. EMBO J 20:6485-6498.

Cotta-Ramusino, C., D. Fachinetti, C. Lucca, Y. Doksani, M. Lopes, J. Sogo, and M. Foiani. (2005). Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol Cell*. *17*:153-9.

Cha, R.S., and N. Kleckner. (2002). ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science. 297*:602-6.

Chan, K.L.; Palmai-Pallag, T.; Ying, S.; Hickson, I.D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell. Biol., 11*, 753-760

Chang DJ, Cimprich KA. (2009). DNA damage tolerance: when it's OK to make mistakes. *Nat Chem Biol; 5*:82-90

Chanoux, R.A., Yin, B., Urtishak, K.A., Asare, A., Bassing, C.H., and Brown, E.J. (2009). ATR and H2AX cooperate in maintaining genome stability under replication stress. *J. Biol. Chem. 284*, 5994–6003.

Christopher D. Putnam and Richard D. Kolodner. (2010). Determination of gross chromosomal rearrangement rates. *Cold Spring Harb Protoc.1*; (9)

Cleaver JE. (1968). Defective repair replication of DNA in xeroderma pigementosum. *Nature 218*:652–56

Datta, A. & Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* 268, 1616–1619. This work, using budding yeast as a model system, was the first demonstration of transcription-associated mutagenesis in eukaryotic cells.

D'Amours, D. and Jackson, S.P. (2002). The Mre11 complex: at the crossroads of DNA repair and checkpoint signaling. *Nat. Rev. Mol. Cell. Biol.*, *3*, 317–327.

Debacker K. and Kooy F. (2007). Fragile sites and human disease. *Hum. Mol. Genet. 16*, 150-158.

Deckbar, D., Jeggo, P.A., and Lobrich, M. (2011). Understanding the limitations of radiation-induced cell cycle checkpoints. *Crit. Rev. Biochem. Mol. Biol. 46*, 271–283.

De Bont, R., and N. van Larebeke. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*. *19*:169-85. Review

de Laat, W.L., N.G. Jaspers, and J.H. Hoeijmakers (1999). Molecular mechanism of nucleotide excision repair. *Genes Dev. 13(7)*: p. 768-85.

De Lange. (2005). Telomere-related genome instability in cancer. *Cold Spring Harb Symp Quant Biol,* 70: 197-204

DeRose, E. F., Perera, L., Murray, M. S., Kunkel, T. A., & London, R. E. (2012). Solution structure of the dickerson DNA dodecamer containing a single ribonucleotide. *Biochemistry , 51* (12), 2407-2416.

de Saro, F.J.L. (2009). Regulation of interactions with sliding clamps during DNA replication and repair. *Current Genomics 10,* 206-215.

Deshpande, A.M., and C.S. Newlon. (1996). DNA replication fork pause sites dependent on transcription. *Science.* 272:1030-3.

DiTullio, R.A., Jr, Mochan, T.A., Venere, M., Bartkova, J., Sehested, M., Bartek, J. and Halazonetis, T.D. (2002). 53BP1 functions in an ATM dependent checkpoint pathway that is constitutively activated in human cancer. *Nat. Cell Biol., 4*, 998–1002.

Dubarry M, Loiodice I, Chen CL, Thermes C, Taddei A. (2011). Tight protein–DNA interactions favor gene silencing. *Genes Dev 25:*1365–1370

Durant, S.T., and Nickoloff, J.A. (2005). Good timing in the cell cycle for precise DNA repair by BRCA1.*Cell Cycle Georget. Tex 4*, 1216–1222.

Durkin, S. G., & Glover, T. W. (2007). Chromosome fragile sites. *Annu Rev Genet , 41*, 169-192.

Dronkert ML, Kanaar R (2001) Repair of DNA interstrand cross-links. *Mutat Res 486*:217–247

Ellis NA, Groden J, Ye T-Z, Straughen J, Lennon DJ. (1995). The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell 83*:655–66

Falck, J., Mailand, N., Syljuasen, R.G., Bartek, J., and Lukas, J. (2001). The ATM-Chk2-Cdc25A checkpoint pathway guards against radio resistant DNA synthesis. *Nature 410,* 842–847.

Fortini P, Pascucci B, Parlanti E, D'Errico M, Simonelli V, Dogliotti E. (2003). The base excision repair: mechanisms and its relevance for cancer susceptibility.

Foster, P. L. (2006). Methods for determining spontaneous mutation rates. *Methods Enzymol 409*, 195–213.

Francino, M. P.; Ochman, (2001). Deamination as the basis of strand-asymmetric evolution in Transcribed *Escherichia coli* Sequences. *H. Mol. Biol. Evol.*, 18, 1147.

Friedberg, E.C., Lehmann, A.R., and Fuchs, R.P. (2005). Trading places: how do DNA polymerases switch during translesion DNA synthesis*? Mol Cell 18*, 499-505.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T. (2006). DNA repair and mutagenesis. ASM press, Washington D.C.

Furnari, B., Rhind, N., and Russell, P. (1997). Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science 277*, 1495–1497.

Gacy, A.M., G. Goellner, N. Juranic, S. Macura, and C.T. McMurray. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell. 81*:533-40

Gaillard H, Herrera-Moyano E, Aguilera A. (2013). Transcription-associated genome instability. *Chem Rev. 113* (11), 8638–8661

Gisselsson, D., Jonson, T., Petersen, A., Strombeck, B., Dal Cin, P., Hoglund, M., Mitelman, F., Mertens, F., and Mandahl, N. (2001). Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc Natl Acad Sci U S A, 98:* 12683-12688.

Gisselsson, D., Pettersson, L., Hoglund, M., Heidenblad, M., Gorunova, L., Wiegant, J., Mertens, F., Dal Cin, P., Mitelman, F., and Mandahl, N. Chromosomal breakage-fusion bridge events cause genetic intratumor heterogeneity. (2000). *Proc Natl Acad Sci U S A, 97:*5357-5362.

Glover, T.W.; Stein, C.K. (1987). Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet, 41,* 882-890

Glover, T.W.; Stein, C.K. (1988). Chromosome breakage and recombination at fragile sites. *Am. J. Hum.Genet., 43,* 265-273

Glover, T.W. (2007). Chromosomes fragile sites. *Annu. Rev. Genet, 4*1, 169-192

Glover TW, Wilson TE. (2013). Breaking news on fragile sites in cancer. *Cancer Cell. Feb 11; 23(*2):137-9

Goldberg, M., Stucki, M., Falck,J., D'Amours, D., Rahman, D., Pappin, D., Bartek, J. and Jackson,S.P. (2003) MDC1 is required for the intra-S-phase DNA damage checkpoint. *Nature, 421*, 952–956.

Gorgoulis, V. G., Vassiliou, L.-V. F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature , 434* (7035), 907-913.

Gottipati, P., Cassel, T. N., Savolainen, L., & Helleday, T. (2008). Transcription-associated recombination is dependent on replication in Mammalian cells. *Mol Cell Biol , 28* (1), 154-164.

Grabczyk E, Mancuso M, Sammarco MC. (2007). A persistent RNA·DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res.* 35(16):5351–59

Green, P.; Ewing, B.; Miller, W.; Thomas, P. J.; Green, E. D. (2003). Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet., 33*, 514.

Green, C.M., Erdjument-Bromage, H., Tempst, P. and Lowndes, N.F. (2000). A novel Rad24 checkpoint protein complex closely related to replication factor C. *Curr. Biol., 10*, 39–42.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell, 97*: 503-514.

Griffith, J.D., Lindsey-Boltz, L.A. and Sancar, A. (2002). Structures of the human Rad17-replication factor C and checkpoint Rad 9-1-1 complexes visualized by glycerol spray/low voltage microscopy. *J. Biol. Chem., 277*, 15233–15236.

Guo C, Kosarek-Stancel JN, Tang TS, Friedberg EC. (2009). Y-family DNA polymerases in mammalian cells. *Cell Mol Life Sci. Jul; 66(14)*:2363-81.

Halazonetis, T. D., Gorgoulis, V. G., & Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science , 319* (5868), 1352-1355.

Hanawalt, P.C., Ford, J.M., and Lloyd, D.R. (2003). Functional characterization of global genomic DNA repair and its implications for cancer. *Mutat. Res. 544*, 107–114.

Harley, C. B., Futcher, A. B., and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature, 345:* 458-460

Harfe, B.D. and S. Jinks-Robertson. (2000). DNA mismatch repair and genetic instability. *Annu Rev Genet, 34*: p. 359-399.

Harrison JC, Haber JE. (2006). Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet*.40:209-35.

A. Helmrich, M. Ballarino, and L. Tora, (2011). "Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes," *Molecular Cell*, vol. 44, 966–977

Heller RC, Marians KJ. (2006). Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol*.7 (12):932-43.

Herr, A. J., Ogawa, M., Lawrence, N. A., Williams, L. N., Eggington, J. M., Singh, M., et al. (2011). Mutator suppression and escape from replication error-induced extinction in yeast. *PLoS Genetics, 7(10)*

Heyer, W.D. (2006). Rad54: the Swiss Army knife of homologous recombination? *Nucleic Acids Res, 34(15):* p. 4115-25.

Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature, 411(6835)*: p. 366-74

Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. *N Engl J Med*. 361:1475-85.

Horejsi, Z., Falck, J., Bakkenist, C.J., Kastan, M.B., Lukas, J. and Bartek, J. (2004). Distinct functional domains of Nbs1 modulate the timing and magnitude of ATM activation after low doses of ionizing radiation. *Oncogene, 23*, 3122–3127.

Huang ME, Rio AG, Nicolas A, Kolodner RD (2003). A genome wide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A 100:11529{11534.

Huertas P, Aguilera A. (2003). Cotranscriptional formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* 12(3):711–21

Ikegami, S., Taguchi, T., Ohashi, M., Oguro, M., Nagano, H., and Mano, Y. (1978). Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-alpha. *Nature 275,* 458–460.

Ivessa, A. S., Lenzmeier, B. A., Bessler, J. B., Goudsouzian, L. K., Schnakenberg, S. L., & Zakian, V. A. (2003). The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell , 12* (6), 1525-1536.

Irigaray P, Belpomme D. (2010). Basic properties and molecular mechanisms of exogenous chemical carcinogens. *Carcinogenesis 31(2):*135–148

Jackson, S. P. and J. Bartek, (2009). "The DNA-damage response in human biology and disease." *Nature 461*(7267): 1071-8.

Jaishree, T. N., van der Marel, G. A., van Boom, J. H., & Wang, A. H. (1993). Structural influence of RNA incorporation in DNA: quantitative nuclear magnetic resonance refinement of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG). *Biochemistry , 32* (18), 4903-4911.

Jaktaji RP, Lloyd RG. (2003). PriA supports two distinct pathways for replication restart in UV-irradiated Escherichia coli cells. *Mol Microbiol*. (4):1091-100.

Karlseder, J., Smogorzewska. Dai, Y., Hardy, S., and de Lange, T. (1999). P53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science. 283,* 1321-1325

Karlseder, J., Smogorzewska, A., and de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. *Science. 295,* 2446-2449

Kastan MB, Bartek J. (2004). Cell-cycle checkpoints and cancer. Nature. 18; 432 (7015):316-23.

Kawanishi, S. and Oikawa, S. (2004). Mechanism of telomere shortening by oxidative stress. *Ann N Y Acad Sci, 1019*: 278-284

Kenneth Larkin, Michael Schweizer, (1999). Development of a yeast-based assay system for monitoring microsatellite instability. *FEMS Microbiology Letters 176*, 205-212

Khanna KK, Jackson SP. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet 27(3):*247–254

Kitagawa, R., Bakkenist, C.J., McKinnon, P.J. and Kastan, M.B. (2004). Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev., 18,* 1423–1438.

Klarer AC, McGregor W. (2011). Replication of damaged genomes. *Crit Rev Eukaryot. Gene Expr; 21:323*-36

Klein HL., (2001). Spontaneous chromosome loss in Saccharomyces cerevisiae is suppressed by DNA damage checkpoint functions. *Genetics; 159* (4):1501-9.

Kolodner, R. D., Putnam, C. D., & Myung, K. (2002). Maintenance of genome stability in Saccharomyces cerevisiae. *Science, 297(5581)*, 552–557

Kondo, T., Matsumoto, K. and Sugimoto, K. (1999) Role of a complex containing Rad17, Mec3, and Ddc1 in the yeast DNA damage checkpoint pathway. Mol. Cell Biol., 19, 1136–1143.

Kondo, T., Wakayama, T., Naiki, T., Matsumoto, K. and Sugimoto, K. (2001) Recruitment of Mec1 and Ddc1 checkpoint proteins to double-strand breaks through distinct mechanisms. *Science, 294*, 867–870.

Kunkel, T. A. (2004). DNA replication fidelity. *J Biol Chem , 279* (17), 16895-16898.

Kurtzman CP, Fell JW (2005). Biodiversity and Ecophysiology of Yeasts *(in: The Yeast Handbook, Gabor P, de la Rosa CL, eds)*. Berlin: Springer. pp. 11–30.ISBN 3-540-26100-1.

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science 252*:1711–1714

Krokan HE, Kavli B, Slupphaug G., 2004. Novel aspects of macromolecular repair and relationship to human disease. *J Mol Med (Berl). May; 82*(5):280-97

Labib K, Hodgson B., 2007. Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep;8(4):*346-53.

Lambert S, Froget B, Carr AM (2007) Arrested replication fork processing: interplay between checkpoints and recombination. *DNA Repair (Amst) 6*: 1042– 1061

Lange SS, Takata K, Wood RD. (2004). DNA polymerases and cancer. 2011. *Nat Rev Cancer; 11:96*-110

Langerak, P., and Russell, P. (2011). Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. *Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366,* 3562–3571.

Lawrence, C.W., A. Borden, S.K. Banerjee, and J.E. LeClerc. (1990). Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res. 18*:2153-7

Lee KY, Myung K. (2008). PCNA modifications for regulation of post-replication repair pathways. *Mol Cells; 26:5*-11

Lehmann, A.R. et al. (2007). Translesion synthesis: Y-farnily polymerases and the polymerase switch. *DNA Repair 6*, 891-899.

Lieber MR. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem 79*:181–211

Lemoine, F.J., N.P. Degtyareva, K. Lobachev, and T.D. Petes. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell*. 120:587-98.

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachagès, A.-M., Vogt, N., Hansen, R. S. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature , 470* (7332), 120-123.

Li X, Heyer WD. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res 18(1):*99–113

Li, Y., & Breaker, R. R. (1999). Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc. , 121* (23), 5364-5372.

Lin Y, Dent SYR, Wilson JH, Wells RD, Napierala M. (2010). R loops stimulate genetic instability of CTG·CAG repeats. *Proc. Natl. Acad. Sci. USA* 107(2):692–97

Lindahl T. (1993). Instability and decay of the primary structure of DNA. *Nature.362*:709-715

Lindahl T, Nyberg B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry 11(19)*:3610-3618

Lindsey, J., McGill, N. I., Lindsey, L. A., Green, D. K., and Cooke, H. J., (1991). In vivo loss of telomeric repeats with age in humans. *Mutat Res, 256:* 45-48

Liu, L. F.; Wang, J. C. (1987). Supercoiling of the DNA template during transcription, *Proc. Natl. Acad. Sci. U.S.A., 84*, 7024-7027

Liu, G., Myers, S., Chen, X., Bissler, J.J., Sinden, R.R., and Leffak, M. (2012). Replication fork stalling and checkpoint activation by a PKD1 locus mirror repeat polypurine-polypyrimidine (Pu-Py) tract. *J. Biol. Chem.287*, 33412–33423.

Lukusa T. and Fryns J. P. (2008). Human chromosome fragility. *Biochim. Biophys. Acta 1779*, 3-16.

Loeb, L.A., Springgate, C. F., & Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. *Cancer Res. 1974 Sep; 34 (9):*2311-21.

Loeb, L.A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res*. *51*: 3075-3079

Loeb LA. (2011). Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer. 2011 Jun; 11 (6):*450-7

Longhese,M.P., Foiani,M., Muzi-Falconi,M., Lucchini,G. and Plevani,P. (1998). DNA damage checkpoint in budding yeast. *EMBO J., 17*, 5525–5528.

Lopes, M., C. Cotta-Ramusino, A. Pellicioli, G. Liberi, P. Plevani, M. Muzi-Falconi, C.S. Newlon, and M. Foiani. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. *Nature*. 412:557-61.

López Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. Nat. Rev. Mol. Cell Biol. *11*, 165–170

Lou, Z., Minter-Dykhouse, K., Wu, X. and Chen, J. (2003). MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways. Nature, 421, 957–961.

Magenis R. E., Hecht F., Lovrien E. W. (1970). Heritable fragile site on chromosome 16: probable localization of haptoglobin locus in man. *Science 170,* 85-86.

Makridakis NM, Reichardt JK. (2012). Translesion DNA polymerases and cancer. *Front Genet; 3:174*

Mathieu, N., Pirzio, L., Freulet-Marriere, M. A., Desmaze, C., and Sabatier, L. (2004). Telomeres and chromosomal instability. *Cell Mol Life Sci, 61:* 641-656.

Maser, R. S. and DePinho, R. A. (2004). Telomeres and the DNA damage response: why the fox is guarding the henhouse. *DNA Repair (Amst), 3*: 979-988.

McClintock, B. (1941). The stability of broken ends of chromosomes in Zea mays. Genetics, *1041:* 234-282

McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res , 18* (1), 148-161.

McIlwraith MJ, Vaisman A, Liu Y, Fanning E, Woodgate R, West SC. (2005). Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell 20(5):*783–792

McKinnon PJ. (2009). DNA repair deficiency and neurological disease. *Nat Rev Neurosci 10(2):*100–112

Measday V, et al. (2005) Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. Proc Natl Acad Sci U S A 102:13956{13961.

Melo, J.A., Cohen, J. and Toczyski, D.P. (2001). Two checkpoint complexes are independently recruited to sites of DNA damage in vivo. *Genes Dev., 15*, 2809–2821.

Melo, J. and Toczyski, D. (2002). A unified view of the DNA-damage checkpoint. *Curr. Opin. Cell. Biol., 14*, 237–245.

Millerhe J. H., Suthar, A., Tai, J., Yeung, A., Truong, C., & Stewart, J. L. (1999). Direct selection for mutators in Escherichia coli. *Journal of Bacteriology, 181(5),* 1576–1584

Mirkin EV, Mirkin SM (2007). Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev 71*: 13–35

Mitas, M., A. Yu, J. Dill, and I.S. Haworth. (1995). The trinucleotide repeat sequence d(CGG) 15 forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry*. 34:12803-11.

Moldovan L, Moldovan NI (2004) Oxygen free radicals and redox biology of organelles. Histochem Cell Biol 122:395-412.

Moldovan, G. L., Pfander, B. and Jentsch, S (2007). PCNA, the maestro of the replication fork. *Cell 129*, 665-679

Moore, J.K., and J.E. Haber. (1996). Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature. 383:644-6.

Moore, H., P.W. Greenwell, C.P. Liu, N. Arnheim, and T.D. Petes. (1999). Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc Natl Acad Sci U S A*. 96:1504-9.

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A, 85*: 6622-6626.

Mugal, C. F.; von Grunberg, H. H.; Peifer, 2009. *M. Mol. Biol. Evol., 26*, 131.

Myung K, Chen C, Kolodner RD (2001). Multiple pathways cooperate in the suppression of genome instability in Saccharomyces cerevisiae. Nature 411:1073{1076.

Naiki, T., Shimomura, T., Kondo, T., Matsumoto, K. and Sugimoto ,K. (2000) Rfc5, in cooperation with rad24, controls DNA damage checkpoints throughout the cell cycle in Saccharomyces cerevisiae. *Mol. Cell Biol., 20*, 5888–5896.

Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol , 11* (3), 220-228.

Nicolette ML, Lee K, Guo Z, Rani M, Chow JM, Lee SE, Paull TT. (2010). Mre11-Rad50-Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. *Nat Struct Mol Biol.17 (12):*1478-85

Nickoloff, J. A. (1992). Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol Cell Biol , 12* (12), 5311-5318.

a. Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundström E-B, Burgers PMJ, Johansson E, Chabes A, Kunkel TA. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl Acad. Sci. USA.107*:4949–4954

b. Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundström E-B, Johansson E, Chabes A, Kunkel TA. (2010). Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol. 6*:774–781

Noll DM, Mason TM, Miller PS. (2006). Formation and repair of interstrand cross-links in DNA. *Chem Rev 106(*2):277–301

O'Connell, M.J., Raleigh, J.M., Verkade, H.M., and Nurse, P. (1997). Chk1 is a wee1 kinase in the G2 DNA damage checkpoint inhibiting cdc2 by Y15 phosphorylation. *EMBO J. 16*, 545–554.

Ohki, R., Tsurimoto, T., and Ishikawa, F. (2001). In vitro reconstitution of the end replication problem. *Mol Cell Biol, 21*: 5753-5766

Ouspenski II, Elledge SJ, Brinkley BR (1999) New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. Nucleic Acids Res 27:3001-3008.

Paciotti V, Clerici M, Lucchini G, Longhese MP (2000) The checkpoint protein Ddc2, functionally related to S. pombe Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. Genes Dev 14:2046{2059.

Pan X, et al. (2006). A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell 124:1069{1081.

Paulovich, A.G., Armour, C.D. and Hartwell, L.H. (1998). The Saccharomyces cerevisiae RAD9, RAD17, RAD24 and MEC3 genes are required for tolerating irreparable, ultraviolet-induced DNA damage. *Genetics, 150*, 75–93.

Parrilla-Castellar, E.R., Arlander, S.J. and Karnitz, L. (2004). Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair (Amst.), 3*, 1009–1014.

Petermann, E., Woodcock, M., and Helleday, T. (2010). Chk1 promotes replication fork progression by controlling replication initiation. *Proc. Natl. Acad. Sci. U. S. A. 107*, 16090–16095.

Petrini, J.H. and Stracker, T.H. (2003). The cellular response to DNA double-strand breaks: defining the sensors and mediators. *Trends Cell Biol., 13,* 458–462.

Possoz C, Filipe SR, Grainge I, Sherratt DJ. (2006). Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo. *EMBO J 25*:2596–2604

Postow L, Crisona NJ, Peter BJ, Hardy CD, Cozzarelli NR. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci U S A., 98(15)*:8219-26.

Prado, F., & Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J , 24* (6), 1267-1276Prakash, S. and L. Prakash, (2000). Nucleotide excision repair in yeast. *Mutat Res, 451*(1-2): p. 13-24.

Prakash S, Johnson RE, Prakash L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem. 74*:317-53.

Ravanat J-L. (2005). Measuring oxidized DNA lesions as biomarkers of oxidative stress: an analytical challenge *FABAD. J Pharm Sci* 30(2):100–113

Raveendranathan, M., S. Chattopadhyay, Y.T. Bolon, J. Haworth, D.J. Clarke, and A.K. Bielinsky. (2006). Genome-wide replication profiles of S-phase checkpoint mutants reveal fragile sites in yeast. EMBO J. 25:3627-39.

Roberts, M.J., Wondrak, G.T., Laurean, D.C., Jacobson, M.K. and Jacobson, E.L. (2003). DNA damage by carbonyl stress in human skin cells. *Mutat. Res.*, 522, 45±56

Rouse J, Jackson SP (2000) LCD1: an essential gene involved in checkpoint control and regulation of the MEC1 signaling pathway in Saccharomyces cerevisiae. EMBO J 19:5801{5812.

Richter, T. and Proctor, C. (2007). The role of intracellular peroxide levels on the development and maintenance of telomere-dependent senescence. *Exp Gerontol, 42:* 1043-1052

San Filippo, J., P. Sung, and H. Klein, (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem. 77: p*. 229-57.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol. 13*, 141–152.

Saka, Y., Esashi, F., Matsusaka, T., Mochida, S. and Yanagida, M. (1997). Damage and replication checkpoint control in fission yeast is ensured by interactions of Crb2, a protein with BRCT motif, with Cut5 and Chk1.*Genes Dev., 11*, 3387–3400.

Schaaper, R. M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *J. Biol. Chem. 268*, 23762–23765

Scholes, D.T., M. Banerjee, B. Bowen, and M.J. Curcio. (2001). Multiple regulators of Ty1 transposition in Saccharomyces cerevisiae have conserved roles in genome maintenance. *Genetics. 159*:1449-65.

Schmitt MW, Prindle MJ, Loeb LA. (2012). Implications of genetic heterogeneity in cancer. *Ann N Y Acad Sci. 1267:110*-6.

Schmidt KH, Pennaneach V, Putnam CD, Kolodner RD. (2006). Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae. *Methods Enzymol.409*:462-76.

Schultz, L.B., Chehab, N.H., Malikzay, A. and Halazonetis, T.D. (2000). p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell. Biol., 151*, 1381–1390.

Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nat Rev Mol Cell., 5(2):*148-57.

Sharma S, Helchowski CM, Canman CE. (2013). The roles of DNA polymerase ζ and the Y family DNA polymerases in promoting or preventing genome instability. *Mutat Res.* 743-744:97-110

Shechter D, Costanzo V, Gautier J. (2004). ATR and ATM regulate the timing of DNA replication origin firing. *Nature Cell Biol*, 6(7): 648-655.

Shiraishi, **T.,** T. Druck, K. Mimori, J. Flomenberg, L. Berk, H. Alder, W. Miller, K. Huebner, and C.M. Croce. (2001). Sequence conservation at human and mouse orthologous common fragile regions, FRA3B/FHIT and Fra14A2/Fhit. Proc Natl Acad Sci U S A. 98:5722-7

Sinha BK. (1995). Topoisomerase inhibitors. A review of their therapeutic potential in cancer. *Drugs 49(1):*11–19

Shiomi,Y., Shinozaki,A., Nakada,D., Sugimoto,K., Usukura,J., Obuse,C. and Tsurimoto,T. (2002). Clamp and clamp loader structures of the human checkpoint protein complexes, Rad9-1-1 and Rad17-RFC. *Genes Cells, 7*, 861–868.

Smith DJ, Whitehouse I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature. 483 (7390):*434-8

Smith S, et al. (2004). Mutator genes for suppression of gross chromosomal rearrangements identi_ed by a genome-wide screening in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 101:9039{9044.

Sofueva S, Osman F, Lorenz A, Steinacher R, Castagnetti S, Ledesma J, Whitby MC. (2011). Ultrafine anaphase bridges, broken DNA and illegitimate recombination induced by a replication fork barrier. *Nucleic Acids Res* 39:6568–6584

Sogo, J.M., M. Lopes, and M. Foiani. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science*. 297:599-602.

Spell RM, Jinks-Robertson S. (2004). Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae. *Methods Mol Biol. 262*:3-12.

Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M. and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature, 421*, 961–966.

Stirling PC, et al. (2011). The complete spectrum of yeast chromosome instability genes identi_es candidate CIN cancer genes and functional roles for ASTRA complex components. PLoS Genet 7:e1002057.

Storchov_a Z, et al. (2006). Genome-wide genetic analysis of polyploidy in yeast. Nature 443:541-547.

Strome ED,Wu X, Kimmel M, Plon SE (2008). Heterozygous screen in Saccharomyces cerevisiae identi_es dosage-sensitive genes that affect chromosome stability. Genetics 178:1193{1207.

Sugiyama H, Fujiwara T, Ura A. (1994). Chemistry of thermal degradation of abasic sites in DNA. Mechanistic investigation on thermal DNA strand cleavage of alkylated *DNA. Chem Res Toxicol 7(5)*:673–683

Takeuchi, Y., T. Horiuchi, and T. Kobayashi. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev. 17:1497-506.

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C., (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science; 294(5550)*:2364-8.

Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J,Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson

JN, Lu H, Ménard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. (2004). Global mapping of the yeast genetic interaction network. *Science. 303(5659)*:808-13.

Tourriere, H., and P. Pasero. (2007). Maintenance of fork integrity at damaged DNA and natural pause sites. *DNA Repair (Amst).* 6:900-13.

Thoma. F. (1999). Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. *The EMBO Journal*, 18, 6585 - 6598

Thomas, B. J., & Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. *Cell , 56* (4), 619-630

Tran PT, Erdeniz N, Symington LS, Liskay RM. (2004). EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair (Amst) 3(12):*1549–1559

Umezu, K., M. Hiraoka, M. Mori, and H. Maki. (2002). Structural analysis of aberrant chromosomes that occur spontaneously in diploid Saccharomyces cerevisiae: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. *Genetics*.160:97-110.

Ungar L, et al. (2009). A genome-wide screen for essential yeast genes that a_ect telomere length maintenance. Nucleic Acids Res 37:3840-3849.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L. and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. EMBO J., 22, 5612–5621.

van Gent DC, Hoeijmakers JH**,** Kanaar R (2001). Chromosomal stability and the DNA double stranded break connection. *Nat Rev Genet 2(3):*196–206

Vernon, M., K. Lobachev, and T.D. Petes. (2008). High rates of "unselected" aneuploidy and chromosome rearrangements in tel1 mec1 haploid yeast strains. *Genetics*. 179:237-47.

Wakayama T, Kondo T, Ando S, Matsumoto K, Sugimoto K (2001) Pie1, a protein interacting with Mec1, controls cell growth and checkpoint responses in Saccharomyces cerevisiae. Mol Cell Biol 21:755-764.

Wang, B., Matsuoka, S., Carpenter, P.B. and Elledge, S.J. (2002). 53BP1, a mediator of the DNA damage checkpoint. *Science, 298,* 1435–1438.

Wang, **N.D**.; Testa, J.R.; Smith, D.I. (1993). Determination of the specificity of aphidicolin-induced breakage of the human 3p14.2 fragile site. *Genomics, 17*, 341-347

Ward, I.M., Minn, K., van Deursen, J. and Chen, J. (2003) p53 Binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. *Mol. Cell Biol., 23*, 2556–2563.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature , 171* (4356), 737-738.

Wells, R.D. (1996). Molecular basis of genetic instability of triplet repeats. *J Biol Chem*. 271:2875-8.

Wierdl, M.; Greene, C. N.; Datta, A.; Jinks-Robertson, S.; Petes, T. D. (1996). Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics. 143*, 713.

Wist, E., and Prydz, H. (1979). The effect of aphidicolin on DNA synthesis in isolated HeLa cell nuclei. *Nucleic Acids Res. 6*, 1583–1590.

Wogan GN, Hecht SS, Felton JS, Conney AH, Loeb LA. (2004).Environmental and chemical carcinogenesis. *Semin Cancer Biol 14(6):*473–486

Yamane. K, Wu, X. and Chen, J. (2002). A DNA damage-regulated BRCT-containing protein, TopBP1, is required for cell survival. *Mol. Cell Biol., 22*, 555–566.

Ye X, Franco AA, Santos H, Nelson DM, Kaufman PD, Adams PD. (2003). Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. *Mol Cell. 11(2):*341-51.

Yonekura S, Nakamura N, Yonei S, Zhang-Akiyama QM. (2009). Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J Radiat Res (Tokyo) 50(1):*19–26

Yuen KWY, et al. (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci U S A 104:3925{3930.

Zhang Y, et al. (2012). Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Non dividing Yeast Cells. Mol Cell 48:254-265.

Acknowledgements

I would like to express my sincere gratitude to Prof. Paolo Plevani and Prof. Marco-Muzi Falconi, first of all for giving me an opportunity to work in their laboratory and secondly for the continuous guidance throughout my Ph.D. studies. Special thanks to doctor Novarina. D, for his continuous support and guidance. My research project would not have proceeded without his non-stop technical help.

I would like to thank all my lab members, not only for their vital support and advices in the research work, but also showing their healthy friendship towards me.

I would like to thank my best friends behind the screen, who always been right there to make sure I did not forget to have fun during my seemingly endless pursuit of a higher education.

Most importantly, endless thanks to my mother and father for their encouragement and support throughout my studies.

PART II

Invited review paper

Daniele Novarina, Muthu K. Shanmugam, Paolo Plevani*, Marco Muzi-Falconi

Controlling genome integrity: The Dr. Jekyll and Mr. Hyde of DNA replication and transcription

Subtitle
Nucleic acid metabolism as the main endogenous threat to genome integrity

Invited review paper

Title

Controlling genome integrity: The Dr. Jekyll and Mr. Hyde of DNA replication and transcription

Subtitle
Nucleic acid metabolism as the main endogenous threat to genome integrity

Authors
Daniele Novarina, Muthu K. Shanmugam, Paolo Plevani*, Marco Muzi-Falconi

* to whom correspondence should be addressed:
paolo.plevani@unimi.it

Affiliations
Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy.

Keywords
genome stability; replication; transcription; chromosomal fragile sites; ribonucleotides incorporation; cancer.

Abbreviations
DSBs, double-strand breaks; ssDNA, single-stranded DNA; MIN, micro-satellite instability; GCRs, gross chromosomal rearrangements; G4s, G-quadruplexes; rNTPs, ribonucleotides; rNMPs, ribonucleoside monophosphates; TNRs, trinucleotide repeats; CFSs, common fragile sites;

ERFSs, early-replicating fragile sites; TAM, transcription-associated mutation; TAR, transcription associated recombination.

Summary

Despite the need to maintain the integrity of the genome to guarantee the stability of genetic information, DNA is not an inactive storage molecule. Indeed, normal cellular metabolism entails complex DNA transactions to transcribe, duplicate and repair the genetic material; paradoxically, one of the major threats to genome integrity comes from the DNA and RNA metabolisms themselves. Here we provide a comprehensive overview of the mechanisms through which nucleic acid metabolism impacts on genome integrity, highlighting the frequent crosstalks among them and presenting recent studies that shed new light on some of these processes, such as ribonucleotides incorporation during genome replication, fragile sites expression and transcription-replication interference. We also discuss potential benefits, which might explain why some of the mechanisms that jeopardize the stability of the genome were not counter-selected during evolution.

Introduction: DNA damage and maintenance of genome integrity

The integrity of DNA molecules, which are the major depositary of genetic information in living organisms, is continuously challenged by multiple endogenous and exogenous agents. First, as a consequence of its intrinsic instability in an aqueous environment, DNA molecules can undergo spontaneous hydrolysis, which results in depurination and subsequent formation of abasic sites, or deamination, yielding miscoding bases [1]. Second, DNA can be severely damaged by several byproducts of normal cellular metabolism. Reactive oxygen, nitrogen and carbonyl species, together with endogenous alkylating agents, estrogen and cholesterol metabolites can cause a wide spectrum of DNA lesions, such as DNA strand breaks, abasic sites, oxidized bases, DNA adducts, DNA cross-links and replication-blocking lesions [2]. Third, chemical or physical environmental agents–above all ionizing radiation (IR) and the ultraviolet (UV) component of sunlight, but also cigarette smoke and several chemotherapeutic agents – can cause hazardous alterations in DNA structure, such as bulky DNA adducts, double-strand breaks (DSBs) and single-strand breaks (SSBs),oxidative damage, interstrand or intrastrand crosslinks (ICLs) [3, 4]. DNA lesions are extremely harmful because they can be either mutagenic (alteration of the genetic information) or cytotoxic (impairment of cell viability). The extent of DNA damage occurring in living organisms is surprisingly high, since it was stimated that each cell experiences about 10^5 DNA lesions per day [1]. Therefore, a serious task posed to all cells is to maintain the integrity of the genome despite all the attacks to which it is continuously subjected, to ensure cell survival and preserve the genetic information that needs to be faithfully transmitted across generations. Increased genome instability causes the accumulation of a wide spectrum of genetic alterations, ranging from point mutations to gross chromosomal rearrangements. Different classes of genomic instability have been described: i) instability leading to mutations, including base substitutions, micro-insertions and micro-deletions; ii) mini- and micro-satellite instability (MIN), leading to expansion or contraction of repetitive DNA sequences; iii) gross chromosomal rearrangements (GCRs), including aberrations in chromosome structure like translocations, duplications, inversions or deletions; iv) chromosomal instability (CIN), defined as a persistently high rate of loss and gain of whole chromosomes, which results in aneuploidy [5]. Increased genome instability is a characteristic of most human cancers and it is considered a hallmark and a key driving force in tumorigenesis [6, 7] (Box 1). Despite the need to counteract all these genotoxic insults and maintain the integrity of the genome, DNA is anything but an inactive storage molecule. Instead, DNA undergoes complex transactions, which also constitute a very serious threat to genome integrity (summarized in Fig. 4):

indeed, the two main processes involving DNA, namely transcription and replication, are potential sources of chromosome breakage. In accordance with the oncogene-induced DNA damage model for cancer development (Box 1), replication stress and subsequent replication errors or failures appear to be the main origins of genome instability [5, 8–10]. Moreover, highly transcribed regions are often responsible for replication impairments [11]. In addition, specific chromosomal loci, such as fragile sites, repetitive sequences, DNA secondary structures-forming regions and telomeres pose specific challenges to genome integrity contributing to the onset of genomic instability [5, 12–14]. In recent years, new sources of genome instability arising from nucleic acid transactions were identified [11, 15, 16], and deeper insights were achieved on known factors, in some cases dramatically altering our perception of these processes in relation to genome integrity [17].

Box 1

Genome instability and cancer

Cancer is a multistep process, characterized by the gradual accumulation of genetic alterations. Two main models were put forward to elucidate the process of tumorigenesis: the "mutator hypothesis" and the "oncogene-induced DNA damage model". According to the mutator hypothesis, at the beginning of carcinogenesis, due to endogenous or environmental DNA damage, a "mutator mutation" occurs in a gene responsible for genome integrity maintenance, resulting in an overall increase of the mutation rate. This enhanced mutagenesis will favour the occurrence of "driver mutations" in oncogenes or tumor suppressor genes [121], which provide a proliferative advantage and will thus be selected within the precancerous cell population, according to changes in the microenvironment. This is expected to be a positive feedback mechanisms, since in a genetically unstable cell new mutations may occur which further increase genomic instability. Subsequent multiple rounds of selection and mutation will direct the evolution of the tumor up to a malignant cancer [121]. The main argument against this model is that mutations in genome stability genes (also called "caretaker genes") are usually recessive. Therefore, two independent mutations are needed to get an unstable genome, and the occurrence of this event before the onset of genomic instability is very unlikely [122]. Accordingly, high-throughput studies on cancer cell lines in many cases failed to detect mutations in known caretaker genes or to identify novel putative caretaker genes frequently mutated in tumors [6]. For this reason, an alternative model was proposed, the so-called "oncogene-induced DNA damage model for cancer development", which still relies on genomic instability, but places oncogene driven replication stress at the first stage of tumorigenesis [10]. This model postulates the activation of an oncogene which deregulates entry into the cell cycle as the key initial step of cancer development. Activated oncogenes can induce a state of replication stress, causing frequent replication fork collapse which, in turn, leads to DSBs. Consistently, both precancerous and cancerous lesions exhibit a persistent DNA damage response indicative of DSBs [8]. The genomic instability thus generated can subsequently lead to the loss of growth restrictions (typically by checkpoints, apoptosis and senescence), which marks the transition from precancerous to cancerous lesion. These two apparently conflicting models can be reconciled taking into account the differences between hereditary and sporadic cancers, [6]. The mutator hypothesis, can explain very effectively the origin of hereditary cancers: here, a germline mutation in a genome stability gene is already present in all patient's cells, and therefore a single mutation is required to inactivate the other allele, resulting in genomic instability. And

indeed, caretaker genes (mainly DNA repair or mitotic checkpoint genes) were often found mutated in hereditary cancers [6]. Instead, the oncogene-induced DNA replication stress model is best suitable to explain the genesis of sporadic cancers, not lastly due to the fact that activated oncogenes are generally dominant [123]. In this view, deregulation of a growth-regulating gene leads to replication stress and DNA damage, which cause genomic instability and subsequently all the other cancer hallmarks. Accordingly, in sporadic cancers the most frequently mutated or deregulated genes were found to be classical oncogenes or tumor suppressors [6]. It is worth noting that, despite differing in the initial event placed at the basis of carcinogenesis, both models rely upon genomic instability as a key factor for tumor development.

1. The replication fork at the center of genome instability

During replication, DNA is most vulnerable, and its integrity is jeopardized by a series of events that may perturb replication fork progression. Accordingly, replication failures emerged as one of the main sources of genomic instability, due to the generation of both ssDNA gaps and DSBs by multiple mechanisms [5] (Fig. 1). Moreover, replication of a nicked template inevitably results in the generation of a DSB [18]. When a replication fork encounters an obstacle on the leading strand that prevents its progression (such obstacle may be a DNA adduct, a protein, a DNA secondary structure or the transcription machinery), uncoupling between replicative helicases and polymerases occurs and large ssDNA stretches are generated: this situation is defined as "replication fork stalling" [19]. If the replisome remains associated with the stalled fork, resumption of DNA synthesis can occur after removal of the obstacle. Conversely, if the stalled fork is not properly stabilized or the obstacle is not removed, the replisome disassembles, resulting in "replication fork collapse", with the subsequent generation of ssDNA gaps and DSBs [20, 21]. Moreover, in case of replication fork stalling or uncoupling between leading-strand and lagging-strand synthesis [22], the fork can reverse forming a Holliday junction-like structure known as "chicken foot" [21, 23]: this structure can revert back to a normal fork, it can be cleaved resulting in a DSB [24], or it can be processed by nucleases to generate a stretch of ssDNA [25] (Fig. 1). Alternatively, replication forks can encounter a lesion on the template strand which prevents DNA synthesis without impairing fork progression. If the lesion is on the lagging strand, a ssDNA gap is left between two neighboring Okazaki fragments; if the lesion is on the leading strand DNA, synthesis can resume past the obstacle, leaving a ssDNA gap behind [26]. DSBs are potential sources of GCRs, due to the different ways in which they are processed and repaired. Direct repair by non-homologous end joining (NHEJ) or DSB

processing (resection) and subsequent channeling in a recombination-mediated repair pathway – such as classical homologous recombination (HR), synthesis-dependent strand annealing (SDSA), break-induced replication (BIR) or single-strand annealing (SSA) – may result in translocations, interstitial deletions, inversions, duplications and insertions [5]. Moreover, a DSB end can undergo de-novo telomere addition by telomerase, resulting in a terminal deletion [27] (Fig. 1). Also ssDNA generated at replication forks is a potential source of genomic instability. Indeed, experimental observations suggest that ssDNA itself can be recombinogenic, even without being converted to a DSB [28, 29]: therefore, the long ssDNA stretches generated as a consequence of a perturbed replication could be a source of hyper-recombination (Fig. 1).

2. Unusual secondary structures

The DNA structure described by Watson and Crick is the canonical right-handed double helical structure called B-form DNA. Non-B-form secondary structures have been found to occur at specific DNA sequences. Hairpins can form at inverted repeats or trinucleotide repeats (TNRs) on single-stranded DNA [30]; similarly, inverted repeats longer than 6 nucleotides can adopt a cruciform structure, which is made up of two hairpin-loop arms and a 4-way junction, resembling a Holliday junction [31]. Three-stranded triplex DNA structures are formed when a single-stranded DNA region binds in the major groove of purine-rich double-stranded B-DNA, leaving its complementary strand unpaired [32]. Repetitive G-rich sequences can form G-quartets, in which 4 guanines are arranged in a planar square, and multiple stacks of G-quartets yield a G-quadruplex (G4) DNA structure, which is usually stabilized by monovalent cations [13]. These non-canonical secondary structures are hotspots for genomic instability: hairpin formation is involved in TNRs instability [33]; triplex DNA structures are intrinsically mutagenic, likely because they cause DSBs that result in translocations [34]; cruciform structures formed at palindrome sequences are implicated in DSB-induced translocations [35]; G-quadruplexes stabilization induces DNA damage, and some rearrangement breakpoints have been mapped at sequences prone to G quadruplex formation [36, 37]. The main mechanism through which all these unusual secondary structures jeopardize genomic stability has been deduced through the observation that they can impair replication fork progression, inducing DSBs and ssDNA gaps [37–40]. Interestingly, non-canonical secondary structures-induced genomic instability was detected also in non-proliferating cells, suggesting alternative mechanisms to replication fork impairment. DNA repair processes expose ssDNA tracts, favoring the formation of non-B-form secondary structures, which might interfere with the repair process itself, resulting in GCRs. Moreover, unusual secondary structures may be recognized as helix-distorting lesions, and cleavage by the repair machinery may occur, yielding deletions and GCRs. The ssDNA which is exposed after triplex DNA formation may itself be a recombinogenic intermediate. In addition,

noncanonical secondary structures seem to be more susceptible to DNA damage, likely because they impair nucleosome positioning [41]. Finally, unusual secondary structures interfere with transcription leading to accumulation of R-loops (see ahead) which are linked to increased genome instability [11]. Importantly, telomeric regions are particularly prone to secondary structure formation, which is one of the causes of telomere instability.

2.1.The two faces of G-quadruplexes

Recent studies on G-quadruplexes (G4s) demonstrated their occurrence *in vivo*, as well as their concomitant physiological roles and detrimental effects on genome integrity [16, 42]. Importantly G4s formation requires DNA denaturation, thus connecting both their beneficial and harmful effects to ssDNA-exposing processes, such as replication, transcription and recombination. The main hazard associated to G4s resides in their ability to impede replication fork progression, and in their highly recombinogenic nature: indeed, G4-dependent GCRs have been observed [37, 43]. Nonetheless, G4-forming sequences are enriched at specific genomic regions, and specific helicases capable of unwiding G4s assist the replication process at these loci, to counteract G4-associated genomic instability [16]. Actually, it seems that cells have learned how to tolerate the deleterious effects of G4s in order to exploit their potential as regulators of genome function. For instance, G4s form at repeated telomeric regions, where they contribute to telomere capping, thus preventing their recognition as DSBs (see further) [44, 45]. Additionally, G4s occurring at non-coding RNAs transcribed from telomeric regions (TERRAs) are likely important for binding of telomere-capping proteins [46]. The striking observations that over 90% of replication origins in higher eukaryotes contain G4 motifs [47], and that G4s are bound by the Origin Recognition Complex [48] suggest that they are likely structural determinants of replication origins, and they may also act as regulators of origin efficiency. Furthermore, transcription-dependent production of RNA:DNA hybrids at C-rich template sequences, and the concomitant formation of G4s on the displaced untranscribed DNA strand (G-loops) seems to be a key determinant of immunoglobulin class switch recombination [49,50]. Finally, G4 formation was proposed to regulate gene expression, both at the transcriptional level, since G4 motifs are enriched at human promoters and other gene regulatory elements [51, 52], and at the post-transcriptional level, due to the potential translation regulation ability of G4s observed in the 5'-UTR of many mRNAs [53].

3. Telomeres and genome instability

The ends of linear eukaryotic chromosomes, called telomeres, are made up of long tracts of repeated sequences (referred to as TG repeats), extending from few hundreds bps in yeast to several Kbs in humans, and terminating in a 3' single-

stranded overhang (also known as “G-tail”) [54]. These structures pose multiple problems for genome integrity. First of all, if not properly protected, they can be mis-recognized as DSBs (discussed further). Moreover, the semi-conservative mode of replication is responsible for the loss of single-stranded DNA on the lagging strand at the chromosome end (a situation known as “end-replication problem”) [54], which, together with resection of the 5' end to generate the 3' single-stranded overhang [55], results in loss of telomeric sequences at each round of replication (telomere erosion). To overcome these threats to genome integrity, multiple mechanisms exist to ensure proper telomere maintenance: on one hand, a nucleoprotein structure protects the telomere from unscheduled reactions and masks it from recognition by the DNA damage response (a phenomenon defined “telomere capping”); on the other hand, a specific complex named telomerase adds short TG-rich repeats to chromosome ends, restoring proper length [54].
Telomere erosion actually occurs in somatic cells due to insufficient expression of telomerase [56]. When a telomere shortens below a certain threshold, it loses its protective cap, it is recognized as a DSB, and repaired through the non-homologous end joining (NHEJ) pathway [57]; the same effect is obtained after telomere uncapping, linked to defects in telomere-capping proteins [58].

Typically, repair of an uncapped telomere results in telomere fusion, either with the sister chromatid, or with another uncapped chromosome end; alternatively, chromosome fusion may occur between the uncapped telomere and a DSB end, producing a translocation. These events may give rise to a series of chromosomal aberrations through a mechanisms named breakage-fusion-bridge (BFB). This mechanisms, which involves the fusion of two chromosomes resulting in a dicentric chromosome, followed by breaking during chromosome segregation, can generate translocations and gene amplification [14]. In addition to progressive telomere erosion, also sporadic telomere deletions occur, in which large tracts of telomeric repeats are lost in a single deletion event [59]. These events are due to the repetitive nature of telomeric sequences, and may result from unequal sister chromatid exchange or replication slippage [60]. Moreover, repetitive telomeric sequences are particularly sensitive to oxidative lesions [61] and are prone to formation of secondary structures [62]; both events lead to replication fork stalling, with subsequent DSBs and telomeric deletion events [63]. Furthermore, it was demonstrated that mammalian telomeric regions resemble fragile sites (see ahead), which are prone to breakage upon replication stress, again resulting in telomere loss and GCRs [64]. As in the case of telomere shortening, sporadic telomere deletions may as well trigger BFB cycles [65].

4. DNA replication fidelity

The accuracy of DNA synthesis according to the Watson-Crick base pairing rules is a key aspect in the transmission of an intact genetic information. For this reason, multiple biochemical mechanisms ensure the fidelity of replicative polymerases (eukaryotic Pol α, Pol δ and Pol ε): selectivity for the insertion of the correct nucleotide is provided by base-base hydrogen bonding, water

exclusion from the catalytic site, and a steric selection on base pair shape and size within the active site [9]. Moreover, many DNA polymerases possess intrinsic proofreading activity, based on a higher efficiency of these enzymes in extending a matched primer compared to an unmatched one, and on an exonuclease activity that allows the excision of the mispaired nucleotide [9]. Despite these fidelity-ensuring systems, errors in base incorporation may occur during replication [66]: it was estimated that the base substitution error rate of replicative polymerases *in vivo* is in the range of 10^{-7} to 10^{-8} [7]. Furthermore, insertions or deletions of single bases may result from strand misalignement, a process that is strongly favored during replication of repetitive sequences [67]. In addition, to replicate past a damaged template, cells in most cases use error-prone translesion synthesis (TLS) polymerases (namely Pol ζ, Pol η, Pol ι and Pol κ), which can accommodate helix distorting modified bases in their active site, at the expense of fidelity in nucleotide incorporation [68].

4.1. Ribonucleotides misincorporation in DNA

Recently, another potential source of genome instability linked to replication was described. Replicative polymerases can incorporate at high rates ribonucleotides (rNTPs) instead of deoxyribonucleotides (dNTPs) during DNA synthesis; this is facilitated by the higher rNTPs levels over dNTPs in the cell (300-3,000 μM for rNTPs and 5-30 μM for dNTPs both in yeast and mammalian cells [69]). Recent *in vitro* estimates yielded a likely incorporation of one ribonucleotide every ~700 bp of replicated DNA in budding yeast and ~1/ 7,600 in mammalian cells [70, 71], which makes incorporation of ribonucleoside monophosphates (rNMPs) the most frequent DNA lesion. rNMPs embedded in DNA can jeopardize genome stability in multiple ways: first, due to the reactive hydroxyl group at the 2' position, RNA is 100000-fold more prone to hydrolysis than DNA under physiological conditions [72]. Moreover, the presence of rNMPs alters DNA helix parameters [73, 74], and this distortion may constitute an obstacle to replication fork progression, resulting in replication stress [71, 75, 76]. Indeed Pols α, δ and ε can replicate past a single ribonucleotide with reduced efficiency [77, 78]. Finally, it was directly demonstrated that increased ribonucleotide incorporation in DNA causes genomic instability at short tandem repeats in yeast, due to mutagenic enzymatic processing of the incorporated ribonucleotides by Topoisomerase I activity [79, 80]. Given the genotoxicity of chromosomal rNMPs, it is somewhat surprising that replicative polymerases did not evolve to avoid these misincorporation events in DNA. Conversely, the addition of rNMPs during DNA replication might have been tolerated due to possible physiological advantages. Indeed, this may be envisioned as a system to distinguish the parental and the newly synthesized DNA strand in eukaryotic cells. In fact, differently from what happens in *E. coli* and other Gram-negative bacteria, this distinction does not rely on DNA methylation as a strand discrimination signal. This hypothesis was proven true in the case of mismatch repair (MMR), where rNMPs processing by RNaseH activities on the daughter leading strand creates an entry

point for exonuclease activities involved in mismatch removal, thus ensuring that the new and not the template strand is corrected [81, 82]. It is tempting to envisage a role for genomic rNMPs in other processes requiring DNA strand discrimination, such as the nonrandom segregation of sister chromatids during stem cell renewal: indeed, mounting evidence suggests asymmetrical inheritance of DNA strands between the new stem cell and the committed-to-differentiation daughter cell [83,84]. In this regard, both the "immortal strand hypothesis" (the undamaged template strand is always retained in the stem cell to protect it from replication errors) [83] and the "silent sister hypothesis" (epigenetic differences due to asymmetric chromosome segregation determine the daughter cells' fate after division of a stem cell) [84] postulate mechanisms of discrimination between DNA strands at the centromere, thus enabling strand-specific segregation of chromatids. Such mechanisms may (also) rely on rNMPs embedded in the daughter DNA strand. Moreover, hotspots of rNMPs incorporation, if found, might represent a form of "genomic imprinting" for specific loci, similarly to what happens for mating-type switch initiation in *Schizosaccharomyces pombe*, where the incorporation of two rNMPs at a specific position within the *mat1* locus triggers a replication-coupled recombination event ultimately resulting in cell-type change [85].

5. Chromosomal fragile sites

Chromosomal fragile sites are defined as specific loci which undergo frequent gaps or breaks under replication stress, and are hotspots of chromosome rearrangements in tumor cells [12, 86]. Fragile sites are conserved from yeast to mammals [12]. Chromosomal fragile sites have been divided in three classes with distinct features. Rare fragile sites are observed in less than 5% of the cases and are inherited in a mendelian fashion: they are constituted of microsatellite trinucleotide repeats (TNRs) or AT-rich minisatellite repeats, and are often associated with genetic diseases caused by repeat expansion [12, 33]. Common fragile sites (CFSs), instead, are present in all individuals, where they represent normal components of chromosome structure; they generally contain AT-rich sequences, but not nucleotide repeats [12, 87]. Recently, a third class was identified and named "early-replicating fragile sites" (ERFSs), which have opposite properties compared to CFSs (see below) [88]. The mechanisms underlying fragility of these sites and consequent genome instability differ for the three classes mentioned above (Fig. 2).

5.1. Rare fragile sites

The instability of TNRs and AT-rich minisatellites relies upon their ability to form unusual secondary structures (such as hairpins, stem-loops or DNA triplexes) during replication (Fig. 2A), which trigger microsatellite instability (MIN): in particular, hairpin structures at the 5'end of a displaced Okazaki fragment during lagging strand synthesis can promote repeat expansion [89];

similarly, secondary structures on the lagging strand can cause replication slippage events, which result in repeat deletions [5]. Moreover, these secondary structures can be processed by nucleases yielding a DSB [90]. Additionally, they can perturb replication fork progression, possibly resulting in gaps or breaks [86].

5.2. Common fragile sites

CFSs are defined as chromosomal regions particularly susceptible to form breaks or gaps during early mitosis (an event defined as "fragile site expression") following replication stress (such as treatment with the DNA polymerase inhibitor aphidicolin) [12]. The molecular mechanisms underlying CFSs fragility was not clearly understood until very recently, when the critical features of CFSs and the mechanisms leading to CFSs breakage have been reported [17, 91] (Fig. 2B). CFSs were initially described as late-replicating regions [92, 93]. Indeed, AT-rich sequences within CFSs are characterized by a high degree of DNA torsional flexibility, which has the potential to form secondary structures capable of impairing replication fork progression [94]. Therefore, a possible mechanism for CFSs instability predicts CFSs as slow-replication genomic regions which favor uncoupling between replicative helicases and polymerases (especially in conditions of replication stress). The subsequent generation of long ssDNA tracts allows secondary structure formation in AT-rich tracts, acting as replication fork barriers, ultimately resulting in fork stalling and/or collapse [12, 86].

Further observations highlighted another feature of CFSs: a paucity in replication initiation. Due to the absence or low efficiency of replication origins, CFSs are often replicated from forks fired in flanking regions, resulting in incomplete replication followed by chromosome breakage under replication stress [95–97]. Interestingly, the density of initiation events is epigenetically determined, which accounts for cell-type differences observed in CFSs fragility [97]. Moreover, a correlation was found between the level of transcription of very large genes at CFSs and the instability of the corresponding site [98]; this mechanism likely relies on the interference between transcription and replication (see ahead).

Strikingly, recent reports revealed that CFSs expression is indeed an active mechanisms relying on the activity of the MUS81-EME1 resolvase complex, which cleaves replication intermediates at CFSs in late G2-mitosis, allowing separation of entangled chromatids and proper chromosome segregation, thus promoting rather than jeopardizing genome integrity [99, 100]. Attempts to segregate unresolved replication intermediates at CFSs give rise to the so-called ultra-fine DNA bridges (UFBs) in anaphase, where a backup mechanism, depending on BLM helicase in complex with TOPIIIa and RMI1/2, can dissolve the intertwined intermediates promoting chromosome segregation [101]. In case an unreplicated intermediate persists, completion of mitosis results in uncontrolled chromosome breakage and/or chromosome mis-segregation (Fig. 2B). Interestingly, these breaks are detectable in the next G1 phase in the form of 53BP1 nuclear bodies, where 53BP1 is thought to protect them from processing.

The exact function of these 53BP1 bodies is still unclear, as it is the mechanisms through which these breaks are eventually repaired [17, 102, 103]. Importantly, the mechanism of CFSs fragility is in line with the oncogene-induced DNA replication stress model for cancer development (Box 1). During the early stages of cancer development, oncogene activation induces replication stress, which may result in chromosome breakage, deletions and rearrangements particularly at CFSs, which precede and likely drive instability in other genomic regions [104]. The model is further reinforced by the finding that some CFSs lie within tumor suppressor genes, suggesting a mechanisms for inactivation of recessive tumor suppressor genes during tumor development [105].

5.3. Early-replicating fragile sites

ERFSs were identified very recently in mouse B lymphocytes, and correspond to stalled and/or collapsed forks in the vicinity of early replication origins, resulting in chromosome breakage during normal replication and especially after replication stress [88]. ERFSs display opposite features than CFSs: they are detected at the beginning of S-phase, are enriched in GC content, repetitive sequences, genes and replication origins. Nevertheless, similarly to CFSs, they are often associated with active transcription, and their fragility is increased after oncogenic stress (Box 1), S-phase checkpoint inhibition and homologous recombination (HR) defects. A likely mechanisms for ERFSs fragility relies on higher origin activation as a consequence of replication perturbation, which is expected to increase the interference between transcription and replication at the highly transcribed genes clustered at ERFSs. In addition, depletion of the dNTPs pool due to a larger number of active origins could contribute to replication failures [88] (Fig. 2C). Strikingly, a substantial proportion of ERFSs identified in mouse B lymphocytes overlaps with rearrangements observed in human B cell lymphomas, pointing towards a conservation of this class of fragile sites in mammals [88], and a causative role for ERFSs in B-cells carcinogenesis. Interestingly, fragile sites corresponding to early-replicating regions had been previously identified in yeast cells defective in S-phase checkpoint, and named “compromised early origins” (CEOs) [106]. Such finding suggests that ERFSs may be a widespread feature of eukaryotic genomes. Moreover, given that the transcriptional landscape and the replication timing distribution changes among different cell types, it is postulated that cell-line specific ERFSs may account for rearrangements observed in other cancers [88]. Recent years, experimental evidence demonstrated unexpected connections between transcription and genomic instability [107]: indeed, high rates of transcription at a genomic locus correlate with increased mutations, a phenomenon known as “transcription-associated mutation” (TAM) [108,109]. Similarly, highly transcribed regions show a greater recombination frequency, a phenomenon referred to as “transcription-associated recombination” (TAR) [110, 111]. Both outcomes are likely related to interference between transcription and replication, as well as to the generation of ssDNA [5] (Fig. 3).

As a consequence of local negative supercoiling during transcription, DNA-strand opening and ssDNA generation occurs behind an elongating RNA polymerase. Given that ssDNA is less stable and more susceptible to mutagenic damage from endogenous or environmental sources than dsDNA [1, 107], this is a potential mechanism for TAM, but likely not the only one, since it cannot explain the observation that the non-transcribed strand (NTS) is more prone to mutations than the transcribed strand (TS) [109, 112]. A further possibility involves the formation of R-loops, a three strand nucleic acid structure formed by annealing the just transcribed RNA on its template, resulting in an RNA:DNA hybrid plus a displaced DNA strand (ssDNA) [11]. Co-transcriptional R-loops occurrence is favored by negative supercoiling and high G content, both of which promote DNA
duplex unwinding; moreover, uncoupling between transcription and mRNA processing and/or nuclear export (caused by defects in protein complexes involved in these activities) also stimulates R-loop formation [11, 113]. These structures may account for the preferential mutagenesis of the NTS, as the displaced NTS is single-stranded, while the TS forms the RNA:DNA hybrid.
Moreover, since persistent RNA:DNA hybrids can induce replication [114], it was proposed [11] that R-loops might trigger unscheduled DNA synthesis, which is expected to be highly mutagenic, in agreement with the finding that break-induced replication (BIR) is extremely inaccurate [115]. Collisions between the replication fork and the transcription machinery cause replication fork impairment, DSBs and TAR [116, 117]. Specifically, due to generation of positive DNA supercoiling by both processes, head-on collisions between transcription and replication have deleterious effects, since topological constraints are induced, which promote replication fork stalling and fork reversal. Such topological constrains may be further increased by the observed tethering of highly transcribed genes at the nuclear pore: this process, named “gene gating”, if on one side facilitates mRNA export, thus limiting R-loop formation, on the other induces torsional stress by restraining DNA mobility [118]. Moreover, other possible mechanisms for the induction of TAR and transcription-associated GCRs rely upon the ability of co-transcriptional R-loops to cause replication fork blockage in multiple ways: i) unrepaired damage on the displaced strand might impede DNA polymerase progression; ii) replication fork progression could be impaired by the RNA:DNA hybrid itself, or by a RNA polymerase blocked by the R-loop; iii) R-loop formation could allow the occurrence of secondary structures on the displaced NTS, resulting in a barrier to DNA polymerase; iv) torsional stress generated in front of a R-loop could cause replication fork reversal, thus generating a highly recombinogenic “chicken foot” structure. In all the indicated cases, the final outcome would be replication fork stalling and/or collapse, with generation of DSBs or ssDNA gaps capable of triggering hyper-recombination and GCRs [11]. Furthermore, it was suggested that the attempt to bypass a R-loop occurring between direct repeats through a template-switch mechanism, inevitably results in the deletion of the intervening region [113].

7. Conclusions and prospects

Physiological DNA and RNA transactions pose specific challenges to genome integrity: indeed, specific chromosomal regions (such as telomeres, sequences prone to secondary structure formation and chromosomal fragile sites) or processes (such as transcription and rNTP incorporation in the genome) interfere with DNA replication leading to DSBs and genomic rearrangements. It is worth noting that all the mechanisms described above cannot be considered as separate causes of genome instability. Rather, these processes are heavily intertwined and might be envisioned as different aspects of the multifaceted "dark side" of normal nucleic acid metabolism. An overview of the mechanisms and their interplay through which nucleic acid metabolism impacts on genome integrity is depicted in Figure 4.

Despite recent insights, several aspects of this complex picture remain to be elucidated. In particular, random versus biased rNTPs incorporation in the genome needs to be assessed, and the putative detection of hotspots could open new perspectives on the physiological roles of rNMPs in the genome. Moreover, the exact nature and function of 53BP1 bodies arising after chromosome breakage at CFSs have to be investigated. Finally, the recently detected correlation between cotranscriptional R-loops and chromatin condensation [119] may suggest that R-loops are not just transcriptional by-products, but they may have regulatory functions which deserve to be explored. The same studies may shed new light on the connection(s) between transcription, R-loops, chromatin condensation and CFSs, since high transcription levels and premature chromatin condensation were both shown to increase breakage at CFSs [87, 120]. These findings could also provide a deeper mechanistic understanding of CFSs fragility. The answer to these questions will give us a better picture of the interplay between nucleic acid metabolism and genomic integrity, and will impact on our understanding of the carcinogenic process and, possibly, on its treatment.

Acknowledgements

The authors apologize for the many interesting papers that they were not able to discuss or acknowledge due to space limitations. Work in the authors' lab is supported by grants from MIUR AIRC, Regione Lombardia - Progetto NEPENTE. The financial support of Telethon-Italy (grant number GGP11003) is gratefully acknowledged.

Bibliography

1. **Lindahl T**. 1993. Instability and decay of the primary structure of DNA. *Nature* **362**: 709–15.
2. **De Bont R, van Larebeke N**. 2004. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**: 169–85.
3. **Lindahl T, Wood RD**. 1999. Quality control by DNA repair. *Science* **286**: 1897–905.
4. **Schärer OD**. 2005. DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* **6**: 27–32.
5. **Aguilera A, Gómez-González B**. 2008. Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* **9**: 204–17.
6. **Negrini S, Gorgoulis VG, Halazonetis TD**. 2010. Genomic instability – an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**: 220–8.
7. **Loeb LA**. 1991. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**: 3075–9.
8. **Gorgoulis VG, Vassiliou L-VF, Karakaidos P, Zacharatos P,** et al. 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*
434: 907–13.
9. **Kunkel TA**. 2004. DNA replication fidelity. *J Biol Chem* **279**: 16895–8.
10. **Halazonetis TD, Gorgoulis VG, Bartek J**. 2008. An oncogene-induced DNA damage model for cancer development. *Science* **319**: 1352–5.
11. **Aguilera A, García-Muse T**. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**: 115–24.
12. **Durkin SG, Glover TW**. 2007. Chromosome fragile sites. *Annu Rev Genet* **41**: 169–92.
13. **Bochman ML, Paeschke K, Zakian VA**. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**: 770–80.
14. **Murnane JP**. 2012. Telomere dysfunction and chromosome instability. *Mutat Res* **730**: 28– 36.
15. **Dalgaard J**. 2012. Causes and consequences of ribonucleotide incorporation into nuclear DNA. *Trends Genet.* **28**: 592–7.
16. **Maizels N, Gray LT**. 2013. The G4 genome. *PLoS Genet* **9**: e1003468.
17. **Pellicioli A, Muzi-Falconi M**. 2013. A blooming resolvase at chromosomal fragile sites. *Nat Cell Biol* **15**: 883–5.
18. **Cortés-Ledesma F, Aguilera A**. 2006. Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep* **7**: 919–26.
19. **Carr AM, Paek AL, Weinert T**. 2011. DNA replication: failures and inverted fusions. *Semin Cell Dev Biol* **22**: 866–74.
20. **Lopes M, Cotta-Ramusino C, Pellicioli A, Liberi G,** et al. 2001. The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412**: 557–61.

21. **Sogo JM, Lopes M, Foiani M**. 2002. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297**: 599–602.
22. **Pagés V, Fuchs RP**. 2003. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo. *Science* **300**: 1300–3.
23. **Postow L, Ullsperger C, Keller RW, Bustamante C,** et al. 2001. Positive torsional strain causes the formation of a four-way junction at replication forks. *J Biol Chem* **276**: 2790–6.
24. **Heller RC, Marians KJ**. 2006. Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol* **7**: 932–43.
25. **Cotta-Ramusino C, Fachinetti D, Lucca C, Doksani Y,** et al. 2005. Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol Cell* **17**:153–9.
26. **Lopes M, Foiani M, Sogo JM**. 2006. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* **21**: 15–27.
27. **Pennaneach V, Putnam CD, Kolodner RD**. 2006. Chromosome healing by de novo telomere addition in Saccharomyces cerevisiae. *Mol Microbiol* **59**: 1357–68.
28. **Fabre F, Chan A, Heyer W-D, Gangloff S**. 2002. Alternate pathways involving Sgs1/Top3, Mus81/Mms4, and Srs2 prevent formation of toxic recombination intermediates from singlestranded gaps created by DNA replication. *Proc Natl Acad Sci U S A* **99**: 16887–92.
29. **Lettier G, Feng Q, de Mayolo AA, Erdeniz N,** et al. 2006. The role of DNA double-strand breaks in spontaneous homologous recombination in S. cerevisiae. *PLoS Genet* **2**: e194.
30. **Nag DK, Petes TD**. 1991. Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in Saccharomyces cerevisiae. *Genetics* **129**: 669–73.
31. **Palecek E**. 1991. Local supercoil-stabilized DNA structures. *Crit Rev Biochem Mol Biol* **26**: 151–226.
32. **Htun H, Dahlberg JE**. 1988. Single strands, triple strands, and kinks in H-DNA. *Science* **241**: 1791–6.
33. **López Castel A, Cleary JD, Pearson CE**. 2010. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* **11**: 165–70.
34. **Wang G, Vasquez KM**. 2004. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A* **101**: 13448–53.
35. **Inagaki H, Ohye T, Kogo H, Kato T,** et al. 2009. Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res* **19**: 191–8.
36. **Bacolla A, Jaworski A, Larson JE, Jakupciak JP,** et al. 2004. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* **101**: 14162–7.

37. **Lopes J, Piazza A, Bermejo R, Kriegsman B,** et al. 2011. G-quadruplex-induced instability during leading-strand replication. *EMBO J* **30**: 4033–46.
38. **Krasilnikova MM, Mirkin SM**. 2004. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol* **24**: 2286–95.
39. **Mirkin E V, Mirkin SM**. 2007. Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* **71**: 13–35.
40. **Voineagu I, Narayanan V, Lobachev KS, Mirkin SM**. 2008. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A* **105**: 9936–41.
41. **Wang G, Vasquez KM**. 2009. Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol Carcinog* **48**: 286–98.
42. **Biffi G, Tannahill D, McCafferty J, Balasubramanian S**. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5**: 182–6.
43. **Paeschke K, Bochman ML, Garcia PD, Cejka P,** et al. 2013. Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**: 458–62.
44. **Schaffitzel C, Berger I, Postberg J, Hanes J,** et al. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A* **98**: 8572–7.
45. **Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM,** et al. 2011. Rudimentary G-quadruplexbased telomere capping in Saccharomyces cerevisiae. *Nat Struct Mol Biol* **18**: 478–85.
46. **Biffi G, Tannahill D, Balasubramanian S**. 2012. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J Am Chem Soc* **134**: 11974–6.
47. **Gilbert DM**. 2012. Replication origins run (ultra) deep. *Nat Struct Mol Biol* **19**: 740–2.
48. **Hoshina S, Yura K, Teranishi H, Kiyasu N,** et al. 2013. Human Origin Recognition Complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J Biol Chem* **288**: 30161–71.
49. **Duquette ML, Handa P, Vincent JA, Taylor AF,** et al. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **18**: 1618–29.
50. **Larson ED, Duquette ML, Cummings WJ, Streiff RJ,** et al. 2005. MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Curr Biol* **15**: 470–4.
51. **Huppert JL, Balasubramanian S**. 2007. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**: 406–13.
52. **Du Z, Zhao Y, Li N**. 2009. Genome-wide colonization of gene regulatory elements by G4 DNA motifs. *Nucleic Acids Res* **37**: 6784–98.
53. **Bugaut A, Balasubramanian S**. 2012. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**: 4727–41.

54. **Blackburn EH, Greider CW, Szostak JW**. 2006. Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nat Med* **12**: 1133–8.
55. **Lingner J, Cooper JP, Cech TR**. 1995. Telomerase and DNA end replication: no longer a lagging strand problem? *Science* **269**: 1533–4.
56. **Harley CB, Futcher AB, Greider CW**. 1990. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**: 458–60.
57. **Smogorzewska A, Karlseder J, Holtgreve-Grez H, Jauch A,** et al. 2002. DNA ligase IVdependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr Biol* **12**: 1635–44.
58. **Van Steensel B, Smogorzewska A, de Lange T**. 1998. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**: 401–13.
59. **Lustig AJ**. 2003. Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat Rev Genet* **4**: 916–23.
60. **Baird DM, Jeffreys AJ, Royle NJ**. 1995. Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J* **14**: 5433–43.
61. **Von Zglinicki T**. 2002. Oxidative stress shortens telomeres. *Trends Biochem Sci* **27**: 339–44.
62. **Parkinson GN, Lee MPH, Neidle S**. 2002. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**: 876–80.
63. **Lansdorp PM**. 2005. Major cutbacks at chromosome ends. *Trends Biochem Sci* **30**: 388–95.
64. **Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL,** et al. 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**: 90–103.
65. **Lo AWI, Sabatier L, Fouladi B, Pottier G,** et al. 2002. DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. *Neoplasia* **4**: 531–8.
66. **McCulloch SD, Kunkel TA**. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**: 148–61.
67. **Streisinger G, Okada Y, Emrich J, Newton J,** et al. 1966. Frameshift mutations and the genetic code. *Cold Spring Harb Symp Quant Biol* **31**: 77–84.
68. **Friedberg EC, Wagner R, Radman M**. 2002. Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**: 1627–30.
69. **Nick McElhinny SA, Watts BE, Kumar D, Watt DL,** et al. 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* **107**: 4949–54.
70. **Sparks JL, Chon H, Cerritelli SM, Kunkel TA,** et al. 2012. RNase H2-initiated Ribonucleotide Excision Repair. *Mol Cell* **47**: 980–6.

71. **Reijns M, Rabe B, Rigby R, Mill P**. 2012. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**: 1008–22.
72. **Li Y, Breaker RR**. 1999. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.* **121**: 5364–72.
73. **Jaishree TN, van der Marel GA, van Boom JH, Wang AH**. 1993. Structural influence of RNA incorporation in DNA: quantitative nuclear magnetic resonance refinement of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG). *Biochemistry* **32**: 4903–11.
74. **DeRose EF, Perera L, Murray MS, Kunkel TA,** et al. 2012. Solution structure of the Dickerson DNA dodecamer containing a single ribonucleotide. *Biochemistry* **51**: 2407–16.
75. **Lazzaro F, Novarina D, Amara F, Watt DL,** et al. 2012. RNase H and Postreplication Repair protect cells from ribonucleotides incorporated in DNA. *Mol Cell* **45**: 99–110.
76. **Hiller B, Achleitner M, Glage S**. 2012. Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J. Exp. Med.* **209**: 1419–26.
77. **Nick McElhinny S a, Watts BE, Kumar D, Watt DL,** et al. 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 4949–54.
78. **Watt DL, Johansson E, Burgers PM, Kunkel TA**. 2011. Replication of ribonucleotidecontaining DNA templates by yeast replicative polymerases. *DNA Repair* **10**: 897–902.
79. **Nick McElhinny SA, Kumar D, Clark AB, Watt DL,** et al. 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* **6**: 774–81.
80. **Kim N, Huang SN, Williams JS, Li YC,** et al. 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* **332**: 1561–4.
81. **Ghodgaonkar MM, Lazzaro F, Olivera-Pimentel M, Artola-Borán M,** et al. 2013. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell* **50**: 323–32.
82. **Lujan SA, Williams JS, Clausen AR, Clark AB,** et al. 2013. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol Cell* **50**: 437–43.
83. **Rando TA**. 2007. The immortal strand hypothesis: segregation and reconstruction. *Cell* **129**: 1239–43.
84. **Lansdorp PM**. 2007. Immortal strands? Give me a break. *Cell* **129**: 1244–7.
85. **Vengrova S, Dalgaard J**. 2005. The wild-type *Schizosaccharomyces pombe mat1* imprint consists of two ribonucleotides. *EMBO Rep.* **7**: 59–65.
86. **Lukusa T, Fryns JP**. 2008. Human chromosome fragility. *Biochim Biophys Acta* **1779**: 3– 16.

87. **Debatisse M, Le Tallec B, Letessier A, Dutrillaux B,** et al. 2012. Common fragile sites: mechanisms of instability revisited. *Trends Genet.* **28**: 22–32.
88. **Barlow JH, Faryabi RB, Callén E, Wong N,** et al. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152**: 620–32.
89. **Freudenreich CH, Kantrow SM, Zakian VA**. 1998. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**: 853–6.
90. **Lobachev KS, Gordenin DA, Resnick MA**. 2002. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**: 183–93.
91. **Ozeri-Galai E, Bester AC, Kerem B**. 2012. The complex basis underlying common fragile site instability in cancer. *Trends Genet.* **28**: 1–8.
92. **Le Beau M, Rassool F**. 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. *Hum. Mol. Genet.* **7**: 755–61.
93. **Hellman A, Rahat A, Scherer SW, Darvasi A,** et al. 2000. Replication delay along *FRA7H*, a common fragile site on human chromosome 7, leads to chromosomal instability. *Mol Cell Biol* **20**: 4420–7.
94. **Zlotorynski E, Rahat A, Skaug J, Ben-Porat N,** et al. 2003. Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol* **23**: 7143–51.
95. **Lemoine FJ, Degtyareva NP, Lobachev K, Petes TD**. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell* **120**: 587–98.
96. **Lemoine FJ, Degtyareva NP, Kokoska RJ, Petes TD**. 2008. Reduced levels of DNA polymerase delta induce chromosome fragile site instability in yeast. *Mol. Cell. Biol.* **28**: 5359–68.
97. **Letessier A, Millot GA, Koundrioukoff S, Lachagès A-M,** et al. 2011. Cell-type-specific replication initiation programs set fragility of the *FRA3B* fragile site. *Nature* **470**: 120–3.
98. **Helmrich A, Ballarino M, Tora L**. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**: 966–77.
99. **Naim V, Wilhelm T, Debatisse M, Rosselli F**. 2013. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat Cell Biol* **15**: 1008–15.
100. **Ying S, Minocherhomji S, Chan KL, Palmai-Pallag T,** et al. 2013. MUS81 promotes common fragile site expression. *Nat Cell Biol* **15**: 1001–7.
101. **Chan KL, Palmai-Pallag T, Ying S, Hickson ID**. 2009. Replication stress induces sisterchromatid bridging at fragile site loci in mitosis. *Nat Cell Biol* **11**: 753–60.
102. **Lukas C, Savic V, Bekker-Jensen S, Doil C,** et al. 2011. 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat Cell Biol* **13**: 243–53.

103. **Harrigan J, Belotserkovskaya R**. 2011. Replication stress induces 53BP1-containing OPT domains in G1 cells. *J. Cell Biol.* **193**: 97–108.
104. **Tsantoulis PK, Kotsinas A, Sfikakis PP, Evangelou K,** et al. 2008. Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene* **27**: 3256–64.
105. **Bignell GR, Greenman CD, Davies H, Butler AP,** et al. 2010. Signatures of mutation and selection in the cancer genome. *Nature* **463**: 893–8.
106. **Raveendranathan M, Chattopadhyay S, Bolon Y-T, Haworth J,** et al. 2006. Genomewide replication profiles of S-phase checkpoint mutants reveal fragile sites in yeast. *EMBO J* **25**: 3627–39.
107. **Aguilera A**. 2002. The connection between transcription and genomic instability. *EMBO J* **21**: 195–201.
108. **Datta A, Jinks-Robertson S**. 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**: 1616–9.
109. **Beletskii A, Bhagwat A**. 1996. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 13919–24.
110. **Thomas BJ, Rothstein R**. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**: 619–30.
111. **Nickoloff JA**. 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol Cell Biol* **12**: 5311–8.
112. **Skandalis A, Ford BN, Glickman BW**. 1994. Strand bias in mutation involving 5- methylcytosine deamination in the human *hprt* gene. *Mutat Res* **314**: 21–6.
113. **Gómez-González B, Felipe-Abrio I, Aguilera A**. 2009. The S-phase checkpoint is required to respond to R-loops accumulated in THO mutants. *Mol. Cell. Biol.* **29**: 5203–13.
114. **Kogoma T**. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* **61**: 212–38.

Figure legends

Fig. 1. The replication fork at the heart of genome instability.
Replication fork stalling and replication fork collapse are main sources of genome instability, due to the generation of both ssDNA gaps and DSBs by multiple mechanisms. Dependening on the different repair pathways to which they are channeled, ssDNA gaps and DSBs are sources of gross chromosomal rearrangements (GCRs), interstitial deletions, gene amplification or hyperrecombination (see main text for details).

Fig. 2. Mechanisms of instability at chromosomal fragile sites.
A: Rare fragile sites. Microsatellite trinucleotide repeats (TNRs) or AT-rich minisatellite repeats are prone to formation of secondary structures (such as hairpins, cruciforms, triplexes and Gquadruplexes). These structures may induce replication slippage and expansion or contraction of the repeats (microsatellite instability), impair replication fork progression with the subsequent generation of ssDNA and DSBs, or can be cleaved by nucleases yielding DSBs and GCRs.
B: Common fragile sites (CFSs). Different features contribute to the instability of CFSs, namely secondary structure formation, paucity in replication origin firing, and high rates of transcription. The combination of these elements results in replication fork impairment and in late replication of these regions. The MUS81-EME1 resolvase cleaves unreplicated intermediates preventing the formation of ultra-fine DNA bridges (UFBs) in anaphase, when the BLM helicase complex can still act to resolve entangled intermediates thus promoting chromosome segregation and preventing DSBs formation and consequent GCRs.
C: Early-replicating fragile sites (ERFSs). Increased origin firing after replication stress in early
S-phase and the interference between transcription and replication cause replication fork block and
GCRs at ERFSs (see main text for details).

Fig. 3. Transcription-associated genomic instability.
Highly transcribed genes are a source of genomic instability, mainly due to the formation of cotranscriptional R-loops. These structures expose tracts of ssDNA (which is more susceptible to genotoxic lesions) and impair replication fork progression causing hyper-recombination, DSBs and subsequent GCRs (see main text for details).

Fig. 4. Nucleic acid metabolism impacts on genome integrity through multiple interdependent mechanisms.
A compendium of all the mechanistic causes of genome instability arising from normal DNA transactions (such as replication and transcription) is presented, highlighting the interconnections between these processes and their specific chromosomal outcomes. Polymerase errors, rNTP misincorporation during replication, intense transcription, telomeres, chromosomal fragile sites and DNA secondary structures threaten genome integrity, mainly by impacting on replication fork progression, with the subsequent generation of ssDNA gaps and DSBs. Specific classes of genomic instability due to nucleic acid metabolism include microsatellite instability (MIN), gross chromosomal rearrangements (GCRs), point mutations, hyper-recombination, gene amplification and interstitial deletions (see main text for details).

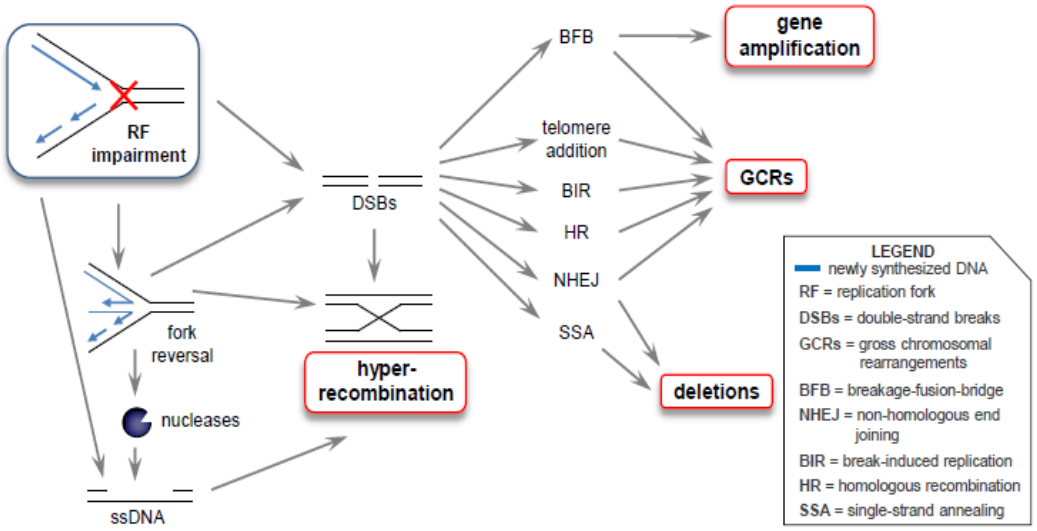


Fig. 1. The replication fork at the heart of genome instability.
Replication fork stalling and replication fork collapse are main sources of genome instability, due to the generation of both ssDNA gaps and DSBs by multiple mechanisms. Dependening on the different repair pathways to which they are channeled, ssDNA gaps and DSBs are sources of gross chromosomal rearrangements (GCRs), interstitial deletions, gene amplification or hyper-recombination (see main text for details).
82x39mm (300 x 300 DPI)

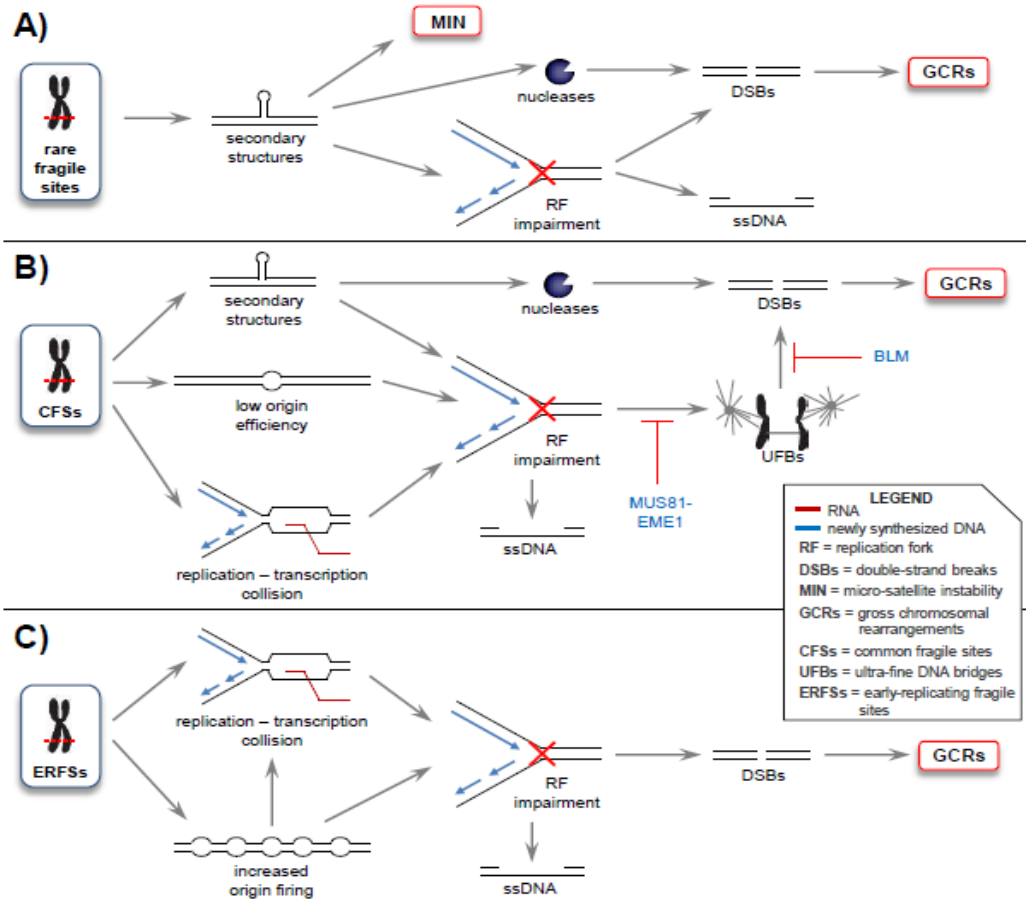


Fig. 2. Mechanisms of instability at chromosomal fragile sites.
A: Rare fragile sites. Microsatellite trinucleotide repeats (TNRs) or AT-rich minisatellite repeats are prone to formation of secondary structures (such as hairpins, cruciforms, triplexes and G-quadruplexes). These structures may induce replication slippage and expansion or contraction of the repeats (microsatellite instability), impair replication fork progression with the subsequent generation of ssDNA and DSBs, or can be cleaved by nucleases yielding DSBs and GCRs.
B: Common fragile sites (CFSs). Different features contribute to the instability of CFSs, namely secondary structure formation, paucity in replication origin firing, and high rates of transcription. The combination of these elements results in replication fork impairment and in late replication of these regions. The MUS81-EME1 resolvase cleaves unreplicated intermediates preventing the formation of ultra-fine DNA bridges (UFBs) in anaphase, when the BLM helicase complex can still act to resolve entangled intermediates thus promoting chromosome segregation and preventing DSBs formation and consequent GCRs.
C: Early-replicating fragile sites (ERFSs). Increased origin firing after replication stress in early S-phase and the interference between transcription and replication cause replication fork block and GCRs at ERFSs (see main text for details).
161x152mm (300 x 300 DPI)

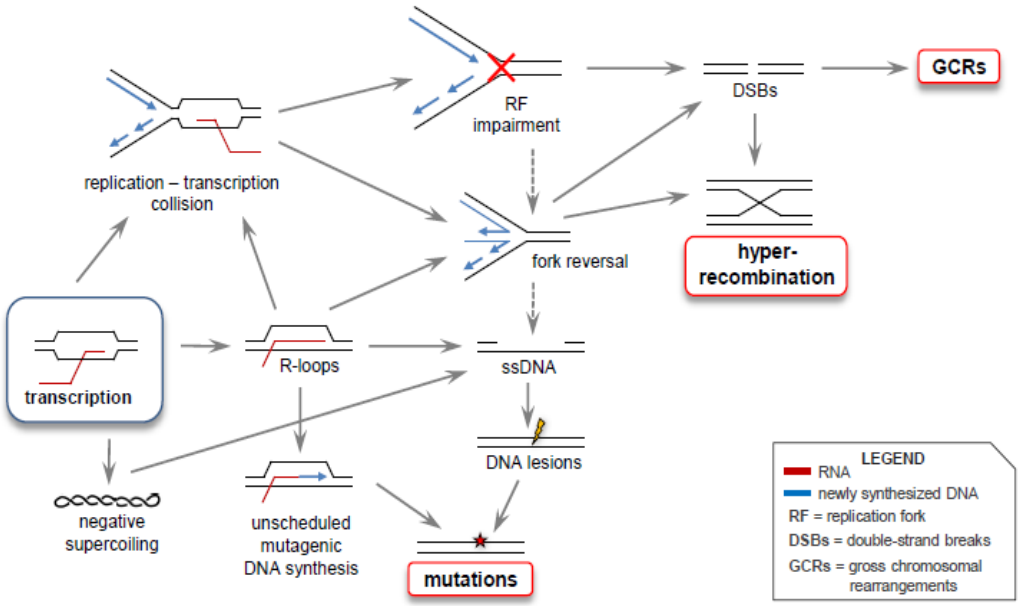


Fig. 3. Transcription-associated genomic instability.
Highly transcribed genes are a source of genomic instability, mainly due to the formation of cotranscriptional R-loops. These structures expose tracts of ssDNA (which is more susceptible to genotoxic lesions) and impair replication fork progression causing hyper-recombination, DSBs and subsequent GCRs
(see main text for details).
94x53mm (300 x 300 DPI)

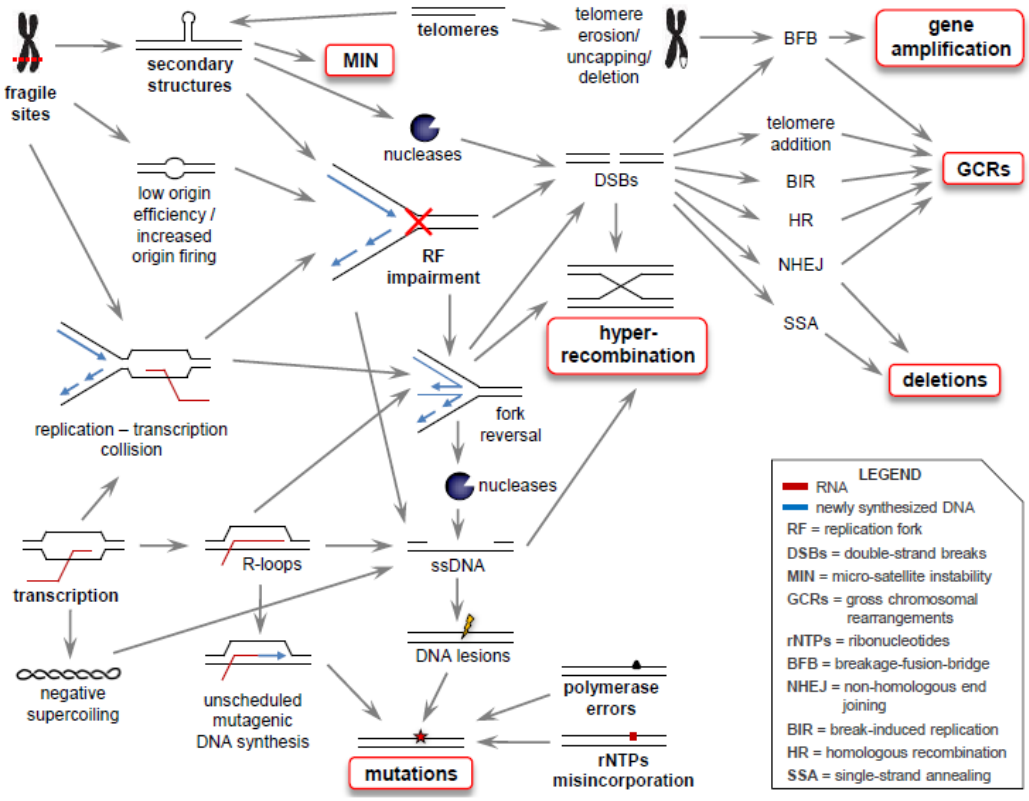


Fig. 4. Nucleic acid metabolism impacts on genome integrity through multiple interdependent mechanisms.
A compendium of all the mechanistic causes of genome instability arising from normal DNA transactions (such as replication and transcription) is presented, highlighting the interconnections between these processes and their specific chromosomal outcomes. Polymerase errors, rNTP misincorporation during replication, intense transcription, telomeres, chromosomal fragile sites and DNA secondary structures threaten genome integrity, mainly by impacting on replication fork progression, with the subsequent generation of ssDNA gaps and DSBs. Specific classes of genomic instability due to nucleic acid metabolism include microsatellite instability (MIN), gross chromosomal rearrangements (GCRs), point mutations, hyper recombination, gene amplification and interstitial deletions (see main text for details).
129x99mm (300 x 300 DPI)

Part III

Appendix

In this appendix I will summarize the results obtained during the first year of my Ph.D. program when I have been working on a project aimed at the identification of RNase H2 interactors using the yesat two-hybrid technology.

Aim of the project

Investigating the role of RNase H in the maintenance of genome stability in *S. cerevisiae*.

Ribonucleases H are enzymes that cleave the RNA mojety in RNA/DNA hybrids that form during replication and repair and that may lead to DNA instability (Stein H, Hausen P. et al., 1969).

Based on amino acid sequence and biochemical properties, two main types of RNases H are found in eukaryotes and classified RNase H1 or RNase H2. The two class of enzymes differ for their substrate specificity. In fact, RNase H1 requires at least four ribonucleotides for enzymatic activity, while RNase H2 is unique in its capacity to excise single ribonucleotides embedded in genomic DNA (Bjoern Hiller et al., 2012; Crouch RJ & Dirksen ML 1982; Lazzaro et al., 2012; Cerritelli SM & Crouch RJ. 2009 and Ohtani N. et al., 1999). Mutations in the *RNASEH2A, RNASEH2B, RNASEH2C* genes encoding the three subunits of the RNase H2 protein complex are the most frequent genetic alterations identified in Aicardi-Goutieres (AGS) patients, an autosomal recessive disorder phenotypically mimicking congenital viral infection with elevated interferon-α level in cerebrospinal fluid (Aicardi J, Goutieres F. 1984; Goutieres F, Aicardi J, Barth PG, Lebon P. 1998; Goutieres F. 2005, Crow etal., 2006; Crow and Rehwinkel , 2009).

Previous research carried out in our lab, showed that RNases H are important for removal of genomic rNTPs incorporated during replication in the yeast *S. cerevisiae* (REF), revealing new role for RNases H in genome stability maintenance mechanisms. We were thus interested in identifying potential physical and genetic interactors with RNases H in *S. cerevisiae*. To start addressing this question, I used the yeast two-hybrid system.

RNase H in *S. cerevisiae*. To address this question, I used yeast two hybrid system.

The yeast two-hybrid system

The yeast two-hybrid system (Y2H) was originally developed by Fields and Song in 1989 and it is the most widely used method to identify an interaction between two proteins (Young K. 1998). The method is based on the properties of several transcription factors, such as the GAL4 protein, which consists of two separable domains responsible for DNA-binding and transcriptional activation: a N-terminal domain which binds to specific DNA sequences (binding domain/BD), and a C-terminal domain containing acidic regions, necessary to activate transcription (activator domain/AD). In the system we used, a LexA DNA-binding domain (BD) is fused to a protein "X" (bait), and the B42 activating region (AD) is fused to a protein "Y" (prey). Thus, if the proteins X and Y interact, the BD and AD domains are brought in closed proximity through the interaction between proteins X and Y resulting in the induced transcription of a reporter gene (fig. 1). Conversely, no reporter expression will be observed if the two transcription factor modules are not brought into close contact.

Basic steps to follow for two hybrid system:

1. Construction of the BD-X fusion (Bait-LexA) plasmid
2. Transformation of BD-X fusion plasmid into yeast
3. Transactivation test of reporter gene of BD-X fusion plasmid in yeast
4. Checking the expression of X protein from BD-X fusion plasmid
5. Cross with yeast library carrying AD-Y prey plasmids
6. Screening of putative positives for the activation of reporter genes
7. Identification of false positives and analysis of true positives
8. DNA seqence analysis of positives

Merits and demerits of yeast two hybrid:

Advantages of the yeast two hybrid system:

1. Direct identification of DNA sequence of interacting protein
2. Protein purification not necessary
3. *In vivo*-protein in native conformation
4. Detection of low affinity or transient interactions

Disadvantages:

1. Gene encoding target protein must be available
2. Bait and prey must be soluble for nuclear localization
3. Independent verification of interaction is recommended
4. False positives can be found, as the bait or prey can transactivate per se the reporter gene.
5. Stable expression of fusion protein might be a problem

In our yeast two-hybrid screening we used: plasmid pSH18-34 as a LacZ reporter plasmid with upstream 8 LexA operators; plasmid pEG202 as the bait plasmid (LexA-Bait fusion protein) and pJG4-5 as the prey plasmid

with B42 activation domain under the control of the *GAL1* promoter. In this system, expression of the prey is induced in galactose containing media and it is repressed in glucose. We use the EGY48 (*MAT alpha, his3, trp1, ura3, 6LexAop-LEU2*) yeast strain throughout the yeast two-hybrid procedure.

The following steps have been carried out in thre yeast two-hybrid procedure:

1. We started with the construction of our bait by cloning full length genes (RNH1, RNH201, RNH202 and RNH203) coding, respectively, for RNase H1 and the three subunits of RNase H2 in plasmid pEG202. The selected baits were constructed downstream to the LexA binding protein region in order to produce a LexA-Bait fusion protein. All the constructed genes in pEG202 were transformed in the EGY48 strain.
2. We checked the expression of all fusion-proteins (LexA-Rnh1, LexA-Rnh201, LexA-Rnh202, and LexA-Rnh203) with an empty vector as control. The fusion proteins were checked using antibody alpha-LexA (fig. 2).
3. We checked the transactivation of reporter genes using a leucine assay (for the Leu2 reporter) and in X-gal (for the LacZ reporter) by transformating the constructed bait plasmid with reporter plasmid pSH18-34 and empty prey plasmid pJG4-5 (fig. 3) in EGY48 yeast strain. In leucine assays, we observed a very weak transactivation by LexA-Rnh1, weak transactivation by LexA-Rnh201, medium transactivation by LexA-Rnh202 and strong transactivation by LexA-Rnh203 (fig. 4) compared to the positive and negative controls. In X-Gal condition from the patch, we observed no transactivation by RNH1 and RNH201 and very weak transactivationo with RNH202 and strong transactivation with RNH203 (fig. 5) compared to the controls (positive controls: EGY48 , pEG202-p53 pJG4-5-Ta-SV40, pSH18-34

and negative control: EGY48, pEG202 pSH18-34 pJG4-5 .We decided to start with RNH202 because it is coding for the catalytic subunit.

4. To resolve the transactivation problem with leucine reporter, we decided to add X-Gal in the screening plates instead to do the patch assay. In this way the colonies growing in the absence of leucine and blue are the putative positives, while the transactivated colonies are white.

   We found that the colonies were blue in positive control plates and RNH202 plated colonies were white (fig. 6), suggesting that the bait RNH202 is not transactivating enough the LacZ reporter and that RNH202 can be used for a two-hybrid screening. We performed the screening with bait RNH202 (fig. 7) by mating the yeast Mat a strain (Bait + DBD) with the Mat α strain (cDNA Library + AD) and subsequent growth on selective plates.

Main Results

1. The number of diploids = -H-U-T Raff = 13000000
2. The number of colonies obtained from a total 43 plates (-H-U-T-L Gal / Raff + X-Gal) = 9249
3. From 9249 colonies, 112 blue colonies were selected for isolation from –H-U-T-L Gal/Raff.

10 blue colonies (i.e., 4, 7, 8, 15, 24, 29, 28, 37, 39 and 76) were selected which were showing higher blue staining in +L Gal/Raff + X-gal plates and white color in +L Raff + X-gal. Unfortunately the other patches (other then 10 selected blue colonies) were blue also in +L Raff + X-gal plates (fig. 8) compared to the controls used for the two-hybrid screening (positive controls: EGY48 , pEG202-p53 pJG4-5-Ta-SV40,pSH18-34 and negative control: EGY48, pSH18-34 pMH16-1 pSG4-5. Unfortunately, after purification and transformation of putative positives prey plasmid, they did not reconfirm the results. We decided to stop the project.

Figures

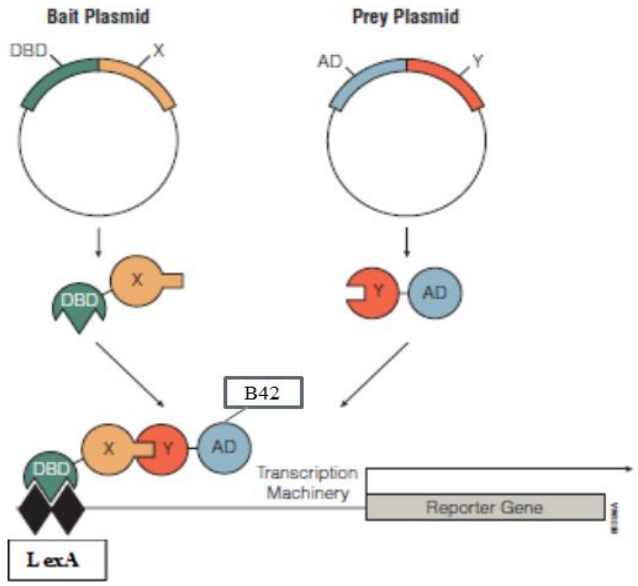
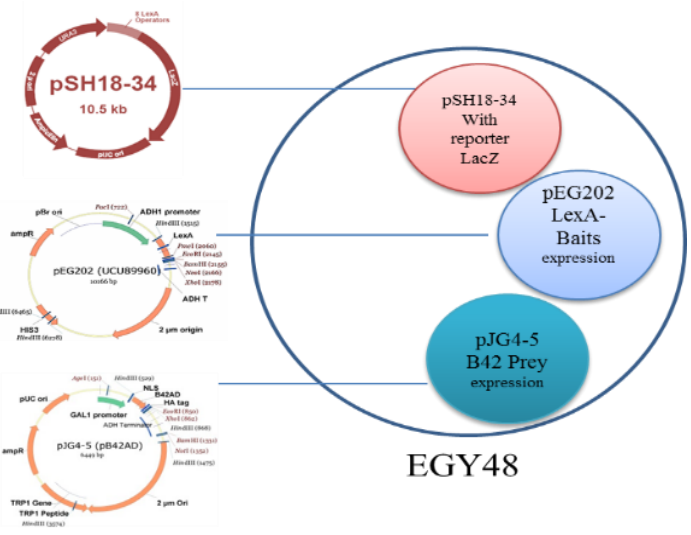


Figure 1. Principle of the yeast two-hybrid system: The protein of interest (X-Bait protein) is fused to the DNA-binding domain (DNA-BD) of a transcription factor. Protein Y is fused to the transcriptional activation domain (TA-Prey protein). Both fusions are targeted to the nucleus. If protein X and Y interact, the transcription factor activity is reconstituted and activate the expression of a reporter gene.



Size of the plasmids

- pEG202: 10160bp
- pSH18-34: 10484bp
- pJG4-5: 6449bp

Figure 2. ***Transformation of the pEG202 bait plasmid together with the pSH18-34 reporter plasmid and pJG4-5 empty prey plasmid in yeast EGY48***

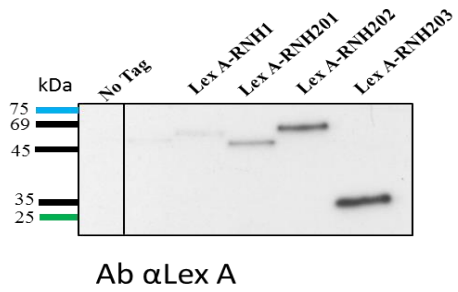


Figure 3. ***Analysis of the expression of Baits (*RNH1, RNH201, RNH202 *&* RNH203*) and no tag as control by western blotting.***

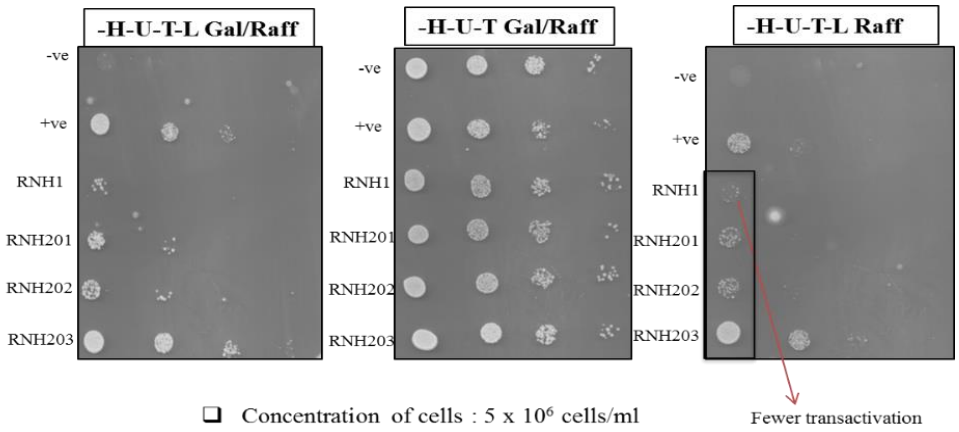


Figure 4. Leucine assay: ***RIGHT: in the absence of –H-U-T-L Raff, transactivation of genes were as follows: a very weak transactivation of*** **RNH1,** ***weak transactivation of*** **RNH201,** ***medium transactivation of*** **RNH202** ***and strong transactivation of*** **RNH203** ***compared to both positive (+ve) and negative (–ve) control. LEFT: (-H-U-T-L Gal/Raff) shows higher transactivation of all baits compared to –H-U-T-L Raff plate (RIGHT) and MIDDLE plates (-H-U-T Gal/Raff), as control plate shows complete growth of all genes with control in the presence of leucine.***

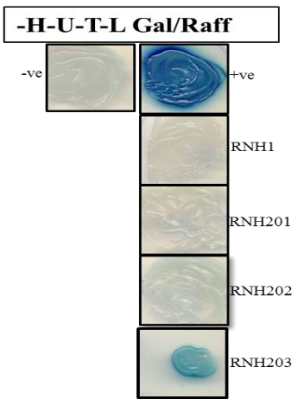


Figure 5. X-Gal assay ***shows no transactivation of*** **RNH1** ***and*** **RNH201,** ***weak transactivationo with*** **RNH202** ***and strong transactivation of*** **RNH203** ***compared to both +ve and -ve control***

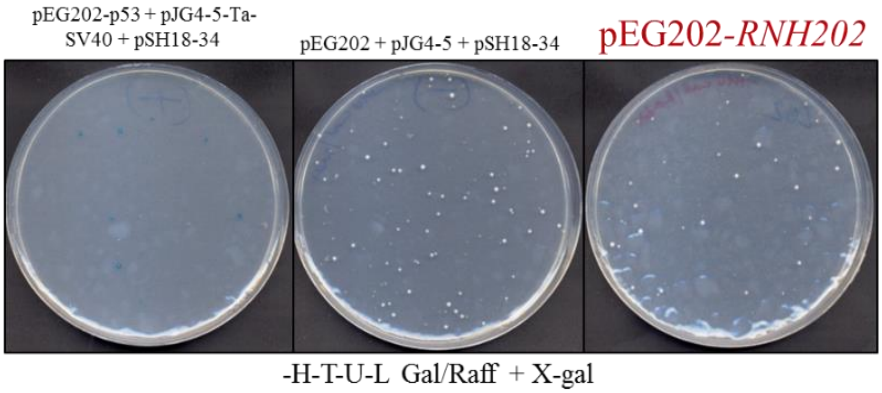


Figure 6. ***In the plates it is possible to distinguish the positive blue control (left side), the middle is a negative control showing white color and*** **RNH202** ***(right side), shows white colonies***

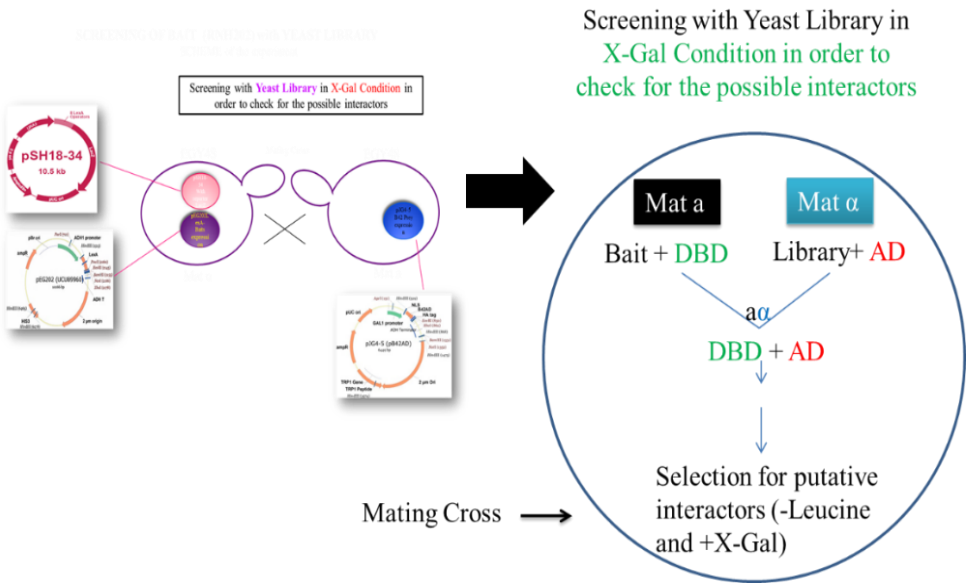


Figure 7. ***Screening of Bait (*****RNH202*****) Mat 'a'with Mat 'α' yeast library in X-gal condition***

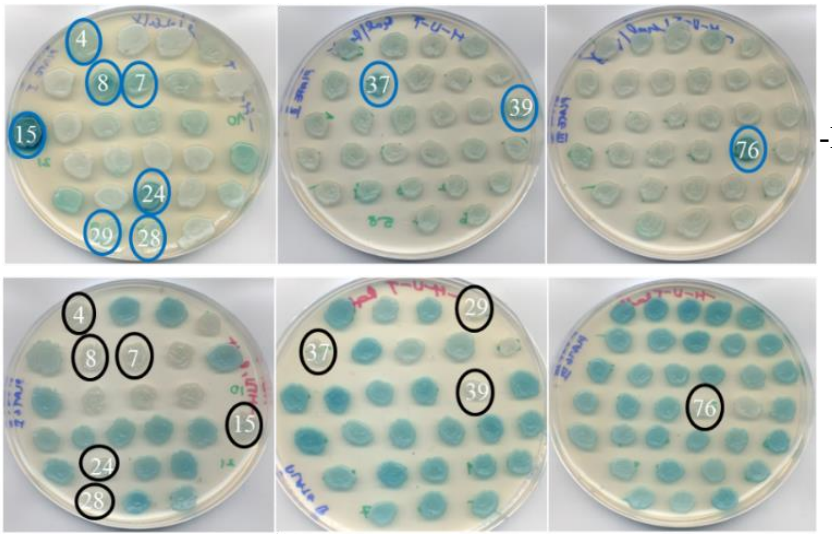


Figure 8. ***Interaction of bait (*RNH202*) wth prey after yeast two-hybrid screening;circles represents the interaction (blue color) in –H-U-T Gal and no inteaction (white color) in –H-U-T Raff***