CHAPTER 2

DNA Repair: a preview

What is DNA Repair?

DNA repair refers to a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome. In human cells, both normal metabolic activities and environmental factors such as UV light and radiation can cause DNA damage, resulting in as many as 1 million individual molecular lesions per cell per day. Many of these lesions cause structural damage to the DNA molecule and can alter or eliminate the cell's ability to transcribe the gene that the affected DNA encodes. Other lesions induce potentially harmful mutations in the cell's genome, which affect the survival of its daughter cells after it undergoes mitosis. Consequently, the DNA repair process is constantly active as it responds to damage in the DNA structure.

The rate of DNA repair is dependent on many factors, including the cell type, the age of the cell, and the extracellular environment. A cell that has accumulated a large amount of DNA damage, or one that no longer effectively repairs damage incurred to its DNA, can enter one of three possible states:

1. an irreversible state of dormancy, known as senescence
2. cell suicide, also known as apoptosis
3. unregulated cell division, which can lead to the formation of cancer

The DNA repair ability of a cell is vital to the integrity of its genome and thus to its normal functioning and that of the organism. Many genes that were initially shown to influence lifespan have turned out to be involved in DNA damage repair and protection. Failure to correct molecular lesions in cells that form gametes can introduce mutations into the genomes of the offspring and thus influence the rate of evolution.

DNA damage

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 1,000 to 1,000,000 molecular lesions per cell per day. While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs), unrepaired lesions in critical genes (such as tumor suppressor genes) can impede a cell's ability to carry out its function and appreciably increase the likelihood of tumor formation.

The vast majority of DNA damage affects the primary structure of the double helix; that is, the bases themselves are chemically modified. These modifications can in turn disrupt the molecules' regular helical structure by introducing non-native chemical bonds or bulky adducts that do not fit in the standard double helix. DNA is, however, supercoiled and wound around "packaging" proteins called histones (in eukaryotes), and both superstructures are vulnerable to the effects of DNA damage.

If the rate of DNA damage exceeds the capacity of the cell to repair it, the accumulation of errors can overwhelm the cell and result in early senescence, apoptosis or cancer. Inherited diseases associated with faulty DNA repair functioning result in premature aging, increased sensitivity to carcinogens, and correspondingly increased cancer risk

Types of DNA Damage

All four of the bases in DNA (A, T, C, and G) can be covalently modified at various positions. One of the most frequent is the loss of an amino group ("deamination") resulting, for example, in a C being converted to a U.

Mismatches of the normal bases because of a failure of proofreading during DNA replication can occur. An example is incorporation of the pyrimidine U (normally found only in RNA) instead of T.

Breaks in the backbone which can be limited to one of the two strands (a single-stranded break, SSB) or on both strands (a double-stranded break (DSB). Ionizing radiation is a frequent cause, but some chemicals produce breaks as well.

Crosslinks or covalent linkages can be formed between bases on the same DNA strand ("intrastrand") or on the opposite strand ("interstrand"). Several chemotherapeutic drugs used against cancers crosslink DNA.

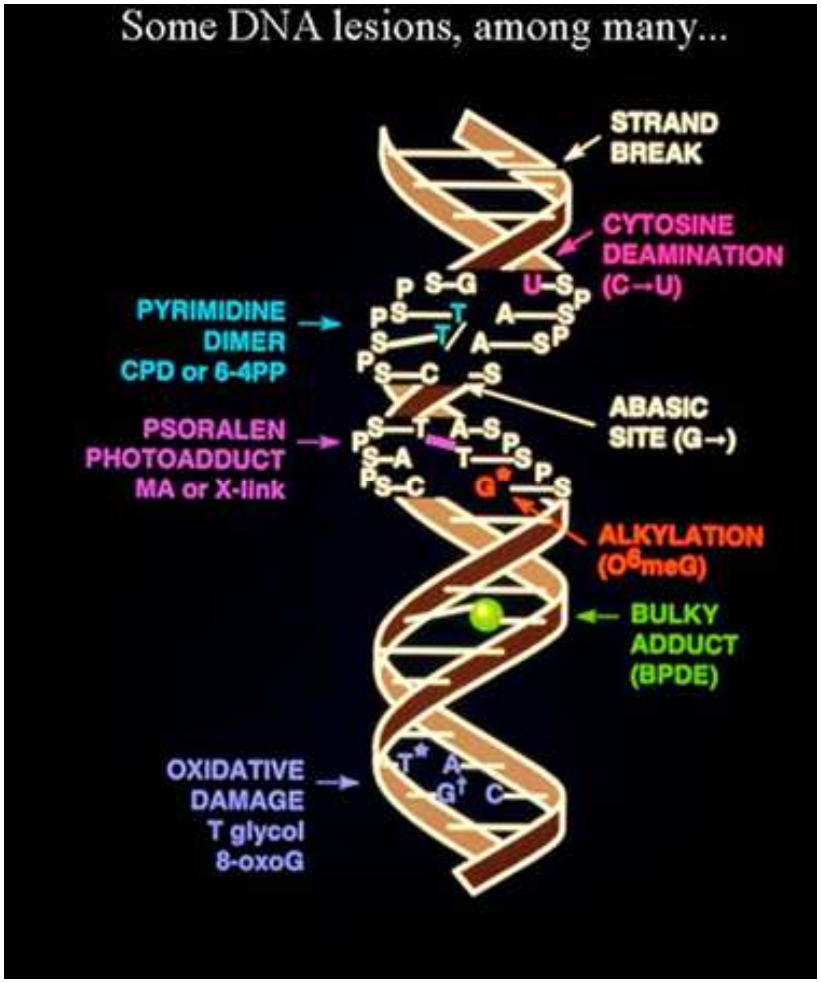

Fig 2.1

Types of DNA damage

Damages are physical abnormalities in the DNA, such as single and double strand breaks, 8-hydroxydeoxyguanosine residues and polycyclic aromatic hydrocarbon adducts (Fig 2.1). DNA damages can be recognized by enzymes, and thus they can be correctly repaired if redundant information, such as the undamaged sequence in the complementary DNA strand or in a homologous chromosome, is available for copying. If a cell retains DNA damage, transcription of a gene can be prevented and thus translation into a protein will also be blocked. Replication may also be blocked and/or the cell may die.

DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair and these errors are a major source of mutation. DNA damages are a special problem in non-dividing or slowly dividing cells, where unrepaired damages will tend to accumulate over time. On the other hand, in rapidly dividing cells, unrepaired DNA damages that do not kill the cell by blocking replication will tend to cause replication errors and thus mutation. The great majority of mutations that are not neutral in their effect are deleterious to a cell's survival. Thus, in a population of cells comprising a tissue with replicating cells, mutant cells will tend to be lost. However infrequent mutations that provide a survival advantage will tend to clonally expand at the expense of neighboring cells in the tissue. This advantage to the cell is disadvantageous to the whole organism, because such mutant cells can give rise to cancer. Thus DNA

damages in frequently dividing cells, because they give rise to mutations, are a prominent cause of cancer. In contrast, DNA damages in infrequently dividing cells are likely a prominent cause of aging.

DNA repair mechanisms

Cells cannot function if DNA damage corrupts the integrity and accessibility of essential information in the genome (but cells remain superficially functional when so-called "non-essential" genes are missing or damaged). Depending on the type of damage inflicted on the DNA's double helical structure, a variety of repair strategies have evolved to restore lost information. If possible, cells use the unmodified complementary strand of the DNA or the sister chromatid as a template to losslessly recover the original information. Without access to a template, cells use an error-prone recovery mechanism known as translesion synthesis as a last resort.

Damage to DNA alters the spatial configuration of the helix and such alterations can be detected by the cell. Once damage is localized, specific DNA repair molecules bind at or near the site of damage, inducing other molecules to bind and form a complex that enables the actual repair to take place. The types of molecules involved and the mechanism of repair that is mobilized depend on the type of damage that has occurred and the phase of the cell cycle that the cell is in.

Damaged or inappropriate bases can be repaired by several mechanisms:

1) Direct chemical reversal of the damage
2) Excision repair, in which the damaged base or bases are removed and then replaced with the correct ones in a localized burst of DNA synthesis. There are three modes of excision repair, each of which employs specialized sets of enzymes
    a) Base Excision Repair (BER) which repairs damage to a single nucleotide caused by oxidation, alkylation, hydrolysis, or deamination. The base is removed with glycosylase and ultimately replaced by repair synthesis with DNA ligase.
    b) Nucleotide Excision Repair (NER) which repairs damage affecting longer strands of 2–30 bases. This process recognizes bulky, helix-distorting changes such as thymine dimers as well as single-strand breaks
    c) Homologous Recombination Repair (HRR) in which both strands in the double helix are severed, are particularly hazardous to the cell because they can lead to genome rearrangements

Base Excision Repair (BER)

The pathway most commonly employed to remove incorrect bases (like uracil) or damaged bases (like 3-methyladenine) is called base excision repair. Actually, it's misleading to talk about this as a pathway, because there are numerous variations, each specific for a different type of incorrect base.

Nevertheless, all of the variant pathways have features in common, and each of the pathways can be considered to consist of 3 steps, with steps 2 and 3 being common for all pathways:

1. Removal of the incorrect base by an appropriate DNA N-glycosylase to create an AP site
2. Nicking of the damaged DNA strand by AP endonuclease upstream of the AP site, thus creating a 3'-OH terminus adjacent to the AP site
3. Extension of the 3'-OH terminus by a DNA polymerase, accompanied by excision of the AP site

Specificity of the various pathways is conferred by the DNA N-glycoslyases. These hydrolyze the N-glycosylic bond between the base and the deoxyribose,

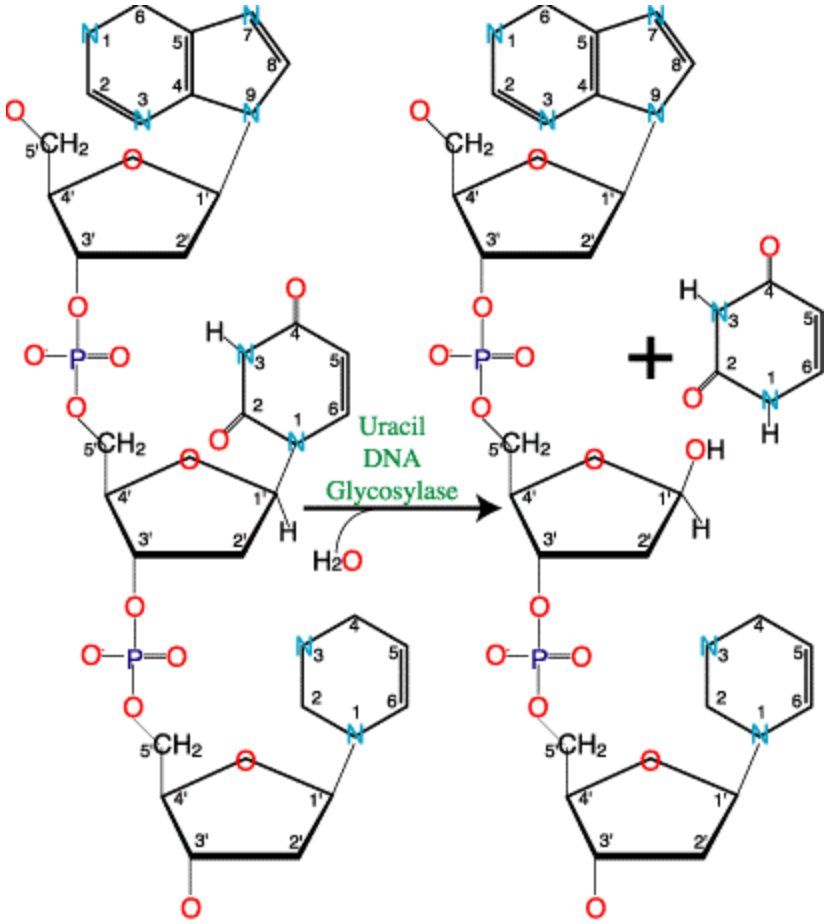

Fig 2.2 Action of uracil DNA N-glycosylase in BER

The next step in BER is catalyzed by an AP endonuclease, which cleaves the DNA backbone on the 5' side of the AP site. The best characterized AP endonuclease is, the *E. coli* enzyme called exonuclease III. The AP endonuclease activity of exonuclease III is important and in fact accounts for about 90% of the total AP endonuclease in the *E. coli* cell. The remaining 10% is accounted for by a different enzyme, endonuclease IV. Similar enzymes are also found in eukaryotic organisms. In *S. cerevisiae*, the major AP endonuclease is related to endo IV, but in mammalian cells, AP

endonucleases related to exonuclease III are more abundant. The final ligation step is carried out by DNA ligase III in partnership with the XRCC1 protein.

Nucleotide Excision Repair

DNA constantly requires repair due to damage that can occur to bases from a vast variety of sources including chemicals but also ultraviolet (UV) light from the sun. Nucleotide excision repair (NER) is a particularly important mechanism by which the cell can prevent unwanted mutations by removing the vast majority of UV-induced DNA damage (mostly in the form of thymine dimers and 6-4-photoproducts). The importance of this repair mechanism is evidenced by the severe human diseases that result from in-born genetic mutations of NER proteins including xeroderma pigmentosum and Cockaynes syndrome. While the base excision repair machinery can recognize specific lesions in the DNA and can correct only damaged bases that can be removed by a specific glycosylase, the nucleotide excision repair enzymes recognize bulky distortions in the shape of the DNA double helix. Recognition of these distortions leads to the removal of a short single-stranded DNA segment that includes the lesion, creating a single-strand gap in the DNA, which is subsequently filled in by DNA polymerase, which uses the undamaged strand as a template.

Nucleotide excision repair can be categorized into two classes, global genome NER (GG-NER) and Transcription Coupled NER (TC-NER). Two different sets of proteins are involved in the distortion and recognition of the DNA damage in the two types of NER. In GG-NER, the XPC-Rad23B complex is responsible for distortion recognition, and DDB1 and DDB2 (XPC) can also recognize some types of damage caused by UV light. Additionally, XPA performs a function in damage recognition that is as yet poorly defined. In TC-NER, CS proteins CSA and CSB bind some types of DNA damage instead of XPC-Rad23B. The subsequent steps in GG-NER and TC-NER are similar to each other. XPB and XPD, which are subunits in transcription factor TFIIH have helicase activity and unwind the DNA at the sites of damages. XPG protein has a structure-specific endonuclease activity, which makes an incision 3’ to the damaged DNA. Subsequently XPF protein, which is associated with ERCC1 makes the 5' incision during the NER. The dual-incision leads to the removal of a ssDNA with a single strand gap of 25~30 nucleotides.

The resulting gap in DNA is filled by DNA pol δ or ε by copying the undamaged strand. Proliferating Cell Nuclear Antigen (PCNA) assists the DNA polymerase in the reaction and Replication Protein A (RPA) protects the other DNA strand from degradation during the NER. Finally, DNA ligase seals the nicks to finish NER.

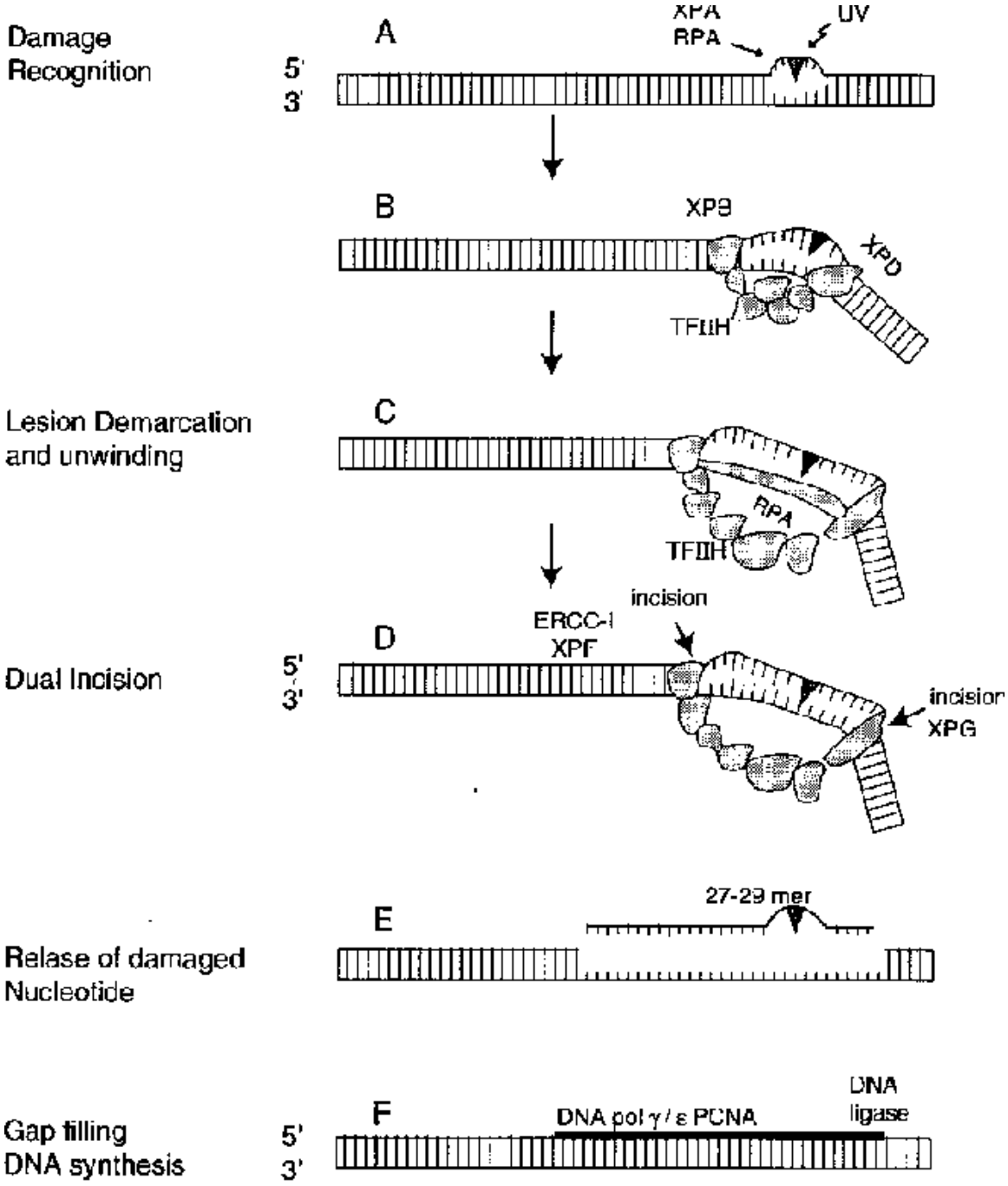

Figure 2.3 - Nucleotide Excision Repair scheme

Homologous Recombination Repair

Double-stranded DNA breaks (DSBs), the most harmful DNA lesions, cause cell death and genome instability. Homologous recombination (HR) is a major pathway that repairs double-stranded DNA breaks (DSBs). The salient feature of the HR mechanism is that it uses homologous DNA sequences as a template to achieve accurate repair of DSBs. In eukaryotes, the initial steps of HR involve processing of broken DNA by exonucleases to generate ssDNA tails, binding of Rad51 protein (RecA homolog) to these tails to form the Rad51 nucleoprotein filament, and searching by this filament for the homologous DNA template to form joint molecules (D-loops.) Once joint molecules are formed, the 3'-ssDNA tails of a broken chromosome are extended by DNA polymerase, restoring the lost information. Afterward, the joint molecules dissociate, leading to re-joining of the broken chromosome through the synthesis-dependent strand annealing (SDSA) pathway. Recently, by reconstituting the process of DSB repair in vitro we demonstrated that Rad54 protein can promote dissociation of D-loops.

Double-strand breaks (DSBs), in which both strands in the double helix are severed, are particularly hazardous to the cell because they can lead to genome rearrangements. Two mechanisms exist to repair DSBs: non-homologous end joining (NHEJ) and recombinational repair (also known as template-assisted repair or homologous recombination repair).

In NHEJ, DNA Ligase IV, a specialized DNA Ligase that forms a complex with the cofactor XRCC4, directly joins the two ends. To guide accurate repair, NHEJ relies on short homologous sequences called microhomologies present on the single-stranded tails of the DNA ends to be joined. If these overhangs are compatible, repair is usually accurate. NHEJ can also introduce mutations during repair. Loss of damaged nucleotides at the break site can lead to deletions, and joining of nonmatching termini forms translocations. NHEJ is especially important before the cell has replicated its DNA, since there is no template available for repair by homologous recombination. There are "backup" NHEJ pathways in higher eukaryotes.

Recombinational repair requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is nearly identical to the machinery responsible for chromosomal crossover during meiosis. This pathway allows a damaged chromosome to be repaired using a sister chromatid (available in G2 after DNA replication) or a homologous chromosome as a template. DSBs caused by the replication machinery attempting to synthesize across a single-strand break or unrepaired lesion cause collapse of the replication fork and are typically repaired by recombination.

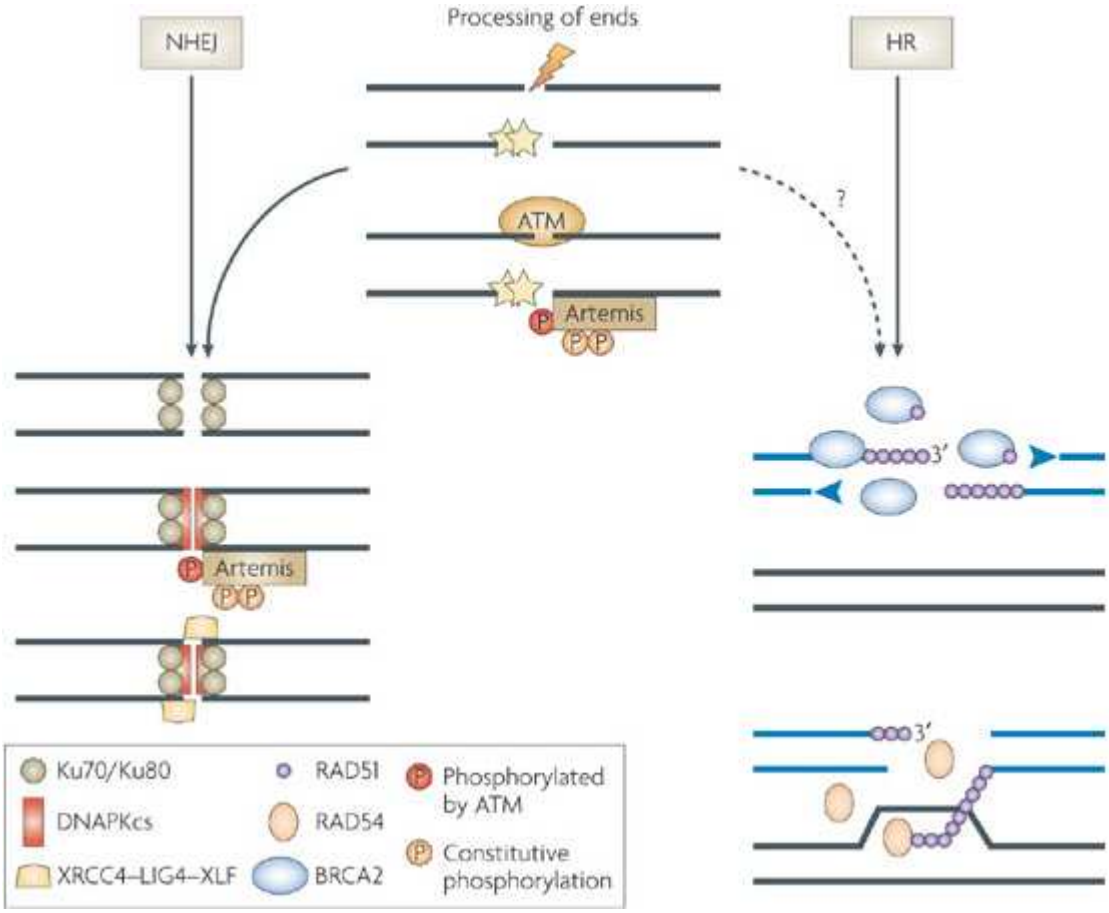

Fig 2.4 Double strand repair pathway

Homologous Recombination in eukaryotes has the additional role of protecting organisms from cancer. In organisms that possess mutations in the pathway, there is a tendency to develop certain cancers.

Single Nucleotide Polymorphism and Cancer

Cancer as a disease involves not only discrete genetic and environmental causes, but also interactions between the two. Even though any two unrelated people share about 99.9% of their DNA sequences, the remaining 0.1% is important because it contains the genetic variants that influence how people differ in their risk of disease or their response to drugs and environmental exposures. Discovering the DNA sequence variants that contribute to disease risk factors offers a good opportunity for understanding disease etiology.

Major heritable disease genes are uncommon in the general population. Much more common are variants of genes termed polymorphisms, which influence various metabolic processes and susceptibility to various environmental exposures. These genetic polymorphisms will not modify an individual's risk of developing disease by themselves, but rather modify the effect of exposure and the damage.

A variation is termed a polymorphism if it is present at an allele frequency higher than 1% in the general population. Genetic polymorphisms are common throughout the genome. The most common type, single nucleotide polymorphisms (SNPs) can occur as frequently as 1 out of every 300 base pairs. Polymorphism can occur in introns and exons of genes and may sometimes have an impact on the structure and function of the protein coded

for by the gene, especially when the polymorphism leads to an amino acid exchange in evolutionarily conserved domains.

A polymorphism that leads to an amino acid exchange, and is found within a active site, at a DNA binding site or any other area of importance for the protein function, may influence the activity of the encoded protein. The structure or affinity of the protein may change and make it nonfunctional or more or less efficient than the original one. If the protein is involved in processes such as DNA repair, cell cycle control or metabolism of toxic substances, a change in function may be associated with a different susceptibility to disease in a carrier of the variant allele. For example, a less efficient DNA repair may increase the risk of having a permanent damage, like a mutated base, chromosomal aberration or micronuclei, if the cell has not had enough time to repair before going through mitosis.

Polymorphisms also help to identify the genetic traits of different diseases. Another aspect of studying polymorphisms is that they may help to identify at risk groups of individuals that are more susceptible to certain environmental exposures than others. This type of knowledge may in future allow us to give individualized preventive advice before disease diagnosis, or offering personalized treatment after the disease has been diagnosed.

Investigated polymorphisms in DNA Repair genes

The repair of damaged DNA is essential to prevent mutations to become fixed in the genome of a cell and thus prevent mutagenic transformations. Individuals with DNA repair capacity (DRC) below the population mean may accumulate genetic alterations and may be at increased risk of developing different kinds of cancer. Polymorphism may result in subtle structural alterations of the repair enzymes and modulation of cancer susceptibility. Functional polymorphisms in DNA repair genes are thought to be responsible for inter-individual differences in repair capacities that exist within populations. In this study, seven polymorphisms in five DNA repair genes were studied.

X-ray cross complementing group 1 protein (XRCC1) is important for genetic stability and for embryonic viability and is involved in the repair of DNA single strand breaks and base damage from a variety of oxidants. XRCC1 acts as a scaffold protein or coordinator in single strand break repair in BER through its interaction with at least three other enzymes, poly-ADP-ribose polymerase (PARP), DNA ligase III and DNA polymerase B. Three coding polymorphisms that could alter the XRCC1 function have been identified, at codon 194 (Arg/Trp) within a hydrophobic core in exon 6, at codon 280 (Arg/His) in exon 9 and at codon 399 (Arg/Gln) in exon 10.The Arg300Gln variation which occurs within the PARP binding domain is believed to affect complex assembly and repair efficiency. This polymorphism is also located in the BRCT domain that is required for efficient single strand repair.

XPD gene (originally named excision repair cross complementation group 2 or ERCC2) is one of the seven genetic complementation groups encoding for proteins involved in the NER pathway. The XPD protein has a dual function, in NER and in basal transcription. It functions as an evolutionary conserved ATP-dependent helicase within the multisubunit transcription repair factor complex, TFIIH. The XPD gene is highly conserved in eukaryotes with homology to Rad3 and Rad15. Mutations in the XPD gene can completely prevent DNA opening and dual incision steps that lead to the repair of DNA adducts. The DNA repair function of XPD is critical in the repair of genetic damage from tobacco and other carcinogens .The amino acid substitution *Lys*751*Gln* in exon 23 does not reside in a known helicase/ATPase domain, but is at an amino acid residue identical in human, mouse, hamster and fish XPD, suggesting functional relevance for such a highly evolutionary conserved sequence.

The X-ray cross complementing group 3 protein (XRCC3) participate in the homologous recombination repair pathway, where is is a member of the Rad 51-related proteins. During recombination, XRCC3 is required for the accumulation of RAD 51 protein at sites of DNA damage and it is also involved in the resolution of Holliday junctions during the later stages of repair. The polymorphism XRCC3 Thr241Met is a nonconservative change which does not reside in the functional ATP-binding domains of the enzyme and the role of the XRCC3 gene in cancer is still uncertain. The variant 241Met allele has been associated with reduced repair of bulky DNA adducts

and predisposition to breast cancer, melanoma, bladder cancer and head and neck cancer. Thus this polymorphism may have biological implications for the functionality of the enzyme and the interaction with other proteins involved in the repair pathway.

During nucleotide excision repair, the damaged DNA is recognized by xeroderma pigmentosum group C complementing protein (XPC)-hHR23B heterodimer, which recruits transcription factor IIH, XPA, and replication protein A. Then, the helicases XPB and XPD, components of the transcription factor IIH complex, separate the damaged strand from the undamaged, and the nucleases XPG and XPF-excision repair cross complementing rodent repair deficiency, complementation group 1 (ERCC1) cleave at 3' and 5' to the damage, respectively, removing a 27- to 30-mer oligonucleotide containing the modified nucleotides. The release of the damaged oligonucleotide allows a new strand of DNA to be synthesized by the replication polymerases Pol δ and Pol ε, recovering an intact strand followed by ligation (64). Among the around 30 identified polypeptides that participate in the nucleotide excision repair process, the ERCC1 protein has a crucial role in the incision process by stabilizing the XPF endonuclease also called Excision repair cross complementing factor 4 (ERCC4) . A coding polymorphisms that could alter the ERCC4 function has been identified, exon 8, codon 415 Arg/Gln.

Besides its predominant role in nucleotide excision repair, ERCC1 is involved in other cellular processes. ERCC1 knockout mice are runted at birth and die before weaning due to liver failure, an unusual phenotype that is shared with XPF knockout mice, but not with any of the other nucleotide excision repair–deficient mice (67-69). In mammals, the ERCC1 protein also acts in a complex with XPF to remove nonhomologous 3V ends in the single-stranded annealing pathway of homologous recombination (70). ERCC1 acts in homologous recombination not only in the removal of nonhomologous tails but also in homology-dependent gene targeting events (71, 72). A synonymous polymorphism at codon 118 converting a common codon usage (AAC) to an infrequent one (AAT), both coding for asparagine, has recently been proposed to impair ERCC1 translation and to affect the response to cisplatin chemotherapy (65, 66).

The ERCC4/XPF-ERCC1 complex is essential for interstrand cross-link repair mediated by recombination repair mechanisms. Chinese hamster cell lines defective in *ERCC1* or *ERCC4/XPF* are not only hypersensitive to UV but are also extremely sensitive to DNA interstrand crosslink (73).