Identification of Single Nucleotide Polymorphisms in Two DNA Double-Strand Break Repair Genes – *XRCC4* and *RAD51*

by

Victoria Lea Kyle
B.Sc., Texas A&M University, 1993
B.Sc., Texas A&M University, 1995

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

We accept this thesis as conforming
to the required standard

[redacted]
Dr. B.W. Glickman, Supervisor (Department of Biology)

[redacted]
Dr. B.F. Koop, Departmental Member (Department of Biology)

[redacted]
Dr. F.Y.M. Choy, Departmental Member (Department of Biology)

[redacted]
Dr. M.E. Brackley, Additional Member (Department of Biology)

[redacted]
Dr. C.C. Helbing, External Examiner (Department of Biochemistry and Microbiology)

University of Victoria

Supervisor: Dr. Barry W. Glickman

Abstract

The most common variations in human gene sequences are single nucleotide polymorphisms (SNPs). SNPs may be responsible for susceptibility or sensitivity to various human diseases, including cancer. The goal of this study is to assess variability within double-strand break repair (DSBR) genes and the distribution of these genes in the healthy and cancer-affected populations. The hypothesis is that polymorphic variants of these repair genes with adverse functional consequences will be more closely associated with the cancer population in comparison to the control population. Using the heterozygote sequencing protocol, SNPs present in the sequences of two DNA double-strand break repair genes, *XRCC4* and *RAD51*, have been identified in cancer samples and healthy control samples. Two SNPs are found in *XRCC4*, while no variation is seen in *RAD51*. Possible associations of DSBR sequence variation to disease sensitivity can be assessed relative to the baseline of variation seen in these genes.

Examiners:

[redacted]

Dr. B.W. Glickman, Supervisor (Department of Biology)

[redacted]

Dr. B.F. Koop, Departmental Member (Department of Biology)

[redacted]

Dr. F.Y.M. Choy, Departmental Member (Department of Biology)

[redacted]

Dr. M.E. Brackley, Additional Member (Department of Biology)

[redacted]

Dr. C.C. Helbing, External Examiner (Department of Biochemistry and Microbiology)

Table of Contents

List of Abbreviations

AT, ataxia telangiectasia
ATM, ataxia telangiectasia mutated gene
ATR, ataxia telangiectasia and rad3-related protein
BER, base-excision repair
BRCA1, breast cancer susceptibility gene 1
Brca1, breast cancer susceptibility protein 1
BRCA2, breast cancer susceptibility gene 2
Brca2, breast cancer susceptibility protein 2
BRCT, *BRCA1* c-terminus
CBS, calf bovine serum
DEPC, diethyl pyrocarbonate
DNA-PK_{CS}, DNA dependent protein kinase catalyzing subunit
DNA-PK_{RS}, DNA dependent protein kinase, regulating subunit
DREJ, direct repeat end joining
DSB, double-strand break
DSBR, double-strand break repair
ERCC1, excision repair cross-complementing group 1 gene
ERCC2, excision repair cross-complementing group 2 gene
HRR, homologous recombination repair
HSP, heterozygote sequencing protocol
Ku, autoantigen, a protein that plays a role in DNA repair and in immunoglobulin gene recombination
MMR, mismatch repair
NBS, Nijmegen Breakage Syndrome
NBS1, Nijmegen Breakage Syndrome gene
NER, nucleotide-excision repair
NHEJ, non-homologous end joining
p53, protein 53
P53, protein 53 gene
PCNA, proliferating cell nuclear antigen
PCR, polymerase chain reaction
RPA, replication protein A gene
SSA, single strand annealing pathway
SNP, single nucleotide polymorphism
TE, tris-ethylenediaminetetraacetic acid
V(D)J, variable, diversity, joining gene segments
XPA, xeroderma pigmentosum group A gene
XRCC1, x-ray cross complementing group 1 gene
XRCC4, x-ray cross complementing group 4 gene
XRCC5, x-ray cross complementing group 5 gene
XRCC7, x-ray cross complementing group 7 gene

List of Tables

List of Figures

Acknowledgements

I would like to take this opportunity to recognize the various types of assistance I have received during my studies at University of Victoria. Members of the Glickman laboratory have offered their knowledge and expertise with the techniques used in my project. I wish to acknowledge David Walsh, Ken Sojonky, and Jana Kangas for beautiful sequencing, as well as assistance with other technical aspects of the project. I would like to thank Roderick Haesevoets for many hours of technical support in all areas from computers to broken pipetters, as well as helpful ideas to try for trouble shooting. Thanks to Stephen Dann for days worth of technical assistance and advice. I also wish to acknowledge Cindy Ruttan for technical help and introduction to the project. Recognition goes to Dr. Barry Ford for development of the project idea. I wish to thank Shulin Zhang for the various techniques he introduced me to at the beginning of my time here. I would like to acknowledge the administrative support of Pauline Tymchuk, Veronica Anthony, and Eleanore Floyd for all their help with my endless questions. Thanks go to Dr. Greg Stuart and Dr. Johan de Boer for scientific input on difficult points of the work and their willingness to pass on relevant findings they have made in the literature. I wish to acknowledge Dr. Moyra Brackley's help with everything from running statistics to proofreading multiple rough drafts. Additionally I wish to acknowledge the time and dedication contributed by my M.Sc. supervisory committee: Drs. Barry Glickman, Ben Koop, Francis Choy, and Moyra Brackley. Special thanks to my supervisor Dr. Barry Glickman for the opportunity to work on this project, as well as his time and sound advice during the course of my graduate education.

On a personal level, I would like to thank my family and friends back in Texas for their understanding of my time spent far from home for this opportunity. Thank you also to David Walsh, Jana Kangas, Ken Sojonky, Roderick Haesevoets, Lisa Sharp, Trina de Monye, and other coffee addicts for introducing me to their vice and for showing me how to appreciate the west coast lifestyle. Special thanks to Stephen Dann for excellent proofreading, listening, being there, and becoming my best friend. You will all be missed.

Introduction

Single nucleotide polymorphisms (SNPs) are the most common variation seen in the human genotype, occurring approximately once every 350 nucleotides (Cargill *et al.*, 1999). As of May 2000, the National Center for Biotechnology Information SNP database contained over 135,000 SNPs (NCBI SNP Data Base, 2000). There are three major advantages to studying SNPs. The first advantage is that detection and genotyping can be accomplished using a single method. Second, all types of sequence changes can be detected, including SNPs, insertions, deletions, and copy number variation. Finally, if genomic DNA is sequenced, variation in both coding regions and putative regulatory regions can be identified (Halushka *et al.*, 1999).

SNPs are responsible for susceptibility or sensitivity to various diseases. Well-documented examples of such diseases are sickle cell anemia and cystic fibrosis, both conditions caused by a single base substitution. SNPs have also been implicated in cancer susceptibility or sensitivity (NCBI SNP Data Base, 2000). Variants in cancer sensitivity genes increase individual risk of developing cancer without apparent environmental exposure. These genes have low frequency, but high penetrance. Variants in cancer susceptibility genes increase an individual's risk of developing cancer in the presence of environmental exposure, these genes have higher frequency, but lower penetrance (Preston, 1996).

Cancer cells are characterized by acquired genetic alterations consisting of mutations and chromosomal aberrations (Sheer and Squire, 1996). In the past few years, evidence of specific genomic rearrangements in human malignancies has been accumulating rapidly (Yunis, 1986). One source of rearrangement thought to contribute

to carcinogenesis is unrepaired DNA double-strand breaks (DSBs). DSBs can have exogenous and endogenous origins (Lieber, 1991), including ionizing radiation and attempts at DNA repair. Unresolved DSBs may block DNA replication and transcription, leaving exposed ends vulnerable to nuclease attack (Rufer and Morgan, 1992). Without an efficient repair system for these breaks, chromosomal fragmentation, translocation, deletion, and ultimately cell lethality may occur (Jeggo, 1998). Incorrectly repaired DSBs in somatic cells can lead to carcinogenesis, and when in the germ line inborn defects may occur (Yu *et al.*, 1999).

DSB repair occurs through either homologous or non-homologous recombinational repair pathways (Roth and Wilson, 1986; Pfeiffer *et al.,* 1994), with many repair genes playing various roles in DSB repair. *XRCC4* is part of the DNA end-joining pathway, a homologous recombination mechanism. The gene product forms a complex with DNA LIGASE IV to prepare strands for rejoining. *RAD51* is part of the homologous recombination pathway; the *RAD51* gene product's role is to search for areas of DNA sequence homology.

DNA repair pathways are among the most critical components that mediate individual response to exposure to environmental carcinogens (Hanawalt, 1998). It is possible that variation in repair genes due to SNPs may make some individuals more likely to develop cancer than other individuals. A single base change may result in a different amino acid sequence, which might substantially affect activity. Some genetic variants result in the complete absence of a gene or protein (Lai and Shields, 1999). Although DNA repair deficiency often arises from mutations that result in a loss of the DNA repair protein, DNA polymorphisms may alter the structure of the DNA repair

enzyme and contribute to cancer susceptibility (Lunn *et al.*, 1999). Therefore, defects in DNA repair likely represent a risk factor for many types of cancer (Yu *et al.*, 1997; Rajewsky *et al.*, 1998). To test this hypothesis, variation in these genes in a population should be determined.

Conventional analysis of variants involves preparation of cDNA, polymerase chain reaction (PCR) to amplify the sequence, and cloning in order to isolate individual alleles for sequencing (Edelmann *et al.*, 1997). This process is very labor, time, and cost prohibitive in a population study. The Heterozygote Sequencing Protocol (HSP), developed in the Glickman laboratory (Ford *et al.*, 2000) enables moderate-scale population surveys of SNPs. HSP uses fluourescently tagged primers and exploits the large dynamic range and low background of automated fluorescent sequencing. HSP may be used for any sequence that can be PCR amplified.

This study analyzes two DSB repair genes, *XRCC4* and *RAD51*, for SNP variants in both a healthy control sample and a small cancer patient sample. In the sections that follow, detailed background is given for DNA pathways, DSBR in particular, the genes studied, DSBR and cancer, and the results found in this study.

Chapter 1: Background

1.1 Environment, DNA damage, and DNA repair

Cancer, aging, and human genetic disease develop through a complex association involving hereditary and environmental factors that can include genomic instability induced by DNA damage. Determining the relative contribution to biological endpoints from endogenous versus environmental sources of DNA damage is an important goal of a great deal of scientific work. Such determinations are relevant to risk estimation based upon exposure to genotoxic agents in the environment, such as ionizing radiation and chemical carcinogens (Hanawalt, 1998).

Repair pathways must operate in all living cells to respond to endogenous threats to the integrity of DNA, as well as to invasive toxic agents present in the external environment. DNA damage does occur, but the cell produces enzymes that attempt to repair this damage. If these enzymes function efficiently, damage incurred to DNA will be minimized. Although DNA structure may be repaired, this repair can sometimes be incomplete or unsuccessful (Thacker, 1999). The chemical instability of the DNA molecule itself also threatens its persistence (Lindahl, 1993), and enzymes that act upon DNA during replication, transcription, and recombination are not error-free (Hanawalt, 1998). When DNA is not correctly repaired, the cell may die or may suffer alteration and loss of genetic information. Subsequent alterations may determine heritable genetic defects if they occur in germline cells, and in somatic cells are thought to be important in the development of radiation-induced cancer (Thacker, 1999).

The hypothesis that a tumorigenic state is produced by multiple genetic stresses is well accepted (Knudson, 1996). Proficient DNA repair mechanisms and cell cycle

regulators have been implicated as guarding a cell against becoming tumorigenic (Jasin, 2000). Inappropriate cell cycle regulation can lead to uncontrolled proliferation or the absence of cell death. Loss of DNA repair functions can lead to a mutator phenotype, causing a cell to become hypermutable (Cascalho *et al.*, 1998). Disruption of DNA repair pathways in hereditary cancer-prone disease strongly supports a role for hypermuation in tumorigenesis (Kim *et al.*, 1997). The prediction that various types of mutagenic base DNA damage is a significant risk factor for cancer has been substantially verified in human patients genetically defective in DNA repair pathways, including nucleotide excision repair (NER) and mismatch repair (MMR) (Lengauer *et al.*, 1998).

1.2 Overview of DNA repair systems

DNA repair pathways are a group of mechanisms involving cellular responses to restore genetic information in the event of damage. Numerous DNA repair genes have been characterized in prokaryotic and eukaryotic cells, many of which show a degree of homology through evolution. The repair pathways can be classified into six systems: direct repair, base excision repair, nucleotide excision repair, mis-match repair, double-strand break repair, and cell cycle regulation.

Direct repair is the simplest response to DNA damage. This involves the removal or reversal of the lesion in a single-step reaction, restoring the sequence to its original state (Yu *et al.*, 1999). This pathway is a light-dependent mechanism that primarily repairs damage from UV-radiation. Direct repair enzymes include O^6-

methylguaine DNA methyl transferase (Hazra *et al.,* 1997) and photolyases (van der Spek *et al.*, 1996).

Single-stranded DNA damage is commonly repaired by the base-excision repair pathway (BER), where a group of enzymes makes use of the undamaged strand as a template for repair (Thacker, 1999). Base damage is excised, the DNA backbone is cut, and the damaged site is rectified. The gap in the strand is then filled, and strand ligation returns the DNA to its original state (Sancar, 1996). BER pathways are normally quick and efficient. Because a considerable fraction of radiation damage and spontaneous damage is of this type, loss of the capacity to repair single-strand damage can be disastrous.

DNA LIGASE IV is a BER enzyme, that is also considered part of the double-strand break repair pathway. Few human disorders associated with BER enzymes have been found, explained by studies that have found that mice with BER enzymes knocked out are early embryonic lethal (Wilson and Thompson, 1997). An example of a BER enzyme is uracil DNA glycosylase, loss of which leads to Bloom's syndrome (Duncan and Weiss, 1982). Bloom's syndrome is a rare autosomal recessively transmitted disorder, the main feature of which is small body size. Bloom's syndrome cells are hypermutable, and excessive numbers of somatic mutations are responsible for many of the clinical features including cancer and chronic lung disease (German, 1995).

Another excision repair pathway is nucleotide-excision repair (NER). NER repairs DNA abnormalities caused by methylation, oxidation, reduction, or fragmentation of bases by ionizing radiation or oxidative damage (Yu *et al.*, 1999).

NER removes a large section of single-stranded DNA containing a site of damage, usually a bulky DNA adduct causing distortion of the double helix (Thacker, 1999).

Several human diseases are caused by defects in NER genes, such as xeroderma pigmentosum and Cockayne's syndrome (Hoeijmakers, 1993). Individuals with defective NER genes are usually sensitive to sunlight and chemical agents that cause bulky damage in DNA, and a few of these individuals show cross-sensitivity to ionizing radiation (Russell *et al.*, 1995).

A small fraction of damage induced by gamma rays is not repaired in cells from individuals with xeroderma pigmentosum, suggesting that ionizing radiation induces some bulky damage that cannot be removed by the base-excision pathway (Satoh *et al.*, 1993). NER enzymes include xeroderma pigmentosum group A (XPA), which is associated with the disease by the same name, and replication protein A (RPA) (Li *et al.*, 1995).

Mismatches in DNA base pairs occur through processes such as replication and the formation of heteroduplexes (Bishop *et al.*, 1985). Another source of mismatches is the result of deamination of 5-methylcytosine to uracil, which then escapes detection and removal by the BER enzyme uracil N-glycosylase, and results in a G:T mispair. Mismatch repair (MMR) deals with such errors. The pathway's essential strategy is similar to that of excision repair, as a patch of nucleotides is removed from one strand, followed by repair synthesis and ligation. There are two types of MMR, long-patch and short-patch, which are both observed in human cells. In humans, at least six MMR enzymes have been characterized, including hMSH2 (human mutS homolog 2) and hMSH3 (human mutS homolog 3) (Kolodner and Marsischky, 1999). Many individuals

who suffer from hereditary non-polyposis colon cancer (HNPCC) have MMR defects, and those with MMR defects frequently develop cancers of other organs (Friedberg *et al.,* 2000).

If damage affects both strands of a DNA molecule at the same site, a double-strand break (DSB) results. These breaks arise through the direct action of ionizing radiation or chemicals. This type of break is more difficult to repair than other DNA damage types and requires specialized enzymatic pathways. Recently several DSB repair pathways have been studied (reviewed in Chu, 1997; Sancar, 1995). In mammalian cells at least three pathways have been elucidated: homologous recombination repair (HRR), non-homologous end joining (NHEJ), which is DNA-PK dependent, and direct repeat end joining (DREJ).

DNA damage responses and cell cycle regulation are closely coupled. Cells that contain damaged genomic DNA arrest at G_1 to S and G_2 to M transition points of the cell cycle. This arrest gains time for repair and provides for the avoidance of repair during replication and cell division. p53 can be induced by DNA damage and may play a central role in arresting the cell cycle, inducing DNA repair, and apoptosis if damage is severe, or repair is not feasible (Canman *et al.*, 1994). The expression levels of some repair genes, such as *UNG* (the gene encoding uracil-DNA glycosylase) fluctuate during cell cycles under the control of cell cycle-dependent transcription factors, for example, E2F-1 (Mitra and Kaina, 1993; Dosanjh *et al.*, 1994).

Some DNA repair systems and the genes involved are only induced in the presence of DNA damage, while others are continuously expressed (Grombacher *et al.*,

1996). Each of the repair pathways is essential for genomic integrity and conservation of genetic information.

Chapter 2: Double-Strand Break Repair

Defects in DSBR can lead to DNA or chromosomal rearrangements commonly associated with carcinogenesis (Dasika *et al.,* 1999). Carcinogenesis is a complex process involving the accumulation of genetic changes resulting from endogenous and exogenous insults. Ionizing radiation is well known for both causing DSBs and its association with occurrence of cancer, as recently studied with *ATM* heterozygotes (Easton, 1994). The DSBR pathway is charged with rectifying damage caused by ionizing radiation to prevent carcinogenesis, and can be classified into sub-pathways of homologous recombination repair and non-homolgous repair. Two key enzymes involved in these mechanisms are XRCC4 and RAD51 respectively. Because of the important role played by these proteins in DSBR, their loss or mutation is certain to be associated with mutagenesis.

2.1 DSBR defect causes rearrangements that are commonly part of the cancer profile

DNA double-strand breaks (DSBs) are the most detrimental form of DNA damage because they lead to chromosomal breakage and rearrangement, and these events can result in apoptosis or tumorigenesis (Dasika *et al.*, 1999). Initiation of tumorigenicity by chemical or virus exposure may result from a subtle chromosome change that precipitates subsequent chromosomal instability. Progressive development of malignancy might then depend on random karyotypic reorganization and selection of new cell types. These new cell types may then lose growth control and gain other traits that facilitate growth in a given environment (Sager, 1979).

Because DNA repair processes are not perfectly effective, decay of the covalent structure of DNA provides a major contribution to spontaneous mutagenesis (Smith, 1992). The accumulated effect of replication and transcription of damaged DNA templates has been determined to be an important factor in carcinogenesis (Lutz, 1990).

There are also problems inherent in repair of closely opposed base damaged sites, since such attempted repair may lead to an error in base sequence. Sequence errors also may occur from rejoining incorrect ends. In this case the separation of the ends of a DSB followed by rejoining with the wrong end is possible, giving rise to deletion mutations and/or chromosomal aberrations. This misjoining is deleterious whether it occurs in an exon or an intron (Ward *et al.,* 1994).

Many different types of cancer result from chromosomal or DNA rearrangements that are associated with DSB. The majority of patients with chronic myeloid leukemia manifest a reciprocal translocation between chromosomes nine and twenty-two. Most Burkitt lymphoma cases exhibit a reciprocal translocation between chromosomes eight and fourteen, and these shufflings are situated at the sites of chromosomal breaks (Seemayer and Cavenee, 1989). In addition to these few examples, Yunis (1986) reviews thirty different cancer types and their associated chromosomal defects. Previous research has clearly established that tumorigenesis is associated with rearrangements. The question remains, however, as to why only some individuals are subject to chromosomal transformations, which then lead to tumor formation.

2.2 Cancer, radiation, and double-strand break repair

Carcinogenesis is a complex multistage process involving the accumulation of genetic changes in the genome. Therefore, cancer can be regarded and researched as a genetic disease. Most sporadic cancers occur due to carcinogen exposure mediated by inherited predisposition susceptibility to carcinogen, metabolism, and DNA repair. Substantial evidence now exists that implicates gene-environment interactions for lung, breast, prostate, and oral cavity cancers (Lai and Shields, 1999).

The model DNA DSB inducing agent is ionizing radiation. This agent is present in the environment mainly from radon gas decay, which accumulates in homes at different levels depending upon the uranium content of the underlying soil (Lyman, 1992; Lubin, 1988). Ionizing radiation is also used for diagnostic x-rays and for treatment of cancers (Chu, 1997).

DSBs have been implicated as being the mutagenic and carcinogenic effects of radiation (Yunis, 1990). Mutagenic effects of radiation result primarily from damage to cellular DNA. Although radiation can induce a variety of DNA lesions, including base damage, unrejoined DSBs are a common result of radiation's toxic effects. Large-scale changes in DNA structure, such as residual DSBs and incorrectly rejoined DSBs, are characteristic of most radiation-induced mutations (Little *et al.*, 1987). Residual unrejoined DSBs are lethal to the cell, whereas incorrectly rejoined breaks may produce important mutagenic lesions. In many cases, DNA misrepair leads to DNA deletions and rearrangements (Little, 1993).

Based on various studies (Boice, 1988; Shimizu *et al.,* 1990), including epidemiological investigations, several conclusions have been made concerning

radiation carcinogenesis in human populations. Radiation is a universal carcinogen in that it will induce cancer in most tissues at all ages. However, it is a relatively weak carcinogen and mutagen as compared to many chemicals, the estimated average relative risk increment is 0.9%/rad (Granath *et al.*, 1999). While cancer can be induced by radiation in many tissues, there is variability in the sensitivity of different tissues to its carcinogenic effects. Additionally, genetic predisposition to radiation induced cancer may be a confounding factor in the prediction of its effects in any given individual (Little, 1993).

Ionizing radiation causes a large number of chemically different types of DNA damage (Ward, 1988). The complexity of the damage will depend on ionization density at the point of interaction of the radiation track with DNA (Ward, 1994). The products of the ataxia telangiectasia mutated gene (*ATM*) and the Nijmegen breakage syndrome gene (*NBS1*) have been implicated upstream from the p53-mediated cellular DNA damage response to ionizing radiation (Artuso *et al.*, 1995; Jongmans *et al.*, 1997). Mutations in these genes confer sensitivity to ionizing radiation and predisposition to cancer. Additionally, several other genes, including *XRCC1* and *XRCC3* (X-ray repair cross complementing genes 1 and 3 respectively), have also been shown to be involved in the cellular DNA damage response to ionizing radiation (Zdzienicka, 1995).

Previous research has shown that cells derived from healthy individuals show variability in their sensitivity to radiation. This may be due to different levels of expression of genes that have an effect on cell radiosensitivity. Incidence of breast cancer is increased in individuals following exposure to ionizing radiation, suggesting that a radiosensitive subgroup can be found among the general population (Goss and

Sierra, 1998). This group might face hazards from routine levels of diagnostic and therapeutic radiation. Therefore, a better understanding of the possible relationship between genes that influence radiosensitivity and breast cancer development is clearly necessary (Jongmans and Hall, 1999). Individual responses to radiation exposure resulting from breast cancer screening and/or treatment protocols are most likely governed by variants in pre-disposing genes (susceptibility and sensitivity loci). Identification and understanding of such susceptibility genes may have positive effects in the clinical setting by enhancing ability to predict individual radiation responses or to customize therapy treatments to specific patients' needs (Bennett, 1999).

It is important to consider the possibility that genetically predisposed individuals may have increased sensitivity to environmental exposures such as radiation. Addressing the question of gene-environment interactions is particularly important in sub-populations who have inherited mutations in genes that play a role in DNA-damage response, cell-cycle control, or tumor suppression (Bennett, 1999).

Linkage analysis of families with a high risk of breast cancer has identified two important susceptibility genes, *BRCA1* and *BRCA2* (Ford and Easton, 1995). Germ-line mutations in the *P53* gene cause a high risk of pre-menopausal breast cancer as a part of the Li-Fraumeni syndrome (Sedlacek *et al.*, 1998). However, epidemiological studies show that these mutations are relatively rare in the general population and account for less than two percent of the total number of breast cancer cases (Ford and Easton, 1995). A greater proportion of breast cancer cases within the population may be accounted for by more common sensitivity genes that may have a relatively low penetrance with regard to breast cancer (Jongmans and Hall, 1999).

The *ATM* gene mutated in ataxia telangiectasia (AT) patients is a candidate for a breast cancer sensitivity gene. Individuals with the autosomal recessive disorder AT are homozygous or compound heterozygous for a mutation in the *ATM* gene (Savitsky *et al.*, 1995). The typical phenotype is caused by *ATM* null alleles that either truncate or severely destabilize the ATM protein (Gilad *et al.*, 1998).

A number of epidemiological studies have indicated that female *ATM*-heterozygous carriers have an excess risk of breast cancer, with a reported estimated relative risk of 3.9 times that of non-carrier frequencies (95% confidence interval 2.0-6.0) (Easton, 1994). However, other studies have found *ATM* heterozygous mutations are not linked to cancer occurrences (Nichols *et al.*, 1999; Bay *et al.*, 1998). Although results have not been consistent, several independent investigators confirmed this increased risk. Broeks, *et al.,* (2000) used the protein truncation test and found a high percentage of *ATM* germline mutations were identified among selected patients with breast cancer. One mutation, the exon eleven splice site mutation IVS10-6T→G, was found frequently in this study and data suggest that this mutation might be associated with an increased risk of breast cancer in *ATM* heterozygotes. Results show that *ATM* heterozygotes do have an increased risk of development of breast cancer (Broeks *et al.,* 2000).

It is known that AT is associated with high sensitivity to ionizing radiation and that fibroblast strains or cultures from *ATM* heterozygotes display moderate levels of radiation sensitivity (West *et al.*, 1995). All patients with breast cancer who were included in the Broeks *et al.*, (2000) study received low-dosage diagnostic radiation at a young age, as part of a screening program for tuberculosis. Moreover, all these patients

had received radiation treatment for their first breast tumor, and a non-significant trend toward a higher incidence of *ATM* mutations was noted in women with bilateral breast cancer. Thus, radiation might be an induction trigger for development of breast cancer in *ATM* heterozygotes. These data strongly suggest that the role of *ATM* heterozygosity in the pathogenesis of radiogenic cancers needs further investigation.

2.3 DSBR Pathways

2.3.1 Double-strand break repair

DNA double-strand breaks (DSBs) occur because of exogenous and endogenous insults. Exogenous causes include ionizing radiation and oxidative effects. Endogenous sources of DSBs include somatic recombination and overlapping of extensive excision repair tracts (Lieber, 1991). Breaks may arise as intermediates during processing of other DNA damage such as interstrand crosslinks or as intermediates of processes such as V(D)J recombination, the process of assembling the variable (v), diversity (d), and joining (j) gene segments of the immunoglobulin and T cell receptor variable region genes during development of B and T lymphocytes (Weaver, 1995). Other processes that may result in breaks in this way are meiosis and telomere maintenance that requires rearrangement of DNA sequences (Jeggo, 1998).

If DNA DSBs remain unresolved, replication and transcription of gene sequences is blocked. Additionally, exposed ends of gene fragments are prone to nuclease attack and destruction (Rufer and Morgan, 1992). Lack of or incorrect repair can lead to chromosomal fragmentation, translocations, deletions, and ultimately cell lethality. DSBs represent a significant lesion, potentially causing lethality, mutagenesis,

and carcinogenesis. In the soma incorrectly repaired DSBs can lead to carcinogenesis through inactivation of tumor suppressors, activation of oncogenes, and loss of heterozygosity (Jeggo, 1998). Two human diseases associated with DSB repair are Ataxia Telangiectasia (AT) and Nijmegen Breakage Syndrome (NBS). Both diseases are rare, autosomal recessive disorders. AT is characterized by progressive cerebellar ataxia beginning between one and three years of age, telangiectasias of the conjuctivae, choreoathetosis, and increased sensitivity to ionizing radiation followed by immunodeficiency and malignancy (in 38% of AT patients), particularly leukemia and lymphoma (85% of malignancies), in many patients (Gatti, 2000). NBS is characterized by short stature, progressive microcephaly with loss of cognitive skills, ovarian failure in females, recurrent sinopulmonary infections, and an increased risk of cancer (found in 35% of NBS patients 34 years of age or under), particularly lymphoma (88% of malignancies) (Concannon and Gatti, 2000). There are phenotypic similarities between the AT and NBS diseases on the cellular level suggesting that the corresponding gene products function in the same pathway. Both cell types fail to induce p53 at the G_1/S checkpoint and fail to suppress DNA synthesis in response to ionizing radiation (Stumm et al, 1997).

Efficient DSB repair is required for genomic integrity and proper gene expression (Yu *et al.,* 1999). This brief description of DSBR gene mutation symptomology illustrates the DSBR pathway's vital role in cellular physiology.

2.3.2 Overview of two types of DSBR

In the repair of double-strand DNA breaks, the processes fall into two main categories: homologous recombination repair (HRR) and non-homologous repair.

Simply stated, the repair mechanism occurs between parts that line up alike (HRR), or parts that do not (NHEJ and DREJ).

HRR requires extensive regions of DNA homology. This pathway accurately repairs DSBs using information from the undamaged sister chromatid or homologous chromosome. Non-homologous repair occurs by DNA end joining, or more specifically, non-homologous end-joining (NHEJ). NHEJ uses little or no sequence homology to rejoin juxtaposed ends, in a manner that is not necessarily error free. This pathway is also referred to as illegitimate recombination. NHEJ differs from HRR in that reactions are typically non-conservative, so that the joining process can result in a loss or gain in nucleotide base number (Roth and Wilson, 1986). The end-joining product may have either nucleotide insertions or deletions, depending on how the overlapping single-stranded ends are processed. Because the overlap intermediates would have only a few base pairings, alignment proteins are thought to stabilize these intermediates (Pfeiffer *et al.,* 1994).

In homologous recombination (see Figure 1) the signal which initiates the transduction pathway following exposure to DNA-damaging agents is the DNA double-strand break itself. All normal cells have low levels of p53, however following exposure to DNA damaging agents such as ionizing radiation a dramatic increase in p53 levels occurs (Kastan *et al.*, 1991). Both ATM and ATR proteins phosphorylate p53 upon DNA damage (Banin *et al.,* 1998). Once p53 is activated, the cell either undergoes apoptosis or growth arrest followed by repair. This outcome is determined in part by the current stage of cell cycle. Early in G_1, p53 triggers a check point that blocks further progression through the cell cycle so damage can be repaired prior to the S phase.

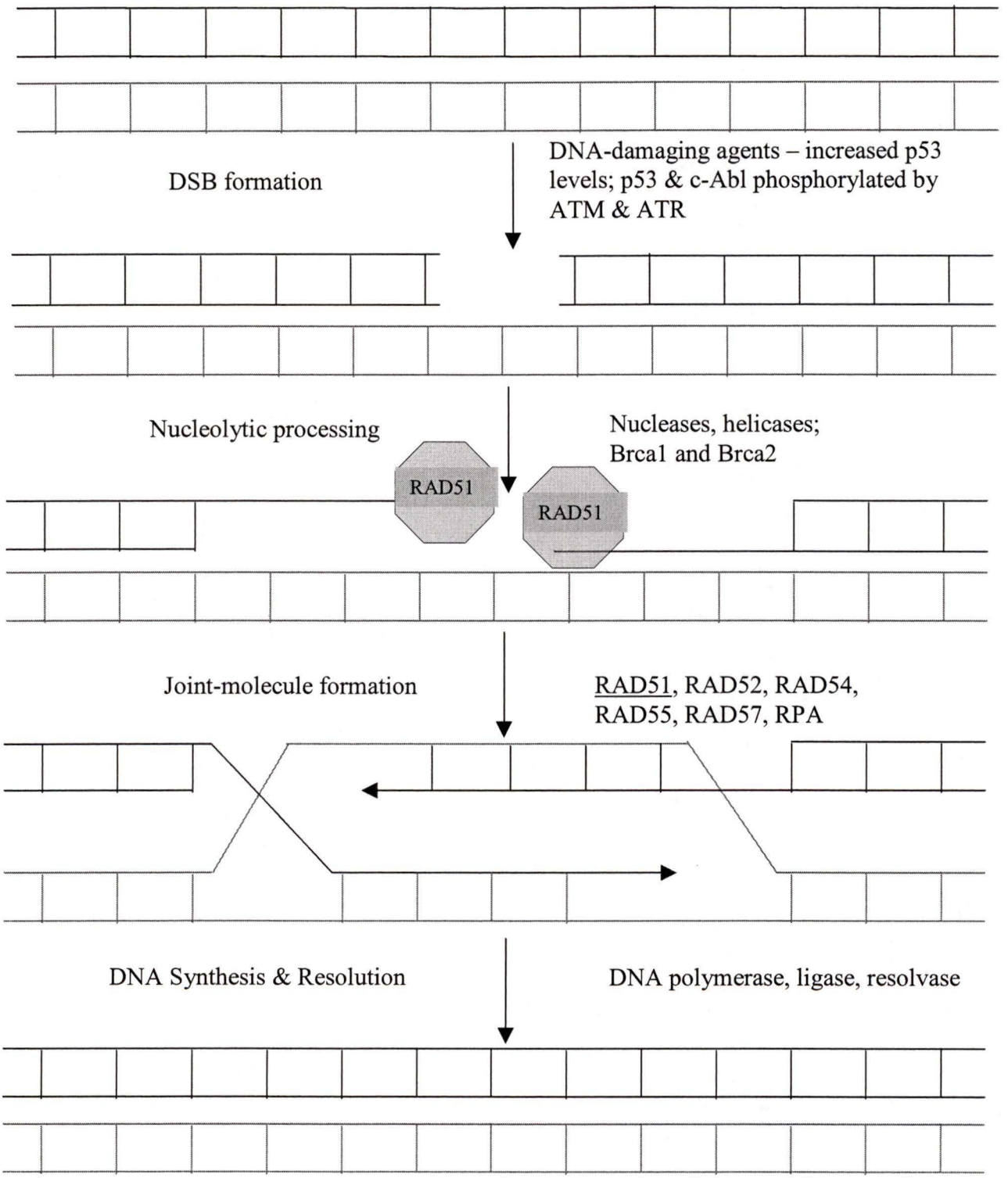

Figure 1: Model for DSBR through homologous recombination. Note the role of RAD51 in forming a nucleoprotein filament that searches for homologous duplex DNA.

However, if the cell is already committed to division, p53 triggers cell death. The C-terminal domain of p53 recognizes and binds to damaged, single stranded regions in DNA (Lewin, 1997). Cells that are in S phase at time of growth arrest only arrest briefly and continue to G_2 where they arrest again. p53 has been found to be required for arrest to occur in G_1 (Kuerbitz *et al.,* 1992). RAD51 has been found to co-localize with BRCA1 and BRCA2. Following DNA damage, the three proteins are relocated to structures which also contain PCNA (proliferating cell nuclear antigen) and may represent sites of active repair (Chen *et al.*, 1998). RAD51 then polymerizes onto the single-strand DNA to form a nucleoprotein filament that searches for homologous DNA. After homologous DNA is found, DNA strand exchange generates a joint molecule between homologous damaged and undamaged double-strand DNA. After the joint molecule is formed, DNA synthesis, utilizing DNA polymerase and DNA ligase, restores the missing information. Then resolution of the crossed DNA strands occurs by resolvase, producing two intact double-stranded DNA molecules.

NHEJ begins with the DNA-Protein Kinase regulating sub-unit (DNA-PK_{RS}) which is a Ku heterodimer encoded by *XRCC5* and *XRCC7*. Ku is a relatively abundant nuclear protein that is recognized by certain human autoimmune sera (Mimori and Hardin, 1986). DNA-PK_{RS} binds to the DNA ends and attracts DNA-Protein Kinase catalyzing sub-unit (DNA-PK_{CS}). The DNA ends are brought together by the DNA-PK_{CS} and its promoters phosphorylate each other producing a structural change in the complex and the removal of DNA-PK_{CS}. The complex consisting of RAD50, MRE11, NBS1 complex is attracted which processes the ends of DNA fragments (Jackson,

1997). Then the XRCC4-DNA LIGASE IV complex rejoins the ends as shown in Figure 2.

2.3.3 DSBR in yeast and mammals

Most of the current understanding of double-strand break repair originated through the study of the yeast, *Saccaromyces cerevisiae*. Organisms as distantly related as *S. cerevisiae* and *Homo sapiens* have similar DNA repair processes. Both HRR and NHEJ occur in yeast, mice, and humans with slight variations in pathways and genes between the organisms. In yeast, most double-strand breaks are repaired by HRR (Game, 1993). Since DNA end-joining is often preceded by processing that deletes or adds nucleotides, it is understandable that the relatively error-free HRR pathway would predominate in yeast, with its high proportion of coding DNA (Thompson and Schild, 1999). HRR in yeast can occur between sister chromatids or homologous chromosomes. In mammalian cells, recombination between homologous chromosomes in mitotic cells in much lower than in meiotic cells. This low rate of recombination in mitotic cells serves to prevent loss of heterozygosity, which can contribute to tumorigenesis or other instability (Kadyk and Hartwell, 1992). The similar DSBR pathways exhibited by such disparate eukaryotic organisms is evidence of DSBR's vital role in cell survival and conservative replication.

2.3.4 HRR in Yeast

HRR takes advantage of the sequence identity between certain regions of DNA, such as sister chromatids of replicated chromosomes or the maternal and paternal copies of the

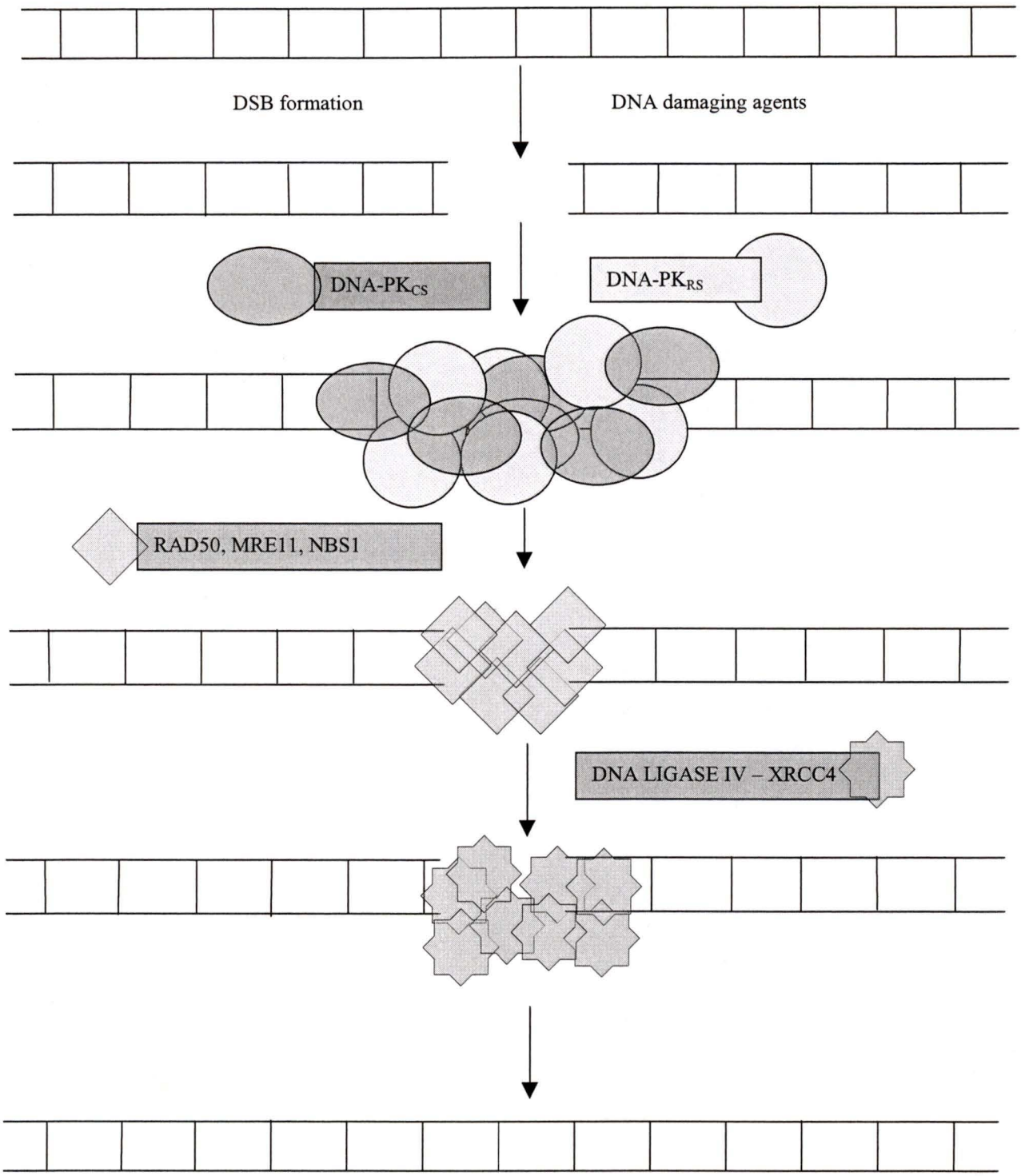

Figure 2: Model for DSBR through end-joining with DNA
Note the role of XRCC4 (complexed DNA LIGASE IV) to rejoin the strands.

same chromosome, from which sequence information is derived to repair the damaged copy. Identity of the DNA sequence is required over a considerable length, greater than 200 base pairs (Shinohara and Ogawa, 1995). The broken 3' end of the DNA strand invades an unbroken double-stranded homolog, and synthesis on this template reforms the damaged strand. Separation of the joint product of this reaction requires the activity of enzymes cutting and rejoining the newly synthesized DNA strands (Thacker, 1999). In yeast, HRR is mediated by members of the *RAD52* epistasis group of genes. Loss of *RAD52* gene function commonly leads to extreme sensitivity to ionizing radiation (Game, 1993). The gene products of this group are ScRAD51, ScRAD52, ScRAD54, and ScRAD57 (Zdzienicka, 1999). Yeast mutants defective in HRR are profoundly sensitive to ionizing radiation. Haploid yeast cells, which have a limited capacity to carry out HRR, are markedly more sensitive than are diploid cells (Sacki *et al.,* 1980).

Two sub-pathways of HRR have been found in yeast – gene conversion and single-strand annealing (SSA). Both gene conversion and SSA have the same first step of processing the DNA ends with 5' → 3' degradation of DNA strands that yield single stranded 3' ends. The removal of nonhomologous 3' ends requires the *RAD1* and *RAD10* gene products, but not other excision repair genes (Ivanov and Haber, 1995). A *rad1* mutant fails to repair a DSB by both mechanisms when its ends contain a small region of non-homology. Recombination proficiency of the *rad1* mutant is then restored when the ends of the DSB are made homologous to donor sequences. Therefore, the gene product of *RAD1* is suggested to be required to remove non-homologous sequences from the 3' ends of recombining DNA (Fishman-Lobell and Haber, 1992). RAD10 protein was shown to promote the renaturation of complementary DNA strands (Sung *et*

al.,1992) and RAD1 was reported to catalyze cleavage of synthetic holliday junctions (Habraken *et al.*, 1994).

Gene conversion occurs in yeast as the result of gap repair. According to current models of gene conversion, the 3' ends can invade the intact homologous template and act as primers of new DNA synthesis that will restore the damaged region (Szostak *et al.*,1983). *RAD52* mutation was shown to have more severe effects on gene conversion than on SSA.

SSA is truly a homologous process. It displays essentially the same homology requirements between interacting sequences as gene conversion (Sugawara and Haber, 1992). SSA differs from gene conversion in that it does not involve such molecular steps as strand invasion and the formation and resolution of holliday structures (Ivanov *et al,* 1996). Break repair is accompanied by a deletion of one of the flanking direct repeats along with the loss of intervening sequences (Fishman-Lobell *et al.*,1992). The current model of SSA implies that 5' → 3' degradation of the DSB ends followed by the annealing of exposed complementary sequences leaves behind the 3' end regions of non-homology that must then be removed to complete the formation of the deletion product (Ivanov and Haber, 1995). The outcome of this reaction is creation of a deletion of one repeat along with the intervening sequence (Lin, *et al.*, 1990). In *S. cerevisiae*, efficient double-strand break repair by SSA requires a minimum of 60-90 base pairs of perfect homology in the regions flanking the DSB (Sugawara and Haber, 1992). RAD1 and RAD10 are required as they are involved in the removal of the 3' end regions of non-homology that are created during annealing (Ivanov and Haber, 1995). Efficient SSA requires RAD52, but not the other members of the *RAD52* epistasis group (Sugawara

and Haber, 1992). The need for RAD52 can be overcome by increased homology between recombining sequences (Ozenberger and Roeder, 1991).

2.3.5 NHEJ in yeast

While the major DSB repair process in yeast is HRR, non-homologous repair does occur. Recent gene disruption studies have revealed that crucial components of NHEJ are conserved between yeast and mammalian cells. While the essential mechanism of NHEJ in yeast and mammals is the same, yeast homologs of the genes encoding DNA-PK_{CS} have not been found (Jeggo, 1998), however yeast does have proteins Tel1p and Mec1p which are related to DNA-PK_{cs} in sequence. The two proteins are homologs of ATM and ATR respectively (Jackson, 1996). Yeast homologs of Ku70 and Ku80 have been identified (Boulton and Jackson, 1996). The key enzymatic step of NHEJ is DNA ligation, for which yeast have two ATP-dependent ligases. Yeast LIGASE I, required for rejoining Okazaki fragments, which is homologous to mammlian DNA LIGASE I. The other is Lig4p, demonstrated to function epistatically with Ku in NHEJ, which is homologous to mammalian DNA LIGASE IV. Lig4p has also been shown to interact with ligase-interacting factor 1, a homolog of human XRCC4 (Teo and Jackson, 1997). Both the yeast and human version of DNA LIGASE IV function as a part of NHEJ.

NHEJ is strongly influenced by the stage of the cell cycle in which DNA breaks are induced. NHEJ G_1 repair occurs predominantly by deletion. Non-homologous repair of chromosomal DSBs is thirty-fold less efficient when DSBs occur in the G_1 stage of the cell cycle than at other times. It is suggested that in yeast, DSBs are

degraded by 5' → 3' exonucleases to create long stable single-stranded tails (Sugawara *et al.*,1995). NHEJ non-G_1 repair proceeds by end preservation. There is an increase in cell viability when DSBs are induced in non-G_1stages of the cell cycle; the increase possibly due to the requirement of replication or repair enzymes whose activity or synthesis may cause them to be available during non-G_1 stages. These proteins could include DNA polymerases or end-binding proteins analogous to Ku protein in mammals (Taccioli *et al.*, 1994). Another explanation for improved NHEJ in G_2 and S is the presence of sister chromatids that might play an important role in stabilizing the ends of the DSB and preventing the degradation that leads to deletions (Miyazaki and Orr-Weaver, 1994).

Work with yeast has provided links between chromatin structure and NHEJ. A yeast two-hybrid screen revealed that the yeast silencing protein SIR4p interacts with YKup (Tsukamoto *et al.*, 1997). There are several possible reasons why it may be beneficial to package DNA in the vicinity of a DSB into transcriptionally inactive or silenced heterochromatin. The condensed nature of this chromatin prevents the transcription or replication machinery from interfering with the DNA repair process. The formation of condensed chromatin at the site of a DSB might exclude non-specific nucleases, preventing loss of genetic information. DNA ends may be stopped from involvement in undesirable recombination reactions with other DNA molecules. Condensation of the DNA might help to juxtapose the two ends, facilitating their ligation and preventing the DSB from giving rise to a chromosomal break, which would be harder to handle (Tsukamoto *et al.*, 1997). SIR2p related proteins also exist in mice

and humans, it will be interesting to determine whether or not these potential mammalian homologs function in a similar way in DSB repair (Brachmann *et al.,* 1995).

2.3.6 HRR in mammals

Mammalian homologs of yeast genes operating in HRR have been identified and knock-out cell lines have been characterized. In mammalian cells, homologous recombination occurs at a much lower frequency than it does in yeast. However, although the efficiency of recombining ends of DNA molecules with chromosomal DNA occurs much more readily by NHEJ than by HRR, this bias does not accurately reflect the biological contribution of HRR in maintaining mammalian genomic integrity (Kirchgessner *et al.,* 1995). Evidence exists that mammalian cells rely on HRR to repair DNA damage and maintain stability. Human and mouse genes have been identified based on similarity to the yeast HRR genes. Also, genes such as *XRCC2* (X-ray repair cross complementing gene 2) have been found to complement radiation sensitive mutants whose phenotypes strongly point toward recombination defects. These two findings have merged as shown by the identification of a family of RAD51 proteins whose members are conserved across vertebrates (Kirchgessner *et al.,* 1995).

2.3.7 NHEJ in mammals

The majority of DSBs in mammals are rejoined by NHEJ. This is a potentially error-free process because it is not dependent on DNA repeats, however the degree of accuracy depends on the nature of the break and whether or not it has lost any DNA bases (Thacker, 1999). Four factors are involved in this process, three of them are

components are the DNA-dependent protein kinase, DNA-PK_{CS}, and the Ku70/Ku80 heterodimer which is a DNA binding subunit. DNA-PK is a nuclear protein serine/threonine kinase that is activated upon binding to a DNA DSB or other perturbation of the DNA double helix (Jackson, 1997). The fourth factor is XRCC4, a thirty-eight kDa nuclear phosphoprotein which complements the DSB repair defect in the CHO mutant, XR-1 (Li *et al.,* 1995). DNA-PK_{cs}, Ku70/Ku80, and XRCC4 are also all required for V(D)J recombination, indicating that the two cellular processes share common components.

2.3.8 DREJ in mammals

Another non-homologous repair pathway is direct repeat end joining (DREJ). This is an error prone process since a DNA deletion will always be associated with the repaired break. Even when DSBs have complementary ends, a fraction of the breaks are rejoined with a loss of sequence around the break sites. This mis-joining process is found to occur by a specific non-conservative mechanism which entails the deletion of DNA bases between short (two to six base pairs) direct repeat sequences, such that one of the repeats is also lost (Thacker *et al.,* 1992). This process can be more complex with an insertion of DNA at the deletion site. The proteins involved in the process have not yet been identified, but it has been determined that they are not components of DNA-PK (Mason *et al.,* 1996). Recent data with yeast also supports the existence of this pathway. When both the HRR and NHEJ pathways were knocked out, breaks were rejoined through the DREJ pathway (Boulton and Jackson, 1996).

2.4 *XRCC4 & RAD51*

2.4.1 *XRCC4*

The *XRCC4* gene has been identified by its ability to confer normal V(D)J recombination activity and partially restore the DSBR defect in XR-1 cells. The gene encodes a small protein of 334 amino acid residues with a molecular weight of thirty-eight kDa. The human and mouse *XRCC4* homologs are approximately seventy-five percent identical (Li *et al.,* 1995). Mammalian DNA ligase, DNA LIGASE IV, has distinct biochemical properties and substrate specificities from other ligases (Wei *et al.,* 1995). Through a combination of both immunological and biochemical approaches, DNA LIGASE IV has been shown to tightly and specifically associate with XRCC4, and the association is not mediated by a DNA intermediate.

Bacterially expressed XRCC4 and LIGASE IV bind to one another tightly, revealing that their interaction is direct, the proteins are only able to be resolved by the addition of harsh ionic detergents (Critchlow *et al.,* 1997). The unique carboxy-terminal domain of DNA LIGASE IV interacts with XRCC4. This region of LIGASE IV contains two tandem copies of the BRCT homology domain (Callebaut and Mornon, 1997). A detailed deletional analysis of DNA LIGASE IV determined that the region in the carboxy-terminal tail of DNA LIGASE IV is located between, rather than within its BRCT domains is necessary and sufficient to confer binding to XRCC4 (Grawunder *et al.,* 1998).

The effect of the absence of XRCC4 on DNA LIGASE IV levels by Western blot analysis found that DNA LIGASE IV levels were almost undetectable in XR-1 cells, but returned to wild type levels in XR-1 cells transfected with the *XRCC4* cDNA. The study

also found that this effect is not mediated through reduced expression of the *DNA LIGASE IV* gene since *DNA LIGASE IV* mRNA was present in XR-1 cells and the introduction of *XRCC4* had no effect on these levels. This strongly suggests that XRCC4 is required for the stabilization of the DNA LIGASE IV protein, a function that effectively regulates cellular levels of DNA LIGASE IV. The decrease in DNA LIGASE IV levels in XR-1 cells is most likely due to proteolytic degradation of the protein in the absence of its functional partner XRCC4 (Bryans *et al.,* 1999). *In vitro* studies show that the interaction of XRCC4 with DNA LIGASE IV results in a seven to eight-fold stimulation of DNA LIGASE IV activity (Grawunder *et al.,* 1997). This same protein-protein interaction resulting in stabilization has been described for XRCC1 and DNA LIGASE III (Caldecott *et al.,* 1995). Regulation of levels of repair proteins may be important to the cell's response to DNA damage.

Yeast homologs of *DNA LIGASE IV* and *XRCC4* have been identified. These homologs have been shown to mediate NHEJ in yeast and to be part of the same pathway as the Ku80 homolog indicating that the process of NHEJ and at least some of the components are evolutionarily conserved (Boulton and Jackson, 1996). The *DNA LIGASE IV* yeast homolog (*LIG4*) was found using multiple sequence alignment. The identification was based on clear conservation of the catalytic domain, the reactive site lysine, and the unique carboxy-terminal extension of human *DNA LIGASE IV* (Wei *et al.,* 1995). In human *DNA LIGASE IV*, the second ATG in the overall open reading frame has been assigned the start codon. However, there is meaningful alignment with *LIG4* starting with the first methionine, arguing that this larger open reading frame is used in some if not all mammalian cells (Wilson *et al.,* 1997). All ATP-dependent DNA

ligases form a covalent ligase-adenylate intermediate as part of the catalytic mechanism. Only human DNA LIGASE IV forms this complex intracellularly in a stable fashion, so that purified protein is adenylated completely with ATP *in vitro* only after removal of bound AMP. This is also seen in *LIG4* demonstrating a functional as well as structural sequence homology (Robins and Lindahl, 1996).

Recent experiments have shown that XRCC4 is associated with and can be phosphorylated by DNA-PK *in vitro* and may be part of the DNA end-binding complex mediated by the Ku70/Ku80 dimer. XRCC4 does not appear to bind to DNA itself (Leber *et al.,* 1998).

Mice deficient in *XRCC4* die during late embryonic development. *XRCC4* deficient embryos display extensive apoptotic death of newly generated, postmitotic neurons throughout the developing nervous system (Gao *et al.*, 1998). Additionally, progenitor lymphocyte development is arrested due to impaired V(D)J recombination. *XRCC4* deficient embryos are also growth retarded and their fibroblasts exhibit decreased proliferation and premature senescence in culture (Gao *et al.*, 1998).

2.4.2 *RAD51*

Human genes implicated in the HRR pathway have been cloned by their structural homology to the yeast *RAD52* group of genes (*RAD50*, *RAD51*, *RAD52*, *RAD54*, and *MRE11*). However, human homologs of some genes, such as *RAD55* and *RAD57*, have not been identified. Twice as many *RAD51* related genes have been identified in humans as in yeast. While HRR is not as common in humans as in yeast, it is a major repair form for DNA crosslinks, as there is no other mechanism for repairing

this type of damage (Thacker, 1999). The sequence of human RAD51 (HsRAD51) protein is highly conserved, 68% identical to yeast RAD51 (ScRAD51) (Shinohara *et al.,* 1993). There is low, but significant similarity between HsRAD51 and the RECA protein of *E. coli* (Yoshimura *et al.,* 1993). Both HsRAD51 and RECA act as central players in recombination by forming filaments on DNA and mediating strand transfer. Yeast *RAD51* is non-essential for growth, however its homologs in vertebrate cells are essential for viability in the context of cell proliferation. Disruption of both alleles of *RAD51* in mice proved to be early embryonic lethal (Tsuzuki *et al.,* 1996). This viability disparity might reflect a difference in the roles of RAD51 related proteins in uni- and multi-cellular organisms. Also, the possibility exists that a process of recombination is essential for the reproduction of mammalian cells. In addition, it is possible that the RAD51 protein is involved in a certain step of DNA replication (Tsuzuki *et al.,* 1996). High levels of mouse RAD51 found in testis, ovary, spleen, and thymus suggests that RAD51 protein might play important roles in meiosis and V(D)J type and/or class switch type recombination (Shinohara *et al.,* 1993).

Chapter 3: Materials and Methods

3.1 Introduction

In order to identify SNPs on a population scale, a novel technique has been developed, the Heterozygote Sequencing Protocol (HSP) (Ford *et al.*, 2000), which enables the characterization of both alleles without need for prior cloning (Edelmann *et al.*, 1997; Price *et al.*,1997). An overview of the method is presented in Figure 3. This technique is significantly more efficient and less costly than previous methods of identifying SNPs (Ucla *et al.*, 1987).

In this study using HSP, each base pair substitution observed was verified by the traditional cloning and sequencing methodology. SNPs observed at polymorphic frequencies, once verified by traditional cloning and sequencing, could then be recorded with confidence.

HSP capitalizes on the low natural background of infrared fluorescent dyes used with the Li-Cor automated DNA sequencers. The high signal to noise ratio allows for visualization of both alleles simultaneously on the sequencing gel. HSP can be used to cleanly sequence up to 1700 base pairs in one run. Increased sequencing runs result from the use of forward and reverse fluorescent labeled primers that emit at different wavelengths. Thus, data can be obtained for both strands simultaneously. This method can also be used to identify samples with insertions or deletions. Such samples are marked by an initial single sequence followed by the overlapping of two distinct sequences. The second allele sequence can either be read off the gel or obtained by cloning and resequencing.

Sample collection from healthy volunteers and colon / breast cancer patients

- Sampling protocol approved by the University of Victoria Committee on Human Subjects Research, British Columbia Cancer Agency and Capital Health Region
- Controls n=128, mean age 25 years (range 17-60); cancer patients n=67 (colon n=53, breast n=14), mean age colon cancer patients 68 (range 38-89), breast cancer patients 60 (range 39-82)
- 10 ml blood collected from each volunteer

↓

Lymphocyte Collection & Culture

- Isolation of lymphocytes
- Cells cultured and grown for 7-10 days
- Cells harvested and resuspended in Trizol for storage in -80°C.

↓

RNA Extraction

- Standard Trizol-chloroform RNA extraction
- DNAase treatment

↓

RT-PCR/Nested PCR Reaction

- First round PCR reaction combined with the reverse transcription reaction
- Second round (nested) PCR reaction performed using product of the first round as a template
- Second round PCR products purified
- Individual primers and annealing temperatures shown Table 4

↓

Sequencing & Data Analysis

- Products sequenced on Li-Cor machines using Li-Cor fluorescent labelled primers
- Digital image files compared to reference sequences obtained from Genbank for each gene

↓

Verification of Variants

- A selection of each type of variant observed was cloned into TOPO-TA 2.1 directly from PCR products
- For each variant, ten clones were sequenced and analyzed as above to confirm polymorphism

Figure 3: Method Overview

3.2 Population Sample

Control blood samples were obtained following a protocol approved by the University of Victoria Committee on Human Subjects Research. The individual samples analyzed here were drawn from a pool of samples collected from healthy anonymous volunteer participants, n=206. Samples analyzed were chosen randomly from the collection. Mean age of the sample was twenty-five, with an age range from seventeen to sixty years of age. Seventy-two percent of participants indicated that cancer had occurred in their families. Ninety-six percent of volunteers indicated ancestry from Europe (Britain, Western Europe, Eastern Europe; four percent were of Asian ancestry (Chinese, Japanese, East Indian, Korean).

Patient samples were obtained following a protocol approved by the University of Victoria Committee on Human Subjects Research, British Columbia Cancer Agency, and Capital Health Region. Individual samples analyzed were drawn from a pool of samples collected from 137 anonymous volunteer colon and breast cancer patients. Samples analyzed were selected randomly from the collection. Mean age of the colon cancer patients was sixty-eight, with an age range from thirty-eight to eighty-nine years of age. Mean age of the breast cancer patients was sixty, with an age range from thirty-nine to eighty-two years of age. Of the samples selected from that group for this study, seventeen of the forty-eight *XRCC4* samples were breast cancer patients, and thirty-one were colorectal cancer patients. Of the samples selected from the patient group for this study, fourteen of the sixty-seven *RAD51* samples were breast cancer patients, and fifty-three were colorectal cancer patients. Eighty-eight percent of participants screened for SNPs in *XRCC4* indicated that cancer had occurred in their families; eighty-six percent

of those screened for SNPs in *RAD51* indicated occurrence of familial cancer. All patients for whom SNPs were analyzed indicated ancestry from Europe (Britain, Western Europe, or Eastern Europe). See Table 1 for cancer patient sample information.

3.3 Lymphocyte Collection and Culture

Ten ml of blood was collected in heparinized Vacutainer tubes (Becton Dickinson) from each volunteer. Lymphocytes were isolated using Ficoll-Paque (Pharmacia Biotech) centrifugation following the package instructions. Cells were then transferred to growth media and incubated at 37°C for seven to ten days.

Growth media consisted of 40% HL-1 (BioWhittaker), 38% RPMI, 10% CBS, 1% Fungizone (all from GibcoBRL), 8% 4+ mixture containing 2535 U/ml penicillin, 2903 U/ml streptomycin sulfate, 7.3 mg/ml l-glutamine and 5.45 mg/ml pyruvic acid, 2% human serum (all from Sigma), 0.3 μg/ml PHA (Murex Biotech) and 5 U/ml IL-2 (GibcoBRL). When sufficient growth was observed, cells were harvested by centrifuging culture for 12 min. at 150 X g, resuspended in 3 ml of Trizol (GibcoBRL), aliquoted into three 1.5 ml cryovials (Sarstedt) and stored at –80 °C.

3.4 RT-PCR/Nested PCR Reaction

Trizol samples were thawed, then incubated at room temperature for 5 minutes before a standard phenol-chloroform (GibcoBRL) RNA isolation was performed. In isolating RNA of patient samples, 175 μg molecular biology grade glycogen (Sigma) was used as a carrier agent. RNA extracts in 20 μl DEPC H_2O were

Table 1: Cancer Patient Sample Data

Ancestry	100% European
Mean age (age range)	Breast cancer patients: 60 (39-82) Colon cancer patients: 68 (38-89)
% familial cancer	*XRCC4* samples: 88% *RAD51* samples: 86%
XRCC4 patient cancer type (n=48)	17 breast cancer patients 31 colorectal cancer patients
RAD51 patient cancer type (n=67)	14 breast cancer patients 53 colorectal cancer patients

stored at -80°C until needed. Gene specific primers (AlphaDNA) were designed for *XRCC4* and *RAD51*, see Table 2 for sequences and annealing temperatures.

The first round PCR reaction was combined with the reverse transcriptase (RT) reaction in single 200 µl tube. The 10x RT-PCR buffer contains 100mM Tris-HCl (pH 9.0), 500 mM KCl and 1.0% Triton X-100 (Sigma) in sterile, deionized water. The RT-PCR reaction mixture was 1xRT-PCR buffer, 0.25 mM each dNTP, 6.25 mM $MgCl_2$. 2.0 µl Forward Outside Primer, 2.0 µl Reverse Outside Primer, 13.5 U RNAguard (Amersham Pharmacia), 0.3 µl Taq DNA polymerase, 100 U M-MLV Rtase (Gibco BRL), 1 µl RNA extract, with DEPC H_2O to a final volume of 50 µl. The RT-PCR method was: 42°C x 60 min., 94°C x 4 min., [94° x 20 s., 55°C x 60s., 72°C x 60s.] x 30 cycles, 72°C x 10 min., 20°C x 15 min., 4°C. Nested PCR reaction conditions were 1x RT-PCR buffer, 0.25 mM each dNTP, 1.75 mM $MgCl_2$, 0.2 µl Taq DNA polymerase, 4.4 µl forward nested primer, 4.4 µl reverse nested primer, 9.0 µl round 1 RT-PCR product, and sterile deionized H_2O to a final volume of 110 µl. The nested PCR method was: 94°C x 4 min., [94° x 20 s., 60°C x 60s., 72°C x 60s.] x 30 cycles, 72°C x 10 min., 20°C x 15 min., 4°C. Nested PCR products were run on 1% agarose gel to demonstrate template purity. If a clean band was not observed for a given sample, a gel plug of the band of interest was used as a template in a third round of PCR, using the nested primers and PCR conditions. PCR products were purified using either the Wizard PCR Preps DNA purification system (Promega) or Quiaquick (Qiagen) PCR product purification kits. The DNA was eluted in sterile deionized H_2O and stored at 4°C before sequencing. Any long-term storage of PCR products was at -20°C.

Table 2: Primers, annealing temperatures, and product lengths for the RT-PCR/Nested PCR Reactions

Gene	Forward Primer	Reverse Primer	Annealing Temp (°C)	Product Length (# base pairs)
XRCC4				
Outside	gtggggctgcctcttta	ccagacaggatgttggacagc	55	1331
Nested	ctgaggtattaagaaatggagag	gtgccagtgtcatcatcaaatc	60	1170
Sequencing	ctgaggtattaagaaatggagag	gtgccagtgtcatcatcaaatc	50	1170
RAD51				
Outside	gcgagtagagaagtggagcgta	cttcataccctcctccaaaacca	55	1361
Nested	cgaccgagtaatggcaatgc	ggcagtcacaacaggaagagg	60	1091
Sequencing	cgaccgagtaatggcaatgc	ggcagtcacaacaggaagagg	50	1091

3.5 Sequence Analysis

Sequencing was performed using LiCor Long ReadIR 4200 instruments (LiCor Biotech). Primers used for sequencing were identical to those used in the nested PCR reaction. Primers were obtained from LiCor as 1.0 nmol lyophilized pellets that were reconstituted with 1ml sterile ice-cold 1 mM TE to a final concentration of 1.0 pmol/μl. The primer solution was protected from light to prevent loss of efficiency from degradation. The forward and reverse primers were 5' labeled with two distinct infrared dyes with fluorescence of 700 nm for the forward primer, and 800 nm for the reverse primer.

SequiTherm EXCEL II Long-Read DNA Sequencing Kits-LC (Epicentre Technologies) were used. Sequencing reactions were set up with four reactions for each sample. Each reaction contains both infrared dye labeled forward and reverse primers, all four deoxynucleotides, purified PCR template, DNA polymerase, and one of the four di-deoxynucleotides. Sequencing reactions contained a final volume of 9 μl. Primers are annealed to the cDNA by running the following PCR reaction: 95°C x 5 min., [95° x 30 s., 50°C x 15s., 70°C x 60s.] x 30 cycles, 20°C x 15 min., 4°C. 3 μl stop/loading buffer was added to each reaction.

The sequencing gel is prepared from 50% Long Ranger (FMC Bioproducts) and 1-1.5 μl of sequencing reaction are loaded into each well. HSP gels are grouped into four sets of sixteen lanes representing individual bases A, T, C, and G. This approach facilitates interpretation of the gels, as variants are immediately visible.

Digital image files (TIFF format) of all sequencing runs were archived and used for image analysis. Images were converted to JPEG format for image analysis with Paintshop Pro (Jasc Software) software. Sequence variants were compared to reference sequences obtained from Genbank accession numbers for each cDNA.

The HSP method allows the rapid and unambiguous identification of SNPs in cDNA directly from purified PCR products. Example gel images from HSP runs showing SNPs for *XRCC4* are shown in Figures 4 and 5.

3.6 Verification/Reconstruction Experiments

Conventional cloning and sequencing was utilized to confirm variant sequences. For each novel SNP observed, the PCR products from a random heterozygous individual and a random homozygous individual were chosen to be cloned using the TOPO-TA 2.1 (Invitrogen) cloning kit. Clones were selected on LB plates with ampicillin (50 μl/ml), and cloned DNA prepared for sequencing following standard methods. Ten clones were sequenced for each sample to be verified. The frequency of PCR-induced sequence artifacts was recorded from the cloned verification sequences. In each case the variant sequence was confirmed.

3.7 Statistical analysis

Individual genotypes for each variable base pair in each gene were entered into Excel '97 (Microsoft Office, 1997) spreadsheets. These data were imported into SAS (SAS Institute Inc., 1996) to run summary statistics and preliminary tests. GENEPOP (Raymond and Rousset, 1999; http://www.cefe.cnrs-mop.fr/) was utilized to calculate allele frequencies and assess exact *p* values for Hardy-Weinberg equilibrium tests

(Brackley, 2000). Tests for association between types of cancer and specific SNPs, as well as ancestry and SNPs were run using Instat (http://graphpad.com) and Fisher's Exact two-tailed *p*-value test.

Chapter 4: Results and Discussion

4.1 *XRCC4* Data

Heterozygote Sequencing Protocol (HSP) was used to analyze sequence variation in the double-strand break repair gene *XRCC4* in 128 healthy control samples (n=128) and forty-eight cancer patient samples (n=48). While working with the control samples was very successful, it was not possible to obtain adequate PCR products for the desired number of patient samples. Of the pool of 137 patient samples, only the forty-eight reported here produced adequate products. Although multiple attempts of both RNA extraction and RT-PCR were tried with the remaining eighty-nine samples, it was unsuccessful. Possible explanations for the difference in quality of the patient and control RNA are the disease state of cancer patients, effects of radiation and chemotherapy treatments on the individuals' cells, and the more advanced age of patients when compared to controls.

Variants are identified as base pairs counting from the first A of the first codon in the coding region. The Genbank sequence is defined as the reference. Deviation from this sequence is defined as a variant, regardless of the allele frequency. Of the control samples, 127 carried 148 variants from the Genbank reference sequence at two sites. For the cancer patient sample population, forty-four samples carried forty-six variants at two sites.

Variants from the Genbank reference sequence, not previously reported, were found in *XRCC4* in two positions: 401 and 921. Both variants were found in control and patient samples. Examples of the HSP gels showing these polymorphisms are shown in Figures 4 and 5 respectively. Results for *XRCC4* are summarized in Table 3.

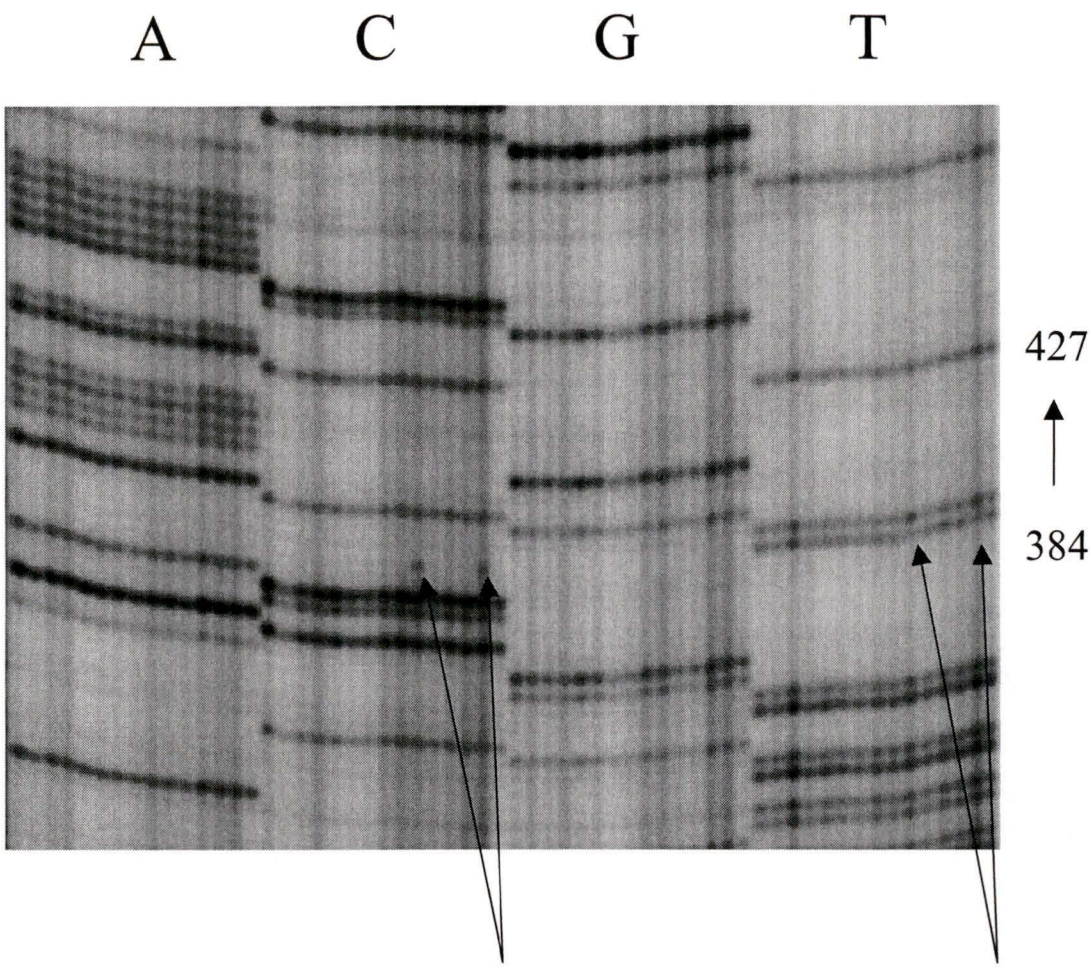

Figure 4: Section of LiCor Sequencing Gel of *XRCC4* (positions #384-427)

***XRCC4* sequence variant, a T → C substitution at position 401.** This gel represents 15 individual samples. The HSP samples were loaded grouped by base such that the first 15 lanes show A bands, the second set of 15 are C bands, the third set are G bands, and the final set of 15 lanes are the T bands. An individual sample's sequence can be determined by looking at the respective lane in each A, C, G, and T section. The arrows located in lanes 10 and 14 of the C and T sections of the gel show two samples that are heterozygous for the T → C substitution at position 401. The other 13 samples shown display the Genbank reference sequence for *XRCC4* at this position. The heterozygotes are indicated with fainter band strength in the respective lanes for both the C and T sections. A homozygous substitution, while not shown here, is seen when the band is absent from the expected base section and present in another section for a respective lane.

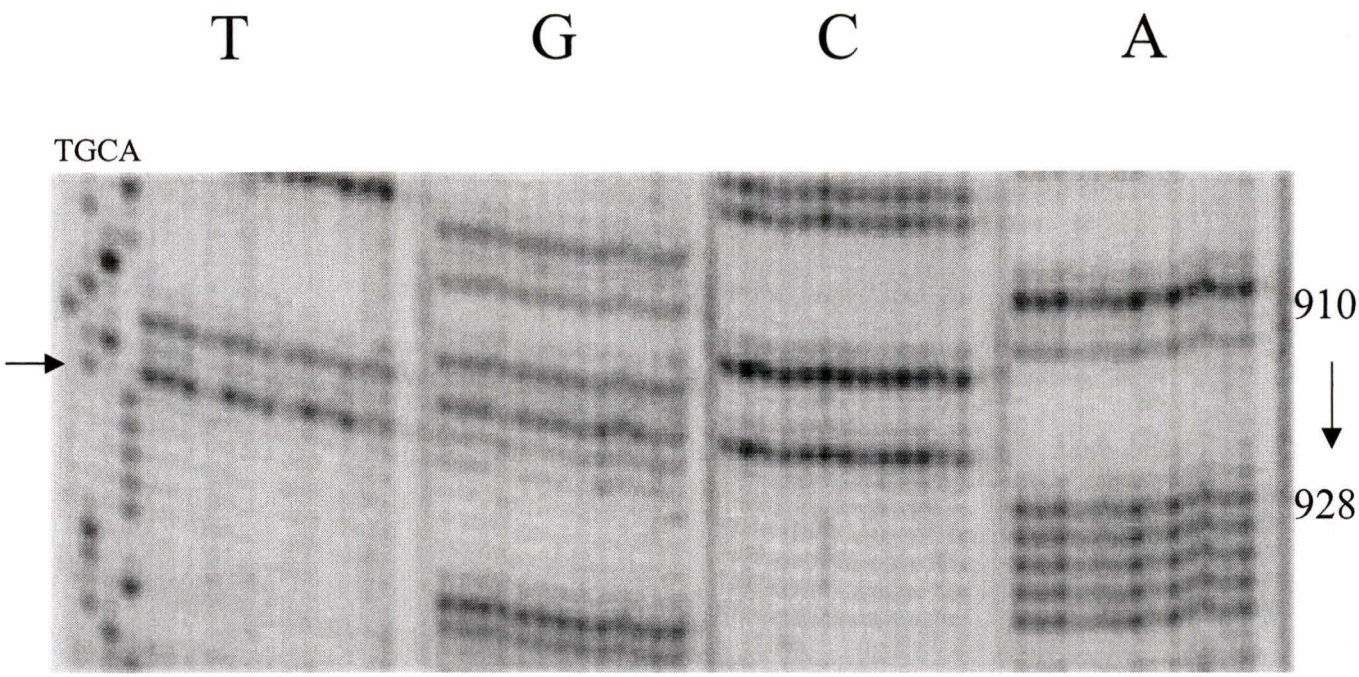

Figure 5: Section of LiCor Sequencing Gel of *XRCC4* (positions #910-928)

***XRCC4* sequence variant, a G → T substitution at position 921.** This gel represents 16 individual samples. The first 4 lanes of the gel represent one sample with bands T, G, C, A. The remainder of the gel was loaded such that the next 15 lanes show A bands, the following set of 15 C bands, the third set G bands, and the final set of 15 lanes the T bands. An individual sample's sequence can be determined by looking at the respective lane in each A,C,G, and T section. The arrow indicates position 921 where the SNP is found. Note how several samples in the G section are faint and are also faint in the T section, these are the samples that are heterozygous for the substitution. Also, those samples that do not show any banding in the G section, do show banding in the T section, these are the homozygous substitutions.

Table 3: Coding sequence variants by gene and base pair position for *XRCC4.*

Position	Change	Observed	n (H)	n(Q)	Amino acid change	Codon #	Exon #
			Controls n=128				
Reference sequence		1					
401	T → C	12	11	1	ATT (ile) → ACT (thr)	134	IV
921	G → T	126	20	106	TCG (ser) → TCT(ser)	307	VIII
127 samples carried 148 variants at 2 sites							
			Cancer patients n=48				
Reference sequence		4					
401	T → C	2	0	2	ATT (ile) → ACT (thr)	134	IV
921	G → T	44	13	31	TCG (ser) → TCT(ser)	307	VIII
44 samples carried 46 variants at 2 sites							

Sequence changes are in comparison to the reference sequence, defined as the sequence obtained from the accession numbers for each cDNA sequence as noted in the text, numbered from the A in the initiation codon.
H: heterozygotes, Q: substitution homozygotes

The polymorphism at position 401 is a T → C transition. This mutation causes a non-conservative amino acid change of isoleucine to threonine. Of the 128 control samples, one homozygous substitution and eleven heterozygous substitutions of this type were observed. Of the forty-eight patient samples, two homozygous substitutions and no heterozygous variants of the position 401 T → C substitutions were found. The polymorphism at position 921 is a G → T transversion. This is a silent mutation, meaning that the variation does not result in an amino acid change. Of the 128 control samples, 106 homozygous substitutions and twenty heterozygous substitutions were observed. Interestingly, only two of the 128 control samples exhibited the Genbank reference sequence for position 921. Of the forty-eight patient samples, thirty-one homozygous substitutions and thirteen heterozygous substitutions of this type were found.

4.2 Discussion of two SNPs identified in *XRCC4*

Allele frequencies and Hardy-Weinberg calculations for the SNPs found in *XRCC4* are shown in Table 4. The two variants found in *XRCC4* are in Hardy-Weinberg Equilibrium in the control group, while only the SNP at position 921 is in equilibrium in the patient group. Equilibrium will exist in the absence of mutation, gene flow, selection, and drift, as predicted by evolutionary theory. It is a reasonable assumption that allelic association with cancer should constitute a powerful selective force and therefore alleles that are in equilibrium do not represent likely associations to cancer risk. However, since cancer occurs predominantly as a post-reproductive event, the

Table 4: Allele frequency and Hardy-Weinberg calculations for each sequence variant in *XRCC4*.

Position	Sample	Allele Frequency		HWE exact *p*
		reference	substitution	
401	Healthy Controls	0.95	0.05	0.27
	Cancer Patients	0.96	0.04	0.0003
921	Healthy Controls	0.09	0.91	0.30
	Cancer Patients	0.22	0.78	0.19

alleles may still be associated with disease sensitivity even though they are in equilibrium (Ford *et al.,* 2000).

The frequency of silent substitution at position 921 was found to be at a higher rate than the non-synonymous change at position 401. This may be because given two mutations, the more overt change in coding sequence should be preferentially eliminated. See Table 5 for the frequencies of SNPs to base pairs screened.

Analysis of mouse and human gene sequences of *XRCC4* (Genbank) indicated that SNPs observed in the present study occurred in locations conserved during evolution. This observation suggests that the amino acid residues coded for by these regions are important in maintaining normal protein structure and integrity. It is also suggestive that the resultant codon substitutions could produce a protein with altered function in either repair capacity or fidelity.

Allele frequencies of the two substitutions across both study groups ranged from 0.04 to 0.22 when examining the least common base observed in each position. These frequencies are consistent with those found in a study by Shen *et al.*, (1998) examining polymorphisms in the DNA repair genes *XRCC1*, *XPD*, *XRCC3*, and *XPF* with allele frequencies that ranged from 0.04 to 0.42.

Associations between SNPs and type of cancer patients experienced, as well as between SNPs and ancestry of controls were investigated using Instat (http://graphpad.com) and Fisher's Exact two-tailed *p*-value test. Tables with association *p*-values and Odds Ratios (95% confidence interval) are in the Appendix. No association was found between SNPs occuring in *XRCC4* and type of cancer that patients experienced.

Table 5: Frequencies of SNPs (from the most common sequence observed) to base pairs screened for *XRCC4*

	# Base Pairs Screened	SNPs from Most common sequence observed	Frequency (SNP/base pair)
Conservative	205,920	14	6.79×10^{-4}
Silent	205,920	39	1.89×10^{-3}
Total	205,920	53	2.57×10^{-3}

SNPs have been found in genes such as *P53* to cluster in certain ethnic populations (Sjalander *et al.*, 1996). However, the present study includes only a small Asian population in the control sample (n=5) and none in patient samples. Associations were tested, but none was observed between the Asian sub-sample population and the SNPs in the control samples.

4.3 Implications of position 401 non-conservative substitution

The T → C transition at position 401 of the coding region of *XRCC4* results in an amino acid change of isoleucine to threonine, a non-conservative mutation. Isoleucine is a hydrophobic amino acid, while threonine is polar. This type of change can result in differences in protein folding and interaction. Also, threonine is capable of being phosphorylated by kinases. This threonine phosphorylation can result in anomalous protein function as well as changes in interaction of XRCC4 with other DNA repair proteins.

Several studies have recently shown the phenotypic significance of a SNP that results in an amino acid change in the coding region. In *XRCC1* three coding polymorphisms resulting in non-conservative amino acid changes were found (Shen *et al.*, 1998). The resulting arginine → glycine change at codon 399 was shown to be significantly associated with higher levels of aflatoxin B_1-DNA adducts and somatic gycophorin A mutations (Lunn *et al.*, 1999). *XRCC1* shows functional similarity to *XRCC4*, in that both gene products bind to DNA ligases in the rejoining of DNA strand breaks. In another study, mean DNA adduct levels among individuals with two copies

of the *XRCC1* codon 399 glycine allele were higher when compared with individuals heterozygous for the SNP (Duell *et al.*, 2000).

The action of natural selection during evolution is evident from the comparatively lower rate of non-synonymous coding region SNPs (cSNPs), especially those that create non-conservative changes (Cargill *et al.,* 1999). Non-synonymous cSNPs not only occur less often, but also have lower minor allele frequencies for this class.

In *XRCC4* at position 401, control samples are in Hardy-Weinberg Equilibrium, while cancer patient samples are not. This result is suggestive of a selective event. It should be noted however, that the result reflects an expectation of four heterozygotes in the cancer sample, which displayed none.

4.4 Implications of position 921 silent substitution

The G → T substitution at position 921 of *XRCC4* displayed an interesting frequency in comparison to the Genbank reference sequence. Only two individuals in the control group and four patients in the cancer patient group exhibited the sequence listed in Genbank for that position of the cDNA. In comparison to the population studied here, the Genbank sequence is a variant and the change is actually a T → G transversion, instead of vice versa.

If the least common allele is taken to be the variant sequence, then in the control group 106 individuals displayed the non-variant sequence, twenty were heterozygous for the position 921 change, and two were homozygous for the substitution. In the cancer patient group, thirty-one exhibited the non-variant sequence, thirteen were heterozygous,

and four were homozygous. The data for the G → T transversion at position 921 adjusted for the least frequent allele is shown in Tables 8 and 9.

In *XRCC4* at position 921, there is a significant difference between controls and cancer patients for both genotype (p=0.005) and alleles (p=0.005) (GENEPOP genic differentiation and genotypic differentiation, Fisher's Exact two-tailed test). While silent mutations are neutral with regard to the protein, they can affect gene expression via sequence change in the RNA. For example, a change in secondary structure of the RNA can influence transcription, processing, or translation. Another possibility is that a change in synonym codons calls for a different tRNA to respond, influencing efficiency of translation (Lewin, 1997).

Recent studies (Yu *et al.*, 1997; Yu *et al.*, 2000) of another DNA repair gene, *ERCC1*, indicate that a base substitution that does not result in a change of amino acid may still affect DNA repair. A polymorphism has been found in the coding region of the NER gene *ERCC1* that results in a silent substitution. Although the same amino acid is coded, the polymorphism is associated with a fifty percent reduction in codon usage for that amino acid. This reduction could affect the rate of transcription and/or translation (Yu *et al.*, 1997). Further study found that this silent substitution polymorphism is associated with a reduced DNA repair capacity in human ovarian cancer cells (Yu *et al.*, 2000). Similar to *XRCC4*, the presence of functional *ERCC1* is essential for life (McWhir *et al.*, 1993).

XRCC4 is thought to serve as a molecular bridge to link DNA LIGASE IV with DNA DSBs, perhaps through additional interactions between XRCC4 and other components of the DNA NHEJ system. Immunoprecipitation studies indicate that

XRCC4 can interact with DNA-PK, although the interaction is weak in comparison to that with DNA LIGASE IV (Critchlow *et al.*, 1997). In the initial analysis of XRCC4, a potential DNA-PK phosphorylation motif was noted (Li *et al.*, 1995). Further biochemical studies have confirmed that XRCC4 is phosphorylated when incubated *in vitro* with purified DNA-PK and linearized plasmid DNA. DNA-PK dependent and independent serine/threonine phosporylation of XRCC4 *in vivo* appears to occur between amino acid residues 204 and 334 (Leber *et al.*, 1998).

The substitution at cDNA position 921 in *XRCC4* results in a silent substitution in codon 307 coding for serine. The references above give rationale as to why this silent SNP could cause the type of DNA repair deficiency that results in cancer. Since a serine residue in that region is phosphorylated by DNA-PK, a reduced rate of transcription or translation of that residue because of codon efficiency may result in reduced double-strand break repair in the NHEJ pathway.

Mouse and human XRCC4 proteins are highly conserved, with 75% sequence identity and 84% similarity (Leber *et al.,* 1998). The SNP at position 921 occurs in a region of sequence identity shared between the mouse and human sequences (Genbank). Thus, the serine residue coded for by codon 307 potentially could be the one involved with the interaction between the DNA-PK and XRCC4 proteins.

4.5 Summary of *XRCC4* SNPs

If amino acid differences, especially at conserved sites, in these enzymes impair their function in the repair of DNA damage, these polymorphisms theoretically could predispose an individual to environmentally induced cancer. While studies of actual

protein function and related DNA repair phenotype are needed to confirm that these amino acid differences do indeed result in changes in repair proficiency, several lines of evidence presented herein support this biologically plausible concept.

4.6 No variation observed in *RAD51*

No variation was observed for *RAD51* in the 127 control samples (n=127) and the sixty-seven patient samples (n=67) screened. While working with the control samples was very successful, it was not possible to obtain adequate PCR products for the desired number of patient samples. Of the pool of 137 patient samples, only the sixty-seven reported here produced adequate products. Although multiple attempts of both RNA extraction and RT-PCR were tried with the remaining seventy samples, it was unsuccessful. Possible explanations for the difference in quality of the patient and control RNA are the disease state of cancer patients, effects of radiation and chemotherapy treatments on the individuals' cells, and the more advanced age of patients when compared to controls.

RAD51 is highly conserved in both structure and function through evolution (Baumann and West, 1998). The *Escherichia coli* protein RecA, *Saccharomyces cerevisiae* RAD51, and *Homo sapiens* RAD51 show a high degree of structural and functional homology (Shinohara *et al.*, 1993). Disruption of both alleles of *RAD51* in mice proved to be early embryonic lethal (Tsuzuki *et al.,* 1996).

As described regarding *XRCC4*, a base substitution that changes the codon, but does not result in a change in amino acid sequence, still can affect gene expression. The

fact that not even a silent mutation was observed in *RAD51* is suggestive that the resulting gene product is essential for DNA repair, and ultimately life.

Chapter 5: Summary and Conclusions

5.1 Results

Two SNPs were observed in *XRCC4*, while no variation was seen in *RAD51*. The transition at cDNA position 401 in *XRCC4* resulted in a non-conservative amino acid mutation. This polymorphism is in Hardy-Weinberg Equilibrium in the control group, while the cancer samples are out of equilibrium. Although this difference in equilibrium is indicative of a selective event, the sample size should continue to be increased in order to reach a firmer conclusion.

The transversion at cDNA position 921 in *XRCC4* resulted in no change in the corresponding amino acid. However, studies have shown that a silent substitution may still cause a change in gene expression due to differential transcription and translation of the amino acid codon affected. This polymorphism is in Hardy-Weinberg Equilibrium in both the control and cancer patient samples, but there is a significant difference between the samples for both genotypic frequency and allelic frequency.

RAD51 is highly conserved across kingdoms and is embryonic lethal in mice. The findings of this study of *RAD51* showing no variation in a screening of 176 individuals provides additional evidence for the gene's crucial role in DSBR in humans.

5.2 Environment, DNA damage, and DNA repair

Environmental insults such as chemical carcinogens and ionizing radiation cause various types of DNA damage. Additionally, the unstable structure of DNA itself allows breaks and other DNA injury to occur as a result of endogenous origins (Lutz,

1990). Mutagenesis by chemical or physical agents requires the persistence of structural lesions in genomic DNA for a period of time sufficient to give rise to heritable changes in nucleotide sequence (Rajewsky *et al.*, 1998). Specialized pathways have evolved to correct lesions in genetic information sequence to prevent mutagenesis and carcinogenesis. While DNA repair systems are efficient and effective, errors do occur. A defect in a mechanism will increase error rates and lead to a higher chance of a resultant problem. SNPs in gene sequences of involved DNA repair proteins are a source of inefficient participants to the pathway.

5.3 Cancer, radiation, and DSBR

Carcinogenesis is a complex multistage process involving an accumulation of information changes in the genome. The model DNA DSB inducing agent is ionizing radiation, which is present in the environment, as well as used as a treatment method for cancers. DSBs have been implicated as being the mutagenic and carcinogenic effects of radiation (Yunis, 1990). Genetic predisposition to radiation induced cancer may be a confounding factor in the prediction of its effects in any given individual (Little, 1993). Genetically predisposed populations may have increased sensitivity to radiation. Addressing the question of gene-environment interactions is particularly important in sub-populations who have inherited mutations in genes that play a role in DNA-damage response (Bennett, 1999). Therefore, it will be necessary for future studies to identify those mutations that may affect sensitivity of individuals and become useful for screening those susceptible.

5.4 Future work to address unknowns

Similar studies are being conducted in our laboratory on genes in other repair pathways. See Table 10 for a list of genes and the respective pathways being studied.

This study includes a relatively small number of subjects in the various subgroups such as cancer type and ancestry and, therefore, needs to be duplicated. Further studies with a larger sample size are needed to clarify these complex interactions between genotype, exposure, and demographic characteristics.

As a preliminary test of functionality, the association between of each of the *XRCC4* polymorphisms and proficiency for repair of ionizing radiation induced double-strand breaks should be examined. A similar study was conducted on *XPD* polymorphisms associated with X-ray induced breaks (Lunn *et al.*, 2000). Once a subject's polymorphism status for a given site is established, DNA repair efficiency can be assessed using an assay that measures unrepaired DNA in cytogenetic preparations of lymphocytes isolated from blood. The lymphocytes would be exposed to ionizing radiation and then incubated to allow for repair. The existence of residual or misrepaired breaks can then be assessed.

It should be determined if the mutation at the serine residue at codon 307 truly does disrupt the phosphorylation rate of XRCC4 by DNA-PK. This can be assessed by performing a phosphorylation assay on the *XRCC4* protein product with and without the transversion at position 921 present. Clontech's Phosphoamino Acid Monoclonal Antibodies assay run on *XRCC4* and *DNA-PK* protein products would quantify the ability of DNA-PK to phosphorylate XRCC4 with and without the SNP.

It would also be interesting to perform phosphorylation assays on the isoleucine to threonine mutation at codon 134 versus the reference sequence, as an increase in phosphorylation can affect protein function and interaction. If an increase in phosphorylation is found, then assays on DSBR efficiency should follow.

The collection of an older cancer-free control sample set will be beneficial to additional study. The control population sampled for the current study is significantly younger than the cancer patient population. While the control subjects are cancer free, it may be said that they are cancer free for now. Inevitably a normal percentage of those individuals will contract cancer later in life.

Literature Cited

Artuso, M., Esteve, A., Bresil, H., Vuillaume, M., and Hall, J. (1995). The role of the Ataxia telangiectasia gene in the p53, WAF1/CIP1(p21)- and GADD45-mediated response to DNA damage produced by ionising radiation. *Oncogene* **11**:1427-1435.

Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C. W., Chessa, L., Smorodinsky, N. I., Prives, C., Reiss, Y., Shiloh, Y., and Ziv, Y. (1998). Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**:1674-1677.

Baumann, P. and West, S. C. (1998). Role of the human RAD51 protein in homologous recombination and double- stranded-break repair. *Trends in Biochemical Sciences* **23**:247-251.

Bay, J. O., Grancho, M., Pernin, D., Presneau, N., Rio, P., Tchirkov, A., Uhrhammer, N., Verrelle, P., Gatti, R. A., and Bignon, Y. J. (1998). No evidence for constitutional ATM mutation in breast/gastric cancer families. *International Journal of Oncology* **12**:1385-1390.

Bennett, L. M. (1999). Breast cancer: genetic predisposition and exposure to radiation. *Molecular Carcinogenesis* **26**:143-149.

Bishop, J. O., Selman, G. G., Hickman, J., Black, L., Saunders, R. D., and Clark, A. J. (1985). The 45-kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. *Molecular and Cellular Biology* **5**:1591-1600.

Boice, J. D., Jr. (1988). Carcinogenesis--a synopsis of human experience with external exposure in medicine. *Health Physics* **55**:621-630.

Boulton, S. J. and Jackson, S. P. (1996). Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Research* **24**:4639-4648.

Brachmann, C. B., Sherman, J. M., Devine, S. E., Cameron, E. E., Pillus, L., and Boeke, J. D. (1995). The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes and Development* **9**:2888-2902.

Brackley, M. E. (2000). Personal Communication.

Broeks, A., Urbanus, J. H., Floore, A. N., Dahler, E. C., Klijn, J. G., Rutgers, E. J., Devilee, P., Russell, N. S., van Leeuwen, F. E., and van't Veer, L. J. (2000). ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. *American Journal of Human Genetics* **66**:494-500.

Bryans, M., Valenzano, M. C., and Stamato, T. D. (1999). Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutation Research* **433**:53-58.

Caldecott, K. W., Tucker, J. D., Stanker, L. H., and Thompson, L. H. (1995). Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells. *Nucleic Acids Research* **23**:4836-4843.

Callebaut, I. and Mornon, J. P. (1997). From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *Federation of Experimental Biologists Society Letters* **400**:25-30.

Canman, C. E., Chen, C. Y., Lee, M. H., and Kastan, M. B. (1994). DNA damage responses: p53 induction, cell cycle perturbations, and apoptosis. *Cold Spring Harbor Symposium of Quantitative Biology* **59**:277-286.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q., and Lander, E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* **22**:231-238.

Cascalho, M., Wong, J., Steinberg, C., and Wabl, M. (1998). Mismatch repair co-opted by hypermutation. *Science* **279**:1207-1210.

Chen, J., Silver, D. P., Walpita, D., Cantor, S. B., Gazdar, A. F., Tomlinson, G., Couch, F. J., Weber, B. L., Ashley, T., Livingston, D. M., and Scully, R. (1998). Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells. *Mol Cell* **2**:317-328.

Chu, G. (1997). Double strand break repair. *Journal of Biological Chemistry* **272**:24097-24100.

Concannon, P. J. and Gatti, R. A. Gene Clinics: Nijmegen Breakage Syndrome Website. http://www.geneclinics.org/profiles/nijmegen /details.html . (2000).

Critchlow, S. E., Bowater, R. P., and Jackson, S. P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology* **7**:588-598.

Dasika, G. K., Lin, S. C., Zhao, S., Sung, P., Tomkinson, A., and Lee, E. Y. (1999). DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. *Oncogene* **18**:7883-7899.

Dosanjh, M. K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., and Singer, B. (1994). All four known cyclic adducts formed in DNA by the vinyl chloride

metabolite chloroacetaldehyde are released by a human DNA glycosylase. *Proceedings of the National Academy of Sciences U S A* **91**:1024-1028.

Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. (2000). Polymorphisms in the DNA repair genes XRCC1 and ERCC2 and biomarkers of DNA damage in human blood mononuclear cells. *Carcinogenesis* **21**:965-971.

Duncan, B. K. and Weiss, B. (1982). Specific mutator effects of ung (uracil-DNA glycosylase) mutations in Escherichia coli. *Journal of Bacteriology* **151**:750-755.

Easton, D. F. (1994). Cancer risks in A-T heterozygotes. *International Journal of Radiation Biology* **66**:S177-S182.

Edelmann, W., Yang, K., Umar, A., Heyer, J., Lau, K., Fan, K., Liedtke, W., Cohen, P. E., Kane, M. F., Lipford, J. R., Yu, N., Crouse, G. F., Pollard, J. W., Kunkel, T., Lipkin, M., Kolodner, R., and Kucherlapati, R. (1997). Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. *Cell* **91**:467-477.

Fishman-Lobell, J. and Haber, J. E. (1992). Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. *Science* **258**:480-484.

Fishman-Lobell, J., Rudin, N., and Haber, J. E. (1992). Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Molecular and Cellular Biology* **12**:1292-1303.

Ford, B. N., Ruttan C.C., Kyle, V. L., Brackley, M. E., and Glickman, B. W. (2000). Identification of Single Nucleotide Polymorphismsin Human DNA Repair Genes. *Carcinogenesis (In press)*.

Ford, D. and Easton, D. F. (1995). The genetics of breast and ovarian cancer. *British Journal of Cancer* **72**:805-812.

Friedberg, E. C., Bond, J. P., Burns, D. K., Cheo, D. L., Greenblatt, M. S., Meira, L. B., Nahari, D., and Reis, A. M. (2000). Defective nucleotide excision repair in xpc mutant mice and its association with cancer predisposition. *Mutation Research* **459**:99-108.

Game, J. C. (1993). DNA double-strand breaks and the RAD50-RAD57 genes in Saccharomyces. *Seminars in Cancer Biology* **4**:73-83.

Gao, Y., Sun, Y., Frank, K. M., Dikkes, P., Fujiwara, Y., Seidl, K. J., Sekiguchi, J. M., Rathbun, G. A., Swat, W., Wang, J., Bronson, R. T., Malynn, B. A., Bryans, M., Zhu, C., Chaudhuri, J., Davidson, L., Ferrini, R., Stamato, T., Orkin, S. H., Greenberg, M. E., and Alt, F. W. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**:891-902.

Gatti, R. A. Gene Clinics: Ataxia-Telangiectasia Website. http://www.geneclinics.org/profiles/ataxia-telangiectasia/details.html. (2000).

German, J. (1995). Bloom's syndrome. *Dermatology Clinics* **13**:7-18.

Gilad, S., Chessa, L., Khosravi, R., Russell, P., Galanty, Y., Piane, M., Gatti, R. A., Jorgensen, T. J., Shiloh, Y., and Bar-Shira, A. (1998). Genotype-phenotype relationships in ataxia-telangiectasia and variants. *American Journal of Human Genetics* **62**:551-561.

Goss, P. E. and Sierra, S. (1998). Current perspectives on radiation-induced breast cancer. *Journal of Clinical Oncology* **16**:338-347.

Granath, F. N., Vaca, C. E., Ehrenberg, L. G., and Tornqvist, M. A. (1999). Cancer risk estimation of genotoxic chemicals based on target dose and a multiplicative model. *Risk Analysis* **19**:309-320.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., and Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**:492-495.

Grawunder, U., Zimmer, D., and Leiber, M. R. (1998). DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. *Current Biology* **8**:873-876.

Grombacher, T., Mitra, S., and Kaina, B. (1996). Induction of the alkyltransferase (MGMT) gene by DNA damaging agents and the glucocorticoid dexamethasone and comparison with the response of base excision repair genes. *Carcinogenesis* **17**:2329-2336.

Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1994). Holliday junction cleavage by yeast Rad1 protein. *Nature* **371**:531-534.

Halushka, M. K., Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., and Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genetics* **22**:239-247.

Hanawalt, P. C. (1998). Genomic instability: environmental invasion and the enemies within. *Mutation Research* **400**:117-125.

Hazra, T. K., Roy, R., Biswas, T., Grabowski, D. T., Pegg, A. E., and Mitra, S. (1997). Specific recognition of O6-methylguanine in DNA by active site mutants of human O6-methylguanine-DNA methyltransferase. *Biochemistry* **36**:5769-5776.

Hoeijmakers, J. H. (1993). Nucleotide excision repair. II: From yeast to mammals. *Trends in Genetics* **9**:211-217.

Ivanov, E. L. and Haber, J. E. (1995). RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **15**:2245-2251.

Ivanov, E. L., Sugawara, N., Fishman-Lobell, J., and Haber, J. E. (1996). Genetic requirements for the single-strand annealing pathway of double- strand break repair in Saccharomyces cerevisiae. *Genetics* **142**:693-704.

Jackson, S. P. (1996). DNA damage detection by DNA dependent protein kinase and related enzymes. *Cancer Surveys* **28**:261-279.

Jackson, S. P. (1997). DNA-dependent protein kinase. *The International Journal of Biochemistry and Cell Biology* **29**:935-938.

Jasin, M. (2000). Chromosome breaks and genomic instability. *Cancer Investigations* **18**:78-86.

Jeggo, P. A. (1998). Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiation Research* **150**:S80-S91.

Jongmans, W., Vuillaume, M., Chrzanowska, K., Smeets, D., Sperling, K., and Hall, J. (1997). Nijmegen breakage syndrome cells fail to induce the p53-mediated DNA damage response following exposure to ionizing radiation. *Molecular and Cellular Biology* **17**:5016-5022.

Jongmans, W. and Hall, J. (1999). Cellular responses to radiation and risk of breast cancer. *European Journal of Cancer* **35**:540-548.

Kadyk, L. C. and Hartwell, L. H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics* **132**:387-402.

Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R. W. (1991). Participation of p53 protein in the cellular response to DNA damage. *Cancer Research* **51**:6304-6311.

Kim, N., Kage, K., Matsuda, F., Lefranc, M. P., and Storb, U. (1997). B lymphocytes of xeroderma pigmentosum or Cockayne syndrome patients with inherited defects in nucleotide excision repair are fully capable of somatic hypermutation of immunoglobulin genes. *The Journal of Experimental Medicine* **186**:413-419.

Kirchgessner, C. U., Patil, C. K., Evans, J. W., Cuomo, C. A., Fried, L. M., Carter, T., Oettinger, M. A., and Brown, J. M. (1995). DNA-dependent kinase (p350) as a candidate gene for the murine SCID defect. *Science* **267**:1178-1183.

Knudson, A. G. (1996). Hereditary cancer: two hits revisited. *Journal of Cancer Research and Clinical Oncology* **122**:135-140.

Kolodner, R. D. and Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. *Current Opinion in Genetics and Development* **9**:89-96.

Kuerbitz, S. J., Plunkett, B. S., Walsh, W. V., and Kastan, M. B. (1992). Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proceedings of the National Academy of Sciences U S A* **89**:7491-7495.

Lai, C. and Shields, P. G. (1999). The role of interindividual variation in human carcinogenesis. *Journal of Nutrition* **129**:552-555.

Leber, R., Wise, T. W., Mizuta, R., and Meek, K. (1998). The XRCC4 gene product is a target for and interacts with the DNA- dependent protein kinase. *Journal of Biological Chemistry* **273**:1794-1801.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* **396**:643-649.

Lewin, B. (1997). "Genes VI," Oxford University Press, New York.

Li, L., Lu, X., Peterson, C. A., and Legerski, R. J. (1995). An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Molecular and Cellular Biology* **15**:5396-5402.

Li, Z., Otevrel, T., Gao, Y., Cheng, H. L., Seed, B., Stamato, T. D., Taccioli, G. E., and Alt, F. W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**:1079-1089.

Lieber, M. R. (1991). Site-specific recombination in the immune system. *The FASEB Journal: official publication of the Federation of American Societies for Experimental Biology* **5**:2934-2944.

Lin, F. L., Sperle, K., and Sternberg, N. (1990). Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. *Molecular and Cellular Biology* **10**:103-112.

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

Little, J. B., Yandell, D. W., and Liber, H. L. (1987). Molecular Analysis of Mutations at the *tk* and *hgprt* loci in human cells. *In* "Mammalian Cell Mutagenesis" (M. M. Moore, Ed.), pp. 225, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Little, J. B. (1993). Cellular, molecular, and carcinogenic effects of radiation. *Hematology/Oncology Clinics of North America* **7**:337-352.

Lubin, J. H. (1988). Models for the analysis of radon-exposed populations. *Yale Journal of Biology and Medicine* **61**:195-214.

Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. (1999). XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. *Cancer Research* **59**:2557-2561.

Lunn, R. M., Helzlsouer, K. J., Parshad, R., Umbach, D. M., Harris, E. L., Sanford, K. K., and Bell, D. A. (2000). XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis* **21**:551-555.

Lutz, W. K. (1990). Endogenous genotoxic agents and processes as a basis of spontaneous carcinogenesis. *MutationResearch* **238**:287-295.

Lyman, G. H. (1992). Risk factors for cancer. *Primary Care* **19**:465-479.

Mason, R. M., Thacker, J., and Fairman, M. P. (1996). The joining of non-complementary DNA double-strand breaks by mammalian extracts. *Nucleic Acids Research* **24**:4946-4953.

McWhir, J., Selfridge, J., Harrison, D. J., Squires, S., and Melton, D. W. (1993). Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nature Genetics* **5**:217-224.

Mimori, T. and Hardin, J. A. (1986). Mechanism of interaction between Ku protein and DNA. *Journal of Biological Chemistry* **261**:10375-10379.

Mitra, S. and Kaina, B. (1993). Regulation of repair of alkylation damage in mammalian genomes. *Progress in Nucleic Acids Research and Molecular Biology* **44**:109-142.

Miyazaki, W. Y. and Orr-Weaver, T. L. (1994). Sister-chromatid cohesion in mitosis and meiosis. *Annual Review of Genetics* **28**:167-187.

National Center for Biotechnology Information. NCBI Single Nucleotide Polymorphism Data Base. http//www.ncbi.nlm.nih.gov/SNP/. (2000).

Nichols, K. E., Levitz, S., Shannon, K. E., Wahrer, D. C., Bell, D. W., Chang, G., Hegde, S., Neuberg, D., Shafman, T., Tarbell, N. J., Mauch, P., Ishioka, C., Haber, D. A., and Diller, L. (1999). Heterozygous germline ATM mutations do not contribute to radiation- associated malignancies after Hodgkin's disease. *Journal of Clinical Oncology* **17**:1259.

Ozenberger, B. A. and Roeder, G. S. (1991). A unique pathway of double-strand break repair operates in tandemly repeated genes. *Molecular and Cellular Biology* **11**:1222-1231.

Pfeiffer, P., Thode, S., Hancke, J., and Vielmetter, W. (1994). Mechanisms of overlap formation in nonhomologous DNA end joining. *Molecular and Cellular Biology* **14**:888-895.

Preston, R. J. (1996). Interindividual variations in susceptibility and sensitivity: linking risk assessment and risk management. *Toxicology* **111**:331-341.

Price, E. A., Bourne, S. L., Radbourne, R., Lawton, P. A., Lamerdin, J., Thompson, L. H., and Arrand, J. E. (1997). Rare microsatellite polymorphisms in the DNA repair genes XRCC1, XRCC3 and XRCC5 associated with cancer in patients of varying radiosensitivity. *Somatic Cell and Molecular Genetics* **23**:237-247.

Rajewsky, M. F., Engelbergs, J., Thomale, J., and Schweer, T. (1998). Relevance of DNA repair to carcinogenesis and cancer therapy. *Recent Results in Cancer Research* **154**:127-146.

Robins, P. and Lindahl, T. (1996). DNA ligase IV from HeLa cell nuclei. *Journal of Biological Chemistry* **271**:24257-24261.

Roth, D. B. and Wilson, J. H. (1986). Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Molecular and Cellular Biology* **6**:4295-4304.

Rufer, J. T. and Morgan, W. F. (1992). Potentiation of DNA damage by inhibition of poly(ADP-ribosyl)ation: a test of the hypothesis for random nuclease action. *Experimental Cell Research* **200**:506-512.

Russell, N. S., Arlett, C. F., Bartelink, H., and Begg, A. C. (1995). Use of fluorescence in situ hybridization to determine the relationship between chromosome aberrations and cell survival in eight human fibroblast strains. *International Journal of Radiation Biology* **68**:185-196.

Sacki, T., Machida, I., and Nakai, S. (1980). Genetic control of diploid recovery after gamma-irradiation in the yeast *Saccharomyces cerevisiae*. *Mutation Research* **73**:251-265.

Sager, R. (1979). Transposable elements and chromosomal rearrangements in cancer--a possible link. *Nature* **282**:447-449.

Sancar, A. (1995). DNA Repair in Humans. *Annual Review of Genetics* **29**:69-105.

Sancar, A. (1996). DNA excision repair. *Annual Review of Biochemistry* **65**:43-81.

Satoh, M. S., Jones, C. J., Wood, R. D., and Lindahl, T. (1993). DNA excision-repair defect of xeroderma pigmentosum prevents removal of a class of oxygen free radical-induced base lesions. *Proceedings of the National Academy of Sciences U S A* **90**:6335-6339.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., and Sfez, S. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**:1749-1753.

Sedlacek, Z., Kodet, R., Seemanova, E., Vodvarka, P., Wilgenbus, P., Mares, J., Poustka, A., and Goetz, P. (1998). Two Li-Fraumeni syndrome families with novel germline p53 mutations: loss of the wild-type p53 allele in only 50% of tumours. *British Journal of Cancer* **77**:1034-1039.

Seemayer, T. A. and Cavenee, W. K. (1989). Molecular mechanisms of oncogenesis. *Laboratory Investigations: a journal of technical methods and pathology* **60**:585-599.

Sheer, D. and Squire, J. (1996). Clinical applications of genetic rearrangements in cancer. *Seminars in Cancer Biology* **7**:25-32.

Shen, M. R., Jones, I. M., and Mohrenweiser, H. (1998). Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Research* **58**:604-608.

Shimizu, Y., Kato, H., and Schull, W. J. (1990). Studies of the mortality of A-bomb survivors. 9. Mortality, 1950-1985: Part 2. Cancer mortality based on the recently revised doses (DS86). *Radiation Research* **121**:120-141.

Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N., Ikeo, K., and Ogawa, T. (1993). Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and recA. *Nature Genetics* **4**:239-243.

Shinohara, A. and Ogawa, T. (1995). Homologous recombination and the roles of double-strand breaks. *Trends in Biochemical Sciences* **20**:387-391.

Sjalander, A., Birgander, R., Saha, N., Beckman, L., and Beckman, G. (1996). p53 polymorphisms and haplotypes show distinct differences between major ethnic groups. *Human Heredity* **46**:41-48.

Smith, K. C. (1992). Spontaneous mutagenesis: experimental, genetic and other factors. *Mutation Research* **277**:139-162.

Stumm, M., Sperling, K., and Wegner, R. D. (1997). Noncomplementation of radiation-induced chromosome aberrations in ataxia-telangiectasia/ataxia-telangiectasia-variant heterodikaryons. *American Journal of Human Genetics* **60**:1246-1251.

Sugawara, N. and Haber, J. E. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Molecular and Cellular Biology* **12**:563-575.

Sugawara, N., Ivanov, E. L., Fishman-Lobell, J., Ray, B. L., Wu, X., and Haber, J. E. (1995). DNA structure-dependent requirements for yeast RAD genes in gene conversion. *Nature* **373**:84-86.

Sung, P., Prakash, L., and Prakash, S. (1992). Renaturation of DNA catalysed by yeast DNA repair and recombination protein RAD10. *Nature* **355**:743-745.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* **33**:25-35.

Taccioli, G. E., Gottlieb, T. M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A. R., Alt, F. W., Jackson, S. P., and Jeggo, P. A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science* **265**:1442-1445.

Teo, S. H. and Jackson, S. P. (1997). Identification of Saccharomyces cerevisiae DNA ligase IV: involvement in DNA double-strand break repair. *European Molecular Biology Organization Journal* **16**:4788-4795.

Thacker, J., Chalk, J., Ganesh, A., and North, P. (1992). A mechanism for deletion formation in DNA by human cell extracts: the involvement of short sequence repeats. *Nucleic Acids Research* **20**:6183-6188.

Thacker, J. (1999). Repair of ionizing radiation damage in mammalian cells. Alternative pathways and their fidelity. *C R Acad Sci III (Comptes rendus de l'Academie des sciences)* **322**:103-108.

Thompson, L. H. and Schild, D. (1999). The contribution of homologous recombination in preserving genome integrity in mammalian cells. *Biochimie* **81**:87-105.

Tsukamoto, Y., Kato, J., and Ikeda, H. (1997). Silencing factors participate in DNA repair and recombination in Saccharomyces cerevisiae. *Nature* **388**:900-903.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., and MoritaT (1996). Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. *Proceedings of the National Academy of Sciences U S A* **93**:6236-6240.

Ucla, C., van Rood, J. J., Gorski, J., and Mach, B. (1987). Analysis of HLA-D micropolymorphism by a simple procedure: RNA oligonucleotide hybridization. *Journal of Clinical Investigation* **80**:1155-1159.

van der Spek, P. J., Kobayashi, K., Bootsma, D., Takao, M., Eker, A. P., and Yasui, A. (1996). Cloning, tissue expression, and mapping of a human photolyase homolog with similarity to plant blue-light receptors. *Genomics* **37**:177-182.

Ward, J. F. (1988). DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Progress in Nucleic Acids Research and Molecular Biology* **35**:95-125.

Ward, J. F. (1994). The complexity of DNA damage: relevance to biological consequences. *International Journal of Radiation Biology* **66**:427-432.

Ward, J. F., Milligan, J. R., and Jones, G. D. D. (1994). Biological consequences of non-homogeneous energy depositions by ionizing radiation. *Radiation Protection Dosimetry* **52**:271-276.

Weaver, D. T. (1995). V(D)J recombination and double-strand break repair. *Advances in Immunology* **58**:29-85.

Wei, Y. F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J., Yu, G. L., Wang, R. P., Shell, B. K., Nash, R. A., and Schar, P. (1995). Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Molecular and Cellular Biology* **15**:3206-3216.

West, C. M., Elyan, S. A., Berry, P., Cowan, R., and Scott, D. (1995). A comparison of the radiosensitivity of lymphocytes from normal donors, cancer patients, individuals with ataxia-telangiectasia (A-T) and A-T heterozygotes. *International Journal of Radiation Biology* **68**:197-203.

Wilson, D. M., III and Thompson, L. H. (1997). Life without DNA repair. *Proceedings of the National Academy of Sciences U S A* **94**:12754-12757.

Wilson, T. E., Grawunder, U., and Lieber, M. R. (1997). Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**:495-498.

Yoshimura, Y., Morita, T., Yamamoto, A., and Matsushiro, A. (1993). Cloning and sequence of the human RecA-like gene cDNA. *Nucleic Acids Research* **21**:1665.

Yu, J. J., Mu, C., Lee, K. B., Okamoto, A., Reed, E. L., Bostick-Bruton, F., Mitchell, K. C., and Reed, E. (1997). A nucleotide polymorphism in ERCC1 in human ovarian cancer cell lines and tumor tissues. *Mutation Research* **382**:13-20.

Yu, J. J., Lee, K. B., Mu, C., Li, Q., Abernathy, T. V., Bostick-Bruton, F., and Reed, E. (2000). Comparison of two human ovarian carcinoma cell lines (A2780/CP70 and MCAS) that are equally resistant to platinum, but differ at codon 118 of the ERCC1 gene. *International Journal of Oncology* **16**:555-560.

Yu, Z., Chen, J., Ford, B. N., Brackley, M. E., and Glickman, B. W. (1999). Human DNA repair systems: an overview. *Environmental and Molecular Mutagenesis* **33**:3-20.

Yunis, J. J. (1986). Chromosomal rearrangements, genes, and fragile sites in cancer: clinical and biologic implications. *Important Advances in Oncology*93-128.

Yunis, J. J. (1990). Fragile sites, genomic rearrangements and cancer predisposition. *Birth Defects Original Article Series* **26**:67-83.

Zdzienicka, M. Z. (1995). Mammalian mutants defective in the response to ionizing radiation- induced DNA damage. *Mutation Research* **336**:203-213.

Appendix

Fisher's Exact Test (95% confidence interval) for association between breast or colon cancer patients and 401 SNP

	Breast cancer patients	Colorectal cancer patients	Total
401 variant	1 (2%)	1 (2%)	2 (4%)
401 reference	16 (33%)	30 (63%)	46 (96%)
Total	17 (35%)	31 (65%)	48 (100%)

The two-sided *p*-value is 1.0000, considered not significant. The row/column association is not statistically significant.

Odds Ratio = 1.875 (0.1098 to 32.032)

Calculated with Instat (http://graphpad.com) Fisher's two-tailed *p*-value Exact test

Fisher's Exact Test (95% confidence interval) for association between breast or colon cancer patients and 921 SNP

	Breast cancer patients	Colorectal cancer patients	Total
921 variant	16 (33%)	28 (58%)	44 (92%)
921 reference	1 (2%)	3 (6%)	4 (8%)
Total	17 (35%)	31 (65%)	48 (100%)

The two-sided *p*-value is 1.0000, considered not significant. The row/column association is not statistically significant.

Odds Ratio = 1.714 (0.1642 to 17.897)

Calculated with Instat (http://graphpad.com) Fisher's two-tailed *p*-value exact test

Fisher's Exact Test (95% confidence interval) for association between control sample ancestry and 401 SNP

	Asian	European	Total
401 variant	0 (0%)	12 (9%)	12 (9%)
401 reference	5 (4%)	111 (87%)	116 (91%)
Total	5 (4%)	123 (96%)	128 (100%)

The two-sided *p*-value is 1.0000, considered not significant. The row/column association is not statistically significant.

Odds Ratio 0.8109 (0.4226 to 15.559)

Since at least one value is zero, 0.5 was added to each value to make calculations possible.

Calculated with Instat (http://graphpad.com) Fisher's two-tailed *p*-value exact test

Fisher's Exact Test (95% confidence interval) for association between breast or colon cancer patients and 921 SNP

	Asian	European	Total
921 variant	5 (4%)	121 (95%)	126 (98%)
921 reference	0 (0%)	2 (2%)	2 (2%)
Total	5 (4%)	123 (96%)	128 (100%)

The two-sided *p*-value is 1.0000, considered not significant. The row/column association is not statistically significant.

Odds Ratio = 0.2263 (0.009650 to 5.309)

Since at least one value is zero, 0.5 was added to each value to make calculations possible.

Calculated with Instat (http://graphpad.com) Fisher's two-tailed P-value exact test

VITA

Surname: Kyle Given Names: Victoria Lea

Place of Birth: San Antonio, Texas, United States

Educational Institutions Attended:

University of Victoria	1999-2000
Texas A&M University	1990-1995

Degrees Awarded:

B.Sc.	Texas A&M University	1995
B.Sc.	Texas A&M University	1993

Honors and Awards:

Texas A&M University College of Education Distinguished Student Award 1994-1995

Texas A&M University College of Agriculture Distinguished Student Award 1992-1993

Publications:

Presented poster: *Identification of Single Nucleotide Polymorphisms in Two Double Strand Break Repair Genes – XRCC4 and RAD51* at Environmental Mutagen Society Conference; New Orleans, LA; 2000.

In press *Carcinogenesis*: *Identification of Single Nucleotide Polymorphismsin Human DNA Repair Genes;* Authors: Barry N. Ford, Cindy C. Ruttan, Victoria L. Kyle, Moyra E. Brackley and Barry W. Glickman; 2000.

Presented poster: *Identification of Single Nucleotide Polymorphisms in Two Double Strand Break Repair Genes* at Bridging the Strait of Georgia Conference; Victoria, B.C.; 2000.

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

Identification of Single Nucleotide Polymorphisms in Two DNA Double-Strand Break Repair Genes – *XRCC4* and *RAD51*

Author 

Victoria Lea Kyle

September 5, 2000