DNA Repair Capacity as a Marker of Breast Cancer Susceptibility

Maya A. Kappil

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Public Health
under the Executive Committee
of the Mailman School of Public Health

COLUMBIA UNIVERSITY
2013

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

Introduction: The wide-ranging prognostic implications of a breast cancer diagnosis highlight the need to better enable women to make informed decisions regarding screening and treatment options. As several cancer susceptibility syndromes have been linked to germline mutations resulting in defective DNA repair, including the predisposition to breast cancer due to *BRCA1* and *BRCA2* mutations, more subtle defects in DNA repair capacity may contribute to the components driving differential susceptibility within the general population. Hence, understanding the role of DNA repair capacity in breast cancer onset may aid in the development of a more comprehensive risk profile, thereby furthering the effort to target relevant populations for early screening.

In the studies undertaken for this dissertation, we employed various methodologies capturing endpoints across different repair pathways detectable in blood to both further elucidate the etiologic basis of breast cancer development and leverage the information into the potential development of a screening biomarker.

Methods: For the phenotypic assessment of nucleotide excision repair (NER) capacity, we developed an ELISA-based method to determine benzo(a)pyrene diolepoxide (BPDE)-DNA adduct capacity in lymphoblastoid cell lines. Gene expression levels were assessed with pre-designed Taqman kits in RNA-derived cDNAs from mononuclear cells using a real-time PCR-based platform. Methylation analysis was conducted with in-house designed assays on bisulfite-converted DNA from mononuclear cells using a pyrosequencing platform. Finally, single nucleotide polymorphisms (SNP) genotyping was assessed in DNA derived from white blood cells with pre-designed Taqman SNP genotyping assays using a real-time PCR-based platform. All studies were conducted in sister-sets enrolled in the New York site within the Breast Cancer Family Registry and all statistical analysis was conducted using the R Foundation for Statistical Computing (2011).

Results: We did not detect an association between the ELISA-based phenotypic assessment of NER capacity in the lymphoblastoid cells lines of the sister-sets (n=246, 114 sister-sets) and breast cancer

risk (OR = 1.0, 95%CI=0.95, 1.04). Furthermore, we did not observe a correlation with previously determined NER capacity in the same population using an immunohistochemical-based method (r= -0.01, p=0.86).

In our gene expression study (n=569, 218 sister-sets), women in the lowest tertile of ATM expression had a heightened risk of breast cancer compared to women in the highest tertile of expression, adjusted for age at blood draw and smoking status (OR=2.12, 95%CI=1.09, 4.12). This association was largely restricted to women with an extended family history of breast cancer ($p_{interaction}$ = 0.06). Additionally, women in the lowest tertile of MSH2 expression also had a heightened risk of breast cancer compared to women in the highest tertile of expression, adjusted for age at blood draw and smoking status (OR=2.75, 95%CI=1.31, 5.79). The association observed between reductions in ATM expression level and breast cancer risk was lost upon incorporating previously determined end-joining capacity of EcoRI-generated sticky end substrates (OR=1.28, 95%CI=0.15, 11.2) and HincII-generated blunt end substrates (OR=1.55, 95%CI=0.15, 15.5) into the model, suggesting that the impact on risk due to reductions in ATM expression maybe partially driven by the reduction in double strand break repair capacity.

In our study investigating breast cancer risk due to the impact of epigenetic modulation on DNA repair gene activity (n=569, 218 sister-sets), no association with risk was observed due to differential promoter methylation levels of *BRCA1* (OR=1.09, 95%CI=0.98, 1.20), *MLH1* (OR=1.19, 95%CI=0.91, 1.55) or *MSH2* (OR=0.89, 95%CI=0.48, 1.64). Furthermore, no correlation between *BRCA1* methylation and expression (r=-0.05, p=0.39) or *MSH2* methylation and expression (r=-0.04, p=0.39) was observed.

Finally, our mismatch repair genotyping study (n=714, 313 sister-sets) indicated an association between the variant MutY_rs3219489 (OR=2.23, 95%CI=1.10, 4.52) and breast cancer risk, as well as a borderline association with risk due to the variant MSH2_rs2303428 (OR=1.71, 95%CI=0.99, 2.95). Furthermore, a protective effect was observed due to the variant MLH3_rs175080, restricted to women without an extended family history of breast cancer ($p_{interaction}$ = 0.03).

Conclusion: These studies suggest that the deregulation of targets spanning various DNA repair pathways contribute to the risk of familial breast cancer.

Table of Contents

List of Tables and Figures

Chapter I

Figures

Chapter II

Figures

Tables

Chapter III

Figures

Tables

Chapter IV

Figures

Tables

Chapter V

Figures

Tables

Chapter VI

Figures

List of Abbreviations

8oxoG	8-oxoguanine
AP	apurinic/apyrimidinic
APE	Apurinic/apyrimidinic endonuclease class
AT	Ataxia telangiectasia
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
B(a)P	Benzo(a)pyrene
BER	Base excision repair
BMI	Body mass index
BPDE	Benzo(a)pyrene diol epoxide
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
BSA	Bovine serum albumin
CETN2	Centrin-2
CSA	Cockayne syndrome group A protein
CSB	Cockayne syndrome group B protein
dAMP	Deoxyadenylic monophosphate
dCMP	Deoxycytidine monophosphate
dTMP	Deoxythymidylic monophosphate
DDB1	DNA damage-binding protein 1
DDB2	DNA damage-binding protein 2
DMSO	Dimethyl sulfoxide
d-loop	displacement loop
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dNTP	Deoxyribonucleotide triphoshpate
DSBR	Double-strand break repair
EJ	End-joining
ERCC1	Excision repair cross-complementating rodent repair deficiency, complementation group 1
Exo1	Exonuclease 1
FBS	Fetal bovine serum
GEE	Generalized estimating equations
GWAS	Genome-wide association study
GGR	Global genome repair
HW	Hardy-Weinberg
HNPCC	Hereditary nonpolyposis colorectal cancer
HCR	Host cell reactivation
HR	Homologous repair
HRT	Hormone replacement therapy
IDL	insertion/deletion loop
miRNA	micro RNA
MLH1	MutL homolog 1
MMR	Mismatch repair
MRN	Mre11-Rad50-NBS
MSH2	MutS homolog
MSH3	MutS homolog 3
MSH6	MutS homolog 6
MSI	Microsatellite instability
NER	Nucleotide excision repair
NHEJ	Non-homologous end-joining
NTC	Non-template controls

NY-BCFR	New York site of Breast Cancer Family Registry
OGG1	8-oxoguanine DNA glycosylase
OR	Odds ratio
PAH	Polycyclic aromatic hydrocarbons
PALB2	Partner and localizer of BRCA2
PBC	Peripheral Blood Cells
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PIKK	PI3-kinase-like phosphor-inositide 3-kinase family
PHA	Phytohemagglutinin
PMS2	Postmeiotic segregation increased 2
ROS	Reactive oxygen species
RPA	Replication Protein A
SAM	*S*-adenosylmethionine
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
TCR	Transcription-coupled repair
TdT	Terminal deoxynucleotidyl transferase
TFIIH	Transcription factor II Human
WBC	White Blood Cells
XP	*Xeroderma Pigmentosum*
XPA	Xeroderma Pigmentosum, complementation group A
XPB	Xeroderma Pigementosum, complementation group B
XPC	Xeroderma Pigmentosum, complementation group C
XPD	Xeroderma Pigmentosum, complementation group D
XPF	Xeroderma Pigmentosum, complementation group F
XPG	Xeroderma Pigmentosum, complementation group G
XRCC1	X-ray complementing defective repair in Chinese hamster cells 1
XRCC4	X-ray complementing defective repair in Chinese hamster cells 4

Acknowledgements

First and foremost, I would like to thank my mentor, *Dr. Regina Santella*. She took a gamble on me, taking me in as an orphaned technician during my Master's degree, and then again, suggesting I continue my tenure in the lab as a doctoral student. My decision to apply was largely shaped by her example as a role model of a woman who forged a successful career in this field while maintaining the ability to balance her work and family life. As a mentor, her door has always been open to me, as well as all other students who have passed through her class and lab. Despite her hectic schedule through the many hats she wears, she has always taken the time to provide feedback ranging from experimental protocol, written drafts and career paths. Her guidance throughout my time in the lab has been invaluable.

I would like to thank my additional committee members, *Dr. Greg Freyer*, *Dr. Mary Gamble*, *Dr. Jeanine Genkinger* and *Dr. Mary Beth Terry* for taking the time and effort to meet with me and provide me with feedback throughout this process.

This work would not have been possible without the support of fellow Santella lab members, past and present. I would especially like to mention *Qiao Wang*, who in addition to having worked with me side-by-side through all these years, teaching me invaluable skills on the bench, has been one of my closest friends and leading member of my lab-mom squad, supported by other core facility members, *Iryna Sirosh* and *Irina Gurvich*. I would like to thank *Gail Garbowski* for readily attending to administrative needs, ensuring among other things that all necessary supplies were always readily available, and to our lab manager, Irina Gurvich, for providing me with all necessary database-related information. I am also grateful to the advice I received from post-doctoral fellows past and present in our lab, including *Hui-Chen Wu*, *Lissette Delgado-Cruzata*, *Jennifer Zipprich* and *Hulya Yazici*.

I would like to thank the cohort of EHS students who have been part of my experience during my tenure here. I would especially like to thank *Maria Rosa*, *Megan Niedzwiecki*, *Nada Hamade, Allan Just*, *Rob Prins*, *Caitlin Howe*, *Tiffany Sanchez* and *Brandilyn Peters* for their feedback and fellowship. I feel honored to count you among my peers.

Finally, I would like to thank my parents, for the sacrifices they made in coming to this country to provide my sister and I the opportunities we had to pursue our passions and interests. I would like to thank my sister, *Cynthia*, for being my voice of reason. And, *M*, for the steady supply of prized Butter Kekse and IT support.

Dedication

for my parents

Chapter I. Literature review

Breast cancer demographics

Breast cancer is the most common type of cancer and second leading cancer-death among women in the US, with one in every six diagnosed women succumbing to the disease[1]. Treatment options include lumpectomy or mastectomy, often followed by a regimen including any or all of chemotherapy, radiation therapy and hormone-therapy. As with all cancers, breast cancer is most treatable prior to the clinical manifestation of the disease. In addition to a successful overall outcome, the state of progression of the disease at the time of diagnosis also impacts the quality of life, dictating the duration and severity of the regimen required during treatment. Hence, early detection of this disease is of vital importance.

Routine mammography is the recommended screening method of choice for early detection of breast cancer. Using this technique, 80-90% of women with breast cancer can be diagnosed prior to the presentation of clinical symptoms[1]. However, there are limitations to the technology. Patient characteristics, including age, presence of micro-calcification and dense breast tissue, can obscure detection, contributing to the up to 20% of asymptomatic cases who will remain undiagnosed despite regular screening[2]. In addition to sensitivity concerns, there are also issues regarding the specificity of the methodology. Fifty percent of women receiving yearly mammograms will receive a false positive diagnosis during their lifetime[3]. In fact, a majority of women with abnormal mammograms do not have cancer, necessitating biopsies to confirm the test result[4]. The implications of a positive diagnosis and need for further invasive tests likely contributes to undue stress and anxiety.

Lastly, there is a suggested trend towards overdiagnosis and overtreatment based on mammography. Up to 30% of cancers detected are, in fact, non-invasive cancers and pre-malignant lesions with a low proliferative index[5]. Intervening on such cases with the prescribed treatment of surgery, chemotherapy, radiation and/or hormone therapy likely provides little benefit, and may, in actuality, be harmful due to the duress involved[6,7]. Consequently, the reliance on the diagnostic capabilities of mammograms has mainly led to the treatment of cases with generally favorable prognosis, at the cost of missing women who present at a younger age and with rapid tumor growth[8].

In the US, annual mammograms are recommended starting at the age of 40. In addition to the aforementioned inherent limitations in the technology, implementation of the current standard is complicated by the fact that risk is not equivalently distributed throughout the population. For a woman with a lower likelihood of developing aggressive breast cancer prior to menopause, screening less frequently and starting at a later age may be sufficient and decrease the likelihood of overdiagnosis and overtreatment. However, for others, more active surveillance than the current recommendations may be necessary for successful intervention. Such measures could include the option of prophylactic mastectomy and tamoxifen treatment. In fact, tamoxifen has been shown to reduce breast cancer incidence by up to 50% among high-risk women. However, the side-effects associated with long-term treatment with such chemopreventive agents are not trivial, including risk of endometrial cancer, pulmonary emboli and menopausal symptoms[9]. Hence, there exists a need to provide adequate information regarding the costs and benefits relevant to the given situation, enabling women to make more informed decisions regarding screening and intervention strategies. Fulfilling this imperative will require a more thorough understanding of the underpinnings of the disease to better profile women into more clearly defined risk categories.

Known risk factors of breast cancer

A number of risk factors contributing to breast cancer risk have been identified thus far[10]. As with many other cancer-subtypes, breast cancer risk increases with age, as few women below the age of 30 present with the disease. A number of theories have been put forward to explain the observed relationship. These include age-related hormonal changes, failing immunologic surveillance over time and the effect of environmental insults accumulated throughout the life-span[11].

Approximately 20% of risk is attributed to having a family history of breast cancer[12,13]. The increasing risk with increasing numbers of first-degree relatives is further enhanced the earlier the relative presented at the age of diagnosis[14]. The observed clustering of cases within families could be attributed to shared inherited and/or lifestyle and environmental factors. Twin studies have sought to address the degree to which genetic and non-genetic factors contribute to familial risk by capitalizing on the fact that monozygotic twins have identical copies of genetic information while dizygotic twins

share only half the information. Meanwhile, the degree to which non-genetic factors are shared is likely not differentiated by zygosity. Hence, an overriding genetic influence should be reflected by greater concordance among monozygotic twins, while an equivalent concordance would suggest that shared non-genetic factors play a more active role. A number of such studies have, in fact, shown a greater concordance for breast cancer among monozygotic twins than dizygotic twins, suggesting that familial presentation of the disease is mainly driven by genetic components[15,16].

The identification of germline mutations in the tumor suppressor genes *BRCA1* and *BRCA2* have been key in delineating the genetic components mediating a role in familial breast cancer[17]. Women who inherit mutations in these genes have up to an 85% chance of developing the disease by the age of 70[18]. However, due to the rarity of such mutations, only 5% of the risk posed by family history of the disease is believed to be attributed to the loss of function of these genes, highlighting the large proportion of risk that has yet to be accounted for[17].

Various reproductive factors have been shown to play a role in risk as well. In a case-control study drawn from a nationwide screening program, increasing age at menarche was associated with decreasing risk, with a 23% reduction in risk among women with an age at menarche after the age of 15 compared to women with menarche before the age of 12. The same study also demonstrated a heightened risk for pre-menopausal women compared to post-menopausal women, and induced menopause through bilateral ooppherectomy afforded even greater reduction in risk[19]. Parity also exerts a protective effect[20]. However, the reduction in risk among parous women compared to nulli-parous women is attenuated with increasing age at parity[21]. It has been suggested that the means by which these reproductive factors mediate an effect on breast cancer risk may involve the influence on mammary gland differentiation. During development, undifferentiated epithelial cells in the mammary gland undergo rapid proliferation, heightening the region's susceptibility to carcinogenic agents. This undifferentiated proliferative state continues until the final stage of pregnancy, at which point terminal differentiation is initiated. Hence, early age at menarchy and late age at first pregnancy are factors which may broaden the window of susceptibility during which the accrual of hits can lead to transformation[22].

In addition to intrinsic traits, efforts have been made to identify environmental and lifestyle factors as potential intervention endpoints to reduce risk. One such factor includes lifetime exposure to estrogen, which has been implicated through various routes. Exogenous administration of estrogen was commonly prescribed with hormone replacement therapy (HRT) for the management of menopausal symptoms. Several studies have now shown that HRT use within the past 5 years is significantly associated with increased breast cancer risk[23,24].

Estrogen exposure may also exert a more indirect influence on breast cancer risk by mediating the observed association with obesity. While higher body mass index (BMI) is considered a risk factor among postmenopausal women, it is considered a protective factor among premenopausal women[25,26]. This differential effect may be due to the bioavailability of circulating estrogen. In premenopausal women, estrogen production is primarily mediated by the ovaries, which maintain the level of freely circulating estrogen under tight homeostatic regulation[27]. In postmenopausal women, there is an increase in extraglandular estrogen production, including in adipose tissue, where the level of freely circulating estrogen is not as strictly regulated. Hence, higher BMI leads to a concomitant rise in estrogen bioavailability, which in turn, has been linked to an increase in breast cancer risk among postmenopausal women[28]. The association is further substantiated by the fact that postmenopausal women are more likely to develop estrogen-dependent tumors compared to premenopausal women, and may, therefore, be more responsive to the mitogenic effect of an obesity-induced increase in estrogen level[29].

Radiation exposure is the most significant environmental agent associated with breast cancer risk, though the risk is far more attenuated than what has been reported with other more typically linked radiation-induced cancer sub-types, including leukemia and thyroid cancer. Still there are special implications for vigilance despite the incremental increase in risk in breast cancer, as both monitoring and treating disease require radiation exposure[30,31].

Using these currently known risk factors, models have been developed to generate approximate risk profiles with the purpose of allowing women and practitioners to make more informed choices in selecting screening and intervention strategies. The Gail model is one such popularly used model

which incorporates established risk factors and accounts for possible synergy between them. Included in the model are family history, late age at childbirth, early menarche, multiple benign biopsies, all modified by age[32,33]. However, as each of these factors account for only a modest proportion of risk, the models generated based on these risk factor lack adequate predictive power[10].

Much attention has been paid to identifying additional factors, particularly genetic components, to improve upon such models. However, isolating areas of deregulation that have a meaningful impact on disease susceptibility requires an understanding of the underlying processes involved in the onset and progression of breast cancer.

Hallmarks of cancer

Like all forms of carcinogenesis, breast cancer arises from a multi-step process, enabling the out of control growth of abnormal cells. This entails the sequential transformation of normal cells into a premalignant lesion, a subset of which progress to a localized tumor, a further subset of which evolve into an invasive tumor, out of which an additional set of changes can result in ultimate metastasis. While the specific steps from initiation to progression vary from case to case, they can be categorized according to a common set of hallmarks met to transform a normal cell into a neoplastic one. The first among these is the ability to continually proliferate. Under normal conditions, entry and progression through the cell-cycle is tightly regulated, with specific growth signaling cues necessary to initiate replication. In addition to the specific signals needed to trigger replication, the regulation of cell-cycle entry is further enforced by growth suppressor signaling. These signals are also key in ensuring that replication only progresses once proper functioning of cellular processes has been verified. Transformed cells acquire the ability to overcome both checkpoints, and the ensuing abnormal proliferation of normal cells can be observed in the breast as hyperplasia[34]. In the presence of insurmountable physiologic stressors, such as the accumulation of irreversible DNA damage, normal cells are wired to trigger apoptosis to prevent further deregulation and propagation of the resulting instability. During the transformation process, cells acquire the ability to desensitize themselves from this signaling cascade, thereby, furthering the transformative potential of these cells by enabling the continued unchecked accrual of genetic alterations. Normal cells are also restricted in the number of

divisions they undergo prior to senescence, a state in which the cells remain viable but no longer proliferative. Transformed cells overcome this restriction on cell cycles, enabling replicative immortality. Once a mass of transformed cells has developed, a means of sustenance and waste removal is required. For this purpose, angiogenesis, a process normally quiescent in adult cells, is induced to provide nutrient and oxygen access to the tumor as well as waste and CO_2 removal from the tumor. In the breast, atypical growths that have accumulated these qualities present as non-invasive in situ tumors[34]. Higher grade tumors acquire the additional ability to delocalize and invade neighboring tissues, a process facilitated by the reduction in expression of adhesion molecules. This occurs locally at first, then distally with the entrance and transit through the blood and lymphatic systems, providing the metastasized cancer access to more distant sites[35].

Mutator phenotype

As indicated by the outlined steps for the hallmarks of cancer, a sequential accrual of alterations in key genes is required for transformation to go to completion. However, based on the estimated 10^{-10} spontaneous mutations/nucleotide/division mutation rate, the number of alterations accumulated within a lifetime would fall short of the multiplicity of mutations typically observed in tumors. Hence, it has been suggested that the presence of a mutator phenotype may facilitate the process early on by increasing the inherent mutation rate[36]. Several potential candidates whose altered state could lead to a mutator phenotype have been suggested. Key among them are genes involved in maintaining the fidelity of the genome.

DNA repair as a candidate target for a mutator phenotype

Instructions for the essential functions of each cell are held within the genome through the specific arrangement of a four-letter code consisting of adenine, thymine, guanine and cytosine. As the blueprint of life, it is vital that these instructions are preserved and faithfully transmitted for the proper functioning and propagation of all living-things. These instructions are held within DNA molecules, consisting of two polynucleotide chains wound around each other in a double-stranded helix. The chains are held together by hydrogen bonds between the purine and pyrimidine side-chains of DNA, with adenines always matched with thymines and guanines always matched with cytosines. Strict

adherence to these base-pairing rules allows the sequence along one chain to be determined based on the sequence of the other. Through this redundancy in information, replication can progress with the separation of the duplex and use of the parental strands as templates to synthesize complimentary strands. Hence, the inherent structure of DNA facilitates the transmission of the genome during the numerous replication cycles cells undergo throughout a lifetime[37].

However, maintenance of this structure is compromised through various processes, including endogenous sources such as spontaneous deamination, cellular metabolism and replication error as well exogenous sources of oxidative stress, ionizing radiation and pollution[38–43]. To retain genomic integrity in the face of these various threats, organisms have evolved repair processes that recognize and remove such changes in the chemistry of DNA. Hence, mutations in these genes would tend towards greater overall genomic instability, increasing the baseline mutation rate and facilitating the accumulation of changes necessary to develop cancer.

Initial findings outlining the effects associated with deficiencies in DNA damage recovery stem from experiments conducted in mutant *E.Coli* strains[44–46]. These repair-deficient strains were shown to be hyper-responsive to environmental agents. Establishing a similar link within a human population, however, was complicated by the fact that such mutants are unlikely to be found in a natural setting due to the severity of the phenotype associated with such a defect. However, a few rare inherited human disorders do present with a heightened sensitivity to environmental agents. One such example is *Xeroderma Pigmentosum* (XP), a condition that results in an up to 1000-fold increase in risk of developing skin cancer due to sunlight exposure[47]. Studies have shown that while cells from normal and XP individuals start out with an equivalent number of lesions following UV-irradiation, a greater proportion of the lesions persist over time in cells from XP individuals compared to normal individuals, indicating that the observed sensitivity is due to a defect in damage repair[48].

Subsequent studies established that repair consisted of multiple components, and that different types of damages are targeted by specialized pathways[49,50,51]. Currently, four distinct, yet partially overlapping, types of repair pathways have been identified: Nucleotide excision repair, double-strand break repair, base excision repair and mismatch repair.

Nucleotide excision repair

Nucleotide excision repair (NER) is regarded as one of the most versatile types of the repair pathways, recognizing a wide range of lesions. Unlike other excision repair pathways, NER does not rely on the specificity of enzymes to uniquely recognize individual lesions, instead focusing on more general bulky distortions to the helix, enabling the versatility of the pathway. The major form of defense is directed against DNA lesions that result from UV light exposure, which induces covalent bonds between adjacent pyrimidines to form dimers. Also targeted are lesions due to chemicals that covalently bind DNA, forming adducts. Exposure routes for such chemicals include food-intake, as in the case of the mold-produced aflatoxin, as well as air inhalation, as in the case of smoke-related benzo(a)pyrene. In many of these instances, the parent chemical compound itself is innocuous. The toxic effects are, in fact, mediated by the intermediary metabolites produced during the process of eliminating these compounds from the system[52].

NER mediates detection and removal of such lesions via two processes: The global genome repair (GGR) arm of the pathway is initiated upon damage-detection during a genome-wide survey, whereas the transcription-coupled repair (TCR) arm of NER is initiated upon damage-detection during the specific interrogation of actively transcribed regions[53]. In GGR, XPC complexed with Rad23-centrin and CETN2 facilitates recognition of damage in conjunction with the UV-damaged DNA binding protein complex containing DDB1 and DDB2[54,55]. The DNA is unwound around the lesion by the XPB and XPD subunits of TFIIH[56], while XPA and RPA act to stabilize the ssDNA, forming a 27 nucleotide bubble[57,58]. The endonucleases XPG and ERCC1-XPF cleave the DNA 3' and 5' to the lesion, respectively, generating a 24-32bp fragment[59,60]. The ensuing gap is filled by PCNA-dependent DNA polymerases δ/ε[61–63], and the nick is sealed by DNA ligase III-XRCC1[64]. In contrast to GGR, TCR is believed to be initiated upon RNA polymerase II encountering a lesion during transcription. Stalling of the polymerase triggers the assembly of various factors, including CSB, CSA and TFIIS[65]. Following recognition, the pathway proceeds using the same components outlined above for the removal of damage in GGR[66].

Base excision repair

Minor base alterations, causing slight distortions to the helix, are primarily targeted by base excision repair (BER). Examples of such alterations include depurination, resulting in the formation of abasic sites, deamination of cytosine to uracil, alkylation of adenine to 3-methyl adenine, and oxidation of guanine to 8-oxoguanine. Agents capable of such changes are readily found within the cellular environment in the form of oxidative stress due to normal metabolic activity and the presence of endogenous *S*-adenosylmethionine (SAM) as an alkylating agent, but can also be elicited exogenously through ionizing radiation-induced oxidative stress and tobacco smoke-related alkylation [67]. The resulting modifications cause the replication machinery to stall at the lesions, triggering eventual apoptosis. If bypassed, however, many of these lesions can prove to be mutagenic, with the resemblance to alternate bases resulting in miscoding in subsequent replication cycles. This is especially the case for 8-oxoguanine, which as a thymine-mimic, can base-pair equally well with cytosines and adenines, leading to a potential GC:TA transversion.[68]

BER consists of a six-step process: damage recognition, removal of the distorted base, nicking of the DNA backbone, strand processing, polymerase refilling, and resealing by DNA ligase. Damage recognition is facilitated by a family of damage-specific DNA glycosylases. During this process, the DNA backbone is pinched, such that the nucleotide flips out and into the substrate-binding pocket of the glycosylase. Successful recognition catalyzes a series of reactions, starting with the glycosylase-initiated cleavage of the N-glycosyl bond between the base and DNA backbone, forming an apurinic/apyrimidinic (AP) site. Bifunctional DNA glycosylases have an intrinsic AP lyase activity, enabling them to nick the DNA backbone 3' to the AP site following removal of the base, while in the case of monofunctional DNA glycosylases , an AP endonuclease (APE) enzyme is recruited to incise the DNA backbone 5' to the AP site. In either case, the 3'-phosphodiesterase activity of the APE enzyme further processes the nicked strands to reveal the 3'-OH group necessary for subsequent polymerase β elongation of the strand with the appropriate nucleotide. DNA ligase is then recruited to seal the remaining nick. Subsequent to the initial incision, BER can progress along two possible pathways, short patch repair, involving the replacement of a single nucleotide, or long patch repair,

resulting in the re-filling of a widened 2-8 nucleotide gap and involving the recruitment of additional enzymes for processing .[69]

Mismatch repair

Mismatch repair (MMR) is initiated in the presence of base-pairing that contradicts the standard A-T/C-G Watson-Crick base-pair patterns (GA/CA/TC/TG/GG/CC/TT/AA). These can arise in instances of replication-induced errors, spontaneous deamination, as well as cellular metabolism-related oxidative stress. Replication-related errors occur through the misincorporation of a base, leading to a single base:base mismatch, as well as through a misalignment of the polymerase with the template, leading to insertion/deletion loops (IDLs). Mutations that arise as a consequence of such replication-related errors are kept to a minimum by the fidelity of DNA polymerases to incorporate correct nucleotides, which is reinforced by an inherent proofreading mechanism that enables detection and excision of any mispaired bases prior to the incorporation of subsequent nucleotides, thereby, ensuring maintenance of base-pair integrity. Still, in a genome consisting of $3x10^9$ bases, the resulting estimated error rate of 1 in every 10^7 bases translates to the introduction of hundreds of alterations with each replication cycle[70]. Further reduction of this error rate to 1 in every 10^9-10^{11}, one that is more tolerant to human cells, is realized downstream of the replication fork through the complimentary activities of the MMR system.

A key distinguishing characteristic of post-replicative MMR from other types of excision repair is that while the response is initiated by distortions in the helix, the mispaired constituents of the distortions are both chemically-normal bases. Hence, correctly identifying the base for removal adds a layer of complexity to MMR not faced by the other pathways. In replication-associated errors, the key lies in discriminating the newly synthesized strand from the parent template. How this is accomplished in eukaryotic cells remains to be confirmed, however, the presence of strand discontinuities in the nascent strand may play a role[71].

The presence of chemically-modified bases is an additional source of mispairing. Such modifications can either result in altering the base to a form that favors alternative binding partners from the one prescribed by traditional Watson-Crick base-pairing or converting the base into another naturally occurring base, maintaining Watson-Crick base-pairing rules in subsequent replication cycles, albeit

different from the ones prescribed for in the original code. Alkylating agents, for example, can induce methylation of guanines to form O6-methyl guanine, which binds equally well to dcAMP and dtAMP, and oxidation of guanines generates 8-oxoguanines, which favor binding daAMP over the originally dictated dcAMP. Meanwhile, 5-methyl cytosines base-paired with guanines can spontaneously deaminate to generate thymines, altering the original GC basepair to a TA basepair in subsequent replication cycles.

Chemical modifications can occur in bases already incorporated into the genome, as well as to nucleotides in the dNTP pool. Hence such modifications threaten the integrity of the genome on two separate fronts: through the incorporation of altered bases during replication as well as subsequent alteration of correctly incorporated nucleotides following replication.

Depurination and depyrimidation present special case scenarios for generating mismatches. The resulting non-instructive abasic sites cause DNA polymerases to stall during replication. In such instances, random nucleotides are incorporated and tolerated to avoid collapse of the replication fork.

There are various components involved in the pathway of mismatch repair. MMR is initiated through the recognition of the mismatch by one of two heterodimers. The MSH2:MSH6 heterodimer preferentially targets base:base mismatches and short IDLs while the MSH2:MSH3 heterodimer identifies longer IDLs. Binding of either heterodimer recruits the MLH1:PMS2 heterodimer. The ATP-driven conformational change allows this heterodimer to slide along the strand away from the mismatch. PMS2, in the presence of RPA, PCNA and ATP, introduces nicks in the mismatch-containing strand, with the generated fragment degraded by EXO1. The gap is filled by DNA polδ and sealed by DNA ligase I[71].

Double-strand break repair

Double-strand breaks present as both a necessary means of genetic diversification, during meiosis[72] and V(D)R recombination[73], as well as the most lethal form of DNA damage, in which the information lost cannot be recovered from the same molecule, as with other types of damages[74]. Two major pathways, homologous repair (HR) and non-homologous end-joining (NHEJ) repair, remediate double-strand breaks. The type of response invoked is dictated by the presentation of the damage. A break

may be encountered during replication, at which time the intact sister chromatid will likely be within the vicinity of the replication fork, increasing the feasibility of homologous recombination[75]. A homologous template, however, may not be as readily accessible for the repair of double-strand breaks generated during other stages of the cell cycle. Direct induction of such breaks, incurred through such sources as ionizing radiation and mechanical stress[76], generally produce two ends in close proximity to one another, which can be rejoined without the reliance on a repair template. Although sequence information can be lost in this process, this loss is tolerated over the potentially more grievous deleterious effects of further damage and instability resulting from the genomic rearrangements of unattended broken ends[74].

NHEJ is initiated by the binding of the ku7/ku80 heterodimer to the termini[77]. Once the ends have been aligned, the heterodimer recruits DNA-dependent protein kinase catalytic subunit (DNA-PKcs). Upon binding, the kinase is activated[78]. If the ends are directly ligatable, the XRCC4-DNA ligase IV complex is recruited to join the ends[79]. However, various different ends can be produced, as is the case with ionizing radiation where reactive oxygen species (ROS) can lead to additional base and sugar damage[76], and the necessary 3'-hydroxyl and 5'-phosphate groups may not be readily available for ligation. In such instances, an intervening processing step prior to ligation is included, in which endonucleases such as Artemis are recruited to produce the required ligation substrates[80] . Any remaining single-stranded gaps are filled in by polymerases λ and μ and terminal deoxynucleotidyl transferase (TdT), which can perform template-independent nucleotide extensions[81].

In HR, broken ends are first detected upon interaction with the Mre11-Rad50-NBS (MRN) complex. Upon contact with the DNA, this complex bridges and initiates processing of the termini into 3' ssDNA overhangs through 5'-end resection[82,83]. It also recruits and activates the damage signal transducer ATM, which then phosphorylates several downstream components of the pathway[84]. RPA binds to the single-stranded overhangs eliminating secondary structures at the termini[85]. Subsequently, a BRCA1/BRCA2/PALB2 containing complex facilitates loading of the Rad51 recombinase protein[86], displacing RPA and enabling the ends to invade the nearby sister-chromatid[87,88]. The displacement of one strand and pairing with the other forms the heterduplex displacement loop (d-loop). Once a

homologous region has been found, the invading end can be used as a primer for polymerase synthesis. Through this strand exchange, the lost information of the damaged strand can be recovered from the undamaged strand.

DNA repair and cancer

In addition to XP, several other genetic syndromes manifest in individuals with heightened sensitivities to environmental agents. Individuals with ataxia telangiectasia are sensitive to x-ray exposures and present with progressive ataxia, dilated vessels of the eye and immune deficiencies. Individuals with Fanconi's anemia are sensitive to DNA cross-linking agents and present with hematological abnormalities. Despite heterogeneity in clinical symptoms, all three are linked by the fact that they result from defects in factors involved in DNA repair and are all cancer-prone, establishing the carcinogenic potential of defective DNA repair[89]. The rarity of such genetic syndromes, however, limits the population-wide significance of the findings from such studies. Still, less severe deficiencies in components related to DNA damage repair may be more prevalent in the population and are likely to have a more public health-relevant impact on the risk of various cancer sub-types, including breast cancer.

Assays determining DNA repair

The ability to assess inter-individual variations in repair is limited by the technology available to detect such differences. Early studies linking genetic defects in repair with carcinogenesis focused on gross abnormalities resulting in the complete loss of function of genes involved in repair processes. Once the association was established, more refined analyses could be conducted to identify factors with more subtle influences on repair. The genetic defects leading to the aforementioned genetic syndromes typically arise due to missense substitutions and frameshift deletions in critical regions of the genes, often resulting in premature truncation[90,91]. These alterations could be pin-pointed based on genetic analyses comparing individuals with clear-cut presentations of phenotypic defects and apparently normal individuals. Alterations in genetic sequences resulting in a more attenuated reduction in repair-related activity are less likely to have a readily apparent phenotype associated with it, making this approach of identifying relevant mutations less feasible in such situations. However, with the aid of

technological advancements in sequencing, additional variants have now been mapped. These single nucleotide polymorphisms (SNPs), occurring at greater frequencies but with less apparent phenotypic implications, can now be evaluated for health effects.

The genotyping assays utilized to evaluate SNPs can be carried out in a high-throughput format, hence, utilizing this approach serves as an attractive option to determine repair capacity within a population-based setting. Consequently, a number of studies have attempted to identify polymorphisms that may account for inter-individual differences in repair. However, few relevant SNPs have thus far been identified, often with borderline associations. This is likely due to the complex interplay between various gene products, introducing the difficulty of extrapolating final phenotype from the given genotype.

An alternative to focusing on the genotypic level rests on the use of phenotypic assays to query variations at the level of protein function, integrating interacting pathways and enzymes, and thereby, providing a more comprehensive determination of repair capacity[92]. Indeed associations observed in studies utilizing phenotyping assays are generally much more robust than what is typically observed in SNP studies. However, currently available phenotyping methods, including the host cell reactivation (HCR) assay, are time-consuming and tedious to carry out, limiting their potential to be applied in a population-based setting.

Factors impacting repair

Both genotyping and phenotyping methods have been implemented to determine inter-individual variability in repair. The presence of such variability has been observed even within apparently healthy populations. In a study conducted among 102 non-Hispanic white control individuals enrolled in a case-control study, a 4-fold variation in the ability to repair UV-induced damage, a substrate targeted by NER, was determined using HCR[93]. This variability has been observed in other repair pathways as well. In an antioxidant intervention study, 48 healthy adults between the ages of 18-30 were enrolled to study the impact on BER capacity following micronutrient supplementation. BER was measured using a modified COMET assay, in which lysed and plated HeLa cells were treated with UV light and incubated with lymphocyte extracts from the participants. Differences in the amount of single-strand

breaks, as determined by the COMET assay, between cells incubated with and without extracts were used as indicators of BER capacity. Baseline measurements using this approach showed considerable amount of variation, with the extent of tail DNA, a measure of DNA damage, ranging from 0.6-25%[94].

Given the presence of variability in repair capacity, a number of studies have sought to identify factors involved in the modulation of the repair response within the general population. The impact on NER capacity due to the presence of polymorphisms in NER-relevant genes was assessed among 102 healthy individuals. Homozygosity in several of the tested SNPs led to a reduction in repair capacity as determined by the HCR assay, reaching statistical significance for XPC[poly(AT) intron 9][95]. The potential impact of polymorphisms in these genes on repair capacity is further substantiated by the higher adducts levels observed among individuals with SNP variants in NER-related genes[96,97].

Several studies have indicated an association between increasing age and decreasing DNA repair capacity, possibly suggesting a means by which cancer incidence increases with age. The existence of a link between DNA repair and aging is most clearly seen in the DNA repair related genetic syndromes Ataxia Telangiectasia and Cockayne Syndrome, both of which include premature aging as a characteristic clinical feature among affected individuals. Another means by which aging and DNA repair have been associated with one another is through the free radical theory of aging, which suggests that manifestations of aging are related to the accumulation of damage due to oxidative stress. Reductions in repair capacity could overwhelm the system with the challenge imposed by free radicals, enabling the accumulation of DNA damage over time to account for the overall decline associated with aging. Such a reduction in repair with time has indeed been shown in a case-control study investigating risk for basal cell carcinoma, where a nearly 25% decline in NER capacity was noted among 60-year old controls compared to 20-year old controls[98].

In addition to these intrinsic factors, several modifiable factors have also been identified. Since nutritional status has a putative role in cancer risk, studies have sought to assess whether dietary factors also modulate DNA repair capacity. To date, there have been varied reports of stimulatory, inhibitory and no effects on repair, depending on the pathway and dietary factor tested. However, in one of the largest studies addressing this question, 559 non-Hispanic white cancer-free individuals

from the Houston metropolitan area were analyzed for the impact of folate status, derived from food frequency questionnaire data. DNA repair capacity in these same individuals was assessed using the host cell reactivation assay. In this study, an 18% reduction in repair capacity was observed among individuals in the lowest tertile of folate levels compared to individuals in the highest tertile[99].

While the increase in oxidative stress resulting from exercise-induced aerobic metabolism is well-recognized, few studies have addressed whether the accumulation of oxidative stress-related damage as a result of physical activity also has implications on DNA repair capacity. In a study conducted among marathon runners, the activity of OGG1 was increased in skeletal muscle following a marathon race, suggesting an adaptative response of increased 8-oxoG excision to meet the challenge imposed by the increased presence of free radicals[100]. However, the study relied on measurements in only 6 subjects, limiting the conclusiveness of the data.

Finally, exposures to various environmental agents have also been shown to mediate an effect on DNA repair. In a lung cancer case-control study, current smokers were observed to have the highest DNA repair capacity among both cases and controls, suggesting that exposure to damage-inducing environmental agents may induce repair activity[101].

Defects in DNA repair pathways have implications for breast cancer

Using these approaches, a number of studies have established links between deficiencies in DNA repair capacity and breast cancer risk. Most notably, the major breast cancer susceptibility genes *BRCA1* and *BRCA2* are both associated with DSBR. In fact, these two genes are implicated in facilitating the HR branch of the pathway. Interestingly, no such high-penetrant effects have been identified for genes involved primarily in the NHEJ branch of DSBR, although a modest increase in risk associated with several SNP variants located in NHEJ genes have been observed[102].

Phenotypic assessments of repair have also demonstrated a role for double-strand breaks in breast cancer risk. Several studies have shown a greater accrual of damage among breast cancer cases compared to controls following exposure to DSB-inducing agents. In one such study, the number of chromatid breaks in peripheral blood lymphocytes immediately following X-irradiation was similar between breast cancer cases and controls. However, following a recovery period, the number of

breaks detected in the controls diminished over time, while a greater number of breaks were retained in the cases, indicating a deficiency in the ability to repair such damage[103]. Additionally, a number of case-control studies assessing overall EJ capacity using an *in vitro* functional assay have observed an up to 3-fold increase in breast cancer risk among women with the poorest EJ capacity[104,105].

The major hallmark shared by all NER-related genetic syndromes[106,107], which in addition to XP include Cockayne syndrome and trichothiol dystrophy[106], is photosensitivity. While the resulting predisposition to skin cancer due to NER-related deficiencies has been well-established, associations with additional cancer subtypes, including breast cancer, have been observed as well. A study conducted in Puerto Rico measured NER capacity among breast cancer cases and controls using the HCR assay and reported a 36% reduction in repair capacity ($p<0.001$) among the cases compared to controls[108]. Similarly, a study was conducted within discordant sister-sets enrolled in the New York site of the Breast Cancer Registry which assessed NER activity using an immunohistochemical method to measure the removal capacity of benzo(a)pyrene diol epoxide (BPDE)-DNA adducts, a substrate for NER. This study also reported a significant reduction in repair capacity among cases compared to their control sisters, with a more than two-fold increase in breast cancer risk among subjects with repair levels below the median of the controls[109].

The earliest indications linking MMR to carcinogenesis were recognized in studies focusing on patients with Lynch syndrome, a predisposition towards hereditary nonpolyposis colorectal cancer (HNPCC) as well as several other extracolonic cancers including ovarian, endometrial and gastric cancer[110]. These studies frequently observed a particular microsatellite phenotype, with the short repetitive motifs distributed throughout the genome often varying in lengths in the tumors of these patients[111]. This type of replication error occurs as a result of polymerase slippage, with the primer and strand dissociating from one another, only to reassociate with one another misaligned, leading to potential insertions or deletions of repeats in the nascent strand. The resulting errors observed within such sites may themselves promote carcinogenesis, if located within critical genes whose function are disrupted by the altered alignment, or may be indicative of a general genome-wide instability. Hence, while such errors may constitute part of a normal phenomenon during replication of these sites, such mistakes are likely

readily corrected with a mechanism that seems to be impaired in these individuals. This mechanism was subsequently identified in linkage analysis studies of tumors derived from Lynch family patients which pinpointed the phenotype to three particular loci in the genome, later identified as *MLH1, MSH2* and *PMS2*, key components of MMR[112, 113,114]. It has now been established that cancers arising in this particular population arise as a consequence of a genetic predisposition leading to loss of function in genes critical to MMR function.

While cancer susceptibility due to failures in MMR has been most strongly associated with Lynch syndrome related cancers, there are some indications that deficiencies of this repair pathway may also have implications for breast cancer. The greatest weight of evidence is derived from studies observing that a substantial proportion of analyzed breast tumors present with microsatellite instability (MSI)[115]. However, unlike Lynch syndrome-associated cancers, few studies have been able to implicate a genetic predisposition due to defects in MMR genes to breast cancer susceptibility.

Unlike other DNA repair pathways, genes involved in BER have, thus far, not been implicated with a genetic susceptibility to cancer. While an increased mutator phenotype has been observed in bacterial and yeast mutants, mice with BER targeted knock out genes present with a near normal phenotype. This has been attributed to the likely redundancy of the various enzymes, with the overlap in functionality minimizing the effect of a deficiency in any one single component.[69] Still, a few epidemiological studies do indicate a potential role for BER deficiency and human carcinogenesis. In a case-control study conducted in Israel, individuals in the lowest tertile of enzymatic activity of OGG1, the DNA glycosylase responsible for targeting oxidative stress-induced 8-oxoguanines, were found to be at an increased risk of lung cancer.[116]

Significance of study

1. To improve risk assessment by better identifying individuals with suboptimal repair, thereby enhancing the targeted screening of susceptible individuals
2. To provide a more thorough understanding of the means by which deficiencies in DNA repair bear relevance on breast cancer susceptibility, facilitating identification of interventional endpoints.

DNA repair mechanisms are vital to maintain the integrity of the genome in the face of continual exposure to damaging agents in the exogenous as well as endogenous environment. Reduced repair capacity compromises the stability of the genome, allowing incurred alterations to persist. As these changes serve as potential initiators of carcinogenesis, DNA repair capacity can serve as a marker of breast cancer susceptibility.

In fact, it has already been shown that women at high risk due to a family history of breast cancer have five times the likelihood of suboptimal DNA repair compared to women with minimal family history of breast cancer[117]. These preliminary observations call for a need to establish high-throughput functional assays capable of assessing repair as a means to screen individuals with an increased risk of cancer due to a heightened sensitivity to carcinogenic exposure.

Current phenotypic assays measuring DNA repair are labor intensive and not applicable to large scale population studies. To address this need, we intend to develop a NER capacity assay using a competitive ELISA-based method to determine the *in vitro* removal of DNA adducts induced by BPDE, a metabolite of the polycyclic aromatic hydrocarbon benzo(a)pyrene, commonly found in cigarette smoke, exhaust fumes and char-boiled food[118].

Various groups, including our own, have already reported findings implicating suboptimal repair capacity, as indicated by mutational analysis, genotyping studies, epigenetic studies and gene expression studies of specific DNA repair-related endpoints, as well as phenotyping assays indicative of overall repair capacity, with breast cancer susceptibility. However, how alterations in these individual components integrate into the overall observed reduction in pathway activity, and how all of these factors ultimately translate into the observed increase in breast cancer risk, has not yet been

established. For this purpose, select intervening factors that may explain currently available data implicating DNA repair and breast cancer risk were selected for further analysis. Specifically, the presence of SNP variants and gene expression and promoter methylation levels of DNA-relevant genes were queried for an association with observed repair capacity and breast cancer case status (Figure 1).

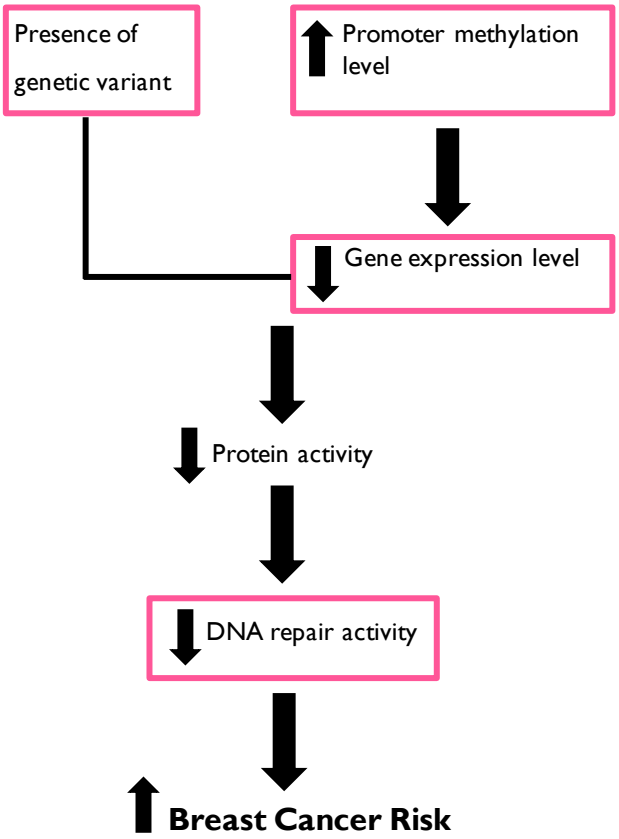


Figure 1. Overall study hypothesis
The undertaken studies focus on assessing the risk of breast cancer risk due to a compromise in DNA repair activity as determined by the indicated endpoints

All studies were carried out among participants enrolled in the New York site of the Breast Cancer Family Registry, one of six sites established to further understanding of the genetic epidemiology of breast cancer. Participants enrolled in the registry had to meet one or more of the following eligibility criteria: 1) A female relative who had been diagnosed with either breast or ovarian cancer prior to the age of 45; 2) A female relative who has been diagnosed with breast and ovarian cancer at any age; 3)

Two or more female relatives who had been diagnosed with breast or ovarian cancer after the age of 45; 4) A male relative diagnosed with breast cancer at any age; 5) A known carrier of *BRCA1* or *2* mutation. Epidemiologic risk factors, family history of cancer and food frequency information were determined through questionnaires, which gathered information on demographics, ethnicity, history of all cancers, smoking, alcohol consumption, reproductive history, hormone use, height, weight, physical activity, and dietary intake. Blood was collected at the time of recruitment, on average 5 years after diagnosis for cases[127].

We conducted familial based case-control studies, selecting sisters discordant for breast cancer among the families participating in the New York site of the registry. For each study, we selected the subset of discordant sister-sets for whom relevant biospecimens were available (Figure 1).

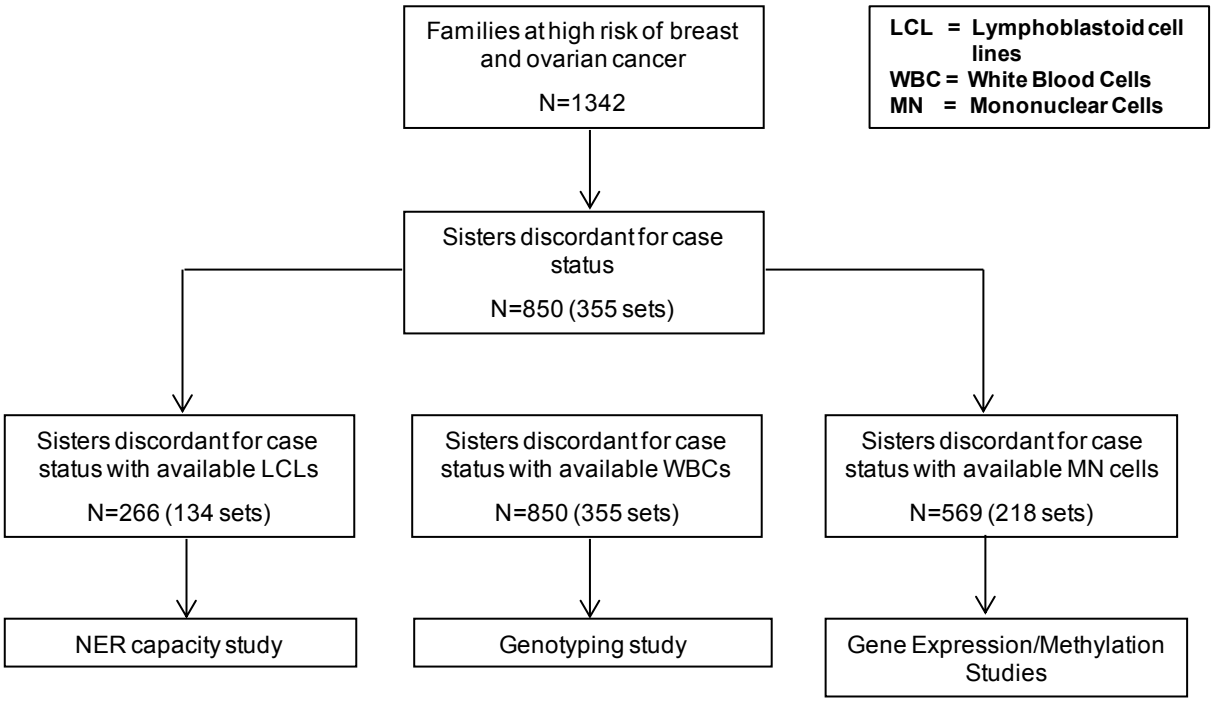


Figure 2. Overall study design
Each familial-based case control study was conducted among sisters discordant for disease with available biospecimens relevant for each study.

Hypothesis 1: Breast cancer cases are more likely to have deficient NER capacity than their unaffected sisters

Specific Aim 1a: Develop a competitive ELISA applicable in a population-wide setting with sufficient sensitivity and specificity to evaluate inter-individual differences in NER capacity using nuclear extracts derived from relatively NER proficient and deficient lymphoblastoid cell lines

Specific Aim 1b: Apply the developed ELISA method to compare repair distribution between cases and controls using nuclear extracts of lymphoblastoid cell lines derived from sisters discordant for breast cancer (N=266) who are participants in the Breast Cancer Family Registry, most of whom were previously assayed for NER capacity using an immunohistochemical method

Hypothesis 2: Breast cancer cases are more likely to present with variants in DNA repair-relevant genes than their unaffected sisters.

Specific Aim 2a: Assess single nucleotide polymorphisms in DNA repair-relevant genes and determine the association with case-status

Specific Aim 2b: Assess the impact of extended family history on the association between the presence of variants in DNA repair-relevant genes and breast cancer risk

Hypothesis 3: Breast cancer cases are more likely to have reduced expression levels of DNA repair-relevant genes than their unaffected sisters

Specific Aim 3a: Assess expression levels of DNA repair-relevant genes and determine the association with case-status

Specific Aim 3b: Assess the impact of extended family history on the association between the expression levels of in DNA repair-relevant genes and breast cancer risk

Hypothesis 4: Breast cancer cases are more likely to have increased promoter methylation levels in DNA repair-relevant genes than their unaffected sisters

Specific Aim 4a: Assess promoter methylation levels of DNA repair-relevant genes and determine the association with case-status

Specific Aim 4b: Determine the correlation between promoter methylation levels and gene expression levels of DNA repair-relevant genes for which information is available for both in the sister-sets

Chapter II. Phenotypic assessment of NER capacity

Abstract

Several studies have indicated an association between suboptimal NER capacity and breast cancer risk. While these findings have implications for early detection, validating these initial studies and developing a potential screening tool assessing NER capacity in a population-wide setting is limited by the technology available to measure DNA repair. Current phenotypic assays are labor intensive and not applicable for large scale population studies. To address this need, we have developed an NER capacity assay using a competitive ELISA-based method to determine the *in vitro* removal of benzo(a)pyrene diol epoxide (BPDE)-DNA adducts. This method was subsequently applied in a family-based case-control study among sisters recruited in the New York site of the Breast Cancer Family Registry (NY-BCFR) previously assessed for NER capacity using the immunohistochemical method. Unlike previous studies, the current study did not detect an association between NER activity and breast cancer risk (OR = 1.01, 95%CI=0.97,1.05). The findings of the current study also do not correlate with the findings within the same population based on the immunohistochemical method (r=-0.01, p=0.86).

The previously suggested role of suboptimal NER capacity in breast cancer susceptibility and the discrepancy in findings among the various investigations highlight the need for additional studies to validate the implied association. Furthermore, since currently available methodologies to assess overall NER capacity are not designed to be applied in a population-wide setting for the necessary validation or subsequent development as a screening tool, other avenues, potentially focused on more specific endpoints, will have to be explored.

Introduction

The NER pathway targets bulky, helix-distorting DNA lesions arising from environmental exposures including UV radiation and tobacco smoke[119]. Studies investigating the impact of NER on cancer risk have mainly focused on skin and lung cancer as outcomes of interest since exposures known to induce NER-targeted lesions are also typically associated with these cancer sub-types[120]. Though not as strongly associated, several studies have indicated that DNA damage arising from these types of exposures may also be relevant for breast cancer. Of particular note, DNA adducts have been shown to present at higher levels in breast tissue derived from breast cancer cases compared to breast tissue derived from non-cancer controls[121,122]. The adduct types observed in these studies are indicative of exposure to polycyclic aromatic hydrocarbons (PAH), a class of chemical carcinogens contained within cigarette smoke, incomplete combustion of fossil fuel, and charred meats[123].

The link between adduct levels and risk of breast carcinogenesis was further established in several *in vitro* studies. Human breast epithelial cells treated with benzo(a)pyrene (B(a)P), a type of PAH, undergo changes in cell morphology, increased growth rate, and exhibit anchorage-independent growth and invasiveness in agar-methocel. Subsequent inoculation of clones isolated from agar in SCID mice led to tumor formation, indicating the tumorogenic potential of the transformed cells[124].

This transformation process is triggered through the incorporation of incorrectly specified bases opposite helix-distorting lesions, such as DNA adducts, during replication. Failure of the NER pathway to remove such lesions, especially if present in key genes related to cell cycle progression and proliferation, can lead to persistent mutations that trigger carcinogenesis[125]. Hence, the observed link between breast cancer susceptibility and adduct persistence may be related to inherent deficiencies in NER capacity to repair damage resulting from environmental exposures.

The ability to assess the relationship between NER and breast cancer susceptibility is limited by the technology available to detect inter-individual differences in repair. One approach focuses on the genotypic level, querying for the presence of SNPs within regions in specific DNA repair genes. While genotyping assays can be carried out in a high-throughput format, studies assessing the impact of DNA repair SNPs on breast cancer risk have identified few relevant SNPs thus far, often with borderline

associations. Associations observed in studies utilizing phenotyping assays are generally much more robust than what is typically observed in SNP studies, likely due to the more comprehensive measure of pathway activity captured by such assays.

A few studies have reported on the risk of breast cancer due to suboptimal NER capacity utilizing phenotyping assays. In a retrospective case-control study conducted in Puerto Rico, a 36% reduction in NER capacity, as determined by the HCR assay, was observed among breast cancer cases compared to cancer-free controls[108].

Our own group has developed an assay to measure NER activity using an immunologic-based approach. In a study conducted in sister-pairs discordant for breast cancer enrolled in the NY-BCFR, this method was implemented to demonstrate a significantly lower capacity to repair NER-specific damage among breast cancer cases compared to their control sisters, with a more than two-fold increase in risk among subjects with repair levels below the median of controls[109].

While findings based on both methodologies indicate that reduced NER capacity may be a potential risk factor for breast cancer, the small sample size and retrospective nature of these studies necessitate further inquiries in larger prospective cohorts to substantiate the results. However, as both these methods are time-consuming and tedious to carry out, utilizing them in such settings would not be feasible.

Additionally, given a true association between NER capacity and breast cancer risk, an assay capable of detecting such a risk factor would be instrumental as a potential screening tool. Availability and implementation of such a tool could enable the reduction in breast cancer incidence by focusing prevention strategies to relevant populations and improving prognosis through early detection. However, the aforementioned characteristics of currently existing phenotyping assays would impede translating them for use within this desired context.

In this study, we set out to develop a high-throughput competitive ELISA method capable of identifying susceptible individuals with suboptimal NER capacity, as indicated by the ability to remove PAH-induced BPDE-DNA adducts[126] (Figure 1). Following optimization, we conducted a case-control study among sisters discordant for breast cancer case status, previously assayed for NER capacity using the

immunohistochemical method, using the newly developed competitive ELISA assay to further characterize the parameters of the assay.

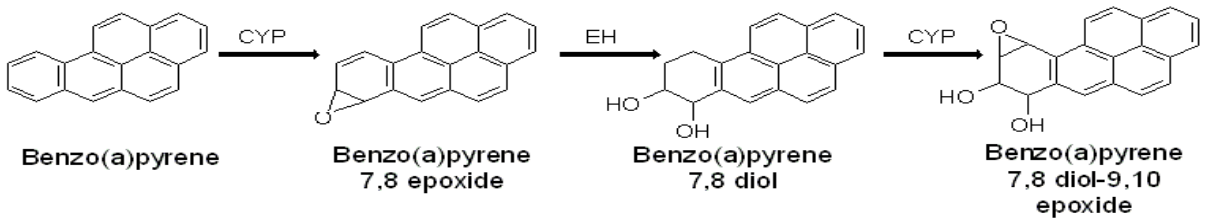


Figure 1. Formation of Benzo(a)pyrene diol epoxide
Carcinogenic effect associated with PAH exposure is mediated by Benzo(a)pyrene diol epoxide (BPDE), the intermediate by-product of B(a)P. Metabolic activation occurs through a series of steps following the internalization of B(a)P, commencing with the initial epoxidation by cytochrome P450 enzymes to generate B(a)P-7,8-epoxide. Epoxide hydrolase converts this intermediate into B(a)P-7,8 diol. A second epoxidation step produces the final carcinogen B(a)P diol epoxide (BPDE), which binds DNA at guanine residues to produce BPDE-DNA adducts.

Materials and Methods

Lymphoblastoid cell culture

Lymphoblastoid cell lines were thawed from -140°C and spun down. After removal of the DMSO-containing freezing medium, 10% of the cells were re-suspended in fresh medium containing RPMI 1640 (Fisher Scientific, Pittsburgh, PA), 1% Penicillin/Streptomyocin (Mediatech, Manassas, VA), and 15% fetal bovine serum (Atlanta Biologicals, Lawrenceville, GA) and placed in a 37°C incubator. Samples were deemed ready for extraction once a confluence of at least $8x10^6$ cells was reached.

Nuclear protein extraction

Cells were spun down and re-suspended in 1mL PBS. Nuclear protein was extracted from the cells using an NE-PER nuclear and cytoplasmic extraction kit (Fisher Scientific, Pittsburgh, PA). Following extraction, the samples were buffer exchanged into NE buffer (10mM Tris-HCl, 200mM KCl, 1mM EDTA) using Zeba spin columns (Fisher Scientific, Pittsburgh, PA). Protein concentration was determined using a Bradford Protein Assay kit (Bio-Rad Laboratories, Hercules, CA, USA), with aliquots of samples diluted 1:8 in water read at 595nm absorbance against a bovine serum albumin (BSA) based standard curve ranging from 7.5ug/mL to 75ug/mL. All samples were aliquoted at a final concentration of 0.75ug/uL in 20% glycerol and stored at -80°C.

BPDE modified-DNA substrate generation

0.05mg/mL diolepoxide in ethanol was mixed with 1mg of calf-thymus DNA (0.5mg/mL TE) in a 1:2 volume ratio. The reaction was left to incubate in the dark at 37°C for 1 hr. To remove any remaining unreacted BPDE derivatives, the modified DNA was extracted 5X using equal volumes of TE-saturated ethyl acetate (1 part TE:8 parts ethyl acetate). Finally, the DNA was precipitated by adding 3M sodium acetate (pH 5.0) to a final concentration of 0.3M and 3 volumes 100% ethanol. Precipitation was repeated until the A350/A260 ratio reached equilibrium. The concentration of the total DNA amount extracted was calculated using the following formula:

[DNA, umol/ul]=(A260-0.2(A350))/coeff ext A260

The amount of modified guanine present in the extracted sample was calculated using the following formula:

[BP-G, umol/ul]=A350/coeff ext. A350

The percent modification of the generated substrate ([BP-G]/[DNA]*100) was determined to be 0.95%.

BPDE adduct removal activity assay

Each repair reaction mixture consisted of 200ng nuclear extract and 250fmol sonicated BPDE modified DNA in a 100uL volume of NE buffer. Following a 1hr incubation at 37°C, the reaction mixtures, as well as a set of standard curve dilutions in the range of 0-250fmols of sonicated BPDE-modified DNA, were heat denatured at 95°C for 15min and then cooled on ice.

BPDE-adduct removal competitive ELISA

The wells of UV transparent flat bottom plates (Corning, Tewksbury, MA) were coated with 5ng BPDE-DNA/50ul PBS/well. The plates were dried down in an overnight incubation at 37°C and stored the following day at -20°C until time of assay.

Prior to use, plates were thawed to room temperature and rinsed with wash buffer (1X PBS, 0.05% Tween, 0.2% NaN_3) using an ELx405 washer (BioTek, Winooski, VT). The plates were then blocked using blocking buffer (1X PBS, 0.05% Tween, 1% fetal calf serum (FCS)) at 200uL/well and incubated at 37°C for 1 hr. During this blocking period, the extract repair reactions were set up. A 1:1000 dilution of the in-house generated monoclonal antibody 5D11 was prepared and added to the inactivated reaction mixtures and standard curve in a 1:1 ratio for a final 1:2000 antibody concentration. Following the blocking incubation and washing of the plates, the samples, in the presence of 5D11, were added to the plates in duplicate at 100uL/well. Following a 1.5 hr incubation at 37°C, the plates were washed and secondary goat anti-mouse IgG alkaline phosphatase (Sigma-Aldrich, St. Louis, MO) was added in a 1:1000 dilution in blocking buffer at 100uL/well. Following a subsequent 1.5 hour incubation at 37°C, the plates were washed and rinsed with 1% diethanolamine. p-nitrophenyl phosphate substrate was added to the plates at 100uL/well. After a final incubation at 37°C for 60min, the plates were read at 450nm using a SpectraMax Plus UV spectrophotometer (Molecular Devices, Sunnyvale, CA). The amount of BPDE-adduct remaining in the wells following the repair reaction was quantitated by referencing the optical density (OD) values from the samples against the values generated by the standard curve. Each plate was set up to include duplicates of each sample, zero-antibody

background control wells to which only NE Buffer was added, and an internal control sample run on each plate to monitor for batch effects. Five percent of the samples were repeated during each run to assess intra-assay variability.

Study subjects

Subjects consisted of 114 sister-sets (N=246) discordant for breast cancer enrolled in the NY-BCFR previously assayed for NER capacity using the immunohistochemical method (Figure 2).

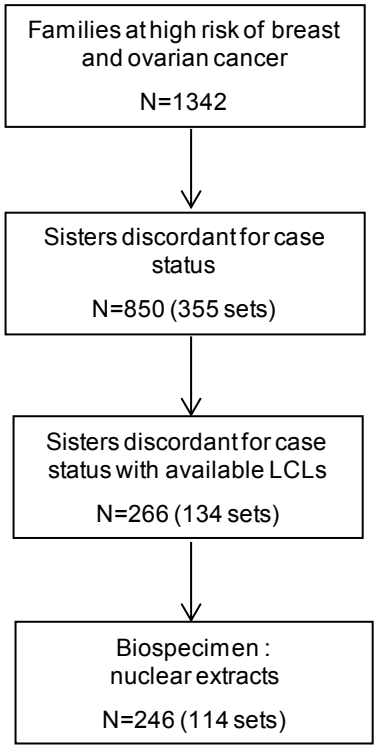


Figure 2. NER capacity study design
The study was conducted among sisters discordant for disease with nuclear extracts available from LCLs.

Statistical analysis

Chi-Square tests were performed to compare distributions of demographic variables between cases and controls. Linear regression was conducted to assess the impact of known breast cancer risk factors on adduct removal capacity. Student's t-test was performed to compare mean case-control differences in BPDE-DNA removal capacity. Conditional logistic regression analysis was conducted,

adjusting for age at blood draw and smoking status, and odds ratios were calculated using maximum likelihood methods. Adduct removal capacity was analyzed as a continuous variable, with change in risk assessed due to 1% decrease in adduct removal capacity, as well as a categorical variable, with cut-offs based on tertiles within unaffected sisters. Spearman correlation was calculated to determine the concordance between findings based on the competitive ELISA and immunohistochemical methods. All tests were performed using R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing (2011), Vienna, Austria.

Results

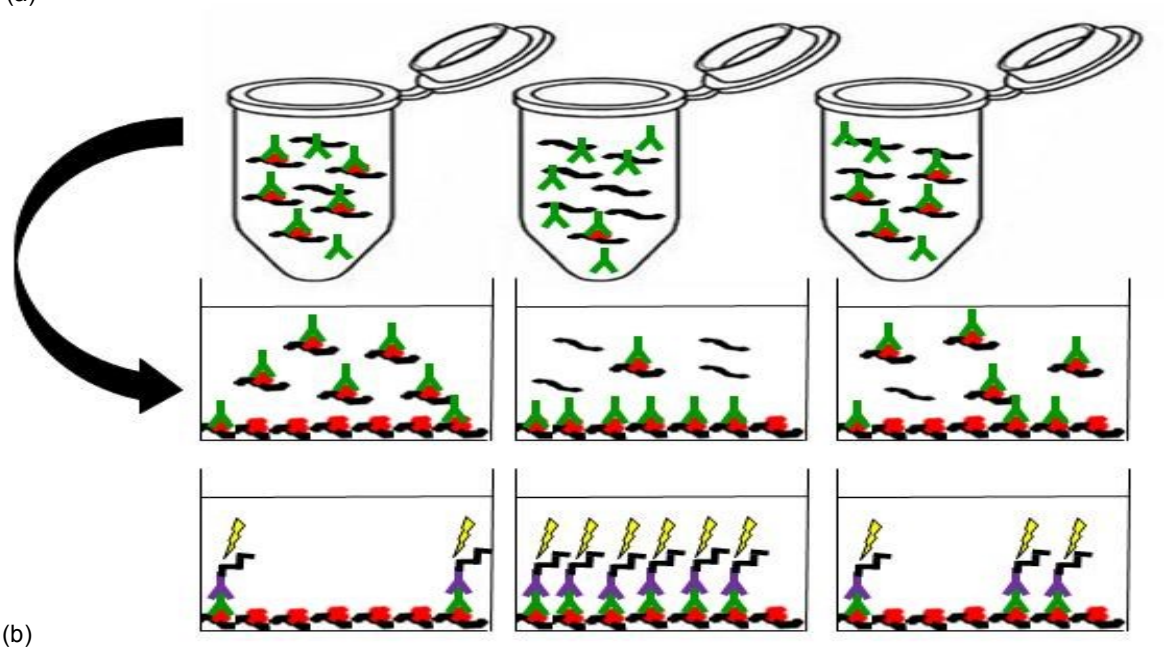
Principles of the competitive ELISA-based method to determine the *in vitro* removal of BPDE-induced DNA adducts

The paradigm of the developed assay rests on detecting the enzymatic activity of the host's repair machinery in the presence of its target substrate. The development of the methodology involves optimizations on two fronts: a) mimicking the *in vivo* presentation of damage to induce a response by the NER machinery and b) detecting the level of the subsequent response. To induce NER activity, BPDE- DNA adducts are exposed to nuclear proteins containing the relevant repair enzymes. Differences in the levels of adduct removal, indicative of repair capacity, is then determined through a competitive ELISA. During this detection phase, 5D11, a BPDE- DNA adduct specific antibody previously established by our group, is added to a BPDE-DNA adduct coated micro-titer plate in the presence of the deactivated repair reaction mixture. In this context, unrepaired BPDE-DNA adducts remaining in the reaction mixture compete with BPDE-DNA adducts bound to the micro-titer plate antibody binding. Successful micro-titer plate antibody binding can then be captured by the subsequent tethering of conjugated secondary antibody through the interaction with its alkaline phosphatase substrate, resulting in a colorimetric reaction. Referenced against wells in which the reaction mixture contains no extract or competitor, high antibody detection indicates less inhibition by the competitor BPDE-DNA adduct present in the reaction mixture, which in turn, indicates high BPDE-adduct removal activity in the presence of nuclear extract. Running a standard curve of BPDE-DNA adduct competitor alongside the reactions during the ELISA provides a means to quantify level of adduct removal capacity. Using the standard curve, the amount of adduct level present in the wells following the reaction can be calculated based on the level of observed antibody binding. Given the known input amount of BPDE-DNA adduct substrate added to each reaction, the level of adduct removal, indicative of NER capacity, can be determined (Figure 3).

Development of protocol

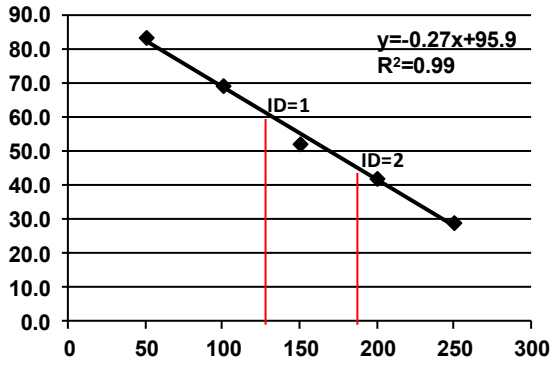
The initial set of experiments focus on optimizing the assay conditions. Setting up the ELISA detection system includes titrating the BPDE-DNA adduct competitor, the BPDE-DNA adduct micro-titer plate

(a)



(b)

fmol	OD1	OD2	Average - Blank	%Binding
0	0.69	0.72	0.60	100
50	0.62	0.59	0.50	83.4
100	0.50	0.54	0.42	69.2
150	0.41	0.42	0.31	52.2
200	0.35	0.36	0.25	42.1
250	0.28	0.28	0.18	29.1



ID	OD1	OD2	Average-Blank	% Binding	Adduct remaining (fmol)	% adduct removed
1	0.463	0.450	0.357	59.3	135.0	46.0
2	0.363	0.370	0.266	44.2	190.4	23.8

Figure 3. Principles of NER-ELISA assay

(a) Following incubation between nuclear extracts and BPDE-DNA adduct substrate, the heat-inactivated reaction is mixed with 5D11 monoclonal antibody and added to a BPDE-DNA adduct coated micro-titer plate. Unrepaired substrate DNA in solution acts as a competitor for antibody binding to the bottom of the plate. Percentage binding is subsequently determined through an alkaline phosphatase conjugated secondary antibody detection system. (b) Percent binding is calculated based on optical density readings of wells containing reaction mixtures referenced against maximum antibody binding no-extract control wells. BPDE-DNA adduct competitior standard curve can then be used to convert these values into quantities of adduct substrate remaining in the wells following the reaction. Finally, % adduct removed, can be determined based on the original input amount of adduct substrate added to the reaction. Example adduct removal calculations for two extracts (ID 1 and 2), based on an original reaction input of 250fmol BPDE-DNA adduct, are shown.

coating and the primary BPDE-DNA adduct specific 5D11 monoclonal antibody amounts which will result in an acceptable OD absorbance reading range within a feasible amount of development time.

When testing varying levels of the BPDE-DNA adduct competitor, the linear range of inhibition due to the substrate ranged between concentrations of 50-250fmoles (Figure 4).

Linear levels of inhibition were observed at all tested levels of BPDE-DNA adduct coating (Figure 5). While linearity was also maintained at varying levels of 5D11 monoclonal antibody titrations, there was a distinct positive shift in percent inhibition at the lowest dilution of 5D11 (Figure 6). However, for every decrease in antibody titration dilution, there was a concomitant increase in the time required for colorimetric detection.

Hence, an antibody dilution of 1:2000, the lowest dilution allowing for a reasonable window of colorimetric development, was selected. Finally, sufficient time for the development of the colorimetric reaction was determined by examining a series of optical density readings following the termination of the assay. Readings typically reached consistency within one hour of the commencement of the colorimetric reaction (Figure 7).

Based on this series of titration experiments, the competitive ELISA conditions were set at 5ng BPDE-DNA adduct coating, 250fmol BPDE-DNA adduct substrate, a 1:2000 dilution of 5D11 primary monoclonal antibody, and a colorimetric reaction time of 1hr.

Repair reaction conditions that needed to be determined include repair reaction time, repair reaction temperature, nuclear extract amount, and buffer components of the reaction mixture. For this set of experiments, a single batch of nuclear proteins was extracted from the commercially-purchased lymphoblastoid cell line of an apparently normal individual (Coriell Cell Repositories, Camden, NJ) to test the repair reaction parameters of interest. The competitive ELISA for the detection of adduct removal following all repair reaction experiments was conducted according to the previously outlined established protocol.

Adequate repair reaction time was determined through a time course analysis. Increasing adduct removal was observed with increasing reaction time, with an eventual plateau reached after 4hrs. As the relationship was most linear within the one-hour time-frame, this reaction time was used for all

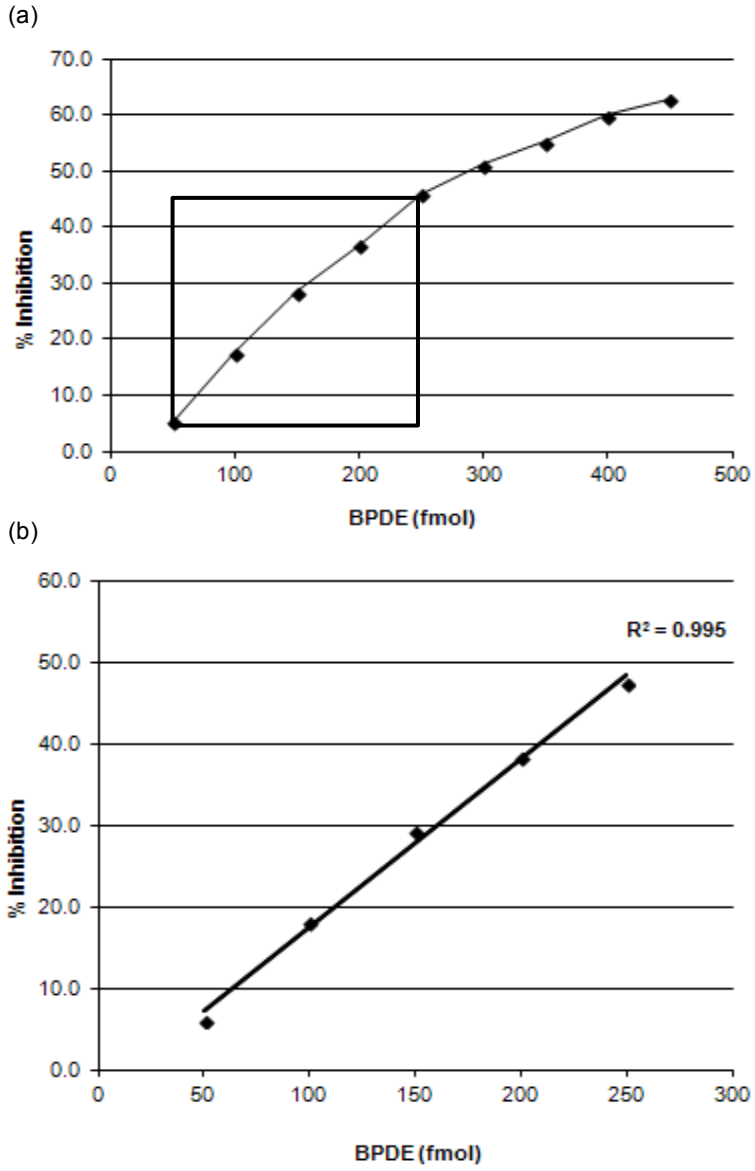


Figure 4. Generated BPDE-adduct DNA substrate acts as inhibitor in competitive ELISA assay
(a) Degree of inhibition observed due to varying concentrations of BPDE-DNA adduct competitor substrate (b) Range of BPDE-DNA adduct substrate concentrations falling within linear phase of inhibition

further experiments (Figure 8). The appropriate nuclear extract amount was determined by conducting a concentration curve. As expected, increasing amounts of extract afforded greater BPDE-DNA adduct removal. While the presence of the highest tested levels of extract lead to the highest observed removal of adducts, a slightly lower concentration of nuclear extract at 200ng was selected for all further

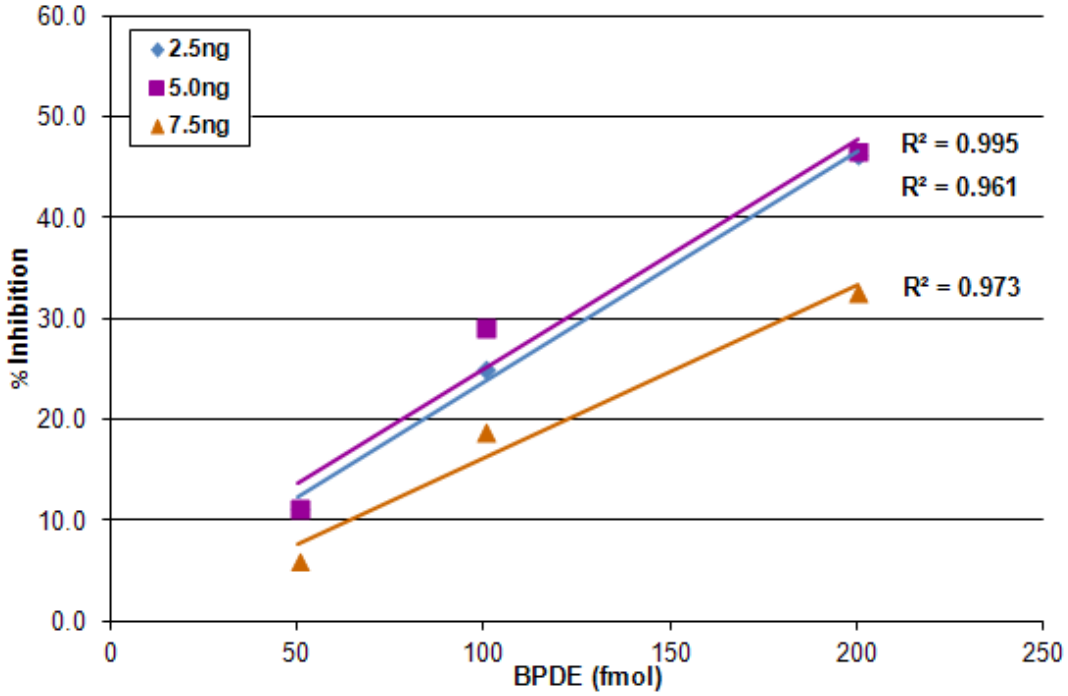


Figure 5. Linearity of substrate inhibition at varying levels of BPDE-DNA adduct coating
Linearity was maintained across all shown levels of BPDE BPDE-DNA adduct coating. All other components of the competitive ELISA were as described.

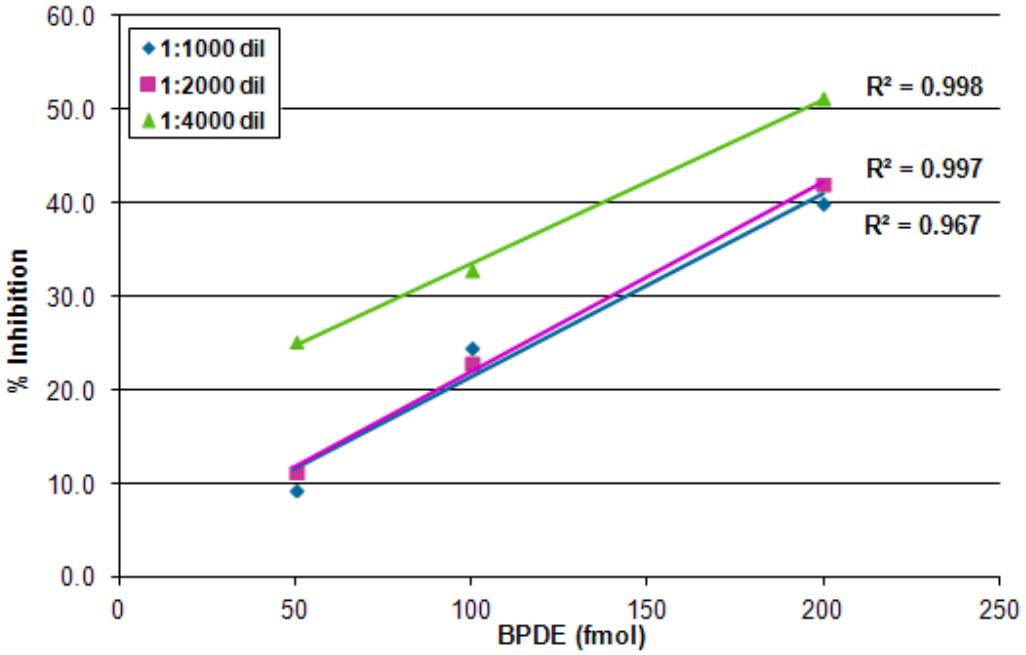


Figure 6. Impact on BPDE-DNA adduct substrate inhibition due to 5D11 monoclonal antibody titration
All BPDE-DNA adduct substrate standard curves were generated using microtiter plates coated with 5ng BPDE-DNA adduct/well. All other elements of the competitive ELISA were as previously described.

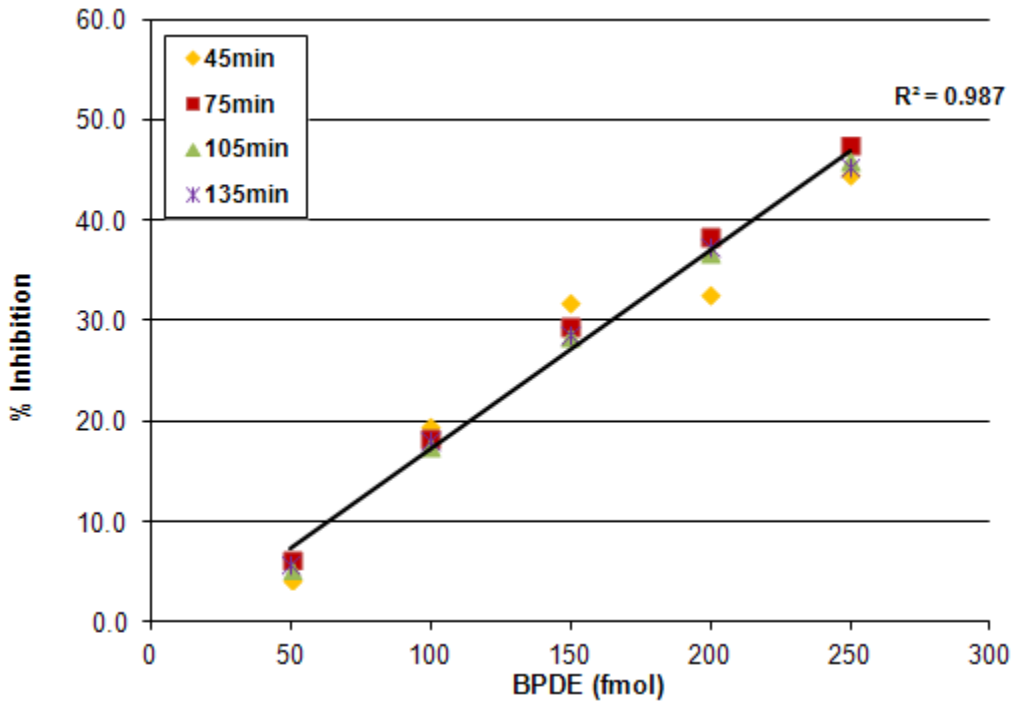


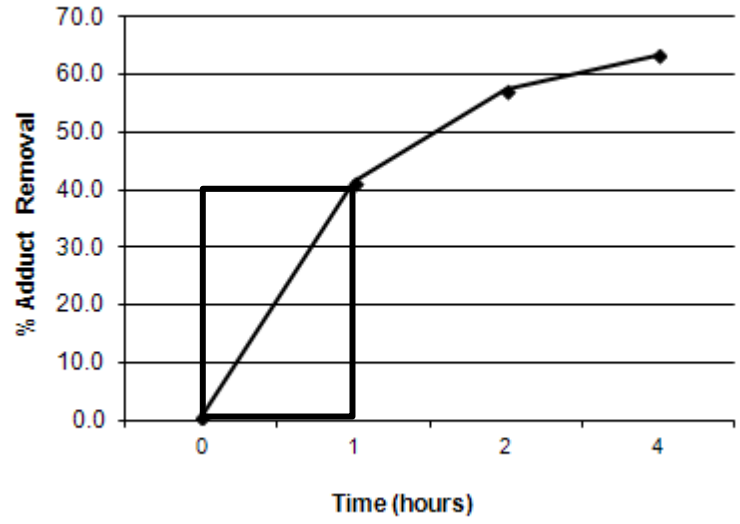
Figure 7. Stability of optical density readings reached over time
Standard curve of BPDE-DNA adduct substrate at 1/2000 antibody titration and 5ng BPDE DNA coating. All other elements of the competitive ELISA were as previously described.

experiments as this amount lead to the most consistent measurements (Figure 9).

Slight variations in repair activity were also observed by varying repair reaction temperatures. The highest levels of adduct removal were observed between a temperature range of 25°C and 37°C (Figure 10). The latter was selected for further experiments, as this temperature most closely reflects *in vivo* conditions. Additional *in vivo* conditions that would have to be met in order to recover the cellular NER activity of the isolated nuclear proteins includes the potential presence of cofactors that may aid in the activation of components in the NER pathway. To test whether this is in fact the case, a set of repair reactions was carried out, one in a reaction mixture buffered according to conditions previously established as appropriate for assays determining nuclear protein activity (NE buffer) and the other in the physiological standard phosphate buffered saline (PBS). As can be seen in Figure 11,

buffer constituents do impact adduct removal capacity of the nuclear proteins. Observed adduct removal activity was indeed higher in NE buffer as compared to PBS.

(a)



(b)

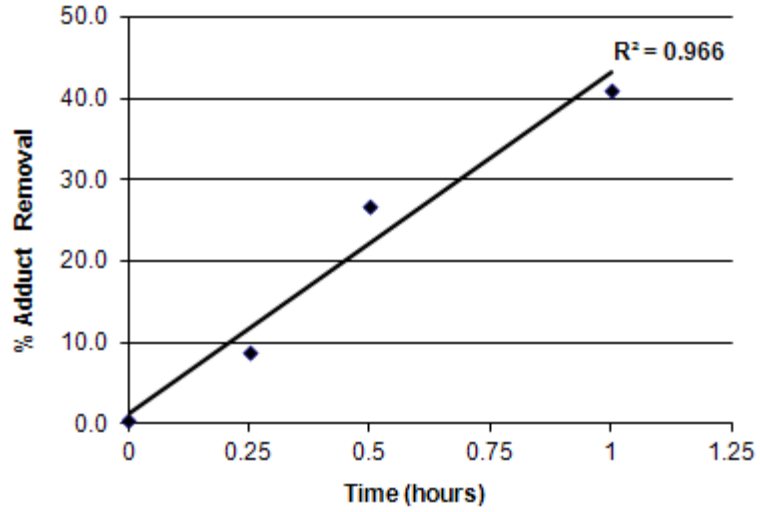


Figure 8. Time course analysis of extract activity
200ng aliquots of nuclear extracts from the lymphoblastoid cell line of an apparently healthy individual were prepared and incubated with 250fm BPDE-DNA substrates for the differing indicated time-spans. All samples were then run using the competitive ELISA method as previously described.
a) BPDE removal capacity measured across a four-hour time course
b) Linearity of BPDE removal activity within a one-hour time course

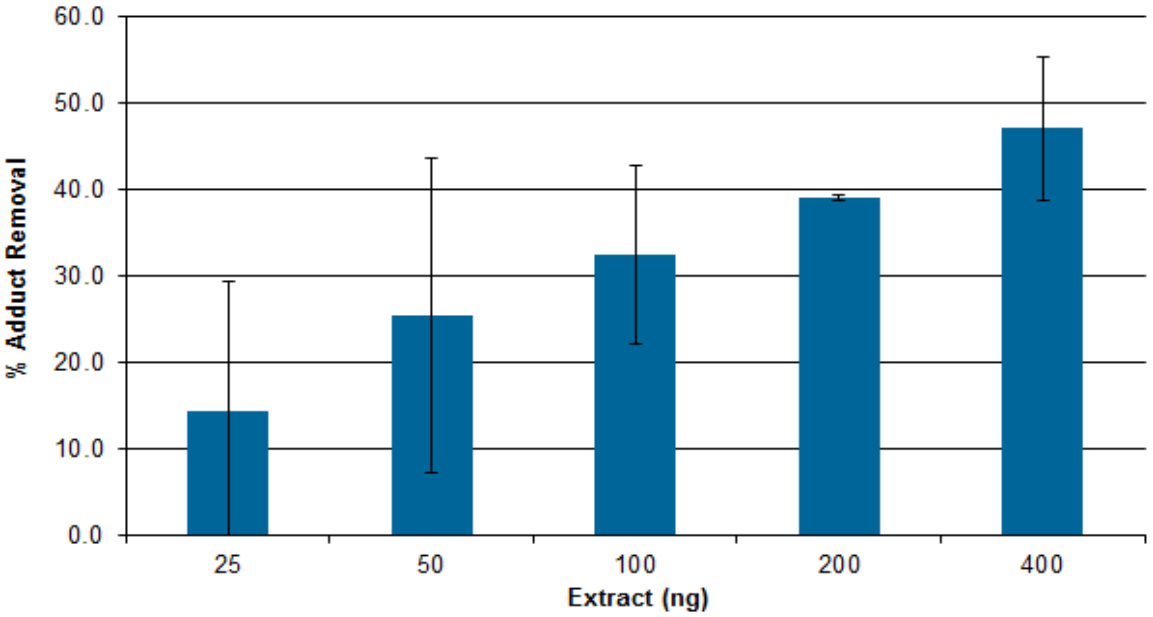


Figure 9. Percent adduct removal increases with increasing amounts of nuclear extract
Indicated concentrations of nuclear extract were incubated with 250fmol BPDE-DNA adduct substrate for one hour. Following the reaction, the competitive ELISA was carried out using established conditions .

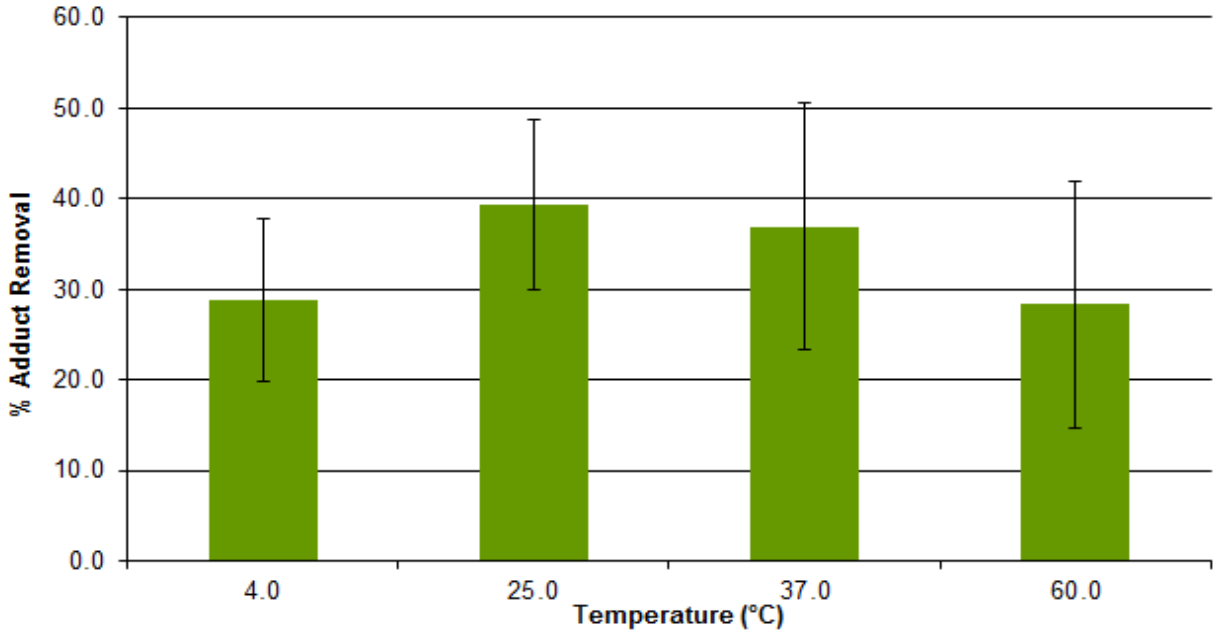


Figure 10. Nuclear extract activity varies across repair reaction temperatures
Four parallel sets of reactions were set up with 200ng nuclear extracts and 250fmol BPDE-DNA adduct substrate at varying temperatures for one hour. Following the reaction, the competitive ELISA was carried out using established conditions

Once the optimized conditions were determined and set, the working parameters of the assay, including specificity, range of detection and inherent variability of the assay were assessed.

The specificity of nuclear extract activity was tested by examining the impact on adduct removal capacity upon deactivating proteins present in the reaction mixture. For this purpose, a set of experiments was conducted in which a normal repair reaction was run alongside a reaction in which proteinase K, an agent that degrades protein, was introduced at the commencement of the reaction. In the presence proteinase K, adduct removal capacity is indeed knocked down compared to the uninhibited repair reaction (Figure 12). This indicates that the observed reduction in adduct levels following a repair reaction is largely due to the presence of active nuclear proteins.

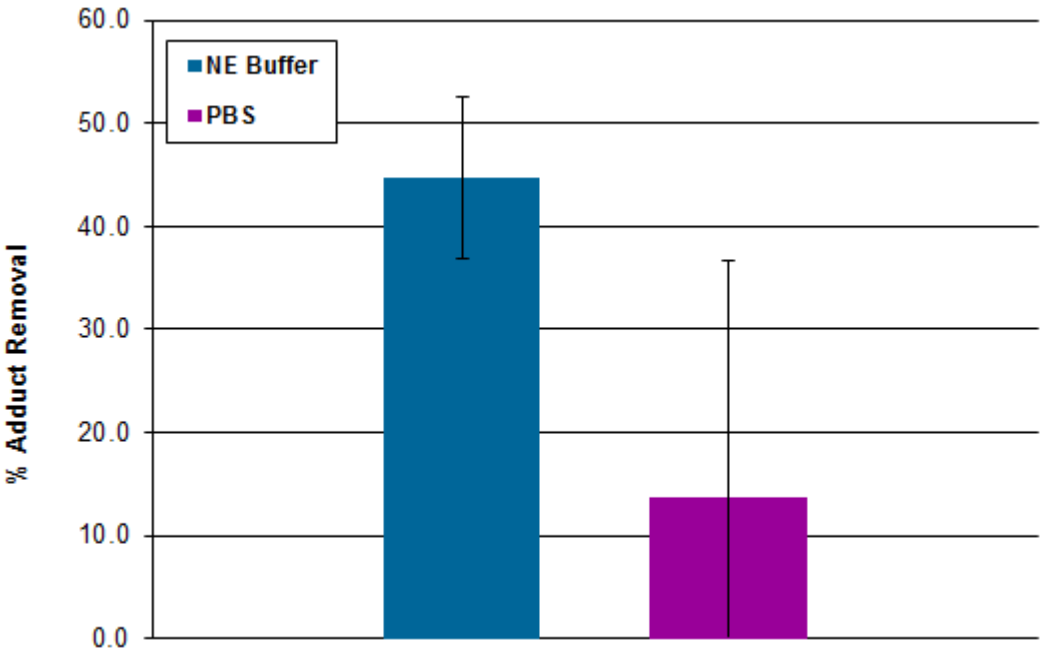


Figure 11. Buffer components impact level of adduct removal activity during repair reaction
Two repair reactions were set up in parallel, each with 200ng nuclear extracts and 250fmol BPDE-DNA adduct substrate but differing in buffer constituents. Higher levels of activity were consistently observed in reaction mixtures containing NE buffer, a buffer which has been previously described for assays determining nuclear extract activity, compared to reaction mixtures containing phosphate buffer saline (PBS).

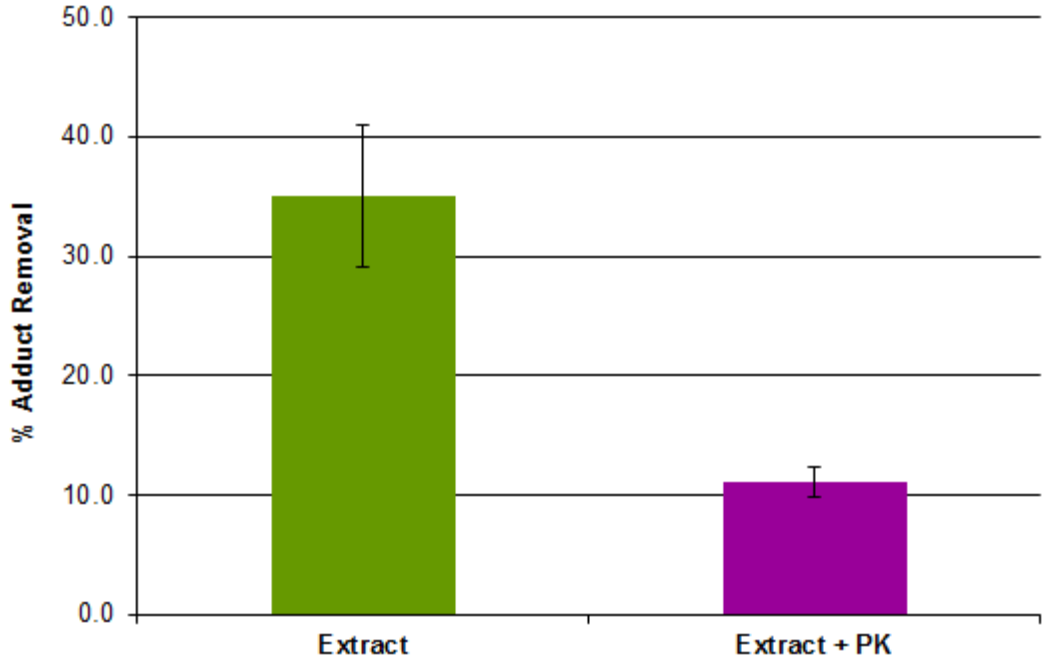


Figure 12. Active nuclear proteins are required for adduct removal activity
Two repair reactions were set up in parallel with 200ng nuclear extracts and 250fmol BPDE-DNA adduct substrate. Proteinase k (100ng/uL) was added to one of the two reactions at the commencement of the 1hr reaction. Following the reaction, the competitive ELISA was carried out using established conditions

Finally, the ability of the assay to distinguish between proficient and deficient NER activity was assessed in a set of experiments utilizing extracts derived from commercially-purchased (Coriell Cell Repositories, Camden, NJ) lymphoblastoid cell lines from an apparently normal individual, an XP individual deficient in XPC and an XP individual deficient in XPD. The parallel repair reactions and competitive ELISA experiments were carried out according to the outlined established conditions. As hypothesized, the assay indicated a knock-down in adduct removal capacity among both XP individuals compared to adduct removal observed in the apparently normal individual (Figure 13).

The inherent variability of the ELISA was determined by comparing the adduct removal capacity of the same extract run in separate independent experiments. The calculated coefficient of variation of 12.3% among the four identically conducted trials (data not shown) falls within an acceptable range of inter-assay variation commonly reported for ELISAs.

Distribution of adduct removal capacity in a group of healthy individuals

The effect of several factors known to impact DNA repair capacity was evaluated among 122 unaffected women enrolled in the NY-BCFR. Adduct removal capacity among these women ranged from 9.85-53.9%. The mean adduct removal capacity of 34.0% with standard deviation of 8.5% is in line with the previously reported adduct removal capacity among these women based on the immunohistochemical assay (35.1% +/- 24.4%)[109] (Figure 14).

None of the selected variables were observed to significantly impact adduct removal capacity within the unaffected women. However, several borderline associations were observed in the expected direction. As seen in Table 2, women above the median age at menopause had reduced adduct removal capacity compared to women below the median age at menopause (p=0.05). Similarly, women with an extended family history of breast cancer had reduced adduct removal capacity compared to women without an extended family history (p=0.07).

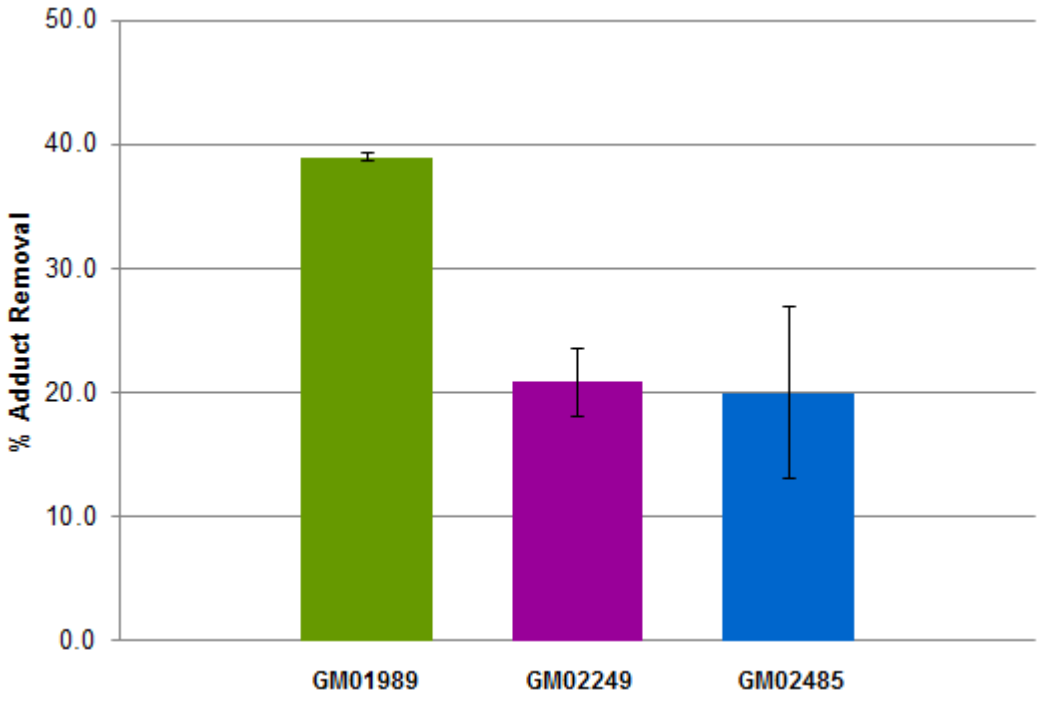


Figure 13. Assay range of adduct removal detection
Three parallel repair reactions were set up with 250fmol BPDE-DNA adduct substrate. Nuclear extracts for each reaction were derived from lymphoblastoid cell lines of an apparently normal individual (GM01989), an XP individual deficient in XPC (GM02249) and an XP individual deficient in XPD.

Case-control study

To determine whether case-control differences in NER activity could be observed using the competitive ELISA method, we assessed the adduct removal capacity from nuclear extracts derived from the LCLs of sisters previously analyzed using the immunohistochemical method in a family-based case control study.

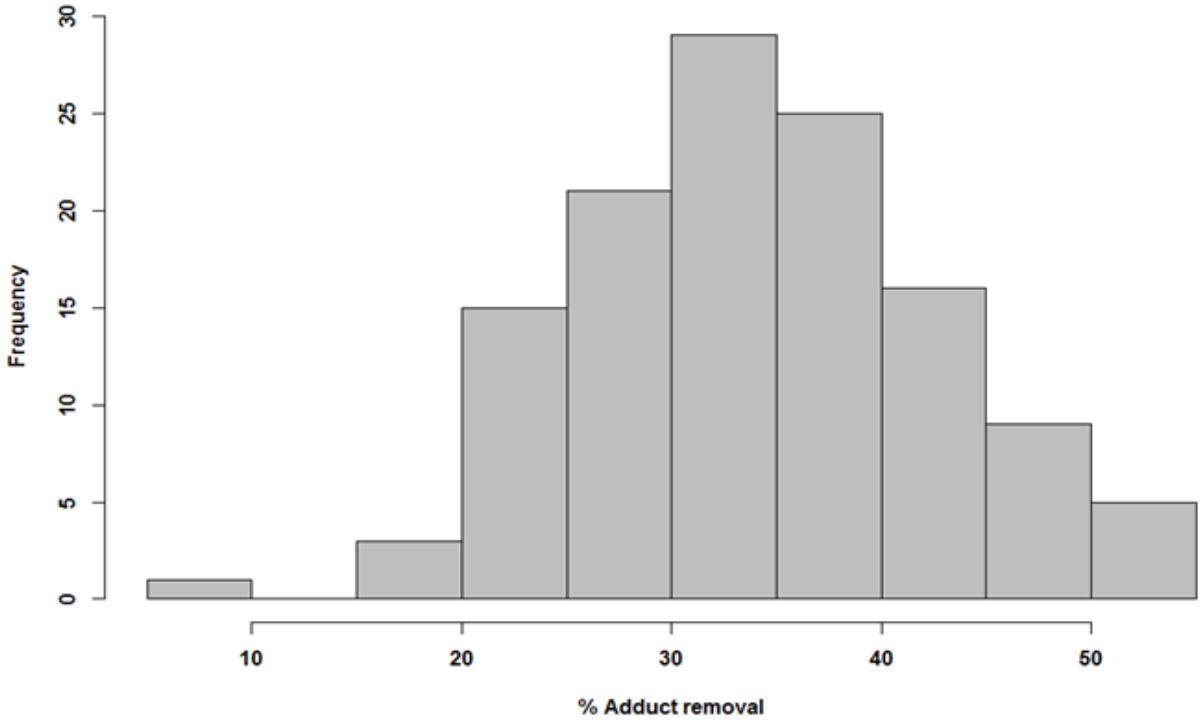


Figure 14. Distribution of adduct removal capacity among the unaffected sisters

The range in adduct removal capacity observed among cases spanned from 13.22% to 54.42% with a mean and standard deviation of 33.98% and 8.45%, respectively. The difference in mean adduct removal capacity between cases and controls was not statistically significant (p=0.98) (Table 3). Conditional logistic regression analysis also did not indicate a significant association adduct removal capacity and breast cancer risk, adjusted for age at blood draw and smoking status (Table 4). Furthermore, NER capacity measured within sister-sets as indicated by the NER-ELISA method did not

correlate (r=-0.01, p=0.86) with the NER capacity previously determined within the same group using the immunohistochemical method (Figure 15).

Table 1. Association between selected variables and adduct removal capacity among unaffected sisters

	Estimate	St.Err	t-value	Pr(>|t|)
Age (>50 vs ≤50)	-3.02	1.52	-1.99	0.05
Smoke (Ever vs. Never)	-2.52	1.53	-1.65	0.10
BMI (kg/m^2) (≥25 vs <25)	-0.20	1.54	-0.13	0.90
Family History (>1 vs. ≤1)	-2.80	1.55	-1.81	0.07

Table 2. Distribution of selected variables among sisters affected and unaffected with breast cancer

Variable	No. cases (%) (n=121)	No. controls (%) (n=124)	p-value[α]
Age at blood draw			
≤50	69 (57.0)	65 (53.3)	0.65
>50	52 (43.0)	57 (46.7)	
BRCA1 (+) mutation	10 (8.3)	6 (4.8)	0.41
BRCA2 (+) mutation	3 (2.5)	2 (1.6)	0.98
Ethnicity			
Non-hispanic White	93 (77.5)	90 (72.6)	0.67
African American	1 (0.80)	1 (0.80)	
Other	26 (21.7)	33 (26.6)	
Smoking Status			
Never	68 (56.2)	68 (53.8)	0.93
Ever	53 (43.8)	56 (45.2)	
BMI (kg/m^2)			
<25	62 (52.1)	60 (48.4)	0.65
≥25	57 (47.9)	64 (51.6)	
Family History			
=1	71 (58.7)	75 (60.5)	0.87
>1	50 (41.3)	49 (39.5)	

[α] Pearson's X^2 test

Table 3. Adduct removal capacity among sisters affected and unaffected with breast cancer

Variable	Cases		Controls		p-value[δ]
	N	Mean	N	Mean	
All	120	34.0 ± 8.39	123	34.0 ± 8.52	0.98
Age at blood draw					
≤50	68	34.5 ± 8.54	66	35.4 ± 8.49	0.56
>50	51	33.3 ± 8.20	56	32.4 ± 8.33	0.54
p[δ]		0.44		0.05	
Smoke Status					
Never	67	34.7 ± 8.03	67	35.1 ± 8.61	0.76
Ever	52	33.2 ± 8.82	55	32.6 ± 8.27	0.74
p[δ]		0.32		0.10	
BMI (kg/m^2)					
<25	61	32.9 ± 8.05	59	34.1 ± 9.61	0.46
≥25	57	35.3 ± 8.63	63	33.9 ± 7.43	0.32
p[δ]		0.11		0.90	
Family History					
=1	70	34.3 ± 9.45	74	35.1 ± 8.13	0.60
>1	49	33.6 ± 6.65	48	32.3 ± 8.90	0.42
p[δ]		0.62		0.07	

[δ] P-value determined based on Student's t-test

Table 4. Conditional logistic regression analysis of the relationship between adduct removal capacity and breast cancer risk among sisters discordant for breast cancer

Variable	Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)*
Adduct Removal Capacity (%)		
Continuous	0.99 (0.95,1.04)	1.0 (0.95,1.04)
Tertile 1: (≤28.0)	0.85 (0.39,1.87)	0.87 (0.39,1.97)
Tertile 2: (28.0-39.6)	0.89 (0.33,2.36)	0.91 (0.33,2.46)
Tertile 3: (>39.6)	1.0 (referent)	1.0 (referent)

*Model adjusted for age at blood draw and smoking status

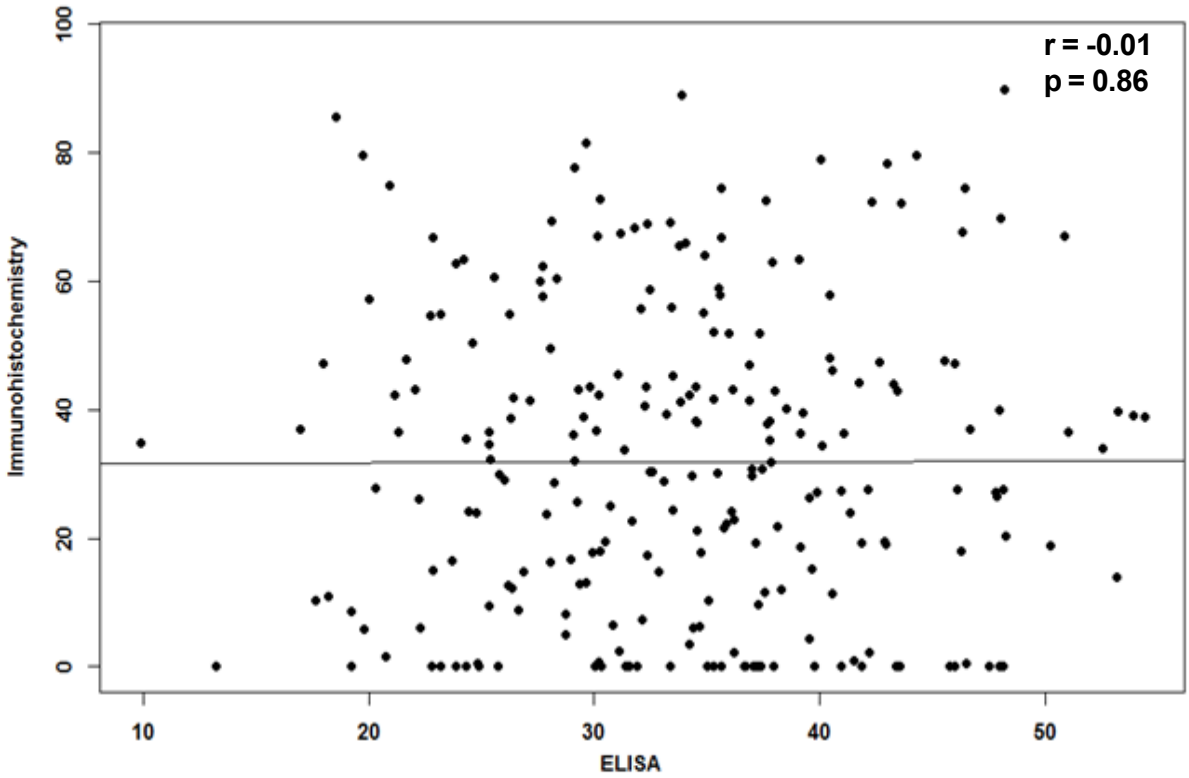


Figure 15. Correlation between immunohistochemical and ELISA assessments of adduct removal capacity among analyzed sister-sets

Discussion

The current study was undertaken to develop a high-throughput DNA repair phenotyping methodology for potential implementation as a screening tool within a population-wide setting. While several phenotyping assays are currently available, constraints on scalability have limited the application of these assays to describe repair capacity-associated implications for cancer risk within small population subsets.

Two such methodologies include HCR and the *in vitro* BPDE-DNA adduct removal capacity. While studies have suggested associations between NER capacity and breast cancer risk utilizing these assays, both require a series of time-consuming and cumbersome steps, a common issue with phenotyping in general.

In addition to the multi-day time-frame entailed for the combined processes of repair reaction and detection, phenotyping assays are often carried out within living cell systems, adding a period of cell culturing until exponential growth is reached prior to the actual commencement of the repair reaction. This presents an additional challenge for matched case-control studies where concern over potential batch effects necessitates cell lines within the same set to be run together, further limiting the flexibility of the assay by requiring entire matched sets to reach exponential growth. Lastly, reliable quantification of repair capacity is also an issue of concern for these assays. The reliance on subjective assessments based on relative fluorescent or luminescent intensity only allows for crude measurements of repair capacity that are non-standardized and liable to bias.

The development of the competitive NER ELISA addresses several of the challenges presented by currently available phenotyping assays. In place of cultured cell lines, the repair reaction for the ELISA utilizes nuclear proteins. As these cellular extracts can be isolated and stored as required, the investigator is afforded added flexibility in determining assay runs, rather than having the timing of the runs determined by the growth rate of the cell lines. Scalability of the detection method is facilitated by the use of BPDE-induced damage specific 5D11, a monoclonal antibody previously isolated by our group and assessed to be sensitive enough to detect levels of damage that can be expected from carcinogen exposures typically observed within the population-based setting[128]. Unlike the BPDE-DNA

damage–specific polyclonal antiserum utilized in the immunohistochemical method, monoclonal antibodies can be reliably obtained from cultured cell-lines, and would allow for a theoretical limitless and consistent supply, facilitating scalability and reproducibility of the assay. Finally, the implementation of a standard curve enables detection of absolute adduct levels, improving upon the reliance on relative and subjective quantification of repair utilized by existing methods.

While studies utilizing HCR and the *in vitro* BPDE adduct removal assay have both shown that suboptimal NER capacity impacts breast cancer risk, a similar association was not observed in the current study using the competitive ELISA method. Several factors may explain the observed discrepancy in findings. While the intended use of the assay will be to assess repair capacity within extracts derived from readily accessible primary mononuclear cells, the findings of the current study were based on assessing repair activity of LCLs. These immortalized cell lines are able to evade senescence experienced by normal cells in culture and, therefore, provide the consistent supply of cells needed for assay development. However, immortalization is achieved through Epstein Barr virus transformation of the cells, a process which may artificially alter cellular processes including DNA repair to an extent where they no longer truly reflect the original host functionality.

While the study utilizing the immunohistochemical method also assessed NER capacity in LCLs, the ELISA method is distinguished by the further isolation of nuclear proteins from this cell source for the repair reaction. The nuclear extracts, while affording additional flexibility to the assay, may also serve to increase the deviation from true *in vivo* conditions by focusing the assessment of protein activity within a context that is removed from potentially essential conditions and components present in the original setting in which repair activity takes place. This artificial setting is in contrast to the entirety of the process captured in the viable cells of the immunohistochemical assay.

The observed null findings may also be partially due to the introduction of measurement error. A major potential for this type of bias can arise from the various opportunities to introduce batch effects into the assay, all of which may not be accounted for by post-hoc analysis. The 247 cell lines were grown and extracted over a period of 1.5 years, and assessment of repair activity of the cell lines using the ELISA method commenced once nuclear proteins from all cell lines had been extracted. As a result, a portion

of the cell lines had been stored frozen for close to 2 years at the time of assay, while others were stored for a relatively shorter period of time. As protein activity degrades over time, it is possible that variations in protein activity measurements may be partially confounded by storage time. However, since complete sister-set cell lines were grown simultaneously, difference in storage time within the same sets did not vary by more than a few days. Still, as all lines stemming from the same set did not necessarily reach adequate cell density for extraction at the same time, extraction batch may have varied among individuals within the same sister-sets. Finally, the ELISA assessment of repair capacity of the complete batch of cell lines was necessarily conducted over separate assay runs. Potential variability between plates could introduce artificial variability between assayed samples. Indeed, variation based on an internal control extract run alongside samples on every plate did indicate an inter-plate variability of 12%. However, as sister-sets were run together on the same plate, variation due to this factor was accounted for.

The inability to distinguish cases from controls based on NER capacity maybe due to limitations in the sensitivity and specificity of the assay. Degradation of protein activity in the presence of proteinase K led to a knock-down but not a complete knock-out in observed activity. Similarly, low levels of repair activity were observed in nuclear extracts derived from individuals with mutations in *XPC* and *XPD*. This noise, likely non-specific and unrelated to repair activity, limits the possible range of detection of the assay, as can be observed by the modest differences in repair capacity observed between extracts derived from an apparently normal individual compared to individuals with XP. Given that these individuals represent the extreme ends of the spectrum of repair capacity detectable with this assay, the observed narrow range in values distinguishing these individuals may render the assay unable to distinguish more subtle differences in repair more typically observed within the population.
The observed narrow range in detection is at least partially attributed to background noise of the assay. Repair activity, as determined by the ELISA, relies on the assumption that an increase in signal detected in the presence of nuclear extracts reflects a reduction in competitor level available to inhibit antibody binding to BPDE-DNA adduct coating. However, the reduction in competitor availability could also potentially arise without actual repair-related removal of substrate adducts from the competitor pool. An observed increase in BPDE-DNA adduct coating binding may arise in the presence of a

differential in antibody binding capacity between BPDE-DNA adduct coating and competitor. For example, it is possible that during the reaction with nuclear extracts, components in the reaction mixture mask adducts in such a way that they are not recognized by the antibody, increasing the availability of the antibody to bind to the coating.

Another assumption in the NER ELISA assessment of repair capacity is that BPDE-adduct removal capacity is reflective of overall NER activity. In fact, in addition to ELISA-targeted damage recognition and damage removal, the completion of the pathway also encompasses insertion of the correct base and ligation to seal the nick. It is, therefore, conceivable that the observed lack of an association actually reflects an underrepresentation of NER defects captured by the assay. In addition to limiting the focus of repair capacity to substrate removal, the substrate specification can also introduce error in capturing true NER activity. While BPDE-DNA adducts are major targets of the NER pathway, it has been shown that BER can also mediate these lesions[129]. Hence, focusing on one specific lesion may prevent detecting deficiencies in NER capacity, as they may be masked by the complimentary activities of other pathways.

The discrepant findings between the current and previous studies assessing the impact of NER capacity on breast cancer susceptibility highlight the complications involved in developing a phenotyping assay capable of capturing overall pathway activity. Adequately mimicking *in vivo* conditions to effectively capture the intricacies of interacting components entails processes difficult to scale up into a high-throughput methodology, while the current attempt to simplify and standardize the required steps and components comes at the cost of adequately assessing repair activity. Still, initial findings based on phenotyping have been informative in suggesting a role for NER capacity on breast cancer risk. Validating and translating these findings into a screening tool will require additional studies, likely ones focused on more specific endpoints within the pathway that are more amenable to be assessed in a high-throughput setting.

Chapter III: Mismatch repair polymorphisms as indicators of breast cancer risk

Abstract

Suboptimal DNA repair capacity is believed to contribute to genomic instability and, thereby, onset of carcinogenesis. In fact, deficient repair has been reported to account for up to a five-fold increase in breast cancer risk. Major cancer susceptibility genes involved in repair, including *BRCA1* and *BRCA2* among breast cancer cases, have been identified. However, mutations in these genes are thought to account for only 5-10% of familial cases. Additional genetic susceptibility factors may play a role in the remaining cases, including polymorphisms in DNA repair pathway genes. MSI, a commonly observed marker of MMR defects, has been observed in skin samples from breast cancer patients, suggesting a putative role for MMR in breast cancer. In this study, we investigated polymorphisms in mismatch repair genes and their association with breast cancer risk among participants of the New York site of the BCFR (N=744). Of the 313 sister-sets included in the registry, 306 sister-sets consisted of sisters discordant for disease. A total of twelve SNPs were chosen and assayed using Taqman SNP genotyping kits from Applied Biosystems. Using conditional logistic regression analysis, an association with breast cancer risk was observed for three of the twelve SNPs assayed. A significant increase in breast cancer risk was observed due to the MutY_rs3219489 variant allele (OR = 2.24, 95% CI =1.11-4.52) while a borderline increase in risk was observed due to the MSH2_rs2303428 variant allele (OR = 1.71, 95% CI = 0.99-2.95). A protective effect was observed due to the MLH3_rs175080 variant allele among women without an extended family history of breast cancer. Future studies will focus on replicating these findings in a larger study population.

Introduction

Breast cancer is one of the leading causes of cancer death among women, second only to lung cancer[143]. Two influential susceptibility genes, *BRCA1* and *BRCA2*, have been identified. These genes encode proteins involved in facilitating DSBR, suggesting a critical role imparted by these genes on the maintenance of genomic integrity[159]. Carriers of mutations in these genes have up to an 85% risk of developing breast cancer by the age of 70[18]. However due to the associated lethality, complete loss of function in these genes is rare, accounting for only 5-10% of all observed cases. Additional coverage of risk may be captured by less dramatic alterations, including SNPs.

While DSBR-targeted damage, such as chromosomal aberrations, constitute the type of genomic instability most frequently associated with breast cancer risk, susceptibility to substrates targeted by other repair pathways may also confer risk. In addition to DSBR, *BRCA1* is also associated with a genome surveillance complex that includes MMR proteins to sensor and repair replication-associated DNA damage that have escaped the DNA polymerase proof-reading mechanism[160]. These types of damage include point mutations that result from single base mismatches following the incorrect incorporation of a nucleotide, as well as frame-shift mutations that occur through errors in the number of bases incorporated at repetitive sequences, resulting IDLs.

Such slippages are prone to occur in regions containing microsatellites, simple repeat sequences scattered throughout the genome. Hence, detection of an alteration in the number of repeats, is a common marker used to identify a defect in MMR[161].

It has been proposed that such MSI-generating defects may be an early event in carcinogenesis that confer a mutator phenotype by inducing genomic instability, thereby enabling the acquisition of additional mutations necessary for tumor progression[162]. Studies have already shown increased MSI within tumors, including breast, compared to normal tissues derived from the same individual, implicating defective MMR with breast cancer[115,163]. Furthermore, in a study including 30 breast cancer patients, all cases with point mutations in either *MLH1* or *MSH2*, two MMR genes, exhibited increased MSI in tumor versus control tissue, indicating that such sequence variations in MMR genes may play a role in breast cancer risk[154].

To further investigate this potential relationship, we carried out a family-based case-control study among sisters discordant for breast cancer enrolled in the New York site of the BCFR to assess the association between 12 variants in MMR-related genes and breast cancer susceptibility.

Materials and Methods

Study population

The subjects for the current study included 313 sister-sets (n=714) consisting of sisters discordant for breast cancer enrolled in the New York site of the BCFR with available WBCs (Figure 1).

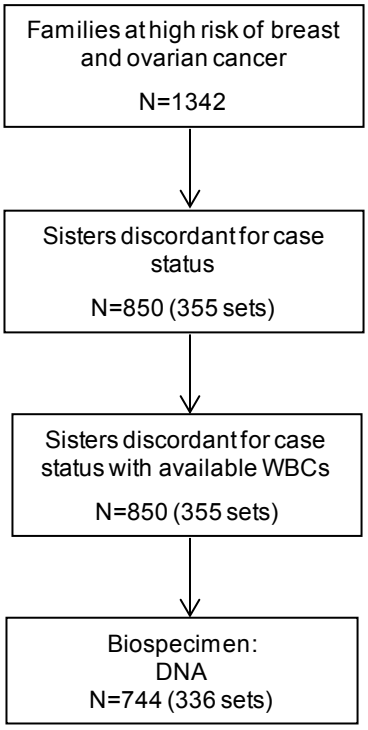


Figure 1. Genotyping study design
The study was conducted among sisters discordant for disease with DNA available from WBCs.

Laboratory methods

DNA was extracted from WBCs using a commercially available kit following the manufacturer's instructions (FIexigene, Qiagen). DNA concentration and quality was determined using a NanoDrop ND-1000 spectrophotometer with absorbance detection at 260nm and 280nm.

Real-time PCR was carried out in 5uL reactions containing 5ng genomic DNA, 1X

Taqman SNP Genotyping Assay Mix, and 1X Taqman Universal PCR Mastermix, containing DNA polymerase, dNTPs, and optimized buffer components. The PCR thermocycling protocol consisted of

95°C for 10min, followed by 45 cycles of 95°C for 15sec and 60°C for 90sec. Assays were conducted in a 7900 Real-time PCR platform (ABI Applied Biosystems, Foster City, CA), and allelic discrimination was determined using software provided by the manufacturer. Each plate contained non-template controls and 10% of the samples were re-assayed to determine concordance. All laboratory personnel involved in sample handling were blinded to case status.

SNP selection: 12 SNPs within genes of the MMR pathway were chosen for analysis. Included SNPs met the following criteria: a) Association with cancer risk reported in epidemiologic studies; b) Minimum 5% minor allele frequency in Caucasians; or c) Missense mutation incurred by presence of polymorphism in the coding region with minimum 1% minor allele frequency in Caucasians.

Statistical analysis

Hardy-Weinberg equilibrium was tested to assess deviations of observed from expected genotype frequencies among cases and controls. Pearson's chi-square tests were conducted to assess case-control differences based on demographic variables. These variables included age at blood draw, smoking status (ever vs. never), BMI (≥25kg/m^2 vs <25kg/m^2, cut-off based on WHO definition of normal range), age at menarchy (>13 vs ≤13 years, based on median age at menarche among unaffected sisters), age at first parity (>30 vs ≤30 years, nulliparous women were assigned age at blood draw), and extended family history (>1 vs 1). Univariate conditional logistic regression was used to determine odds ratios and 95% confidence intervals associating individual queried genotypes with breast cancer risk. Due to the low frequency of homozygous carriers of the variant allele for some of the SNPs assayed, the analysis for each SNP was also conducted upon combining heterozygous and homozygous carriers of the variant allele. A final conditional logistic analysis was conducted to assess the combined effect of the two most strongly associated genotypes from the univariate analysis. These findings were also adjusted for age at blood draw and smoking status. A stratified conditional logistic regression analysis was also conducted to assess the presence of effect modification due to age (≤50 vs. >50) and extended family history (>1 vs 1). All analysis was performed using R: A Language and Environment for Statistical Computing (2011), R Foundation for Statistical Computing, Vienna, Austria.

Results

The SNPs chosen for the current study are shown in Table 1. Seven of the 12 assayed SNPs are located within exons in which the presence of the variant allele induces a non-synonymous amino acid alteration. Two of the selected SNPs (MLH1_1800734 and MSH6_rs3136228) are located upstream of the transcription start site (TSS), and the remaining three SNPs are located in introns (MLH1_rs1799977, MSH2_rs2303428, and MSH3_rs863221).

Table 1. Candidate MMR SNPs selected for analysis

Gene	dbSNP ID	NCBI Accession No.	Polymorphic nucleotide [amino acid] change	Life Technologies Assay No.
MLH1	rs1799977	NM_000249.2	715A>C [I219V]	C_1219076_20
MLH1	rs1800734	NM_000249.3	106G>A	C__7535141_1_
MLH1	rs2286940	NG_007109.1	40128C>T	C__16181046_10
MLH3	rs175080	NM_001040108.1	2531C>T [P844L]	C_1082805_10
MSH2	rs2303428	NG_007110.1	78238T>C [IVS12-6T>C]	C__11804019_1_
MSH3	rs26279	NM_002439.3	3133G>A [T1048A]	C_800002_1_
MSH3	rs184967	NM_002439.2	2846G>A [R952Q]	C_907914_10
MSH3	rs863221	NG_007109.1	40128C>T	C___3103297_10
MSH4	rs5745325	NM_002440.2	289G>A [A97T]	C_3286081_30
MSH4	rs5745549	NM_002440.2	2741G>A [N914S]	C_1184803_10
MSH6	rs3136228	NG_007111.1	4531T>G	C__28985526_10
MutY	rs3219489	NM_001128425.1	1014G>C [H335Q]	C_27504565_10

The demographics of the study population are shown in Table 2. A significantly higher number of BRCA1 mutation carriers were cases compared to controls. No other case-control differences were observed.

Table 2. Distribution of selected variables among sisters affected and unaffected with breast cancer

Variable	No. cases (%) (n=336)	No. controls (%) (n=408)	p-value[α]
Age at blood draw (in years)			
≤50	205 (61.0)	263 (64.5)	0.35
>50	131 (39.0)	144 (35.3)	
BRCA1 (+) mutation	24 (7.14)	14 (3.43)	0.03
BRCA2 (+) mutation	15 (4.46)	13 (3.19)	0.47
Ethnicity			
Non-hispanic White	210 (63.1)	235 (57.6)	0.30
African American	6 (1.8)	10 (2.5)	
Other	117 (35.1)	163 (40.0)	
Smoking Status			
Never	199 (59.2)	233 (57.1)	0.61
Ever	137 (40.8)	175 (42.9)	
BMI (kg/m^2)			
<25	171 (51.0)	217 (53.8)	0.49
≥25	164 (49.0)	186 (46.2)	
Age at menarche (in years)			
<13	153 (46.0)	185 (42.6)	0.98
≥13	181 (54.0)	222 (57.4)	
Parity			
Nulliparous	77 (23.0)	84 (20.6)	0.50
Parous	259 (77.0)	324 (79.4)	
Age at first parity			
<30	196 (63.2)	261 (67.4)	0.13
≥30	140 (36.8)	147 (32.6)	
Family History			
=1	192 (57.1)	258 (63.2)	0.11
>1	144 (42.9)	150 (36.8)	

[α] Pearson's X^2 test

A representative example of the amplification and allelic discrimination profile and plot for a set of samples genotyped for MutY_rs3219489 is shown in Figure 2. As observed in the depicted run, a call rate of greater than 95% was observed for all assays.

.

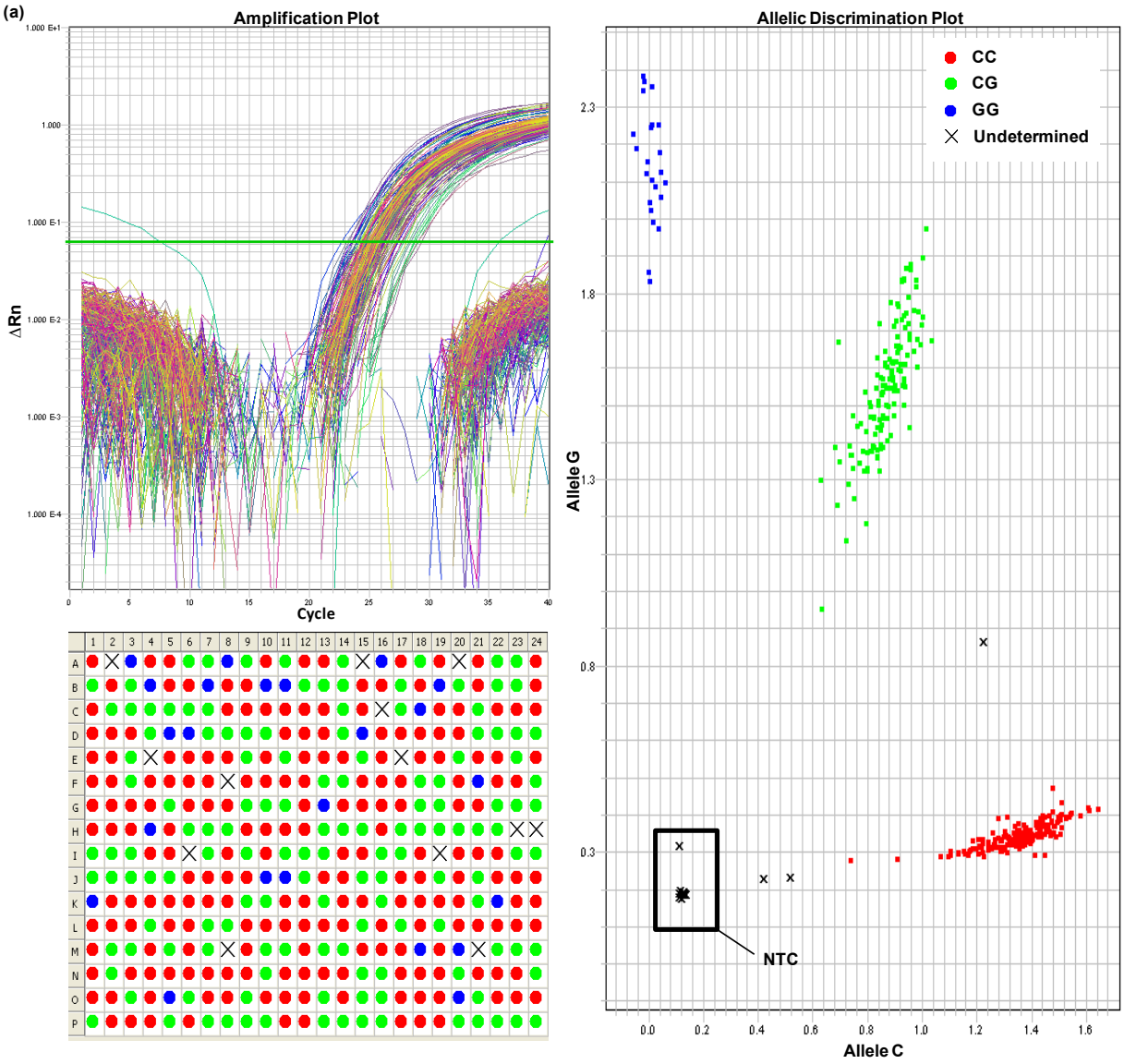


Figure 2. Representative amplification plot (a) and allelic discrimination profile (b) and plot (c) of samples assayed for Muty_rs3219489
Each plate was also run with non-template controls (NTC), as indicated on allelic discrimination plot.

As seen in Table 3, several of the tested SNPs deviated significantly from Hardy-Weinberg (HW) equilibrium (MLH1_rs1799977, p=0.01, MLH1_rs2286940, p<0.01, MSH4_rs5745325, p=0.03). This is not surprising as the expected proportions of allele frequencies under a state of equilibrium presuppose

Table 3. Test for Hardy Weinberg equilibrium of variants genotyped within sister sets

Gene	SNP	Hardy Weinberg ChiSq	p-value
MLH1	rs1799977	6.12	0.01
MLH1	rs1800734	0.00	0.97
MLH1	rs2286940	14.7	<0.01
MLH3	rs175080	0.01	0.91
MSH2	rs2303428	1.03	0.31
MSH3	rs26279	0.02	0.89
MSH3	rs184967	0.09	0.76
MSH3	rs863221	4.00	0.05
MSH4	rs5745325	4.52	0.03
MSH4	rs5745549	1.22	0.27
MSH6	rs3136228	3.29	0.07
MutY	rs3219489	0.89	0.34

random mating within the population. As the current study is a family based case-control study, the allele frequencies may be skewed due to the presence of related individuals in the study population.

Conditional logistic regression analysis revealed a significant association between the presence of the variant allele MutY_rs3219489 and breast cancer risk (OR=2.23, 95% CI = 1.10-4.52), upon adjusting for age at blood draw and smoking status. Similarly, after adjusting for age at blood draw and smoking status, a borderline association was also observed for the presence of the variant allele MSH2_rs2303428 and breast cancer risk, upon combining homozygous carriers with heterozygous carriers (OR=1.71, 95% CI =1.00-3.00) (Table 4).

Including both MSH2_rs2303428 and MutY_rs3219489 in the same analysis, an increase in the odds of breast cancer was observed with increasing number of variant alleles in the two genes. Individuals who are either heterozygous or homozygous carriers of the variant allele for both genes have a two-fold increase in the odds of breast cancer over individuals with no variant allele for either gene, adjusted for age at blood draw and smoking status (Table 5).

A stratified analysis was conducted to assess the presence of effect modification due to age and extended family history on the relationship between MMR genotype and breast cancer risk. Among

Table 4. Association between MMR genotype and breast cancer risk

Gene	dbSNP ID	Genotype	Affected Sister No. (%)	Unaffected Sister No. (%)	OR (95% CI)	OR[a] (95% CI)
MLH1	rs1799977	AA	183 (55.3)	226 (55.8)	1.0 (Ref)	1.0 (Ref)
		AG	118 (35.6)	143 (35.3)	1.05 (0.66-1.66)	1.07 (0.67-1.70)
		GG	30 (9.1)	36 (8.9)	0.96 (0.45-2.06)	1.00 (0.46-2.14)
		AG/GG	148 (44.7)	179 (44.2)	1.04 (0.66-1.64)	1.07 (0.67-1.69)
		Freq(G)	0.27	0.27		
	rs1800734	GG	178 (53.6)	224 (55.2)	1.0 (Ref)	1.0 (Ref)
		GA	131 (39.5)	154 (37.9)	1.00 (0.66-1.52)	0.99 (0.67-1.51)
		AA	23 (6.9)	28 (6.9)	1.07 (0.50-2.31)	1.07 (0.49-2.32)
		GA/AA	154 (46.4)	182 (44.8)	1.01 (0.67-1.52)	1.00 (0.66-1.51)
		Freq(A)	0.27	0.26		
	rs2286940	CC	135 (40.7)	164 (40.4)	1.0 (Ref)	1.0 (Ref)
		CT	133 (40.1)	168 (41.4)	0.96 (0.61-1.52)	0.97 (0.61-1.54)
		TT	64 (19.3)	74 (18.2)	1.08 (0.58-2.00)	1.08 (0.58-2.00)
		CT/TT	197 (59.4)	242 (59.6)	0.97 (0.62-1.53)	0.99 (0.63-1.56)
		Freq(T)	0.39	0.39		
MLH3	rs175080	GG	104 (31.9)	111 (27.4)	1.0 (Ref)	1.0 (Ref)
		GA	158 (48.5)	207 (51.1)	0.68 (0.44-1.06)	0.69 (0.45-1.08)
		AA	64 (19.6)	87 (21.5)	0.59 (0.32-1.08)	0.60 (0.33-1.10)
		GA/AA	222 (68.1)	294 (72.6)	0.67 (0.43-1.03)	0.68 (0.44-1.05)
		Freq(A)	0.44	0.47		
MSH2	rs2303428	TT	242 (75.6)	310 (78.1)	1.0 (Ref)	1.0 (Ref)
		TC	77 (24.1)	81 (20.4)	1.71 (0.99-2.95)	1.73 (1.00-3.00)
		CC	1 (0.3)	6 (1.5)		
		TC/CC	78 (24.4)	87 (21.9)	1.71 (0.99-2.95)	1.73 (1.00-3.00)
		Freq (C)	0.12	0.12		
MSH3	rs26279	AA	160 (48.2)	185 (46.1)	1.0 (Ref)	1.0 (Ref)
		AG	143 (43.1)	171 (42.6)	0.85 (0.56-1.27)	0.84 (0.55-1.26)
		GG	29 (8.7)	45 (11.2)	0.73 (0.38-1.42)	0.73 (0.37-1.42)
		AG/GG	172 (51.8)	216 (53.8)	0.83 (0.56-1.25)	0.82 (0.55-1.24)
		Freq(G)	0.30	0.33		
	rs184967	GG	236 (71.7)	285 (70.4)	1.0 (Ref)	1.0 (Ref)
		GA	82 (24.9)	111 (27.4)	0.76 (0.49-1.19)	0.76 (0.49-1.19)
		AA	11 (3.3)	9 (2.2)	1.34 (0.50-3.57)	1.34 (0.50-3.58)
		GA/AA	93 (28.2)	120 (29.6)	0.80 (0.52-1.24)	0.80 (0.51-1.23)
		Freq(A)	0.16	0.16		
	rs863221	TT	149 (44.5)	179 (44)	1.0 (Ref)	1.0 (Ref)
		TG	155 (46.3)	194 (47.7)	0.89 (0.62-1.29)	0.88 (0.60-1.27)
		GG	31 (9.3)	34 (8.4)	1.04 (0.51-2.10)	1.02 (0.50-2.07)
		TG/GG	186 (55.6)	228 (56.1)	0.90 (0.63-1.30)	0.89 (0.62-1.28)
		Freq(G)	0.32	0.32		
MSH4	rs5745325	GG	163 (49.4)	199 (49)	1.0 (Ref)	1.0 (Ref)
		GA	131 (39.7)	159 (39.2)	1.01 (0.67-1.53)	0.99 (0.65-1.50)
		AA	36 (10.9)	48 (11.8)	0.94 (0.47-1.89)	0.89 (0.44-1.80)
		GA/AA	167 (50.6)	207 (51)	1.0 (0.67-1.51)	0.98 (0.65-1.48)
		Freq(A)	0.31	0.31		
	rs5745549	GG	292 (89.8)	363 (89.6)	1.0 (Ref)	1.0 (Ref)
		GA	33 (10.2)	42 (10.4)	1.01 (0.50-2.05)	1.07 (0.53-2.18)
		AA	0 (0)	0 (0)		
		GA/AA			1.01 (0.50-2.05)	1.07 (0.53-2.18)
		Freq(A)	0.05	0.05		
MSH6	rs3136228	TT	164 (50.8)	191 (47.5)	1.0 (Ref)	1.0 (Ref)
		TG	118 (36.5)	171 (42.5)	0.83 (0.53-1.31)	0.83 (0.53-1.32)
		GG	41 (12.7)	40 (10)	1.44 (0.67-3.08)	1.44 (0.67-3.09)
		TG/GG	159 (49.2)	211 (52.5)	0.88 (0.56-1.37)	0.88 (0.56-1.38)
		Freq(G)	0.31	0.31		
MutY	rs3219489	CC	168 (50.8)	211 (52.4)	1.0 (Ref)	1.0 (Ref)
		CG	127 (38.4)	162 (40.2)	1.05 (0.69-1.58)	1.02 (0.67-1.54)
		GG	36 (10.9)	30 (7.4)	2.19 (1.08-4.44)	2.23 (1.10-4.52)
		CG/GG	163 (49.3)	192 (47.6)	1.16 (0.78-1.73)	1.14 (0.76-1.70)
		Freq(G)	0.30	0.28		

[a] Adjusted for age at blood draw and smoking status

Table 5. Association between combined MSH2 and MutY genotype and breast cancer risk

		Affected Sister No. (%)	Unaffected Sister No. (%)	OR (95% CI)	OR[a] (95% CI)
MSH2_rs2303428	MutY_rs3219489				
TT	CC	125 (39.3)	170 (43.1)	1.0 (Ref)	1.0 (Ref)
TT	CG_GG	114 (35.8)	138 (35.0)	1.20 (0.75-1.91)	1.19 (0.74-1.89)
TC_CC	CC	35 (11.0)	39 (9.90)	1.65 (0.79-3.47)	1.70 (0.81-3.60)
TC_CC	CG_GG	44 (13.8)	47 (11.9)	2.03 (1.04-3.95)	2.02 (1.03-3.94)

[a] Adjusted for age at blood draw and smoking status

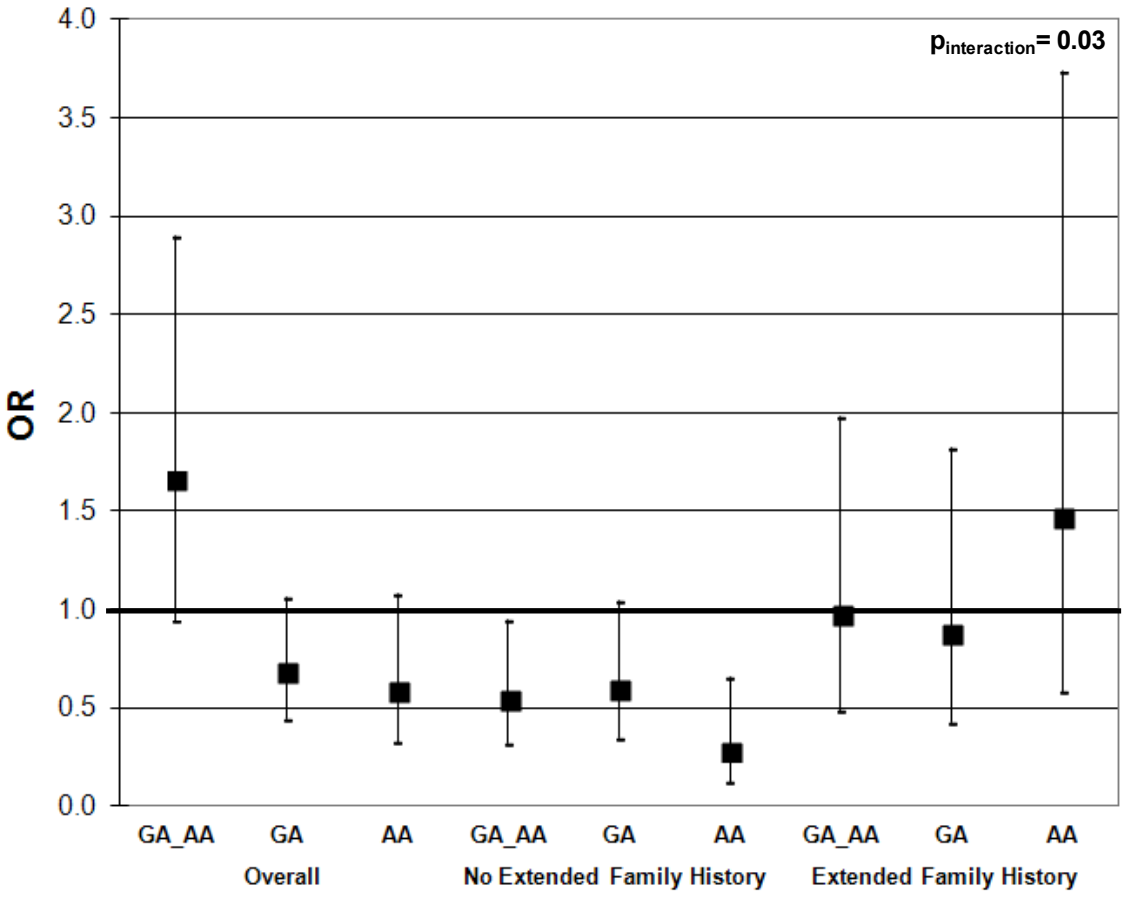


Figure 3. MLH3_rs175080 genotyping analysis stratified by extended family history
A protective effect due to the variant allele is observed among women without an extended family history of breast cancer.

women without an extended family history of breast cancer, a protective effect due to the MLH3_rs175080 variant allele was observed (OR=0.28, 95% CI = 0.12-0.66) (Figure 3). This protective effect was not observed among women with an extended family history of breast cancer ($p_{interaction}$ = 0.03).

Discussion

In this study, an association was observed between the risk of breast cancer and the presence of two MMR-related SNPs, MSH2_rs2303428 and MutY_rs3219489.

MSH2 is part of the MSHα heterodimer that initiates MMR upon recognition of post-replicative errors. This heterodimer consists of MSH2 complexed with either MSH6 to identify base mismatches and small IDLs or with MSH3 to identify larger IDLs. Similarly, MutY is also involved in the recognition of mismatches, specifically the mispairing between adenine and 8-Oxoguanine. The subsequent removal of the mispaired adenine by MutY followed by the correct replacement of the adenine with a cytosine allows the initiation of BER to remove the oxidized base and replace it with an intact guanine. As such, MutY is classified as a member of both MMR and BER.

Deficiencies in MMR have most often been reported in association with the Lynch syndrome family of cancers, while no clear link between any type of cancer and reduced BER capacity has thus far been established. Though few studies implicate either MMR or BER with breast cancer, one hospital based case-control study conducted in a Portuguese population has shown an association between SNPs in the MMR pathway and breast cancer risk[164]. Similar to our study, a main effects finding associated the MLH3_rs175080 variant with a decrease in risk of breast cancer among women without an extended family history. Contrary to our study, however, an increase in risk due to either MSH2_rs2303428 or MutY_rs3219489 variants was not observed in the study reported by Conde et al. This is likely due to the inherent differences in the populations interrogated by the respective studies. While our study selected high risk families, the Portuguese study focused on cases and unrelated controls without a family history of breast cancer.

Although the variants identified in the current study have not been directly implicated with breast cancer, both MSH2_rs2303428 and MutY_rs3219489 have been associated with other cancer sub-types. MSH2_rs2303428 was shown to be associated with gastric cancer in a case-control study conducted within a Chinese population[165]. Upon stratifying the study population, the observed effect was restricted to those with a family history of gastric cancer. However, unlike our study, the strength of the association was elevated among those with onset of the disease prior to age 50. Fewer studies

have implicated a role for MutY_rs3219489 in cancer. However, it has most recently been associated with rectal cancer in a colorectal case-control study conducted in Sweden [166].

While the MutY_rs3219489 SNP is located in exon 12 and results in a non-synonymous amino acid change, the MSH2_rs2303428 SNP is located in the intron region between exons 12 and 13. Based on computational analysis and functional assays, this polymorphism has been shown to be located near a splice acceptor site at the exon-intron boundary, and the presence of the variant leads to partial exon 13 skipping, resulting in the translation of an altered protein from the alternatively spliced message[167].

Due to the low frequency of homozygous recessive individuals for the MSH2_rs2303428 SNP, we were not able to assess the risk of breast cancer due to the presence of two variant alleles. However, the borderline association observed in the presence of one or more variant alleles suggests the presence of this allele functions through a dominant model. Hence, the deleterious effect imparted by just one variant allele may also account for the low frequency of individuals with two copies of the variant.

While an association with breast cancer risk was observed due to the presence of two MutY_rs3219489 variant alleles, the effect was lost upon combining heterozygous and homozygous carriers, suggesting this variant acts through a recessive model. This is in agreement with the findings from the Swedish colorectal case-control study, in which the association observed between the presence of two variant MutY_rs3219489 alleles and rectal cancer was lost upon combining heterozygous and homozygous carriers of the allele[166].

In conclusion, the increased risk in breast cancer observed in our family-based case control study due to alleles typically associated with Lynch syndrome associated cancers suggests that while polymorphisms in MMR have thus far not been associated with sporadic breast cancer, deficiencies in this pathway may be relevant in familial breast cancer.

There are several limitations to the candidate SNP approach utilized in the current study. Numerous SNPs exist across genes involved in the MMR pathway. Hence, although an association with breast cancer was observed with several SNPs, the subset of SNPs selected in this study may have excluded SNPs with potentially greater relevance for breast cancer. Restricting the focus on a choice few targets also results in examining the impact of each SNP in isolation. This approach does not take into

account the possibility that the presence of other SNPs may magnify or counteract the impact of the SNP under study. Due to the small sample size of our study, we did not have enough power to assess possible interactions among the variants interrogated. Even if these interactions could have been investigated, we would not have been able to account for the impact of SNPs not interrogated in our study. This can be addressed by genome-wide association studies (GWAS), which, unlike candidate SNPs studies, can not only query the entire genome for all SNPs of relevance, but can also examine the impact of each SNP in the context of all other known variants across the genome. However, the magnitude of the association observed with SNPs identified using this hypothesis-free driven approach has generally been small in size and findings have often failed to be replicated. Furthermore, the variants identified in GWAS studies are often of unknown function, complicating the inferences that can be drawn from such findings. This is in contrast to candidate SNP studies, as selection of SNPs is informed based on known functional implications of targets in genes and pathways of relevance, facilitating the mechanistic understanding underlying the observed association.

Another area of concern is the use of a case-control study design to assess the association between genetic variants and the outcome of interest. The underlying assumption of this approach is that observed differences in allele frequencies are related to differences in case-status. However, differences in allele frequencies between populations can also arise due to differences in ancestral patterns and mating practices, among other factors. If disease status is also differentiated by these factors, the presence of population admixture in the study population can introduce the potential for confounding. Hence a major strength of the current study is the ability to minimize this type of confounding through the use of a family-based case-control study, ensuring that cases and controls stem from a genetically homogenous population.

Still, the aforementioned limitations restrict the inferences that can be drawn from the current study. Hence, replication of the reported findings in other study populations is warranted.

Chapter IV. DNA repair gene expression levels as indicators of breast cancer risk

Abstract

Reduced DNA repair capacity contributes to genomic instability, a hallmark of carcinogenesis. The link between DNA repair capacity and cancer is especially relevant in breast cancer, where women with suboptimal repair capacity have been reported to have up to a five-fold increase in breast cancer risk compared to women with greater repair capacity. We investigated the expression level of several DNA repair-related genes and their association with breast cancer risk among participants of the NY-BCFR. We examined RNA from viable mononuclear cells in 213 sisters (N=511 women). A total of five genes, *ATM, BRCA1, MSH2, MUTYH* and *XPC*, were selected and assayed using Taqman gene expression kits from Applied Biosystems. Using generalized estimating equations (GEE), we found that individuals in the lowest tertile of expression for *ATM,* a gene involved in DNA double-strand break repair, and *MSH2, a* gene involved in mismatch repair*,* had significantly higher odds of breast cancer compared to individuals in the highest quartile of expression, adjusting for age at blood draw and smoke status (OR=1.51, 95% CI=1.03, 2.20) and (OR=2.12, 95% CI=1.34, 3.34), respectively. Individuals with reduced expression of both genes had an over 3-fold increase in breast cancer risk, adjusted for age at blood draw and smoking status (OR=3.85, 95% CI=1.69, 8.77). Upon stratifying the GEE model, the observed risk due to reductions in ATM expression level was limited to women with an extended family history of breast cancer ($p_{interaction}$ = 0.04). Furthermore, an increase in risk due to reductions in XPC expression was observed among women with an age at menarche after age 13 (OR=2.07, 95% CI=1.21, 3.54). Finally, the observed association between reductions in ATM expression level and breast cancer risk is no longer significant after incorporating previously determined DSBR capacity of either sticky-end damage substrates (OR=1.28, 95% CI=0.15, 11.2) or blunt-end damage substrates (OR=1.55, 95% CI=0.15, 15.5), substantiating that deficiencies in ATM functionality impacts breast cancer risk through its role in DSBR. These findings suggest that reduced expression of these genes compromise DNA repair capacity and, thereby, increase breast cancer susceptibility.

Introduction

The importance of DNA repair in breast tumorigenesis has been most clearly demonstrated by the identification of *BRCA1* and *BRCA2*, genes now known to be involved in DSBR, as contributors to the familial presentation of breast cancer[130]. In addition to these genes a handful of other less penetrant breast cancer susceptibility genes have been identified, including *ATM*, a gene also involved in DSBR. Women who have inherited mutations in either *BRCA1* or *BRCA2* have up to an 80% likelihood of developing breast cancer by age 75, while women who are carriers of mutations in *ATM* are at a 2-fold increased risk of breast cancer compared to non-carriers[131,132]. Due to the severity of the deleterious effects associated with mutations in these genes, defects leading to a complete loss of function of such critical genes occur at low frequencies and make minor contributions towards the attributable risk observed within a population. Still, these findings highlight the importance of DSBR capacity in breast cancer susceptibility. A number of investigations have suggested a possible role of other repair pathways as well. Several studies have reported on the presence of microsatellite instability (MSI), a phenotype commonly associated with deficiencies in MMR, in tumors of breast cancer cases[115,133]. Deficiencies in overall NER capacity, as assessed by various phenotyping assays, has also been implicated in contributing to breast cancer risk[108,109]. These findings suggest that determination of DNA repair capacity may serve to identify women at elevated risk for breast cancer for targeted screening and intervention. While promising, translating these findings for such a purpose is limited by the methods of assessment currently available to indicate repair capacity. Validation studies in larger populations and subsequent development and application of a screening tool within a population-wide setting demand a means to detect inherent variability in repair within easily accessible surrogate biospecimens using a scalable methodology. The extent of manipulation required to carry out phenotyping assays for the global assessment of repair, limits their utility in this desired context. Validated high-throughput technologies geared towards more specific endpoints within various repair pathways, including genotyping of SNPs, have been applied in population studies. However, findings based on these studies have been mixed, with positive studies capturing only a small proportion of the observed effect sizes associating phenotypic assessments of repair capacity with breast cancer risk. This limitation of SNP studies, resulting from the inherently restrictive focus on individual factors within

a pathway, is further magnified by the fact that even within the context of individual genes, a variety of upstream factors other than genetic alterations, including epigenetic marks and transcriptional modifications, can impact the activity of the specific component of interest. Hence, a more high-throughput means of assessing repair than is feasible through phenotyping while also more comprehensive than genotyping may be afforded by determining expression levels of genes participating in DNA repair pathways. In addition to the potential development of a risk biomarker, findings based on such intermediary targets can also be of etiologic consequence, filling in the gaps in current knowledge by linking specific upstream endpoints and repair phenotypes known to be associated with breast cancer.

In this study, we assess the association between breast cancer risk and the expression levels of 5 genes, *ATM*, *BRCA1*, *MSH2*, *MUTYH*, and *XPC*, known to be involved in various DNA repair pathways in PBMCs from sister-sets enrolled in the NY-BCFR.

Materials and Methods

Expression in stimulated vs. unstimulated PBMCs

14ml of blood was collected in BD vacutainers from 5 control individuals. The blood was immediately processed to isolate PBMCs via a Ficoll gradient. A subset of the cells consisting of $5x10^5$ cells were pelleted and frozen at -140°C in RPMI 1640 media supplemented with 40% fetal bovine serum (FBS) and 15% dimethyl sulfoxide (DMSO), while a second subset of $2x10^6$ cells was grown in culture in RPMI 1640 media supplemented with 15% FBS, 1% Pen-strep and phytohemagglutinin (PHA, 150uL/10mL; 5ug/mL) for 52 hr, then frozen as for un-stimulated cells. At the time of assay, both PHA-stimulated and un-stimulated mononuclear cells were thawed, pelleted, washed with phosphate buffered saline (PBS), and counted. RNA was subsequently extracted using an RNeasy Mini kit (Qiagen, Valencia, CA).

250ng of RNA was reverse transcribed using a High Capacity cDNA reverse transcription kit (Life Technologies, Carlsbad, CA). The reaction consisted of 10X RT buffer, 25X dNTP Mix (100mM), 10X RT random primers, and MultiScribe Reverse Transcriptase. An RT (-) control, in which the reaction mixture included a randomly selected RNA template and all RT reagents with the exception of the reverse transcription enzyme, was included as a means to assess the presence of genomic DNA contamination. The thermocycling conditions followed 25°C for 10min, 37°C for 120min, and 85°C for 5min. Level of β-actin expression was compared between stimulated and un-stimulated cells at varying concentrations of 2.5, 5, and 10ng in duplicate. Expression of β-actin was assessed using pre-designed Taqman assays: Hs01112347_m1 (ATM), Hs01556194_m1 (BRCA1), Hs00953523_m1 (MSH2), Hs01014856_m1 (MUTYH), Hs00190295_m1 (XPC) (Life Technologies). In addition to the cDNA template, the reaction mixture also consisted of 20X Gene expression assay and 2X Taqman Universal Master Mix. The thermocycling conditions including an initial incubation at 95°C for 10min, followed by a 95°C for 15sec and 60°C for 1min, repeated for 40 cycles. Fluorescence signal of β -actin was determined using a FAM-reporter labeled probe.

RNA isolation and reverse transcription

PBMCs were thawed from -140°C. DNA and RNA were extracted from the cells using an All-prep DNA/RNA 96 kit (Qiagen). The plates were designed such that complete sister-sets were present on the same plate, while investigators remained blinded to case-control status. Purified RNA and DNA were quantified using a Nanodrop 2000 (Thermo Scientific, Pittsburgh, PA).

60ng of extracted RNA was reverse transcribed as above. Each reverse transcribed plate also included an RT (-) control as described above.

Gene expression

Gene expression assays were run with plates set up to include 2.5ng template per sample, run in duplicate for each target assay and β-actin endogenous control. Each assay was also run with cDNA derived from commercially purchased human total RNA (Life Technologies) as a calibrator sample, an RT (-) control, and a non-template control. Relative expression was determined using the $2^{-\Delta\Delta Ct}$ method.

Study population

The study participants consisted of 218 sister-sets (n=569) with available viable PBMCs enrolled at the NY- BCFR (Figure 1).

Statistical methods

Pearson's chi-square tests were performed to assess the distribution of the study population according to selected measured variables. Student's t-test was conducted to compare mean case-control differences in relative expression level as a continuous variable. Linear regression models were conducted to assess the impact of breast cancer risk factors on gene expression level on a continuous scale, and logistic regression models were conducted to assess the impact of these factors on differentiating women in the lowest tertile of expression from women with higher levels of expression. Breast cancer risk factors that were included in the analyses included age at blood draw, smoking status (ever vs. never), BMI (≥25kg/m^2 vs. <25kg/m^2, cut-off based on WHO definition of normal range), age at menarche (>13 vs. ≤13 years, based on median age at menarche among unaffected sisters),

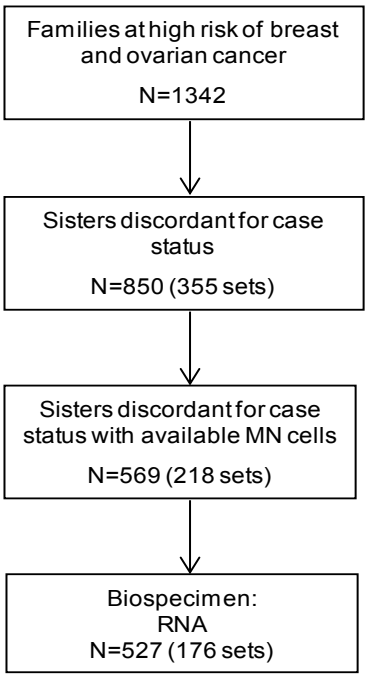


Figure 1. Gene expression study design
The study was conducted among sisters discordant for disease with RNA available from MN cells.

age at first parity (>30 vs ≤30 years, nulliparous women were assigned age at blood draw), and extended family history (>1 vs 1). Conditional logistic regression models and generalized estimating equations were run to assess the associations between repair gene expression level. Risk was determined evaluating expression level as a continuous variable, with a change in risk assessed due to a 1 unit decrease in expression level, and evaluating expression level partitioned into tertiles with cut-offs based on controls. To adjust for potential confounding, additional known risk factors for disease, including age at blood draw and smoking status, were included in the models. Effect modification was assessed to determine whether extended family history impacted the association between lowest tertile of expression in the assayed genes and breast cancer risk. Spearman correlations were assessed between ATM expression levels and previously determined End-joining capacity of EcoRI-generated and HincII-generated damage substrates. The presence of mediation due to ATM expression level on the association between end-joining capacity and breast cancer risk was analyzed using conditional

logistic regression. All statistical analyses were conducted using R: A Language and Environment for Statistical Computing (2011), R Foundation for Statistical Computing, Vienna, Austria.

Results

Selection of candidate genes

Candidate genes were selected based on information available regarding DNA repair-related findings in studies conducted within the NY-BCFR and otherwise reported in the literature. Within the BCFR, DNA repair-related breast cancer risk assessments spanning from high and low penetrant genetic alterations to overall phenotype have been conducted (Figure 2). These findings were used in the current study to identify alternative intermediate endpoints that may account for a greater amount of variability in repair capacity than could be attributed to the previously investigated genetic components in the pathway while also providing greater scalability than the identified phenotypic endpoints. Final selection of targets (Figure 3) was also informed based on studies in other breast cancer-relevant populations reported in the literature.

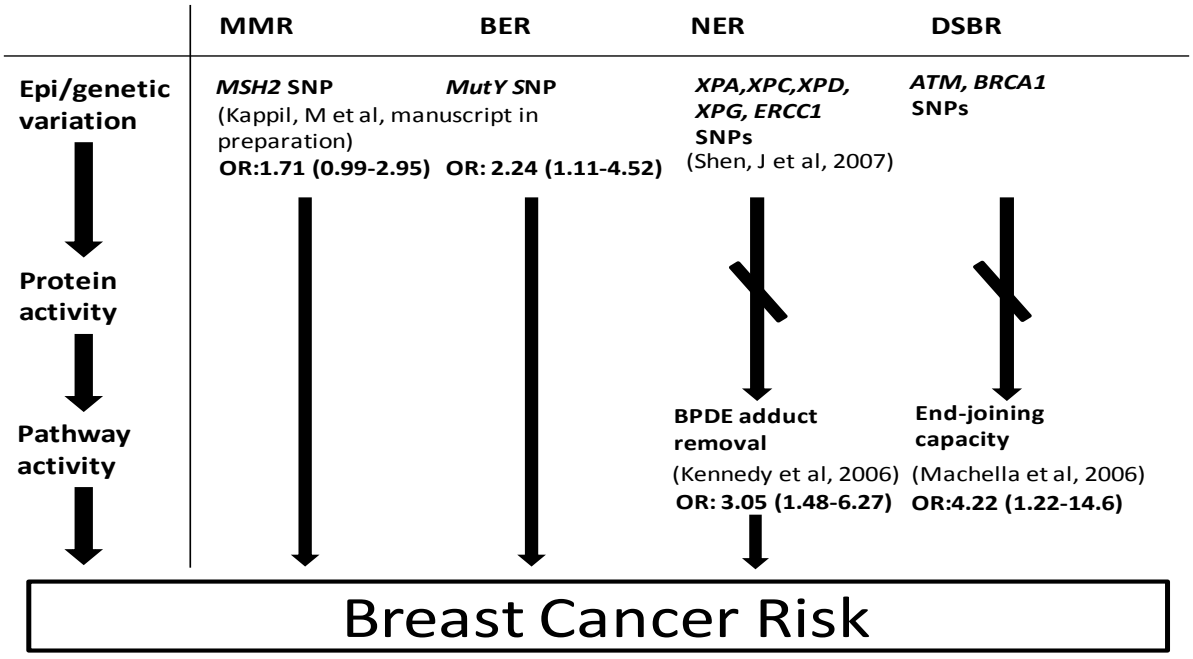


Figure 2. DNA repair-related findings within the NY-BCFR. Deficiencies in all four repair pathways have been implicated with breast cancer risk in this population. SNP genotyping studies have implicated genes involved in MMR and BER with breast cancer risk. Phenotyping assays measuring NER and DSBR have suggested suboptimal capacity in both pathways are associated with breast cancer risk in this population. However genotyping of SNPs in genes in these pathways have failed to reveal causal genetic determinants driving the observed association.

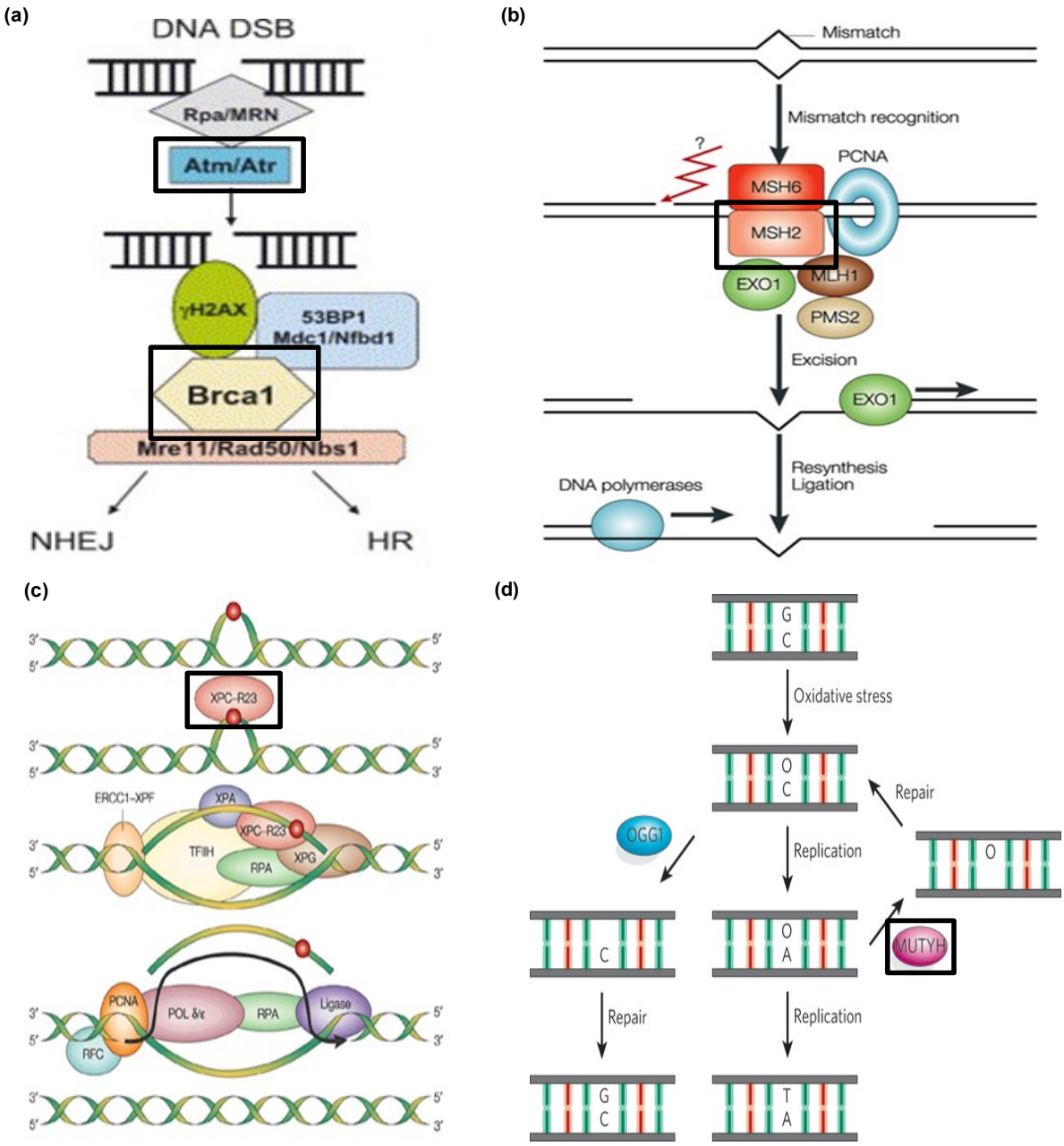


Figure 3. Candidate genes chosen for analysis (a) DSB pathway: ATM transduces the signal of DSB damage to downstream components of the pathway. BRCA1 acts as part of a scaffolding complex tethering participants of the pathway together. (b) MMR pathway: MSH2 takes part in the mismatch recognition complex. (c) NER pathway: XPC participates in the helix-distortion recognition complex. (d) BER pathway: MUTYH recognizes and removes adenines in adenine:8oxoguanine mispairs. Reprinted from Ting and Lee, 2004, Martin and Scharff, 2002, Friedberg, 2001, and David et al, 2007.

Biospecimen considerations

As the current gene expression study was conducted in PBMCs, we first investigated the need to accommodate biospecimen-specific adjustments to the existing standardized protocol for the

assessment of gene expression levels. For this purpose, a number of assay parameters were determined prior to commencing the study. Since isolated cells likely have limited protein activity, we first assessed the need to reactivate cells through mitogen stimulation to attain detectable levels of transcriptional activity.

While level of activity was higher in stimulated cells, as observed by the earlier threshold cycle, detectable levels of activity were also observed among un-stimulated cells. Furthermore, greater inter-individual variation in expression level was observed among the un-stimulated cells, suggesting that stimulation may artificially homogenize transcriptional activity (Figure 4). As a result, all further experiments were conducted in un-stimulated cells, eliminating the need to culture cells prior to analysis, and, thereby, streamlining the process.

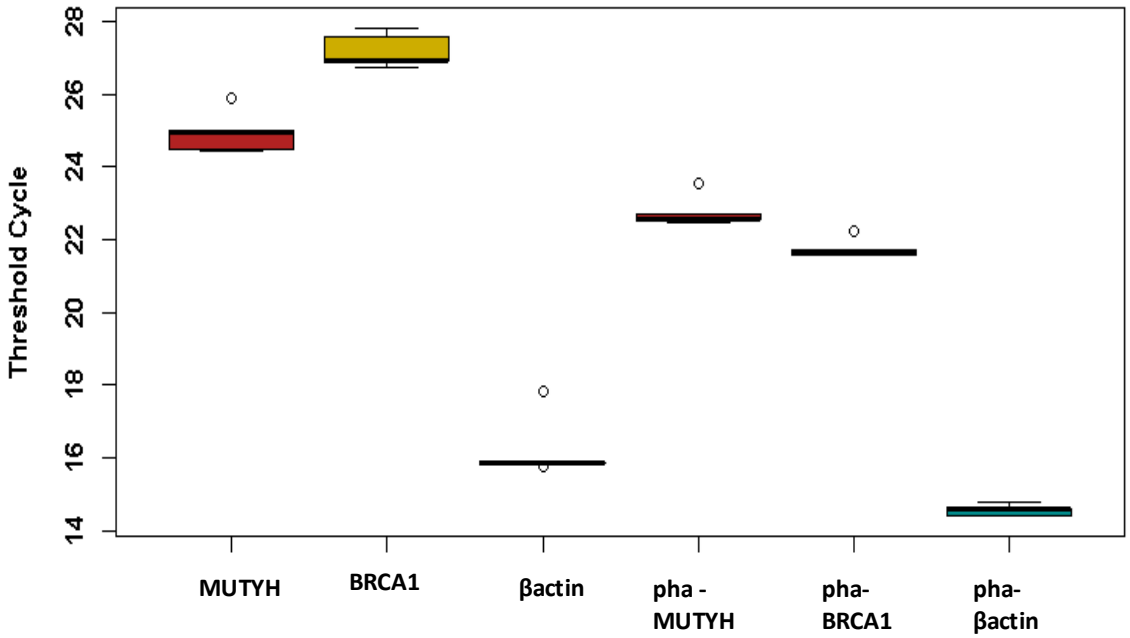


Figure 4. Expression levels of βactin in unstimulated and pha-stimulated PBMCs
Level of expression was assessed within five individuals stemming from the same sister-set.

An additional area of concern in regards to this cell type is ensuring that an appropriate housekeeping gene is selected as the endogenous control. Housekeeping genes typically selected for this purpose in gene expression studies include *βactin* and *GAPDH*. However, investigators have disputed the relevance of these genes as endogenous controls in studies involving PBMCs as these genes may not be uniformly expressed at steady-state levels in this cell population[134]. It has been suggested that

RPLO may be a more appropriate choice. To test this hypothesis, we compared expression levels of *βactin*, *GAPDH*, and *RPLO* in stimulated and unstimulated cells in our investigation. *RPLO*, while expressed consistently across samples with equivalent RNA template amounts, presented at levels too low for reliable implementation as an endogenous control. Hence, as *βactin* was expressed at the highest levels among the tested genes, this gene was selected as endogenous control (Figure 5).

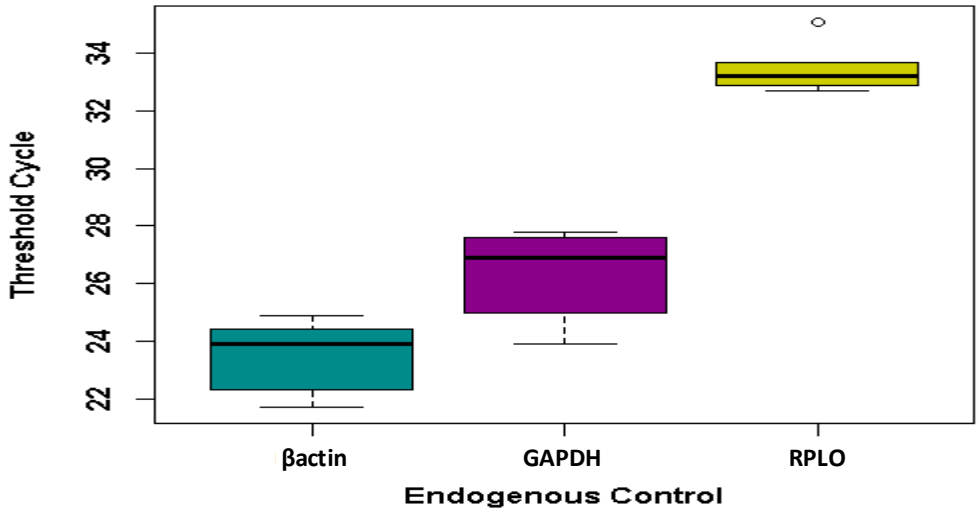


Figure 5. Comparison of endogenous control expression level in PBMCs
Level of expression was assessed within five individuals stemming from the same sister-set.

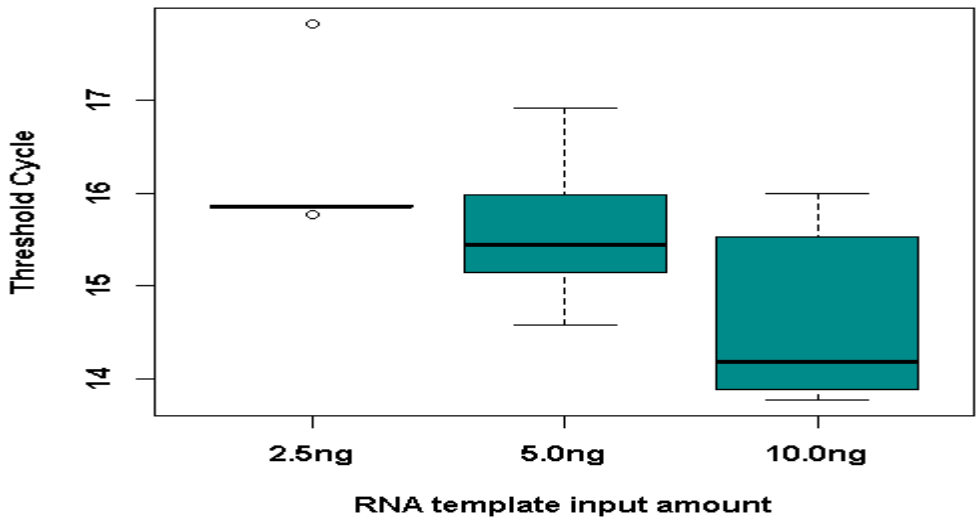


Figure 6. Expression levels of βactin at various template concentrations
Level of expression was assessed within five individuals stemming from the same sister-set.

Finally, limitations in availability of viable PBMCs prompted us to ascertain the limits of input RNA amounts feasible for detection. As expected, higher input amounts led to higher levels of detection. However, as the lowest tested RNA amount of 2.5ng also yielded detectable levels of RNA, this input template amount was selected for the study (Figure 6).

Study population characteristics

The distributions of various risk factors known to be relevant to breast cancer are shown in Table 1. No significant differences between cases and controls were observed based on age at blood draw, BRCA1/2 mutation status, ethnicity, smoke status, BMI, age at menarche or parity.

Out of the five genes assayed, a significant difference between cases and control was observed in mean levels of *ATM* expression (Figure 7). Differences in gene expression level were also observed based on additional variables (Table 2). Significantly lower mean ATM expression level was observed in cases compared to controls among women with a history of smoking, women with BMI below 25 kg/m^2, age at first parity < 30 years of age, and borderline lower levels in cases compared to controls were observed among women with an age at menarche ≥ 13 years.

While no overall case-control differences were observed for the other assayed genes, case-control differences were observed within certain sub-populations for several of the genes. Cases had lower mean expression levels of *MSH2* compared to controls among women above age 50 and among smokers. Similarly, within cases, smokers had lower expression of *MSH2* compared to non-smokers. No case control differences in expression level were observed for *BRCA1, MUTYH* or *XPC*. However, among unaffected women, there was a borderline decrease in mean expression levels of *MUTYH* among women with an age at menarche prior to age 13 compared to women with a later age at menarche, and mean expression level of *XPC* was significantly higher among smokers than non-smokers.

The observed differences in mean expression level based on the selected variables suggest that these risk factors may impact the level of gene expression (Table 3). Unaffected women with a history of smoking were more likely to be in the lowest tertile of expression for *ATM* and to have lower overall expression of *XPC*. Women with BMI greater than 25kg/m^2 were more likely to be in the lowest tertile

of expression for both *BRCA1* and *MSH2*. Finally, a borderline association was observed for women above age 50 being more likely to be in the lowest tertile of expression for *BRCA1*.

Table 1. Distribution of selected variables among sisters affected and unaffected with breast cancer

Variable	No. cases (%) (n=239)	No. controls (%) (n=298)	p-value[α]
Age at blood draw (in years)			
≤50	142 (59.4)	196 (65.8)	0.14
>50	97 (40.6)	102 (34.2)	
BRCA1 (+) mutation	11 (4.62)	7 (2.36)	0.22
BRCA2 (+) mutation	13 (5.46)	7 (2.36)	0.10
Ethnicity			
Non-hispanic White	127 (54.0)	142 (47.8)	0.32
African American	5 (2.12)	8 (2.69)	
Other	103 (43.8)	147 (49.5)	
Smoking Status			
Never	143 (59.8)	186 (62.4)	0.60
Ever	96 (40.2)	112 (37.6)	
BMI (kg/m^2)			
<25	114 (47.7)	142 (48.3)	0.96
≥25	125 (52.3)	152 (51.7)	
Age at menarche (in years)			
<13	109 (46.0)	127 (42.6)	0.49
≥13	128 (54.0)	171 (57.4)	
Age at first parity			
<30	151 (63.2)	201 (67.4)	0.35
≥30	88 (36.8)	97 (32.6)	
Family History			
=1	137 (57.3)	193 (64.8)	0.09
>1	102 (42.7)	105 (35.2)	

[α] Pearson's X^2 test

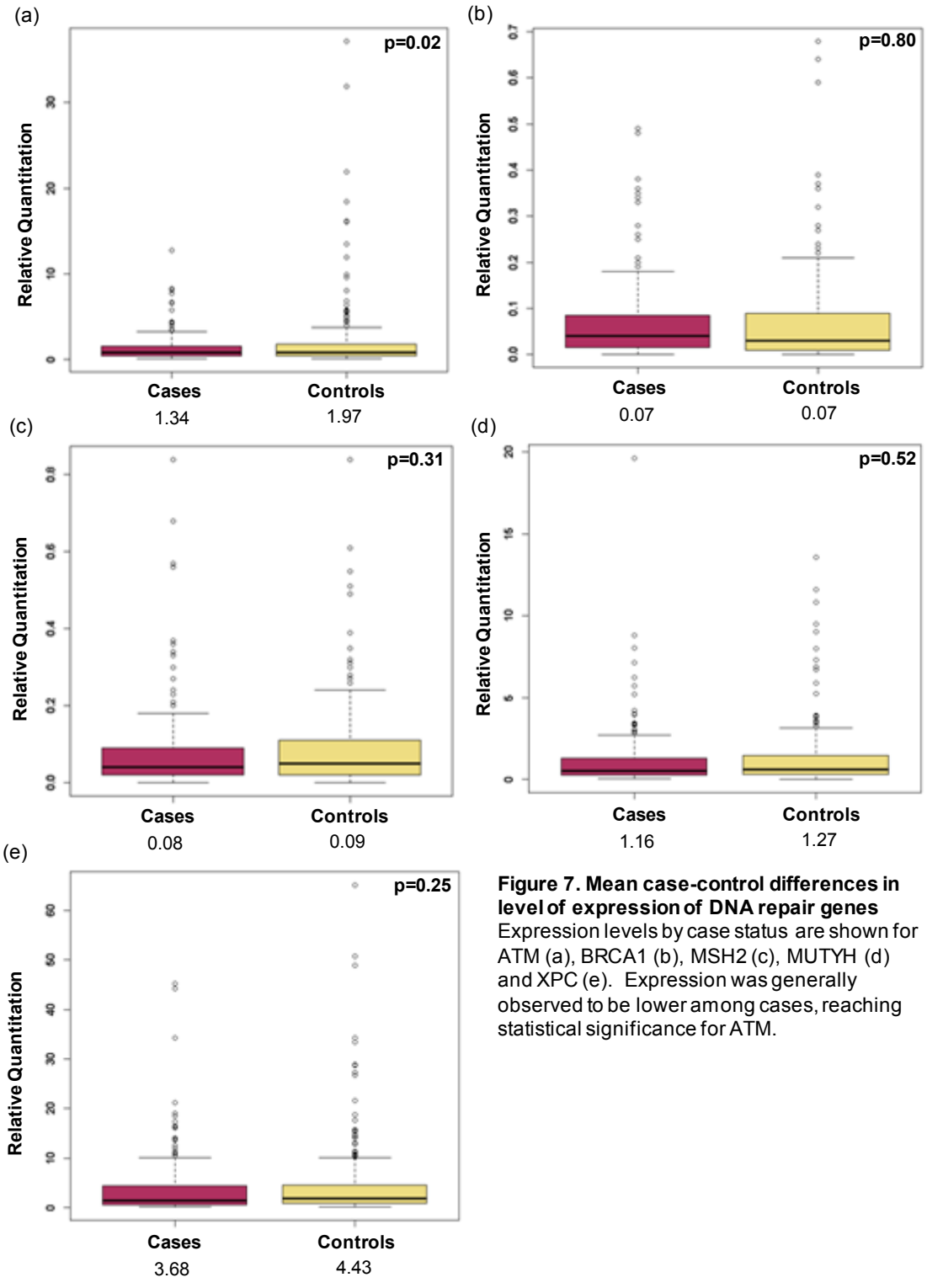


Figure 7. Mean case-control differences in level of expression of DNA repair genes
Expression levels by case status are shown for ATM (a), BRCA1 (b), MSH2 (c), MUTYH (d) and XPC (e). Expression was generally observed to be lower among cases, reaching statistical significance for ATM.

Table 2. Expression level of DNA repair genes among sisters affected and unaffected with breast cancer

Variable	ATM					BRCA1					MSH2					MUTYH					XPC				
	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]
	N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD	
All	196	1.34 ± 1.64	255	1.97 ± 4.02	**0.02**	135	0.08 ± 0.10	171	0.08 ± 0.11	0.37	171	0.08 ± 0.12	207	0.10 ± 0.12	0.22	193	1.16 ± 1.92	252	1.28 ± 1.87	0.47	217	3.68 ± 6.07	251	4.44 ± 7.88	0.24
Age at blood draw																									
≤50	117	1.25 ± 1.39	164	1.62 ± 2.51	0.12	84	0.08 ± 0.10	105	0.08 ± 0.11	0.97	95	0.10 ± 0.14	133	0.09 ± 0.11	0.94	113	0.93 ± 1.19	165	1.26 ± 1.74	0.07	126	3.80 ± 6.14	164	4.52 ± 8.14	0.39
>50	78	1.48 ± 1.95	90	2.61 ± 5.80	0.08	50	0.09 ± 0.11	65	0.08 ± 0.11	0.86	75	0.06 ± 0.09	73	0.10 ± 0.13	**0.04**	79	1.47 ± 2.61	86	1.33 ± 2.09	0.69	90	3.52 ± 6.01	86	4.30 ± 7.42	0.44
p[δ]		0.39		0.12			0.64		0.76			0.08		0.70			0.09		0.79			0.74		0.83	
Smoking Status																									
Never	115	1.41 ± 1.85	155	1.64 ± 3.14	0.44	83	0.08 ± 0.11	106	0.08 ± 0.09	0.80	106	0.10 ± 0.14	130	0.09 ± 0.10	0.50	115	1.23 ± 2.24	152	1.27 ± 1.94	0.87	127	3.61 ± 5.65	157	3.51 ± 5.43	0.88
Ever	80	1.25 ± 1.28	99	2.49 ± 5.08	**0.02**	51	0.09 ± 0.09	64	0.09 ± 0.14	0.93	64	0.06 ± 0.06	76	0.11 ± 0.13	**<0.01**	77	1.04 ± 1.31	99	1.30 ± 1.76	0.27	89	3.79 ± 6.66	93	6.01 ± 10.7	0.09
p[δ]		0.49		0.14			0.75		0.56			**0.01**		0.13			0.47		0.92			0.83		**0.04**	
BMI (kg/m^2)																									
<25	92	1.45 ± 1.88	114	2.49 ± 4.93	**0.04**	67	0.08 ± 0.09	73	0.09 ± 0.13	0.77	78	0.08 ± 0.10	88	0.09 ± 0.12	0.39	93	1.39 ± 2.44	118	1.45 ± 2.27	0.87	106	4.00 ± 5.61	111	5.23 ± 9.49	0.24
≥25	103	1.25 ± 1.38	137	1.57 ± 3.08	0.28	67	0.08 ± 0.11	95	0.08 ± 0.10	0.65	92	0.09 ± 0.13	116	0.10 ± 0.11	0.36	99	0.93 ± 1.22	130	1.14 ± 1.42	0.23	110	3.37 ± 6.49	137	3.82 ± 6.32	0.59
p[δ]		0.41		0.08			0.86		0.59			0.69		0.63			0.10		0.21			0.45		0.18	
Age at Menarche																									
<13	92	1.33 ± 1.53	107	1.74 ± 3.71	0.30	68	0.08 ± 0.11	73	0.07 ± 0.10	0.54	81	0.09 ± 0.14	87	0.09 ± 0.10	0.84	89	1.20 ± 1.58	110	1.03 ± 1.25	0.28	98	3.89 ± 6.63	112	4.70 ± 7.46	0.41
≥13	102	1.36 ± 1.74	148	2.13 ± 4.22	**0.05**	65	0.08 ± 0.10	98	0.09 ± 0.12	0.84	88	0.07 ± 0.10	120	0.10 ± 0.13	**0.05**	102	1.13 ± 2.18	142	1.47 ± 2.21	0.23	117	3.54 ± 5.61	139	4.21 ± 8.21	0.44
p[δ]		0.90		0.45			0.97		0.43			0.22		0.48			0.79		**0.05**			0.67		0.62	
Age at First Parity																									
<30	124	1.27 ± 1.52	175	2.06 ± 4.49	**0.03**	85	0.08 ± 0.11	120	0.08 ± 0.11	0.95	110	0.08 ± 0.11	147	0.10 ± 0.12	0.09	121	1.19 ± 2.22	169	1.22 ± 1.77	0.90	136	3.67 ± 6.44	175	4.52 ± 8.00	0.30
≥30	71	1.48 ± 1.84	80	1.77 ± 2.71	0.44	49	0.08 ± 0.09	51	0.07 ± 0.11	0.78	60	0.09 ± 0.13	60	0.09 ± 0.10	0.73	71	1.09 ± 1.27	83	1.39 ± 2.04	0.27	80	3.70 ± 5.44	76	4.22 ± 7.62	0.63
p[δ]		0.39		0.51			0.76		0.59			0.38		0.36			0.69		0.51			0.98		0.78	
Family History																									
=1	113	1.38 ± 1.68	167	2.00 ± 4.09	0.08	82	0.08 ± 0.10	114	0.09 ± 0.12	0.78	93	0.09 ± 0.13	136	0.10 ± 0.12	0.55	111	1.01	162	1.36	0.07	121	3.51 ± 4.87	162	4.88 ± 9.05	0.10
>1	82	1.30 ± 1.59	88	1.91 ± 3.89	0.17	52	0.08 ± 0.10	57	0.07 ± 0.08	0.39	77	0.07 ± 0.10	71	0.09 ± 0.10	0.28	81	1.35	90	1.12	0.50	95	3.90 ± 7.35	89	3.61 ± 5.04	0.75
p[δ]		0.74		0.87			0.91		0.26			0.32		0.54			0.28		0.29			0.65		0.15	

[δ] P-value determined based on Student's t-test

Table 3. Association between selected variables and DNA repair gene expression levels among unaffected women

	ATM				BRCA1				MSH2				MUTYH				XPC			
Age (>50 vs ≤50)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)
Intercept	1.61	0.31	5.21	<0.01	0.07	0.01	6.87	<0.01	0.09	0.01	9.14	<0.01	1.24	0.14	8.64	<0.01	4.50	0.61	7.35	<0.01
Continuous	1.00	0.52	1.92	0.06	0.01	0.02	0.85	0.40	0.01	0.02	0.65	0.52	0.08	0.25	0.34	0.74	-0.20	1.04	-0.19	0.85
Intercept	-0.81	0.17	-4.83	<0.01	-0.24	0.18	-1.33	0.18	-0.76	0.18	-4.23	<0.01	-0.95	0.17	-5.50	<0.01	-0.93	0.17	-5.39	<0.01
Tertile	-0.21	0.29	-0.73	0.46	-0.61	0.32	-1.93	**0.05**	-0.34	0.32	-1.07	0.28	0.30	0.28	1.07	0.28	0.24	0.29	0.83	0.41
Smoking (Ever vs. Never)																				
Intercept	1.63	0.32	5.12	0.00	0.07	0.01	6.95	<0.01	0.08	0.01	8.48	<0.01	1.26	0.15	8.38	<0.01	3.49	0.62	5.65	<0.01
Continuous	0.86	0.51	1.68	0.09	0.00	0.02	0.28	0.78	0.02	0.02	1.30	0.20	0.04	0.24	0.17	0.87	2.52	1.01	2.49	**0.01**
Intercept	-0.67	0.17	-3.99	0.00	-0.43	0.19	-2.27	0.02	-0.81	0.19	-4.34	<0.01	-0.68	0.17	-4.02	<0.01	-0.72	0.17	-4.27	<0.01
Tertile	-0.59	0.29	-2.01	**0.04**	-0.04	0.30	-0.15	0.88	-0.17	0.31	-0.55	0.58	-0.42	0.29	-1.45	0.15	-0.35	0.29	-1.20	0.23
BMI (kg/m^2) (≥25 vs <25)																				
Intercept	2.47	0.37	6.62	<0.01	0.07	0.01	6.15	<0.01	0.08	0.01	7.14	<0.01	1.44	0.17	8.42	<0.01	5.19	0.74	6.99	<0.01
Continuous	-0.90	0.51	-1.77	0.08	0.00	0.02	0.11	0.91	0.02	0.02	1.18	0.24	-0.30	0.24	-1.28	0.20	-1.37	1.00	-1.37	0.17
Intercept	-0.76	0.20	-3.81	<0.01	-0.13	0.21	-0.63	0.53	-0.55	0.21	-2.65	0.01	-0.77	0.20	-3.92	<0.01	-0.76	0.20	-3.77	<0.01
Tertile	-0.25	0.28	-0.89	0.37	-0.62	0.30	-2.09	**0.04**	-0.61	0.30	-2.03	**0.04**	-0.14	0.27	-0.50	0.62	-0.14	0.28	-0.49	0.62
Menarche (<13 vs ≥13)																				
Intercept	2.13	0.33	6.47	<0.01	0.07	0.01	7.26	<0.01	0.10	0.01	9.36	<0.01	1.46	0.16	9.42	<0.01	4.21	0.67	6.32	<0.01
Continuous	-0.39	0.51	-0.76	0.45	0.00	0.02	-0.27	0.79	-0.01	0.02	-0.59	0.55	-0.42	0.23	-1.80	0.07	0.49	1.00	0.50	0.62
Intercept	-1.04	0.19	-5.56	<0.01	-0.34	0.19	-1.83	0.07	-0.72	0.19	-3.81	<0.01	-0.82	0.18	-4.54	<0.01	-0.99	0.19	-5.20	<0.01
Tertile	0.34	0.28	1.24	0.21	-0.26	0.30	-0.86	0.39	-0.38	0.31	-1.25	0.21	-0.04	0.28	-0.14	0.89	0.31	0.28	1.12	0.26
Age at first parity (>=30 vs. <30)																				
Intercept	2.06	0.30	6.80	<0.01	0.07	0.01	7.98	<0.01	0.10	0.01	10.47	<0.01	1.22	0.14	8.55	<0.01	4.52	0.59	7.61	<0.01
Continuous	-0.30	0.54	-0.55	0.58	-0.01	0.02	-0.60	0.55	-0.02	0.02	-1.10	0.27	0.15	0.25	0.60	0.55	-0.30	1.08	-0.28	0.78
Intercept	-0.95	0.17	-5.66	<0.01	-0.40	0.17	-2.29	<0.01	-1.01	0.18	-5.56	<0.01	-0.93	0.17	-5.48	<0.01	-0.92	0.17	-5.53	<0.01
Tertile	0.20	0.29	0.70	0.49	-0.15	0.32	-0.46	0.65	0.45	0.31	1.45	0.15	0.27	0.29	0.96	0.34	0.25	0.29	0.85	0.39
Family History (>1 vs=1)																				
Intercept	2.00	0.31	6.44	<0.01	0.07	0.01	7.94	<0.01	0.10	0.01	9.89	<0.01	1.35	0.15	9.31	<0.01	4.88	0.62	7.92	<0.01
Continuous	-0.09	0.53	-0.16	0.87	-0.01	0.02	-0.66	0.51	-0.01	0.02	-0.62	0.54	-0.23	0.24	-0.94	0.35	-1.27	1.03	-1.23	0.22
Intercept	-0.75	0.17	-4.52	<0.01	-0.30	0.17	-1.72	0.09	-0.85	0.18	-4.69	<0.01	-0.82	0.17	-4.86	<0.01	-0.87	0.17	-5.09	<0.01
Tertile	-0.43	0.30	-1.43	0.15	-0.46	0.32	-1.44	0.15	-0.04	0.31	-0.14	0.89	-0.04	0.29	-0.14	0.89	0.08	0.29	0.28	0.78

Using conditional logistic regression analysis, an increased risk for breast cancer was observed among women in the lowest tertile of expression for *ATM* (OR=2.12, 95%CI=1.09,4.12) as well as *MSH2* (OR=2.75, 95%CI=1.31,5.79) (Table 4b). Including age at blood draw and smoking status as covariates in the model did not appreciably alter the estimates from the crude model (Table 4b). Analysis conducted using GEE similarly indicated a significant association with breast cancer risk among women in the lowest tertile of expression for *ATM* (OR=1.51, 95% CI=1.03, 2.20) and *MSH2 (*OR=2.12, 95% CI=1.34, 3.34) (Table 5b). Additionally, an inverse association with breast cancer risk was also observed with *ATM* expression level as a continuous variable (OR=1.10, 95% CI=1.02, 1.18). Again, adjusting for age at blood draw and smoking status did not appreciably alter the estimates from the crude model (Table 5a). Women who were in the lowest tertile of expression for both *ATM* and *MSH2* had a nearly four-fold increase in risk compared to their unaffected sisters, adjusted for age at blood draw and smoking status (Figure 8). However, no significant interaction was observed, suggesting an additive effect.

To further characterize the risk associated with reduced expression of DNA repair genes, we examined the presence of effect modification due to family history. Using both conditional logistic regression and GEE analyses, the presence of extended family history impacted the association observed between the lowest tertile of *ATM* expression and breast cancer risk (Figures 9 and 10). The interaction term reached statistical significance using GEE analysis, suggesting that the association with breast cancer risk among women in the lowest tertile of ATM expression is largely restricted to women with an extended family history of disease.

Based on previously reported findings within the sister-sets indicating an increase in breast cancer risk due to deficiencies in DSBR as determined by measurements of end-joining (EJ) capacity, we set out to determine the relationship between our assessments of ATM expression and the previously determined levels of EJ capacity within our study population. A significant correlation was observed between ATM expression level and EJ capacity of EcoRI-generated sticky end substrates (r=0.23,p=0.02), while a borderline association was observed between ATM expression level and EJ capacity of HincII-generated blunt end substrates (r=0.20,p=0.06) (Figure 11). Furthermore, incorporating EJ capacity into the model assessing the impact of ATM expression level and breast

cancer risk, the observed association between ATM expression and breast cancer risk was knocked down (Table 7).

Table 4a. Conditional logistic regression analysis of the relationship between expression levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer

ATM (n=188 sets)	β	StdErr	Odds Ratio	95% Confidence Interval		Pr>z
Continuous	-0.11	0.06	1.11	0.98	1.26	0.10
T3 (>1.22)	ref	ref	ref	ref	ref	ref
T2 (0.54-1.22)	0.00	0.30	1.00	0.56	1.78	0.99
T1 (≤0.54)	0.74	0.34	2.09	1.07	4.07	**0.03**
BRCA1 (n=119 sets)						
Continuous	-0.67	1.36	1.96	0.14	28.3	0.62
T3 (>0.05)	ref	ref	ref	ref	ref	ref
T2 (0.02-0.05)	0.36	0.35	1.44	0.72	2.88	0.31
T1 (≤0.02)	-0.23	0.37	0.79	0.39	1.63	0.53
MSH2 (n=145 sets)						
Continuous	-0.86	1.18	2.36	0.23	24.0	0.47
T3 (>0.08)	ref	ref	ref	ref	ref	ref
T2 (0.03-0.08)	0.16	0.33	1.17	0.61	2.26	0.63
T1 (≤0.03)	1.05	0.38	2.87	1.36	6.02	**0.01**
MUTYH (n=168 sets)						
Continuous	0.00	0.06	1.00	0.89	1.12	0.97
T3 (>0.91)	ref	ref	ref	ref	ref	ref
T2 (0.36-0.91)	0.04	0.31	1.04	0.57	1.92	0.89
T1 (≤0.91)	0.21	0.36	1.23	0.61	2.51	0.57
XPC (n=196 sets)						
Continuous	-0.01	0.02	1.01	0.97	1.05	0.66
T3 (>3.05)	ref	ref	ref	ref	ref	ref
T2 (0.93-3.05)	-0.03	0.28	0.97	0.56	1.68	0.92
T1 (≤0.93)	0.48	0.35	1.62	0.81	3.25	0.17

Table 4b. Conditional logistic regression analysis of the relationship between expression levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer*

ATM (n=188 sets)	β	StdErr	Odds Ratio	95% Confidence Interval		Pr>z
Continuous	-0.10	0.06	1.11	0.98	1.26	0.11
T3 (>1.22)	ref	ref	ref	ref	ref	ref
T2 (0.54-1.22)	0.01	0.30	1.01	0.57	1.81	0.97
T1 (≤0.54)	0.75	0.34	2.12	1.09	4.12	**0.03**
BRCA1 (n=119 sets)						
Continuous	-0.61	1.37	1.84	0.13	26.9	0.66
T3 (>0.05)	ref	ref	ref	ref	ref	ref
T2 (0.02-0.05)	0.36	0.36	1.44	0.72	2.89	0.31
T1 (≤0.02)	-0.25	0.37	0.78	0.38	1.60	0.49
MSH2 (n=145 sets)						
Continuous	-0.79	1.20	2.20	0.21	23.3	0.51
T3 (>0.08)	ref	ref	ref	ref	ref	ref
T2 (0.03-0.08)	0.11	0.34	1.12	0.58	2.17	0.74
T1 (≤0.03)	1.01	0.38	2.75	1.31	5.79	**0.01**
MUTYH (n=168 sets)						
Continuous	0.00	0.06	1.00	0.89	1.12	0.95
T3 (>0.91)	ref	ref	ref	ref	ref	ref
T2 (0.36-0.91)	0.05	0.31	1.05	0.57	1.94	0.88
T1 (≤0.91)	0.22	0.37	1.25	0.61	2.55	0.55
XPC (n=196 sets)						
Continuous	-0.01	0.02	1.01	0.97	1.05	0.67
T3 (>3.05)	ref	ref	ref	ref	ref	ref
T2 (0.93-3.05)	-0.06	0.28	0.94	0.54	1.64	0.83
T1 (≤0.93)	0.46	0.36	1.58	0.79	3.18	0.20

*Adjusted for age at blood draw and smoking status

Table 5a. GEE analysis of the relationship between expression levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer

ATM (n=172 sets)	β	StdErr	L'β	95% Confidence Interval		Pr>Wald
Continuous	-0.09	0.04	0.91	0.85	0.98	0.02
T3 (>1.22)	ref	ref	ref	ref	ref	ref
T2 (0.54-1.22)	0.02	0.22	1.02	0.67	1.56	0.93
T1 (≤0.54)	0.40	0.20	1.49	1.01	2.18	0.04
BRCA1 (n=114 sets)						
Continuous	-0.36	1.23	0.70	0.06	7.81	0.77
T3 (>0.05)	ref	ref	ref	ref	ref	ref
T2 (0.02-0.05)	0.19	0.26	1.21	0.73	2.02	0.46
T1 (≤0.02)	-0.30	0.27	0.74	0.43	1.26	0.26
MSH2 (n=134 sets)						
Continuous	-0.74	0.75	2.09	0.48	9.03	0.32
T3 (>0.08)	ref	ref	ref	ref	ref	ref
T2 (0.03-0.08)	0.14	0.26	1.14	0.69	1.90	0.60
T1 (≤0.03)	0.73	0.23	2.08	1.32	3.28	<0.01
MUTYH (n=155 sets)						
Continuous	0.01	0.05	1.01	0.92	1.12	0.81
T3 (>0.91)	ref	ref	ref	ref	ref	ref
T2 (0.36-0.91)	-0.02	0.24	0.98	0.61	1.58	0.95
T1 (≤0.91)	0.07	0.20	1.08	0.72	1.60	0.71
XPC (n=178 sets)						
Continuous	-0.01	0.01	0.99	0.97	1.02	0.69
T3 (>3.05)	ref	ref	ref	ref	ref	ref
T2 (0.93-3.05)	-0.08	0.22	0.92	0.60	1.41	0.71
T1 (≤0.93)	0.21	0.19	1.23	0.85	1.78	0.28

Table 5b. GEE analysis of the relationship between expression levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer*

ATM (n=172 sets)	β	StdErr	L'β	95% Confidence Interval		Pr>Wald
Continuous	-0.09	0.04	1.10	1.02	1.18	0.02
T3 (>1.22)	ref	ref	ref	ref	ref	ref
T2 (0.54-1.22)	0.01	0.22	1.01	0.66	1.54	0.98
T1 (≤0.54)	0.41	0.19	1.51	1.03	2.20	0.03
BRCA1 (n=114 sets)						
Continuous	-0.42	1.23	1.52	0.14	16.9	0.73
T3 (>0.05)	ref	ref	ref	ref	ref	ref
T2 (0.02-0.05)	0.21	0.26	1.23	0.73	2.06	0.43
T1 (≤0.02)	-0.29	0.27	0.75	0.44	1.27	0.29
MSH2 (n=134 sets)						
Continuous	-0.78	0.74	2.17	0.51	9.24	0.29
T3 (>0.08)	ref	ref	ref	ref	ref	ref
T2 (0.03-0.08)	0.13	0.26	1.14	0.68	1.92	0.61
T1 (≤0.03)	0.75	0.23	2.12	1.34	3.34	<0.01
MUTYH (n=155 sets)						
Continuous	0.01	0.05	0.99	0.90	1.09	0.85
T3 (>0.91)	ref	ref	ref	ref	ref	ref
T2 (0.36-0.91)	-0.02	0.24	0.98	0.61	1.58	0.94
T1 (≤0.91)	0.07	0.20	1.08	0.72	1.60	0.72
XPC (n=178 sets)						
Continuous	-0.01	0.01	1.01	0.98	1.03	0.70
T3 (>3.05)	ref	ref	ref	ref	ref	ref
T2 (0.93-3.05)	-0.08	0.22	0.93	0.61	1.42	0.72
T1 (≤0.93)	0.21	0.19	1.23	0.85	1.79	0.27

*Adjusted for age at blood draw and smoking status

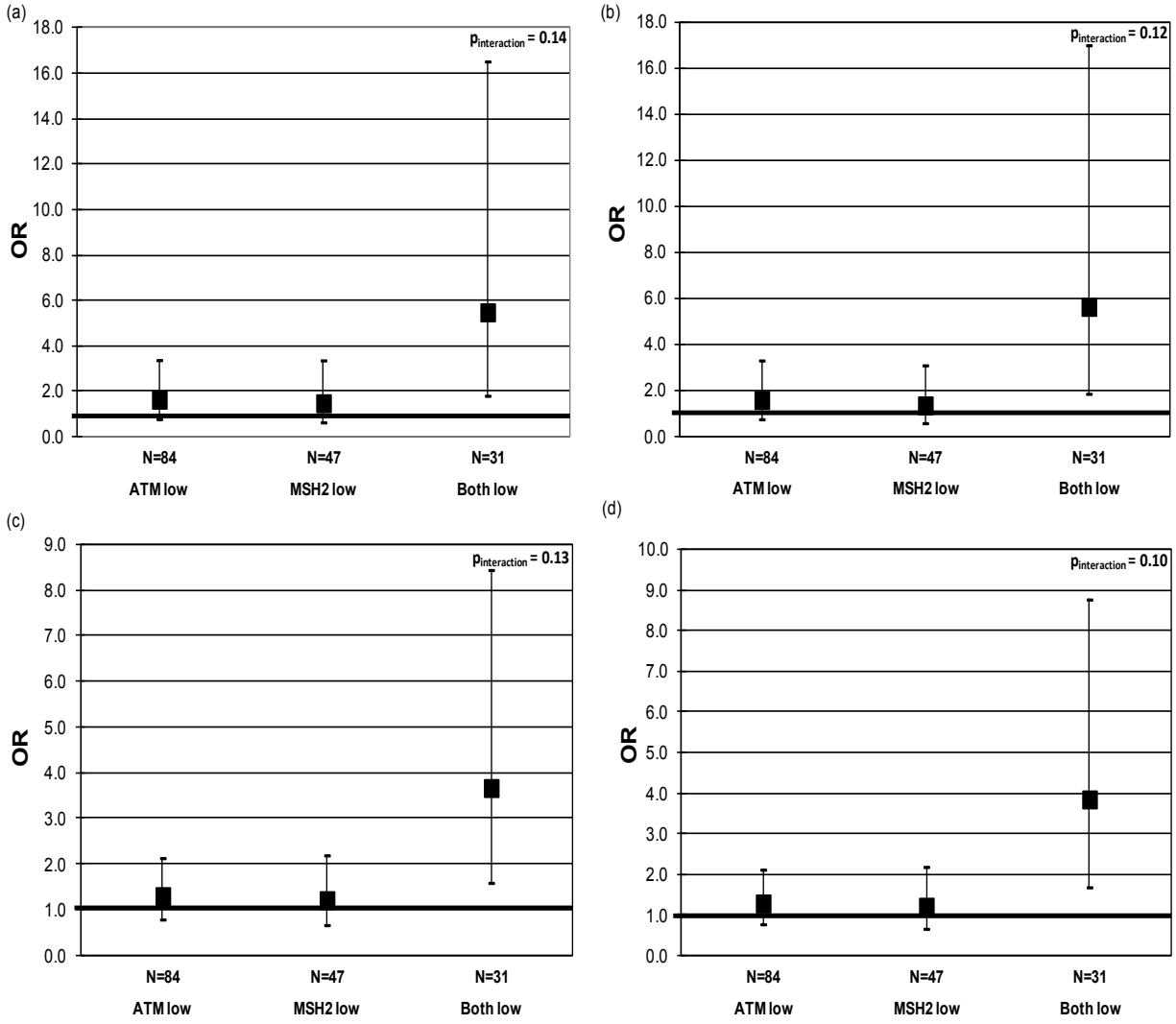


Figure 8. Combined effect of reduced ATM and MSH2 expression level on breast cancer risk among sisters discordant for breast cancer
a) crude conditional logistic regression model b) conditional logistic regression model adjusted for age at blood draw and smoke status c) crude GEE model d) GEE model adjusted for age at blood draw and smoking status

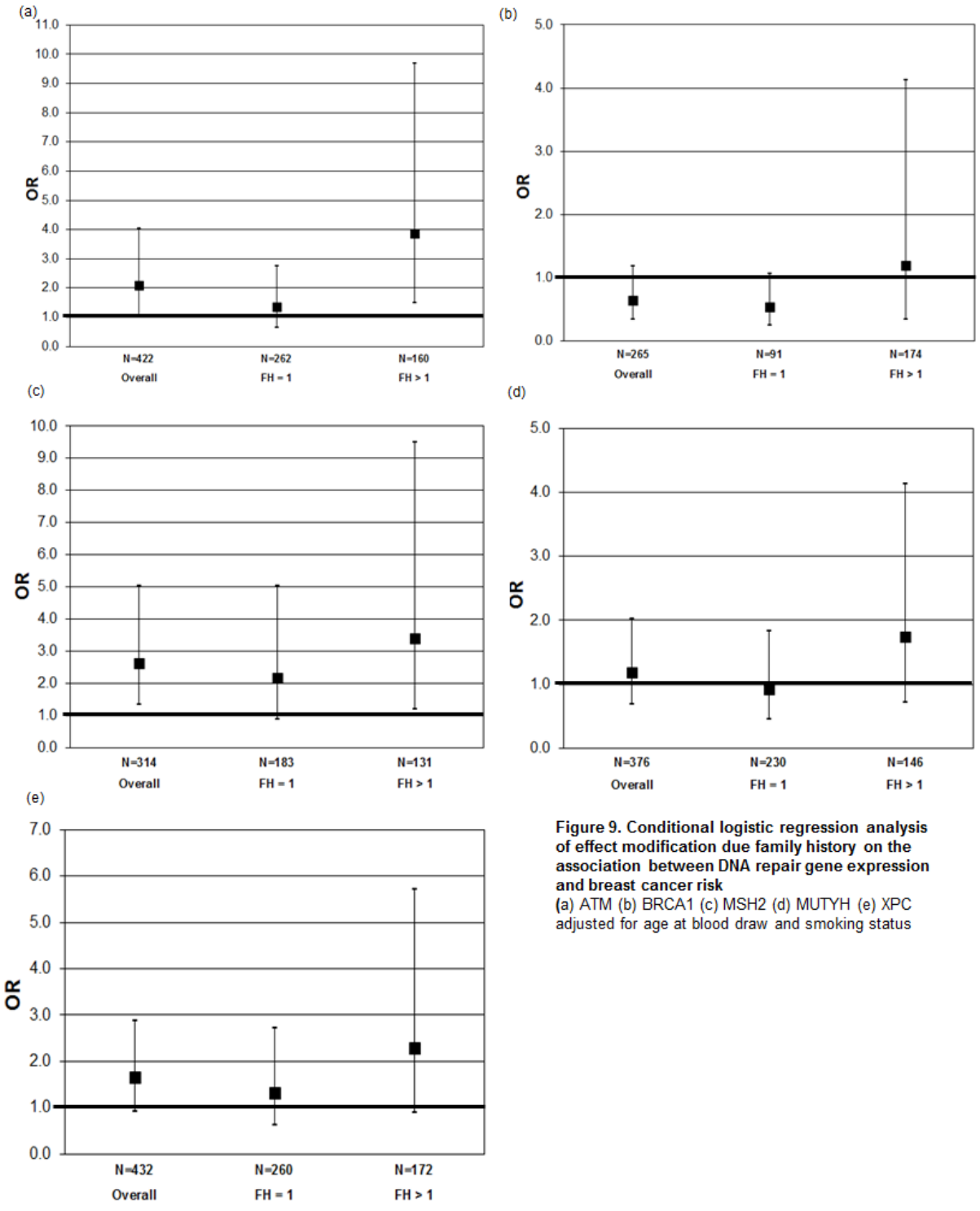


Figure 9. Conditional logistic regression analysis of effect modification due family history on the association between DNA repair gene expression and breast cancer risk
(a) ATM (b) BRCA1 (c) MSH2 (d) MUTYH (e) XPC adjusted for age at blood draw and smoking status

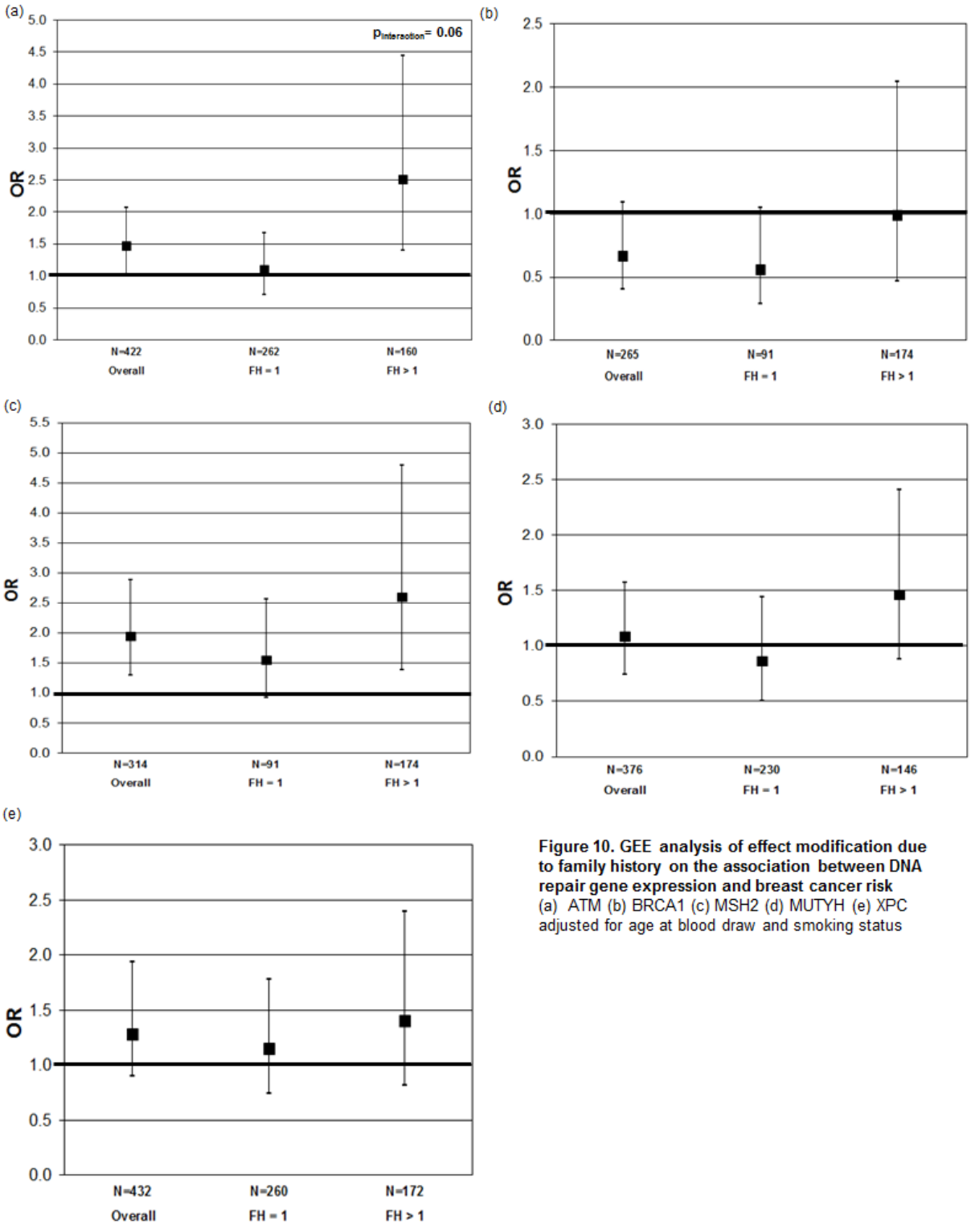


Figure 10. GEE analysis of effect modification due to family history on the association between DNA repair gene expression and breast cancer risk
(a) ATM (b) BRCA1 (c) MSH2 (d) MUTYH (e) XPC adjusted for age at blood draw and smoking status

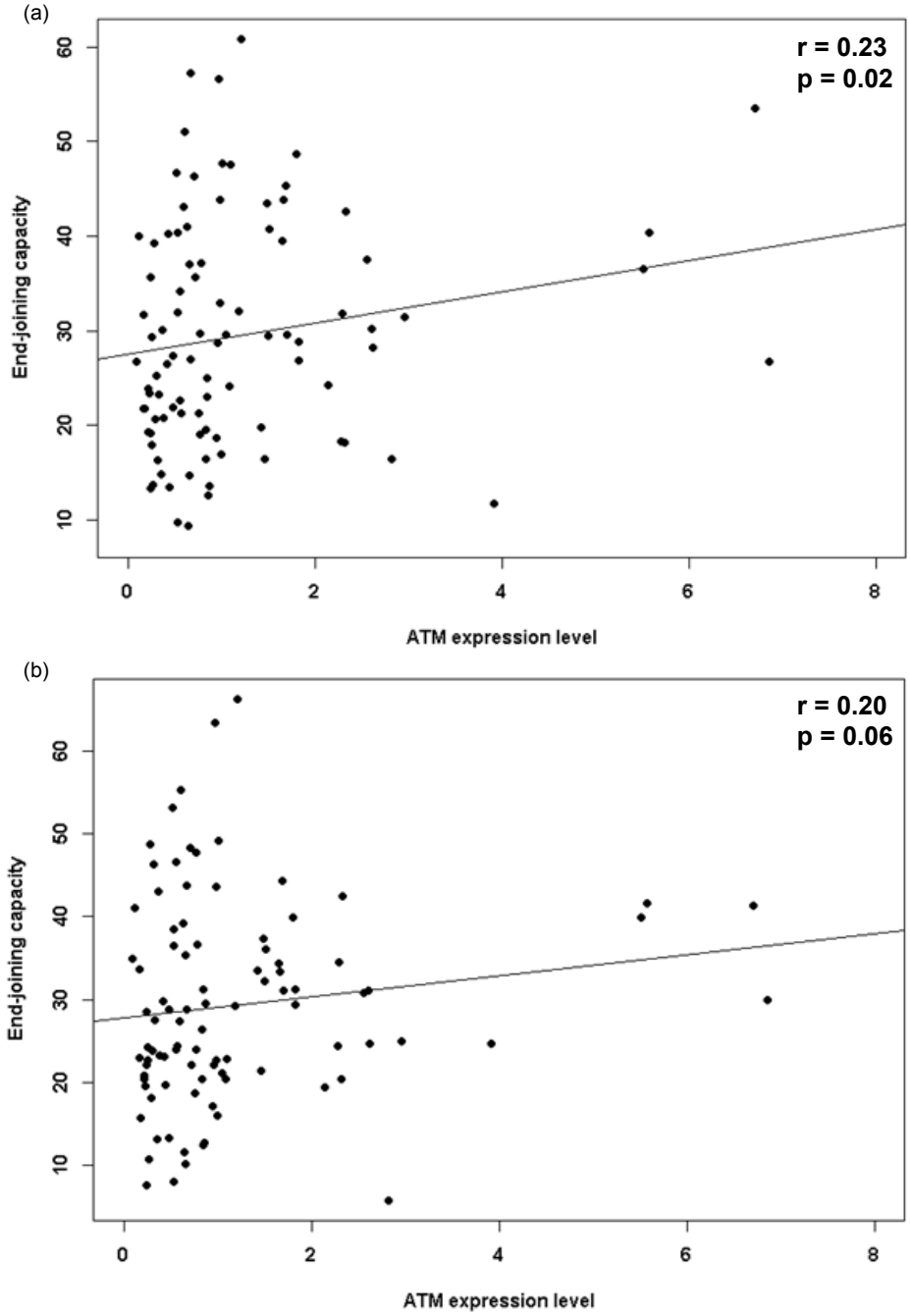


Figure 11. Spearman correlation between ATM expression level and EJ capacity of EcoRI-generated sticky ends (a) and HincII-generated blunt ends (b) among sisters discordant for breast cancer

Table 6. Impact of EJ capacity on the association between ATM expression level and breast cancer risk

Variable	Model 1[α]	Model 2[δ]	Model 3[ψ]
ATM T1 (≤0.54)	2.12 (1.09, 4.12)		
ATM T2 (0.54-1.22)	1.01 (0.57, 1.81)		
ATM T3 (>1.22)	1.00 (referent)		
ATM T1 (≤0.54)		1.28 (0.15, 11.2)	
ATM T2 (0.54-1.22)		1.00 (0.19, 5.23)	
ATM T3 (>1.22)		1.00 (referent)	
ATM T1 (≤0.54)			1.55 (0.15, 15.5)
ATM T2 (0.54-1.22)			1.30 (0.22, 7.64)
ATM T3 (>1.22)			1.00 (referent)

[α] Conditional logistic regression model assessing the relationship between tertiled levels of ATM expression and breast cancer risk, adjusted for age at blood draw and smoking status (ever vs. never)

[δ] Model includes tertiled EJ capacity of EcoRI-generated damage substrates in addition to all other explanatory variables of Model 1

[ψ] Model includes tertiled EJ capacity of HincII-generated damage substrates in addition to all other explanatory variables of Model 1

Discussion

In the current study, we demonstrate an association between reduced *ATM* and *MSH2* expression levels and breast cancer risk. Similar to the major breast cancer predisposition genes, *BRCA1* and BRCA2, *ATM* is also involved in the double-strand break repair pathway as a sensor of double-strand break damage. As a member of the P13K related protein kinases, the inactive serine-threonine kinase dimer phosphorylates into active monomers in the presence of double-strand breaks and transduces the signal to recruit downstream components of the pathway. Individuals with inherited mutations resulting in the inactivation of this kinase activity present with ataxia telangiectasia (AT), a disorder whose clinical features include a predisposition to cancer, with breast cancer leading as the predominant sub-type among women. While AT is a relatively rare genetic syndrome, approximately 1% of the population are carriers of the mutation, and heterozygotic relatives of AT individuals have been shown to have up to a 5-fold increase in risk of breast cancer, indicating a sizable attributable risk in the population setting[135].

While reports focusing on the impact of individual, more highly penetrant variants in *ATM* on breast cancer risk have been conflicting, a mutational screen of the *ATM* gene in a familial-based breast cancer case-control study accounting for truncating, splicing and missense mutations demonstrated a significantly higher count of variations among cases than among controls[136]. However, as *ATM* spans 62 exons, such mutational screens seem infeasible to adopt in large-scale studies. mRNA transcript levels provide an alternative means of comprehensively capturing functional activity, and several studies have shown reduced expression levels of ATM in breast tumors compared to normal tissue. To our knowledge, only one other breast cancer case-control study has reported on ATM expression levels in blood. Contrary to our findings, the study was not able to distinguish differences in expression levels based on case-control status[137]. However, this may be related to the small sample size of the study population.

In our study, the association between ATM expression level and breast cancer was modified by family history. This is in line with findings from a mutational screen of *ATM* within a breast cancer case-control study conducted at the Fred Hutchinson Cancer Research Center in which a higher mutation frequency in *ATM* among women with breast cancer was observed with increasing extent of family

history[138]. Furthermore, as in this study, other studies implicating ATM deficiency with breast cancer risk have primarily been restricted to study populations which selected for cases with a family history of breast cancer while population based case-control studies have failed to demonstrate similar associations. This further suggests that this identified risk factor may be of particular relevance to women with a family history of breast cancer and may, therefore, have implications for targeted genetic counseling.

MSH2 acts in conjunction with either MSH6 to recognize small loops and insertions or MSH3 to recognize single base mismatches in the mismatch repair pathway. While microsatellite instability, a marker of MMR phenotype, has been observed in breast tumor tissues, unlike ATM, a role for the specific deficiency in MSH2, or other constituents of the mismatch repair pathway, has not been commonly reported in association with breast cancer risk. Mismatch repair has been much more clearly implicated in colorectal cancer and other Lynch syndrome affiliated outcomes. However, more recently, several groups have been calling for the inclusion of breast cancer among the cancer-subtypes that constitute Lynch syndrome. In support of this claim, a recently conducted prospective study within the Colon Cancer Family Registry has shown a nearly four-fold increase in risk of breast cancer over the general population among unaffected carriers of MMR mutations while no increase in risk was observed among their non-carrier relatives[139].

In addition to the increase in breast cancer incidence observed due to overall MMR deficiency, upon stratifying risk by MMR genotype, several studies have shown a heightened risk among individuals with mutations specifically in the *MSH2* gene. In a study following Swiss Lynch syndrome families with *MLH1*/*MSH2* germline mutations, no overall increase in breast cancer incidence over the general population was observed. However, a significant increase in risk was observed within families with *MSH2* mutations[140]. Similarly, in a study within the Colon Cancer Family Registry examining the incidences of cancer among individuals with a previous diagnosis of colon cancer, the nearly two-fold elevation in breast cancer risk compared to the general population was mainly driven by the presence of *MSH2* mutations[141].

In a post-hoc analysis, upon combining women in the lowest tertile of expression for both genes, we observed an increase in risk above the risk associated with either gene alone. This indicates that

women with deficiencies in multiple DNA repair pathways are especially prone to develop breast cancer.

We did not observe differences in terms of age at diagnosis or tumor subtype based on the expression level of *ATM* or *MSH2* among breast cancer cases. However, this does not preclude the possibility that women with deficient expression of either gene who go on to develop breast cancer present with disease traits which we did not assess in the current study. A case series study assessing the clinicopathologic characteristics among individuals with breast cancer from families with early-onset colorectal cancer determined that invasive tumors arising in cases with an MMR-deficient phenotype are more poorly differentiated, have a high mitotic index and are more likely to be ER/PR negative than cases with tumors in which no MMR deficiency was detected[142].

An unexpected finding following stratification was the increase in breast cancer risk observed due to reductions in XPC expression level among women with an age at menarche after age 13. This is in contradiction with the convention that, early age at menarche, indicative of a longer exposure period to hormonal factors, increases risk of breast cancer. This finding may, therefore, indicate that the means through which *XPC* mediates its impact on risk is hormone-independent.

Finally, our findings suggest that the means through which reductions in ATM capacity impacts breast cancer risk is through the downregulation of DSBR capacity, substantiating the link between ATM and DNA repair. In addition to establishing the etiologic mechanism through which *ATM* exerts its impact on breast cancer risk, discerning the risk due to the specific genes responsible for an observed phenotype also provides targets that are more readily developed into screening biomarkers than phenotypic endpoints.

The risk due to EJ capacity that is explained by ATM expression level is less than substantial, as indicated by the correlation coefficient. However, this may be explained by a number of factors, including the fact that *ATM* is one of multiple components involved in the DSBR pathway as well as the fact that *ATM* may exert a role in additional pathways with relevance to breast cancer. Assay specific considerations may impact the level of agreement between the two findings as well. While gene expression studies are indicative of transcriptional regulation, assessments of overall DNA capacity reflect post-translational regulation as well. Furthermore, ATM expression levels were measured in

primary mononuclear cells, while EJ capacity was measured in transformed LCLs, likely reducing the comparability between the two assays. Still, the knockdown in the association observed between ATM expression and breast cancer risk in the presence of EJ capacity indicates a vital role imparted by ATM expression in the performance on the DSBR pathway and ultimate breast cancer risk within this population.

In summary, we report reductions in *ATM* and *MSH2* expression level in blood is associated with risk of familial breast cancer. Given a true relationship between these genes and breast cancer incidence, the ability to detect meaningful differences in blood indicates a potential utility as a non-invasive marker of risk. Detection outside of the target site also suggests, on a mechanistic level, that the observed reduction in expression represents a constitutive process. Furthermore, given the known sensitivity to ionizing radiation observed among AT individuals, ATM deficiency could also have prognostic implications, resulting in a greater likelihood of subsequent disease within this population of cases due to treatment of the primary tumor.

The major strength of this study lies in the matched sister-set study design, ensuring comparability between cases and controls. However, the implications of the findings from the current study are also hindered by several inherent limitations. Primarily, the retrospective nature of the study precludes us from ruling out that our observations may be driven by reverse causality. This issue is of special concern in this population, as biospecimens were often collected following onset of treatment.

Due to the limitations in inferences that can be drawn from this study, additional studies will have to be conducted to validate these findings within a prospective setting. Given the importance of these genes in determining risk, factors influencing the varying levels of expression of these genes also remain to be determined. Finally, given an ultimate motivation to better define risk groups, further clarification on the implication for prognosis among women who present with deficiencies in these genes will need to be addressed.

Chapter V: Promoter methylation levels of DNA repair genes as indicators of breast cancer risk

Abstract

Aberrant modulation of epigenetic patterns regulating DNA repair genes may contribute to the observed link between reduced DNA repair capacity and breast cancer risk. In this study, we investigated the promoter methylation levels of DNA repair-related genes and their association with breast cancer risk among participants of the New York site of the BCFR, comparing sisters affected with breast cancer to their unaffected sisters. We analyzed DNA extracted from viable mononuclear cells from 218 sisters (N=538 women) using pyrosequencing assays designed to target CpG islands in the promoter regions of *BRCA1, MSH2,* and *MLH1*. Based on conditional logistic regression analysis, breast cancer risk is not significantly associated with the promoter methylation level of *BRCA1* (OR=1.09, 95% CI = 0.98-1.20), *MLH1* (OR=1.18, 95% CI = 0.91-1.54) or *MSH2* (OR=0.87, 95% CI = 0.48-1.60). We assessed spearman correlations between methylation and previously assessed expression levels of *BRCA1* and *MSH2*, where we had observed a significant increase in breast cancer risk among women in the lowest tertile of expression for *MSH2*. No significant correlation was observed between methylation and expression level for either *BRCA1* (r=-0.05, p=0.39) or *MSH2* (-0.04, p=0.39). These findings suggest that while *MSH2* may have implications for breast cancer risk, as suggested by the previously observed increase in risk with decrease in expression level of *MSH2*, the lack of an association between gene promoter methylation levels and breast cancer risk and the lack of a correlation with the previously determined expression level of this gene, indicate that other mechanisms impacting the reduction in expression level of these genes may be more relevant to explain the observed increase in breast cancer risk. Similarly, the lack of an association observed between promoter methylation levels and breast cancer risk for both *BRCA1* and *MLH1* suggest that the selected CpG sites may not be relevant in influencing breast cancer risk.

Introduction

Breast cancer is the most common malignancy among women in the developing world and claimed nearly 40,000 lives in the US within the past year[143]. Reducing the burden in incidence and mortality will require the development of more targeted diagnostic and prognostic markers based on furthering the elucidation of the etiologic mechanism underlying the breakdown of normal cellular processes. In addition to genetic alterations, this breakdown, characterized by the activation of oncogenes and knockdown of tumor suppressor genes, can also result from epigenetic aberrations. DNA methylation, in which a methyl group is incorporated at the 5' site of a cytosine residue, is the most commonly studied epigenetic mark[144]. In humans, such modifications are typically restricted to CpG dinucleotides, which are found to cluster in the genome in CpG-rich islands[145]. In fact, it has been established that cancer cells present with an altered epigenetic landscape. These alterations follow a common pattern across cancer sub-types, with overall global reductions in methylation levels, facilitating genomic instability, as well as localized increases in methylation at specific gene promoter regions, often resulting in the inactivation of tumor suppressor genes[146].

As DNA repair genes are crucial in the maintenance of genomic integrity, downregulation of these genes can be critical in the onset and progression of carcinogenesis[147]. The importance of these genes in cancer has long been recognized as germline mutations of several DNA repair-related genes, including *BRCA1*, *MLH1* and *MSH2*, have been linked to the familial presentation of cancer[148–150]. Furthermore, tumors in individuals with deficiencies in these genes present with a distinct phenotype. BRCA1-deficient individuals, for example, develop breast cancers that are more likely to be basal-like, triple-negative and of medullary histology[151,152]. Similarly, MLH1 and MSH2 deficient individuals develop Lynch syndrome related cancer sub-types with MSI, the presence of which is indicative of unrepaired errors due to polymerase slippage during replication[153]. However, outside of the familial setting, genetic mutations account for relatively few cases, even among those who present with phenotypes typically associated with deficiencies in these genes[152]. Consequently, more recent studies have been motivated to determine whether epigenetic downregulation of DNA repair activity through gene-specific hypermethylation may further contribute to risk. Indeed, studies have reported on the presence of methylated promoters of *BRCA1*, *MLH1* and *MSH2*, in tumors, and the

clinicopathological features of these tumors resemble those typically observed in cases with germline mutations of these genes[152,154].

While a number of studies have examined the methylation profile of DNA repair genes in tumors, only a few studies focusing on *BRCA1* have investigated the presence of promoter hypermethylation of these genes in peripheral blood[155,151]. Among these, only one study sought to determine case-control differences due to promoter methylation of *BRCA1* in peripheral blood cells (PBCs). In this study, an increase in breast cancer risk was observed among those individuals who had detectable *BRCA1* methylation levels in blood[151].

To further determine the contribution to breast cancer risk attributable to the gene-specific methylation status of DNA repair genes, we conducted the current study to assess methylation levels in the promoter regions of *BRCA1*, *MLH1* and *MSH2* in PBMCs from discordant sister-sets enrolled in the NY-BCFR as potential markers of breast cancer risk.

Materials and Methods

Study population

The subjects for the current study included 218 sister-sets (n=569) consisting of sisters discordant for breast cancer enrolled in the New York site of the BCFR with available viable PBMCs (Figure 1).

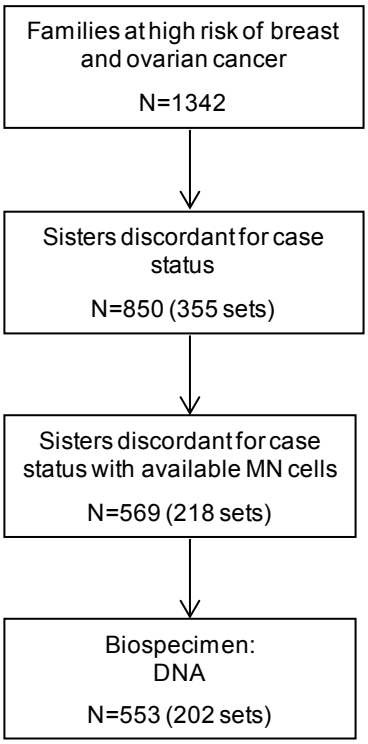


Figure 1. Methylation study design
The study was conducted among sisters discordant for disease with DNA available from MN cells.

DNA extraction

DNA was extracted using an All-prep DNA/RNA 96 kit (Qiagen, Valencia, CA). The plates were designed such that complete sister-sets were present on the same plate, while blindedness to case-control status was maintained. Purified DNA was quantified using Nanodrop 2000 (Thermo Scientific, Pittsburgh, PA)

Pyrosequencing

For bisulfite conversion, 250ng of DNA from each sample in the DNA extraction plate was transferred onto new plates. In addition to the intact sister-sets, each plate also contained 5% repeat samples, a

fully methylated positive control (Millipore, Billerica, MA), a whole genome amplified negative control, and a non-template control.

Bisulfite conversion of the plates was carried out using EZ-96 DNA Methylation Gold kits (Zymo research, Irvine, CA). Replica plates containing 2uL template bisulfite-converted DNA were prepared from the original master plates.

PCR and sequencing primers targeting CpG sites in the promoter regions of each gene of interest were designed using Pyromark Assay Design Software 2.0 (Qiagen) (Table 1). A PCR reaction mixture including 1X Pyromark master mix (Qiagen), 1X Coral loading dye (Qiagen), and 0.2uM forward and reverse primers was added to the DNA template-containing replica plates, yielding a total volume of 25uL. The cycling conditions for the PCR reactions were as follows: 95°C 15min, [94°C 30sec, 56°C 30sec, 72°C 30sec] x 45 cycles, 72°C 10 min.

Following the PCR, 2uL sepharose beads (GE Healthcare, Pittsburgh, PA), 40uL Binding Buffer (Qiagen) in a total volume of 60uL was added to each reaction mixture. After a 5 minute incubation period of vigorous shaking, the samples were transferred to a vaccum prep station, where a filter probe tool was used to isolate the sepharose bead-bound amplicons from the reaction mixture. The probes with the bound amplicons were then passed sequentially through 70% ethanol, 0.5N NaOH and wash buffer (Qiagen) before the vacuum was turned off and the amplicons were transferred into a sequencing plate, with each well containing a 25uL volume of 0.3uM sequencing primer diluted in annealing buffer (Qiagen).

Table 1. PCR and sequencing primers

Assay	Forward Primer	Reverse Primer	Sequencing Primer
BRCA1	GGGTAGATTGGGTGGTTAATT	CTATCCCTTTCCCCAAACTCTACTACCTT (5'biotin)	GGGAATTATAGATAAATTAAAATTG
MLH1	GGTATTTTTGTTTTTATTGGTTGGATATT	AATACCAATCAAATTTCTCAACTCTAT (5'biotin)	TTTTAAAAAAGAATTAATAGGAAGA
MSH2	AGGGGTTTTAAGTTTTGTAGTTGAGT	ACCCCACACCCACTAAACTAT (5'biotin)	AAAAGTAGTTTTGGAAGTTGAT

To denature the template DNA, each sequencing plate was incubated at 80°C for 2min and then allowed to cool to room temperature to enable the sequencing primer to bind the target region. A dispensation cartridge was prepared with the required amount of nucleotides, enzyme and substrate, as prescribed by the Pyromark Q24 software. Both cartridge and sequencing plate were then loaded in the Pyromark Q24 instrument, and the run was initiated. Enzyme, substrate and an assay-specific sequence of nucleotides were dispensed by the machine as determined by the Pyromark Q24 software. The diphosphates released upon nucleotide incorporation serve as substrates for ATP generation, which in turn drives luciferase activity to generate a luminescent signal. This signal, which corresponds to nucleotide incorporation, is detected as peaks in the pyrogram. Following the run, percent methylation of each interrogated CpG site was calculated by the Pyromark Q24 software based on the ratio of the area of the peak generated by the dispensation targeting the unconverted base to the total area of the peaks generated by the dispensations targeting both the converted and unconverted base. Default parameters for peak heights expected based on the input sequence as well as control dispensations to indicate bisulfite conversion efficiency were used to determine quality of the runs.

Statistical analysis

For each of the three genes, percent methylation was reported as the mean value calculated across the interrogated CpG sites. Wilcoxon rank sum tests were conducted to compare mean case-control differences in promoter methylation levels of all three genes. Pearson's chi-square tests were performed to assess the distribution of the study population according to selected measured variables.

Linear regression analysis was conducted to assess the impact of breast cancer risk factors on promoter methylation level on a continuous scale, and logistic regression analysis was conducted to assess the impact of these factors on differentiating women in the highest tertile of methylation from women with lower levels of methylation. Breast cancer risk factors that were included in the analyses included age at blood draw, smoking status (ever vs. never), body mass index (BMI) ($\geq 25 kg/m^2$ vs $<25 kg/m^2$, cut-off based on the WHO definition of the normal range), age at menarchy (>13 vs ≤13 years, based on median age at menarche among unaffected sisters), age at first parity (>30 vs ≤30

years, nulliparous women were assigned age at blood draw), and extended family history (>1 vs 1). Conditional logistic regression and GEE models were run to assess the associations between promoter methylation levels. Risk was determined evaluating methylation level as a continuous variable, by assessing the change in risk due to a 1% increase in methylation level, and partitioning methylation level into tertiles with cut-offs based on the promother methylation levels in the unaffected women. To adjust for potential confounding, additional known risk factors for disease, including age at blood draw and smoking status were included in the models. All statistical analyses were conducted using using R: A Language and Environment for Statistical Computing (2011), R Foundation for Statistical Computing, Vienna, Austria.

Results

Assay design

All assays included in the current study target methylation sites in CpG islands located in the promoter region of the genes of interest (Figures 2, 3 and 4).

BRCA1: NC_000017.10 chr.17:41276061-41277800

```
   1 tacttatatt taccgaaact ggagacctcc attagggcgg aaagagtggg ggattgggac
  61 ctcttcttac gactgctttg gacaataggt agcgattctg accttcgtac agcaattact
 121 gtgatgcaat aagccgcaac tggaagagta gaggctagag ggcaggcact ttatggcaaa
 181 ctcaggtaga attcttcctc ttccgtctct ttccttttac gtcatccggg ggcagactgg
 241 gtggccaatc cagagccccg agagacgctt ggctctttct gtccctccca tcctctgatt
 301 gtaccttgat ttcgtattct gagaggctgc tgcttagcgg tagccccttg gtttccgtgg
 361 caacggaaaa gcgcgggaat tacagataaa ttaaaactgc gactgcgcgg cgtgagctcg
 421 ctgagacttc ctggacgggg gacaggctgt ggggtttctc agataactgg gcccctgcgc
 481 tcaggaggcc ttcaccctct gctctgggta aaggtagtag agtcccggga aagggacagg
 541 gggcccaagt gatgctctgg ggtactggcg tgggagagtg gatttccgaa gctgacagat
 601 gggtattctt tgacgggggg taggggcgga acctgagagg cgtaaggcgt tgtgaaccct
 661 ggggaggggg gcagtttgta ggtcgcgagg gaagcgctga ggatcaggaa gggggcactg
 721 agtgtccgtg ggggaatcct cgtgatagga actggaatat gccttgaggg ggacactatg
 781 tctttaaaaa cgtcggctgg tcatgaggtc aggagttcca gaccagcctg accaacgtgg
 841 tgaaactccg tctctactaa aaatacaaaa attagccggg cgtggtgccg ctccagctac
 901 tcaggaggct gaggcaggag aatcgctaga acccgggagg cggaggttgc agtgagccga
 961 gatcgcgcca ttgcactcca gcctgggcga cagagcgaga ctgtctcaaa acaaaacaaa
1021 acaaaacaaa acaaaaaaca ccggctggta tgtatgagag gatgggacct tgtggaagaa
1081 gaggtgccag gaatatgtct gggaagggga ggagacagga ttttgtggga gggagaactt
1141 aagaactgga tccatttgcg ccattgagaa agcgcaagag ggaagtagag gagcgtcagt
1201 agtaacagat gctgccggca gggatgtgct tgaggaggat ccagagatga gagcaggtca
1261 ctgggaaagg ttaggggcgg ggaggccttg attggtgttg gtttggtcgt tgttgatttt
1321 ggttttatgc aagaaaaaga aaacaaccag aaacattgga gaaagctaag gctaccacca
1381 cctacccggt cagtcactcc tctgtagctt tctctttctt ggagaaagga aaagacccaa
1441 ggggttggca gcaatatgtg aaaaaattca gaatttatgt tgtctaatta caaaaagcaa
1501 cttctagaat ctttaaaaat aaaggacgtt gtcattagtt ctttggtttg tattattcta
1561 aaaccttcca aatcttaaat ttactttatt ttaaaatgat aaaatgaagt tgtcatttta
1621 taaacctttt aaaaagatat atatatatgt ttttctaatg tgttaaagtt cattggaaca
1681 gaaagaaatg gatttatctg ctcttcgcgt tgaagaagta caaaatgtca ttaatgctat
```

Figure 2. BRCA1 promoter region
The pyrosequencing assay for the promoter of this gene was designed to target 4 CpG sites, highlighted in yellow, reported by Snell et al. to be detectable in blood of breast cancer patients. Sequences targeted by primers are indicated in green and translation start site is indicated in pink.

Pyrograms of pyrosequencing runs based on the designed assays are shown in Figure 5. Sequencing of fully methylated and unmethylated control DNAs generated the expected peak height patterns for all three genes.

```
MLH1: NC_000003.11 chr.3:37034514-37035173
  1 ggaaaactag agcctcgtcg acttccatct tgcttctttt gggcgtcatc cacattctgc
 61 gggaggccac aagagcaggg ccaacgttag aaaggccgca aggggagagg aggagcctga
121 gaagcgccaa gcacctcctc cgctctgcgc cagatcacct cagcagaggc acacaagccc
181 ggttccggca tctctgctcc tattggctgg atatttcgta ttccccgagc tcctaaaaac
241 gaaccaatag gaagagcgga cagcgatctc taacgcgcaa gcgcatatcc ttctaggtag
301 cgggcagtag ccgcttcagg gagggacgaa gagacccagc aacccacaga gttgagaaat
361 ttgactggca ttcaagctgt ccaatcaata gctgccgctg aagggtgggg ctggatggcg
421 taagctacag ctgaaggaag aacgtgagca cgaggcactg aggtgattgg ctgaaggcac
481 ttccgttgag catctagacg tttccttggc tcttctggcg ccaaaatgtc gttcgtggca
541 ggggttattc ggcggctgga cgagacagtg gtgaaccgca tcgcggcggg ggaagttatc
```

Figure 3. MLH1 promoter region
The pyrosequencing assay for the promoter region of this gene was designed to target 4 CpG sites, highlighted in yellow, reported by Deng et al. to be relevant in transcriptional regulation. Sequences targeted by primers are indicated in green and translation start site is indicated in pink.

```
MSH2: NC_000002.11 chr.2: 47629800-47630399
  1 cgatgttggc agtttgctta gaaagaaaaa gggaggcagt cggagagggg cacacgtttt
 61 aacaaaatac tgggaggagg aggaaggcta gttttttttt tgttttcaag tttccttctg
121 atgttactcc catgcttccg ggcacattac gagctcagtg cctgccggaa atctcccacc
181 tggtggcaac ctacccttgc atacacccca cccaggggct tcaagccttg cagctgagta
241 aacacagaaa ggagctctac taaggatgcg cgtctgcggg tttccgcgcg acctaggcgc
301 aggcatgcgc agtagctaaa gtcaccagcg tgcgcgggaa gctgggccgc gtctgcttat
361 gattggttgc cgcggcagac tcccacccac cgaaacgcag ccctggaagc tgattgggtg
421 tggtcgccgt ggccggacgc cgctcggggg acgtgggagg ggaggcggga aacagcttag
       myo                          EIA-F
481 tgggtgtggg gtcgcgcatt ttcttcaacc aggaggtgag gaggtttcga catggcggtg
541 cagccgaagg agacgctgca gttggagagc gcggccgagg tcggcttcgt gcgcttcttt
```

Figure 4. MSH2 promoter region
The pyrosequencing assay for the promoter region of this gene was designed to target 7 CpG sites, highlighted in yellow, reported by Iwahashi et al. to be within a region relevant for determining promoter activity. These sites also include putative binding sites for transcription factors myo and EIA-F. Sequences targeted by primers are indicated in green and translation start site is indicated in pink.

Case control study

The demographics of the study population are shown in Table 2. No significant case control differences were observed based on age at blood draw, smoking status, BMI, age at first menarche, age at first birth or family history of breast cancer.

Wilcoxon rank sum test revealed no significant overall differences in methylation levels between affected and unaffected sisters among any of the three genes (Figure 6). Furthermore, no case control differences were observed based on known risk factors associated with breast cancer. However, among

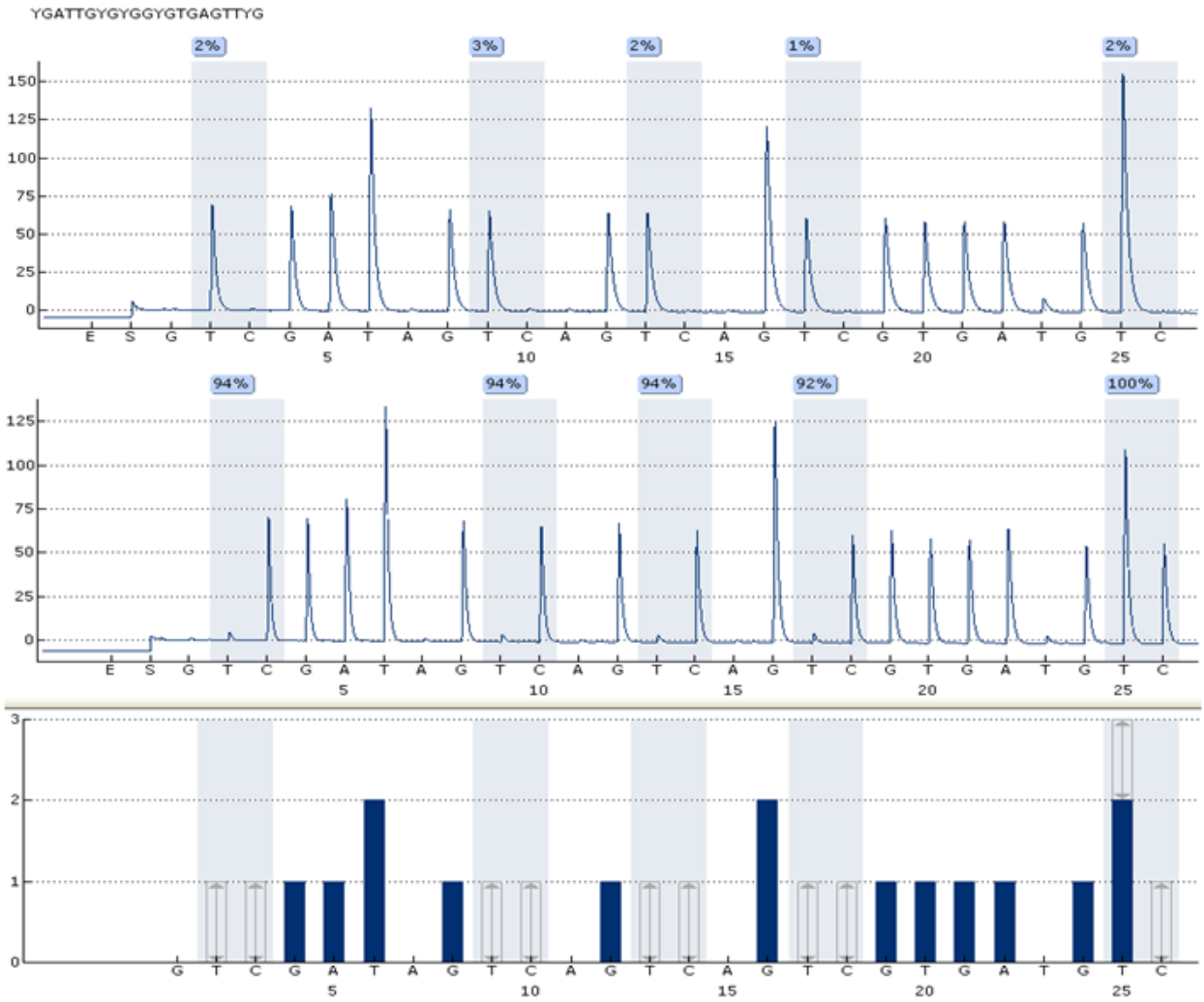


Figure 5a. BRCA1 pyrosequencing assay
BRCA1 pyrograms of a fully unmethylated DNA sample (top panel) and fully methylated DNA sample (middle panel). Expected signal profile is shown in the bottom panel Highlighted blue areas indicate locations of CpG sites.

unaffected women, *BRCA1* methylation levels were significantly higher among non-smokers compared to smokers, while borderline lower levels of MSH2 methylation levels were observed among non-smokers compared to smokers. Additionally, unaffected women with an age at menarche prior to age 13 had higher levels of *MSH2* methylation. Among affected women, *MSH2* methylation levels were significantly higher among those with an extended family history of disease compared to those without and extended family history (Table 2).

Next we determined whether known breast cancer risk factors impact methylation levels in the genes of interest (Table 4). Restricting our analysis to the unaffected population in our study, we observed that

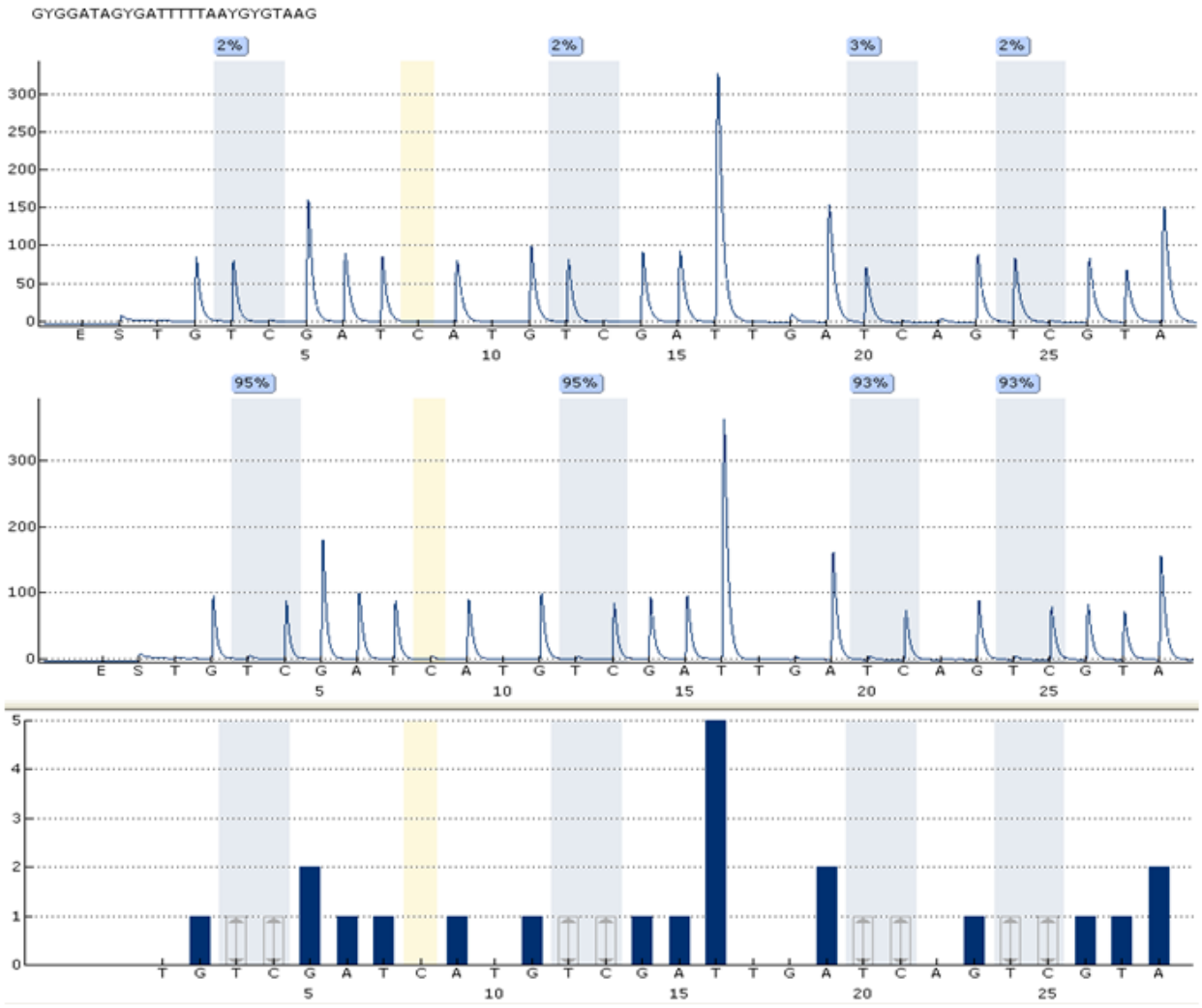


Figure 5b. MLH1 pyrosequencing assay
MLH1 pyrograms of a fully unmethylated DNA sample (top panel) and fully methylated DNA sample (middle panel). Expected signal profile is shown in the bottom panel Highlighted blue and orange areas indicate locations of CpG sites and bisulfite treatment controls, respectively.

non-smokers were more likely to be in the highest tertile of methylation for *BRCA1* (p=0.01). Women with an age at menarche below the median age at menarche in this population were also more likely to be in the highest tertile of methylation for *BRCA1* (p=0.01) as well as *MSH2* (p=0.03) compared to women with a later age at menarche.

Conditional logistic regression analysis revealed no significant association between methylation levels of any of the three genes, measured continuously or tertiled, and case status (Table 5a). Adjusting for age at blood draw and smoking status did not significantly alter any of the models (Table 5b). Similarly, no significant associations with case status were observed using GEE analysis (Table 6a and b).

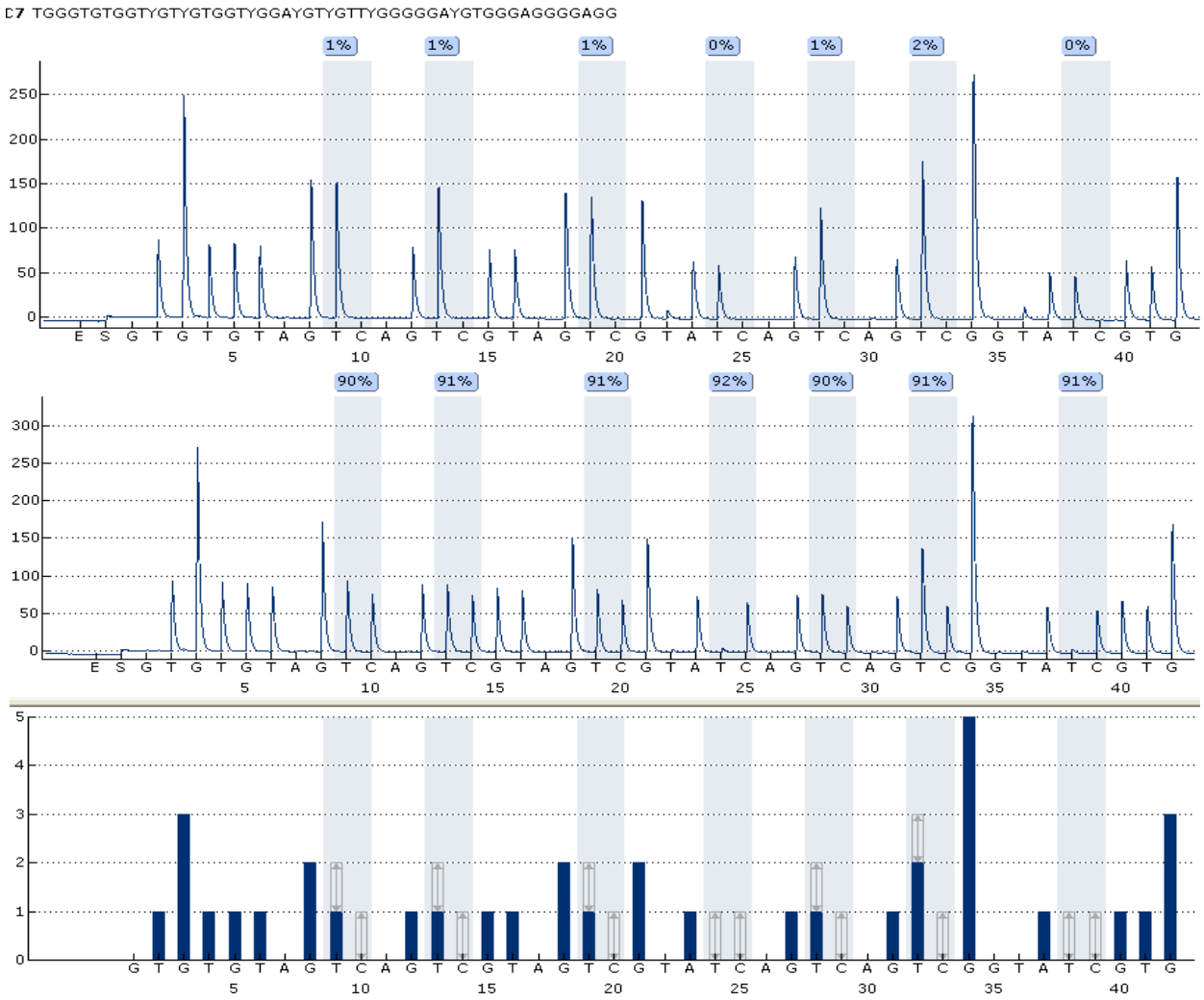


Figure 5c. MSH2 pyrosequencing assay
MSH2 pyrograms of a fully unmethylated DNA sample (top panel) and fully methylated DNA sample (middle panel). Expected signal profile is shown in the bottom panel Highlighted blue areas indicate locations of CpG sites.

Finally, as gene specific promoter methylation is believed to impact cancer risk through its modulation on gene expression levels, we determined the correlation between these endpoints in two genes, *BRCA1* and *MSH2,* for which we had previously determined expression levels in PBMCs from the same individuals. As expected, an inverse relationship was observed for both *BRCA1* (rho=-0.05, p=0.39) and *MSH2* (rho=-0.04, p=0.39), indicating that higher methylation levels correspond with lower expression levels. However, the observed correlations were not significant for either gene (Figure 7).

Table 2. Distribution of selected variables among sisters affected and unaffected with breast cancer

Variable	No. cases (%) (n=239)	No. controls (%) (n=298)	p-value[α]
Age at blood draw (in years)			
≤50	142 (59.4)	196 (65.8)	0.14
>50	97 (40.6)	102 (34.2)	
BRCA1 (+) mutation	11 (4.62)	7 (2.36)	0.22
BRCA2 (+) mutation	13 (5.46)	7 (2.36)	0.10
Ethnicity			
Non-hispanic White	127 (54.0)	142 (47.8)	0.32
African American	5 (2.12)	8 (2.69)	
Other	103 (43.8)	147 (49.5)	
Smoking Status			
Never	143 (59.8)	186 (62.4)	0.60
Ever	96 (40.2)	112 (37.6)	
BMI (kg/m^2)			
<25	114 (47.7)	142 (48.3)	0.96
≥25	125 (52.3)	152 (51.7)	
Age at menarchy (in years)			
<13	109 (46.0)	127 (42.6)	0.49
≥13	128 (54.0)	171 (57.4)	
Age at first parity			
<30	151 (63.2)	201 (67.4)	0.35
≥30	88 (36.8)	97 (32.6)	
Family History			
=1	137 (57.3)	193 (64.8)	0.09
>1	102 (42.7)	105 (35.2)	

[α] Pearson's X^2 test

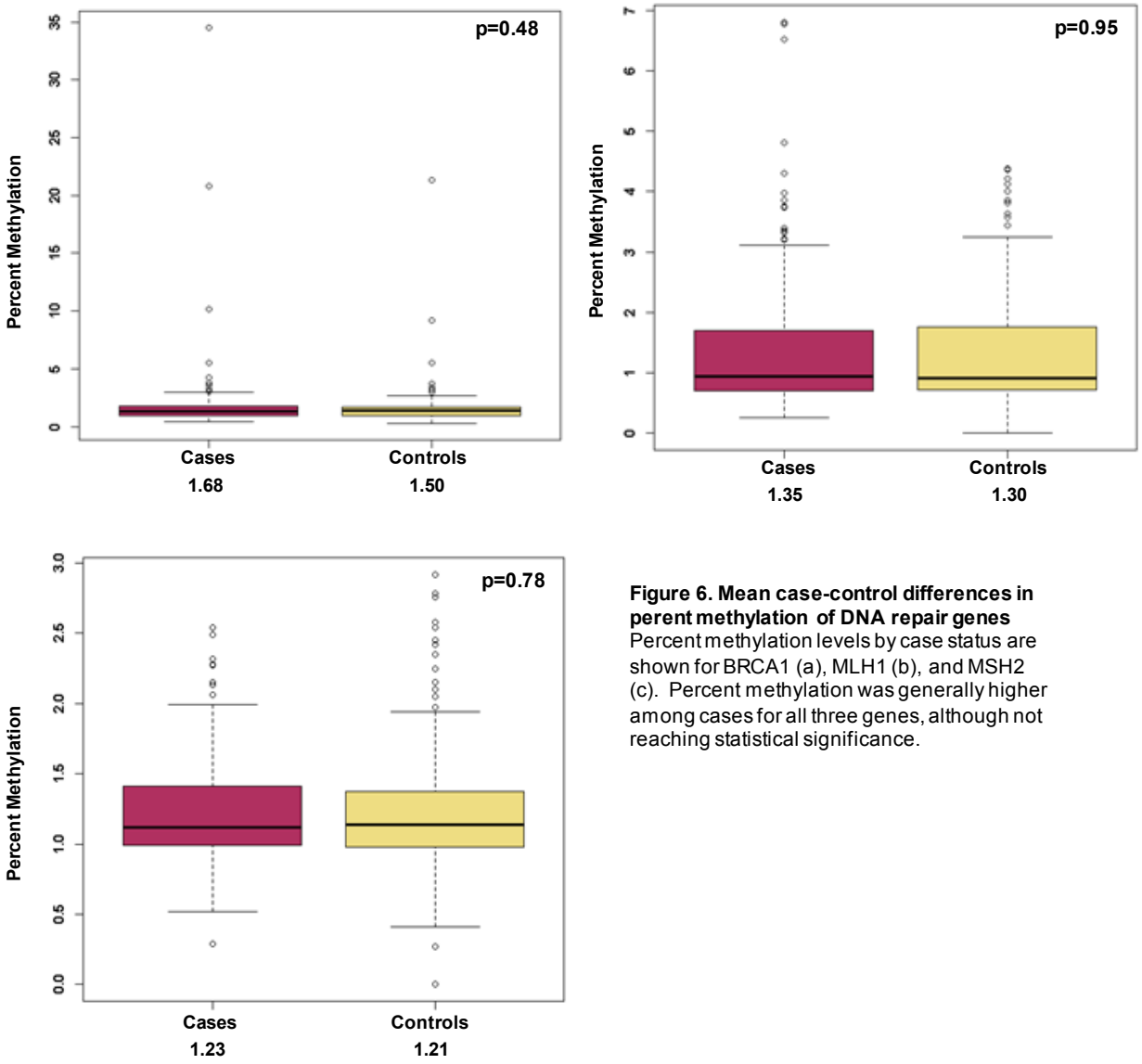


Figure 6. Mean case-control differences in perent methylation of DNA repair genes
Percent methylation levels by case status are shown for BRCA1 (a), MLH1 (b), and MSH2 (c). Percent methylation was generally higher among cases for all three genes, although not reaching statistical significance.

Table 3. Percent methylation of DNA repair gene promoters among sisters affected and unaffected with breast cancer

Variable	BRCA1					MLH1					MSH2				
	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]	Cases		Controls		p-value[δ]
	N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD		N	Mean ± SD	N	Mean ± SD	
All	226	1.68 ± 2.70	281	1.50 ± 1.40	0.48	222	1.35 ± 1.06	276	1.30 ± 0.88	0.95	218	1.26 ± 0.37	216	1.23 ± 0.40	0.78
Age at blood draw															
≤50	134	1.81 ± 3.37	187	1.43 ± 0.60	0.61	130	1.37 ± 1.17	185	1.32 ± 0.90	0.80	134	1.20 ± 0.14	179	1.21 ± 0.11	0.79
>50	91	1.48 ± 1.14	93	1.66 ± 2.27	0.68	91	1.31 ± 0.87	89	1.28 ± 0.87	0.76	93	1.27 ± 0.09	93	1.24 ± 0.13	0.61
p[δ]		0.62		0.64			0.67		0.82			0.15		0.54	
Smoking Status															
Never	137	1.81 ± 3.33	175	1.62 ± 1.71	0.23	136	1.39 ± 0.11	177	1.28 ± 0.09	0.39	139	1.22 ± 0.14	171	1.19 ± 0.10	0.38
Ever	88	1.47 ± 1.15	105	1.31 ± 0.54	0.74	85	1.27± 0.09	97	1.35 ± 0.14	0.26	88	1.23 ± 0.06	101	1.27 ± 0.13	0.64
p[δ]		0.34		**0.02**			0.25		0.59			0.86		0.05	
BMI (kg/m^2)															
<25	104	1.69 ± 3.33	130	1.34 ± 1.71	0.77	108	1.27 ± 1.03	131	1.22 ± 0.88	0.70	134	1.20 ± 0.10	179	1.21 ± 0.12	0.32
≥25	121	1.67 ± 1.15	148	1.64 ± 0.54	0.24	113	1.41 ± 1.11	140	1.38 ± 0.90	0.83	93	1.27 ± 0.13	93	1.24 ± 0.11	0.58
p[δ]		0.52		0.15			0.47		0.23			0.99		0.21	
Age at Menarche															
<13	102	1.41 ± 1.05	121	1.35 ± 0.66	0.82	103	1.40 ± 1.19	121	1.38 ± 0.95	0.90	103	1.24 ± 0.36	120	1.26 ± 0.39	0.52
≥13	122	1.90 ± 3.53	159	1.62 ± 1.76	0.50	117	1.30 ± 0.93	153	1.25 ± 0.83	0.92	123	1.22 ± 0.37	152	1.18 ± 0.40	0.37
p[δ]		0.12		0.06			0.56		0.50			0.63		**0.02**	
Age at First Parity															
<30	142	1.58 ± 1.92	187	1.50 ± 1.55	0.21	136	1.35 ± 1.02	183	1.29 ± 0.88	0.83	143	1.21 ± 0.35	183	1.24 ± 0.41	0.66
≥30	83	1.85 ± 3.67	93	1.50 ± 1.05	0.71	85	1.33 ± 1.12	91	1.34 ± 0.91	0.78	84	1.25 ± 0.40	89	1.17 ± 0.36	0.35
p[δ]		0.31		0.81			0.79		0.86			0.74		0.15	
Family History															
=1	130	1.87 ± 3.42	184	1.59 ± 1.67	0.62	124	1.39 ± 1.15	178	1.28 ± 0.87	0.90	130	1.18 ± 0.35		1.19 ± 0.34	0.59
>1	95	1.42 ± 1.09	96	1.34 ± 0.58	0.78	97	1.29 ± 0.94	96	1.35 ± 0.91	0.81	97	1.28 ± 0.38		1.27 ± 0.48	0.40
p[δ]		0.17		0.09			0.93		0.65			**0.04**		0.54	

[δ] P-value determined based on Wilcoxon rank sum test

Table 4. Association between selected variables and percent methylation levels in DNA repair genes among unaffected women

	BRCA1				MLH1				MSH2			
Age (>50 vs ≤50)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)	β	St.Err	t-value	Pr(>|t|)
Intercept	1.43	0.10	14.0	<0.01	1.32	0.07	20.2	<0.01	1.20	0.03	40.0	<0.01
Continuous	0.23	0.18	1.31	0.19	-0.05	0.11	-0.44	0.66	0.05	0.05	0.89	0.38
Intercept	0.91	0.16	5.64	<0.01	0.79	0.16	5.00	<0.01	0.65	0.16	4.16	<0.01
Tertile	-0.39	0.27	-1.45	0.15	-0.08	0.27	-0.30	0.77	0.06	0.27	0.21	0.83
Smoke (Ever vs. Never)												
Intercept	1.62	0.11	15.4	<0.01	1.28	0.07	19.3	<0.01	1.18	0.03	38.8	<0.01
Continuous	-0.30	0.17	-1.76	0.08	0.05	0.11	0.42	0.67	0.09	0.05	1.72	0.09
Intercept	1.07	0.17	6.18	<0.01	0.68	0.16	4.27	<0.01	0.53	0.16	3.38	<0.01
Tertile	-0.73	0.26	-2.77	**0.01**	0.25	0.27	0.93	0.35	0.39	0.27	1.45	0.15
BMI (kg/m^2) (≥25 vs <25)												
Intercept	1.34	0.12	11.0	<0.01	1.21	0.08	15.8	<0.01	1.17	0.04	33.4	<0.01
Continuous	0.30	0.17	1.78	0.08	0.17	0.11	1.56	0.12	0.08	0.05	1.64	0.10
Intercept	0.61	0.18	3.36	<0.01	0.67	0.18	3.66	<0.01	0.49	0.18	2.76	<0.01
Tertile	0.29	0.26	1.12	0.26	0.15	0.26	0.59	0.55	0.32	0.26	1.24	0.21
Menarche (<13 vs ≥13)												
Intercept	1.62	0.11	14.7	<0.01	1.25	0.07	17.5	<0.01	1.18	0.03	36.4	<0.01
Continuous	-0.26	0.17	-1.58	0.12	0.12	0.11	1.09	0.28	0.07	0.05	1.44	0.15
Intercept	1.07	0.18	5.88	<0.01	0.76	0.17	4.41	<0.01	0.44	0.17	2.65	<0.01
Tertile	-0.63	0.26	-2.44	**0.01**	0.01	0.26	0.02	0.98	0.55	0.26	2.11	**0.03**
Age at first parity (>=30 vs. <30)												
Intercept	1.50	0.10	14.7	<0.01	1.28	0.07	19.6	<0.01	1.24	0.03	42.0	<0.01
Continuous	-0.01	0.18	-0.03	0.97	0.06	0.11	0.52	0.60	-0.09	0.05	-1.70	0.09
Intercept	0.76	0.16	4.84	<0.01	0.78	0.16	4.95	<0.01	0.75	0.16	4.75	<0.01
Tertile	0.05	0.27	0.18	0.86	-0.06	0.27	-0.21	0.83	-0.23	0.27	-0.87	0.38
Family History (>1 vs=1)												
Intercept	1.59	0.10	15.4	<0.01	1.27	0.07	19.3	<0.01	1.18	0.03	39.0	<0.01
Continuous	-0.24	0.18	-1.40	0.16	0.08	0.11	0.70	0.49	0.09	0.05	1.72	0.09
Intercept	0.89	0.16	5.48	<0.01	0.77	0.16	4.80	<0.01	0.70	0.16	4.37	<0.01
Tertile	-0.31	0.27	-1.18	0.24	-0.01	0.27	-0.05	0.96	-0.08	0.26	-0.31	0.75

Table 5a. Conditional logistic regression analysis of the relationship between methylation levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer

BRCA1 (n=215 sets)	β	StdErr	Odds Ratio	95% Confidence Interval		Pr>z
Continuous	0.08	0.05	1.09	0.98	1.20	0.11
T3 (>1.59)	-0.01	0.31	0.99	0.54	1.82	0.97
T2 (1.06-1.59)	-0.09	0.29	0.92	0.52	1.63	0.77
T1 (≤1.06)	ref	ref	ref	ref	ref	ref
MLH1 (n=215 sets)						
Continuous	0.17	0.14	1.18	0.91	1.54	0.22
T3 (>1.27)	0.16	0.33	1.17	0.61	2.24	0.64
T2 (0.77-1.27)	-0.28	0.25	0.75	0.46	1.22	0.25
T1 (≤0.77)	ref	ref	ref	ref	ref	ref
MSH2 (n=217 sets)						
Continuous	-0.14	0.31	0.87	0.48	1.60	0.66
T3 (>1.25)	-0.20	0.29	0.82	0.47	1.45	0.49
T2 (1.04-1.25)	-0.06	0.23	0.95	0.61	1.48	0.81
T1 (≤1.04)	ref	ref	ref	ref	ref	ref

Table 5b. Conditional logistic regression analysis of the relationship between methylation levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer*

BRCA1 (n=215 sets)	β	StdErr	Odds Ratio	95% Confidence Interval		Pr>z
Continuous	0.08	0.05	1.09	0.98	1.20	0.11
T3 (>1.59)	0.00	0.31	1.00	0.54	1.83	0.99
T2 (1.06-1.59)	-0.10	0.30	0.91	0.51	1.62	0.73
T1 (≤1.06)	ref	ref	ref	ref	ref	ref
MLH1 (n=215 sets)						
Continuous	0.17	0.14	1.19	0.91	1.55	0.21
T3 (>1.27)	0.17	0.33	1.19	0.62	2.27	0.61
T2 (0.77-1.27)	-0.29	0.25	0.75	0.46	1.22	0.24
T1 (≤0.77)	ref	ref	ref	ref	ref	ref
MSH2 (n=217 sets)						
Continuous	-0.11	0.31	0.89	0.48	1.64	0.71
T3 (>1.25)	-0.18	0.29	0.83	0.47	1.48	0.53
T2 (1.04-1.25)	-0.06	0.23	0.94	0.60	1.47	0.79
T1 (≤1.04)	ref	ref	ref	ref	ref	ref

*Adjusted for age at blood draw and ever vs. never smoking status

Table 6a. GEE analysis of the relationship between promoter methylation levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer

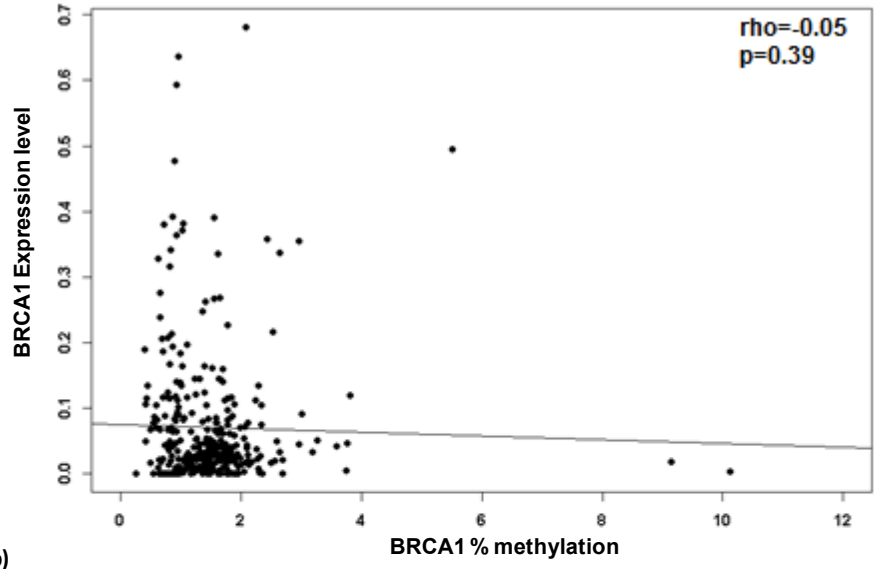
BRCA1 (n=197 sets)	β	StdErr	L'β	95% Confidence Interval		Pr>Wald
Continuous	0.05	0.04	1.05	0.96	1.14	0.27
T3 (>1.59)	-0.18	0.19	0.84	0.58	1.20	0.34
T2 (1.06-1.59)	-0.23	0.19	0.80	0.54	1.16	0.24
T1 (≤1.06)	ref	ref	ref	ref	ref	ref
MLH1 (n=196 sets)						
Continuous	0.07	0.08	1.07	0.93	1.25	0.34
T3 (>1.27)	0.04	0.18	1.04	0.73	1.47	0.83
T2 (0.77-1.27)	-0.16	0.21	0.85	0.56	1.28	0.44
T1 (≤0.77)	ref	ref	ref	ref	ref	ref
MSH2 (n=198 sets)						
Continuous	0.03	0.19	1.03	0.72	1.49	0.86
T3 (>1.25)	0.03	0.19	1.03	0.71	1.48	0.88
T2 (1.04-1.25)	0.02	0.21	1.02	0.67	1.55	0.93
T1 (≤1.04)	ref	ref	ref	ref	ref	ref

Table 6b. GEE analysis of the relationship between promoter methylation levels of various DNA repair genes and breast cancer risk among sisters discordant for breast cancer*

BRCA1 (n=197 sets)	β	StdErr	L'β	95% Confidence Interval		Pr>Wald
Continuous	0.05	0.04	1.05	0.96	1.14	0.28
T3 (>1.59)	-0.15	0.19	0.86	0.60	1.25	0.43
T2 (1.06-1.59)	-0.20	0.20	0.81	0.55	1.20	0.30
T1 (≤1.06)	ref	ref	ref	ref	ref	ref
MLH1 (n=196 sets)						
Continuous	0.07	0.08	1.07	0.92	1.25	0.36
T3 (>1.27)	0.04	0.18	1.04	0.73	1.47	0.84
T2 (0.77-1.27)	-0.17	0.21	0.85	0.56	1.28	0.43
T1 (≤0.77)	ref	ref	ref	ref	ref	ref
MSH2 (n=198 sets)						
Continuous	0.00	0.19	1.00	0.69	1.45	0.99
T3 (>1.25)	0.00	0.19	1.00	0.69	1.45	1.00
T2 (1.04-1.25)	0.00	0.22	1.00	0.65	1.53	1.00
T1 (≤1.04)	ref	ref	ref	ref	ref	ref

*Adjusted for age at blood draw and ever vs. never smoking status

(a)



(b)

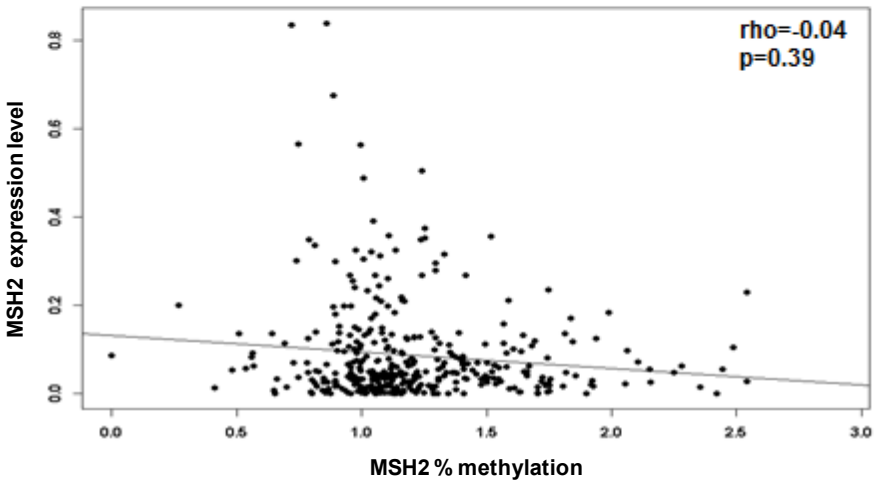


Figure 7. Correlation between expression level and percent methylation for (a) BRCA1 and (b) MSH2 within the sister-sets

Discussion

In this study, we found no association between the methylation levels of the interrogated CpG sites in the promoter regions of *BRCA1*, *MLH1* or *MSH2* and breast cancer risk. Furthermore, we found no significant correlation between methylation levels of either gene with previously assessed gene expression levels in the same individuals.

The lack of an association in our study does not preclude the possibility that variability in the promoter methylation levels of these genes may be relevant to breast cancer risk and expression. The inability to reveal such an association in our study may be a reflection of the inherent limitation in the utilized methodology, as reliable sequencing is constrained to 50 base pairs. It is possible that the resulting narrowed focus on a few CpG sites fell outside a region of greater relevance. Additionally, by design, primers are optimized to bind complementary regions in the template that do not contain CpG sites to avoid introducing methylation-specific amplification bias. Hence the targeted regions necessarily fall outside more CpG-dense areas, which may be of greater relevance for the regulation of transcriptional activity.

Given these considerations, we attempted to maximize the likelihood of targeting the most relevant sites based on previously published reports. However, unlike our study, most of these studies have focused on assessing methylation levels in tissue samples. It may be that the tumors of the cases in our study do exhibit increased methylation levels, however, these levels are not reflected in blood. This would indicate that the epigenetic deregulation of these genes resulting in carcinogenesis did not originate in the germline, but were rather established later in development, resulting in a more localized process that is not observed in other tissues. Since detection in a non-invasive surrogate tissue, such as blood, is the desired specimen for screening, such markers are not suitable for this purpose.

Snell et al did interrogate and determine case-control differences in the methylation levels of *BRCA1* in blood, however, total white blood cells (WBC) were used as the cell source[155]. This cell population is a composite of different blood cell types with unique methylation profiles, and it has been shown that carcinogenesis induces myeloid skewing, impacting the proportion of the constituent cell types present

in blood[156]. Therefore, differential methylation observed in WBC may be an effect rather than a cause of the tumorigenic process.

To avoid this potential bias, we restricted the analysis in the current study to PBMCs, a long-lived blood cell-type. Focusing our analysis on this single cell-type, the methylation levels detected across the interrogated CpG sites for all three genes in our study generally fell below 5%, indicating that these genes may be constitutively expressed outside the target tissue, regardless of case status.

Furthermore, our case population likely differs from the breast cancer cases selected in these previously reported studies. In the Snell et al study, cases were restricted to those with tumor pathologies similar to *BRCA1* mutant tumors but with no known *BRCA1* germline mutations[155]. If, in fact, *BRCA1*-methylation positive tumors account for only a small proportion of all cases, this would explain our inability to detect case-control differences in an unselected population.

While, as in our study, the search for cancer-specific alterations in gene-specific methylation levels typically focus on CpG islands in promoter regions, several studies in recent years reporting on methylation profiles based on genome-wide scans have been challenging this convention[157,158]. These studies contend that the CpG sites most differentially methylated in both tumor vs. nontumor tissues as well as in the bloods of cases and controls, are, in fact, in less dense CpG regions up to 2 kilobases away from the promoter. In these studies, methylation levels of sites located within these "CpG shores" have also been shown to be more closely associated with gene expression level than sites in promoter-region CpG islands[158]. If indeed correct, the findings from these studies suggest that our inability to determine both case-control differences and significant correlation with expression levels may reflect our specification on sites located within CpG islands in the promoter-region of our targets of interest.

While it is possible that expression is regulated by methylation of sites not captured in our study, the lack of a substantial correlation between expression levels of both *BRCA1* and *MSH2* and promoter methylation levels of these genes may also indicate that the downregulation in gene expression observed in our previous study is actually regulated by a mechanism other than promoter methylation of CpG sites. Such processes include alternate means of transcriptional regulation, such as histone

modifications, or post-transcriptional processing, such as microRNA (miRNA) mediated degradation of mRNA transcripts.

Hence, as our study did not determine case-control differences in PBMCs within the interrogated sites for the three genes of interest, future studies should explore alternative routes of expression regulation, including miRNAs. Of particular interest are two DNA repair-related genes, *ATM* and *MSH2*, for which downregulation of gene expression was previously shown to be associated with breast cancer risk in this study population.

Chapter VI: Conclusions and future directions

The studies presented in this dissertation attempted to identify endpoints across DNA repair pathways with the potential to develop into breast cancer screening biomarkers. For this purpose, Chapter II focused on the development of an ELISA phenotyping assay to determine BPDE-DNA adduct removal capacity within LCLs as an indicator of overall NER activity. In this study, we were unable to determine case-control differences previously observed using an immunohistochemical approach to determine NER capacity within the same population. Due to the difficulties in establishing a phenotyping methodology applicable in a population-wide setting, we turned our focus towards more specific endpoints across the various repair pathways for which scalable methodologies for assessment have been established.

In Chapter III, we conducted a gene expression study assessing the impact on breast cancer risk due to reductions in *ATM* (DSBR), *BRCA1* (DSBR), *MSH2* (MMR), *MutY* (BER), and *XPC* (NER) expression levels. In this study, women in the lowest tertile of expression for *MSH2* as well as *ATM* had a two-fold increase in odds of breast cancer compared to women in the highest tertile of expression for these genes. Interestingly, the risk imparted by reduced expression of *ATM* was heightened among women with an extended family history of breast cancer.

As the results in Chapter III indicated a role in breast cancer development due to transcriptional and post-transcriptional regulation of DNA repair gene activity, we attempted to delineate upstream factors which may mediate the transcriptional activity of DNA repair genes. Hence, in Chapter IV, we investigated the impact on breast cancer risk due to differential promoter methylation levels of *BRCA1*, *MLH1* and *MSH2*, DNA repair genes for which promoter methylation levels have previously been shown to differentiate breast cancer cases from controls. While a majority of these studies compared tumor and non-tumor tissues, we found no association between methylation levels of the interrogated CpG sites and breast cancer risk as assessed in PBMCs. This suggests that the sites included in our study, while possibly relevant for tumorigenesis, are not suitable for screening in more accessible biospecimens.

Finally, reduced functional activity can be mediated by means other than transcriptional regulation, including altered configuration of the final protein product. As such modifications in the intended protein structure can be relayed through variants in the genetic code, we examined the impact on breast cancer risk due to the presence of polymorphisms in MMR-related genes in Chapter V. In this study, we observed an increase in risk due to SNP variants in *MSH2* and *MutY*. A protective effect due to a variant in *MSH3* was restricted to women without an extended family history of breast cancer.

In summary, targets relevant to breast cancer were identified spanning several DNA repair pathways (Figure 1). While *ATM,* a component of DSBR*,* has been previously established as a breast cancer susceptibility gene, the other targets, *MSH2* and *MutY*, participate in pathways in which deficiencies are not typically linked with breast cancer. The observation that these endpoints bear relevance to breast cancer risk in our studies may reflect the unique characteristics of our study population. Contrary to other studies which have not been able to convincingly link MMR components to sporadic breast cancer, the participants enrolled in the NY-BCFR stem from high-risk families. The findings in this study population in addition to the role of extent of family history on modulating the risk imparted by several of the identified targets, suggest that susceptibility to familial breast cancer may be partially driven by the deregulation across DNA repair pathways.

Caution needs to be exerted in the inferences drawn from these findings due to the limitations inherent in each of the conducted studies. The null findings of the phenotyping study largely reflects the difficulties in developing a standardized, scalable protocol at the cost of inducing an artificial environment in which comprehensive capture of host activity becomes unfeasible. Similarly, the null findings in the methylation study can also be partially attributed to limitations imposed by the chosen high-throughput platform, restricting the focus of interrogation on a few sites and possibly biasing away from more relevant targets.

While the gene expression and genotyping studies did yield positive findings, the contribution to risk due to the identified endpoints were of moderate effect size. This is likely due to the focus on isolated components within the pathways. Additionally, both studies utilized a candidate approach, narrowing the focus on a few selected sites, at the exclusion of others which may yield a greater impact on risk.

This framework prevents accounting for the actions of other participants within the system. As a result, the actual relevance of the identified targets within the true context in which they mediate their actions, cannot be determined from our studies. Finally, as all findings stem from retrospective case-control studies, the temporality of the associations observed remains a concern.

Still, these findings point towards additional avenues to explore for future studies within a larger, prospectively followed study population. The increase in risk we observed due to deficiencies in *MSH2* and *MutY*, two genes most typically associated with Lynch syndrome related cancers, highlight the need to further clarify the heavily debated notion of whether familial breast cancer should be considered a constituent of the syndrome. To this end, participants within the BCFR could be further characterized for the presence of suboptimal MMR activity beyond the specific endpoints chosen for the current studies, and those found to be deficient could be monitored for the elevated presentation of Lynch syndrome cancer sub-types.

Finally, as reduced ATM activity results in a heightened sensitivity to ionizing radiation, the finding relating ATM deficiency with increased breast cancer risk in our study may also be of consequence to risk of secondary cancers due to treatment of the primary cancer. Hence, following individuals with reduced expression of this gene who go on to develop breast cancer for the elevated risk of additional cancers following treatment will be relevant to develop a more comprehensive risk profile for those who present with this deficiency.

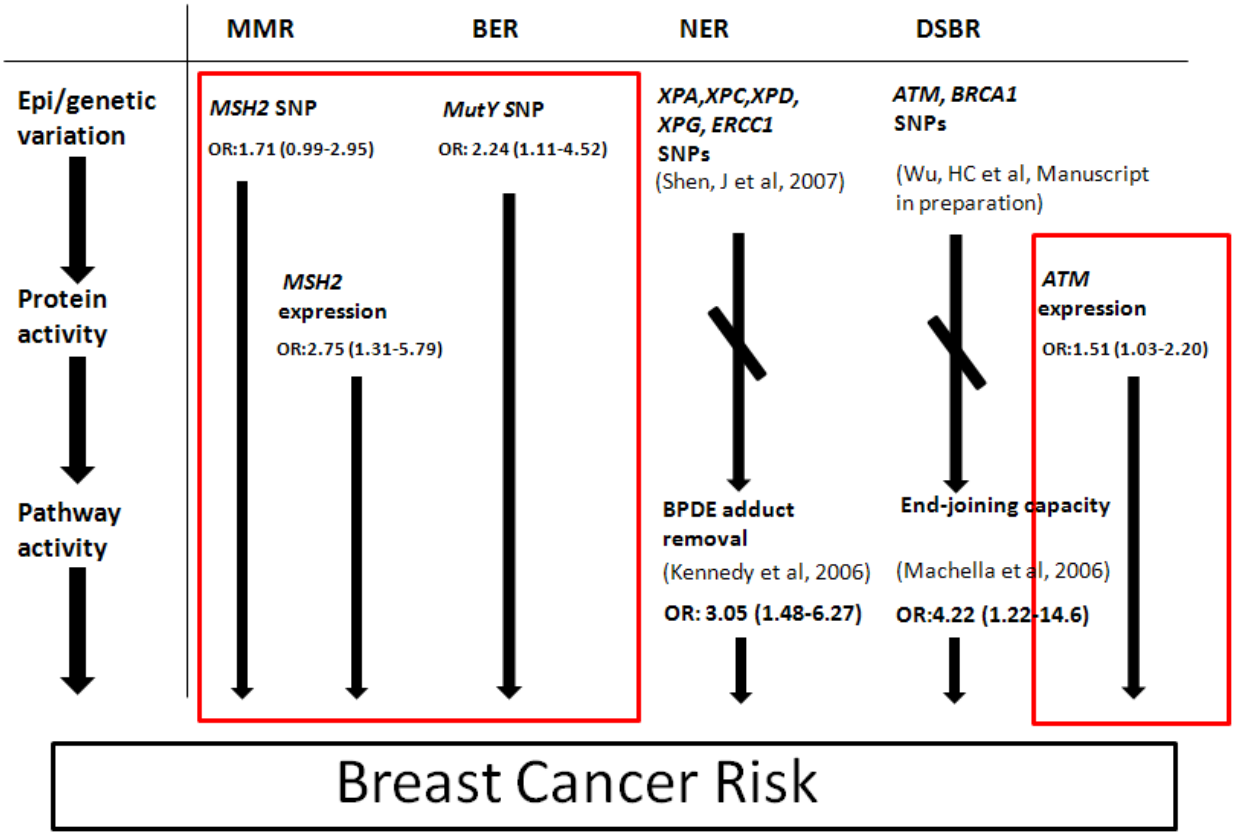


Figure 1. Summary of DNA repair-related findings within the NY-BCFR
Our studies suggest variants in *MSH2* (MMR) and *MUTYH* (BER) contribute to breast cancer risk. Additionally, women in the lowest tertile of expression for MSH2 and ATM (DSBR), are also at heightened risk for breast cancer

References

1. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA: a cancer journal for clinicians* **62**, 10–29

2. Kolb, T. M., Lichy, J. & Newhouse, J. H. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* **225**, 165–75 (2002).

3. Elmore, J. G. *et al.* Ten-year risk of false positive screening mammograms and clinical breast examinations. *The New England journal of medicine* **338**, 1089–96 (1998).

4. Kerlikowske, K. *et al.* Positive predictive value of screening mammography by age and family history of breast cancer. *JAMA : the journal of the American Medical Association* **270**, 2444–50 (1993).

5. Gøtzsche, P. C. & Nielsen, M. Screening for breast cancer with mammography. *Cochrane database of systematic reviews (Online)* CD001877 (2011).doi:10.1002/14651858.CD001877.pub4

6. Christiansen, C. L. *et al.* Predicting the cumulative risk of false-positive mammograms. *Journal of the National Cancer Institute* **92**, 1657–66 (2000).

7. Miller, A. B., Baines, C. J., To, T. & Wall, C. Screening mammography re-evaluated. *Lancet* **355**, 747; author reply 752 (2000).

8. Buist, D. S. M., Porter, P. L., Lehman, C., Taplin, S. H. & White, E. Factors contributing to mammography failure in women aged 40-49 years. *Journal of the National Cancer Institute* **96**, 1432–40 (2004).

9. Sismondi, P., Biglia, N., Giai, M., Sgro, L. & Campagnoli, C. Metabolic effects of tamoxifen in postmenopause. *Anticancer research* **14**, 2237–44

10. Singletary, S. E. Rating the risk factors for breast cancer. *Annals of surgery* **237**, 474–82 (2003).

11. Pitot, H. C. Carcinogenesis and aging--two related phenomena? A review. *The American journal of pathology* **87**, 444–72 (1977).

12. Dite, G. S. *et al.* Familial Risks, Early-Onset Breast Cancer, and BRCA1 and BRCA2 Germline Mutations. *JNCI Journal of the National Cancer Institute* **95**, 448–457 (2003).

13. McClain, Monica R; Palomaki, Glenn E; Nathanson, Katherine L; Haddow, J. E. Adjusting the estimated proportion of breast cancer cases associated with BRCA1 and BRCA2 mutations: public health implications. *Genetics in medicine : official journal of the American College of Medical Genetics* **7**, 28–33 (2005).

14. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* **358**, 1389–99 (2001).

15. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England journal of medicine* **343**, 78–85 (2000).

16. Ahlbom, A. *et al.* Cancer in twins: genetic and nongenetic familial risk factors. *Journal of the National Cancer Institute* **89**, 287–93 (1997).

17. Easton, D. F., Ford, D. & Bishop, D. T. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American journal of human genetics* **56**, 265–71 (1995).

18. Rahman, N. & Stratton, M. R. The genetics of breast cancer susceptibility. *Annu Rev Genet* **32**, 95–121 (1998).

19. Brinton, L. A., Schairer, C., Hoover, R. N. & Fraumeni, J. F. Menstrual Factors and Risk of Breast Cancer. (2009).at <http://informahealthcare.com/doi/abs/10.3109/07357908809080645>

20. MacMahon, B. *et al.* Age at first birth and breast cancer risk. *Bulletin of the World Health Organization* **43**, 209–21 (1970).

21. Lambe, M. *et al.* Maternal risk of breast cancer following multiple births: a nationwide study in Sweden. *Cancer causes & control : CCC* **7**, 533–8 (1996).

22. Russo, J., Moral, R., Balogh, G. A., Mailo, D. & Russo, I. H. The protective role of pregnancy in breast cancer. *Breast cancer research : BCR* **7**, 131–42 (2005).

23. Steinberg, K. K. *et al.* A meta-analysis of the effect of estrogen replacement therapy on the risk of breast cancer. *JAMA : the journal of the American Medical Association* **265**, 1985–90 (1991).

24. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* **350**, 1047–59 (1997).

25. Tretli, S. Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570,000 women in Norway. *International journal of cancer. Journal international du cancer* **44**, 23–30 (1989).

26. Ursin, G., Longnecker, M. P., Haile, R. W. & Greenland, S. A meta-analysis of body mass index and risk of premenopausal breast cancer. *Epidemiology (Cambridge, Mass.)* **6**, 137–41 (1995).

27. Nelson, L. R. & Bulun, S. E. Estrogen production and action. *Journal of the American Academy of Dermatology* **45**, S116–24 (2001).

28. Kaaks, R. *et al.* Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocrine-related cancer* **12**, 1071–82 (2005).

29. Rose, D. P. & Vona-Davis, L. Influence of obesity on breast cancer receptor status and prognosis. *Expert review of anticancer therapy* **9**, 1091–101 (2009).

30. Boice, J. D., Preston, D., Davis, F. G. & Monson, R. R. Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. *Radiation research* **125**, 214–22 (1991).

31. Clemons, M., Loijens, L. & Goss, P. Breast cancer risk following irradiation for Hodgkin's disease. *Cancer treatment reviews* **26**, 291–302 (2000).

32. Gail, M. H. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* **81**, 1879–86 (1989).

33. Prout, M. N. Breast cancer risk reduction: what do we know and where should we go? *Medscape women's health* **5**, E4

34. Beckmann, M. W., Niederacher, D., Schnürch, H.-G., Gusterson, B. A. & Bender, H. G. Multistep carcinogenesis of breast cancer and tumour heterogeneity. *Journal of Molecular Medicine* **75**, 429–439 (1997).

35. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–74 (2011).

36. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer research* **51**, 3075–9 (1991).

37. WATSON, J. D. & CRICK, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–7 (1953).

38. Lindahl, T. & Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **13**, 3405–10 (1974).

39. Mullaart, E., Lohman, P. H., Berends, F. & Vijg, J. DNA damage metabolism and aging. *Mutation research* **237**, 189–210

40. Loeb, L. A. & Cheng, K. C. Errors in DNA synthesis: a source of spontaneous mutations. *Mutation research* **238**, 297–304 (1990).

41. Dizdaroglu, M. & Jaruga, P. Mechanisms of free radical-induced damage to DNA. *Free radical research* **46**, 382–419 (2012).

42. Bonura, T., Smith, K. C. & Kaplan, H. S. Enzymatic induction of DNA double-strand breaks in gamma-irradiated Escherichia coli K-12. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 4265–9 (1975).

43. Perera, F. P. *et al.* Detection of polycyclic aromatic hydrocarbon-DNA adducts in white blood cells of foundry workers. *Cancer research* **48**, 2288–91 (1988).

44. SETLOW, R. B. & CARRIER, W. L. THE DISAPPEARANCE OF THYMINE DIMERS FROM DNA: AN ERROR-CORRECTING MECHANISM. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 226–31 (1964).

45. BOYCE, R. P. & HOWARD-FLANDERS, P. RELEASE OF ULTRAVIOLET LIGHT-INDUCED THYMINE DIMERS FROM DNA IN E. COLI K-12. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 293–300 (1964).

46. PETTIJOHN, D. & HANAWALT, P. EVIDENCE FOR REPAIR-REPLICATION OF ULTRAVIOLET DAMAGED DNA IN BACTERIA. *Journal of molecular biology* **9**, 395–410 (1964).

47. Kraemer, K. H., Lee, M. M., Andrews, A. D. & Lambert, W. C. The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm. *Arch Dermatol* **130**, 1018–1021 (1994).

48. Setlow, R. B., Regan, J. D., German, J. & Carrier, W. L. Evidence that xeroderma pigmentosum cells do not perform the first step in the repair of ultraviolet damage to their DNA. *Proceedings of the National Academy of Sciences of the United States of America* **64**, 1035–41 (1969).

49. Kraemer, K. H. *et al.* Genetic heterogeneity in xeroderma pigmentosum: complementation groups and their relationship to DNA repair rates. *Proc Natl Acad Sci U S A* **72**, 59–63 (1975).

50. Lindahl, T. New class of enzymes acting on damaged DNA. *Nature* **259**, 64–6

51. Modrich, P. DNA mismatch correction. *Annual review of biochemistry* **56**, 435–66 (1987).

52. Nouspikel, T. DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cellular and molecular life sciences : CMLS* **66**, 994–1009 (2009).

53. Hanawalt, P. C. Subpathways of nucleotide excision repair and their regulation. *Oncogene* **21**, 8949–56 (2002).

54. Masutani, C. *et al.* Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *The EMBO journal* **13**, 1831–43 (1994).

55. Nishi, R. *et al.* Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Molecular and cellular biology* **25**, 5664–74 (2005).

56. Schaeffer, L. *et al.* The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *The EMBO journal* **13**, 2388–92 (1994).

57. Lao, Y., Gomes, X. V, Ren, Y., Taylor, J. S. & Wold, M. S. Replication protein A interactions with DNA. III. Molecular basis of recognition of damaged DNA. *Biochemistry* **39**, 850–9 (2000).

58. Buschta-Hedayat, N., Buterin, T., Hess, M. T., Missura, M. & Naegeli, H. Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6090–5 (1999).

59. O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371**, 432–5 (1994).

60. De Laat, W. L., Appeldoorn, E., Jaspers, N. G. & Hoeijmakers, J. H. DNA structural elements required for ERCC1-XPF endonuclease activity. *The Journal of biological chemistry* **273**, 7835–42 (1998).

61. Shivji, M. K., Podust, V. N., Hübscher, U. & Wood, R. D. Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry* **34**, 5011–7 (1995).

62. Ogi, T. *et al.* Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular cell* **37**, 714–27 (2010).

63. Kelman, Z. PCNA: structure, functions and interactions. *Oncogene* **14**, 629–40 (1997).

64. Moser, J. *et al.* Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Molecular cell* **27**, 311–23 (2007).

65. Fousteri, M., Vermeulen, W., Van Zeeland, A. A. & Mullenders, L. H. F. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Molecular cell* **23**, 471–82 (2006).

66. Lainé, J.-P. & Egly, J.-M. When transcription and repair meet: a complex system. *Trends in genetics : TIG* **22**, 430–6 (2006).

67. Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M. & Slupphaug, G. Base excision repair of DNA in mammalian cells. *FEBS Letters* **476**, 73–77 (2000).

68. Fortini, P. *et al.* The base excision repair: mechanisms and its relevance for cancer susceptibility. *Biochimie* **85**, 1053–71 (2003).

69. Nilsen, H. & Lindahl, T. *Encyclopedia of Biological Chemistry*. *null* **null**, 603–608 (Elsevier: New York, 2004).

70. McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell research* **18**, 148–61 (2008).

71. Kunz, C., Saito, Y. & Schär, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cellular and molecular life sciences : CMLS* **66**, 1021–38 (2009).

72. Keeney, S. & Neale, M. J. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochemical Society transactions* **34**, 523–5 (2006).

73. Soulas-Sprauel, P. *et al.* V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. *Oncogene* **26**, 7780–91 (2007).

74. Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annual review of genetics* **40**, 363–83 (2006).

75. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell* **47**, 497–510 (2012).

76. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181–211 (2010).

77. Blier, P. R., Griffith, A. J., Craft, J. & Hardin, J. A. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *The Journal of biological chemistry* **268**, 7594–601 (1993).

78. Yaneva, M., Kowalewski, T. & Lieber, M. R. Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *The EMBO journal* **16**, 5098–112 (1997).

79. Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Molecular and cellular biology* **20**, 2996–3003 (2000).

80. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–94 (2002).

81. Moon, A. F. *et al.* The X family portrait: structural insights into biological functions of X family polymerases. *DNA repair* **6**, 1709–25 (2007).

82. Hopfner, K.-P. *et al.* The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**, 562–6 (2002).

83. Paull, T. T. & Gellert, M. The 3′ to 5′ Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. *Molecular Cell* **1**, 969–979 (1998).

84. Lee, J.-H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science (New York, N.Y.)* **308**, 551–4 (2005).

85. Zou, Y., Liu, Y., Wu, X. & Shell, S. M. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *Journal of cellular physiology* **208**, 267–73 (2006).

86. Esashi, F., Galkin, V. E., Yu, X., Egelman, E. H. & West, S. C. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nature structural & molecular biology* **14**, 468–74 (2007).

87. Chen, J. *et al.* Stable Interaction between the Products of the BRCA1 and BRCA2 Tumor Suppressor Genes in Mitotic and Meiotic Cells. *Molecular Cell* **2**, 317–328 (1998).

88. Sy, S. M. H., Huen, M. S. Y. & Chen, J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7155–60 (2009).

89. Setlow, R. B. Repair deficient human disorders and cancer. *Nature* **271**, 713–7 (1978).

90. Broughton, B. C. *et al.* Molecular and cellular analysis of the DNA repair defect in a patient in xeroderma pigmentosum complementation group D who has the clinical features of xeroderma pigmentosum and Cockayne syndrome. *American journal of human genetics* **56**, 167–74 (1995).

91. Kobayashi, T. *et al.* Mutations in the XPD gene leading to xeroderma pigmentosum symptoms. *Human mutation* **9**, 322–31 (1997).

92. Decordier, I., Loock, K. V & Kirsch-Volders, M. Phenotyping for DNA repair capacity. *Mutat Res* **705**, 107–129 (2010).

93. Qiao, Y. *et al.* Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes. *Mutation research* **509**, 165–74 (2002).

94. Caple, F. *et al.* Inter-individual variation in DNA damage and base excision repair in young, healthy non-smokers: effects of dietary supplementation and genotype. *The British journal of nutrition* **103**, 1585–93 (2010).

95. Qiao, Y. *et al.* Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic XPC and XPD/ERCC2 genotypes. *Carcinogenesis* **23**, 295–9 (2002).

96. Tang, D. *et al.* Polymorphisms in the DNA repair enzyme XPD are associated with increased levels of PAH-DNA adducts in a case-control study of breast cancer. *Breast cancer research and treatment* **75**, 159–66 (2002).

97. Hou, S.-M. *et al.* The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. *Carcinogenesis* **23**, 599–603 (2002).

98. Wei, Q., Matanoski, G. M., Farmer, E. R., Hedayati, M. A. & Grossman, L. DNA repair and aging in basal cell carcinoma: a molecular epidemiology study. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 1614–8 (1993).

99. Wei, Q. *et al.* Association between low dietary folate intake and suboptimal cellular DNA repair capacity. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **12**, 963–9 (2003).

100. Radák, Z. *et al.* Marathon running alters the DNA base excision repair in human skeletal muscle. *Life sciences* **72**, 1627–33 (2003).

101. Shen, H. *et al.* Smoking, DNA repair capacity and risk of nonsmall cell lung cancer. *International journal of cancer. Journal international du cancer* **107**, 84–8 (2003).

102. Fu, Y.-P. *et al.* Breast Cancer Risk Associated with Genotypic Polymorphism of the Nonhomologous End-Joining Genes: A Multigenic Study on Cancer Susceptibility. *Cancer Res.* **63**, 2440–2446 (2003).

103. Patel, R. K., Trivedi, A. H., Arora, D. C., Bhatavdekar, J. M. & Patel, D. D. DNA repair proficiency in breast cancer patients and their first-degree relatives. *International journal of cancer. Journal international du cancer* **73**, 20–4 (1997).

104. Bau, D.-T., Mau, Y.-C., Ding, S.-L., Wu, P.-E. & Shen, C.-Y. DNA double-strand break repair capacity and risk of breast cancer. *Carcinogenesis* **28**, 1726–30 (2007).

105. Machella, N. *et al.* Double-strand breaks repair in lymphoblastoid cell lines from sisters discordant for breast cancer from the New York site of the BCFR. *Carcinogenesis* **29**, 1367–1372 (2008).

106. Bootsma, D. *et al.* Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **347**, 75–81 (1995).

107. Kraemer, K. H. *et al.* Xeroderma pigmentosum and related disorders: examining the linkage between defective DNA repair and cancer. *J Invest Dermatol* **103**, 96S–101S (1994).

108. Ramos, J. M. *et al.* DNA repair and breast carcinoma susceptibility in women. *Cancer* **100**, 1352–7 (2004).

109. Kennedy, D. O. *et al.* DNA repair capacity of lymphoblastoid cell lines from sisters discordant for breast cancer. *J Natl Cancer Inst* **97**, 127–132 (2005).

110. Lynch, H. T. & Smyrk, T. Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. *Cancer* **78**, 1149–67 (1996).

111. Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science (New York, N.Y.)* **260**, 816–9 (1993).

112. Papadopoulos, N. *et al.* Mutation of a mutL homolog in hereditary colon cancer. *Science (New York, N.Y.)* **263**, 1625–9 (1994).

113. Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–38 (1993).

114. Nicolaides, N. C. *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**, 75–80 (1994).

115. Yee, C. J., Roodi, N., Verrier, C. S. & Parl, F. F. Microsatellite instability and loss of heterozygosity in breast cancer. *Cancer research* **54**, 1641–4 (1994).

116. T. Paz-Elizur, M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, Z. L. DNA repair activity for oxidative damage and risk of lung cancer. *J. Natl. Cancer Inst.* **95**, 1312–1319 (2003).

117. Helzlsouer, K. J. *et al.* DNA repair proficiency: potential susceptiblity factor for breast cancer. *J Natl Cancer Inst* **88**, 754–755 (1996).

118. Phillips, D. H. Fifty years of benzo(a)pyrene. *Nature* **303**, 468–472 (1983).

119. Friedberg, E. C. How nucleotide excision repair protects against cancer. *Nat Rev Cancer* **1**, 22–33 (2001).

120. Neumann, A. S., Sturgis, E. M. & Wei, Q. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Molecular carcinogenesis* **42**, 65–92 (2005).

121. Li, D., Wang, M., Dhingra, K. & Hittelman, W. N. Aromatic DNA adducts in adjacent tissues of breast cancer patients: clues to breast cancer etiology. *Cancer research* **56**, 287–93 (1996).

122. Rundle, A. *et al.* Molecular epidemiologic studies of polycyclic aromatic hydrocarbon-DNA adducts and breast cancer. *Environmental and molecular mutagenesis* **39**, 201–7 (2002).

123. el-Bayoumy, K. Environmental carcinogens that may be involved in human breast cancer etiology. *Chem Res Toxicol* **5**, 585–590 (1992).

124. Calaf, G. & Russo, J. Transformation of human breast epithelial cells by chemical carcinogens. *Carcinogenesis* **14**, 483–492 (1993).

125. Tang, M. S., Pierce, J. R., Doisy, R. P., Nazimiec, M. E. & MacLeod, M. C. Differences and similarities in the repair of two benzo[a]pyrene diol epoxide isomers induced DNA adducts by uvrA, uvrB, and uvrC gene products. *Biochemistry* **31**, 8429–8436 (1992).

126. Weinstein, I. B. *et al.* Benzo(a)pyrene diol epoxides as intermediates in nucleic acid binding in vitro and in vivo. *Science* **193**, 592–595 (1976).

127. John, E. M. *et al.* The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast cancer research : BCR* **6**, R375–89 (2004).

128. Santella, R. M., Lin, C. D., Cleveland, W. L. & Weinstein, I. B. Monoclonal antibodies to DNA modified by a benzo[a]pyrene diol epoxide. *Carcinogenesis* **5**, 373–7 (1984).

129. Spitz, M. R., Wei, Q., Dong, Q., Amos, C. I. & Wu, X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **12**, 689–98 (2003).

130. Gudmundsdottir, K. & Ashworth, A. The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* **25**, 5864–5874 (2006).

131. Rahman, N. & Stratton, M. R. The genetics of breast cancer susceptibility. *Annual review of genetics* **32**, 95–121 (1998).

132. Ahmed, M. & Rahman, N. ATM and breast cancer susceptibility. *Oncogene* **25**, 5906–11 (2006).

133. Sourvinos, G., Kiaris, H., Tsikkinis, A., Vassilaros, S. & Spandidos, D. A. Microsatellite instability and loss of heterozygosity in primary breast tumours. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **18**, 157–66 (1997).

134. Dheda, K. *et al.* Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Group* **37**, (2004).

135. Swift, M., Reitnauer, P. J., Morrell, D. & Chase, C. L. Breast and other cancers in families with ataxia-telangiectasia. *The New England journal of medicine* **316**, 1289–94 (1987).

136. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature genetics* **38**, 873–5 (2006).

137. Foroughizadeh, M., Mozdarani, H., Majidzadeh-A, K. & Kaviani, A. Variation of ATM Gene Expression in Peripheral Blood Cells of Sporadic Breast Carcinomas in Iranian Patients. *Avicenna journal of medical biotechnology* **4**, 95–101 (2012).

138. Teraoka, S. N. *et al.* Increased frequency of ATM mutations in breast carcinoma patients with early onset disease and positive family history. *Cancer* **92**, 479–87 (2001).

139. Win, A. K. *et al.* Colorectal and other cancer risks for carriers and noncarriers from families with a DNA mismatch repair gene mutation: a prospective cohort study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 958–64 (2012).

140. Buerki, N. *et al.* Evidence for breast cancer as an integral part of Lynch syndrome. *Genes, chromosomes & cancer* **51**, 83–91 (2012).

141. Win, A. K. *et al.* Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. *Journal of the National Cancer Institute* **104**, 1363–72 (2012).

142. Walsh, M. D. *et al.* Lynch syndrome-associated breast cancers: clinicopathologic characteristics of a case series from the colon cancer family registry. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 2214–24 (2010).

143. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA: a cancer journal for clinicians* **63**, 11–30 (2013).

144. Cooper, D. N. Eukaryotic DNA methylation. *Human genetics* **64**, 315–33 (1983).

145. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–13 (1986).

146. Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M. & Issa, J. P. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Advances in cancer research* **72**, 141–96 (1998).

147. Trosko, J. E. & Chu, E. H. The role of DNA repair and somatic mutation in carcinogenesis. *Advances in cancer research* **21**, 391–425 (1975).

148. Bowcock, A. M. Molecular cloning of BRCA1: a gene for early onset familial breast and ovarian cancer. *Breast cancer research and treatment* **28**, 121–35 (1993).

149. Hemminki, A. *et al.* Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nature genetics* **8**, 405–10 (1994).

150. Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215–25 (1993).

151. Iwamoto, T., Yamamoto, N., Taguchi, T., Tamaki, Y. & Noguchi, S. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. *Breast cancer research and treatment* **129**, 69–77 (2011).

152. Esteller, M. *et al.* Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Journal of the National Cancer Institute* **92**, 564–9 (2000).

153. Thibodeau, S. N. *et al.* Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer research* **58**, 1713–8 (1998).

154. Murata, H., Khattar, N. H., Kang, Y., Gu, L. & Li, G. M. Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. *Oncogene* **21**, 5696–5703 (2002).

155. Snell, C., Krypuy, M., Wong, E. M., Loughrey, M. B. & Dobrovic, A. BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast cancer research : BCR* **10**, R12 (2008).

156. Yamanaka, T. *et al.* The baseline ratio of neutrophils to lymphocytes is associated with patient prognosis in advanced gastric cancer. *Oncology* **73**, 215–20 (2007).

157. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics* **41**, 178–86 (2009).

158. Teschendorff, A. E. *et al.* An epigenetic signature in peripheral blood predicts active ovarian cancer. *PloS one* **4**, e8274 (2009).

159. Tutt, A. & Ashworth, A. The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. *Trends Mol Med* **8**, 571–576 (2002).

160. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev* **14**, 927–939 (2000).

161. Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227–1236 (1993).

162. Loeb, L. A. Microsatellite instability: marker of a mutator phenotype in cancer. *Cancer Res* **54**, 5059–5063 (1994).

163. Shaw, J. A. *et al.* Microsatellite instability in early sporadic breast cancer. *Br J Cancer* **73**, 1393–1397 (1996).

164. Conde, J. *et al.* Association of common variants in mismatch repair genes and breast cancer susceptibility: a multigene study. *BMC Cancer* **9**, 344 (2009).

165. Wang, D. *et al.* Polymorphisms in MSH2 gene and risk of gastric cancer, and interactions with lifestyle factors in a Chinese population. *Cancer epidemiology* **36**, e171–6 (2012).

166. Picelli, S. *et al.* Common variants in human CRC genes as low-risk alleles. *European journal of cancer (Oxford, England : 1990)* **46**, 1041–8 (2010).

167. Tournier, I. *et al.* A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Human mutation* **29**, 1412–24 (2008).