



5-2008

Molecular Profile of Women With and Without Secondary Breast Cancer After the Treatment of Pediatric Hodgkin Lymphoma

Belinda Neal Mandrell

University of Tennessee Health Science Center

Follow this and additional works at: <https://dc.uthsc.edu/dissertations>



Part of the [Neoplasms Commons](#), and the [Nursing Commons](#)

Recommended Citation

Mandrell, Belinda Neal, "Molecular Profile of Women With and Without Secondary Breast Cancer After the Treatment of Pediatric Hodgkin Lymphoma" (2008). *Theses and Dissertations (ETD)*. Paper 165. <http://dx.doi.org/10.21007/etd.cghs.2008.0197>.

This Dissertation is brought to you for free and open access by the College of Graduate Health Sciences at UTHSC Digital Commons. It has been accepted for inclusion in Theses and Dissertations (ETD) by an authorized administrator of UTHSC Digital Commons. For more information, please contact jwelch30@uthsc.edu.

Molecular Profile of Women With and Without Secondary Breast Cancer After the Treatment of Pediatric Hodgkin Lymphoma

Document Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Program

Nursing

Research Advisor

Ann K. Cashion, PhD

Committee

Carolyn Driscoll, PhD Pamela Hinds, PhD Melissa Hudson, MD Geoffrey Neale, PhD Mary Relling, PharmD

DOI

10.21007/etd.cghs.2008.0197

**THE MOLECULAR PROFILE OF WOMEN WITH AND WITHOUT
SECONDARY BREAST CANCER AFTER THE TREATMENT OF
PEDIATRIC HODGKIN LYMPHOMA**

A Dissertation
Presented for
The Graduate Studies Council
The University of Tennessee
Health Science Center

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy
From The University of Tennessee

By
Belinda Neal Mandrell
May 2008

Copyright © Belinda Neal Mandrell, 2008
All rights reserved

DEDICATION

This dissertation is dedicated to the women who participated in this study and shared their treatment experiences. Each of them touched me through their courage and sincere desire to help others who struggle with the morbidity of childhood cancer and late effects.

ACKNOWLEDGMENTS

I would like to thank my dissertation chair, Dr. Ann Cashion, for her support, encouragement, and guidance. I would also like to thank my dissertation committee members, Dr. Carolyn Driscoll, Dr. Pamela Hinds, Dr. Melissa Hudson, Dr. Geoff Neale, and Dr. Mary Relling for their expertise and insightful contribution to my dissertation. I would also like to thank Dr. Stanley Pounds for his contribution to the genomic analysis, Dr. Shes Rai for statistical design of study and excellent selection of matched controls, and Mr. Shengping Yang for computational assistance in statistical analysis. I would also like to acknowledge the support provided by the NIH/NINR through a Ruth L. Kirchstein NRSA predoctoral fellowship, the American Cancer Society through the predoctoral nursing scholarship, and St. Jude Children's Research Hospital.

ABSTRACT

Identification of genetic risk factors associated with the development of secondary cancers would facilitate identification of at risk patients and permit modification of therapy and heightened surveillance that may reduce cancer-related morbidity and mortality. Women survivors of pediatric Hodgkin lymphoma (HL) have an increased risk of morbidity and mortality associated with secondary effects of therapy, with a 35-75 fold excess risk of developing breast cancer over the general population. The mechanism for secondary breast cancer among Hodgkin survivors is not understood. Researchers have postulated that the familial characteristics of HL could be associated with mutations found within familial cancer syndromes; however, these mutations have not been identified. This has led to the exploration of inherent polymorphisms that might impair the patient's capability to detoxify chemotherapy and/or repair DNA damage produced by irradiation. Examinations of candidate polymorphisms indicate that single nucleotide changes may have only a small effect on the development of subsequent cancers. However, multiple studies support the idea that sensitivity to irradiation and the subsequent development of breast cancer is mediated through the interaction of multiple genes or gene complexes. The objective of this case-control study design was to explore the identification of potential candidate genes and polymorphisms that may be risk factors for the development of secondary breast cancer among women who are pediatric HL survivors. Global gene expression and genotyping of women with (n=13) and without (n=36) secondary breast cancer after the treatment of pediatric HL were compared. Differences were found in global gene expression and genotyping between the cases and controls. Additionally, copy number variation in association with gene expression found a locus of interest at 15q11.2 in association with the development of secondary breast cancer.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
Problem	1
Primary Aims	2
Significance	3
Conceptual Model	5
Definition of Terms	7
Assumptions	8
Potential Limitations	9
CHAPTER 2. REVIEW OF LITERATURE	10
Hodgkin Lymphoma	10
Histological Subsets	10
Hodgkin Staging	10
Hodgkin Lymphoma Therapy	11
Hodgkin Lymphoma Therapy of Cases and Controls	11
Secondary Malignancy	13
Demographic and Treatment Characteristics of the Primary Cancer	14
Radiation	14
Age at HL Diagnosis	14
Endocrine Influences	15
Family History	17
Other Breast Cancer Risk Factors	18
Characteristics of Secondary Breast Cancer	18
Candidate Germline Mutations	18
Candidate Polymorphisms	19
Testing Methodologies	20
Microarray Technology	21
Summary of Review	22
CHAPTER 3. METHODS	23
Research Design	23
Research Design and Study Population	23
Eligibility Criteria for Case and Control Participants	23
Exclusion Criteria	23
Procedures for Accomplishment of Aims	25
Case and Control Participation	25
Enrollment of Cases and Controls	26
Blood Collection for RNA and DNA Extraction	28
RNA Extraction	29
DNA Extraction	29

Measurement of RNA: Quality Control.....	31
Processing of Extracted RNA: Hartwell Center, SJCRH.....	31
Processing of Extracted DNA: Hartwell Center, SJCRH	32
Statistical Procedures.....	32
Primary Aim One.....	33
Primary Aim Two	33
Primary Aim Three	34
Secondary Aim	34
 CHAPTER 4. RESULTS.....	 36
Introduction	36
Description of the Sample.....	36
Treatment Characteristics.....	36
Reproductive and Family Characteristics	36
Breast Cancer Characteristics Among Cases	39
Primary Aim One.....	42
Primary Aim Two.....	54
Primary Aim Three	67
Summary of Results.....	67
 CHAPTER 5. DISCUSSION.....	 80
Discussion of Findings.....	80
Demographic Data	80
Primary Aim One.....	81
Pathways of Significance	81
Folate-Methionine Pathway.....	81
t-RNA-Dependent Amino Acid Biosynthesis.....	84
Nicotinate and Nicotinamide Metabolism.....	85
DNA Polymerase	85
Biosynthesis of Secondary Metabolites	85
Global Gene Expression Summary.....	86
Primary Aim Two.....	86
Differences in Genotypes	86
Genotype Summary.....	89
Primary Aim Three	89
Copy Number Variation.....	89
Copy Number Variation Summary	92
Relation of Study Results to Conceptual Framework.....	92
Nursing Implications.....	93
Limitations	93
Recommendations.....	94
Recommendations for Research	94
Recommendations for Practice	95
Conclusion.....	95

LIST OF REFERENCES.....	96
APPENDIX A. INSTITUTIONAL REVIEW BOARD APPROVAL	103
APPENDIX B. INTRODUCTORY LETTER.....	104
APPENDIX C. INFORMED CONSENT STATEMENT.....	105
APPENDIX D. SUMMARY OF RESEARCH AND PRIVACY RIGHTS	109
APPENDIX E. DEMOGRAPHICS QUESTIONNAIRE OF HODGKIN DISEASE SURVIVORS.....	112
APPENDIX F. SINGLE NUCLEOTIDE POLYMORPHISMS OF INTEREST	114
VITA.....	118

LIST OF TABLES

Table 2.1	Ann Arbor Staging Classification.....	12
Table 3.1	Initial Control Contact for 12 Matched Cases	27
Table 4.1	Demographic Data: Current Age, Age at Time of Diagnosis and Years Since Diagnosis For Study Participants Who Did and Did Not Develop Secondary Breast Cancer.....	37
Table 4.2	Demographic Data: Age at Menses, Menopause, and Hormonal Use, According to Study Participants Who Did and Did Not Develop Secondary Breast Cancer.....	38
Table 4.3	Demographic Data: Reproductive History According to Study Participants Who Did and Did Not Develop Secondary Breast Cancer....	40
Table 4.4	Breast Cancer Characteristics	41
Table 4.5	12 Affymetrix-Defined Pathways of Significance.....	43
Table 4.6	Significant Genes Associated with Designated Pathways.....	44
Table 4.7	Genes Associated with Copy Number Variation at 15q11.2.....	71
Table 4.8	Association of First Degree Breast Cancer Family History and Deletion at 15Q11.2	78
Table 5.1	Grouping of Significant Pathways	81

LIST OF FIGURES

Figure 1.1	Conceptual Model Describing the Events Following Mediastinal Radiation in Case and Control, with the Utilization of Gene and Genotype Expression as a Predictor.....	6
Figure 3.1	Study Design.....	24
Figure 4.1	Peak p-Values for Chromosome 1	55
Figure 4.2	Peak p-Values for Chromosome 2	55
Figure 4.3	Peak p-Values for Chromosome 3	56
Figure 4.4	Peak p-Values for Chromosome 4	56
Figure 4.5	Peak p-Values for Chromosome 5	57
Figure 4.6	Peak p-Values for Chromosome 6	57
Figure 4.7	Peak p-Values for Chromosome 7	58
Figure 4.8	Peak p-Values for Chromosome 8	58
Figure 4.9	Peak p-Values for Chromosome 9	59
Figure 4.10	Peak p-Values for Chromosome 10	59
Figure 4.11	Peak p-Values for Chromosome 11	60
Figure 4.12	Peak p-Values for Chromosome 12	60
Figure 4.13	Peak p-Values for Chromosome 13	61
Figure 4.14	Peak p-Values for Chromosome 14	61
Figure 4.15	Peak p-Values for Chromosome 15	62
Figure 4.16	Peak p-Values for Chromosome 16	62
Figure 4.17	Peak p-Values for Chromosome 17	63
Figure 4.18	Peak p-Values for Chromosome 18	63
Figure 4.19	Peak p-Values for Chromosome 19	64

Figure 4.20	Peak p-Values for Chromosome 20	64
Figure 4.21	Peak p-Values for Chromosome 21	65
Figure 4.22	Peak p-Values for Chromosome 22	65
Figure 4.23	Peak p-Values for Chromosome 23	66
Figure 4.24	Prevalence of Inferred Copy Number Variation among Cases and Controls within 14q11.2.....	68
Figure 4.25	Gene Expression Correlated with Copy Number Variation at 14q11.2	69
Figure 4.26	Prevalence of Inferred Copy Number Variation among Cases and Controls within 15q11.2.....	70
Figure 4.27	Correlation of CYFIP1 Gene Expression and Copy Number Variation	72
Figure 4.28	Correlation of PAK2 Gene Expression and Copy Number Variation	73
Figure 4.29	Correlation of NIPA1 Gene Expression and Copy Number Variation	74
Figure 4.30	Correlation of NIPA2 Gene Expression and Copy Number Variation	75
Figure 4.31	Correlation of BCL8 Gene Expression and Copy Number Variation	76
Figure 4.32	Correlation of TUBGCP5 Gene Expression and Copy Number Variation	77
Figure 5.1	DNA Syntheses and Methylation.....	83

CHAPTER 1. INTRODUCTION

Problem

Overall survival of pediatric Hodgkin lymphoma (HL) is 90%; however, due to late treatment complications including subsequent malignancy, survival decreases with time [1, 2]. Women survivors of pediatric HL have an increased morbidity and mortality associated with secondary effects of therapy, most specifically those associated with radiation [3-6]. It is estimated that the relative risk of breast cancer among HL women survivors treated prior to 30 years of age is 6 to 17-fold [7-9], with the largest relative risk ranging from 60-fold to 112-fold for those treated at 16 years of age or younger [7, 8, 10]. Furthermore, these women have reached a 12.9% cumulative incidence of breast cancer by 40 years of age [3], and at 50 years of age many HL survivors treated with chest radiation have exceeded the lifetime risk of developing breast cancer as compared to the general population which is a 13.4 % lifetime cumulative incidence, one in eight women [11, 12]. The relative risk for secondary breast cancer is highest at 15 to 20 years after initial HL radiation [3, 5, 7], while an increased cumulative absolute risk of breast cancer reflects background breast cancer associated with increasing age [11]. In comparison to other high risk breast cancer groups such as BRCA1 and BRCA2 carriers, women treated for HL at a young age with radiation therapy and without alkylating agents may only have a modestly smaller risk than the high risk carriers of the BRCA mutation [11]. Therefore, women HL survivors are at a significantly high risk for the development of breast cancer and remain so throughout their lifetime.

The mechanism for early age on-set breast cancer among Hodgkin survivors is not understood; however, specific characteristics have been identified as being risk factors and include: age at the time of HL diagnosis[3, 5-7, 11, 13, 14], radiation exposure[5, 7, 11, 13-16], endocrine function[13, 14], and familial cancer history [3]. Studies show that other risk factors such as administration of combined modality therapy [7], alkylating agents [3, 13, 14, 17], pelvic radiation [13], pregnancy, and use of hormonal therapy have conflicting associations to secondary breast cancer among HL survivors.

In attempting to better understand the mechanism of secondary breast cancer, researchers have postulated that the familial characteristics of HL and high incidence of secondary breast cancer may result from mutations similar to those identified in inherited breast cancer syndromes. To date, inherited mutations such as TP53, BRCA1, BRCA2, and heterozygous ATM mutations have not been identified among HL survivors with secondary breast cancer [3, 18, 19]. While many of the associated risk factors of secondary breast cancer are well documented, there have been no association studies that have linked an identified genetic risk factor marker.

Failure to identify commonly inherited mutations or markers has led to the exploration of inherent polymorphisms that might impair the patient's capability in detoxification of chemotherapy and/or repair of deoxyribonucleic acid (DNA) damage produced by radiation [20, 21]. Examination of these candidate polymorphisms has

found that single nucleotide changes may have only a small effect on the development of subsequent cancers and the strongest effect on secondary cancer development is perhaps a gene-gene interaction. This is evident in studies that support the idea that sensitivity to radiation and the subsequent development of breast cancer is mediated through the interaction of multiple genes or gene complexes [18, 19, 21].

Thus far, the association of candidate genes and polymorphisms and the development of secondary breast cancer have not been found. To my knowledge, there has not been a whole genome association study examining the global gene expression and single nucleotide polymorphisms associated with HL and secondary breast cancer. To explore the genetic contribution to this complex disease, the purpose of this study was to identify gene expression patterns and candidate polymorphisms that may be risk factors for the development of secondary breast cancer. Study methods focused on identifying differential gene expressions between two groups of women who are pediatric HL survivors; those who developed secondary breast cancer (case) and those who have not (control) developed secondary breast cancer within 20 years after HL diagnosis. Individual global gene expression was characterized and associated with their expressed polymorphism (mutation). The gene expression was measured by microarray analysis and an association study correlated the genotypes at 500,000 single nucleotide polymorphism (SNP) loci with the gene expression levels. The study also examined gene and polymorphism expression differences between the case and controls and identified global gene expression and the associated genotype copy number variation among the cases and controls. Association data will provide candidate genes, polymorphisms and a region of copy number variation for consideration in the etiology of secondary breast cancer after HL therapy. Furthermore, identification of genes, polymorphism, and copy number variations associated with the development of secondary cancers would facilitate identification of at risk patients and permit modification of therapy and heightened surveillance that may reduce cancer-related morbidity and mortality.

Primary Aims

The primary aims were to:

1. Assess whether global gene expression profiles differ between two groups of radiated pediatric female HL survivors with (cases) and without (controls) the development of secondary breast cancer.
 - 1.1. Identify the gene expression profile for cases
 - 1.2. Identify the gene expression profile for controls
 - 1.3. Identify the gene expression profile similarities and differences between cases and controls
2. Assess whether genotype profiles differ between two groups of radiated pediatric female HL survivors with (cases) and without (controls) the development of secondary breast cancer.

- 2.1. Identify the genotype profile for cases
 - 2.2. Identify the genotype profile for controls
 - 2.3. Identify the genotype profile similarities and differences between cases and controls
3. Identify potential candidate genes through identifying copy number variations that associate with the risk of secondary breast cancer.

The secondary aim was to explore the associations among lifestyle characteristics (hormone use, pregnancy, lactation), family history and the occurrence of breast cancer.

Significance

According to the American Cancer Society, over 7,000 new cases of HL were diagnosed in 2005 with 10-15% of these cases occurring in children 16 years and younger [22]. With the combination of chemotherapy and radiation, the 5-year survival for those with low stage disease is 85-95% and 70-90% survival for those with advanced disease [12]. The intensity of therapy is stage-related.

With improved therapy it is now estimated that among the general population 1 of every 300 young adults < 45 years of age is a survivor of childhood cancer [22]. Furthermore, cancer survivors of all ages comprise 3.5% of the United States population. Although these individuals have survived cancer, they are at greater risk for subsequent cancers. Approximately 16% of the annual cancer incidence occurs as a second cancer in a cancer survivor [23]. One of the most devastating event for any cancer survivor is the occurrence of a second cancer, but particularly devastating is the second cancer for a young adult who is starting a career and family. As cancer survival rates increase so will the number of survivors who will experience a secondary cancer.

Recently, there has been an increased interest and concern regarding the health status and late effects associated with childhood cancer care, most specifically the late effect of secondary cancers. Studies have found one of the highest occurrences of second cancers is among women who are HL survivors [3, 4, 6, 7]; therefore, it is of importance that factors contributing to secondary cancers in this group be identified.

While environmental risk factors such as radiation have been identified, inherent risk factors are just beginning to be studied. Identification of secondary breast cancer inherent risk factors could provide an at risk patient profile at the time of initial diagnosis. This profile would be important when considering therapeutic agents that may be added or deleted in the initial treatment of HL resulting in a reduction of late effects. Lastly, one could foresee the high risk profile individual being counseled, screened, and offered preventive therapies such as those of known breast cancer risk associated with BRCA1 and BRCA2 gene mutation.

The primary focus of this study was to begin the development of an inherent genetic risk patient profile. Thus far the candidate gene approach has not found evidence of known inherited mutations among those with secondary breast cancer [3, 18, 19]. Therefore, this study explored the comparison between an individual's genotype in consideration of treatment variables that may contribute to the increased incidence of secondary breast cancer among HL survivors compared to non-secondary breast cancer HL survivors. This study took advantage of rapidly evolving biotechnology in gene profiling and genotyping using microarray technology. The global gene expression and genotypes of HL survivors with breast cancer was explored to identify potential candidate polymorphisms associated with genotype allele frequency and the possible contribution of the genotype to the development of secondary breast cancer. Identification of potential polymorphisms will be valuable in the development of future research projects in replication and validation within larger samples of HL survivors with secondary breast cancer.

Gene profiling and genotyping technology is quickly advancing and showing significant progress in the identification of prognostic indicators for complex disorders. A number of studies have identified prognostic and predictive 'gene' signatures as prognostic indicators in the treatment of breast cancer [24]. Most of the early work with microarray analysis and primary breast cancer has been done with tumor samples taken at the time of breast surgery. These analyses have enabled researchers to analyze breast tissue and obtain a better understanding of the molecular characteristics and clinical behavior of breast cancer. This has been most helpful in the analysis of gene amplification and coamplification in prognosis and treatment of breast cancer [25]. While tissue study is beneficial in determining prognosis and treatment response, molecular expression from a tumor sample presents difficulty in determining the causality of the tumor due to both primary and secondary chromosomal changes. Primary changes are present in all tumor cells, while secondary changes are associated with revascularization or the development of metastatic disease [26]. Examination of breast tissue will give expression at the site of the disease and reflect the characteristics of the individual tumor, but will not give a global picture of the gene interaction that may have predisposed an individual to cancer. With the hypothesis that the second cancer is the result of an inherent gene-gene/environment interaction, breast tissue would not reflect inherent or germline expression. Therefore, this study will explore the use of peripheral blood as a potential predictor of germline genetic risk for the occurrence of secondary breast cancer.

Lastly this study is significant for its interdisciplinary approach. Nurses can play a pivotal role in the era of genomic research through their skill of design, interdisciplinary interaction in completing the methodology and analysis, and the nurses' skill of research dissemination with translation into practice. The nurse-patient rapport provides an excellent navigational environment for the nurse to describe genomic screening, provide a clear translation of the findings, as well as the implications for treatment decision making and follow-up. It is also imperative that the patient have an update family history with particular emphasis on family history of cancer and most specifically breast cancer.

Conceptual Model

This study evaluated the similarities and differences in genotype allele frequency and the associated global gene expression between individuals treated for pediatric HL with and without the development of secondary breast cancer. The conceptual model (see Figure 1.1) depicts an individual with diagnosed HL involving the mediastinum. The standard HL therapy is combination chemotherapy and radiation at the site of disease. With the presence of mediastinal HL, most individuals receive radiation therapy to the mediastinum disease. At the initiation of treatment, each individual has inherent characteristics including familial inherited characteristics and individual genetic differences within their individual nucleotide sequence known as polymorphisms. In addition, each individual has specific treatment related risk factors. These risk factors include: age at HL diagnosis [3, 5-7, 11, 13, 14], dose of mediastinal radiation [5, 7, 11, 13-16], administration of alkylating agents [3, 13, 14, 17] and pelvic radiation [13]. During the therapeutic administration of mediastinal radiation there is an interaction between the inherent and treatment characteristics (gene-environment interaction). Both the case and control have ablation of Hodgkin tumor, but the case will have retention and/or disrepair of mutated deoxyribonucleic acid (DNA) within the non-diseased breast tissue. The control group will have repair and/or apoptosis of mutated DNA in the non-disease breast tissue. Overtime, both groups will undergo additional genetic insults from exposures within the individuals' environment. For each case, penetrance of secondary breast cancer is dependent upon persistent and additional somatic mutations [27]. As a result of the retained mutated DNA and successive mutations, the secondary cancer will occur approximately 15-20 years after the initial DNA mutation initiated by radiation [3, 7, 8].

Determining the genotype differences among the cases and controls was accomplished through the technology of DNA and expression microarray. The human genome is made of 3 billion base pairs with single nucleotide polymorphisms (SNPs) noted between individuals at every 1000 base pairs. Furthermore, it is these polymorphisms that give "biochemical individuality" among the human species [28]. With the interaction of one's biochemical pathways (gene-gene interaction), regardless of one's state of health, there is a genetic predetermined chemical individuality that responds to environmental and pharmacological influences (gene-environment) [28].

Microarray technology has predominantly been utilized in the analysis of expressed genes within and between biologically matched subjects; however, with human genome mapping they can now be utilized for study of genetic variations among individuals and referred to as SNP analysis. SNP analysis is useful in linkage analysis for identification of disease markers, loss of heterozygosity (LOH), uniparental disomy, for analysis of tumor suppressor genes, and association studies to link SNPs to overly or underly expressed genes [29]. Therefore, this proposal utilized whole genome analysis of gene expression and genotyping at polymorphic sites between cases and controls. The gene expression was analyzed from the extracted RNA, while genotyping (SNP) data were analyzed from the extracted DNA.

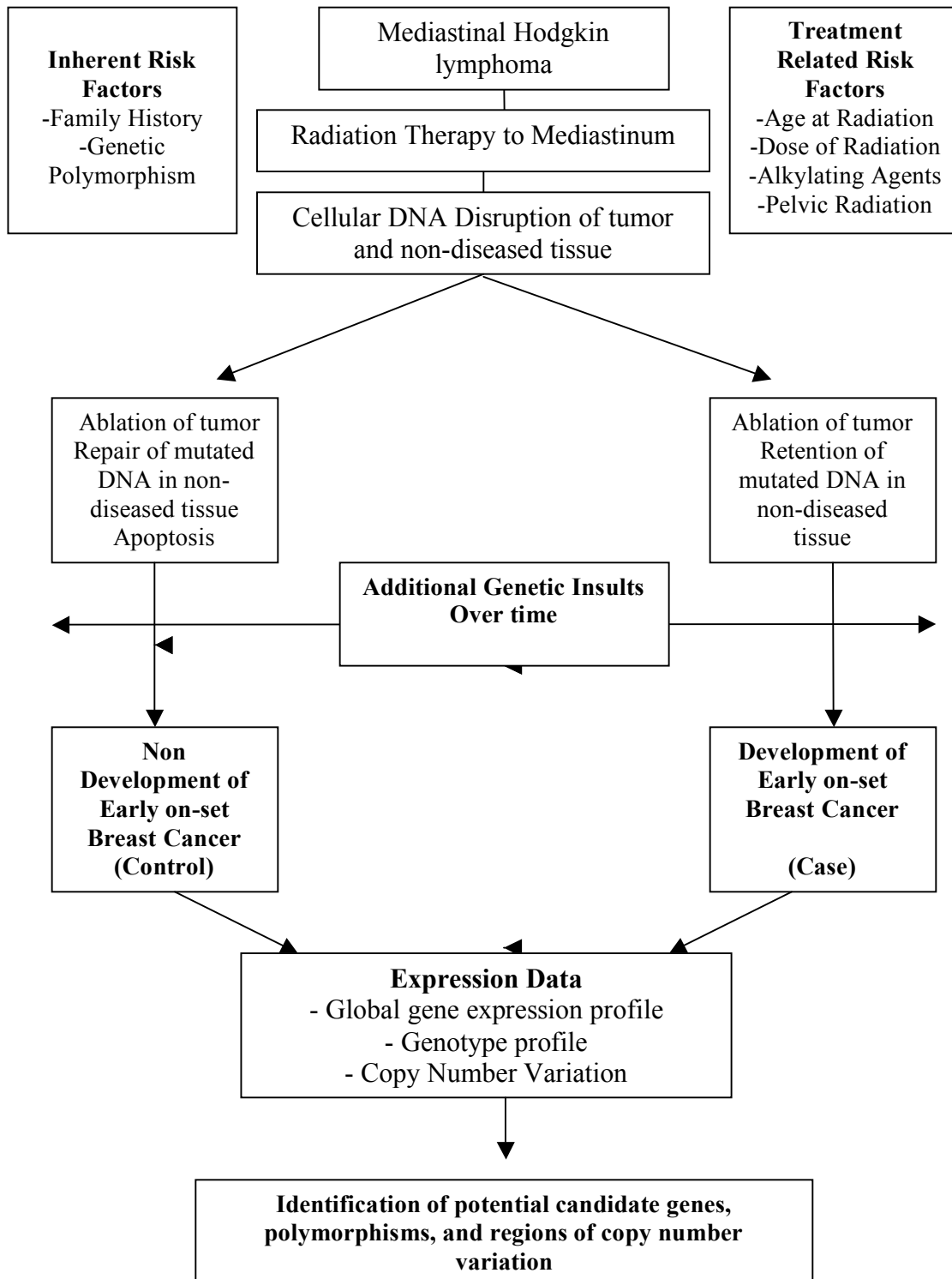


Figure 1.1 Conceptual Model Describing the Events Following Mediastinal Radiation in Case and Control, with the Utilization of Gene and Genotype Expression as a Predictor.

The goal of the study was to acquire an enhanced understanding of the biological pathway leading to the development of secondary breast cancer, and identification of potential biomarkers that may identify women at risk. It is hypothesized that genetic differences between the case and control are responsible for a differential response to initial DNA damage. Accumulation of DNA mutations at different rates would lead to early onset of secondary breast cancer. The gene expression and genotype profiles were compared between cases and controls as well as the copy number variation. This analysis identified potential candidate genes and a site of deletion for consideration in the etiology of secondary breast cancer. Once women are identified as being at risk, treatment may be modified for prevention of therapy related morbidity/mortality thereby improving quality of life.

Definition of Terms

- **Hodgkin lymphoma (HL):** Participants within this study, both cases and controls, were diagnosed with HL prior to 25 years of age. The diagnosis of HL was based upon pathological examination of lymph node tissue and documented in the medical record.
- **Case (Secondary Breast Cancer):** Participants classified as cases were surviving women who were treated for pediatric HL with mediastinal radiation and subsequently developed breast cancer prior to 50 years of age. The diagnosis of secondary breast cancer is based on pathology examination and documented in the medical record.
- **Control (Non-Secondary Breast Cancer):** Participants classified as controls were surviving women who were treated for pediatric HL with mediastinal radiation and have not developed any subsequent cancer. These women will have a negative mammogram within one year of the study and a negative history for proliferative breast disease. This data was obtained through participant reporting.
- **Copy Number Variation:** Copy number variation gives the number of DNA segments within the genome. It was thought that genes were present in two copies but is now known that there is variability and DNA segments may be lost or gained. This gives an insight into the variability among individuals and may contribute to disease.
- **Genotype:** Genotype is an individual's set of alleles that make up one's genetic constitution, either collectively or at a single locus. This study explored differences in genotypes using single nucleotide polymorphisms.
- **Phenotype:** Phenotype is the observable characteristics of an individual, resulting from the interaction of the genotype and environment.
- **Single Nucleotide Polymorphism (SNP):** Single nucleotide polymorphisms are found throughout an individual's genotype and may affect phenotype. A SNP is a DNA marker and designates differences in the DNA sequence of homologous chromosomes. For the DNA differences to be termed a SNP, they must be present in more than 1% of the population [30]. The genotype expression, the presence of single nucleotide polymorphisms (SNP), was measured with the Affymetric GeneChip® Human Mapping 500K Set.

- **Gene Expression:** Gene expression is the process of transcription of DNA into messenger ribonucleic acid (mRNA) and translation of the mRNA into protein. This study will assess an individuals' gene expression and the association of this expression to the genotype. The differences in global gene expression between cases and controls will be measured with the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array.
- **Microarray:** Microarray analysis allows genomic measurement of an individual's gene expression under a variety of genetic and environmental conditions. The mRNA is reverse transcribed to complimentary DNA (cDNA) and labeled with fluorescent dyes and hybridized to the microarray glass slide. The glass slide has oligonucleotides arranged in columns and rows which are constructed for a specific gene. Lasers are then used to generate an image from the sample. The data are expressed in a matrix with columns indicating samples and rows indicating genes [31]. The Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array will analyze gene expression. Microarray technology also has the application for detection of polymorphisms within an individuals' genome. The DNA is labeled with fluorescent dyes and hybridized to the microarray. The array has oligonucleotides that are associated with 500,000 known SNPs within the genome. This study will use the Affymetrix GeneChip® Human Mapping 500K Set for SNP analysis.
- **Gene-Environment Interaction:** A gene to environment interaction occurs when the environmental effects on phenotype differ according to the individual genotype.
- **Gene-Gene Interaction:** The regulated expression of genes is dependent upon complex interrelationships among genes within the genome. Changes within this interrelationship among genes can significantly affect biological pathways [28].

Assumptions

1. Women diagnosed with pediatric HL are a vulnerable host for recurrent malignancies.
2. HL survivors with and without breast cancer have a genetic predisposition that responds differently to radiation exposure.
3. The gene expression profile between the cases and controls may not differ due to cumulative absolute risk of breast cancer among controls due to background breast cancer associated with increasing age.
4. The identified gene pathways will direct future research in the study of candidate genes that had not been previously considered.
5. Ultimately, identification of polymorphisms and copy number variation could be integral in the development of a predictive model for those at risk of secondary breast cancer.

Potential Limitations

1. Much research has focused on the genetic study of primary breast cancer; however, these data may not be generalizable to other breast cancer patients. Most breast cancer research has focused on the genetic characteristics of the tumor, while this study will focus on the inherent individual characteristics that predispose one to the development of breast cancer.
2. Secondary breast cancer is most often diagnosed 15-20 years after the initial HL diagnosis; however, this will vary among individuals [3, 7, 8]. All controls will have undergone the passage of 15 to 20 years since diagnosis and will be matched with the cases on time since diagnosis; however, a control could become a case during the course of the study.
3. Blood samples for gene expression were collected at outlying physician's offices and mailed to a designated lab for RNA extraction. If collection, storage, and mailing were not completed as instructed, there may have been degradation of the RNA in transport. If RNA degradation occurs, there may be loss of transcripts and gene expression. However, each sample will undergo a quality control analysis before processing and another RNA sample will be submitted if needed.
4. Methods are available to compute power and sample size for most gene expression microarray studies [32]. However, power and sample size calculations were not used to design this study due to the limited number of cases. Thus, the statistical power to find meaningful associations may be quite low.

CHAPTER 2. REVIEW OF LITERATURE

Hodgkin Lymphoma

In 1832, Thomas Hodgkin published a paper entitled, “On Some Morbid Appearances of the Absorbent Glands and Spleen” describing lymph node and spleen enlargement in what came to be known as Hodgkin lymphoma [33]. At the turn of the century, diagnosis was dependent upon the microscopic morphology of tissue and Hodgkin tumors were recognized to have abnormally giant cells. These giant cells were first described by Sternberg [34] and Reed [35] who were credited with describing HL histopathology, hence the naming of the Reed Sternberg cells found within HL tumor.

Until recently the malignant cell of HL origin has remained unknown. New technologies such as polymerase chain reaction (PCR) and micro laser dissection have assisted in determining clonal immunoglobulin rearrangements and the origin of malignancy as B-lymphoid cells [36, 37]. The World Health Organization (WHO) has now renamed Hodgkin disease to Hodgkin lymphoma reflecting an understanding of its B-cell origin [38].

It is estimated that only a small portion (0.1% to 10%) of the HL tumor is malignant [38]. HL tumors are primarily comprised of inflammatory cells including lymphocytes, neutrophils, eosinophils, histiocytes, plasma cells and fibrosis. The malignant cells found in HL are the Reed-Sternberg cell and mononuclear/multinucleated variants. The Reed-Sternberg cell is a large binuclei or binuclear lobe cell and within each of the nuclear lobes are large eosinophilic nucleoli larger than neighboring lymphocytes. HL cells may also be highly pleomorphic creating difficulty with diagnosis and a differential to include high-grade sarcomas and diffuse large cell lymphomas with anaplastic features [39].

Histological Subsets

The WHO classification recognizes two major subsets of HL, classical Hodgkin and nodular lymphocyte predominant Hodgkin [38]. This classification is based on the tumor morphology and immunophenotype. Classical Hodgkin is categorized into four subsets and includes: nodular sclerosis subtype, mixed cellularity subtype, lymphocyte-rich subtype, and lymphocyte depleted subtype. All subtypes of HL are equally responsive to treatment [39].

Hodgkin Staging

In 1971, the Ann Arbor staging for HL was developed, this staging was based on the knowledge that HL progressed along lymph node chains [40]. Substage classification is also determined at diagnosis according to clinical symptoms at presentation and includes A, B, and E substage. At presentation, the history includes notation of

constitutional symptoms which include fever $>38^{\circ}\text{C}$ for 3 consecutive days, drenching night sweats, and weight loss of $\geq 10\%$ of body weight over the previous 6 months. Upon presentation, a patient without these symptoms is classified as an A substage; however, if these symptoms are present they are classified as B substage. Patients with E substage have extralymphatic extension of disease directly from nodal disease. The Ann Arbor Staging Classification of HL is described in Table 2.1.

Hodgkin Lymphoma Therapy

Radiotherapy was the first effective treatment for HL, providing high doses of radiation to large fields. Then in 1964, a four-drug regime (MOPP) was introduced as the first effective systemic therapy for HL. The MOPP regime includes mechlorethamine (nitrogen mustard), Oncovin (vincristine), procarbazine, and prednisone. This regimen resulted in a prolonged disease free survival in more than 50% of adult and pediatric patients. However, the treatment was found to have significant late effects including acute myeloid leukemia (AML) and infertility. To reduce the risk of AML, another alkylating agent, cyclophosphamide (COPP) was substituted for mechlorethamine. In 1970, the ABVD (Adriamycin (Doxorubicin), bleomycin, vinblastine, and dacarbazine) regimen was introduced and did not have the associated late effects of AML and infertility. While this regimen has led to superior outcomes in adults, the regimen has been of concern in pediatrics due to the risk of cardiopulmonary toxicity. Through the 1980's, this combined approach of multi-agent chemotherapy and extended volume radiation delivered at standard dose of 35-44 Gray (Gy) was the treatment of choice [39]. Overtime, significant late effects were noted in HL survivors as a result of the radiation. These late effects resulted in investigating the effectiveness of combined multi-agent chemotherapy with low dose involved-field radiation delivered at the dose of 15-25.5 Gy [41, 42].

In the 1990's, investigators found that multi-agent chemotherapy with low dose radiation did not compromise disease-free survival. Children and adolescents are now treated for HL with risk-adapted, combined-modality therapy using multiagent chemotherapy and low-dose involved field radiation [39]. For children with low-stage, nonbulky disease, chemotherapy alone is being investigated [39]. The objective of risk adapted therapy is to provide disease-free survival with minimization of treatment-related sequelae including musculoskeletal deformity, cardiopulmonary, and second malignancy.

Hodgkin Lymphoma Therapy of Cases and Controls

The cases and controls for this study were treated at St. Jude Children's Research Hospital (SJCRH) between 1970 and 1991, according to institutional protocol. These protocols included: HOD 68, HOD 72, HOD 80, and HOD 90. During the years 1969-1972, HL patients were treated according to the HOD 68 protocol. This protocol delivered combined chemotherapy agents; cyclophosphamide, vincristine +/- procarbazine and prednisone plus radiation therapy at doses ranging from 35-44 Gy

Table 2.1 Ann Arbor Staging Classification.

Stage	Definition
I (A/B)	Involvement of a single lymph node region (I) or of a single extralymphatic organ or site (IE)
II (A/B)	Involvement of two or more lymph node regions on the same side of the diaphragm (II) or localized involvement of an extralymphatic organ or site and one or more lymph node regions on the same side of the diaphragm (IIE)
III (A/B)	Involvement of lymph node regions on both sides of the diaphragm (III), which may be accompanied by involvement of the spleen (IIIS) or by localized involvement of an extralymphatic organ or site (IIIE) or both (IIISE)
IV (A/B)	Diffuse or disseminated involvement of one or more extralymphatic organs or tissues with or without associated lymph node involvement

delivered to extended volume fields. The HOD 72 protocol was opened in 1972-1980 and compared standard dose radiation therapy at doses 35-44 Gy alone to radiation therapy at doses 35-44 Gy plus chemotherapy agents cyclophosphamide, vincristine, procarbazine, and prednisone (COPP). The HOD 80 protocol was opened in 1980-1984 and combined chemotherapy COPP with lower dose radiation at 20 Gy extended-volume. In 1984-1990, the HOD 80A protocol was open and continued utilization of lower dose radiation at 20 Gy; however, the radiation was delivered to only the field of disease involvement. Alternating cycles of chemotherapy were administered and included COP(P) and ABVD.

Prednisone was only administered to patients with “B” symptoms. HOD 90 was opened in 1990 and was the first protocol to classify patients according to risk. The low risk arm therapy consisted of 4 cycles of vinblastine, adriamycin, methotrexate, and prednisone (VAMP) and low dose involved field radiation delivered between 15-25.5 Gy. The radiation dose was determined according to disease response after two cycles of VAMP. High risk patients received 6 cycles of chemotherapy which included vincristine, etoposide, prednisone, and adriamycin (VEPA) followed by 25.5Gy radiation therapy. Radiation therapy was not initiated until after the second cycle of chemotherapy except for patients with extensive disease who began radiation after the completion of all chemotherapy. Patients that required more than one treatment field of radiation received the later courses of radiation after the fourth or sixth course of chemotherapy.

Secondary Malignancy

Children and adolescents treated for HL have a higher incidence of long-term late effects secondary to treatment when compared to other childhood cancer survivors [1, 11, 41]. One of the devastating late effect for any cancer survivor is the occurrence of a second cancer. A large HL cohort followed for 27.8 years found the estimated cumulative incidence of any second cancer was 10.6% at 20 years and increasing to 26.3% at 30 years. The estimated cumulative incidence of solid tumors was 7.3% at 10 years and increasing to 23.5% at 30 years, with breast cancer being the most common solid tumor. The most surprising finding was the incidence of a third cancer with the estimated cumulative incidence of 21% at 10 years after the second [10].

Breast cancer is the most common cancer among women and is the second leading cause of death from cancer [1]. Women treated for HL with mediastinal radiation at 30 years of age or younger have the highest incidence of breast cancer, outside familial syndromes [11]. The risk is greatest at 15 to 20 years after initial HL diagnosis, which is more than 20 years before the median age (61years) of breast cancer among the general population [3, 7, 20]. The mechanism for secondary breast cancer following mediastinal radiation is not well established, but specific risk factors have been identified. Specific demographic and treatment characteristics identified as risk factors contributing to secondary breast cancer include: dose of mediastinal radiation, age at the time of HL diagnosis and therapy, endocrine function, and familial cancer history including inherent genetic variability [3-6, 8, 12, 13].

Researchers have developed predictive models of cumulative risk of breast cancer among women treated for HL [11] based on treatment, age at therapy and length of follow-up as well as family history. It could be argued that pediatric HL patients treated on identical treatment regimens are matched on the variables of treatment, age, and follow-up; therefore, all are seen as high risk for secondary breast cancer. Therefore, the genetic variability at the time of therapy may play the greatest role in determining those of increased sensitivity to the prescribed therapy, most specifically radiation. A recent study found an increased relative risk of breast cancer after HL with a positive first degree family history of cancer [43]. While the family history is a significant predictor of future health status, family history in the pediatric patient should also consider the health status of second degree relatives. Many of the first degree relatives of pediatric HL patients may not have developed a cancer due to the background of time, while predictors may be with the secondary degree family members. For the pediatric HL survivor one should consider all the associated risk factors; radiation, chemotherapy regimens, age at therapy, and family history; however, there appears to be no straightforward predictor for the development of secondary breast cancer [44].

Demographic and Treatment Characteristics of the Primary Cancer

Radiation

The risk factor that is most strongly associated with the development of secondary breast cancer is mediastinal radiation including the radiation dose received as part of the treatment for HL. The radiation dose of ≥ 40 Gy has been associated with the highest risk of secondary cancer with a 23.3 fold increase in comparison to 20-39 Gy which is associated with a 5.9 fold increase [15]. While the lower dose of radiation is thought to reduce risk of secondary cancer, women who received radiation therapy alone at doses of 4 Gy to the breast continued to have a 3.2 fold increase in development of breast cancer [13]. Another study found that the risk of secondary breast cancer was increased in HL patients compared to the general population regardless of the radiation dose [5]. Although current treatment protocols use combined modality therapy, the risk of lower dose radiation < 20 Gy has not been established. It is believed that the reduction in volume of exposed breast tissue and radiation dose reduction to 20-30 Gy will change the long-term risk profile of patients cured of HL [45]. While the association has been made with radiation and the occurrence of a secondary breast cancer, there is clearly no radiation dose that is without the risk of breast cancer. All pediatric HL survivors treated with chest radiation appear to be at an increased risk regardless of the radiation dose.

Age at HL Diagnosis

Along with radiation, age at the time of HL diagnosis has been implicated as a risk factor for secondary breast cancer. As stated earlier, the greatest risk of secondary breast cancer has been found in women treated for HL prior to the age of 30 years [7, 13].

Women radiated prior to 30 years of age appear to have an excess risk 15-20 years after therapy and beyond 20 years, with a 2.3-3.7% excess risk per person per year [8].

Young age at the time of radiation and the occurrence of breast cancer parallels the well established concept of increased sensitivity of young breast tissue to ionizing radiation [11]. An earlier age of risk has been proposed by Travis et al. [13] who determined the relative risk (RR = ratio observed/expected) of radiation associated breast cancer for subgroups defined by age and found the risk to be greatest in women who received radiation therapy at the time of puberty, 13-17 years. Another study explored the risk factor of age using the multiple Poisson regression of standard incidence ratios. The regression model analysis found the risk of secondary breast cancer to be greatest in girls treated for HD at 5-9 years of age [6]. To further explore the correlation of age at HL diagnosis and secondary breast cancer, another study used two statistical approaches, the Cox regression and Poisson regression of standard incidence ratio. Each method had similar results and the study concluded that there is little statistical evidence for an association between age at diagnosis of HL and the development of secondary breast cancer [46]. As noted, the studies that have examined the association of secondary breast cancer and age at HL treatment have found varying ages of association. It is known that women undergo an increase of breast tissue proliferation during the early menarche years and first pregnancy [47]. Therefore, an age of radiation therapy prior to 30 years is perhaps more inclusive of those at risk of secondary breast cancer and would include the ages of breast vulnerability. While the relative risk of breast cancer is highest 15-20 years after radiation therapy, this relative risk will decrease with time. However, due to the background risk of breast cancer with increasing age, HL survivors have a decreased relative risk and an increased absolute risk and cumulative incidence of breast cancer with age [7].

Endocrine Influences

Ovulatory function and most specifically the greater number of lifetime ovulatory cycles has been well established as a risk factor associated with breast cancer. This lifetime exposure to sex hormones is determined by age at menarche, age at first full pregnancy, the number of pregnancies, and age at menopause. Early menarche is less than 12 years of age for initiation of menses and is thought to increase breast cancer risk secondary to prolonged breast exposure to estradiol during adolescence [48]. In relation, the delay of menopause maximized the number of ovulatory cycles and may therefore increase breast cancer risk. Many women with significant risk factors for breast cancer are advised to have ovariectomy prior to 45 years to induce menopause and thus decreasing breast cancer risk by 40% [49].

After radiation, the risk of secondary breast cancer increases if the women has a larger number of premenopausal years due to prolonged ovulation and hormonal influences [14]. Women treated for HL have been found to experience early menopause secondary to ovarian irradiation and alkylating chemotherapy which results in ovarian dysfunction, reduced hormonal influence, and a theorized reduction in breast cancer risk

[11, 13, 14]. Reaching menopause, as early as 31 years of age has been associated with the strongest reduction in breast cancer and remains beneficial if achieved by 36 years in comparison to those remaining premenopausal or entering menopause after 45 years of age [14]. Additionally, the relative risk is reduced if there are less than 15 years of ovulation after HL therapy [14]. After HL therapy, ovarian failure may occur at a radiation dose of 5 Gy or more to the ovaries, and ovarian failure is greatest when the HL treatment includes both pelvic radiation and alkylating agents [13, 14]. According to Travis et al. [11], women treated with radiation and alkylating agents had a 1.4 fold increased risk of secondary breast cancer in comparison to a 3.2 fold increase in women treated with radiation alone; therefore, the addition of alkylating agents further decreased the ovarian function and decreased the breast cancer risk. Other data suggest that young women treated with mediastinal radiation without alkylating agents have increased risk, comparable or slightly smaller than those with BRCA mutations who have a 65-80% risk [11]. Another study found a protective effect of pelvic radiation, but did not find a protective effect of alkylating agents even at higher doses [3]. Although the use of alkylating agents has been documented as being protective, the timing of the alkylating exposure may effect the chemotherapies induction of premature menopause [50]. Furthermore, Travis et al found the trend of protection was not statistically significant in HL patients when treated at 21 years of age or younger [13]. While several studies have cited the administration of alkylating agents and pelvic radiation to be breast cancer protective, pelvic radiation may be the most protective among pediatric HL patients. For the patient less than 21 years of age the administration of alkylating agents without pelvic radiation may offer little benefit of early menopause. To determine the effect of chemotherapy and pelvic radiation, dosing of alkylating agents and chemotherapy was determined for each case and control within the study.

The use of exogenous hormones is strongly linked to an increased risk of breast cancer among postmenopausal women within the general population [51]. The link of exogenous hormones and breast cancer has also been made in women using oral contraceptives. A large meta analysis found that risk was strongest while taking oral contraceptive, with a persistent decreasing risk up to 10 years after hormone cessation. Additionally, this analysis found initiation of oral contraceptive prior to 20 years of age increases risk as opposed to initiation at an older age [52]. Studies have examined the association of hormone use after HL therapy and have been unable to clarify the risk of early on-set breast cancer, hormone therapy and HL [14, 53]. The question remains if use of hormone therapy diminishes the beneficial effect of premature menopause and should all HL survivors discontinue hormone therapy by 31 years of age, when menopause is most beneficial.

In addition to chemotherapy and radiation, the endocrine function of ovulation is decreased with pregnancy and is associated with a decreased breast cancer risk, especially with early age at first pregnancy. A study of Japanese bomb survivors found that the dose of radiation and nulliparity act multiplicatively in causing breast cancer with parous women giving birth at a younger age demonstrating lowest risk [54]. Although pregnancy decreases the number of ovulatory cycles and is protective, the exposure of pregnancy hormones on breast tissue results in an increased risk for primary and

secondary breast cancer 3 to 7 years after childbirth in women previously treated for HL [53]. Furthermore, Hill et al [53] found that women with a history of HL had an increased relative risk to other parous women if they delivered within 60 months after HL therapy. This risk is thought to be related to increased production of prolactin and lactation which promote mammary carcinogenesis after radiation. To determine the hormonal influence among the cases and controls, treatment variables of chemotherapy and radiation were collected, as well as age of menarche, menopause, history of pregnancy and use of hormonal therapy.

Family History

Hodgkin lymphoma occurs within families, and siblings of HL patients have an increased risk of developing the same cancer [3]. The family history of women who developed a second cancer after HL has been reviewed to determine if the family history is significant for the development of secondary breast cancer. It has been hypothesized that HL survivors with a family history of breast cancer have a higher risk of developing secondary breast cancer and suggest the presence of cancer predisposing genetic risk factors. In a case-control study of HL survivors treated prior to 30 years, a positive family history of breast or ovarian cancer in a first or second degree family resulted in a 2.5 fold increased breast cancer risk [53]. Furthermore, it has been suggested that a family history of breast or ovarian cancer influences one's response to radiation therapy leading to the development of breast cancer. Hill et al. [53] found that the combined effect of a breast or ovarian family history and increasing radiation dose does not exceed that of women without this history and may be lower. These findings may be more relevant to the age of radiation exposure to the young breast than the cumulative factor of family history and radiation dose. There may also be ascertainment bias in obtaining the breast cancer history in HL survivors with secondary breast cancer in comparison to controls.

In documentation of family history, several studies find that family cancer history is more significant [18, 43, 53] than breast cancer history alone. A study by Nichols et al [18] identified a cohort of Hodgkin survivors and found 21% (11 of 52 patients) with a positive family history of cancer to have developed a secondary breast cancer. Family history in this study included the proband and two other relatives with cancer over 2 generations in the paternal or maternal lineage. This cohort was compared to another cohort of Hodgkin survivors without a secondary cancer and found to have only 4% (3 of 68 patients) with a positive family history of cancer. The study concluded that a subset of HL survivors with a positive cancer family history will have a higher incidence of secondary tumors than those without a cancer family history [18]. A more recent study examined the risk of a secondary cancer among lymphoma patients with a family history of cancer [43] and found an increased relative risk of breast cancer among HL survivors with a positive family history of cancer. It was suggested that this family history presents risk associated with DNA disrepair, shared environment, or interaction. To determine the prevalence of breast cancer and familial cancers among the study cases and controls, a two generation family history of both maternal and paternal lineage was obtained.

Other Breast Cancer Risk Factors

The consumption of alcohol has been associated with breast cancer risk in both pre- and postmenopausal women [55]. There are several mechanisms through which alcohol is thought to increase breast cancer risk. Alcohol may act through its first metabolite, acetaldehyde, leading to procarcinogen activation [56] or increase estrogen levels in both pre- and postmenopausal women [57, 58]. Alcohol also affects the synthesis of nutritional supplements including folate, pyridoxal phosphate (methyl group synthesis and transfer), vitamins B12, D, A, E, retinoids, zinc, and selenium. These nutrients are essential in amino acid metabolism and response to oxidative stress and DNA repair [59] which when depleted enhance carcinogenesis [56]. The use of alcohol was not collected on the cases and controls due to the concern for the lack of recall prior to the onset of breast cancer and ascertainment bias in the cases. Other contributing breast cancer risk factors include dietary fat, physical activity, body mass index (BMI).

Characteristics of Secondary Breast Cancer

Women at greatest risk for secondary breast cancer are those who receive treatment for HL prior to 30 years of age [7-9]. The onset of secondary breast cancer after radiation is 15-20 years, occurring 20 years prior to the median age of onset among the general population [3, 5, 7]. The characteristics of the disease and axillary nodal status are important in determining the long term prognosis. Women who develop secondary breast cancer may have histories of atypical breast hyperplasia or high mammographic density which confers a 3-5 fold excess risk of breast cancer [53] and a ninefold increase if the women also has a family history of breast cancer [49]. In addition, hormone therapy users within the general population are more than twice as likely to have high risk breast density on mammography in comparison to non-users [60]. Therefore, history of atypical breast disease or high mammographic breast density should indicate the need for diligent screening and counseling. In addition to these clinical findings, the breast cancer histological features, stage, and hormone receptors of HL survivors are similar to those in the general population [3]. Furthermore, these women commonly present with tumor of the medial breast and ductal carcinoma in situ. Lastly, secondary breast cancer presents as bilateral disease in 10-30% of the breast cancer cases, while bilateral presentation is only seen in 4-6% of the sporadic breast cancer cases [3].

Candidate Germline Mutations

Common germline mutations associated with familial breast cancer syndromes such as ATM, TP53, BRCA1 and BRCA2 became candidate genes for explanation and examination within the HL survivors with secondary breast cancer. One of the first germline mutations to be examined was the heterozygous inactivating mutation in ATM, the gene defective in ataxia-telangiectasia and associated with defective DNA repair mechanisms and the development of cancer. The study used a protein truncation assay with 52 patients who had a second cancer after treatment for HL. Mutation of the ATM

gene was not found to occur in excess frequency among these survivors with second cancers [18]. Variants within the ATM gene have also been considered. Thirty-seven women with breast cancer after HL disease were matched for age at diagnosis and year of diagnosis with 27 women after HL without breast cancer; heterozygous protein-truncating or missense mutations of ATM were again not associated with second cancers [61].

Mutations in the familial tumor suppressor genes, TP53, BRCA1, and BRCA2 have also been explored [19]. Analysis of Hodgkin survivors with secondary cancers including breast have been conducted. Mutational analysis using genomic DNA was completed on 44 patients for the presence of the TP53 germline mutation. Nineteen of the survivors with secondary breast cancer were also surveyed for presence of the BRCA1 and BRCA2 germline mutations. Importantly, the study was unable to locate germline mutations that have been associated with the development of cancer, specifically TP53, BRCA1, or BRCA2 [19]. To further examine a possible mutation of the BRCA1 and BRCA2, loss of heterozygosity of these genes was evaluated but was not found [62]. This has led researchers to speculate that second cancers among HL survivors is a result of inactivating germline mutations in other DNA repair genes not yet identified that interact with TP53, BRCA1, and BRCA2 [18, 62].

Candidate Polymorphisms

Because the surveillance of leading candidate genes has not provided insight into the etiology of secondary cancers, the focus has shifted to candidate polymorphisms. The study of common population polymorphisms of drug metabolizing enzymes has become an area of interest, most specifically those of the glutathione-S-transferase (GST) [20, 21, 63]. It is hypothesized that a homozygous genetic deletion of the null genotype of (GST) M1 and P1 may impair the detoxification of environmental genotoxins or chemotherapy resulting in a secondary malignancy [21]. The Childhood Cancer Survivor Study (CCSS), a multi-institutional study group conducted a longitudinal study on 14, 054 survivors of childhood cancer and explored the possibility of gene polymorphisms within the GST and XRCC1, a key member of the base excision repair, as contributing to second cancers among HL survivors [7]. From this data base, 650 Hodgkin survivors submitted buccal cell specimen for GST study. The results suggested a modest possible role of the GSTM1 null genotype, loss of function, in the development of secondary cancers; however, the effect was only significant when the GSTM1 and GSTT1 were considered together [7]. The study also examined the polymorphism of XRCC1 which participates in DNA strand break repair and is important in radiation therapy recovery. The results were not statistically significant but suggested that the presence of 1 codon 399 glutamine allele provided protection against thyroid carcinoma and a slight increase in secondary breast cancer. The CCSS research suggests that individual sensitivity may be a result of multiple gene interactions or complexes [7].

Research supports that highly penetrant mutations such as BRCA1 and BRCA2 are associated with only 5% of primary breast cancer occurrences [64]. Therefore,

investigators have hypothesized that the expression of less penetrant functional polymorphisms combined with environmental interactions are associated with the majority of disease. While candidate polymorphisms have been evaluated according to their biochemical and physiological pathways, there appears to be multiple gene interactions that may play a role in breast cancer onset. A recent study explored the association of primary breast cancer with individual genes and with two- and three- gene combinations. The study enrolled 1,716 women with primary breast cancer and 4,435 controls to explore single gene polymorphisms in genes known to be associated with breast cancer. The findings were significant for two-and three-gene combinations suggesting an interaction of genes and the development of breast cancer [64]. While this study was conducted with primary breast cancer patients, it suggested the need for an exploratory study of gene expression and associated polymorphisms in determining gene to gene interaction. This is the approach of this study of women after secondary breast cancer. While the etiology of secondary breast cancer among HL survivors is not the result of a single variable, the cited literature describes both primary and secondary breast cancer as resulting from multiple events or interactions. Thus far, the candidate genes and polymorphisms can not fully explain the multiple event or interaction. The hypothesis guiding this study was intended to could give additional insight into the interaction between treatment variables and inherent characteristics, genetic polymorphisms, of the individual and the role in development of secondary breast cancer.

Testing Methodologies

From the reviewed literature, determining the etiology of secondary breast cancer using a candidate gene approach has not been successful [18, 19, 61, 62]. Additionally, studies have found only weakly associated polymorphisms with an underlying etiology that is polygenic [20, 21, 63, 64]. The candidate gene and polymorphism approach has given limited information into the inherent predispositions; therefore, it is reasonable to undertake an exploratory study to describe and compare the whole genome expression and genotype of HL survivors with secondary breast cancer compared to those without secondary breast cancer. A case-control design was used in the current study and is one of two basic experimental designs used with gene-association studies [65]. This design is an emerging methodology to examine gene-environment interactions in populations with small sample sizes. The difference in global gene expression between groups was measured with the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array. The difference in genotype between groups was measured by the presence of single nucleotide polymorphisms (SNP), between groups with the Affymetrix GeneChip® Human Mapping 500K Set. Combined these analyses were compared to determine the significant global gene expression and genotype differences between the two groups. After determining significant differences between the case and controls, the genotype and global gene expression of only the women with secondary breast cancer were examined to determine candidate genes associated with genotype allele frequency.

Microarray Technology

The technology of microarray analysis has exploded over the last 10 years, with this research technique being applied to numerous studies of medical and clinical relevance [66]. High-throughput microarray technology provides major areas of genomic exploration: genome-wide analysis of gene expression, SNP genotyping, copy number variants, transcript mapping, and resequencing [67]. This technique can now genotype drug metabolizing enzyme genes, resequence tumor suppressor genes, and classify cancer by differential gene expression and prognostic indicators. A number of studies have identified prognostic and predictive ‘gene’ signatures as prognostic indicators in the treatment of breast cancer [24]. Most of the early work with microarray analysis and primary breast cancer has been done with tumor samples taken at the time of surgery. These analyses have enabled researchers to analyze breast tissue and obtain a better understanding of the molecular characteristics and clinical behavior of breast cancer. This has been most helpful in the analysis of gene amplification and coamplification in prognosis and treatment of breast cancer [25]. Examination of breast tissue will give expression at the site of the disease, but will not give a global picture of the gene interaction that may have predisposed individuals to multiple cancers. With the hypothesis that the second cancer is the result of an inherent gene-gene/environment interaction, peripheral blood was the sample used for the global gene expression and genotype.

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array used in this study analyzes the expression level of over 47,400 transcripts and variants including 38,500 human genes. The array is comprised of over 54,000 probe sets and 1,300,000 distinct oligonucleotide features [29]. Microarray analysis allows genomic measurement of an individual’s gene expression under a variety of genetic and environmental conditions and may best capture biological phenomena through the interaction of multiple genes. This gene expression is typically measured as an up or down regulation of genes/gene pathways associated with the studied phenotype. This analysis has been used to study and compare healthy and diseased samples, using the steady state mRNA to infer maladaptive changes associated with disease. The earliest method was fold-change which is a metric for comparing a gene’s mRNA expression between two experimental conditions such as case-control. The challenge with fold change has been generalizability of findings with arithmetic definitions differing between investigators. While these studies have made it possible to compare the expression levels for active genes, challenges remain most specifically with study design, inference, gene classification, and validation [68].

Single nucleotide polymorphism (SNP) analysis was performed using the Affymetrix GeneChip Human Mapping 500K set. This assay comprises two arrays designed for high throughput parallel genotyping of ~500,000 SNPs that span the human genome with a median intermarker distance of 2.5 kb [29]. The sequence of each allele is determined using 24 probes that interrogate the forward and reverse strands at six positions across each SNP. Genotype calls generated from the 500K Mapping assay ([500k_datasheet.pdf](#)) [29] are highly concordant (HapMap agreement, 99.3%) and

consistent (Mendelian consistency, 99.9%). Overall, 85% of the genome is represented within 10 kb of a SNP. Further, with a mean allele heterozygosity of 0.30 across a wide range of populations and ethnicities, the Mapping 500K set is well-suited for genotype analysis within diverse ethnic groups.

The global gene expression from the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array was combined with the SNP genotype from the Affymetrix GeneChip Human Mapping 500K set in an association analysis. The SNP array data also analyzed for amplification, deletion, or no change in copy number. The expression data were treated as quantitative phenotype and the analysis screened the *in silico* copy number of the most significant SNP's and found an association with the expression of a list of genes and pathways. An understanding of the gene expression may bring an understanding to the pathways that connect the genotype to the cellular traits [69] and a better understanding of the biological processes driving the association between expression and response.

Summary of Review

The reviews of literature describe known risk factors associated with secondary breast cancer among HL survivors and include radiation, chemotherapy, age, time from diagnosis, hormonal influence and family history. While the radiation dose has been implicated as a significant risk factor, the risk is inclusive of all who are treated with chest radiation exposure regardless of treatment dose. The age of radiation exposure before 30 years has clearly been established as a risk factor, which would include all pediatric female HL survivors. Additionally, the time from exposure has been cited as being greatest at 15-20 years post therapy; however, these women continue to have an increased absolute and lifetime risk. Administration of alkylating agents has been associated with a decreased breast cancer risk; however, the timing of alkylating exposure in the pediatric population may not be protective. Perhaps pelvic radiation is the most productive endocrine insult, leaving the question and unclear association of hormonal supplementation in HL survivors. The hormonal influence of pregnancy within 60 months after HL therapy has been found to be associated with an increased breast cancer risk and should be discussed with survivors. Lastly, the family history of not only breast cancer, but familial cancer should be reviewed with all HL survivors. The review of risk factors attempt to assist the clinician in identification of survivors at greatest risk for secondary breast cancer. However, it is my assumption that the pediatric HL patient is a vulnerable host due to radiation exposure of growing, immature cells and all are at an increased risk of secondary cancers. Therefore, this study was directed at exploration of inherent factors that impede the DNA repair, facilitate hormonal exposure, or prevent apoptosis leading to the development of secondary breast cancer.

CHAPTER 3. METHODS

Research Design

Research Design and Study Population

This descriptive cross-sectional case-control study assessed the gene expression and genotype profile of women who are survivors of childhood HL (see Figure 3.1). The first group (cases) included women who were treated with mediastinal radiation and developed secondary breast cancer. The second group (controls) included women who were treated with mediastinal radiation and have not developed secondary breast cancer.

The study definition of breast cancer was the development of documented breast cancer after treatment for pediatric HL, which included mediastinal radiation. Estimates regarding the timing of the second cancer range from 15-20 years after radiation exposure [1-2]. The study definition for non-breast cancer was the absence of breast cancer and a negative mammogram within one year of study.

Eligibility Criteria for Case and Control Participants

- Females treated for pediatric HL at St. Jude Children's Research Hospital between 1970 to 1991
- 18 to 55 years of age
- Cases had documented breast cancer prior to 51 years of age
- Controls had documented negative mammogram within one year of participating in the study
- Controls matched treatment variables of the breast cancer case and included: age at diagnosis, mantle/mediastinal radiation dose, para-aortic/splenic radiation dose, alkylating agents dose (cyclophosphamide and procarbazine), and time since diagnosis of HL
- English-Speaking
- Written consent for participation

Exclusion Criteria

- Confirmed diagnosis in the control group of any secondary cancer
- Treatment within the last 6 months for any cancer, excluding basal and/or squamous cell carcinoma, in the case group
- Inability or unwillingness of research participants or legal guardian/representative to give written informed consent

Hodgkin Lymphoma Treated with Radiation Therapy

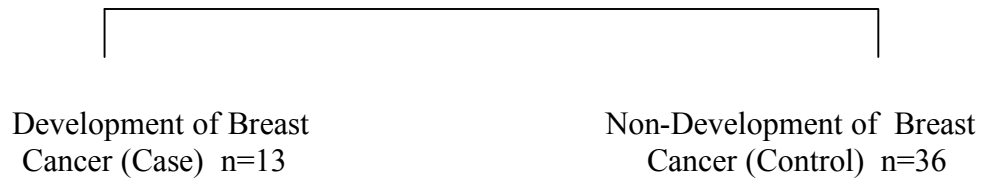


Figure 3.1 Study Design.

Procedures for Accomplishment of Aims

Case and Control Participation

The case/controls were taken from the data base of active alumni who were treated for pediatric HL at St. Jude Children's Research Hospital (SJCRH) between July 1970 and July 1991. After obtaining Institutional Review Board approval (XPD05-108), a retrospective chart review was conducted and found 440 persons treated for pediatric HL within these years, of which 190 are women and 19 (10%) of these women subsequently developed secondary breast cancer. Of the 19 women with secondary breast cancer, 13 women were identified as survivors of secondary breast cancer after HL. All surviving secondary breast cancer patients (cases) were asked to participate in the study, while the control participants were selected based on controlling variables from the treatment characteristics. Once the cases were identified, the protocol was submitted for approval from the SJCRH and University of Tennessee Institutional Review Board (Appendix A).

The 13 HL breast cancer survivors are the study cases and were matched with 3 controls, based on power estimates as discussed in the statistical section. In identifying the controls, 101 women were eligible as controls and were randomly selected by the study biostatistician after matching on the variables of age at diagnosis, mantle/mediastinal para-aortic and pelvic radiation dose, alkylating agent (cyclophosphamide and procarbazine) dose, and time since HL diagnosis. These variables were classified into classes based on Low/None, Medium and High groups. This led to no match for four cases. Since the list of the controls was limited in matching with all cases, the median cut-off for age at diagnosis and time since diagnosis was used. The final age classes were age < 14.7608 years (Age Dx Class =1) and age \geq 14.7606 (Age Dx Class =2) and time since diagnosis < 27.8696 years (Time Dx Class =1) and time since diagnosis \geq 27.8696 (Time Dx Class=2). The dose of maximum chemotherapy was then calculated 'Max Chemo' dose based on the maximum doses of Cyclophosphamide and/or Procarbazine patient received. This chemotherapy dose was defined as high and low dose of Max Chemo dose using a cut-off <7000 mg (Max Chemo Class =1) and \geq 7000 mg (Max Chemo Class=2). Similarly, the dose of radiation was calculated using the same analogy as for Max Chemo class to define Max Radiation Class, where Max Radiation Class =1 when Max Radiation < 30 Gy and Max Radiation Class=2 when Max Irradiation \geq 30 Gy and Max Irradiation is the maximum amount of radiation patient received at Mantle, Mediastinum, Paraarortic and/or Pelvic. When more than 3 possible controls were available for each case, simple random sample was used to select controls. The controls were matched on variable significance with the most important being radiation dose, then chemotherapy, time since diagnosis, and age at diagnosis. Of the 13 cases, one was treated on HOD 68, four cases were treated on HOD 72, five on HOD 80, one on HOD 90 and two according to best clinical management. Eight of the cases received alkylating agents, while four were treated with radiation alone. The dose of mantle/mediastinal radiation ranged from 39-20Gy. The controls were treated with similar doses of chemotherapy and radiation according to their matched case. If a control

was not located or unable to participate, the biostatistician continued the randomization of controls, this process was repeated as needed until all matched controls were considered. If a control was found to be a case, the control was reclassified as a case and matched with 3 controls.

Once the cases and controls were identified, eligible participants were discussed with the SJCRH alumni physician. An introductory letter (Appendix B) and consent form (Appendix C) was mailed to the eligible participants explaining the purpose of the study and requirements of obtaining demographic information (Appendix D) and a blood sample. Two weeks after receiving the letter, each eligible participant received a follow-up phone call from the primary investigator (PI), who responded to questions and obtained verbal consent for completion of a demographic health questionnaire over the phone. After the participant gave verbal consent for the study, the demographic questions were reviewed and responses from each participant recorded by the PI for validation accuracy. The demographic responses were recorded and categorized; accuracy of extraction and categorization were verified for inter-rater reliability with another researcher with 98% agreement between raters. Written consent was obtained from all participants and returned to SJCRH. Once this consent was received, each participant was officially entered on study and mailed a kit for blood collection. To enhance the scope of future research of HL and secondary cancers, a biological repository consisting of stored blood (DNA, serum phase) from all participants was discussed with each participant and written consent was obtained for those agreeing to participate in the biological repository. The stored DNA sample was obtained from the initial 7mLs of blood collected in the EDTA tube.

Enrollment of Cases and Controls

Ten of the 13 known survivors of secondary breast cancer were initially enrolled on study. Three of the survivors were lost to follow-up; however, it was later learned that one of these lost to follow-up had died within the previous two months secondary to bacterial colitis. The matched controls to these 10 cases were then contacted for enrollment. Telephone contact with these controls found two additional Hodgkin survivors with secondary breast cancer. One had developed breast cancer within the previous 6 months, while the other had been diagnosed within the month and had received no breast cancer therapy. These women were then enrolled as cases and 3 controls were matched for each. A study packet was sent to all potential controls with follow-up telephone contact. The number of participants who gave consent, were ineligible, declined, or who were lost to follow-up from the initial case control matched variable randomization are included in Table 3.1.

Fifty one controls were contacted, with 34 found to be eligible and consenting to participate. Two agreed to participate, but were ineligible due to a secondary basal cell carcinoma and another due to pregnancy secondary to gene expression associated with pregnancy. Another who was consented and enrolled on study was not included in the demographic analysis due to subsequent development of hepatocellular carcinoma but

Table 3.1 Initial Control Contact for 12 Matched Cases.

Control Participation	Responses
Controls contacted via mail or phone	n=51
Consented	n=34
Ineligible at time of contact	n=2
Declined	n=3
Lost to follow-up	n=12

was included in the expression and genotype analysis. Three declined participation due to time conflict in arranging blood collection and 12 were lost to follow-up. A total of 34 controls were enrolled from the initial case-control matching, with 33 meeting all eligibility criteria. The list of potential controls was reviewed for additional participants on the criteria of radiation dose and chemotherapy as previously described. Three additional controls were identified and matched as described, with one consenting (35 controls) to study and two lost to follow-up. No other controls were matched to the cases on the variables of radiation and chemotherapy dose.

Approximately 6 months after the study began, the 13th case was reported to the SJCRH database and this survivor was subsequently enrolled. In matching on race (African American) and disease variables, only three controls were identified with one consenting and two lost to follow-up. This resulted in 13 cases and 36 controls. Ten cases have 3 matched controls per case, 2 cases have 2 matched controls per case, and the last case described had only 1 matched control.

Blood Collection for RNA and DNA Extraction

Sample collection was discussed with each participant and arrangements were made accordingly. The samples were collected at SJCRH, an affiliate clinic, local physician's office, or at the participant's residence. Samples were collected by a registered nurse, technician, or physician designee. Collection supplies and instructions were provided to those responsible for blood collection. Instructions regarding overnight mailing and shipping procedure were also outlined. No cost to the participant was associated with the blood collection. Risks associated with the venipuncture were shared with participants and included the possibility of bruising at the venipuncture site.

For RNA extraction, 5 milliliters (mL) of blood were collected directly into two 2.5mL PAXgene™ Blood RNA System (Qiagen, Valencia, CA). The PAXgene tube contains a reagent that stabilizes intracellular RNA by reducing RNA degradation. The blood was collected into the PAXgene tube according to manufacture's instruction. After collection, each tube was inverted 10 times and stored in an upright position for a minimum of two hours at room temperature to allow for incubation. At the completion of incubation, samples were stored on ice for shipment to SJCRH. The samples were stored at -20°C for a minimum of 24 hours and then transferred to a non-frost free freezing unit and stored at -80°C until processing.

For DNA extraction, 7 mLs of whole blood was collected in tubes containing EDTA as an anticoagulant. The blood sample was mixed and placed on ice for shipment to SJCRH. If the sample was not immediately processed, the sample was stored at 4° for up to 72 hours.

RNA Extraction

The RNA extraction was performed under the Processing Clinical Samples for DNA, RNA, Protein and Cryopreservation Procedures Manual within the Department of Pharmaceutical Sciences at SJCRH. Samples collected for RNA extraction were removed from the -80°C freeze and placed in a 37°C water bath for a quick thaw. In the lab, the RNA was extracted using the TRI REAGENT® protocol. After thawing, the tubes were centrifuged for 10 minutes at 5000xg (gravitational acceleration) using a swing-out rotor at room temperature. After centrifuge, the supernatant was decanted and discarded. Five mL of RNase-free water was added to the pellet. The pellet was resuspended into solution and transferred to an 8cc polypropylene tube and capped. The tube was centrifuged for 10 minutes at 5000xg in a swing-out rotor at room temperature. After centrifuge the supernatant was removed. Two mL of TRI reagent was added and mixed well with the pellet. The TRI reagent was allowed to digest for 5 minutes. The homogenate was then supplemented with 0.4 mL chloroform (0.2 mL for 1mL of TRI reagent) and vortexed for 15 seconds. The mixture was stored at room temperature for 5 minutes and then centrifuged at 12000 rcf for 15 minutes at 4°C. A sterile transfer pipette was used to transfer the colorless upper aqueous phase into a new tube. To the colorless aqueous fluid, 1 mL of isopropanol (0.5 mL of isopropanol per 1 mL of TRI reagent used in the initial homogenization) was added. The mixture was vortexed and stored at room temperature for 5 minutes. The mixture was then centrifuged at 12000 rcf for 15 minutes at 4°C. The supernatant was discarded and the tube dried without disruption of the RNA pellet. One mL of cold 75% ethanol was mixed with the pellet into solution and transferred into a 1.5 mL microtube (Axygen). The pellet/ethanol mixture was centrifuged at 10000 rcf for 5 minutes at 4°C. The supernatant was discarded, leaving the RNA pellet to which 10-15 mL of RNase free water was added. The pellet and water were mixed with 11 mL of solution placed in a 0.5 mL microtube (Axygen). Both microtubes were placed on ice, with the 0.5 mL microtube transported to the Hartwell Center for processing. The remaining RNA solution was placed in an 80°C non-frost free freezer for disposal at the completion of the study. This system typically yields 4-10µg of pure RNA from 2.5 mLs of blood.

DNA Extraction

All buffers and reagents used in the extraction process were commercially purchased in the QIAGEN Blood and Cell Culture DNA Kit. The 7 mL blood samples were transferred from storage and each was transferred to a 50 mL screw-cap labeled tube. The lysis protocol began with 1 volume (7mLs.) of ice-cold Buffer C1 and 3 volumes (21mLs.) of ice-cold distilled water. The tubes were then inverted several times and incubated in an ice bucket for 10 minutes. The Buffer C1 lyses the cells but stabilizes and preserves the lymphocyte nuclei. The blood suspension becomes translucent due to lysing of the erythrocytes and release of hemoglobin.

After ice incubation, the 50mL tube of lysed blood was centrifuged at 4°C for 15 minutes at 1300xg. The blood suspension was poured off and the nuclear pellet was

visible. Ten mLs of Buffer G2 was added into the tube and the tube was vortexed for 30 seconds at maximum speed, resuspending the nuclei. The resuspended nuclei was then mixed with 200 μ l of QIAGEN Protease and incubated in a 50°C water bath for 60 minutes. The Buffer G2 and QIAGEN Protease work together in stripping the genomic DNA of all bound proteins, facilitating efficient DNA removal during purification.

Part II of the DNA extraction process was isolation of genomic DNA from the nuclei. Prior to beginning this process, QIAGEN Genomic-tip 500/G was equilibrated with 10 mLs of Buffer QBT and allowed to empty by gravity flow. Once the equilibration buffer emptied from the genomic-tip, the sample was removed from the water bath and vortexed for 10 seconds at maximum speed. This step prevents poor DNA flow rates due to increased viscosity. The DNA was then filtered through the equilibrated genomic-tip. Once the DNA completed gravity flow through the genomic-tip, 2 x 15 mL washings of Buffer QC were filtered through the tip. These two washings are sufficient to remove all contaminants in the majority of DNA preparations.

The QIAGEN Genomic-tip was then placed over a clean 30 mL collection tube and the genomic DNA was eluted with a 15 mL washing of prewarmed Buffer QF. The warming of the buffer increases DNA yield. To precipitate the DNA, 10.5 mLs. of isopropanol was added to the eluted DNA and centrifuged at >5000xg for 15 minutes at 4°C.

Lastly, the DNA pellet was washed with 1 mL. of 70% ethanol, vortexed, and centrifuged at >5000xg for 15 minutes at 4°C. The supernatant was removed and the DNA pellet was allowed to air dry in the tube for 5-10 minutes. The DNA was resuspended in 1 mL of TRIS(TE) EDTA Buffer, pH 8.0. DNA dissolves best under slightly alkaline conditions with a pH 8.0-8.5. To facilitate the concentration equilibration, the suspension was transferred to a microcentrifuge tube and placed on a rotator (Labnet labroller) overnight.

After the DNA samples were rotated overnight, the concentration of each was determined with the technique of PicoGreen® ultrasensitive fluorescent nucleic acid stain use for quantization of double-stranded DNA (dsDNA) in solution protocol. After the DNA concentration for each sample was determined, 50 nanograms (ng) of each sample were run on an electrophoresis gel at a concentration of 0.8% with High DNA Mass™ Ladder by Invitrogen. High DNA Mass™ Ladder is suitable for estimating the mass (quantity) of unknown DNA samples by ethidium bromide staining. This ladder consists of an equimolar mixture of six blunt-ended DNA fragments of 10, 6, 4, 3, 2, and 1 kb. Electrophoresis of 4 μ l of the High DNA Mass™ Ladder results in bands containing 200, 120, 80, 60, 40, and 20 ng (520 ng total) of DNA, respectively. The gels were run at 75 volts for 180 minutes. At the completion, each gel was removed and placed in the Fotodyne transilluminator and camera and analyzed with the Fotodyne Electrophoresis Gel Analysis Software. This software program provides quantification of the sample through the use of the High DNA Mass™ Ladder and provides an accurate molecular weight and DNA quantity calibration. After each sample concentration was determined,

250ng of DNA was plated into a 96 well plate for Nsp analysis and 250ng of DNA was plated into a 96 well plate for Sty analysis and forwarded to the Hartwell Center.

Measurement of RNA: Quality Control

The purity and quantity of RNA was measured using a Spectromax Plus 384 spectrophotometer. The OD260 values were used to quantify the concentration, while the OD260/280 ratios were used to assess the purity of the RNA sample. OD260/280 ratios between 1.9 and 2.1 indicate high purity of RNA; ratios <1.8 were associated with protein contamination, while ratios >2.1 indicated carryover of small oligonucleotides and nucleotides. RNA quality was also evaluated with an Agilent 2100 Bioanalyzer. An electrophoresis profile generated by an Agilent analysis was used to calculate the relative abundance of the 28S and 18S ribosomal RNAs which are surrogate markers of mRNA degradation. A ratio <2.0 indicated RNA degradation which can also be seen as fragments running ahead of the 18S species. RNA samples that passed these quality control measures were processed for Affymetrix GeneChip analysis. Five samples did not pass these quality control measures and the additional PAXgene tube was extracted for RNA and resubmission to the Hartwell Center for quality control and processing. All five passed quality control upon the second process.

Processing of Extracted RNA: Hartwell Center, SJCRH

As discussed, gene expression analysis was performed using the Affymetrix HG-U133Plus2 GeneChip array. RNA samples were processed using the Affymetrix Eukaryote Two-Cycle Target Assay (www.affymetrix.com/support/technical). In the process, 100 ng of total RNA was used as starting template for cDNA synthesis. The RNA was annealed to a T7-oligo(dT)(24) primer containing the recognition site for the bacteriophage T7 RNA polymerase. A first cycle of cDNA synthesis was accomplished using the SuperScript II cDNA synthesis kit according to the manufacturer's conditions (Invitrogen). The resulting cDNA was then used as template for generating antisense cRNA using the MEGAscript T7 in vitro transcription kit (Ambion). The purified cRNA (400 ng) was then primed with random hexamers to initiate a second cycle of cDNA synthesis. The second round cDNA was then used as a template to generate biotin-labeled cRNA using the Affymetrix GeneChip In Vitro Transcription labeling kit. Twenty micrograms of biotin-labeled cRNA was fragmented by heating and metal induced hydrolysis and added to a hybridization cocktail containing probe array controls and blocking agents (BSA and herring sperm DNA), then incubated overnight at 45° C on a GeneChip array. Following hybridization, the arrays were washed automatically using a GeneChip Fluidics Station 400 under high stringency conditions to remove non-hybridized labeled cRNA. To detect bound target, arrays were stained with Streptavidin Phycoerythrin (SAPE, Molecular Probes), washed again, and then scanned using the Affymetrix GeneChip Scanner 3000.

Expression signals for each transcript were calculated by the MAS5 statistical algorithm using the Affymetrix GCOS software (version 1.4). In this method, the global signals on the array were scaled to a 2% trimmed mean of 500. Detection calls (Present, Absent or Marginal) were determined using the default parameters of the software and recommended by the Affymetrix GeneChip protocol.

Processing of Extracted DNA : Hartwell Center, SJCRH

DNA samples submitted for SNP analysis were processed according to the Affymetrix protocol ([affymetrix/500k_assay_manual.pdf](#)). Total genomic DNA (250 ng) was digested with a restriction enzyme (10 units of *NspI* or *StyI*, 2h at 37°C), and then ligated to adapters that recognized the cohesive four bp overhangs (800 units T4 DNA ligase, 0.75 uM adapter, for 3 hours at 16°C). Each adapter contained a unique sequence that permitted amplification by using a single PCR primer. The ligated DNA was amplified using thermocycling conditions (5 uM primer, 1X Clontech TITANIUM Taq polymerase; 30 cycles of 30 sec at 94°C, 30 sec at 60°C, 15 sec at 68°C) that enriched restriction fragments from 250bp to 1100 bp in size. This selective reduction in genomic complexity is key to the success of the mapping array. Following PCR, 90 ug of amplified DNA was fragmented to 25-100 bp by digestion with DNase I (0.25 units enzyme, 35 min at 37°C), end-labeled with 857 uM biotinylated-nucleotide using terminal deoxynucleotidyl transferase (105 units enzyme, 4 hours at 37°C), was added to a hybridization cocktail containing probe array controls and blocking agents (human Cot-1 and herring sperm DNA). This mixture was then incubated overnight at 48°C on a GeneChip array. After hybridization, arrays were washed automatically using a GeneChip Fluidics Station 400 under high stringency conditions to remove non-hybridized labeled DNA. Arrays were incubated with Streptavidin Phycoerythrin (SAPE, Molecular Probes), washed, and then incubated with biotin-conjugated anti-streptavidin antibody. After removal of the antibody solution, the arrays were re-stained using the SAPE, washed again, and then scanned using an Affymetrix GeneChip Scanner 3000.

Genotype calls are initially determined using the Affymetrix Genotyping Software (GTYPE) version 4.0. The GTYPE software uses a dynamic model algorithm to call genotypes on an array by array basis. After processing the complete set of samples, genotype calls were generated using the BRLMM algorithm (also available within the GTYPE software). The BRLMM method [70] uses a multi-chip analysis algorithm which provides superior accuracy and consistency of calls as compared to the GTYPE algorithm.

Statistical Procedures

The study is based on a case-control design. The cases include 13 survivors and for each case there are three matched controls. This 1:3 matched case control design gives an efficiency of 75% for binary measures [71], an increase of 25% from a 1:1 match. Controls were matched to cases based on the previously described variables including; mantle/mediastinal radiation dose, para-aortic/splenic radiation dose,

alkylating agent dose (cyclophosphamide and/or procarbaine), age at initial diagnosis, and time since HL diagnosis. In the surviving patients, 101 controls were identified and stratified based on the matching characteristics from which 3 controls from each stratum were randomly selected. If a control declined participation another control was randomly selected from the stratum of matched controls. During the study if a control became a case, three matched controls (one for the original case and two for the new case) were randomly selected. It was proposed the study would take approximately 18 months. All data were collected and ready for analysis approximately 13 months after opening the study.

Statistical analyses for specific aims are outlined as follows:

Primary Aim One

Assess whether global gene expression profiles differ between two groups of radiated pediatric female Hodgkin lymphoma survivors with and without the development of secondary breast cancer.

Affymetrix U133 Plus 2.0 arrays were utilized to measure extracted RNA gene expression in blood samples taken from both the cases and controls. Normalized gene expression signals were computed using Affymetrix MAS 5.0 software. The normalized signals were log-transformed prior to subsequent statistical analysis [72]. For each probe set, the two-sided Wilcoxon rank sum tests were used to compare the mean/median expression of the cases to that of the controls. To account for multiple testing, the set of resulting p-values was used to estimate the false discovery rate (FDR) with a robust method [73].

Gene-set enrichment analysis was performed with a structured permutation approach [74]. Affymetrix-defined pathways were used to define gene sets and the negative sum of Wilcoxon p-values was used as the gene-set statistic. The significance of the gene set statistic was determined via 10,000 permutations.

Primary Aim Two

Assess whether genotype profiles differ between two groups of radiated pediatric female Hodgkin lymphoma survivors with and without the development of secondary breast cancer.

Affymetrix GeneChip® Human Mapping 500K Set was used to characterize the genotype of each case and control. The BRLMM algorithm [70], as implemented in Affymetrix software, was used to generate a genotype call (AA, AB, BB, or NoCall) for each of the 500K SNP markers and each sample.

First, markers with low genotypic variability were removed due to statistical futility prior to subsequent analysis. For each of the remaining SNPs, the association of genotype with case-control status was explored with a chi-square test. To adjust for multiple-testing, the resulting p-values were used to estimate the FDR with a robust method [73].

Additionally, the data were used to test for association of full-gene genotypes with case-control status using a structured permutation approach [74]. SNPs were mapped to genes on the basis of physical location relative to the start- and stop-transcription positions. SNPs between the start- and stop-transcription positions or within 5,000 base pairs of the start- and stop-transcription positions were mapped to the genes. The negative sum of log-p-values from the asymptotic chi-square test was used as a statistic summarizing the significance of each group of SNPs mapped to a specific gene. The significance of the SNP-set statistic was determined via 10,000 permutations. The method of Benjamini and Hochberg [75] was used to compute FDR-adjusted p-values [76].

Primary Aim Three

Identify potential candidate genes through identifying copy number variations that associate with the risk of secondary breast cancer.

For each SNP marker, summary signals were computed using dChip SNP software [77]. These signals were subsequently reference normalized [78]. For each marker, the median and inter-quartile range of signal values across controls was computed. Then, for each subject, a standardized difference was computed for each marker by subtracting the median among controls and dividing by the inter-quartile range among controls.

A Bayesian segmentation algorithm [79, 80] was used to segment the standardized difference profile of each chromosome in each subject. Segments with fewer than 25% of standardized differences being positive were identified as *in silico* regions of loss and segments with at least 75% of standardized differences being positive were identified as *in silico* regions of gain.

Finally, the segmentation results of individuals were used to segment the genome into regions with distinct patterns of gain/loss across the entire cohort. For each region with at least 5 subjects with an *in silico* gain or loss, we used Fisher's test to explore the association of *in silico* copy number with case-controls status.

Secondary Aim

Explore the association between lifestyle characteristics (hormone use, pregnancy, lactation), family history and the occurrence of breast cancer.

Descriptive analyses of cases and controls were conducted on the following variables: age, age at Hodgkin diagnosis, years since HL diagnosis, radiation, chemotherapy, age at menses, age of hormone initiation, age at menopause, age at first pregnancy and time from HL diagnosis and first pregnancy. Two sample t test was used to compute the p values.

Frequency tables were generated for menses status, use of hormone therapy and status of pregnancy, and simple logistic regression was used to evaluate the association between these variables and case-control status.

In addition, log-linear regression was used to evaluate the differences between case and control in number of pregnancies and number of live births. This part of the analyses was conducted using SAS version 9.1 (SAS Institute, 2003).

CHAPTER 4. RESULTS

Introduction

Data obtained over the course of this study are presented in this chapter. The results are presented according to the study aims.

Description of the Sample

Treatment Characteristics

The study sample consisted of 48 women, from whom blood was collected and analyzed for global gene expression and genotype. The women who developed secondary breast cancer (cases) were matched with women who did not develop secondary breast cancer (controls) according to radiation dose, maximum alkylating chemotherapy dose, age at diagnosis and time since diagnosis. The cases and controls were not statistically different in current age, age at HL diagnosis and time since HL diagnosis as described in Table 4.1. Treatment variables of mediastinal/mantel radiation and maximum dose of alkylating chemotherapy were well matched. Total radiation dose was the most significant matching variable and ranged from 41 to 20 Gy among cases with a median dose of 35.5 Gy.; radiation dose for the controls ranged from 15 to 41.7 Gy with a median dose of 35 Gy. Alkylating chemotherapy total dose was the second most significant matching variable because the induction of premature menopause that is associated with alkylating chemotherapy is thought to be a protective mechanism against the development of breast cancer. Ten of the 13 cases (76.9%) received alkylating therapy; 24 of the 35 controls (68.5 %) received alkylating therapy. The cases and controls were not matched on pelvic and/or para-aortic radiation; however, it is important to determine this potential protective effect. Only 1 case (7%) received pelvic radiation and 8 cases received para-aortic which would have a smaller radiation scatter to the ovaries. In comparison, 16 controls (45.7%) had pelvic radiation of which 4 of the 16 received combined para-aortic and pelvic radiation, 11 controls received para-aortic radiation, 7 controls received no pelvic or para-aortic radiation.

Reproductive and Family Characteristics

Each woman also completed a demographic form that included items regarding her hormonal use, reproductive history, and cancer family history. Differences between cases and controls on variables of menses, menopause, hormonal therapy, age at first pregnancy, and time from HL to first pregnancy are listed in Table 4.2. There were no statistical differences between the cases and controls in the age of menses, current status of menses, or use of hormonal therapy. There is a statistical difference in age of menopause between the cases and controls, 16 controls (45.7%) who received pelvic

Table 4.1 Demographic Data: Current Age, Age at Time of Diagnosis and Years Since Diagnosis For Study Participants Who Did and Did Not Develop Secondary Breast Cancer.

	Study Sample n=48	With Breast Cancer n=13	Without Breast Cancer n=35	p value
Age in yrs (mean \pm SD)	40.2 \pm 5.6	41.9 \pm 6.7	39.6 \pm 5.2	0.2181
Age at Hodgkin Diagnosis (mean \pm SD)	14.3 \pm 3.2	14.6 \pm 2.1	14.3 \pm 3.3	0.8127
Years Since HL Diagnosis (mean \pm SD)	26 \pm 5.9	27.5 \pm 5.2	25.4 \pm 6.01	0.2722

Table 4.2 Demographic Data: Age at Menses, Menopause, and Hormonal Use, According to Study Participants Who Did and Did Not Develop Secondary Breast Cancer.

	Study Sample n=48	With Breast Cancer n=13	Without Breast Cancer n=35	P-value*
Age at Menses (mean \pm SD)	12.7	12.8 \pm 1.4	12.7 \pm 2.3	0.8387
Age at Menopause (mean \pm SD)	34	38.3 \pm 2.5	32.1 \pm 9.6	0.0191
Currently Having Menses	Yes=20 No=28	Yes=4 No=9	Yes=16 No=19	0.3544
Use of Hormones (mean \pm SD)	48	13	35	NA
Age Started Hormones (mean \pm SD)	21.6	22.9 \pm 6.5	21.1 \pm 7.7	0.4995
Age Stopped Hormones (mean \pm SD)	32.6	32.8 \pm 8.3	32.4 \pm 7.7	0.9030
Age at 1 st Pregnancy (mean \pm SD)	25.4	26.5 \pm 3.5	24.9 \pm 4.6	0.3508
Time from HL to 1 st Pregnancy (mean \pm SD)	10.6	11.4 \pm 3.1	9.9 \pm 5.0	0.3992

radiation experienced the earliest onset of menopause with an average of 27.8 years of age (range 16-40), while only 3 of the 16 continue (28.7%) to have menses. All of the cases and controls reported hormonal use with no differences in the ages that hormones were initiated or discontinued. There was no statistical difference in age at first pregnancy or time from Hodgkin diagnosis and first pregnancy between the cases and controls. Reproductive history outlined in Table 4.3 shows that overall 31 (64.5%) of the cases and controls reported at least one pregnancy, with 76.7% of the cases becoming pregnant and 60% of the controls becoming pregnant. Overall the cases and controls reported 69 pregnancies and 53 live births with the controls having the highest number of lost pregnancies. While the controls reported higher rates of pregnancy difficulties with a lower number of pregnancies and live births; differences in reproductive history were not statistically significant.

Family history of cancer was collected by self-report from the cases and controls. Eleven of the 13 cases (84.6%) reported a family history of cancer in a first or second degree family member. Five of the 11 (45%) cases reported among first degree relatives included: breast, prostate, esophageal, and skin. Among these 11 cases, the most common cancers among second degree relatives included: lung, breast, prostate, and lymphoma.

Thirty one of the 35 controls (88.5%) reported a family history of cancer in a first or second degree family member. Fourteen of the 31 (45%) controls had a first degree family member with cancer with breast being the most frequent followed by lymphoma, skin, and lung. The most common cancers among second degree relatives included: lymphoma, breast, and prostate.

Breast Cancer Characteristics Among Cases

For the women with breast cancer, histology and hormonal status of their breast cancer was collected from pathology report if available or self report. In addition, a retrospective chart review collected additional information related to breast cancer occurrence including age at breast cancer diagnosis in comparison to Hodgkin diagnosis, and time from radiation to diagnosis of breast cancer. The average age of secondary breast cancer among the cases occurred at 36.5 ± 6.7 years of age with a range of 27.2-51 years of age. The time from radiation exposure to the development of secondary breast cancer among the cases is 21.3 ± 5.8 years after radiation with a range of 12-34 years. The range of 34 years is the result of an outlier and if excluded the range would be 12-27 years with a mean of 20.2 ± 4.5 years from radiation to the development of breast cancer. There was no difference in the location of the breast cancer (right vs. left breast) and the tumors were predominately infiltrating ductal carcinoma. The tumor estrogen, progesterone, and HER-2 statuses are unknown in 3 of the cases. The location, histology, and expression are detailed for each case in Table 4.4.

Table 4.3 Demographic Data: Reproductive History According to Study Participants Who Did and Did Not Develop Secondary Breast Cancer.

	Study Sample n=48	With Breast Cancer n=13	Without Breast Cancer n=35	P-Value (95% CI)
Achieving Pregnancy	31	Yes=10 No=3	Yes=21 No=14	0.2827 (0.105, 1.931)
Number of Pregnancies (mean ± SD)	69	1.5 ± 1.2	1.4± 1.4	0.7223 (0.492,1.393)
Number of Live Births (mean ± SD)	53	1.3 ± 0.9	1.0 ± 1.0	0.7223 (0.347, 1.101)
Breast Feeding	30	Yes=3 No=7	Yes=8 No=12	0.5934 (0.307, 7.872)

Table 4.4 Breast Cancer Characteristics.

Location of breast cancer	Histology	Estrogen	Progesterone	Her-2
Bilateral	Adenocarcinoma: Infiltrating ductal	+	-	-
Left Breast	Infiltrating ductal	Unknown		
Left Breast	Infiltrating ductal	Unknown		
Left Breast	Infiltrating ductal	+	+	-
Right Breast	Infiltrating ductal	+	+	-
Left Breast	Infiltrating lobular	+	+	+
Right breast	Invasive ductal	-	-	-
Left Breast	Infiltrating ductal	+	+	+
Right Breast	Infiltrating ductal	-	-	Unknown
Right Breast	Infiltrating ductal	+	-	+
Left Breast	Adenocarcinoma	+	-	-
Right Breast	Ductal in Situ	Unknown		
Left Breast	Infiltrating ductal	+	-	+

Primary Aim One

Assess whether global gene expression profiles differ between two groups of radiated pediatric Hodgkin lymphoma survivors with and without the development of secondary breast cancer.

The analysis identified 100 differentially expressed genes ($p \leq 0.0015$; $FDR \approx 0.68$) and 12 pathways that showed significant gene-set enrichment ($p \leq 0.05$; $FDR \approx 0.57$).

In an analysis that ignored matching of individual cases to their selected controls using unmatched ranksum found genes of significance and 12 Affymetrix-defined pathways from 233 identified pathways with a statistical difference between cases and control were observed. The pathways of significance as well as the test statistic, permutation p value for the pathway, and false discovery rate (q) are given in Table 4.5; $p \leq 0.05$; $FRD \approx 0.57$. The pathways were generated from KEGG, a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. KEGG is a recognized bioinformatics data base and is part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo [81, 82]

The genes within the pathways were then explored for significance between the cases and controls. The pathway report files included gene specific results for genes in the pathways and a pathway p-value less than 0.10 per the method of Barry et al [74]. The analysis identified differentially expressed genes associated with the pathways and are reported in Table 4.6 $\approx FDR 0.57$. The table is labeled with the associated pathway and includes the gene name and symbol, the case average of expression, the control average of expression, permutation p value, false discovery rate (q), and the Affymetric gene probe ID. A positive value of the test statistic indicates expression of the gene tends to be higher in controls than cases and is denoted in blue; conversely, a negative value of the test statistic indicated expression of the gene tended to be higher in cases than controls and is denoted in red. A gene summary statement is also given and was retrieved from Entrez Gene; a database designed to give detailed gene information and is linked with relevant databases for additional reference [83]. This analysis provides a list of genes of significant interest for function and their association with the phenotype and became genes of high interest in the genotype analysis. It is noted that many of the genes are involved in multiple pathways that are significant within this analysis. Additionally, the same gene may be reported multiple times within a pathway due to gene splicing and expressed as differences in the gene isoforms, which may be verified with the Affymetric probe ID.

The 12 pathways of significance are involved in amino acid metabolism, metabolism of cofactors and vitamins, biosynthesis of secondary metabolites, folding, sorting, and degradation of proteins, translation, and DNA replication and repair. Within these pathways, the global gene analysis found significant differences in gene expression between the cases and controls; however, each of these genes has a high false discovery

Table 4.5 12 Affymetrix-Defined Pathways of Significance.

Defined Pathway	Test Stat	Perm. p.	False Discovery (q)
Methionine Metabolism	50.67365	0.0015	0.3495
Selenoamino Acid Metabolism	55.48675	0.0051	0.5736
Phenylalanine, Tyrosine and Tryptophan biosynthesis	39.28965	0.014	0.5736
Alkaloid Biosynthesis	24.16919	0.014	0.5736
Aminoacyl-tRNA Biosynthesis	86.91439	0.0177	0.5736
Nicotinate and Nicotinamide Metabolism	70.03519	0.0263	0.5736
One Carbon Pool by Folate	58.43735	0.0314	0.5736
Tetracycline Biosynthesis	19.23147	0.0336	0.5736
Type II Secretion System	7.301847	0.0413	0.5736
Histidine Metabolism	111.7057	0.0415	0.5736
DNA Polymerase	28.78765	0.0452	0.5736
Phenylalanine Metabolism	105.2655	0.0478	0.5736

Table 4.6 Significant Genes Associated with Designated Pathways. Red denotes a higher gene expression in cases, while blue denotes higher gene expression in controls.

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MARS Methionine-tRNA synthetase Location 12q13.2	2098.546	2500.1472	0.0077	0.70	213671_s_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. The protein encoded by this gene belongs to the class I family of tRNA synthetases.						
TRDMT1 tRNA aspartic acid methyl-transferase 1 Location 10p15.1	451.7462	371.95833	0.01	0.70	206308_at	Methionine Metabolism
Gene Summary: CpG methylation is an epigenetic modification that is important for embryonic development, imprinting, and X-chromosome inactivation. This gene encodes a protein with similarity to DNA methyltransferases, but this protein does not display methyltransferase activity. The protein strongly binds DNA, suggesting that it may mark specific sequences in the genome. Alternative splicing results in multiple transcript variants encoding different isoforms.						
MARS Methionine-tRNA synthetase Location 12q13.2	202.5	274.73889	0.02	0.71	213672_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Previously described						
MARS2 methionine-tRNA synthetase 2 (mitochondrial) Location 2q33	182.4846	239.97222	0.023	0.71	243529_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. The protein encoded by this gene belongs to the class I family of tRNA synthetases						
MAT2B methionine adenosyltransferase II, beta Location 5q34-q35.1	21324.31	17708.381	0.027	0.71	217993_s_at	Methionine Metabolism Selenoamino acid Metabolism

Gene Summary: The protein encoded by this gene belongs to the methionine adenosyltransferase (MAT) family. MAT catalyzes the biosynthesis of S-adenosylmethionine from methionine and ATP. This protein is the regulatory beta subunit of MAT. Alternative splicing occurs at this locus and two transcript variants encoding distinct isoforms have been identified.

Table 4.6 *Continued.*

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MARS Methionine-tRNA synthetase Location 12q13.2	1483.392	1741.5194	0.037	0.72	201475_x_at	Methionine Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Previously described.						
MTR 5- methyltetrahydrofol ate- homocysteine methyltransferase Location 1q43	11136.008	1417.133	0.0374	0.727	203774_at	Methionine Metabolism One carbon pool by folate
Gene Summary: MTR encodes the enzyme 5-methyltetrahydrofolate-homocysteine methyltransferase. This enzyme, also known as cobalamin-dependent methionine synthase, catalyzes the final step in methionine biosynthesis. Mutations in MTR have been identified as the underlying cause of methylcobalamin deficiency complementation group G.						
DNMT3A DNA (cytosine-5-)- methyltransferase 3 alpha Location 2p23	431.8462	494.2083	0.0416	0.728	244428_at	Methionine Metabolism
Gene Summary: CpG methylation is an epigenetic modification that is important for embryonic development, imprinting, and X-chromosome inactivation. Studies in mice have demonstrated that DNA methylation is required for mammalian development. This gene encodes a DNA methyltransferase that is thought to function in de novo methylation, rather than maintenance methylation. The protein localizes to the cytoplasm and nucleus and its expression is developmentally regulated. Alternative splicing results in multiple transcript variants encoding different isoforms.						
PAPSS2 3'- phosphoadenosine 5'-phosphosulfate synthase 2 Location 10q23-q24	139.1923	91.25	0.016	0.704	203059_s_at	Selenoamino acid Metabolism
PAPSS2 3'- phosphoadenosine 5'-phosphosulfate synthase 2 Location 10q23-q24	535.6385	469.7139	0.0262	0.717	203060_s_at	Selenoamino acid Metabolism
Gene Summary: Sulfation is a common modification of endogenous (lipids, proteins, and carbohydrates) and exogenous (xenobiotics and drugs) compounds. In mammals, the sulfate source is 3'-phosphoadenosine 5'-phosphosulfate (PAPS), created from ATP and inorganic sulfate. Two different tissue isoforms encoded by different genes synthesize PAPS. This gene encodes one of the two PAPS synthetases.						

Table 4.6 *Continued.*

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MARS Methionine-tRNA synthetase Location 12q13.2	1483.392	1741.5194	0.037	0.72	201475_x_at	Selenoamino acid Metabolism
Gene Summary: Previously described						
YARS tyrosyl-tRNA synthetase Location 1p35.1	1755.892	2400.817	0.0029	0.69	212048_s_at	Phenylalanine, tyrosine tryptophan biosynthesis Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases catalyze the aminoacylation of tRNA by their cognate amino acid. Because of their central role in linking amino acids with nucleotide triplets contained in tRNAs, aminoacyl-tRNA synthetases are thought to be among the first proteins that appeared in evolution. Tyrosyl-tRNA synthetase belongs to the class I tRNA synthetase family. Cytokine activities have also been observed for the human tyrosyl-tRNA synthetase, after it is split into two parts, an N-terminal fragment that harbors the catalytic site and a C-terminal fragment found only in the mammalian enzyme. The N-terminal fragment is an interleukin-8-like cytokine, whereas the released C-terminal fragment is an EMAP II-like cytokine.						
FARS2 phenylalanine- tRNA synthetase 2 (mitochondrial) Location 6p25.1	569.8	697.2167	0.0197	0.70	204282_s_at	Phenylalanine, tyrosine tryptophan biosynthesis Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. This gene encodes a phenylalanine-tRNA synthetase (PheRS) localized to the mitochondrion which consists of a single polypeptide chain, unlike the (alpha-beta) ₂ structure of the prokaryotic and eukaryotic cytoplasmic forms of PheRS. Structure analysis and catalytic properties indicate mitochondrial PheRSs may constitute a class of PheRS distinct from the enzymes found in prokaryotes and in the eukaryotic cytoplasm.						
ENO2 enolase 2 (gamma, neuronal)	215.7615	296.916	0.0199	0.70	201313_at	Phenylalanine, tyrosine tryptophan biosynthesis
Gene Summary: This gene encodes one of the three enolase isoenzymes found in mammals. This isoenzyme, a homodimer, is found in mature neurons and cells of neuronal origin. A switch from alpha enolase to gamma enolase occurs in neural tissue during development in rats and primates.						
DDC dopa decarboxylase (aromatic L-amino acid decarboxylase) Location 7p11	63.95385	28.15833	0.0067	0.69	214347_s_at	Alkaloid biosynthesis Histidine Metabolism Phenylalanine Metabolism

Table 4.6 *Continued.*

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MARS Methionine-tRNA synthetase Location 12q13.2	2098.546	2500.1472	0.0077	0.70	213671_s_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. The protein encoded by this gene belongs to the class I family of tRNA synthetases.						
TRDMT1 tRNA aspartic acid methyl-transferase 1 Location 10p15.1	451.7462	371.95833	0.01	0.70	206308_at	Methionine Metabolism
Gene Summary: CpG methylation is an epigenetic modification that is important for embryonic development, imprinting, and X-chromosome inactivation. This gene encodes a protein with similarity to DNA methyltransferases, but this protein does not display methyltransferase activity. The protein strongly binds DNA, suggesting that it may mark specific sequences in the genome. Alternative splicing results in multiple transcript variants encoding different isoforms.						
MARS Methionine-tRNA synthetase Location 12q13.2	202.5	274.73889	0.02	0.71	213672_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Previously described						
MARS2 methionine-tRNA synthetase 2 (mitochondrial) Location 2q33	182.4846	239.97222	0.023	0.71	243529_at	Methionine Metabolism Selenoamino acid Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. The protein encoded by this gene belongs to the class I family of tRNA synthetases.						
MAT2B methionine adenosyltransferase II, beta Location 5q34-q35.1	21324.31	17708.381	0.027	0.71	217993_s_at	Methionine Metabolism Selenoamino acid Metabolism
Gene Summary: The protein encoded by this gene belongs to the methionine adenosyltransferase (MAT) family. MAT catalyzes the biosynthesis of S-adenosylmethionine from methionine and ATP. This protein is the regulatory beta subunit of MAT. Alternative splicing occurs at this locus and two transcript variants encoding distinct isoforms have been identified.						

Table 4.6 Continued.

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MARS Methionine-tRNA synthetase Location 12q13.2	1483.392	1741.5194	0.037	0.72	201475_x_at	Methionine Metabolism Aminoacyl-tRNA biosynthesis
Gene Summary: Previously described						
MTR 5- methyltetrahydrofol ate- homocysteine methyltransferase Location 1q43	11136.008	1417.133	0.0374	0.727	203774_at	Methionine Metabolism One carbon pool by folate
Gene Summary: MTR encodes the enzyme 5-methyltetrahydrofolate-homocysteine methyltransferase. This enzyme, also known as cobalamin-dependent methionine synthase, catalyzes the final step in methionine biosynthesis. Mutations in MTR have been identified as the underlying cause of methylcobalamin deficiency complementation group G.						
DNMT3A DNA (cytosine-5-)- methyltransferase 3 alpha Location 2p23	431.8462	494.2083	0.0416	0.728	244428_at	Methionine Metabolism
Gene Summary: CpG methylation is an epigenetic modification that is important for embryonic development, imprinting, and X-chromosome inactivation. Studies in mice have demonstrated that DNA methylation is required for mammalian development. This gene encodes a DNA methyltransferase that is thought to function in de novo methylation, rather than maintenance methylation. The protein localizes to the cytoplasm and nucleus and its expression is developmentally regulated. Alternative splicing results in multiple transcript variants encoding different isoforms.						
PAPSS2 3'- phosphoadenosine 5'-phosphosulfate synthase 2 Location 10q23-q24	139.1923	91.25	0.0168	0.704	203059_s_at	Selenoamino acid Metabolism
PAPSS2 3'- phosphoadenosine 5'-phosphosulfate synthase 2 Location 10q23-q24	535.6385	469.7139	0.0262	0.717	203060_s_at	Selenoamino acid Metabolism
Gene Summary: Sulfation is a common modification of endogenous (lipids, proteins, and carbohydrates) and exogenous (xenobiotics and drugs) compounds. In mammals, the sulfate source is 3'-phosphoadenosine 5'-phosphosulfate (PAPS), created from ATP and inorganic sulfate. Two different tissue isoforms encoded by different genes synthesize PAPS. This gene encodes one of the two PAPS synthetases.						

Table 4.6 *Continued.*

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
TYR tyrosinase (oculocutaneous albinism IA) Location 11q14- q21	159.9538	109.6667	0.0133	0.70	206630_at	Alkaloid biosynthesis
AARS alanyl-tRNA synthetase Location 16q22	1105.031	1357.4944	0.0039	0.69	201000_at	Aminoacyl-tRNA biosynthesis
Gene Summary: tRNA synthetases are the enzymes that interpret the RNA code and attach specific aminoacids to the tRNAs that contain the cognate trinucleotide anticodons. They consist of a catalytic udomain which interacts with the amino acid acceptor-T psi C helix of the tRNA, and a second domain which interacts with the rest of the tRNA structure.						
LARS l leucyl-tRNA synthetase Location 5q32	2189.269	2854.6	0.0084	0.70	217810_x_at	Aminoacyl-tRNA biosynthesis
Gene Summary: This gene encodes a cytosolic leucine-tRNA synthetase, a member of the class I aminoacyl-tRNA synthetase family. The encoded enzyme catalyzes the ATP-dependent ligation of L-leucine to tRNA(Leu). It is found in the cytoplasm as part of a multisynthetase complex and interacts with the arginine tRNA synthetase through its C-terminal domain.						
LARS2 leucyl-tRNA synthetase 2, mitochondrial Location 3p21.3	260.8538	297.0306	0.0492	0.73	34764_at	Aminoacyl-tRNA biosynthesis
Gene Summary: This gene encodes a class 1 aminoacyl-tRNA synthetase, mitochondrial leucyl-tRNA synthetase. Each of the twenty aminoacyl-tRNA synthetases catalyzes the aminoacylation of a specific tRNA or tRNA isoaccepting family with the cognate amino acid.						
HARS histidyl-tRNA synthetase Location 5q31.	1129.131	1352.417	0.0554	0.74	202042_at	Aminoacyl-tRNA biosynthesis Histidine Metabolism
Gene Summary: Aminoacyl-tRNA synthetases are a class of enzymes that charge tRNAs with their cognate amino acids. The protein encoded by this gene is a cytoplasmic enzyme which belongs to the class II family of aminoacyl-tRNA synthetases. The enzyme is responsible for the synthesis of histidyl-transfer RNA, which is essential for the incorporation of histidine into proteins.						

Table 4.6 Continued.

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
NMNAT2 nicotinamide nucleotide adenylyltransferase 2 Location 1q25	155.7385	97.59167	0.001	0.67	1552712_a_at	Nicotinate and nicotinamide metabolism
Gene Summary: This gene product belongs to the nicotinamide mononucleotide adenylyltransferase (NMNAT) enzyme family, members of which catalyze an essential step in NAD (NADP) biosynthetic pathway.						
NT5C3 5'- nucleotidase, cytosolic III Location 7p14.3	9632.208	6214.1278	0.032	0.72	223298_s_at	Nicotinate and nicotinamide metabolism
Gene Summary: Pyrimidine 5-prime-nucleotidase (P5N; EC 3.1.3.5), also called uridine 5-prime monophosphate hydrolase (UMPH), catalyzes the dephosphorylation of the pyrimidine 5-prime monophosphates UMP and CMP to the corresponding nucleosides						
PBEF1 pre-B-cell colony enhancing factor 1 pre-B cell enhancing factor 1 pseudogene Location 7q22.2 /// 10p11.21	42190.62	34318.536	0.043	0.7	217739_s_at	Nicotinate and nicotinamide metabolism
Gene Summary: This gene encodes a protein that catalyzes the condensation of nicotinamide with 5-phosphoribosyl-1-pyrophosphate to yield nicotinamide mononucleotide, one step in the biosynthesis of nicotinamide adenine dinucleotide. The protein is an adipokine that is localized to the bloodstream and has various functions, including the promotion of vascular smooth muscle cell maturation and inhibition of neutrophil apoptosis. It also activates insulin receptor and has insulin-mimetic effects, lowering blood glucose and improving insulin sensitivity. The protein is highly expressed in visceral fat and serum levels of the protein correlate with obesity. This gene has a pseudogene on chromosome 10.						
NMNAT2 nicotinamide nucleotide adenylyltransferase 2 Location 1q25	41.99231	20.952778	0.046	0.73	209755_at	Nicotinate and nicotinamide metabolism
Gene Summary: Previously described						
NT5E 5'- nucleotidase, ecto (CD73) Location 6q14-q21	67.71538	99.030556	0.049	0.73	227486_at	Nicotinate and nicotinamide metabolism
Gene Summary: Ecto-5-prime-nucleotidase (5-prime-ribonucleotide phosphohydrolase; EC 3.1.3.5) catalyzes the conversion at neutral pH of purine 5-prime mononucleotides to nucleosides, the preferred substrate being AMP. The enzyme consists of a dimer of 2 identical 70-kD subunits bound by a glycosyl phosphatidyl inositol linkage to the external face of the plasma membrane. The enzyme is used as a marker of lymphocyte differentiation. Consequently, a deficiency of NT5 occurs in a variety of immunodeficiency diseases						

Table 4.6 *Continued.*

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
TYMS Thymidylate synthetase Location 18p11.32	128.5154	187.9	0.006	0.69	243016_at	One carbon pool by folate
Gene Summary: Thymidylate synthase (TYMS, or TS; EC 2.1.1.45) uses the 5,10-methylenetetrahydrofolate (methylene-THF) as a cofactor to maintain the dTMP (thymidine-5-prime monophosphate) pool critical for DNA replication and repair. The enzyme has been of interest as a target for cancer chemotherapeutic agents. It is considered to be the primary site of action for 5-fluorouracil, 5-fluoro-2-prime-deoxyuridine, and some folate analogs.						
DHFR///LOC64350 9 dihydrofolate reductase /// similar to Dihydrofolate reductase Location 5q11.2-q13.2 /// 18q11.2	240.7385	311.76389	0.008	0.70	48808_at	One carbon pool by folate
Gene Summary: Dihydrofolate reductase converts dihydrofolate into tetrahydrofolate, a methyl group shuttle required for the de novo synthesis of purines, thymidylic acid, and certain amino acids. While the functional dihydrofolate reductase gene has been mapped to chromosome 5, multiple intronless processed pseudogenes or dihydrofolate reductase-like genes have been identified on separate chromosomes						
SHMT1 serine hydroxymethyltransferase 1 (soluble) Location 17p11.2	119.7846	80.466667	0.338	0.72	217304_at	One carbon pool by folate
Gene Summary: This gene encodes the cellular form of serine hydroxymethyltransferase, a pyridoxal phosphate-containing enzyme that catalyzes the reversible conversion of serine and tetrahydrofolate to glycine and 5,10-methylene tetrahydrofolate. This reaction provides one carbon units for synthesis of methionine, thymidylate, and purines in the cytoplasm.						
TYMS thymidylate synthetase Location 18p11.32	31.77692	21.111111	0.044	0.73	217684_at	One carbon pool by folate
Gene Summary: Previously described						
ACACA acetyl-Coenzyme A carboxylase alpha Location 17q21	198.8846	352.24167	0.0008	0.67	212186_at	Tetracycline biosynthesis
Gene Summary: Acetyl-CoA carboxylase (ACC) is a complex multifunctional enzyme system. ACC is a biotin-containing enzyme which catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the rate-limiting step in fatty acid synthesis. There are two ACC forms, alpha and beta, encoded by two different genes. ACC-alpha is highly enriched in lipogenic tissues. The enzyme is under long term control at the transcriptional and translational levels and under short term regulation by the phosphorylation/dephosphorylation of targeted serine residues and by allosteric transformation by citrate or palmitoyl-CoA.						

Table 4.6 Continued.

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
ACACB acetyl-Coenzyme A carboxylase beta Location 12q24.11	212.2308	294.34722	0.033	0.72	49452_at	Tetracycline biosynthesis
Gene Summary: Acetyl-CoA carboxylase (ACC) is a complex multifunctional enzyme system. ACC is a biotin-containing enzyme which catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the rate-limiting step in fatty acid synthesis. ACC-beta is thought to control fatty acid oxidation by means of the ability of malonyl-CoA to inhibit carnitine-palmitoyl-CoA transferase I, the rate-limiting step in fatty acid uptake and oxidation by mitochondria. ACC-beta may be involved in the regulation of fatty acid oxidation, rather than fatty acid biosynthesis.						
DCTD dCMP deaminase Location 4q35.1	785.7692	943.88889	0.011	0.70	201572_x_at	Type II section system
DCTD	960.6692	1117.6056	0.105	0.78	210137_s_at	Type II section system
Gene Summary: The protein encoded by this gene catalyzes the deamination of dCMP to dUMP, the nucleotide substrate for thymidylate synthase. The encoded protein is allosterically activated by dCTP and inhibited by dTTP, and is found as a homohexamer. This protein uses zinc as a cofactor for its activity.						
ABP1	61.82308	27.313889	0.005	0.69	242452_at	Histidine Metabolism Phenylalanine Metabolism
ABP1	76.28462	117.71667	0.007	0.70	229156_s_at	Histidine Metabolism Phenylalanine Metabolism
ABP1	940.4462	1124.4139	0.026	0.71	236401_at	Histidine Metabolism Phenylalanine Metabolism
ABP1 Amiloride binding protein 1 (amine oxidase (copper- containing) Location 7q34-q36	2142.508	2590.5944	0.033	0.72	236583_at	Histidine Metabolism Phenylalanine Metabolism
Gene Summary: This gene encodes a membrane glycoprotein that is expressed in many epithelium-rich and/or hematopoietic tissues and oxidatively deaminates putrescine and histamine. The protein may play a role in controlling the level of histamine and/or putrescine in these tissues. It also binds to and is inhibited by amiloride, a diuretic that acts by closing epithelial sodium ion channels.						

Table 4.6 Continued.

Gene Name (Gene Symbol)	Case Average	Control Average	Perm P	q	Probe-set ID	Significant Pathways
MAOA	157.7385	72.35	0.0121	0.70	A 212741_at	Histidine Metabolism Phenylalanine Metabolism
MAOA monoamine oxidase Location Xp11.3	143.4769	104.04167	0.021	0.70	234534_at	Histidine Metabolism Phenylalanine Metabolism
Gene Summary: This gene encodes monoamine oxidase A, an enzyme that degrades amine neurotransmitters, such as dopamine, norepinephrine, and serotonin. The protein localizes to the mitochondrial outer membrane. The gene is adjacent to a related gene on the opposite strand of chromosome X. Mutation in this gene results in monoamine oxidase deficiency.						
POLB Polymerase (DNA directed), beta Location 8p11.2	18.22308	35.7	0.005	0.69	234907_x_at	DNA Polymerase
Gene Summary: In eukaryotic cells, DNA polymerase beta (POLB) performs base excision repair (BER) required for DNA maintenance, replication, recombination, and drug resistance.						
POLS Polymerase (DNA directed) Sigma Location 5p15	1525.169	1910.8167	0.005	0.69	202466_at	DNA Polymerase
Gene Summary: The protein encoded by this gene is a DNA polymerase that is likely involved in DNA repair. In addition, the encoded protein may be required for sister chromatid adhesion.						
POLG Polymerase (DNA directed), gamma Location 15q25	567.2154	754.58056	0.006	0.69	203366_at	DNA Polymerase
PRDX2 Phenylalanine Metabolism peroxiredoxin 2 Location 19p13	203.8846	155.89167	0.039	0.72	201006_at	Phenylalanine Metabolism
Gene Summary: This gene encodes a member of the peroxiredoxin family of antioxidant enzymes, which reduce hydrogen peroxide and alkyl hydroperoxides. The encoded protein may play an antioxidant protective role in cells, and may contribute to the antiviral activity of CD8(+) T-cells. This protein may have a proliferative effect and play a role in cancer development or progression. The crystal structure of this protein has been resolved to 2.7 angstroms.						

rate, giving the expected proportion of false positives among the declared significant results and accounts for the testing of thousands of hypotheses simultaneously [75].

Primary Aim Two

Assess whether genotypes differ between two groups of radiated pediatric Hodgkin lymphoma survivors with and without the development of secondary breast cancer.

A total of 397,877 SNPs passed the variation requirement for detailed statistical analysis. At a p-value threshold of 0.001, the robust FDR estimate is 1, indicating that most or all SNPs are false discoveries. Due to the large number of SNP's and false discovery rate of 1, the SNP's with a chi square p value < 0.001 were transformed into a minuslog P which is 10 based $-\log$ of the chi-square p values and described in Appendix E.

To determine the genes associated with the SNP's of interest, a Geneset enrichment p-value file was created and consisted of 13, 090 genes, of which 574 had a p-value of ≤ 0.05 and false discovery rate of 1. This file gave the gene name, chromosome number, the start and end location of the gene, gamma test statics, permutation p value, and FDR Value. This file provided additional location of the gene chromosome and relationship to SNP.

To visualize the results, a plot was created and corresponds to the $-\log_{10}$ p-value of the SNP which corresponds to having a chi-square p-value < 0.001. Each of the SNP's having a $-\log_{10} p >$ than 3 ($p < 0.001$) is distinguished on the plot with a red dot and displayed for each chromosome in Figures 4.1-4.23. These figures mark clusters of SNP's and their corresponding location and become those of highest interest. The chromosomes which displayed the highest number of significant SNP differences between cases and controls were chromosomes 5, 10, and 11.

Chromosome 5 displays a significant SNP_1907286 ($-\log_{10} p = 4.8$) at position 31786818 and is associated with the PDZD2 gene although not statically significant ($p = 0.49$; $FDR \approx 1$). Chromosome 10 had the largest number of SNP's differentially expressed with a region of 6.1 kilobases containing 11 differentially expressed SNPs between the cases and controls. Two highly significant SNP's within this region are SNP_A-2181577 ($-\log_{10} p = 3.97$) and SNP_A-2006464 ($-\log_{10} p = 3.97$) at position 127481155 and 127485797 and statistically associated with the UROS gene ($p = .001$; $FDR \approx 1$). The most significant SNP within the region of chromosome 10 was SNP_A-1842005 ($-\log_{10} p = 3.98$) at position 127508528 and statistically associated with the BCCIP gene ($p = .001$; $FDR \approx 1$). Within this polymorphic region there were two other genes that were significant in the analysis, MMP21 ($p = .001$; $FDR \approx 1$) and DHX32 ($p = .001$; $FDR \approx 1$). The analysis found the most significant SNPs, SNP_A-2294376 ($-\log_{10} p = 5.1$) and SNP_A-2209464 ($-\log_{10} p = 5.2$), to be located on chromosome 11 at position 119148037 and 119149661 and associated with the genes TRIM29 ($p = 0.845$)

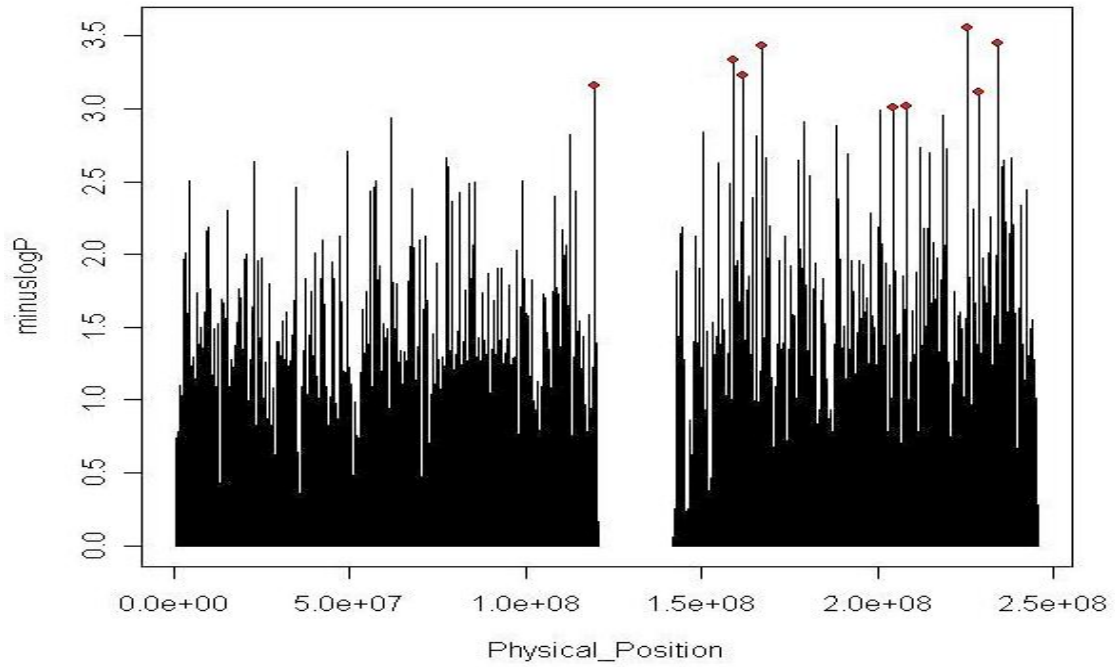


Figure 4.1 Peak p-Values for Chromosome 1.

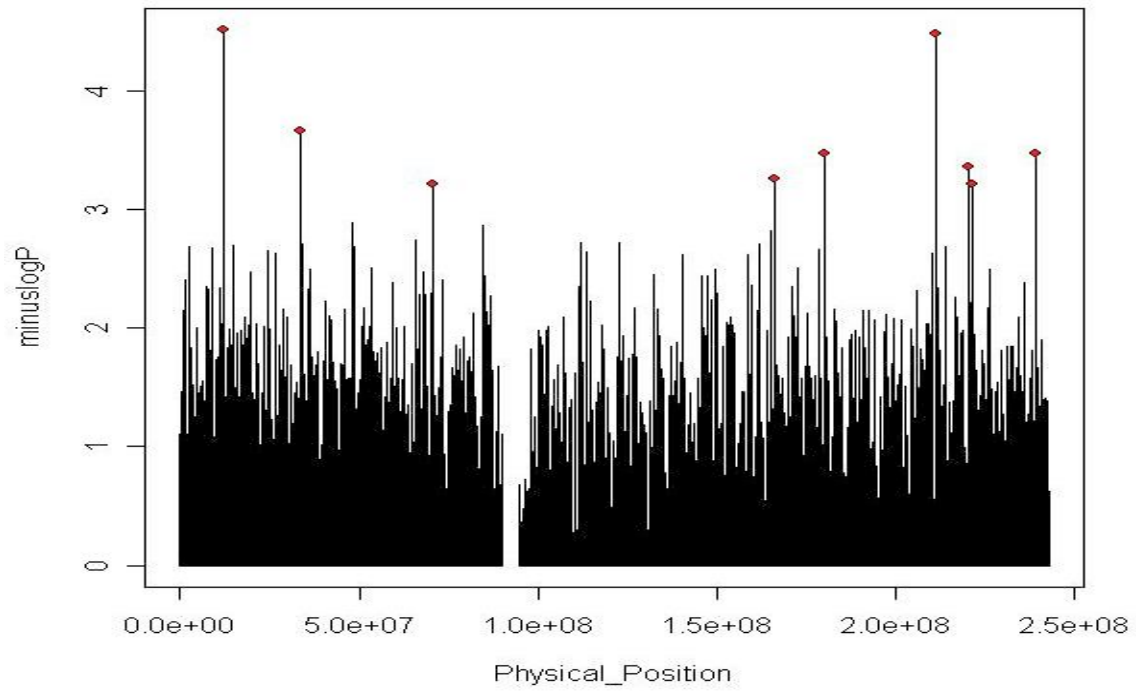


Figure 4.2 Peak p-Values for Chromosome 2.

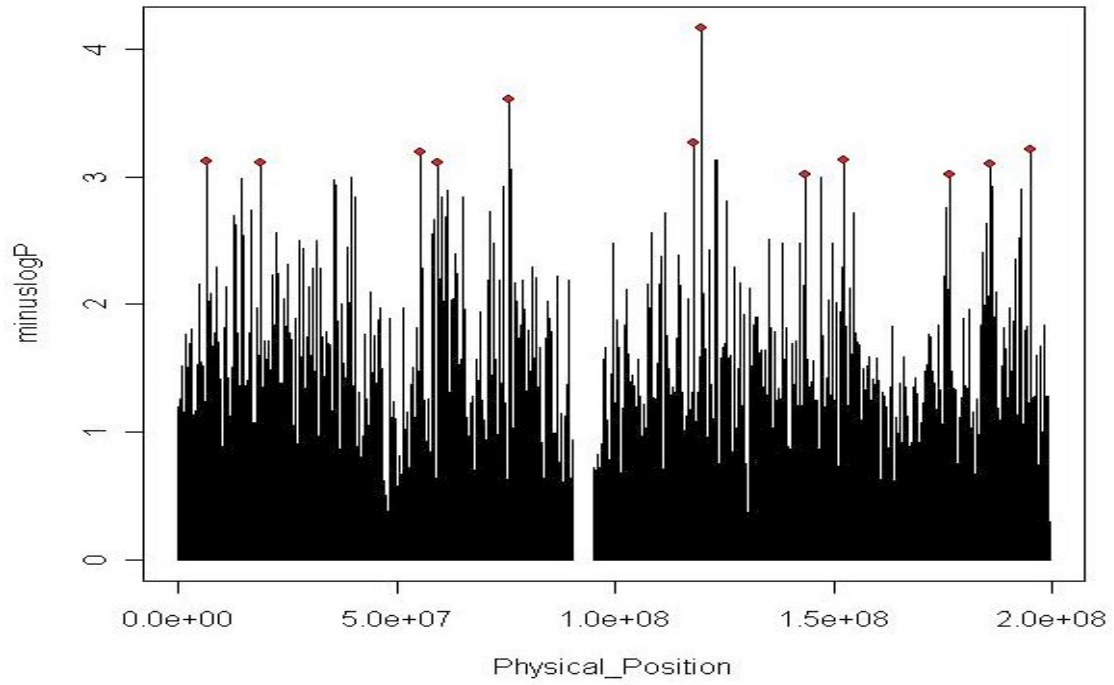


Figure 4.3 Peak p-Values for Chromosome 3.

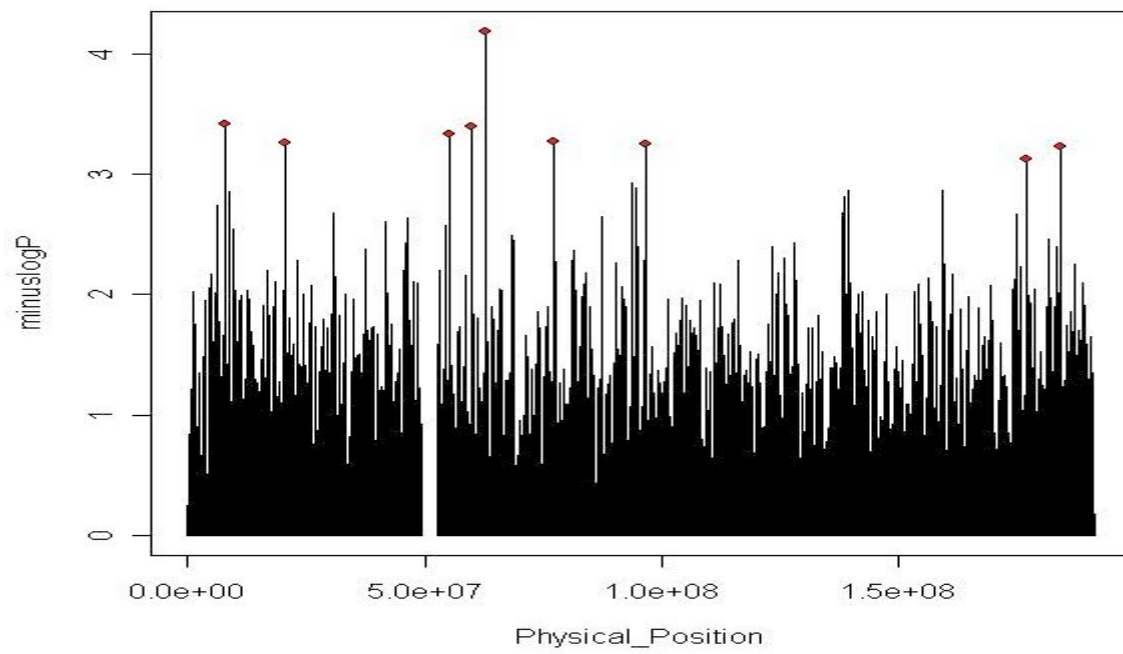


Figure 4.4 Peak p-Values for Chromosome 4.

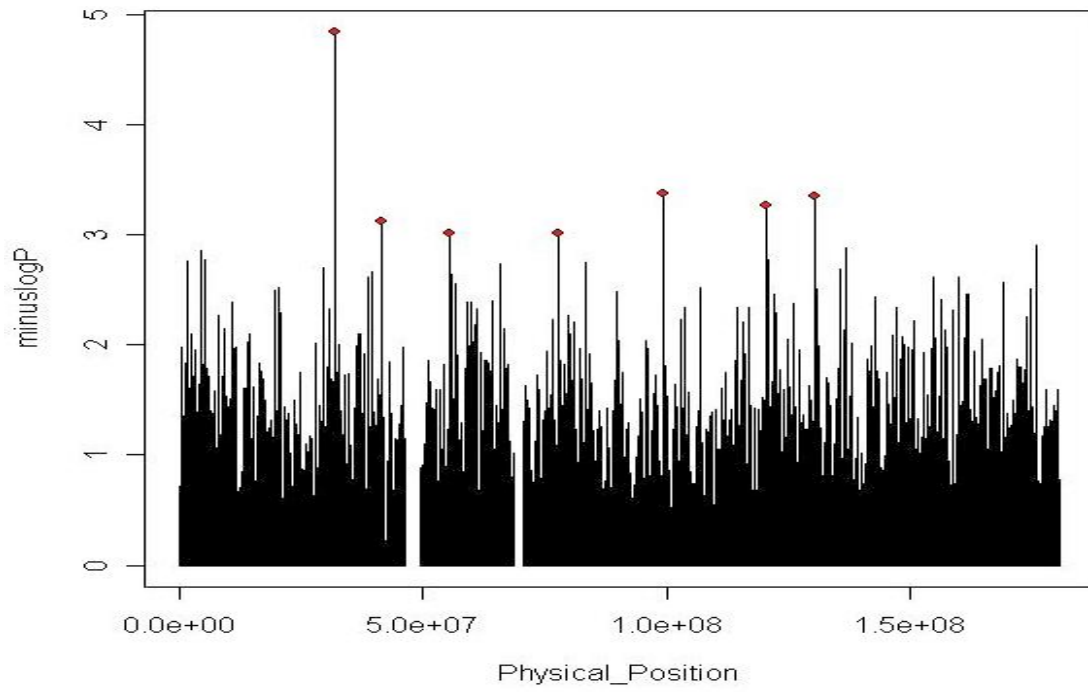


Figure 4.5 Peak p-Values for Chromosome 5.

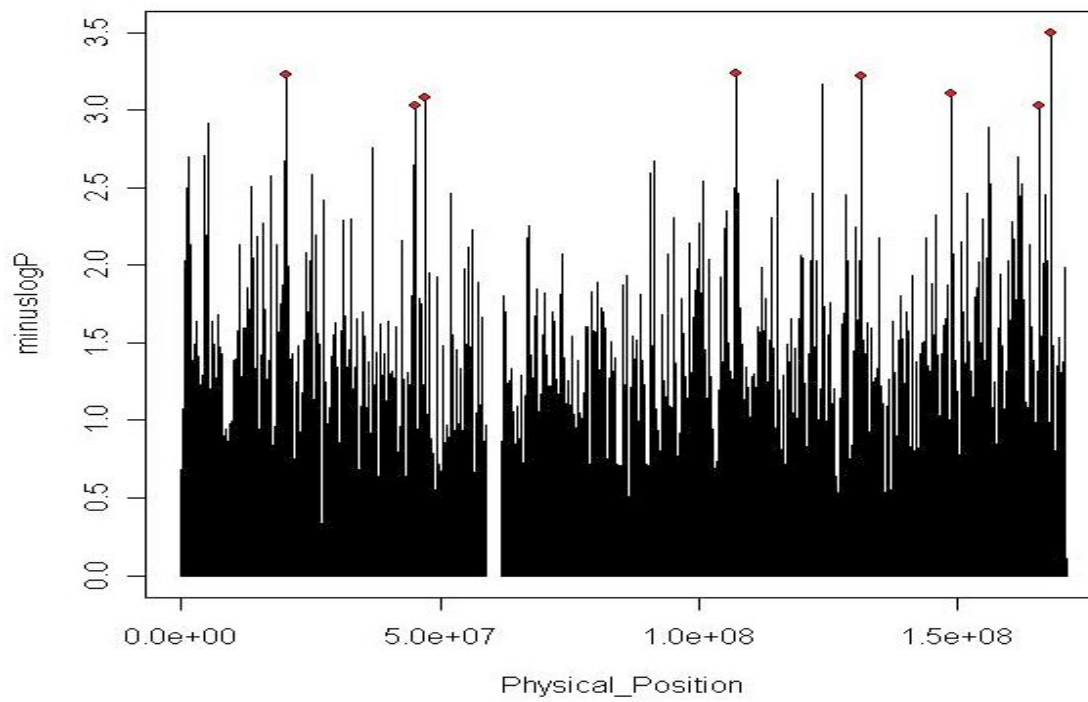


Figure 4.6 Peak p-Values for Chromosome 6.

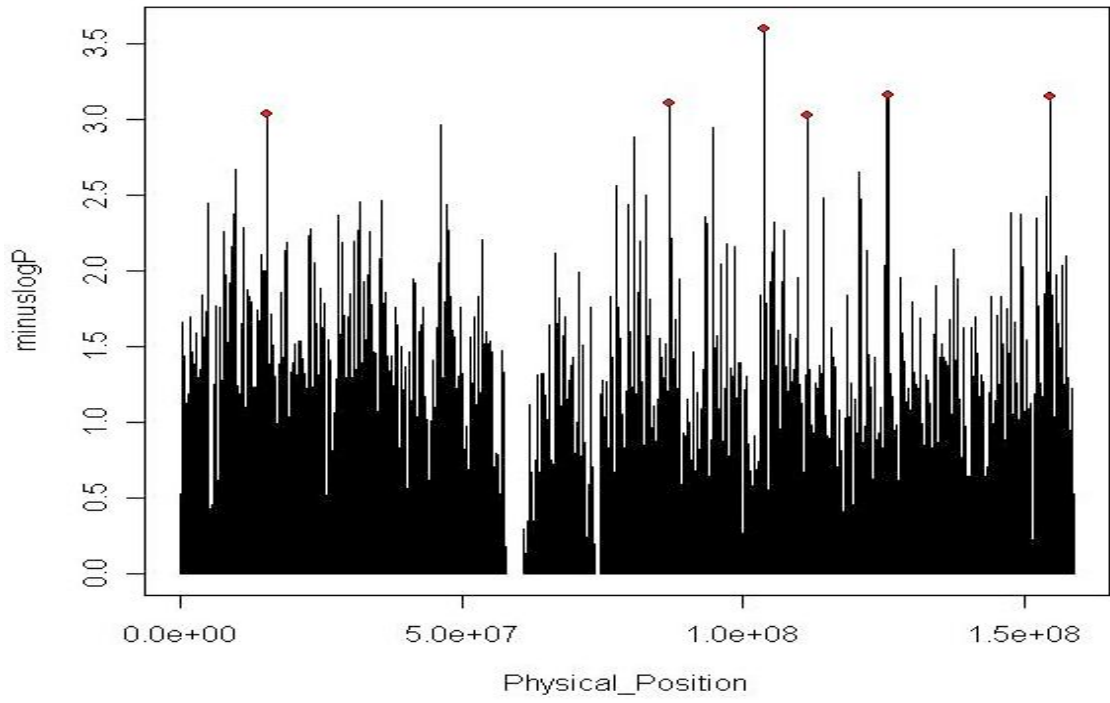


Figure 4.7 Peak p-Values for Chromosome 7.

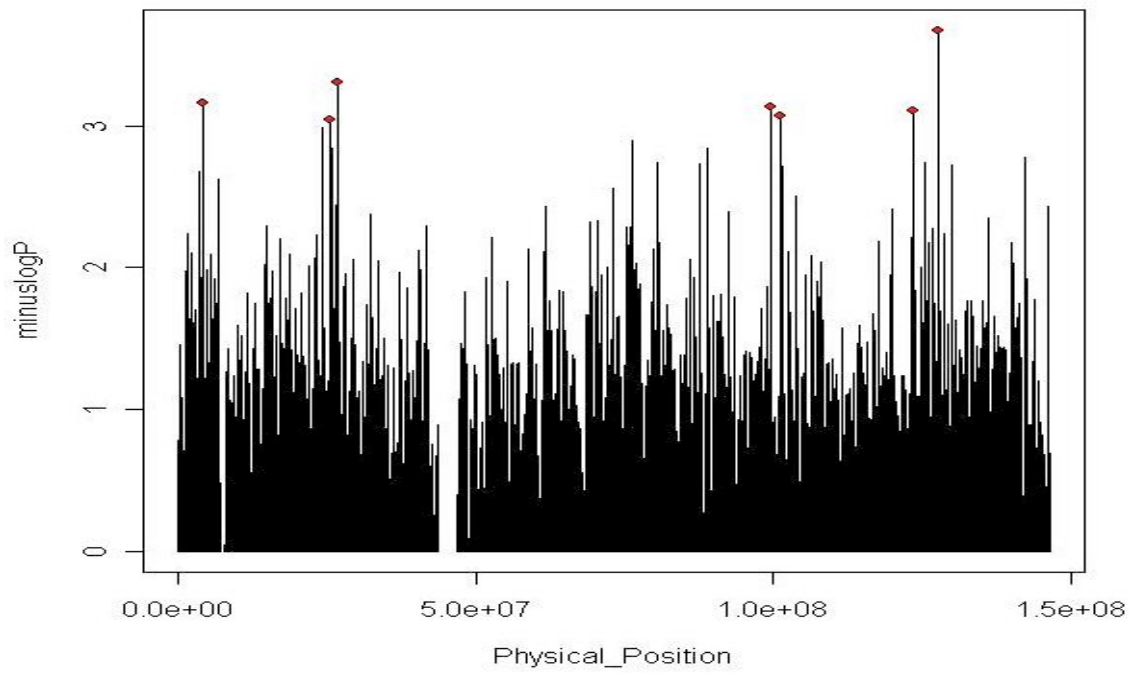


Figure 4.8 Peak p-Values for Chromosome 8.

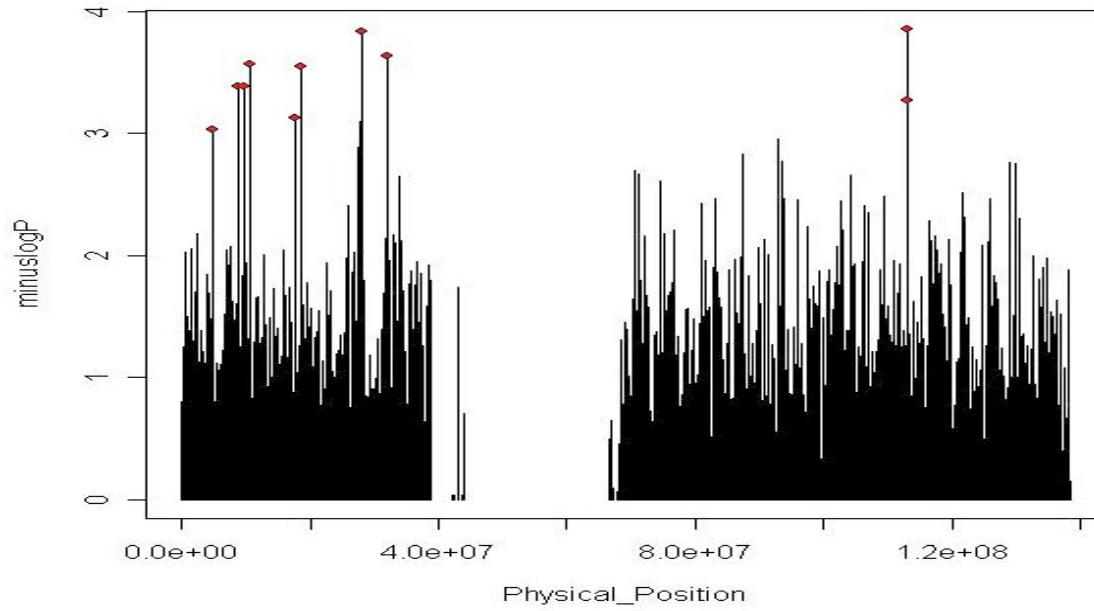


Figure 4.9 Peak p-Values for Chromosome 9.

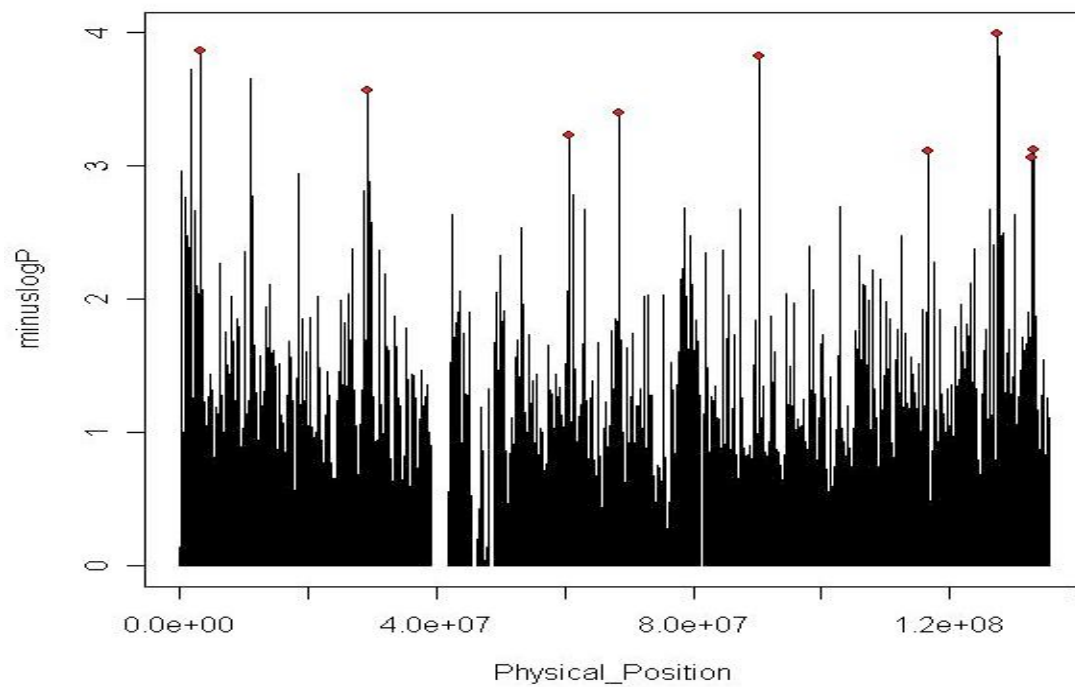


Figure 4.10 Peak p-Values for Chromosome 10.

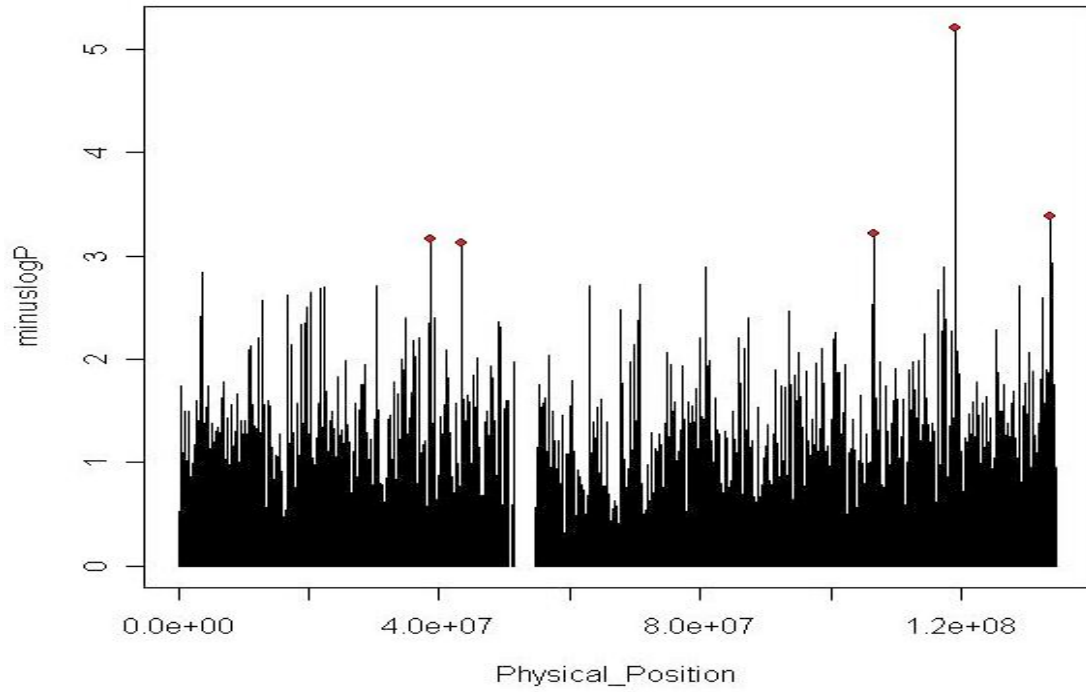


Figure 4.11 Peak p-Values for Chromosome 11.

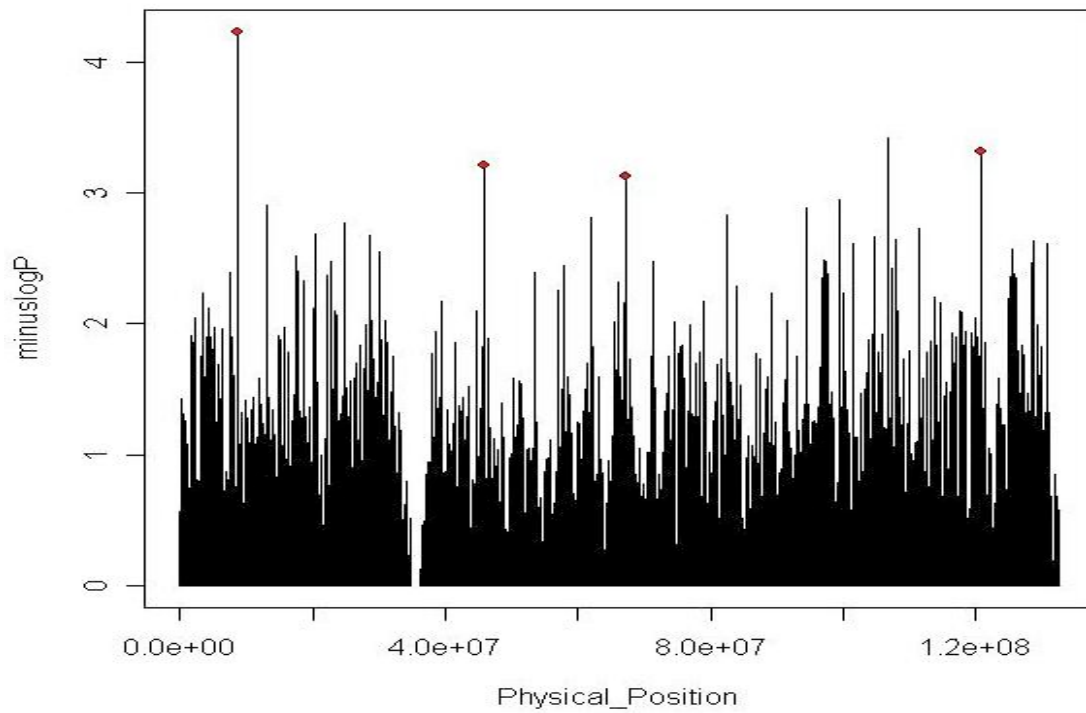


Figure 4.12 Peak p-Values for Chromosome 12

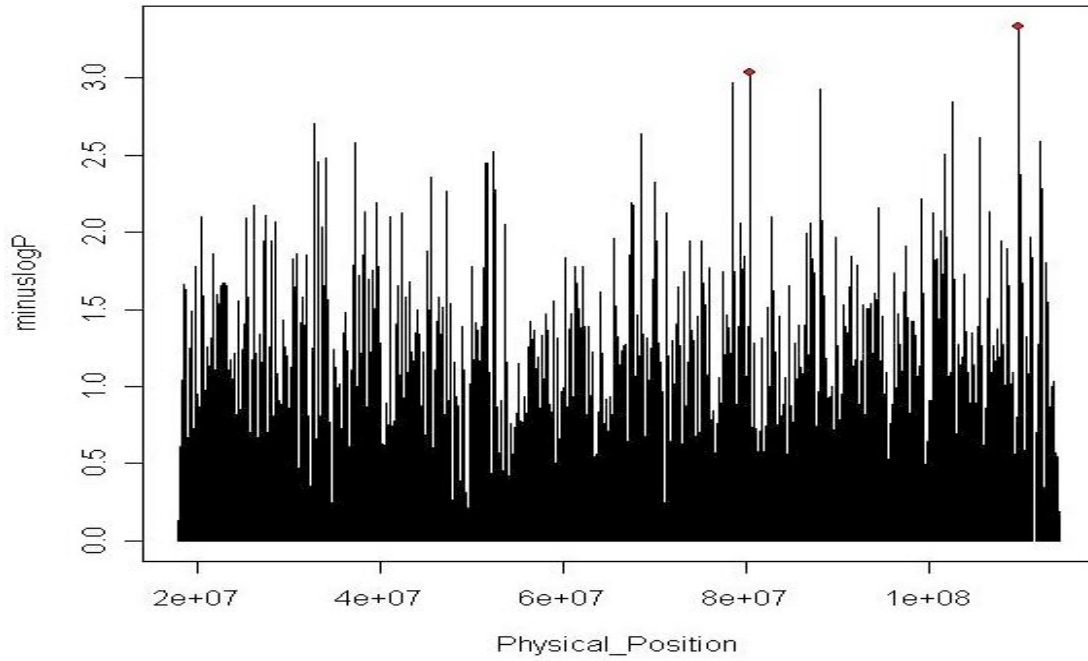


Figure 4.13 Peak p-Values for Chromosome 13.

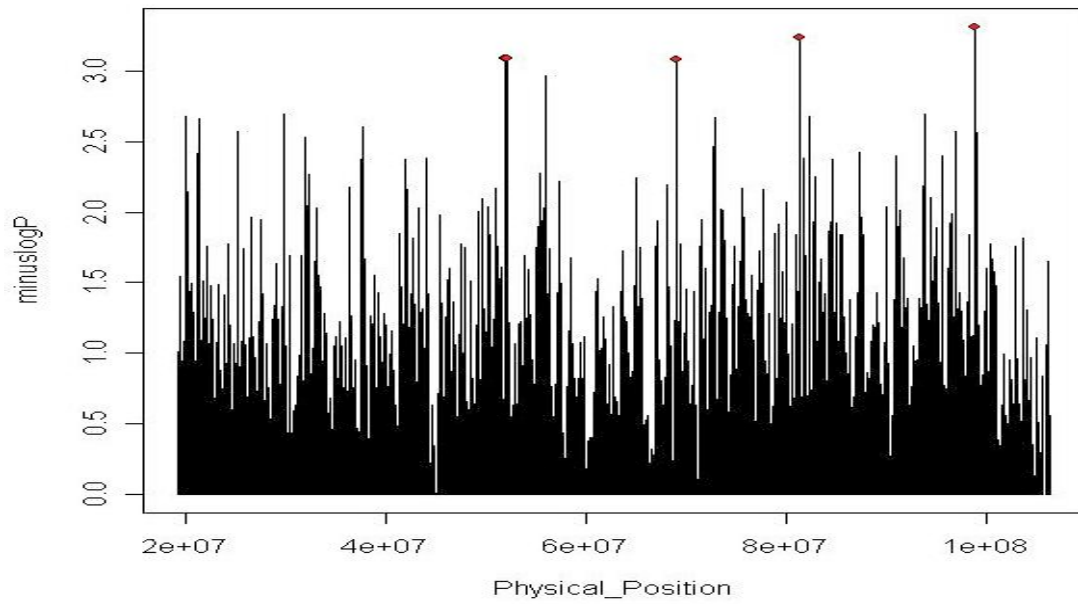


Figure 4.14 Peak p-Values for Chromosome 14.

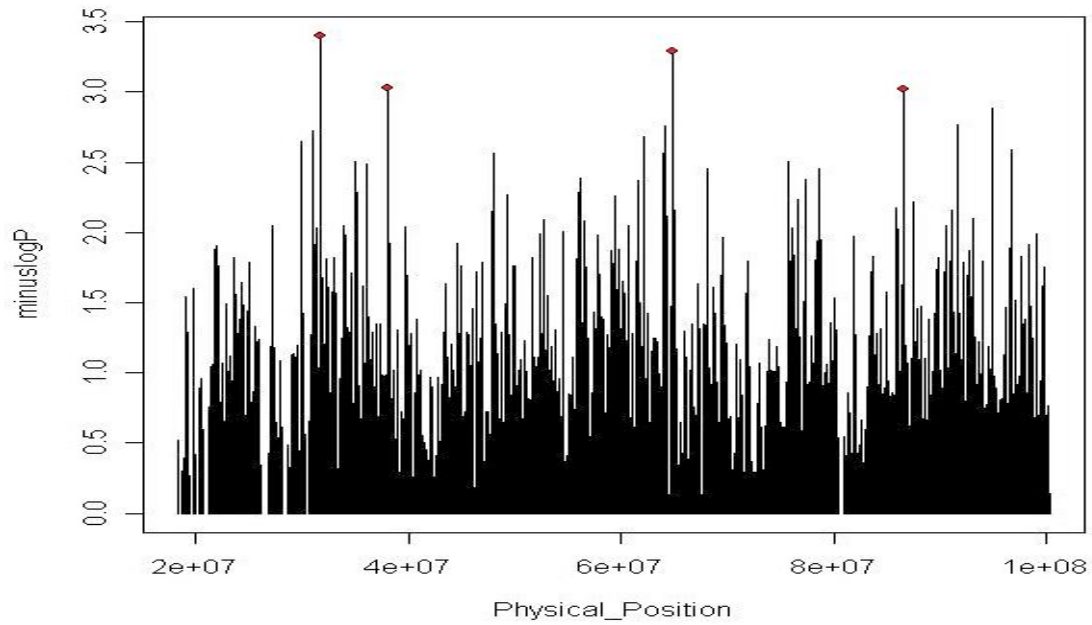


Figure 4.15 Peak p-Values for Chromosome 15.

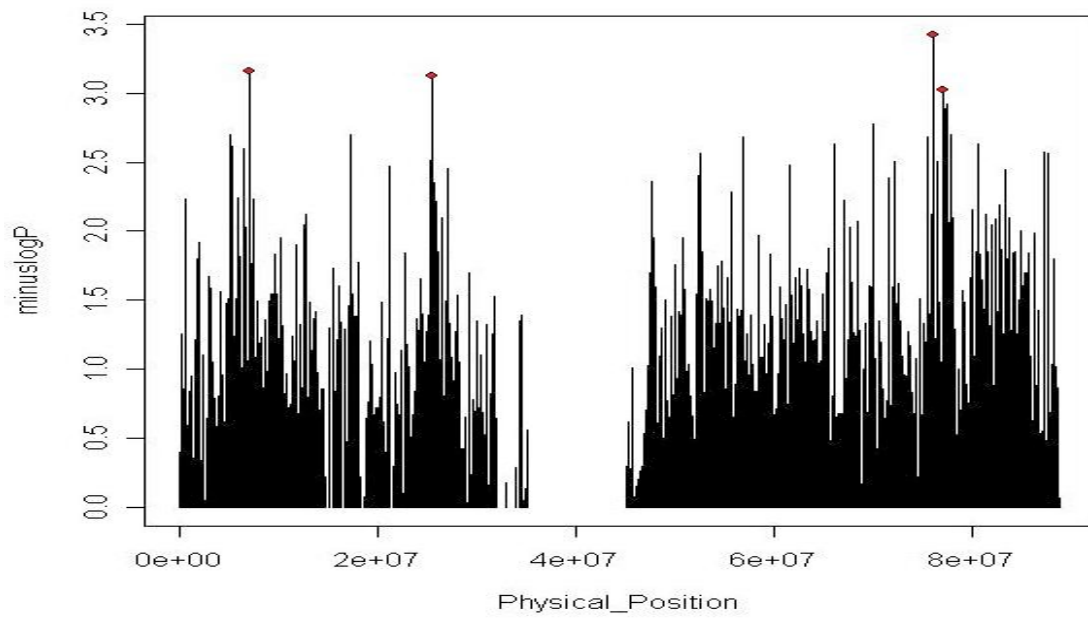


Figure 4.16 Peak p-Values for Chromosome 16.

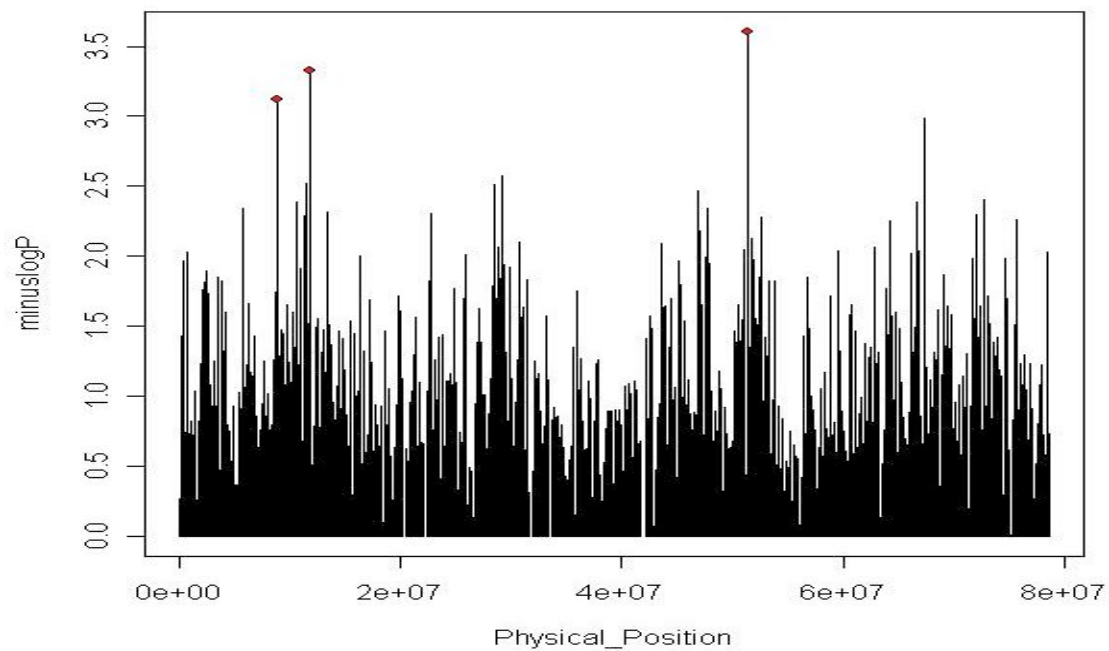


Figure 4.17 Peak p-Values for Chromosome 17.

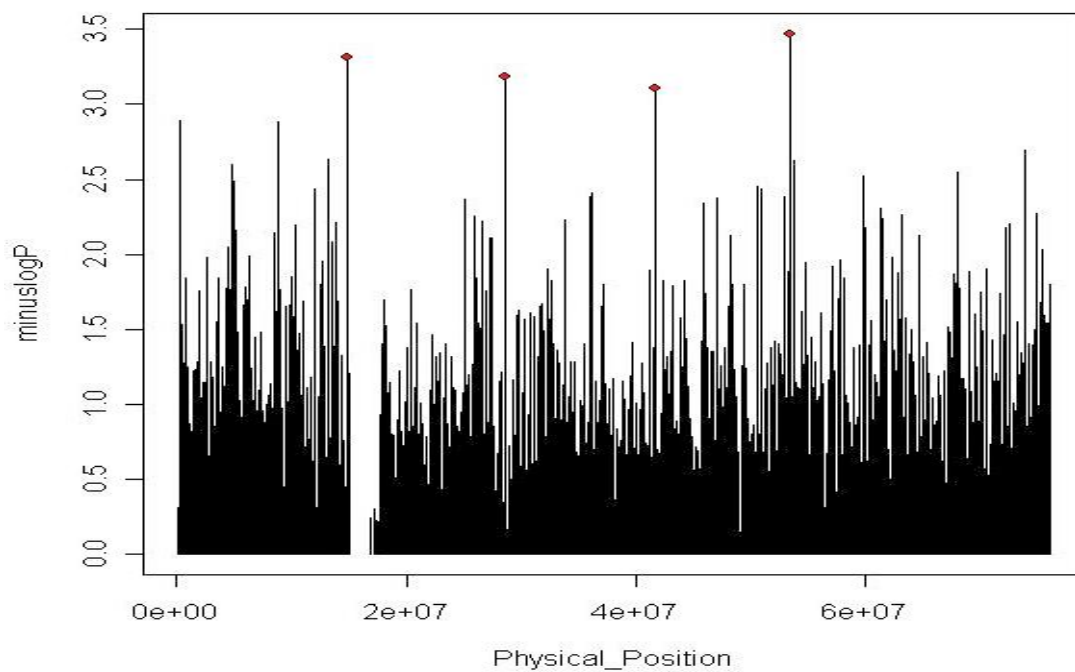


Figure 4.18 Peak p-Values for Chromosome 18.

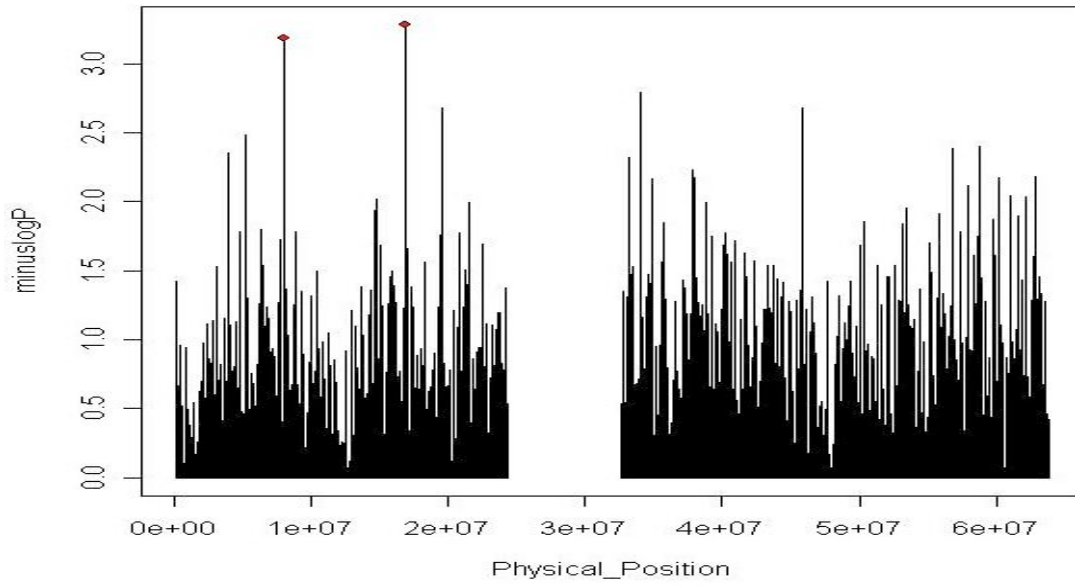


Figure 4.19 Peak p-Values for Chromosome 19.

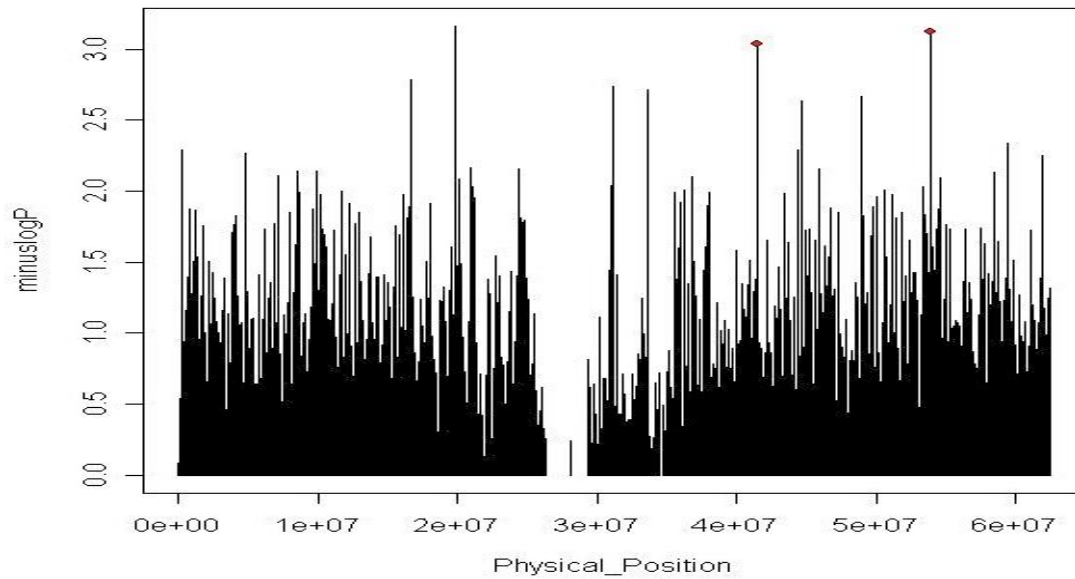


Figure 4.20 Peak p-Values for Chromosome 20.

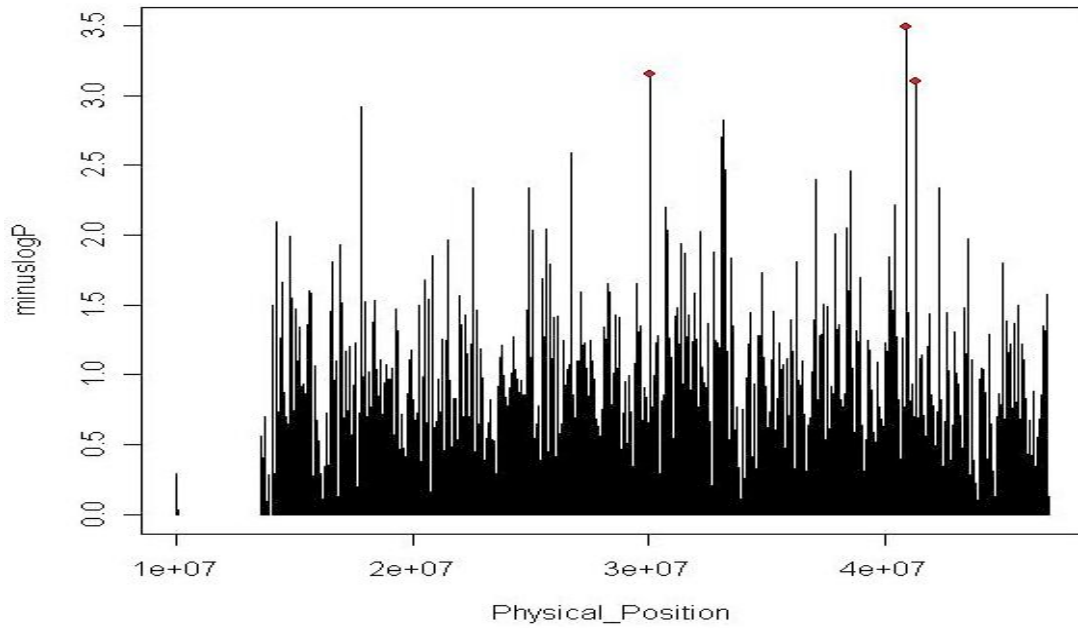


Figure 4.21 Peak p-Values for Chromosome 21.

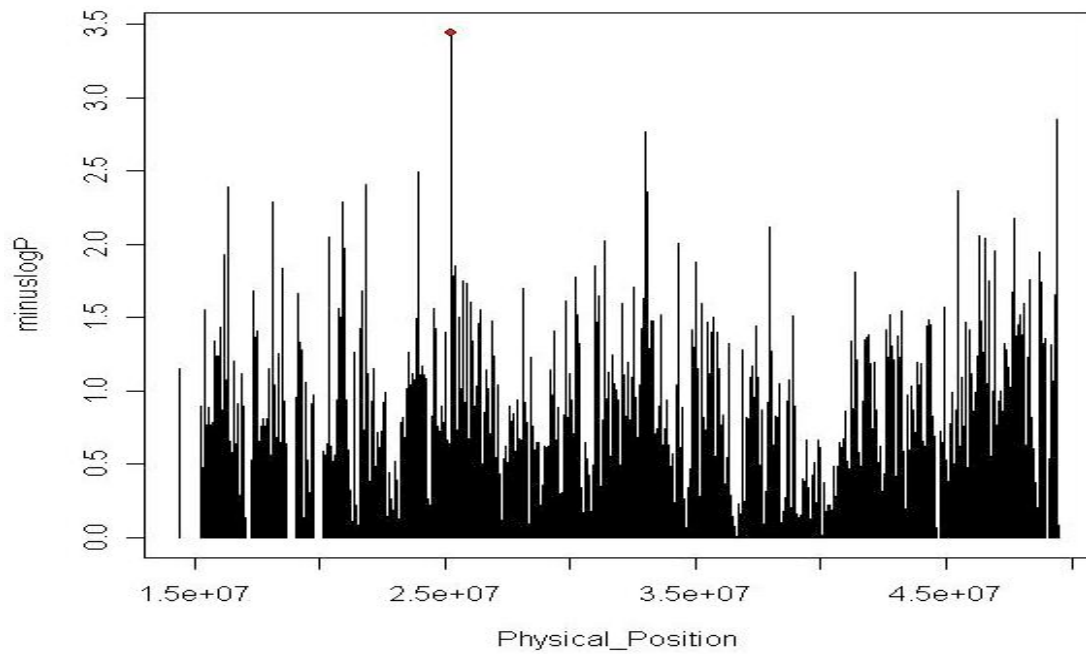


Figure 4.22 Peak p-Values for Chromosome 22.

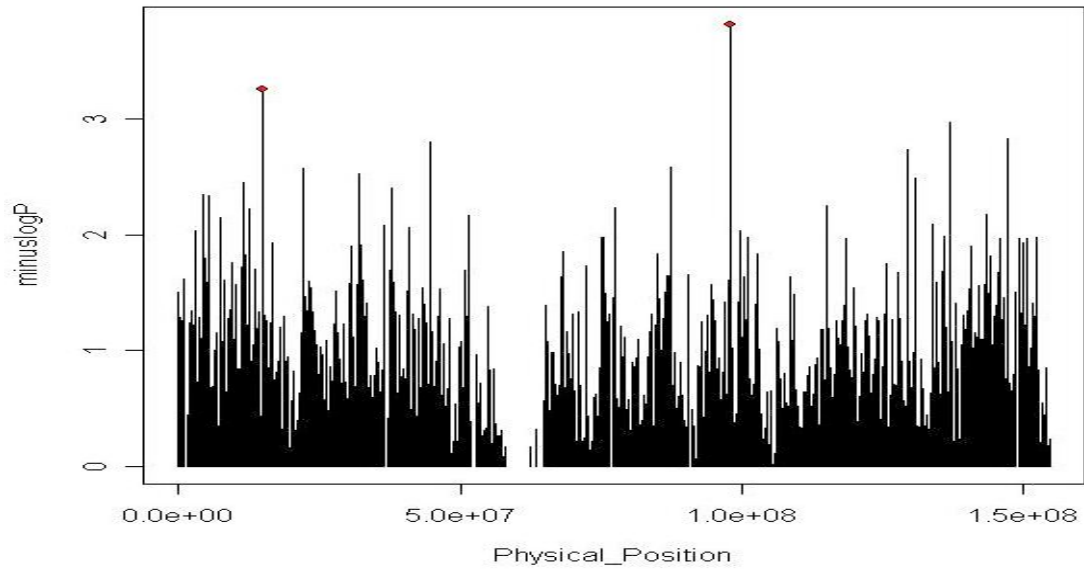


Figure 4.23 Peak p-Values for Chromosome 23. (X chromosome)

and PVRL ($p=0.142$). However, these genes were not statistically associated with the SNP in the Geneset enrichment file.

Primary Aim Three

Identify candidate copy number variants that may be associated with the risk of secondary breast cancer.

This analysis identified 130 *in silico* amplifications or deletions. Multiple cases and controls had common regions of amplification (gain) or deletion (loss) in sections of chromosomes 2, 14, and 15. *In silico* copy number variants were detected in 2q 37.3 for 5 subjects, 14q11.2 for 8 subjects, and 15q11.2 for 24 subjects.

There is some evidence that case-control status is associated with *in silico* copy number of a section of 14q11.2 and 15q11.2. Within 14q11.2, gain was observed in 3 of 13 (23%) cases, 3 of 36 controls (8%), and 1 case had a deletion ($p=0.069$). Within 15q11.2, gain was observed in 11 of 36 controls (30.5%) and no cases (0%), diploid status was inferred for 16 of 36 controls (44%) and 9 of 13 cases (69%) and loss was inferred for 4 of 13 cases (30.7%) and 9 of 36 controls (25%) ($p=0.067$). The copy number prevalence over the region of 14q11.2 is illustrated in Figure 4.24 and the correlated genes expression plots for chromosome 14 are given in Figure 4.25.

The genes associated with copy number variation 14q11.2, OR4N4 and OR4Q3 are olfactory genes known to be the largest gene family within the genome and known to regions of variability and have no known association with the study phenotype.

The copy number prevalence over the region of 15q11.2 is depicted in Figure 4.26. Chromosome 15 had 19 Affymetric gene probe sets found to be within the region of copy number variation between those with a deletion, gain, or diploid within 15q11.2. The Affymetric gene probe set and corresponding gene are described in Table 4.7. From these probe sets, 19 expression plots were generated; however, only the genes with specific interest are depicted in Figures 4.27-32 and include PAK2, CYFIP1, NIPA1, NIPA2, TUBGCP5, and BCL8.

The family history of breast cancer was then analyzed with consideration of copy number variation. The analysis found no association with gain or no change within the region; however, there was an association ($p=0.0392$) as depicted in Table 4.8 with deletion within the 15q11.2 region and having a first degree relative with breast cancer. No association was found with second degree relative and copy number deletion.

Summary of Results

Aim one described the global gene expression between the cases and controls with defined Affymetric pathways and genes that were significantly differentially

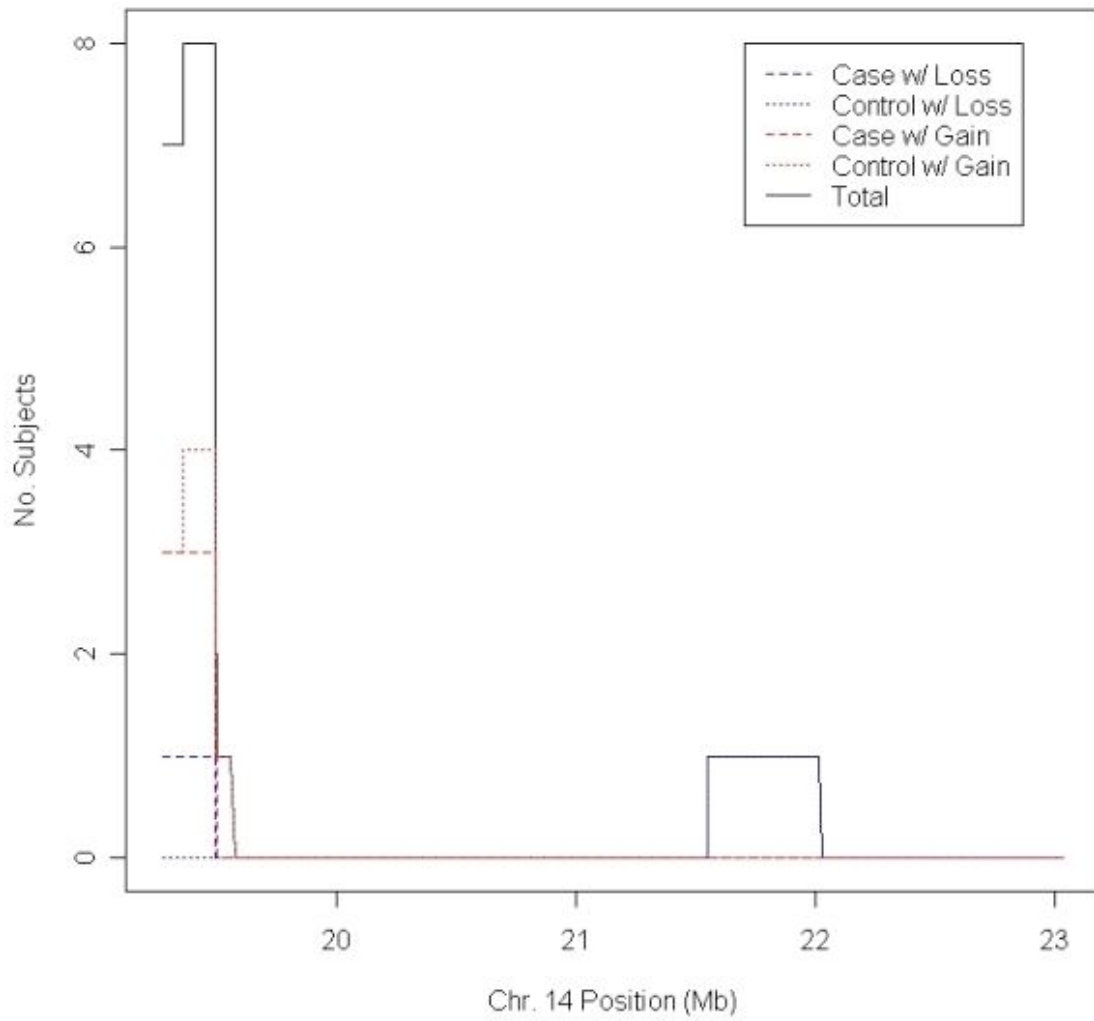


Figure 4.24 Prevalence of Inferred Copy Number Variation among Cases and Controls within 14q11.2.

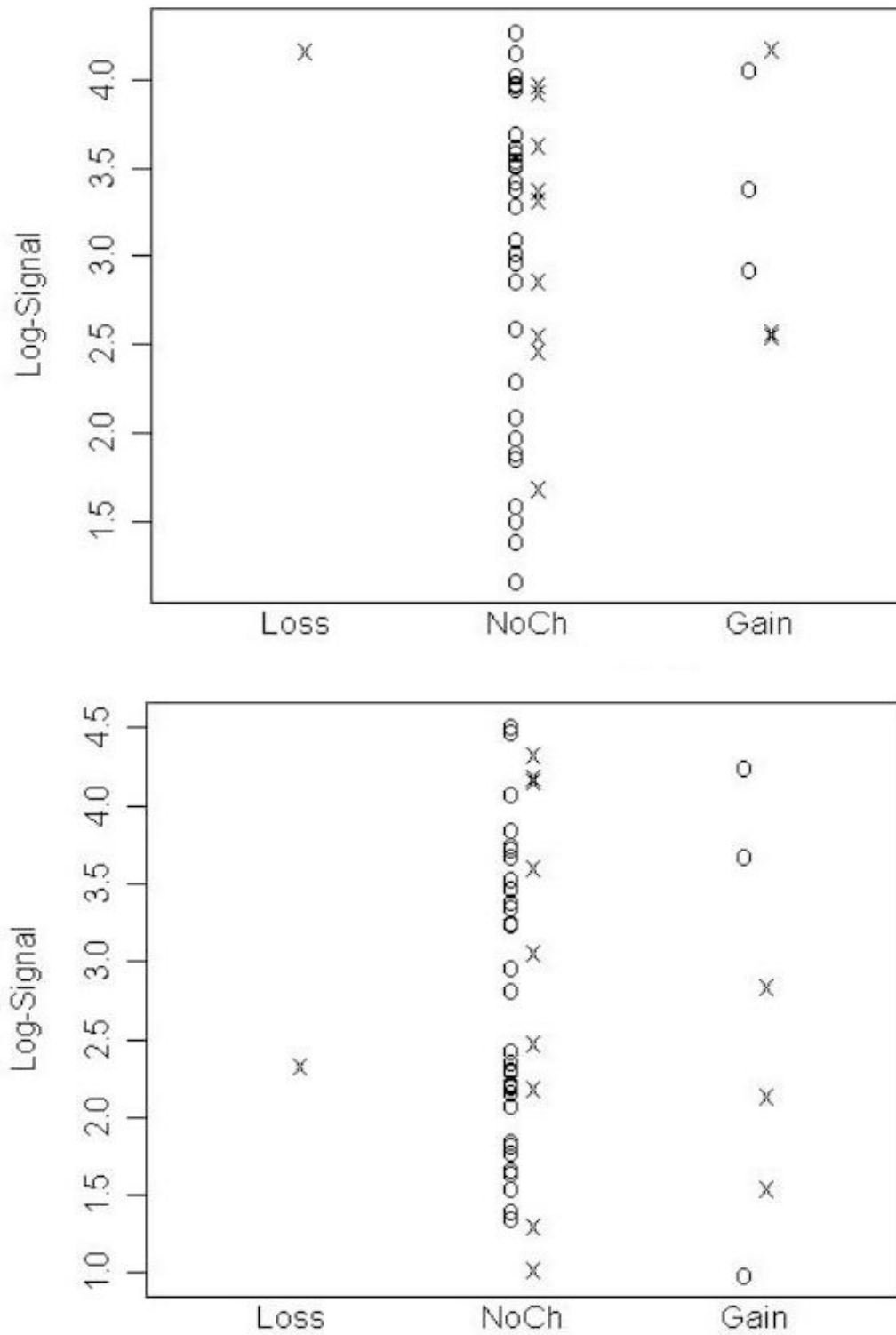


Figure 4.25 Gene Expression Correlated with Copy Number Variation at 14q11.2. (X – Case; O – Control)

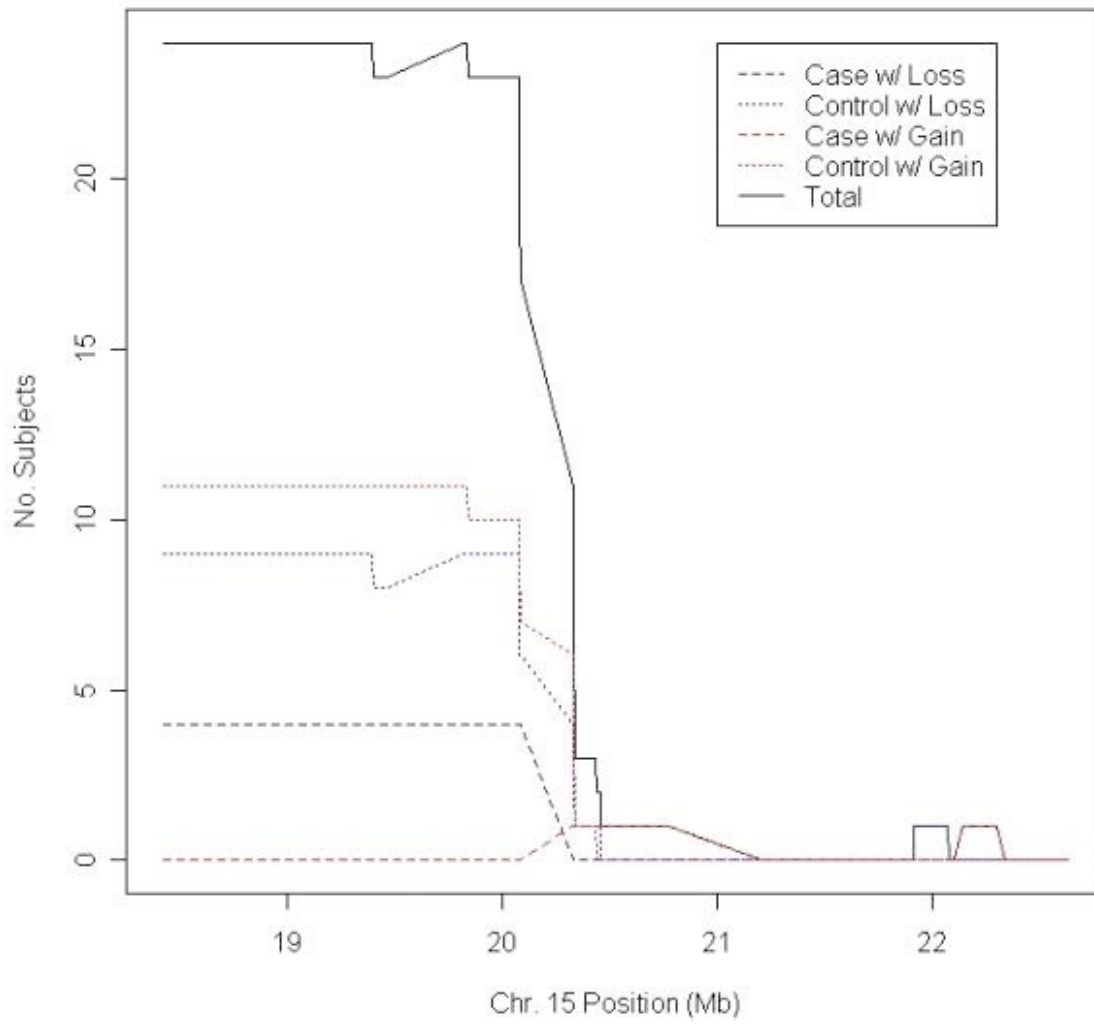


Figure 4.26 Prevalence of Inferred Copy Number Variation among Cases and Controls within 15q11.2. (X – Case; O – Control)

Table 4.7 Genes Associated with Copy Number Variation at 15q11.2.

Probe Set	Gene Symbol / Name
208923_at	CYFIP1: Cytoplasmic FMR1 Interacting protein
235083_at	FLJ38359: hypothetical protein
236283_x_at	PAK2: CDKNIA activated kinase 2
239899_at	RNF145: Ring finger protein 145
1552696_at	NIPA1: non imprinted in Prader Will Angleman Snyderome 1
1556739_at	FLJ35785: hypothetical protein LOC653125: Similar to golgi autoantigen LOC653720: Golgi Subfamily a, 8A
1560683_at	BCL8: B-cell CLL/lymphoma
1560684_x_at	BCL8: B-cell CLL/lymphoma
1564855_at	LOC727924: hypothetical protein
1564856_at	LOC 727924: hypothetical protein
1569013_at	LOC 96610: hypothetical protein
212129_at	NIPA2: non imprinted in Prader Will Angleman Snyderome 2
212133_at	NIPA2: non imprinted in Prader Will Angleman Snyderome 2
214876_s_at	TUBGCP5: Tubulin, gamma complex associated protein 5
216193_at	LOC 440366: hect domain and RLD 2 pseudogene
225752_at	NIPA1: non imprinted in Prader Will Angleman Snyderome 1
226077_at	RNF 145: Ring finger protein 145
227967_at	TUBGCP5: Tubulin, gamma complex associated protein 5
234912_at	LOC 651964: hypothetical protein

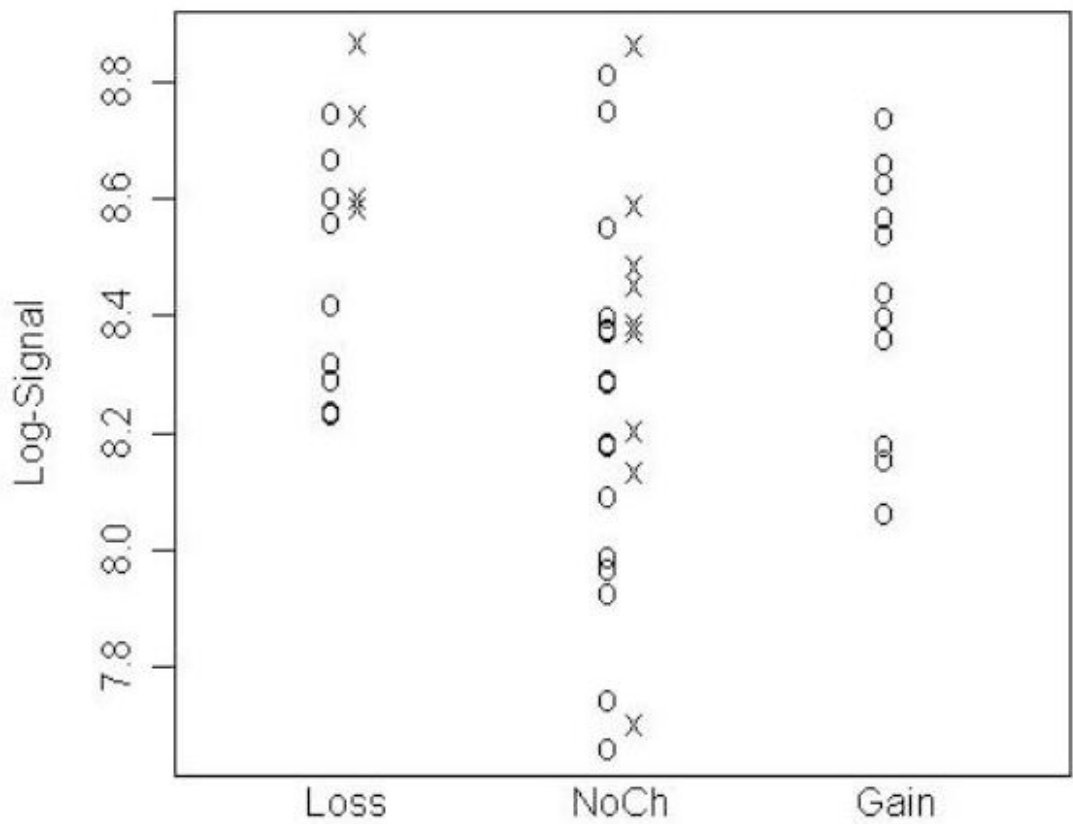


Figure 4.27 Correlation of CYFIP1 Gene Expression and Copy Number Variation. (X – Case; O – Control)

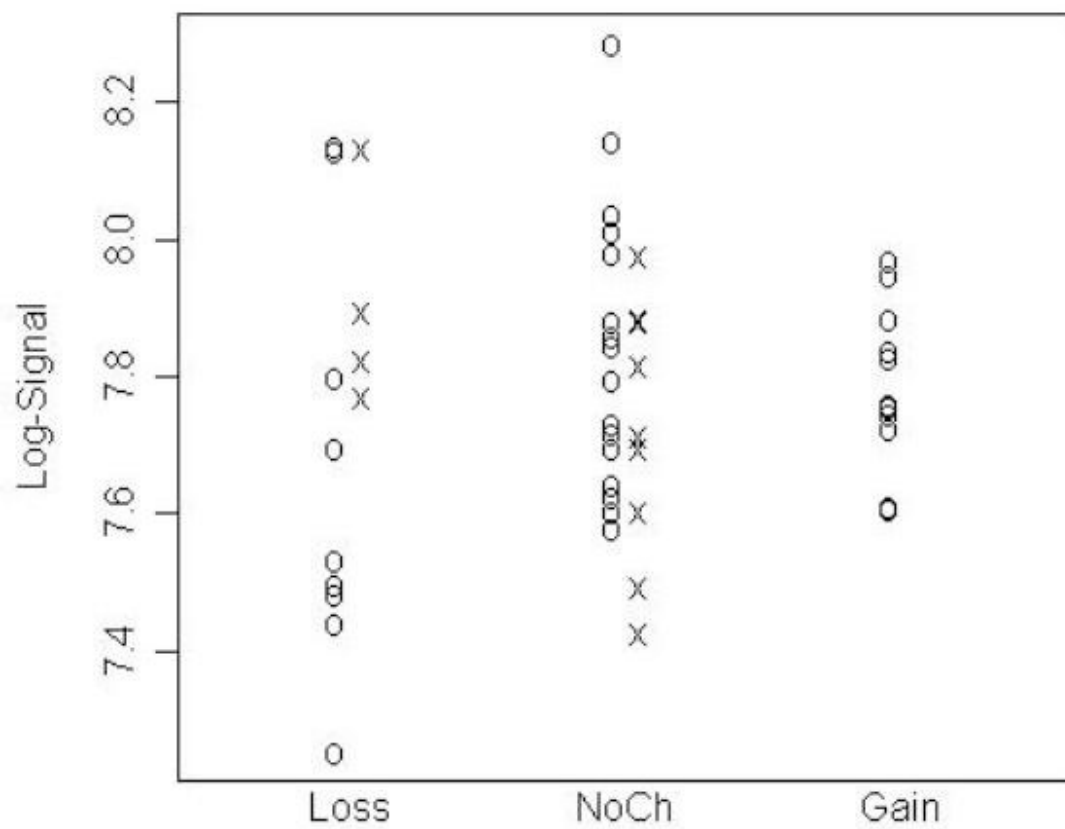


Figure 4.28 Correlation of PAK2 Gene Expression and Copy Number Variation. (X – Case; O – Control)

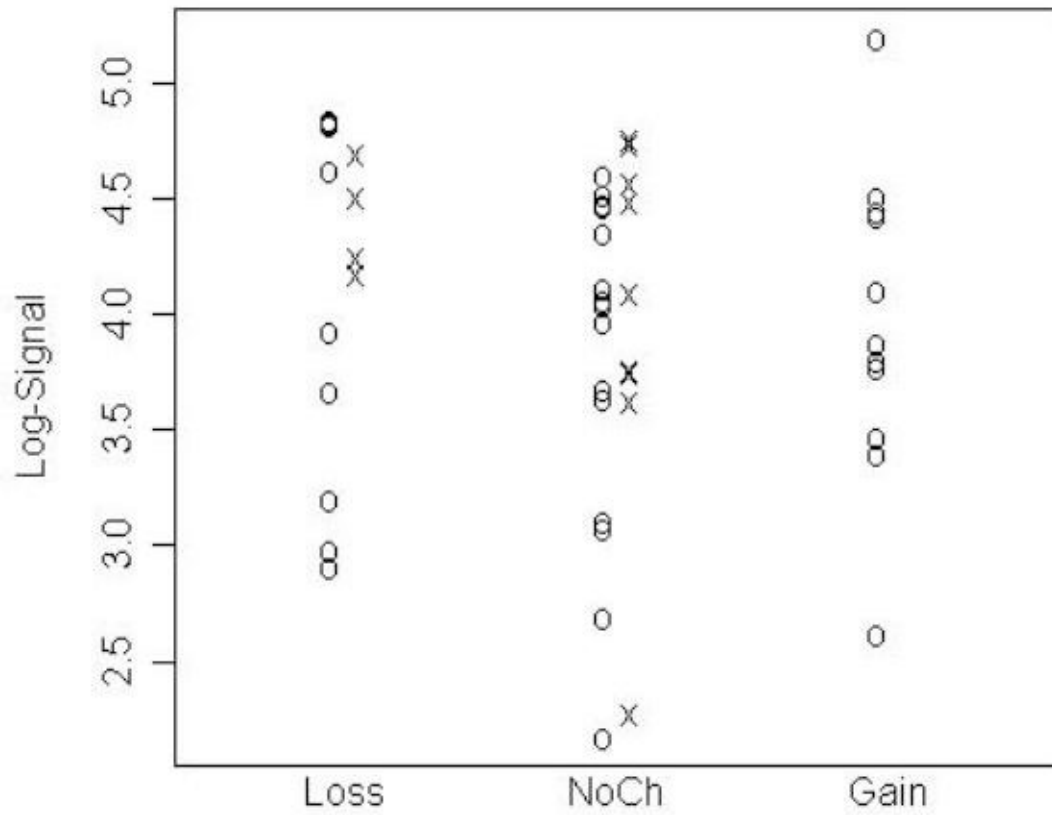


Figure 4.29 Correlation of NIPA1 Gene Expression and Copy Number Variation. (X – Case; O – Control)

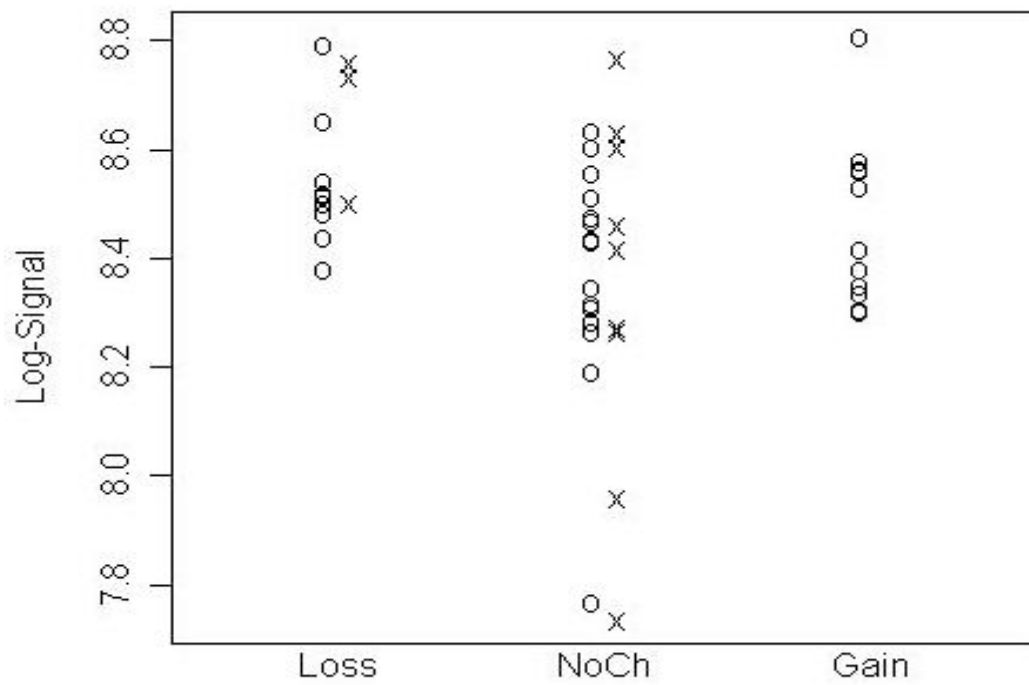


Figure 4.30 Correlation of NIPA2 Gene Expression and Copy Number Variation. (X – Case; O – Control)

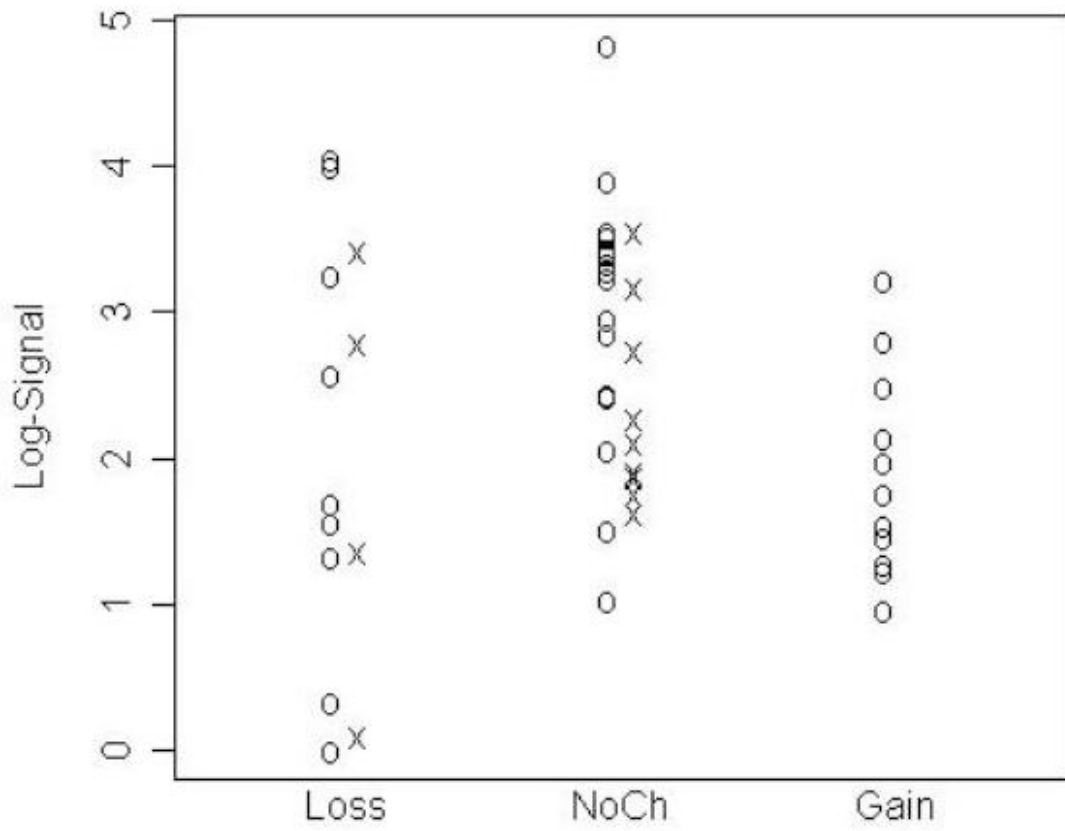


Figure 4.31 Correlation of BCL8 Gene Expression and Copy Number Variation. (X – Case; O – Control)

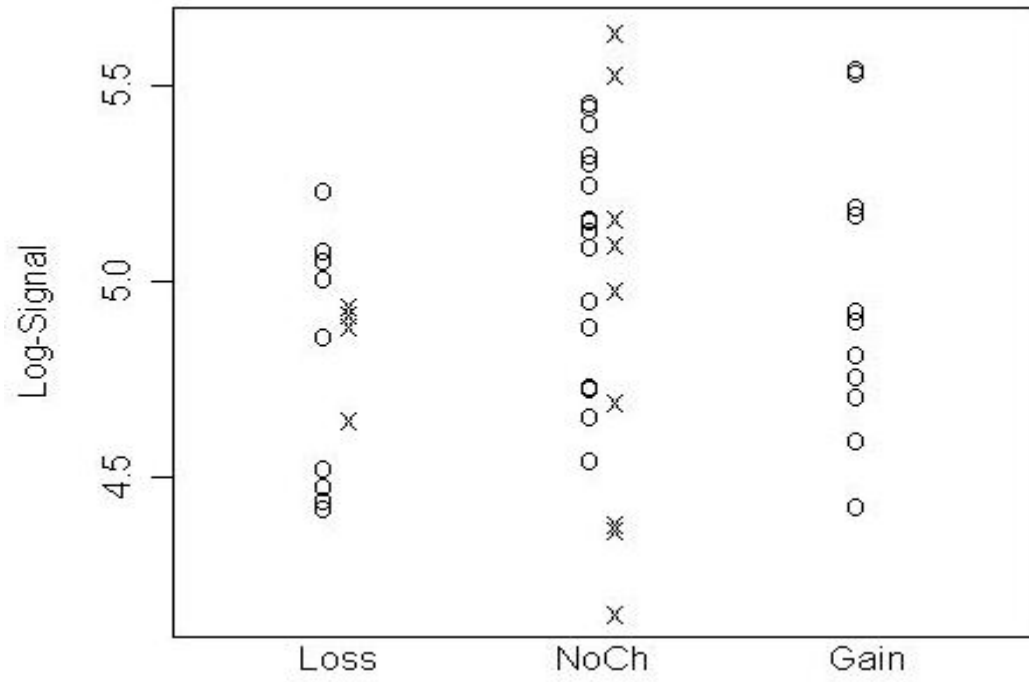


Figure 4.32 Correlation of TUBGCP5 Gene Expression and Copy Number Variation. (X – Case; O – Control)

Table 4.8 Association of First Degree Breast Cancer Family History and Deletion at 15q11.2. 95% Confidence.

Parameter	Estimate	Limits		p-Value
Relative 1	9.499	1.092	135.406	0.0392
Relative 2	2.730	0.380	20.335	0.4154

expressed between the cases and controls. This provided pathways and genes of interest in association with the phenotype of HL and breast cancer. The analysis found significant differences between the cases and controls in genes involved in amino acid metabolism, metabolism of cofactors and vitamins, biosynthesis of secondary metabolites, folding, sorting, and degradation of proteins, translation, and DNA replication and repair. Therefore, there are differences in the global gene expression between those with and without secondary breast cancer.

Aim two explored the genotype variability between cases and controls and found a large number of SNP variation and statistical differences between the genotype of those with and without secondary breast cancer. The most significant SNP variability between cases and controls was noted on chromosomes 5, 10, and 11. A region of genotype variability on chromosome 10 is of great interest due to the significant association with the BCCIP gene ($p=.001$) and the genes known interaction with BRCA2 and the phenotype of breast cancer.

Lastly, the analysis of aim three screened for evidence of amplification or deletion within the SNP array signal data to determine a set of candidate genes and region of interest associated with case/control status. The analysis found inferred regions of 14q11.2 and 15q11.2 through copy number variation to have some association with case/control status. An analysis of association of inferred copy number variation with the expression of individual probe sets resulted in a list of candidate genes within these regions. Region 14q11.2 found two genes within the region of interest with both being from the large olfactory gene family and known to have significant variability. The region of 15q11.2 found 19 candidate genes of interest most specifically genes found to be associated with the phenotype of breast cancer and lymphoma and include PAK2, CYFIP1, NIPA1, NIPA2, TUBGCP5, BCL8, and numerous hypothetical proteins.

CHAPTER 5. DISCUSSION

Discussion of Findings

Chapter 5 presents a discussion of the demographic data and the primary aims. The findings will be discussed in relation to the conceptual framework along with nursing implications and recommendations.

Demographic Data

Cases and controls were well matched with no statistical differences in age at HL diagnosis and time since diagnosis, maximum dose of mantle/mediastinal radiation, or alkylating chemotherapy. Comparison of demographic data for age of menses, menopause, and hormonal use found only the age of menopause ($p=0.0191$) to be statistically different between the cases and controls. This difference is most likely due to the fact that cases and controls were not matched for para-aortic and/or pelvic radiation. For the controls who received pelvic radiation, early age of menopause with onset prior to 31 years of age may be most protective against breast cancer [14]. The age at 1st pregnancy and time of pregnancy after HL have been described as being associated with secondary breast cancer [53]; however, these ages were not statistically different between the cases and controls. Although not significant, the reproductive history did find that the controls had fewer numbers of pregnancies and live births in comparison to the cases. This may be related to the earlier menopause secondary to a higher incidence of pelvic radiation or other factors that contributed to reproductive decisions. Lastly, the cases and controls provided a self report of first and second degree family members with cancer. Both the cases and controls reported a 45% incidence of cancer among first and second degree relatives with breast being a common cancer among both first and second degree relatives.

Breast cancer among the cases was diagnosed on average 36.5 ± 6.7 years of age, with a range of 27.2-51 years, which is significantly younger than the general population median diagnosis age of 61 years [12]. The time from radiation to the development of breast cancer was on average 21.3 ± 5.8 years after radiation and is consistent with the previously described range of 15 to 20 years [3, 5, 7]. The controls are approximately 25.4 ± 6.01 years (range 15.2-33.8 years) from HL diagnosis. This is important in that the majority of controls have exceeded the passage of time and the period of highest relative risk of secondary breast cancer after radiation exposure. However, there is significant variability with a standard deviation ± 6.01 years, meaning that some controls remain at an increased relative risk. The cases were also asked specifics regarding their breast cancer histology and hormonal markers. Information for the cases and chart review revealed no difference between the presentation of those with secondary breast cancer and sporadic disease. While bilateral disease has been reported to occur more frequently among HL survivors [3], only one case reported a bilateral occurrence.

Age at menopause ($p=0.0191$) was the only significant variable with statistical difference between the cases and controls. Overall, the demographic data found these survivors of pediatric HL to be a very homogenous group except for the occurrence of secondary cancer.

Primary Aim One

Aim one sought to determine the global gene expression differences between cases and controls. The global gene expression found significant gene expression differences in genes that are inter-related within 12 molecular pathways. These 12 pathways can be grouped into larger pathways of amino acid metabolism, metabolism of cofactors and vitamins, biosynthesis of secondary metabolites, and genetic information processing including folding, sorting and degradation, translation, and DNA replication/repair. The 12 significant pathways are outlined in Table 5.1. Within these pathways 43 genes and their isoforms were differently expressed between cases and controls with test p-value and permutation p-values of significance. However, the false discovery rate for each of the genes was .70 or greater indicating that 70% of the reported results are expected to be false discoveries. Although the data declare a high false discovery rate, the gene set is of interest in relation to the phenotype of secondary breast cancer.

Pathways of Significance

Folate-Methionine Pathway. One carbon metabolism, DNA methylation, synthesis, and repair have been long studied as mechanisms for the development of carcinogenesis. The methionine and folate metabolizing pathways are integral for DNA synthesis and methylation and are of interest with this study phenotype. During translation and transcription, gene expression is associated with the patterns of modification of DNA through the process of methylation. The process of methylation is not an alteration in the nucleotide sequence, but results from an addition of a methyl group to the cytosine-phosphoguanine (CpG) promoter region of the gene [84]. These CpG islands are located in over half of all human genes [85]. Gene expression can be altered by DNA methylation resulting in both hypomethylation and hypermethylation. DNA hypomethylation has been correlated with cancer through activation of proto-oncogenes, reactivation of transposable elements, and loss of imprinting [86]. Hypermethylation results in gene repression of tumor suppressors, chromatin condensation, and loss of DNA repair [87]. Methylation has been implicated as an early event in breast cancer, with the activation of oncogenes and the suppression of tumor suppressors.

Many of the genes involved in DNA methylation and synthesis were statistically significant in their differential expression between cases and controls as described and depicted in Figure 5.1. The first described is thymidylate synthetase (TS OR TYMS) which binds methylenetetrahydrofolate (5, 10-methylene THF) and serves as a hydroxymethyl donor in the conversion of dUMP to dTMP which is critical in DNA

Table 5.1 Grouping of Significant Pathways.

Pathway Function	Pathway
Amino Acid Metabolism	Methionine Metabolism
	Histidine Metabolism
	Phenylalanine, Tyrosine, and Tryptophan Biosynthesis
	Phenylalanine Metabolism
Metabolism of Other Amino Acids	Selenoamino Acid Metabolism
Metabolism of Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism
Biosynthesis of Secondary Metabolites	One Carbon Folate Pool
	Alkaloid Biosynthesis I Tetracycline Biosynthesis
General Information Processing	
Folding, Sorting, and Degradation	Type II Secretion
Translation	Aminoacyl-tRNA biosynthesis
Replication and Repair	DNA Polymerase

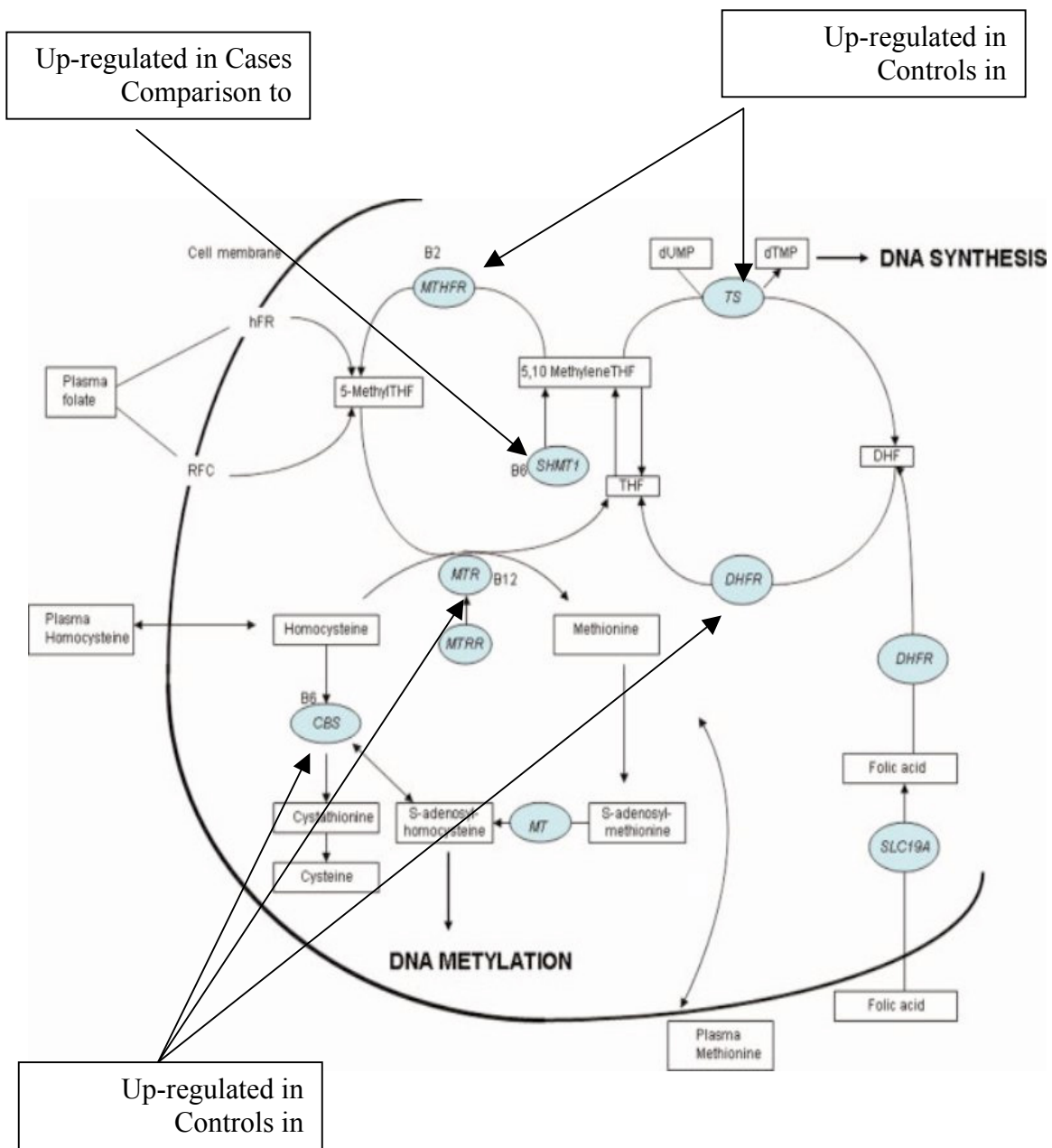


Figure 5.1 DNA Syntheses and Methylation. Adapted with permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc. from Lissowska, et al. Genetic polymorphisms in the one-carbon metabolism pathway and breast cancer risk: a population-based case-control study and meta-analyses. *Int. J. of Cancer*. 2007. 120: p 2697 [122].

synthesis and was expressed higher among the controls. Another gene, dCMP deaminase (DCTD) within the Type II secretion system pathway catalyzes the deamination of dCMP to DUMP a substrate for thymidylate synthase and was expressed higher among the controls. Inter-related is the process of DNA methylation in which methylenetetrahydrofolate reductase (MTHFR) catalyzes the reduction of 5, 10 methylenetetrahydrofolate (methylene THF) to 5-methyl THF, 5-methyl THF is the circulating form of folate and carbon donor for the remethylation of homocysteine to methionine. Therefore, MTHFR is important in homocysteine metabolism, methionine synthesis and methylation reactions. Reduced gene activity may result in accumulation of 5, 10 methylene THF and reduced 5-methyl THF. Although not statistically significant, controls were found to have a higher expression of MTHFR than the cases. From the methionine, methionine adenosyltransferase (MAT2B) catalyzes the biosynthesis of S-adenosylmethionine (SAM) and was expressed higher among the cases.

Two vitamin B dependent enzymes were expressed differently between cases and controls. Methionine synthase (MTR) is a vitamin B-12 dependent enzyme and catalyzes the methyl base from 5-methyl THF to homocysteine for the production of methionine and tetrahydrofolate. MTR is also important in maintaining S-adenosylmethionine (SAM) for DNA methylation and preventing increased homocysteine and was significantly expressed higher among the controls. Vitamin B6 enzyme, serine hydroxymethyltransferase (SHMT) was found to be expressed higher in the cases than controls. SHMT is responsible for conversion of serine and THF to glycine and methylene THF and is important in providing one-carbon units for purine, thymidylate, and methionine synthesis. Lastly, within Figure 5.1 is dihydrofolate reductase (DHFR) which was expressed higher among controls. This gene is within the one carbon folate pathway and required for de novo synthesis of amino acids.

Lastly, two genes of importance within the methionine metabolism were tRNA aspartic acid methyl-transferase (TRDMT1) expressed higher in the cases and DNA methyltransferase 3 alpha (DMT3A) expressed higher in the controls. Both of the these genes regulate CpG methylation. The DNMTs transfer a methyl group from SAM to generate methylation and the silencing of genes [88]. DNMT3a is a de novo methyltransferase and functions to methylate unmethylated DNA.

Dietary intake has been linked with DNA methylation and most specifically deficiencies in folic acid, methionine, zinc, selenium have been implicated [89]. While the nutritional data remains controversial, there is the link of 5-methylenetetrahydrofolate as the precursor of the methyl group needed for methionine to SAM [90]. While adequate nutrients are needed, excess of intake of folic acid, methionine, and selenium have been known to disrupt DNA methylation [87]. There are many factors that are important in maintenance of the folate-methionine pathway with gene expression being differently expressed between the cases and controls.

t-RNA-Dependent Amino Acid Biosynthesis. The aminoacyl-tRNAs are essential substrates in the process of translation and protein synthesis. As ribosome moves along the mRNA aligning the codons, aminoacylated-tRNA is brought into line for the transfer

from the mRNA into the growing amino acid polypeptide chain. Each aminoacyl-tRNA synthetase recognizes its specific amino acid. Therefore, the aminoacylated t-RNA is essential in genetic coding and serves a major function as RNA editors. Most of the aminoacyl-tRNAs can distinguish their corresponding amino acid; however, many can not bringing in the importance of pre and post transfer editing [91]. Mutations within aminoacylated t-RNA result in ambiguity of the genetic code and missense substitutions.

Among the cases and controls, the t-RNA dependent amino acid biosynthesis pathway was statistically significant. Eight aminoacyl-tRNA associated genes were differentially expressed and all were significantly expressed higher among controls than cases. These genes included: three isoforms of MARS which is a methionine tRNA synthetase and MARS2 which is the tRNA synthetase responsible for mitochondrial tRNA, YARS tyrosyl tRNA synthase FARS2 phenylalanine tRNA synthetase mitochondrial, AARS alanyl tRNA synthase, LARS and LARS2 leucyl tRNA synthase, and HARS histidyl tRNA synthase. It could be inferred that lower gene expression among cases may be responsible for impaired protein synthesis.

Nicotinate and Nicotinamide Metabolism. The metabolism of nicotinate and nicotinamide are pathways of interest due to their response to cellular stress for which four genes were expressed higher among the cases than controls. Two isoforms of nicotinamide nucleotide adenylyltransferase 2 (NMNAT2) a family of enzymes which catalyze an important step in nicotinamide adenine dinucleotide (NAD) biosynthesis were differentially expressed. Another gene pre-B-cell colony enhancing factor 1 (PBEF1) was also expressed higher among the cases and is involved in the biosynthesis of NAD. The biosynthesis of NAD is of interest within cases in that DNA damage is found to stimulate NAD biosynthesis and recovery of damage occurs sooner with higher levels of NAD [92].

DNA Polymerase. DNA polymerases are important in the repair of DNA after toxic exposure and prevent mutagenic lesions within the DNA. Many DNA polymerases are known and three were differently expressed higher among the controls than cases and include Poly beta sigma and gamma. Poly beta supports DNA maintenance, replication, and recombination through base excision repair. Poly beta has been found to be involved in DNA repair of UV lesions allowing DNA to repair through translesion synthesis[93]. In tumor formation, an increased expression of Poly beta is consistent with synthesis of genes needed for DNA repair in an attempt to detain the mutagenic lesion [94, 101].

Biosynthesis of Secondary Metabolites. Another gene of interest and associated with breast cancer is Acetyl-CoA carboxylase α which was expressed higher among the controls. This gene has been identified as being a partner of the breast cancer gene BRCA1 and interacts with AAC α via the tandem of BRCT domains in BRCA1 [94]. Therefore, the higher expression among the controls may facilitate the interaction and function of BRCA1. This gene has been found to be expressed in the liver, adipose, brain and mammary gland during lactation. AAC α is highly expressed in breast carcinoma [94].

Global Gene Expression Summary

The study finds that there are differences in the global gene expression between those with and without secondary breast cancer. The expression data defines 12 molecular pathways and 43 genes ($p \leq 0.05$) differentially expressed between the cases and controls within each of the pathways. Most specifically, the cases were found to have lower expression in the folate-, methionine, t-RNA-Dependent Amino Acid Biosynthesis, Nicotinate and Nicotinamide Metabolism. Regardless of the direction of change in gene expression, differences were found between the cases and controls and occurred within key processes of amino acid metabolism, one-carbon metabolism, DNA replication, and RNA translation.

These pathways of interest are very similar to those described by the consensus coding sequences of human breast and colon cancer [95]. This study evaluated genetic mutations of tumors, focusing on protein coding genes or consensus coding sequences. This data generated a list of somatic mutations and candidate cancer genes. Each of these genes was assigned a function based on molecular function or biochemical processes. For breast cancer 18% of the genes were transcriptional regulators, of particular zinc finger transcription factors. Additional pathways included cellular adhesion and motility, signal transduction, transport, cellular metabolism, intracellular trafficking, and RNA metabolism, and response to DNA damage stimulus [95]. Many of the described pathways within the tumor from the consensus coding sequences were found from this study of germline gene expression. Therefore, differential expression within these pathways may be important in the transformation of the breast cancer phenotype in cases.

Primary Aim Two

Aim two determined the genotype differences between the cases and controls, and found there to be SNP's of statistical significance in allele frequency between the cases and controls. As previously discussed, the most significant SNP's with a chi-square p-value < 0.001 were transformed, plotted, and highlighted according to the chromosome location. A Geneset enrichment p-value file was created to determine the genes associated with the SNP's of interest and found 574 genes associated with the corresponding SNP. SNP's of greatest significance and within clusters were determined to be of highest interest and were found on chromosomes 5, 10, and 11. These SNP's are described according to their corresponding gene function and relevance to the phenotype of breast cancer.

Differences in Genotype

As previously described, chromosome 5 displays a significant SNP_1907286 ($-\log_{10} p = 4.8$) at position 31786818. This SNP is associated with the PDZD2 gene and is found in an intron and falls within the outer boundaries of the transcript record [29]. The PDZD2 has a physical position of 31834787-32146794; therefore, the SNP is not

located within the gene thus a Geneset enrichment value of ($p=0.49$; $FDR\approx 1$). Next the known allele frequency was considered. The SNP allele A/T frequency among CEPH (Utah residents with ancestry from northern and western Europe) found the major A allele at a 38% occurrence, minor T allele at a 62% occurrence, and a heterozygosity frequency of 47% [29]. Therefore, this is a commonly occurring polymorphism among the CEPH population studies. The SNP associated gene, PDZD2 has been noted to be associated with prostate tumors but no association with breast cancer. The PDZD2 gene contains PDZ domains that have been shown to bind the C-termini of transmembrane receptors or ion channels and could be involved in intracellular signaling. The protein encoded by this gene contains six PDZ domains and the encoded protein localizes to the endoplasmic reticulum and is thought to be cleaved by a caspase to produce a secreted peptide containing two PDZ domains. The gene has been found to be unregulated in primary prostate tumors [83].

The largest cluster of differentially expressed allele frequencies was found on chromosome 10, over a region of 6.1 kilobases with 11 differentially expressed SNP's between cases and controls. As previously described, two highly significant SNP's within this region are SNP_A-2181577 ($-\log_{10} p=3.97$) and SNP_A-2006464 ($-\log_{10} p=3.97$) at position 127481155 and 127485797 and statistically associated with the UROS gene ($p=.001$; $FDR\approx 1$) located at position 127467141-127501757. Both SNP's are found in the gene intron and fall within the outer boundaries of the transcript record [29]. SNP_A-2181577, allele C/T among the CEPH population has a major allele frequency of 61%, minor allele frequency of 38.6% and a heterozygosity frequency of 47.4%. SNP_A-2006464, allele A/C among the CEPH population has a major allele frequency of 59%, minor allele frequency of 41% and a heterozygosity frequency of 48% [29]. The UROS (uroporphyrinogen II synthase) gene encodes a gene which catalyzes the fourth step of porphyrin biosynthesis in the heme biosynthetic pathway and is found to have no correlation with the phenotype of breast cancer [83].

SNP_A-1842005 ($-\log_{10} p= 3.98$) at position 127508528 was found to be the most significant SNP within the region of chromosome 10 and was statistically associated with the BCCIP gene ($p=.001$; $FDR\approx 1$) located at the physical position of 127502104-127532080. SNP_A-1842005, allele C/T among the CEPH population has a major allele frequency of 59%, minor allele frequency of 40%, and a heterozygosity frequency of 48% [29]. The BCCIP gene is a BRCA2 and CDKN1A interacting protein and was isolated due to its interaction with BRCA2 and p21 proteins. It is an evolutionarily conserved nuclear protein with multiple interacting domains. Functional studies indicate that this protein may be an important cofactor for BRCA2 in tumor suppression, and a modulator of CDK2 kinase activity via p21. Several transcript variants encoding different isoforms have been described for this gene [83]. This gene may prove to be very important within this cohort due to the BCCIP gene interaction with BRCA2 and association with breast cancer phenotype.

The BCCIP gene is of interest due to its interaction with the tumor suppressor BRCA2 and association with the phenotype of breast cancer. The function of BCCIP was first described due to its interaction with BRCA2 and CDKN1A (p21) [96]. BRCA2 is

known to be a tumor suppressor and is involved in DNA repair contributing to genomic stability [97], regulation of mitosis and cytokinesis ensuring numerical chromosomal stability [98]. Lu et al [99] described the chromatin-bound fraction of both BCCIP isoforms colocalize with BRCA2 and contribute to BRCA2 and RAD51 nuclear focus formation. A recently published study found that downregulation of either BCCIP isoform significantly reduces DNA double strand break-induced homologous recombination, impaired G1/S checkpoint activation, abrogates p53 transactivation and down-regulates p21 expression [100, 101]. Therefore, BCCIP expression may regulate genomic stability through multiple pathways including homologous recombination repair, S checkpoint, and cytokinesis. Additionally, BCCIP may regulate centrosome stability through p53 and p21 function, sequestering p53 protein from the p21 promoter despite an elevated total p53 protein level [101]. Due to the importance in genomic stability, BCCIP loss may lead to tumorigenesis and becomes an important candidate gene for the phenotype of breast cancer.

Within this polymorphic region there were two other significant regions of allele frequency differences between cases and controls, SNP_A-1944367 (-log₁₀ p= 3.55) located at position 127555310 and SNP_A-1887998 (-log₁₀ p= 3.47) located at position 127455338. SNP_A-1944367 is located within the intron of the gene DHX32 (p=.001: FDR ≈1) and falls within the boundaries of the transcript record. The allele C/T among the CEPH population has a major allele frequency of 40%, minor allele frequency of 60%, and heterozygosity frequency of 48% [29]. The gene DHX32 (Dead box polypeptide 32) is characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD) which are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome assembly and may be involved in cellular growth and division.

The DHX32 gene is of interest due to its role in RNA metabolism which was found to be a significant pathway within the analysis of global gene expression. DHX32 gene expression has been found to be down-regulated in acute lymphoblastic leukemia, lymphoblastic lymphoma, and follicular lymphoma. Inversely the DHX32 has been found to be up-regulated in mantle cell lymphoma, B-cell lymphoma, and Reed Sternberg cells in cases of nodular sclerosis Hodgkin lymphoma have been found to strongly express DHX32 [102, 103]. Not only has the DHX32 gene expression been association with Hodgkin lymphoma, DHX32 is an anti-sense to the BCCIP gene. The last three exons of the BCCIP gene overlap the 3'-terminal seven exons of the DHX32 gene [102].

The second SNP of significance within this region, SNP_A-1887998 is located downstream 11800 bases from the transcript of the gene UROS and 958 bases upstream from the transcript of the gene MMP21 (p=.001: FDR ≈1). The allele C/T among the CEPH population has a major allele frequency of 69%, minor allele frequency of 30% and heterozygosity frequency of 42%. The MMP21 gene (matrix metalloproteinase 21) encodes a member of the matrix metalloproteinase family involved in the breakdown of extracellular matrix for both normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, and disease processes, such as asthma

and metastasis [82]. The MMP21 gene has been found to contribute to tumor development and polymorphisms within this gene have been associated with an increased risk of cancer [104, 105].

The analysis found the most significant SNPs, SNP_A-2294376 (-log₁₀ p=5.1) and SNP_A-2209464 (-log₁₀ p=5.2), to be located on chromosome 11 at position 119148037 and 119149661 and are associated with the genes TRIM29 and PVRL. SNP_A-2294376 is located upstream at a distance of 43392 nucleotide bases from the transcript of PVRL1 and downstream 339168 nucleotide bases from the transcript of TRIM29. SNP_A-2209464 is located upstream at 45016 nucleotide bases from the transcript of PVRL1 and downstream 337544 nucleotide bases from the transcript of TRIM29. Due to the SNP location in relation to the gene, these genes were not statistically associated with the SNP in the Geneset enrichment file, TRIM29 (p=0.845) and PVRL (p=0.142).

Genotype Summary

Aim two found that there were differences in the genotype of those with and without breast cancer and the most significant genotype differences were found on chromosomes 5, 10, and 11. A region within chromosome 10 was found to have the largest cluster of significant genotype variability and genes that are of importance to the breast cancer phenotype. From this analysis, SNP's within the BCCIP and DHX32 gene were found to be statistically significant, overlap in physical position, and have been described as being associated with the phenotype of breast cancer and Hodgkin lymphoma. Additionally, polymorphisms within the gene MMP21 as described have been associated with an increased cancer risk. The region within chromosome 10 becomes of importance for further validation and analysis due to the strong association with the phenotype not only of breast cancer but of Hodgkin lymphoma.

Primary Aim Three

Copy Number Variation

Aim three identified inferred regions of 14q11.2 and 15q11.2 through copy number variation to have some association with case/control status. It is known that copy number variation is integral in the genomic variation between individuals. However, the extent to which genetic variability contributes to disease is unknown. One study suggests that copy number variations greater than 100 kilobases contribute to genetic variation and found these large scale variations to differ by 11 copy number polymorphisms with an average length of 465 kilobases [106]. Nozawa et al [107] found that 14.2% of the human gene loci have copy number polymorphisms. Therefore, regions of copy number variation are not uncommon but will vary according to those sampled. Although the effect of copy number variation on the phenotype is unknown, it is thought that the effect

is significant given the impact of copy number variation on gene expression [108]. Copy number variation requires additional analysis of the genotype data. This can be explained in that the SNP genotype assumes 2 copies and will force variants to conform to the chip. For example, the region of copy number variation amplification may have a genotype of AAB, which will be recorded as AA or AB on the SNP chip without inference to the additional allele. Therefore, this study analyzed the genotype data for areas of copy number variation. This analysis found regions of copy number variation at 14q11.2 ($p=0.069$) and region 15q11.2 ($p=0.067$). From the copy number polymorphisms the log-transformed Affymetric U133 plus expression signal by inferred copy number status for each of two probe sets mapped to genes that overlap or are located with 5,000 base pairs of the copy number variation region. This provided the candidate genes for the two regions of interest.

The region of 14q11.2 was mapped to two genes OR4N4 and OR4Q3. These genes are olfactory sensory genes and one of the largest gene families with the genome. The sensory genes have been studied due to their large number of copy number variations [109, 110]. It is thought that this variation is the result of physiological requirements in response to ones environment [107]. In addition, the sensory gene families are thought to contain a large number of pseudo genes [103, 104] that appear to have experienced random change in copy number variation and have undergone genetic drift [111]. This region of variation is well documented and most related to evolutionary change.

The region of chromosome 15q11-13 has been found to be one of the most unstable regions within the genome [112]. Clinical cytogenetic analysis have frequently identified a region of amplification on 15q that has not been associated with Prader-Willi/Angleman Syndrome and appears to have no significant effect on phenotype [113, 114]. This region of amplification has also been studied for patterns of inheritance and found inherited from either sex parent with no apparent parent of origin effect [115]. Another study followed 130 families with unbalanced chromosome abnormalities at five regions including 15q11-13 and 70 families with euchromatic variants including 15q11.2 [116]. Among the families with an unbalanced chromosome abnormality, 9 families had a duplication of this region with either an affected child or an affected parent and child. Expression of phenotype was due to involvement of the critical Prader-Willi region and included developmental delay and autism. Thirty two families were found to have euchromatic variants within the region of 15q11.2. Euchromatic variants reflect copy number variation of segments of genes and pseudogenes which are polymorphic in the general population and typically have no phenotypic effect. This study found 32 families to have duplications within 15q11.2. This gain in copy number is thought to be the result of additional copies of the NF1 pseudogene. Additionally, the study notes that similar variation may be expected at other sites of the NF1 pseudogene including 14q11.2 [116]. Another study has also documented the 15q duplication has being inherited from either parent with no effect on phenotype, and related to the polymorphic regions of pseudogene sequences [115]. Within this analysis of copy number variation, 11 controls were found to have duplication at 15q11.2; no case was found with duplication. From the described studies and this duplication only occurring within the controls, it is likely a common inherited variant with no effect on phenotype.

The occurrence of deletion was found in 4 cases and 9 controls, including the one case who has died from breast cancer progression and a control who has died from another secondary tumor. Deletion within this region has typically been described with Prader-Willi and Angelman Syndrome. A search of the literature found no reports of a deletion at this region in association with the phenotype of breast cancer or Hodgkin lymphoma. However, the analysis of family history as described earlier did find an association ($p=0.039$) with the deletion and first degree relative with breast cancer for both the cases and controls.

The remainder of the cases and controls had no change within this region. However, in reviewing the data it is important to keep in mind that this analysis compares signal values from within the cohort to themselves. Therefore, it is most likely that many of the copy number variations were missed. It is likely that many of those found in the analysis to be no change may actually have a duplication or deletion when the validation real-time PCR is complete. It is less likely that those with a duplication or deletion in this analysis become no change (two copies). This conclusion is inferred from the ad hoc statistical approach.

The location of the copy number variations produced a list of candidate genes for the 15q11.2 region. This list included CYFIP1, PAK2, RNF145, NIPA1, NIPA2, BCL8, TUBGCP5, and numerous hypothetical proteins. Expression of these genes was not noted to be directional according to copy number variation. A review of the literature explored the association of these candidate genes with the phenotype of lymphoma, cancer, and breast cancer.

The region of 15q11-13 has been found to be a site of rearrangements in 3-4% of those diagnosed with diffuse cell large cell lymphoma and is the location of BCL8 a B-cell lymphoma gene [117]. The BCL8 was found due to this recurrent site of rearrangements in a small number of large cell lymphoma patients. The function of BCL8 is not understood but it is ectopically activated in lymphomas. Furthermore, it is speculated that this site rearrangement is most likely related to mutations in the regulatory region or deregulation by a trans-acting factor or gene deregulation by changes in the imprinting pattern of the chromosomal region [117].

Two of the candidate genes, PAK2 and CYFIP1 are within the tyrosine kinase family. Tyrosine kinases are enzymes that transfer a phosphate group causing phosphorylation after a factor binds to its receptor. This allows the cell to participate in a chemical reaction such as activation of a signaling pathway. CYFIP1 is a tyrosine kinase which has been found to have differential expression during breast cancer development [118]. The serine threonine protein kinase, PAK2 is activated by cellular stress such as hyperosmolarity, ionizing radiation, and DNA damage. PAK2 executes part of the cellular response to stress by inhibiting activity of Myc and may be a tumor suppressor [119]. Expression of PAK2 is frequently affected by rearrangements associated with hematological malignancies such as chronic lymphocytic leukemia and B-cell lymphoma. The interaction between Myc and PAK2 may be bidirectional, in that Myc may down regulate PAK2 [120]. Lastly, candidate genes, NIPA1 and TUBGP5 have recently been

documented as having potential implication in the development of breast cancer in Li-Fraumeni syndrome animal model [121].

Copy Number Variation Summary

The copy number analysis found some association with case/control status and copy number variation at 14q11.2 and 15q11.2. Amplification with this region in those with no phenotype effect has been described as an euchromatic variant and likely related to the polymorphic segment of the NF1 pseudogene [115, 116]. However, the deletion is not described within the literature and may be of interest with the phenotype of breast cancer. This is further warranted with the association noted between the deletion among the cases and controls and family history of breast cancer. Lastly, the region of 15q11.2 provides another list of potential candidate genes that may be associated with the development of breast cancer among HL survivors. The genes of greatest interest include CYFIP1, PAK2, NIPA2, and TUBGCP5 which have been cited in the literature as being associated with the phenotype of breast cancer as previously described. An analysis is ongoing to validate the presence of amplification or deletion within the 15q11.2 region of interest.

Relation of Study Results to Conceptual Framework

This study evaluated the similarities and differences in global gene expression, genotype, and copy number variation among females treated for pediatric HL with and without the development of secondary breast cancer. The conceptual model (see Figure 1.1) depicts each individual with treatment related risk factors including age at treatment, dose of mediastinal and pelvic radiation, and dose of alkylating agents. There were no statistical differences in age, mediastinal radiation, or alkylating agents; however, there were differences between the cases and controls receiving pelvic radiation and age of menopause. It is these treatment characteristics and their interaction with inherited characteristics that place the individual at an increased risk of secondary breast cancer. Therefore, the family history was assessed between the cases and controls, with no differences noted. However, at the time of therapy the family history is less suggestive of potential breast cancer risk and should be reevaluated with the passage of time. Additionally, at the time of therapy, the individual's genetic influences are unknown; therefore, the individual's ability to maintain genomic stability after the initiation of radiation and chemotherapy at a young age is unknown. Females who maintain genomic stability have ablation of tumor and cellular apoptosis or DNA repair secondary to the genomic insult of radiation and chemotherapy. Females who are unable to maintain genomic stability will retain mutated lesions and with additional insults over time will likely develop secondary breast cancer.

This study found differences in gene expression, genotype, and a region of copy number variation among the cases and controls. These findings will require a step wise validation with further analysis and subsequent study. If validated these findings could be

used as a screening for females at potential risk of secondary breast cancer after HL therapy. While these validation studies are on going, changes in the conceptual model should include a periodic update of the individual's family history. Future studies will continue to match treatment variable between cases and controls and will include pelvic radiation. Lastly, each individual should be made aware of lifestyle characteristics that reduce additional genetic insults such as healthy diet (folic acid, methionine, and selenium), limited sun exposure, non smoking, use of hormonal therapy and limited alcohol intake. For young females undergoing pediatric cancer therapy, their family history may not be noted as positive for breast or other cancer. However, these females should understand the importance of their family history, lifestyle characteristics and the implications these have for their own health risk. There will be an ongoing assessment of the conceptual framework in subsequent studies.

Nursing Implications

This study has implications for nursing research and practice. The knowledge gained could provide a larger framework of research in the exploration of the biological mechanism underlying early on-set breast cancer after a childhood cancer. With an improved understanding of inherent risk factors, the biological mechanism, and the interaction of endocrine influences, this finding could assist in developing a predictive clinical model. It is reasonable that future treatment approaches may require screening for breast cancer-related polymorphisms or copy number variations prior to treatment of pediatric HL. If future screening identifies an individual with a genetic predisposition for the development of breast cancer, therapy may be modified in an effort to decrease this risk. This may include a modification of radiation, chemotherapy, hormonal therapy post-treatment therapy, or the administration of anti-estrogen therapy. The individual and family would be informed of and develop an understanding of potential risk and the measures that are needed to heighten surveillance. This surveillance should also entail an annual review of the individual's family cancer history and most specifically that of breast cancer among primary and secondary degree relatives. In practice the oncology nurse would navigate the individual through a surveillance plan which would include monthly self-breast exam, counseling in obtaining a biannual physician breast exam and annual diagnostic imaging. The nurse should provide counseling that may reduce oxidative stress and genomic insults through lifestyle characteristics such as healthy diet (folic acid, methionine, and selenium), limited sun exposure, non smoking, limited use of hormonal therapy and limited alcohol intake. Lastly, the nurse can explain and disseminate these findings to primary care providers caring for these women.

Limitations

The major limitation of this study is the small sample size of 13 cases and 36 controls. While SJCRH is a large center for patient accrual with secondary breast cancer after pediatric HL, only 14 known secondary breast cancer survivors were eligible at the

time of patient enrollment. To compensate for the small case sample size, matching of three controls did improve the efficiency for binary measures.

Another limitation of the study is the generalizability of the study findings. This study was designed to explore the mechanism of secondary breast cancer after pediatric HL; therefore, this study can not be generalized to sporadic breast cancer and will need further study among those with breast cancer after other pediatric cancers. Cases and controls were screened for active disease prior to blood collection; however, both cases and controls have subsequently reported disease reoccurrence or newly diagnosed disease. Therefore, the gene expression data may be affected by current disease status. Lastly, genomic analysis is moving at a rapid pace and provides numerous biostatistical and bioinformatical methodological limitations.

Recommendations

Recommendations for Research

These data and analysis provide a vast number of future hypotheses generating questions.

1. The study findings must be validated by real time PCR to determine the presence of candidate polymorphisms and a gain, loss or no change at 15q11.2
2. Further analysis of data to determine if variation in gene expression is related to genotype or copy number variation.
3. The study should be replicated in females treated for other pediatric cancers who develop secondary breast cancer and all new cases of HL associated breast cancer.
4. The current data should be compared to other genomic data sets that have evaluated germline polymorphisms among survivors of other pediatric cancers and females with sporadic breast cancer.
5. Further studies should utilize the DNA repository that was initiated through this study and access other polymorphisms of interest that were not included in the 500K genotyping such as the polymorphism associated with MDM2.
6. Additional analysis of the cases and controls in comparison to other genomic data sets to explore genomic etiology that may be associated with pediatric HL.
7. If the findings are replicated in associated studies, prospective screening may be appropriate in the diagnostic HL evaluation.

Recommendations for Practice

Currently, these study findings will not have application in clinical practice. However, if the real time PCR finds the locus of interest to be valid, findings may have clinical implications. The study should first be replicated within a larger HL cases/control designed study. If the findings are consistent and the loss (deletion) is found to be associated with secondary breast cancer, controls that have the deletion will require heightened monitoring and breast cancer surveillance. If the locus of interest is validated, evaluation may become routine for the newly diagnosed HL patient prior to the initiation of therapy.

Conclusion

This study was an unbiased genome scan of germline DNA and RNA with unique and common loci identified and provides a data base for additional analysis. This study explored the global gene expression and genotype between females treated for pediatric HL with and without the development of secondary breast cancer. To the best of my knowledge, this is the first genome wide association study among female HL survivors with and without secondary breast cancer. With the use of two genome wide approaches, this study found differences in global gene expression, genotype, and two regions of copy number variation that were associated with case/control status. The study did find that the global gene expression is different among cases and controls with differences in amino acid metabolism, translation, and DNA repair giving a list of candidate genes with these pathways. Secondly, the study found that there were differences between cases and controls and their genotype expression. Of particular interest are genotype differences with the BCCIP and DHX32 genes which have association with the phenotype of breast cancer and Hodgkin lymphoma. Lastly, the genotype data were screened for amplifications and deletions and found a near significant region of variability. The deletion of the region was found to be associated with a first degree family history of breast cancer; therefore, these women warrant continued follow-up for the development of breast cancer. Lastly a set of genes were described within this region and documented within the literature as being associated with the phenotype of breast cancer. This data analysis warrants validation of findings and further data analysis in determining potential candidate genes and genotypes in association with the development of breast cancer after pediatric Hodgkin lymphoma.

LIST OF REFERENCES

1. Hudson, M.M., et al., *Increased mortality after successful treatment for Hodgkin's disease*. J Clin Oncol, 1998. **16**(11): p. 3592-600.
2. Mertens, A.C., et al., *Late mortality experience in five-year survivors of childhood and adolescent cancer: the Childhood Cancer Survivor Study*. J Clin Oncol, 2001. **19**(13): p. 3163-72.
3. Kenney, L.B., et al., *Breast cancer after childhood cancer: a report from the Childhood Cancer Survivor Study*. Ann Intern Med, 2004. **141**(8): p. 590-7.
4. Lin, H.M. and M.A. Teitell, *Second malignancy after treatment of pediatric Hodgkin disease*. J Pediatr Hematol Oncol, 2005. **27**(1): p. 28-36.
5. Wendland, M.M., et al., *Time interval to the development of breast carcinoma after treatment for Hodgkin disease*. Cancer, 2004. **101**(6): p. 1275-82.
6. Neglia, J.P., et al., *Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study*. J Natl Cancer Inst, 2001. **93**(8): p. 618-29.
7. Ng, A.K., et al., *Second malignancy after Hodgkin disease treated with radiation therapy with or without chemotherapy: long-term risks and risk factors*. Blood, 2002. **100**(6): p. 1989-96.
8. Travis, L.B., R.E. Curtis, and J.D. Boice, Jr., *Late effects of treatment for childhood Hodgkin's disease*. N Engl J Med, 1996. **335**(5): p. 352-3.
9. van Leeuwen, F.E., et al., *Long-term risk of second malignancy in survivors of Hodgkin's disease treated during adolescence or young adulthood*. J Clin Oncol, 2000. **18**(3): p. 487-97.
10. Bhatia, S., et al., *High risk of subsequent neoplasms continues with extended follow-up of childhood Hodgkin's disease: report from the Late Effects Study Group*. J Clin Oncol, 2003. **21**(23): p. 4386-94.
11. Travis, L.B., et al., *Cumulative absolute breast cancer risk for young women treated for Hodgkin lymphoma*. J Natl Cancer Inst, 2005. **97**(19): p. 1428-37.
12. Reis, L.G., Eisner, M.P. Kosary, C., et al, *SEER cancer statistics review, 1973-2001*. Bethesda (MD): National Cancer Institute. Available at:<http://serr.cancer.gov/publicdata>. 2004.
13. Travis, L.B., et al., *Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease*. Jama, 2003. **290**(4): p. 465-75.
14. van Leeuwen, F.E., et al., *Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease*. J Natl Cancer Inst, 2003. **95**(13): p. 971-80.
15. Bhatia, S., et al., *Breast cancer and other second neoplasms after childhood Hodgkin's disease*. N Engl J Med, 1996. **334**(12): p. 745-51.
16. Guibout, C., et al., *Malignant breast tumors after radiotherapy for a first cancer during childhood*. J Clin Oncol, 2005. **23**(1): p. 197-204.
17. Dores, G.M., et al., *Second malignant neoplasms among long-term survivors of Hodgkin's disease: a population-based evaluation over 25 years*. J Clin Oncol, 2002. **20**(16): p. 3484-94.

18. Nichols, K.E., et al., *Heterozygous germline ATM mutations do not contribute to radiation-associated malignancies after Hodgkin's disease*. J Clin Oncol, 1999. **17**(4): p. 1259.
19. Nichols, K.E., et al., *TP53, BRCA1, and BRCA2 tumor suppressor genes are not commonly mutated in survivors of Hodgkin's disease with second primary neoplasms*. J Clin Oncol, 2003. **21**(24): p. 4505-9.
20. Mertens, A.C., et al., *XRCC1 and glutathione-S-transferase gene polymorphisms and susceptibility to radiotherapy-related malignancies in survivors of Hodgkin disease*. Cancer, 2004. **101**(6): p. 1463-72.
21. Kelly, K.M. and J.P. Perentesis, *Polymorphisms of drug metabolizing enzymes and markers of genotoxicity to identify patients with Hodgkin's lymphoma at risk of treatment-related complications*. Ann Oncol, 2002. **13 Suppl 1**: p. 34-9.
22. Society, A.C., *What are the key statistics about hodgkin's disease*. Retrived April 20, 2005, from <http://www/cancer.org>. 2005.
23. Travis, L.B., et al., *Cancer survivorship--genetic susceptibility and second primary cancers: research strategies and recommendations*. J Natl Cancer Inst, 2006. **98**(1): p. 15-25.
24. Murphy, N., E. Millar, and C.S. Lee, *Gene expression profiling in breast cancer: towards individualising patient management*. Pathology, 2005. **37**(4): p. 271-7.
25. Al-Kuraya, K., et al., *Prognostic relevance of gene amplifications and coamplifications in breast cancer*. Cancer Res, 2004. **64**(23): p. 8534-40.
26. Devillee, P., Cleton-Jansen, A.M., & Cornelisse, D.J., *Ever since Knudson*. Trends in Genetics, 2001. **17**: p. 569-573.
27. Knudson, A.G., *Cancer genetics*. Am J Med Genet, 2002. **111**(1): p. 96-102.
28. Nussbaum, R.L., McInnes, R.R., & Willard, H.F., *Thompson & Thompson genetics in medicine*. Vol. 6th rev.ed. 2004, Philadelphia: Saunders.
29. Affematrix.
30. Hartl, D.L., & Jones, E.W., *Genetics: analysis of genes and genomes*. 6th ed. DNA stucture and DNA manipulation. 2005, Sudbury: Jones and Bartlett.
31. Fathallah-Shaykh, H.M., *Microarrays: applications and pitfalls*. Arch Neurol, 2005. **62**(11): p. 1669-72.
32. Pounds, S. and C. Cheng, *Sample size determination for the false discovery rate*. Bioinformatics, 2005. **21**(23): p. 4263-71.
33. Hodgkin, T., *On some morbid appearances of the absorbent gland and spleen*. Medical-Chirurgical Society Transactions, 1832. **17**: p. 68-114.
34. Sternberg, C., *Uber eine Eigenartige unter dem Bilde der Pseudoleukmie verlaufende Tuberculose des lymphatischen*. Apparates Z Heilkd, 1898. **19**: p. 21.
35. Reed, D.M., *On the pathological changes in Hodgkin's disease, with special reference to its relation to tuberculosis*. Johns Hopkins Rep, 1902. **10**: p. 133.
36. Kuppers, R., M.L. Hansmann, and K. Rajewsky, *Clonality and germinal centre B-cell derivation of Hodgkin/Reed-Sternberg cells in Hodgkin's disease*. Ann Oncol, 1998. **9 Suppl 5**: p. S17-20.
37. Marafioti, T., et al., *Hodgkin and reed-sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription*. Blood, 2000. **95**(4): p. 1443-50.

38. Stein, H., Delsol, G., Pileri, S, *Hodgkin lymphoma*. World Health Organization classification of tumors. Tumors of hematopoietic and lymphoid tissues, ed. E.S. Jaffe, Harris, N.L., Stein, H., Vardiman, J.W. 2001, Lyon: IARC Press. 237-253.
39. Hudson, M.O., M, & Donaldson, S, *Hodgkin lymphoma*, ed. P.A.P. Pizzo, D.G.
40. Carbone, P.P., et al., *Report of the Committee on Hodgkin's Disease Staging Classification*. *Cancer Res*, 1971. **31**(11): p. 1860-1.
41. Donaldson, S.S. and H.S. Kaplan, *Complications of treatment of Hodgkin's disease in children*. *Cancer Treat Rep*, 1982. **66**(4): p. 977-89.
42. Donaldson, S.S. and M.P. Link, *Combined modality treatment with low-dose radiation and MOPP chemotherapy for children with Hodgkin's disease*. *J Clin Oncol*, 1987. **5**(5): p. 742-9.
43. Landgren, O., et al., *Risk of second malignant neoplasms among lymphoma patients with a family history of cancer*. *Int J Cancer*, 2007. **120**(5): p. 1099-102.
44. Koontz, B.F., et al., *Combined-modality therapy versus radiotherapy alone for treatment of early-stage Hodgkin's disease: cure balanced against complications*. *J Clin Oncol*, 2006. **24**(4): p. 605-11.
45. Yahalom, J., *Breast cancer after Hodgkin disease: hope for a safer cure*. *Jama*, 2003. **290**(4): p. 529-31.
46. Yeazel, M.W., et al., *The cancer screening practices of adult survivors of childhood cancer: a report from the Childhood Cancer Survivor Study*. *Cancer*, 2004. **100**(3): p. 631-40.
47. Thomas, D.C., *Genetic epidemiology with a capital "E"*. *Genet Epidemiol*, 2000. **19**(4): p. 289-300.
48. Bernstein, L., *Epidemiology of endocrine-related risk factors for breast cancer*. *J Mammary Gland Biol Neoplasia*, 2002. **7**(1): p. 3-15.
49. McPherson, K., C.M. Steel, and J.M. Dixon, *ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics*. *Bmj*, 2000. **321**(7261): p. 624-8.
50. Byrne, J., et al., *Early menopause in long-term survivors of cancer during adolescence*. *Am J Obstet Gynecol*, 1992. **166**(3): p. 788-93.
51. Rossouw, J.E., et al., *Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial*. *Jama*, 2002. **288**(3): p. 321-33.
52. *Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies*. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet*, 1996. **347**(9017): p. 1713-27.
53. Hill, D.A., et al., *Breast cancer risk following radiotherapy for Hodgkin lymphoma: modification by other risk factors*. *Blood*, 2005. **106**(10): p. 3358-65.
54. Land, C.E., et al., *A case-control interview study of breast cancer among Japanese A-bomb survivors. II. Interactions with radiation dose*. *Cancer Causes Control*, 1994. **5**(2): p. 167-76.
55. Singletary, K.W. and S.M. Gapstur, *Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms*. *Jama*, 2001. **286**(17): p. 2143-51.
56. Poschl, G. and H.K. Seitz, *Alcohol and cancer*. *Alcohol Alcohol*, 2004. **39**(3): p. 155-65.

57. Reichman, M.E., et al., *Effects of alcohol consumption on plasma and urinary hormone concentrations in premenopausal women*. J Natl Cancer Inst, 1993. **85**(9): p. 722-7.
58. Coutelle, C., et al., *Risk factors in alcohol associated breast cancer: alcohol dehydrogenase polymorphism and estrogens*. Int J Oncol, 2004. **25**(4): p. 1127-32.
59. Duthie, S.J., *Folic acid deficiency and cancer: mechanisms of DNA instability*. Br Med Bull, 1999. **55**(3): p. 578-92.
60. Sala, E., et al., *High-risk mammographic parenchymal patterns, hormone replacement therapy and other risk factors: a case-control study*. Int J Epidemiol, 2000. **29**(4): p. 629-36.
61. Offit, K., et al., *Rare variants of ATM and risk for Hodgkin's disease and radiation-associated breast cancers*. Clin Cancer Res, 2002. **8**(12): p. 3813-9.
62. Gaffney, D.K., et al., *Breast cancer after mantle irradiation for Hodgkin's disease: correlation of clinical, pathologic, and molecular features including loss of heterozygosity at BRCA1 and BRCA2*. Int J Radiat Oncol Biol Phys, 2001. **49**(2): p. 539-46.
63. Hasse, D., Binder, C., Bunger, J., Fonatsch, C., Strubel, B., Schnittger, S., et al., *Increased risk of therapy-associated hematologic malignancies in patients with carcinoma of the breast and combined homozygous gene deletions of glutathione transferases M1 and T1*. Leukemia Research, 2002. **26**: p. 249-254.
64. Aston, C.E., *Oligogenic combinations associated with breast cancer risk in women under 53 years of age*. Human Genetics, 2005. **116**: p. 208-221.
65. Farrall, M. and A.P. Morris, *Gearing up for genome-wide gene-association studies*. Hum Mol Genet, 2005. **14 Spec No. 2**: p. R157-62.
66. Wu, L., P.M. Williams, and W.H. Koch, *Clinical applications of microarray-based diagnostic tests*. Biotechniques, 2005. **39**(4): p. 577-82.
67. Dalma-Weiszhausz, D.D., et al., *The affymetrix GeneChip platform: an overview*. Methods Enzymol, 2006. **410**: p. 3-28.
68. Allison, D.B., et al., *Microarray data analysis: from disarray to consolidation and consensus*. Nat Rev Genet, 2006. **7**(1): p. 55-65.
69. Li, J. and M. Burmeister, *Genetical genomics: combining genetics with gene expression analysis*. Hum Mol Genet, 2005. **14 Spec No. 2**: p. R163-9.
70. Rabbee, N.S., T.P., *A genotype calling algorithm for Affymetrix SNP arrays*. Bioinformatics, 2006. **22**: p. 7-12.
71. Breslow, N.E. and N.E. Day, *Statistical methods in cancer research. Volume I - The analysis of case-control studies*. IARC Sci Publ, 1980(32): p. 5-338.
72. Long, A.D., et al., *Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12*. J Biol Chem, 2001. **276**(23): p. 19937-44.
73. Pounds, S. and C. Cheng, *Robust estimation of the false discovery rate*. Bioinformatics, 2006. **22**(16): p. 1979-87.
74. Barry, W.T., A.B. Nobel, and F.A. Wright, *Significance analysis of functional categories in gene expression studies: a structured permutation approach*. Bioinformatics, 2005. **21**(9): p. 1943-9.

75. Benjamini, Y., & Hochberg, Y, *Controlling the false discovery rate: practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society Series, 1995. **B(57)**: p. 289-300.
76. Reiner, A., Yekutieli, D., & Benjamini, Y, *Identifying differentially expressed genes using false discovery rate controlling procedure*. Bioinformatics, 2003. **19**: p. 368-375.
77. Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H., Li, C., *dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data*. Bioinformatics, 2004. **20**: p. 1233-40.
78. Mullighan, C.G., et al., *Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia*. Nature, 2007. **446(7137)**: p. 758-64.
79. Thomas, F., *Automated road segmentation using a bayesian algorithm*. Journal of Transportation Engineering, 2005. **131**: p. 591-598.
80. Thomas, F., *Statistical approach to road segmentation*. Journal of Transportation Engineering, 2003. **129**: p. 300-308.
81. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG*. Nucleic Acids Res, 2006. **34(Database issue)**: p. D354-7.
82. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28(1)**: p. 27-30.
83. Ostell, J., *The entrez search and retrieval system*. In The NCBI Handbook. 2003, Bethesda: National Library of Medicine.
84. Bird, A., *DNA methylation patterns and epigenetic memory*. Genes Dev, 2002. **16(1)**: p. 6-21.
85. Baylin, S.B., *DNA methylation and gene silencing in cancer*. Nat Clin Pract Oncol, 2005. **2 Suppl 1**: p. S4-11.
86. Esteller, M., *Cancer epigenetics: DNA methylation and chromatin alterations in human cancer*. Adv Exp Med Biol, 2003. **532**: p. 39-49.
87. Agrawal, A., R.F. Murphy, and D.K. Agrawal, *DNA methylation in breast and colorectal cancers*. Mod Pathol, 2007. **20(7)**: p. 711-21.
88. Ross, S.A., *Diet and DNA methylation interactions in cancer prevention*. Ann N Y Acad Sci, 2003. **983**: p. 197-207.
89. Dreosti, I.E., *Nutrition, cancer, and aging*. Ann N Y Acad Sci, 1998. **854**: p. 371-7.
90. Duthie, S.J., et al., *Impact of folate deficiency on DNA stability*. J Nutr, 2002. **132(8 Suppl)**: p. 2444S-2449S.
91. Splan, K.E., et al., *In vitro assays for the determination of aminoacyl-tRNA synthetase editing activity*. Methods, 2008. **44(2)**: p. 119-28.
92. Jacobson, E.L., W.M. Shieh, and A.C. Huang, *Mapping the role of NAD metabolism in prevention and treatment of carcinogenesis*. Mol Cell Biochem, 1999. **193(1-2)**: p. 69-74.
93. Servant, L., et al., *A role for DNA polymerase beta in mutagenic UV lesion bypass*. J Biol Chem, 2002. **277(51)**: p. 50046-53.
94. Sinilnikova, O.M., et al., *Acetyl-CoA carboxylase alpha gene and breast cancer susceptibility*. Carcinogenesis, 2004. **25(12)**: p. 2417-24.
95. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers*. Science, 2006. **314(5797)**: p. 268-74.

96. Liu, J., et al., *Inhibition of breast and brain cancer cell growth by BCCIPalpha, an evolutionarily conserved nuclear protein that interacts with BRCA2*. *Oncogene*, 2001. **20**(3): p. 336-45.
97. Moynahan, M.E., A.J. Pierce, and M. Jasin, *BRCA2 is required for homology-directed repair of chromosomal breaks*. *Mol Cell*, 2001. **7**(2): p. 263-72.
98. Daniels, M.J., et al., *Abnormal cytokinesis in cells deficient in the breast cancer susceptibility protein BRCA2*. *Science*, 2004. **306**(5697): p. 876-9.
99. Lu, H., Guo, X., Meng, X., Liu, J., Allen, C., Wray, J., et al., *The BRCA2-interacting protein BCCIP functions in RAD51 and BRCA2 focus formation and homologous recombinational repair*. *Mol Cell Biol*, 2005. **25**: p. 1949-1957.
100. Meng, X., et al., *Abrogation of the transactivation activity of p53 by BCCIP down-regulation*. *J Biol Chem*, 2007. **282**(3): p. 1570-6.
101. Meng, X., J. Fan, and Z. Shen, *Roles of BCCIP in chromosome stability and cytokinesis*. *Oncogene*, 2007. **26**(43): p. 6253-60.
102. Abdelhaleem, M., *Do human RNA helicases have a role in cancer?* *Biochim Biophys Acta*, 2004. **1704**(1): p. 37-46.
103. Alli, Z., M. Ho, and M. Abdelhaleem, *Expression of DHX32 in lymphoid tissues*. *Exp Mol Pathol*, 2005. **79**(3): p. 219-23.
104. Shagisultanova, E.I., et al., *The matrix metalloproteinase-21 gene 572C/T polymorphism and the risk of breast cancer*. *Anticancer Res*, 2004. **24**(1): p. 199-201.
105. Ahokas, K., et al., *Matrix metalloproteinase-21, the human orthologue for XMMP, is expressed during fetal development and in cancer*. *Gene*, 2002. **301**(1-2): p. 31-41.
106. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome*. *Science*, 2004. **305**(5683): p. 525-8.
107. Nozawa, M., Y. Kawahara, and M. Nei, *Genomic drift and copy number variation of sensory receptor genes in humans*. *Proc Natl Acad Sci U S A*, 2007. **104**(51): p. 20421-6.
108. Redon, R., et al., *Global variation in copy number in the human genome*. *Nature*, 2006. **444**(7118): p. 444-54.
109. Ache, B.W. and J.M. Young, *Olfaction: diverse species, conserved principles*. *Neuron*, 2005. **48**(3): p. 417-30.
110. Niimura, Y. and M. Nei, *Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates*. *J Hum Genet*, 2006. **51**(6): p. 505-17.
111. Nei, M., *The new mutation theory of phenotypic evolution*. *Proc Natl Acad Sci U S A*, 2007. **104**(30): p. 12235-42.
112. Locke, D.P., et al., *BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications*. *J Med Genet*, 2004. **41**(3): p. 175-82.
113. Jalal, S.M., et al., *Form of 15q proximal duplication appears to be a normal euchromatic variant*. *Am J Med Genet*, 1994. **52**(4): p. 495-7.
114. Mignon, C., et al., *Inherited DNA amplification of the proximal 15q region: cytogenetic and molecular studies*. *J Med Genet*, 1997. **34**(3): p. 217-22.

115. Fantes, J.A., et al., *Organisation of the pericentromeric region of chromosome 15: at least four partial gene copies are amplified in patients with a proximal duplication of 15q*. J Med Genet, 2002. **39**(3): p. 170-7.
116. Barber, J.C., *Directly transmitted unbalanced chromosome abnormalities and euchromatic variants*. J Med Genet, 2005. **42**(8): p. 609-29.
117. Dyomin, V.G., et al., *BCL8, a novel gene involved in translocations affecting band 15q11-13 in diffuse large-cell lymphoma*. Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5728-32.
118. Chen, Y., Choong, L, Lin, Q., Philp, R, Wong, C., Ang, B., Tan, Y., et al., *Differential expression of noval tyrosine kinase substrates during breast cancer development*. Molecular and Cellular Proteomics, 2007. **6**(12): p. 2072-2087.
119. Huang, Z., *Stress signaling and Myc downregulation: implications for cancer*. Cell Cycle, 2004. **3**(5): p. 593-6.
120. Huang, Z., J.A. Traugh, and J.M. Bishop, *Negative control of the Myc protein by the stress-responsive kinase Pak2*. Mol Cell Biol, 2004. **24**(4): p. 1582-94.
121. Koch, J.G., et al., *Mammary tumor modifiers in BALB/cJ mice heterozygous for p53*. Mamm Genome, 2007. **18**(5): p. 300-9.
122. Lissowska, J., et al., *Genetic polymorphisms in the one-carbon metabolism pathway and breast cancer risk: a population-based case-control study and meta-analyses*. Int J Cancer, 2007. **120**(12): p. 2696-703.

APPENDIX A. INSTITUTIONAL REVIEW BOARD APPROVAL



October 11, 2006

Belinda Mandrell, RN, PhD Candidate
Diagnostic Imaging

Dear Ms. Mandrell:

re: **MPBC11D** – *The Molecular Profile Of Women With And Without Secondary Breast Cancer After Treatment Of Pediatric Hodgkin Disease*

This is to certify that, on 10/10/06, the

Informed consent documents dated 7/25/06 and the Memorandum dated 9/29/06 from Ms Mandrell in response to the 9/19/06 IRB review of the new protocol (Initial version) dated 7/25/06

submitted to the Institutional Review Board for consideration were reviewed by the IRB regarding human research as adequately protecting the rights and welfare of participants, using adequate methods of securing informed consent from these individuals and not involving undue risk in relation to potential benefits to be derived therefrom.

IRB Review Status:

The protocol (Initial version) dated 7/25/06 and informed consent documents dated 7/25/06 are approved as a minimal risk study for a period of one year. This approval is effective 10/10/06 through 10/9/07.

Institutional Review Board Signatory Official

STATEMENT OF PRINCIPAL INVESTIGATOR:

As previously signed and certified, I understand that approval of this research involving human subjects is contingent upon my agreement:

1. To report to the Institutional Review Board for Human Research (IRB) any adverse effect or research related injuries which might occur in relation to the human experimentation. I have read and will comply with IRB reporting guidelines.
2. To submit in writing for prior IRB approval any alterations to the plan of human research.
3. To submit timely continuing review reports of this research as requested by the IRB.
4. To maintain copies of all pertinent information related to the research activities in this project, including copies of informed consent agreements obtained from all participants.
5. To notify the IRB immediately upon the termination of this project, and/or the departure of the principal investigator from this institution and the project.

CC: Pamela Hicks, RN, PhD

CPD/MO
IRB MPBC11D 10/11/06
/sb

APPENDIX B. INTRODUCTORY LETTER

Date

Patient's Address

Dear (Patient's Name),

We are writing to you and other women who were treated at St. Jude Children's Research Hospital for pediatric Hodgkin disease to ask your participation in a new research study. We will describe the study in this letter, and we welcome any questions you may have.

Women who were treated for pediatric Hodgkin disease are at an increased risk for development of breast cancer with the greatest risk occurring 15 to 20 years after Hodgkin disease. Researchers do not yet understand all the factors associated with increasing a women's risk of breast cancer after Hodgkin disease therapy. We would like to learn why some female Hodgkin disease survivors develop breast cancer while other female Hodgkin disease survivors do not. We believe that a better understanding may be obtained by exploring the genetic differences that affect a woman's tolerance to radiation and chemotherapy. Therefore, we have chosen to conduct a research study that will examine and compare the genetic differences between pediatric Hodgkin disease survivors who developed breast cancer to those pediatric Hodgkin disease survivors who have not developed breast cancer.

We will contact you by phone within two weeks after you receive this letter. During that call, we will ask if you would like to participate in this study. If you agree to participate, we will ask you to give verbal consent and complete a 10 item questionnaire over the telephone. The verbal consent and questionnaire conversation will be taped; the tape will not begin until you give taping approval. The tape will be in a locked cabinet and can not be linked with you. The tape will be destroyed after the completion of the study and publication of study findings. We have enclosed the questionnaire for your review.

We will also ask that you read, sign, and return the enclosed consent form. The consent is enclosed along with an explanation of this study. Once we have received your signed consent, we will ask that a blood sample be obtained at your local physician's office, a St. Jude affiliate clinic, or at St. Jude Children's Research Hospital. We will pay for the cost of the blood sample.

Please know that you do not need to take part in this study-it is voluntary. Please feel free to contact either of us at the hospital with any questions about this study (901) 495-4209. We would be glad to have the opportunity to discuss it with you. Our very best to you always.

Belinda Mandrell, PhD candidate, RN

Melissa Hudson, MD

RESEARCH PARTICIPANT STATEMENT

APPENDIX C. INFORMED CONSENT STATEMENT

THE MOLECULAR PROFILE OF WOMEN WITH AND WITHOUT SECONDARY BREAST CANCER AFTER TREATMENT OF PEDIATRIC HODGKIN DISEASE

You are being asked to take part in a research study because you were treated for pediatric Hodgkin disease. This study will help researchers understand more about why some women treated for Hodgkin disease develop treatment-related breast cancer while other women do not. This study is not testing for a genetic or inherited disease (passed on in families) but is looking at genetic differences that may help explain how the body responds to radiation and chemotherapy. This research approach may help identify risk factors for secondary breast cancer at the time of Hodgkin disease diagnosis. If so, that information could be used to adjust Hodgkin therapy.

This consent form gives you information about the study, which will be discussed with you. Once you understand the study, and if you agree to take part, you will be asked to sign this consent form. You will be given a copy of the consent form to keep.

Before you learn about the study, it is important that you know the following:

- Whether or not you take part in this study is entirely up to you.
- You may decide not to be in the study, or to withdraw from the study at any time.
- This study is being sponsored by the American Cancer Society, National Institutes of Health, and St. Jude Children's Research Hospital. The sponsors will receive anonymous information about your treatment characteristics as related to the study.

Why is this study being done?

We now know that after treatment for pediatric Hodgkin disease, patients can have problems with late effects from Hodgkin therapy. For women treated for pediatric Hodgkin disease, the risk of developing breast cancer is higher than among the general population. Doctors caring for Hodgkin disease survivors are not sure why some women develop breast cancer and others do not. The purpose of this research study is to compare genes in women treated for pediatric Hodgkin disease who develop breast cancer to the genes of women treated for pediatric Hodgkin disease who did not develop breast cancer.

The study has the following primary goals:

1. To assess for a difference in the RNA (gene appearance) of women who received chest radiation for pediatric Hodgkin disease and later developed breast cancer or did not.

2. To assess for a difference in the DNA of women who received chest radiation for pediatric Hodgkin disease who later developed breast cancer with the DNA of those women who did not develop breast cancer.
3. To assess if genes may be identified through the study of RNA and DNA that may represent a women's higher risk of developing breast cancer after Hodgkin disease therapy.

The study has the following secondary goal:

1. To assess for any association between the use of hormones, childbirth, and family history and the development of secondary breast cancer.

How many research participants will take part in the study?

Forty eight (48) women treated at St. Jude Children's Research Hospital, from 1970-1991, for pediatric Hodgkin disease who subsequently did or did not develop breast cancer will be asked to participate in the research study.

What is involved in this study?

As a survivor of pediatric Hodgkin disease you are invited to participate in this study. This research study has two parts:

- 1). A telephone interview asking questions about
 - a) your use of hormones for birth control, fertility or ovarian failure
 - b) pregnancies
 - c) family cancer history
- 2). Collection of a blood sample.

You will be ask to join this study only after you have been fully informed by way of the enclosed letter and telephone conversation with the primary researcher, your questions have been answered, and you have verbalized an understanding of the study. After obtaining your verbal consent during the phone conversation, you will be asked to respond to 10 questions regarding your use of hormones for birth control, ovarian failure, or fertility. You will also be asked your history/date of pregnancies and your family history of cancer. You will also need to give your written consent using the attached consent forms. These forms will need to be returned to St. Jude in the enclosed envelope. You will also be instructed on blood collection. The blood collection will require 11 mls (2 teaspoons) of blood. This blood may be collected at St. Jude, an affiliate clinic, or at your local doctor's office. All supplies and instructions will be forwarded to your doctor and the blood will be returned to St. Jude by express mail. Participants will only be enrolled once in this study. All samples will be mailed via Federal Express to the RNA/DNA extraction lab of Dr. Mary Relling, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105. You will not pay for any of these costs. All costs will be paid from a grant at St. Jude. The consent process will be documented in the

medical record as per institutional guidelines. We also ask you to allow us to save any left over DNA not used in this study. We seek your permission to store the left over DNA for potential future research questions regarding pediatric Hodgkin disease and breast cancer. This DNA will not be stored with your name and will only be identified by an assigned number. Your DNA will be used only for research and will not be sold.

How long will I be in the study?

Once you have submitted the blood, you will be considered off study. You will receive a follow-up letter after the study is completed with an abstract of the study, in which findings from the study will be described. If you have further questions after reading these findings, you may call the primary research investigator, Belinda Mandrell at (901) 495-4209.

What research studies will be done?

DNA and RNA will be extracted from your blood. The DNA will be analyzed looking for common differences between participants with a technique called Single Nucleotide Polymorphisms (SNPs). The RNA will be analyzed looking at differences in gene expression using a technique, microarray analysis. This is not a diagnostic technique and is a research analysis only.

What will I learn about the results of this study?

Your individual test results are not linked to your personal information, and are unavailable to the investigators, we are unable to report them back to you.

What are the risks of the study?

The researchers in this study will protect your records and blood sample and insure that your samples will be labeled without your name and will not be directly linked to you. The results of this research will not be added to your St. Jude record.

The risk from blood collection will be related to having a needle put into a vein which may cause a feeling of faintness, pain, bruising, and a minimal chance of infection.

What are the benefits of this study?

This research approach may help identify risk factors for breast cancer at the time of Hodgkin disease diagnosis and could result in adjustment of Hodgkin therapy. The future benefit of participation in this study is for others who will be learning more about what causes cancer, how to prevent second cancers, and how to offer better treatment. This research may help children with Hodgkin disease in the future.

What are the consequences of not joining the study?

You can choose to not join or to stop participating in the study at any time. Your participation is voluntary. The alternative to participating in this study is not to participate. You may contact us anytime at (901) 495-4209 and tell us you do not want

your information or blood used for this study. Your information and your blood specimen will be discarded and no longer used for research.

Compensation

There is no cost to participate in this study. There is no payment for participation.

Confidentiality

Your medical records will be kept confidential to the degree allowed by law. Information from your medical records will not be given to anyone outside the hospital unless you agree. You will not be identified in any publication about this study.

Government agencies oversee research studies involving people. Your medical records may be reviewed by such agencies if you take part in this research study. These agencies may include the Food and Drug Administration (FDA) and the National Cancer Institute (NCI). Anonymous information regarding your treatment characteristics and the study results may also be shared with researchers from other institutions that are involved in this study. These include the American Cancer Society and the National Institutes of Health. By signing this consent form, you are allowing your medical records to be reviewed by these persons.

It may be necessary to check parts of your medical record to be sure that your original treatment for Hodgkin disease is recorded correctly and completely. Such a check might be done by the following groups:

1. A federal agency such as the Food and Drug Administration (FDA) or the National Cancer Institute (NCI).
2. St. Jude Children's Research Hospital Institutional Review Board, a committee that reviews the ethics of research studies.
3. American Cancer Society

No information other than what is needed for the study is recorded. Every effort is made to protect your privacy.

APPENDIX D. SUMMARY OF RESEARCH AND PRIVACY RIGHTS

Non-Therapeutic Research

IRB version: November 4, 2003

I have been told the following information:

1. I may talk as much as I want to with the doctors who are responsible for my care about the reasons for this study and about its risks.
2. This study may have risks that the researchers do not know about now.
3. There will be no additional costs to me from taking part in this research study. It is the policy of St. Jude Children's Research Hospital not to bill research participants or their families for the cost of medical care. This includes any costs associated with taking part in research studies.
4. I and my family will not receive any compensation or payment of any kind for being in this study, or for any treatments, products, or any other things of value that may result from this study.
5. The hospital policy is not to provide payment if I am injured or damaged by being in this study
6. I have been told that I can withdraw from this study at any time.
7. I received a copy of the St. Jude Notice of Privacy Practices at the time of registration. That document tells me how my medical information may be used or disclosed (given to someone outside the hospital). I have been told I have the right to review the Notice of Privacy Practices before I sign this form. I have been told that I may request another copy of the Notice of Privacy Practices and that it is posted in the Hospital and on our website (www.stjude.org).
8. I have the right to inspect, copy and/or amend (change) my protected health information that is to be used or disclosed. I have been told in this consent form about any limitations to this right, such as research information that I will not have access to until the end of the study or that will be used strictly for research purposes.
9. My protected health information will be disclosed to or used by the following:
 - Investigators and co-investigators of research study
 - American Cancer Society
 - National Institutes of Health

10. My records may also be reviewed by agencies such as the Food and Drug Administration or the National Institutes of Health, or other agencies as required by state or federal regulations.
11. Information about me that may be disclosed includes the following:
- Complete medical record including information regarding diagnosis, illness, treatment, and information that may be recorded about previous diagnosis or treatment.
 - Information gathered as a part of this research study as explained in the informed consent/authorization.
12. I have been told that once my records are disclosed to or used by others, St. Jude Children's Research Hospital cannot guarantee that information will not be further disclosed. Also, the released information may no longer be protected by federal privacy regulations.
13. I have been told that this authorization for the use and disclosure of my protected health information does not expire.
14. I have been told that I may withdraw my authorization for the disclosure or use of my records at anytime, for any reason, with the following exceptions:
- When that information has already been disclosed or used based on my permission
 - When the information is required to maintain the integrity of the study
15. To withdraw my authorization, I must complete a Revocation of Release of Authorization form. I have been told I may request this form at St. Jude Children's Research Hospital by calling the Privacy Officer at 901-495-2341. The form must be returned by mail or hand delivery to:
- HIPAA Privacy Officer
St. Jude Children's Research Hospital
332 N. Lauderdale
Memphis, TN 38105
16. I know that if I have more questions about this study or about any injury from the treatment, I can call the Principal Investigator of this study, Belinda Mandrell at 901/495-3300.
17. I know that I can get more information about my rights as a research participant by calling the Chairman of the Institutional Review Board at 901/495-4357 or the Research Participant Advocate (Ombudsman) at 495-4644. If I am outside of the Memphis area, I may call 1-866-583-3472 (1-866-JUDE IRB). This is a toll-free call.
18. I will receive a copy of this signed statement.

19. I have asked to allow the principal investigator to store any left over DNA for potential future research questions regarding pediatric Hodgkin disease and/or breast cancer. The DNA will not be stored with my name and will be identified by an assigned number

//_____ I agree to have my DNA stored for potential future research

//_____ I do not agree to have my DNA stored for potential future research

PLEASE FAX WITH CONSENT SIGNATURE PAGE TO PROTOCOL OFFICE #6265

I have read (or have had read to me) the contents of this document and have been encouraged to ask questions. I have received answers to my questions. I give consent to take part in this research study and authorize the disclosure and use of my protected health information for the purposes of that research.

_____ Date _____ Time (AM/PM)
Research Participant

PHYSICIAN/INVESTIGATOR/DESIGNEE STATEMENT

I hereby certify that I have discussed the research project with the research participant. I have explained all the information contained in the informed consent document, including any risks that may be reasonably expected to occur. I further certify that the research participant was encouraged to ask questions and that all questions were answered.

_____ Date _____ Time (AM/PM)
Physician/Investigator/Designee

In case of questions or emergencies in reference to this protocol, please contact:
St. Jude Children's Research Hospital
332 North Lauderdale
Memphis, TN 38105
(901) 495-3300 - FAX Interdepartmental #6265

PLEASE FAX CONSENT FORM TO PROTOCOL OFFICE #6265

APPENDIX E. DEMOGRAPHIC QUESTIONNAIRE OF HODGKIN DISEASE SURVIVORS

1. Do you have menstrual periods? Yes No
2. What was your age at your first period? _____
3. What was your age at your last period or menopause? _____
4. Have you used hormone (estrogen/progesterone) medication for birth control, to stop bleeding, to regulate your periods, or help with symptoms of menopause? Yes No

If yes, give date or age of starting hormone use and reason of use _____

Are you currently on hormone medication? Yes No

If yes, give name of current hormone medication _____

If no, Please give the general date for starting and stopping hormone medication

Started _____

Stopped _____

5. Have you ever been pregnant?
 Number of pregnancies _____, Number of live births _____

Date(s)of Birth	Did You Breast Feed, if yes how long
1.	
2.	
3.	
4.	

6. Please describe your family history of cancer, give blood relative (grandmother, grandfather, mother, father, aunt, uncle, sister, brother, children). Also note if the relative is from your mother's (maternal) or father's (paternal) side of the family.

Relative	Type of Cancer	Age at Cancer Diagnosis	Living or Deceased
1.			
2.			
3.			
4.			
5.			
6.			
7.			

7. Have you had a breast exam by a doctor within the last year? Yes No
8. Have you had a mammogram within the last year? Yes No
9. Have you been diagnosed with any other cancers since treatment of pediatric Hodgkin Disease? Yes No

If yes, list cancer _____

10. Please answer if you have been diagnosed with breast cancer after Hodgkin Disease treatment.

a. What was the breast cancer histology?

Adenocarcinoma Infiltrating Ductal Carcinoma
 Ductal In-Situ Carcinoma Other Not sure

b. Location of breast cancer: Right Breast Left Breast Both Breasts

c. Was the breast cancer positive for any or all of the following?

Estrogen Progesterone HER-2 Dont' know

d. What was your treatment for breast cancer (check all that apply to you)

Surgery Chemotherapy Radiation Hormone Therapy Other

e. Last date of breast cancer treatment _____

f. Are you currently on hormone therapy for breast cancer prevention or treatment?

Yes No

APPENDIX F. SINGLE NUCLEOTIDE POLYMORPHISMS OF INTEREST

probeid	test.stat	test.p	perm.p	q	Probe.Set.ID	Gene Symbol	Chromosomal Location
219812_at	3.985615	6.73E-05	1.00E-04	0.6786	219812_at	MGC2463	chr7q22.1
1559777_at	-3.69122	0.000223	1.00E-04	0.6786	1559777_at	---	---
207873_x_at	-3.55535	0.000377	1.00E-04	0.6786	207873_x_at	SEZ6L	chr22q12.1
237143_at	3.55535	0.000377	1.00E-04	0.6786	237143_at	---	---
1555243_x_at	3.419477	0.000627	1.00E-04	0.6786	1555243_x_at	C8orf59	chr8q21.2
220418_at	3.408154	0.000654	1.00E-04	0.6786	220418_at	IFT52 /// UBASH3A	chr21q22.3
219134_at	-3.39683	0.000682	1.00E-04	0.6786	219134_at	ELTD1	chr1p33-p32
1560661_x_at	3.374186	0.00074	2.00E-04	0.6786	1560661_x_at	KIAA1641	chr2q11.2
201661_s_at	-3.35154	0.000804	2.00E-04	0.6786	201661_s_at	ACSL3	chr2q34-q35
230319_at	-3.26096	0.00111	2.00E-04	0.6786	230319_at	---	---
242368_at	-3.66858	0.000244	3.00E-04	0.6786	242368_at	GTF2I	chr7q11.23
236503_at	3.464768	0.000531	3.00E-04	0.6786	236503_at	LOC38852 6	chr19q12
219868_s_at	-3.44212	0.000577	3.00E-04	0.6786	219868_s_at	ANKFY1	chr17p13.3
211328_x_at	-3.2836	0.001025	3.00E-04	0.6786	211328_x_at	HFE	chr6p21.3
45572_s_at	3.283603	0.001025	3.00E-04	0.6786	45572_s_at	GGA1	chr22q13.31
212480_at	3.600641	0.000317	4.00E-04	0.6786	212480_at	KIAA0376	chr22q11.23
1566094_at	-3.48741	0.000488	4.00E-04	0.6786	1566094_at	ARHGEF1 2	chr11q23.3
203424_s_at	-3.32889	0.000872	4.00E-04	0.6786	203424_s_at	IGFBP5	chr2q33-q36
207264_at	-3.60064	0.000317	0.0004	0.6786	207264_at	KDEL3	chr22q13.1
221790_s_at	3.464768	0.000531	0.0004	0.6786	221790_s_at	LDLRAP1	chr1p36-p35
236341_at	3.453445	0.000553	0.0004	0.6786	236341_at	CTLA4	chr2q33
211219_s_at	-3.35154	0.000804	0.0004	0.6786	211219_s_at	LHX2	chr9q33-q34.1
221973_at	3.623287	0.000291	0.0005	0.6786	221973_at	LOC15075 9	chr2q11.2
219751_at	3.453445	0.000553	0.0005	0.6786	219751_at	SETD6	chr16q21

probeid	test.stat	test.p	perm.p	q	Probe.Set.ID	Gene Symbol	Chromosomal Location
225391_at	3.419477	0.000627	0.0005	0.6786	225391_at	LOC93622	chr4p16.1
230974_at	3.419477	0.000627	0.0005	0.6786	230974_at	DDX19B	chr16q22.1
242208_at	3.374186	0.00074	0.0005	0.6786	242208_at	ZNF37B	chr10q11.21
210383_at	-3.34022	0.000837	0.0005	0.6786	210383_at	SCN1A	chr2q24.3
236436_at	3.340217	0.000837	0.0005	0.6786	236436_at	LOC283130	chr11q13.1
204890_s_at	3.328894	0.000872	0.0005	0.6786	204890_s_at	LCK	chr1p34.3
229495_at	-3.19302	0.001408	0.0005	0.6786	229495_at	ACY1L2	chr6q15
243479_at	-3.4308	0.000602	6.00E-04	0.6786	243479_at	MYST4	chr10q22.2
204204_at	-3.35154	0.000804	0.0006	0.6786	204204_at	SLC31A2	chr9q31-q32
233539_at	3.328894	0.000872	0.0006	0.6786	233539_at	NAPE-PLD	chr7q22.1
200013_at	3.215667	0.001301	0.0006	0.6786	200013_at	RPL24	chr3q12
238099_at	3.193021	0.001408	0.0006	0.6786	238099_at	---	---
205353_s_at	3.170376	0.001522	0.0006	0.6786	205353_s_at	PEBP1	chr12q24.23
227412_at	3.306249	0.000946	0.0007	0.6786	227412_at	PPP1R3E	chr14q11.2
224162_s_at	3.181698	0.001464	0.0007	0.6786	224162_s_at	FBXO31	chr16q24.2
227784_s_at	3.170376	0.001522	0.0007	0.6786	227784_s_at	COG1	chr17q25.1
225696_at	3.102439	0.001919	0.0007	0.6786	225696_at	COPS7B	chr2q37.1
225562_at	3.487413	0.000488	0.0008	0.6786	225562_at	RASA3	chr13q34
212186_at	3.419477	0.000627	0.0008	0.6786	212186_at	ACACA	chr17q21
1561443_at	-3.37419	0.00074	0.0008	0.6786	1561443_at	---	---
217807_s_at	3.14773	0.001645	0.0008	0.6786	217807_s_at	GLTSCR2	chr19q13.3
233771_at	3.14773	0.001645	0.0008	0.6786	233771_at	TRIO	chr5p15.1-p14
216100_s_at	-3.26096	0.00111	9.00E-04	0.6786	216100_s_at	TOR1AIP1	chr1q24.2
227177_at	3.260958	0.00111	9.00E-04	0.6786	227177_at	CORO2A	chr9q22.3
239018_at	3.260958	0.00111	9.00E-04	0.6786	239018_at	WBSCR18	chr7q11.23
208714_at	3.14773	0.001645	9.00E-04	0.6786	208714_at	NDUFV1	chr11q13

probeid	test.stat	test.p	perm.p	q	Probe.Set.ID	Gene Symbol	Chromosomal Location
235697_at	3.14773	0.001645	9.00E-04	0.6786	235697_at	ZNF544	chr19q13.43
1555079_at	-3.12508	0.001778	9.00E-04	0.6786	1555079_at	C9orf68	chr9p24.2-p24.1
1557258_a_at	-3.19302	0.001408	0.0009	0.6786	1557258_a_at	BCL10	chr1p22
216166_at	3.079794	0.002071	0.0009	0.6786	216166_at	RHOH	chr4p13
234339_s_at	3.35154	0.000804	0.001	0.6786	234339_s_at	GLTSCR2	chr19q13.3
221780_s_at	3.260958	0.00111	0.001	0.6786	221780_s_at	DDX27	chr20q13.13
1552712_a_at	-3.21567	0.001301	0.001	0.6786	1552712_a_at	NMNAT2	chr1q25
227915_at	3.215667	0.001301	0.001	0.6786	227915_at	ASB2	chr14q31-q32
221318_at	-3.1817	0.001464	0.001	0.6786	221318_at	NEUROD4	chr12q13.2
1558987_at	3.170376	0.001522	0.001	0.6786	1558987_at	FLJ40473	chr3q21.3
221966_at	-3.14773	0.001645	0.001	0.6786	221966_at	GPR137	chr11cen-q22.3
229813_x_at	3.14773	0.001645	0.001	0.6786	229813_x_at	DAZAP1	chr19p13.3
234116_at	-3.14773	0.001645	0.001	0.6786	234116_at	CX40.1	chr10p11.21
1563539_at	-3.12508	0.001778	0.001	0.6786	1563539_at	---	---
235784_at	3.125085	0.001778	0.001	0.6786	235784_at	MTERF	chr7q21-q22
203245_s_at	3.260958	0.00111	0.0011	0.6786	203245_s_at	FLJ35348	chr9q34
230248_x_at	3.260958	0.00111	0.0011	0.6786	230248_x_at	---	---
208875_s_at	-3.07979	0.002071	0.0011	0.6786	208875_s_at	PAK2	chr3q29
1561442_at	-3.2836	0.001025	0.0012	0.6786	1561442_at	LOC283585	chr14q31.3
211921_x_at	3.260958	0.00111	0.0012	0.6786	211921_x_at	PTMA	chr2q35-q36
1563101_at	-3.20434	0.001354	0.0012	0.6786	1563101_at	ARNTL2	chr12p12.2-p11.2
226433_at	3.204344	0.001354	0.0012	0.6786	226433_at	RNF157	chr17q25.1
222198_at	-3.11376	0.001847	0.0012	0.6786	222198_at	---	---
243854_at	-3.07979	0.002071	0.0012	0.6786	243854_at	APLP2	chr11q23-q25 11q24
213157_s_at	-3.05715	0.002235	0.0012	0.6786	213157_s_at	KIAA0523	chr17p13.2
205079_s_at	-3.01186	0.002597	0.0012	0.6786	205079_s_at	MPDZ	chr9p24-p22

probeid	test.stat	test.p	perm.p	q	Probe.Set.ID	Gene Symbol	Chromosomal Location
221208_s_at	3.306249	0.000946	0.0013	0.6786	221208_s_at	C11orf61	chr11q24.2
1558769_s_at	-3.26096	0.00111	0.0013	0.6786	1558769_s_at	DNAH1	chr3p21.1
1560741_at	3.170376	0.001522	0.0013	0.6786	1560741_at	SNRPN	chr15q11.2
1561604_at	3.14773	0.001645	0.0013	0.6786	1561604_at	---	---
211796_s_at	3.14773	0.001645	0.0013	0.6786	211796_s_at	TRBV21-1 /// TRBV19 /// TRBV5-4 /// TRBV3-1 /// TRBC1	chr7q34
238528_at	-3.12508	0.001778	0.0013	0.6786	238528_at	UBR1	chr15q13
243014_at	-3.09112	0.001994	0.0014	0.6786	243014_at	ZNRF3	chr22q12.1
210757_x_at	-3.07979	0.002071	0.0014	0.6786	210757_x_at	DAB2	chr5p13
219947_at	-3.01186	0.002597	0.0014	0.6786	219947_at	CLEC4A	chr12p13
226602_s_at	3.011857	0.002597	0.0014	0.6786	226602_s_at	BCR /// FLJ42953 /// LOC644165 /// LOC653617	chr22q11 22q11.23 /// chr22q11.21 /// chr22q11.23
240942_at	-2.98921	0.002797	0.0014	0.6786	240942_at	HSMPP8	chr13q12.11
203848_at	3.193021	0.001408	0.0015	0.6786	203848_at	AKAP8	chr19p13.1-q12
215068_s_at	3.170376	0.001522	0.0015	0.6786	215068_s_at	FBXL18	chr7p22.2
224669_at	3.170376	0.001522	0.0015	0.6786	224669_at	DBNDD2	chr20q13.12
1566279_at	-3.15905	0.001583	0.0015	0.6786	1566279_at	---	---
233129_at	-3.14773	0.001645	0.0015	0.6786	233129_at	RYR3	chr15q14-q15
207068_at	3.125085	0.001778	0.0015	0.6786	207068_at	ZFP37	chr9q32
225067_at	3.125085	0.001778	0.0015	0.6786	225067_at	ULK3	chr15q24.1
235322_at	3.102439	0.001919	0.0015	0.6786	235322_at	LOC64973 4	---
222668_at	3.057148	0.002235	0.0015	0.6786	222668_at	KCTD15	chr19q13.11
234106_s_at	-3.05715	0.002235	0.0015	0.6786	234106_s_at	FLYWCH1	chr16p13.3
205255_x_at	3.034502	0.002409	0.0015	0.6786	205255_x_at	TCF7	chr5q31.1
209790_s_at	3.011857	0.002597	0.0015	0.6786	209790_s_at	CASP6	chr4q25

VITA

Belinda Neal Mandrell received a Bachelor's Degree in Public Health from The University of Tennessee Knoxville in 1981 and a Diploma in Nursing from St. Mary's Medical Center, Knoxville in 1984. Her career began with the Upper Cumberland Regional Public Health Agency as an educator, nurse manager, and school health nurse. In 1987, she moved to Baltimore, MD where she was the school nurse for a level five special education academic facility within the Kennedy Institute. At the Kennedy Institute, she also worked in the neuro rehabilitation program as a staff nurse. While in Baltimore, she attended the University of Maryland, Baltimore and completed a Bachelor's Degree in Nursing and Master's Degree in Nursing Science as a pediatric nurse practitioner in 1991. As a nurse practitioner, she began her career and remains at St. Jude Children's Research Hospital. As a nurse practitioner, her practice has concentrated on the care of children with leukemia and lymphoma. In 1999, she became the associate director of the Pediatric Oncology Nurse Practitioner Distant Learning Graduate Study Program with the University of Pennsylvania and St. Jude Children's Research Hospital. This program allowed staff nurses the opportunity to acquire a Master's degree on St. Jude campus while continuing to work full time as a staff nurse. Throughout her clinical and academic endeavors at St. Jude, Belinda has had a keen interest in the genetics of cancer and it is this interest that led to her decision for doctoral studies at The University of Tennessee Health Science Center in Memphis.

While in doctoral studies at The University of Tennessee Health Science Center, Belinda continued as clinical instructor for the St. Jude Distant Learning Graduate Program and as a research associate in the division of Nursing Research at St. Jude. During her doctoral studies, she pursued interest specific to genetics and attended the Oncology Nursing Society Genetic Short Course in 2004 and the National Institutes of Health/National Institute of Nursing Research (NIH/NINR) Summer Genetic Institute in 2005. She was awarded dissertation funding from the NIH/NINR Ruth L. Kirschstein NRSA predoctoral fellowship, the American Cancer Society predoctoral nursing scholarship, and St. Jude Children's Research Hospital. She has written numerous manuscripts published in peer-reviewed journals and book chapters and has presented at local, national, and international conferences. Her current position is research associate in the division of Nursing Research, St. Jude Children's Research Hospital. She is a member of Oncology Nursing Society (ONS), International Society of Nurses in Genetics (ISONG), and the Association of Pediatric Hematology/Oncology Nurses (APHON) and is on the editorial board for Cancer Nursing.