Peroxisome Proliferator-Activated Receptor-Alpha Gene, Obesity, and Breast Cancer Incidence and Survival: A LIBCSP Ancillary Study

Amanda Keatley Golembesky

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Public Health (Epidemiology).

Chapel Hill
2008

Approved by:
Advisor: Marilie Gammon
Reader: Jeannette Bensen
Reader: Kari North
Reader: Regina Santella
Reader: Jane Schroeder

© 2008
Amanda Keatley Golembesky
ALL RIGHTS RESERVED

ABSTRACT

Amanda Keatley Golembesky

Peroxisome Proliferator-Activated Receptor-Alpha Gene, Obesity And Breast Cancer Incidence And Survival: A LIBCSP Ancillary Study
(Under the direction of Dr. Marilie Gammon)

Peroxisome proliferator-activated receptor-alpha (PPARA) has been shown to increase fatty acid oxidation and decrease cytokine levels, and has been implicated in insulin production. Genetic variants of *PPARA* have been associated with cardiovascular disease, obesity and type II diabetes mellitus. Although no research to date has investigated the possible link between *PPARA* and breast cancer incidence and survival, the function of this gene suggests that it could play a role in breast cancer development and prognosis. Six *PPARA* polymorphisms were evaluated in association with incident breast cancer (n=1073 cases, n=1112 controls) and survival (n=1073 cases) in the Long Island Breast Cancer Study Project, a population-based case-control study. The National Death Index was used to determine vital status through December 31, 2002. The case-control study analyses used unconditional logistic and multilevel regression, and haplotype-based analyses while the survival analyses employed Kaplan-Meier curves, Cox regression and haplotype-based analyses for all-cause and breast cancer-specific mortality (n = 132 (12.3%) and 88 (8.2%), respectively). The odds of breast cancer were doubled among women with *PPARA* polymorphism rs4253760 (OR=1.97 for rare vs. common homozygote alleles; 95% CI: 1.14, 3.43). This association remained constant with the inclusion of all interrogated polymorphisms studied in hierarchical models. rs4253760 was also associated

with over a two-fold increase in all-cause mortality at time of disease diagnosis with inclusion of a continuous time interaction (HR=2.25 for rare vs. common homozyote alleles; 95% CI: 1.00, 5.08). This beta coefficient for this time interaction is negative, implying that survival is improving over time, so that the HR is equal to 0.69 (95% CI: 0.36, 1.29) following five years of follow-up. Thus, caution is necessary when interpreting the results for this polymorphism. Haplotype analyses did not reveal any differences between cases and controls or survival. Our results are the first to evaluate the relationship between *PPARA* and breast cancer incidence and survival and suggest that replication in an independent cohort is warranted.

To my husband Kurt and my son Dane.
I could not have done this without their love and support.

Thank you to Dr. Marilie Gammon for her leadership throughout this process. I would also like to thank all the members of my dissertation committee for their time and effort for the past four and a half years. Lastly, thank you to the participants of the Long Island Breast Cancer Project for their willingness to participate in this research.

TABLE OF CONTENTS

LIST OF TABLES xiv

LIST OF FIGURES xviii

LIST OF ABBREVIATIONS AND SYMBOLS xx

CHAPTER 1 BODY SIZE/OBESITY 1

1.1 INTRODUCTION 1

1.2 MEASURES OF BODY SIZE 1

1.3 DEFINITION OF BODY SIZE 3

1.4 DETERMINANTS OF OBESITY 4

1.5 TIMING OF WEIGHT CHANGE 5

1.5.1 Menarche and obesity 5

1.5.2 Pregnancy and obesity 6

1.5.3 Menopause and obesity 6

1.6 OBESITY PREVALENCE 7

1.7 METABOLIC CONSEQUENCES OF OBESITY 9

1.7.1 Overview 9

1.7.2 Endocrine disturbances 9

1.7.3 Obesity Associated disease 10

1.8 CONCLUSION 11

CHAPTER 2 BREAST CANCER INCIDENCE 12

2.1 INTRODUCTION 12

2.2 TRENDS IN INCIDENCE 12

2.3 MECHANISM OF ACTION 14

2.4 RISK FACTOR EPIDEMIOLOGY 16
3.4.1 Reproductive factors 16
3.4.2 Exogenous hormones 17
3.4.3 Diet 19
3.4.4 Environmental Factors 20
3.4.5 Obesity and physical activity 21
3.4.6 Genetic Factors 24
2.5 CONCLUSION 26
CHAPTER 3 BREAST CANCER SURVIVAL 28
3.1 INTRODUCTION 28
3.2 TYPES OF BREAST CANCER 28
3.3 DIAGNOSIS 29
3.3.1 Breast cancer imaging 29
3.3.2 Breast cancer biopsies 30
3.4 PATHOLOGY 30
3.4.1 DCIS 30
3.4.2 Invasive breast cancer 31
3.5 TRENDS IN BREAST CANCER SURVIVAL 35
3.6 EPIDEMIOLOGY OF BREAST CANCER SURVIVAL 37
3.6.1 Reproductive factors 37
3.6.2 Exogenous hormones 38
3.6.3 Diet 39
3.6.4 Smoking 41
3.6.5 Obesity and Physical Activity 41
3.6.6 Genetic Factors 43
3.7 CONCLUSION 45

CHAPTER 4 *PPARA* 46
4.1 THE BIOLOGY OF PPARA 46
4.1.1. Introduction 46
4.1.2 Nuclear Hormone Receptors (NHR) 46
4.1.3 Biologic Mechanisms of PPARα 48
4.2 EPIDEMIOLOGY OF PPARA 51
4.2.1 Introduction 51
4.2.2 Lipoprotein levels 54
4.2.3 Obesity 55
4.2.4 Coronary Heart Disease (CHD) 55
4.2.5 Type II Diabetes Mellitus 56
4.2.6 Fibrate Response 57
4.2.7 Other diseases 58
4.3 CONCLUSION 59
CHAPTER 5 RESEARCH DESIGN AND METHODS 60
5.1 STUDY OVERVIEW 60
5.2 LONG ISLAND BREAST CANCER STUDY PROJECT 61
5.2.1 Subject Eligibility 61
5.2.2 Case Identification 62
5.2.3 Control Identification 63
5.2.4 Subject Recruitment 63
5.2.5 Case-Control Study Interview 64
5.2.6 Case Subjects' Medical Record Retrieval and Abstracting 68
5.2.7 Interview Data Management and Statistical Analyses 68
5.2.8 Summary 69
5.3 LONG ISLAND BREAST CANCER STUDY PROJECT FOLLOW-UP 69

5.3.1 Subject Eligibility for Follow-up 69
5.3.2 Subject Recruitment for Follow-up 69
5.3.3 Follow-up Study Interview 70
5.3.4 Case Subjects' Medical Records and Abstracting 71
5.3.5 Outcome Ascertainment 72
5.3.6 Follow-up Interview Data Management and Statistical Analyses 74
5.3.7 Summary 75
5.4 GENOTYPING 75
5.4.1 SNP selection and Haplotypes 75
5.4.2 Genotyping 77
5.5 COVARIATES 78
5.5.1 Overview 78
5.5.2 Outcome of interest 80
5.5.3 Obesity as an effect modifier 80
5.5.4 Confounders for the case-control study 81
5.5.5 Confounders for the survival analysis 83
5.5.6 Summary 84
5.6 DATA MANAGEMENT FOR CASE-CONTROL AND FOLLOW-UP STUDIES 84
5.7 DATA ANALYSIS 86
5.7.1 Overview 86
5.7.2 Haplotype Analysis 86
5.7.3 Hierarchical Modeling 91
5.7.4 Sample Size and Power 92
5.8 STUDY ADVANTAGES AND LIMITATIONS 97
5.9 SUMMARY 99
CHAPTER 6 CASE CONTROL STUDY MANUSCRIPT 101

6.1 ABSTRACT 101
6.2 INTRODUCTION 102
6.3 MATERIALS AND METHODS 103
6.3.1 Study population 103
6.3.2 Exposure assessment 104
6.3.3 Genetic polymorphisms 105
6.3.4 Statistical methods 106
6.4 RESULTS 108
6.5 DISCUSSION 109
6.6 ACKNOWLEDGEMENTS 112
6.7 TABLES 113
6.8 LEGENDS TO FIGURES 119
CHAPTER 7 SURVIVAL STUDY MANUSCRIPT 120
7.1 ABSTRACT 120
7.2 INTRODUCTION 121
7.3 MATERIALS AND METHODS 122
7.3.1 Study Population 122
7.3.2 Data and Biological Sample Collection 122
7.3.3 Selection of tagging SNPs 122
7.3.4 Laboratory Analysis 123
7.3.5 Statistical Analysis 123
7.4 RESULTS 124
7.5 DISCUSSION 125
7.6 TABLES AND FIGURES 127
CHAPTER 8 DISCUSSION 130
8.1 SPECIFIC AIMS 130

8.2 STRENGTHS AND LIMITATIONS 133

8.2.1 Tag SNP selection and genotyping 133

8.2.2 Recall bias 135

8.2.3 Study Power 136

8.3 FUTURE DIRECTIONS 137

8.3.1 Replication 137

8.3.2 Expansion to the PPAR family 138

8.3.3 Survival 140

8.4 SUMMARY 140

APPENDICES 142

APPENDIX A: TABLE OF PPAR-ALPHA STUDIES 142

APPENDIX B: BREAST CANCER SURVIVAL AND OBESITY LITERATURE REVIEW 154

APPENDIX C: COMPARISON OF TNM STAGING EDITIONS [252] 159

APPENDIX D: METHODS AND DATA ANALYSIS TABLES 161

WORKS CITED 234

LIST OF TABLES

TABLE 1.1.Obesity classifications based on 4 different measures of body composition 4

TABLE 1.2. Common hormonal abnormalities associated with abdominal adiposity according to the World Health Organization [5] 10

TABLE 1.3. Relative risk (RR) of health problems associated with obesity according to the World Health Organization [5] * 11

TABLE 3.1. In situ breast cancer pathology features of interest from Bilous *et al.* [80] 31

TABLE 3.2. Invasive breast cancer pathology features of interest from Bilous *et al.* [80] 32

TABLE 3.3. TNM staging system from the National Comprehensive Cancer Network [78] .. 35

TABLE 4.1. Summary of evidence based on epidemiology studies of *PPARA* and various disease outcomes 58

TABLE 5.1. Distribution of select demographic characteristics by cases and control status, Long Island Breast Cancer Study Project, 1996-1997 [37] 66

TABLE 5.2. NDI death codes used to create the broad categorical variables for underlying, primary and breast cancer-related death 73

TABLE 5.3. Frequencies of causes of death for the 1,083 women who donated blood and completed the baseline case-control interview 74

TABLE 5.4. Frequencies of causes of death for the 823 women who donated blood and completed a follow-up interview 74

TABLE 5.5. Distribution of treatment by vital status for the 823 women who donated blood and completed a follow-up interview 74

TABLE 5.6. Haplotype tagging SNPs of *PPARA* to be evaluated identified using ldSelect and the PGA European-American population as the reference panel 78

TABLE 5.7. Distribution of Body Mass Index among Cases and Controls, Long Island Breast Cancer Study Project, 1996-1997 [37] 81

TABLE 5.8. Study power for main gene effects assuming an alpha of 0.05 and a 1:1 control/case ratio 93

TABLE 5.9. Study power for gene-environment interactions assuming an alpha of 0.05 and a 1:1 case/control ratio 95

TABLE 5.10. Study power for main gene effects assuming an alpha of 0.05 for survival analysis (n=1,061) 95

TABLE 5.11. Study power for survival analysis gene-environment interactions assuming an alpha of 0.05 (n=1,061) 96

TABLE 6.1. PPARA haplotype tagging SNPs identified using LDSelect and the PGA European-American population as the reference panel 113

TABLE 6.2. Summary table of ORs for association between six PPARA polymorphisms and breast cancer risk by menopausal status in LIBCSP....................114

TABLE 6.3. Odds ratios and 95% CIs for six PPARA polymorphisms and breast cancer among post-menopausal women in the LIBCSP for conventional and hierarchical models with varying priors 116

TABLE 6.4. Odds ratios, interaction contrast ratios and 95% CIs for six *PPARA* polymorphisms and breast cancer by aspirin use in all LIBCSP women under the dominant inheritance model....................117

TABLE 6.5. Odds ratios and frequencies by case-control status for 14 PPARA haplotypes relative to all other haplotypes among post-menopausal women in LIBCSP....................118

TABLE 7.1. Hazard ratios and frequencies by vital status for six SNPs in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n=1073) 127

TABLE 7.2. Hazard ratios and frequencies by vital status for haplotypes in relation to all-cause mortality (n=132) among a cohort of women with breast cancer (n=1073)....................128

TABLE 8.1. Minor allele frequencies by race for LIBCSP versus PGA 134

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes....................142

TABLE B.1. Breast cancer survival and obesity literature (2003-Present) 154

TABLE C.1. Comparison of TNM Staging Editions [252] 159

TABLE D.1. Variables and variable definitions included in case-control and survival analyses 161

TABLE D.2. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs135542 (AA vs. AG+GG) in post-menopausal women....................164

TABLE D.3. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs1800206 (CC vs. CG+GG) in post-menopausal women 166

TABLE D.4. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253623 (AA vs. AG+GG) in post-menopausal women....................168

TABLE D.5. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253699 (TT vs. CT+TT) in post-menopausal women....................170

TABLE D.6. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253755 (GG vs. AG+AA) in post-menopausal women....................172

TABLE D.7. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253760 (TT vs. GT+GG) in post-menopausal women................174

TABLE D.8. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, ΔW interaction & BMI 176

TABLE D.9. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, BMI interaction 178

TABLE D.10. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, NSAID interaction...180

TABLE D.11. Summary of odds ratios for rs135542 stratified by three effect modifiers 182

TABLE D.12. Model building procedure for SNP rs1800206 (CC vs. CG+GG) in post-menopausal women, Jewish interaction..183

TABLE D.13. Model building procedure for SNP rs1800206 (CC vs. CG+GG) in post-menopausal women, PA interaction..185

TABLE D.14. Model building procedure for SNP rs4253623 (AA vs. AG+GG) in post-menopausal women, NSAID interaction...187

TABLE D.15. ORs, ICRs and multiplicative p-values for rs4253623 by aspirin use 188

TABLE D.16. Model building procedure for SNP rs4253699 (TT vs. CT+CC) in post-menopausal women, PA interaction..189

TABLE D.17. Model building procedure for SNP rs4253699 (TT vs. CT+CC) in post-menopausal women, NSAID interaction...191

TABLE D.18. ORs, ICRs and multiplicative p-values for rs4253699 by physical activity level and aspirin use 192

TABLE D.19. Model building procedure for SNP rs4253755 (GG vs. AG+AA) in post-menopausal women, PA interaction..193

TABLE D.20. Summary of odds ratios for rs4253755 stratified by PA 194

TABLE D.21. Model building procedure for SNP rs4253760 (TT vs. GT+GG) in post-menopausal women, no interactions...195

TABLE D.22. Haplotypes, frequencies and Odds Ratios by menopausal status for six PPARA SNPs 196

TABLE D.23. Model building procedure for SNP rs135542 (AA vs. AG+GG), looking at overall survival with continuous interaction term 200

TABLE D.24. Model building procedure for SNP rs1800206 (CC vs. CG+GG), looking at overall survival with continuous time interaction term...201

TABLE D.25. Model building procedure for SNP rs4253623 (AA vs. AG+GG), looking at overall survival with continuous interaction term 202

TABLE D.26. Model building procedure for SNP rs4253699 (TT vs. CT+CC), looking at overall survival with continuous interaction term 203

TABLE D.27. Model building procedure for SNP rs4253755 (GG vs. AG+AA), looking at overall survival with continuous time interaction term..204

TABLE D.28. Model building procedure for SNP rs4253760 (TT vs. GT+GG), looking at overall survival with continuous time interaction term..205

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics 206

TABLE D.30. Distribution of select survival covariates by genotyping information 222

TABLE D.31. Model building procedure for SNP rs135542 (AA vs. AG+GG), breast cancer specific mortality with continuous time interaction term 224

TABLE D.32. Model building procedure for SNP rs1800206 (CC vs. CG+GG), breast cancer specific mortality with continuous time interaction term 225

TABLE D.33. Model building procedure for SNP rs4253623 (AA vs. AG+GG), breast cancer specific mortality with continuous time interaction term 226

TABLE D.34. Model building procedure for SNP rs4253699 (TT vs. CT+CC), breast cancer specific mortality with continuous time interaction term 227

TABLE D.35. Model building procedure for SNP rs4253755 (GG vs. AG+AA), breast cancer specific mortality with continuous time interaction term 228

TABLE D.36. Model building procedure for SNP rs4253760 (TT vs. GT+GG), breast cancer specific mortality with continuous time interaction term 229

TABLE D.37. Hazard ratios and frequencies by vital status for six SNPs in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n = 1073) 230

TABLE D.38. Hazard ratios and frequencies by vital status for haplotypes in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n = 1073) 231

TABLE D.39. Hazard ratios and frequencies for all cause mortality by vital status for six SNPs in all cases verus invasive cases only...233

LIST OF FIGURES

FIGURE 1.1. Adipose tissue involvement in the release of many peptides and cytokines (adapted from Bray [6]) ... 2

FIGURE 1.2. Hormonal factors involved in the critical windows for obesity development in women, adapted from Lovejoy [17] ... 7

FIGURE 1.3. Trends in obesity prevalence in women, using data from the National Health Examination Survey 1960-1962, National Health and Nutrition Examination Survey, 1971-1974, 1976-1980, 1988-1994, 1999-2000. ... 8

FIGURE 2.1. Diagram of the Breast [28] ... 13

FIGURE 2.2. Major metabolic pathways of estrogen synthesis from *Endocrinology: An integrated approach* [19] ... 15

FIGURE 2.3. Type II Diabetes Mellitus and Breast Cancer from Wolf *et al.* [63] ... 23

FIGURE 3.1. SEER Relative Survival Rates by Stage at Diagnosis in Whites using data from the SEER 9 Registries for 1988-2001 [88] ... 36

FIGURE 3.2. SEER Stage Distribution Age Group in Whites using data from the SEER 9 Registries for 1988-2001 ... 37

FIGURE 4.1. PPARα mechanism of action to induce gene expression, adapted from Kiec-Wilk *et al.* [145] and Kota *et al.* [142] ... 47

FIGURE 4.2. Schematic depiction of the functional domains of PPARα, adapted from Aranda and Pascual [144] and Kota *et al.* [142] ... 48

FIGURE 4.3. Biologic effects of PPARα activation in different organ systems, adapted from Kota *et al.* [142] ... 49

FIGURE 4.4. A conceptual diagram of PPARα and obesity-mediated breast cancer development, adapted from Calle and Thun [147] ... 50

FIGURE 4.5. Conservation of L162V across 11 mammalian species from the University of California, Santa Cruz genome browser ... 52

FIGURE 5.1. Map of Long Island ... 61

FIGURE 5.2. Participation in the Long Island Breast Cancer Study Project and Follow-up Study ... 65

FIGURE 5.3. Directed acyclic graph for the association between post-menopausal breast cancer risk and *PPARA*. ... 82

FIGURE 5.4. Directed acyclic graph for the association between breast cancer prognosis and *PPARA*. 84

FIGURE 6.1. Linkage disequilibrium (r^2) between six PPARA tag SNPs among LIBCSP controls....119

FIGURE 7.1. Linkage disequilibrium (r^2) between the six *PPARA* polymorphisms....129

FIGURE 8.1. *PPARA* linkage disequilibrium from HapMap version B36 in the CEU population....141

FIGURE D.1. Kaplan-Meier curves for rs135542 (AA vs. AG+GG) for overall survival 197

FIGURE D.2. Kaplan-Meier curve for rs1800206 (CC vs. CG+GG) for overall survival 197

FIGURE D.3. Kaplan-Meier curves for rs4253623 (AA vs. AG+GG) for overall survival 198

FIGURE D.4. Kaplan-Meier curve for rs4253699 (TT vs. CT+CC) for overall survival 198

FIGURE D.5. Kaplan-Meier curve for rs4253755 (GG vs. AG+AA) for overall survival 199

FIGURE D.6. Kaplan-Meier curve for rs4253760 (TT vs. GT+GG) for overall survival......199

LIST OF ABBREVIATIONS AND SYMBOLS

AA = African American

ACS = American Cancer Society

AF = Activation Function

APO = Apolipoprotein

BMI = Body Mass Index (weight in kilograms divided by height in meters squared)

CA repeat = Cytosine-Adenosine Repeat

CATI = Computer Assisted Telephone Interviewing

CHD = Coronary Heart Disease

CI = Confidence interval

COMT = Catechol-O-methyl Transferase

COX-2 = Cyclooxygenase- 2

CYP = Cytochrome P450

DAG = Directed Acyclic Graph

DBD = DNA Binding Domain

DCIS = Ductal Carcinoma In Situ

DNA = Dioxyribonucleic Acid

ED = European Descent

EM = Expectation Maximization

ER = Estrogen Receptor

ERCC2 = Excision Repair Cross-Complementing group 2 gene

FFQ = Food Frequency Questionnaire

GPX1 = Selenium-dependent Glutathione Peroxidase

GST = Glutathione S-transferase

HDL = High Density Lipoproteins

HCFA = Health Care Finance Administration

2HE = 2-hydroxyestrone

16HE = 16α-hydroxyestrone

HER-2 = Human epidermal growth factor receptor

HR = Hazard Ratio

HRT = Hormone Replacement Therapy

HWE = Hardy-Weinberg Equilibrium

IACR = International Agency for Research on Cancer

ICD-9 = International Classification of Diseases – 9th edition

ICR = Interaction Contrast Ratio

ID = Identification

IL = Interleukin

IGF = Insulin-like Growth Factors

IGFBP = Insulin-like Growth Factors Binding Proteins

IR = Ionizing radiation

IRB = Institutional Review Board

κ = Kappa Statistic

KG = Kilogram

LBD = Ligand Binding Domain

LD = Linkage Disequilibrium

LDL = Low Density Lipoproteins

LIBCSP = Long Island Breast Cancer Study Project

LRT = Likelihood Ratio Test

MAF = Minor Allele Frequency

MnSOD = Manganese superoxide dismutase

MRI = Magnetic resonance imaging

NDI = National Death Index

NHANES = National Health and Nutrition Examination Survey

NHES = National Health Examination Survey

NHR = Nuclear Hormone Receptor

NSAID = Non-Steroidal Anti-Inflammatory Drug

NCI = National Cancer Institute

OC = Oral contraceptive

OGG1 = 8-Oxoguanine DNA Glycosylase

OR = Odds Ratio

PA = Physical Activity

PAH = Polycyclic Aromatic Hydrocarbons

PET = Positron Emission Tomography

PGA = Program for Genetic Applications

PL = Partition Ligation

PPARA = Peroxisome Proliferator-Activated Receptor gene-alpha

PPARα = Peroxisome Proliferator-Activated Receptor gene-alpha protein

PPARD = Peroxisome Proliferator-Activated Receptor gene-delta

PPARG = Peroxisome Proliferator-Activated Receptor gene-gamma

PPRE = Peroxisome Proliferator Response Element

PR = Progesterone Receptor

RDD = Random Digit Dialing

ROS = Reactive Oxygen Species

RR = Risk Ratio

RXR = Retinoid X Receptor

SAS = Statistical Analysis Systems

SE = Standard Error

SEER = Surveillance Epidemiology and End Results

SES = Socioeconomic Status

SHBG = Sex Hormone Binding Globulin

SNP = Single Nucleotide Polymorphism

SOD = Superoxide Dismutase

SRB = Sequentially Rejective Bonferroni test

TNFα = Tumor Necrosis Factor-Alpha

UGT = Uridine-5'-diphosphate-glucuronosyl-transferase

UNC = University of North Carolina

UTR = Untranslated Region

WAT = White Adipose Tissue

WC = Waist Circumference

WHO = World Health Organization

WHR = Waist to Hip circumference Ratio

ΔW = Weight Change

XPD = Xeroderma Pigmentosum, Complementation Group D

XRCC1 = X-Ray Cross-Complementing Protein, gene 1

CHAPTER 1 BODY SIZE/OBESITY

1.1 INTRODUCTION

Obesity has now reached epidemic proportions in the United States, but much remains unknown about its effects. In fact, even body size measures remain rudimentary and the most biologically relevant definition of body size has not been clearly established. This section outlines different ways of measuring and defining body size, potential critical windows for disease development with specific application to breast cancer, trends in obesity prevalence and its metabolic consequences.

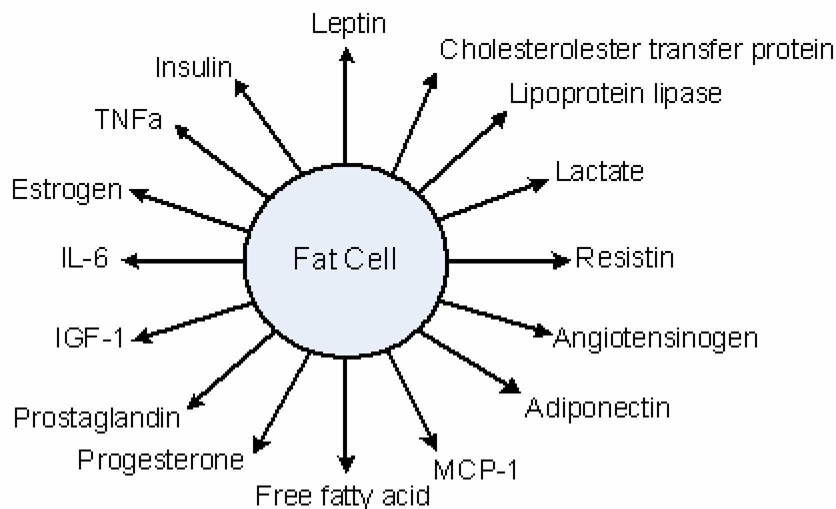
1.2 MEASURES OF BODY SIZE

According to the American Heritage dictionary [1], obesity can be defined as “the condition of being obese; increased body weight caused by excessive accumulation of fat”. This excess fat is deposited subcutaneously, primarily in the hips, thighs and gluteus in women and abdominally in men [2]. To approximate excess body fat, several different measures of obesity have been developed, including body mass index (BMI), waist to hip circumference ratio (WHR), and waist circumference (WC). With the exception of BMI, these measures aim to capture not only excess fat but also the distribution pattern of this fat. Weight change (ΔW), specifically weight gain, is related to obesity, and can describe excess fat when baseline weight is taken into consideration.

Fat tissue consists of active endocrine cells that produce and release a broad range of factors, including cytokines, insulin, progesterone and insulin-like growth factor (IGF)-1 (see Figure 1.1) [3]. In particular, adult weight gain has been linked to central adiposity,

which is associated with high levels of free fatty acids [4], cortisol and androgen receptors [5]. As will be discussed later, central adiposity is also associated with a variety of adverse health outcomes, such as heart disease, Type II diabetes mellitus and cancer. Thus, abdominal adipose tissue accumulation has implications on endocrine levels and disease risk, illustrating the potential importance of capturing fat distribution in obesity measurement.

FIGURE 1.1. Adipose tissue involvement in the release of many peptides and cytokines (adapted from Bray [6])



Obesity can be defined multiple ways and it is not clear which method may be most appropriate for breast cancer. Body mass index (BMI) is the most commonly used and is calculated by taking the ratio of weight in kilograms and height in meters squared. While this measure is extremely easy to calculate and implement using self-report or medical records, BMI has been criticized for inaccurately representing individuals who exercise extensively and lift weights [7]. For most individuals, however, it will be highly correlated with percent body fat [4]. In the Long Island Breast Cancer Study Project (LIBCSP), BMI will accurately measure body fat since most women are in middle age or older and unlikely to have extreme muscle mass.

WHR has been used to better measure body fat distribution (rather than general obesity), although waist circumference alone has been suggested to be as accurate at measuring central adiposity as the WHR and is easier to determine [4].

As an alternative to obesity, weight change over time is also gaining popularity because it allows investigation of potential critical windows for disease development and may reflect changes in hormone levels over time. Case-control studies typically measure self-reported BMI and weight change prior to diagnosis because disease status may influence waist circumference if measured after diagnosis and women are unlikely to know their waist circumference throughout their lives. In addition, participants can generally recall weight accurately even over long periods of time [4]. Lastly, weight cycling, or the repeated loss and regain of body weight [8], has emerged as an important health concern that should be considered in epidemiologic investigations of body size. Like weight change, weight cycling may help identify critical windows for disease development and help elucidate disease pathogenesis. LIBCSP data did not indicate that cycling was important for breast cancer development in this population [9].

In addition to these anthropometric measures of body size, several technological devices have been used to measure body fat percent in the clinical setting. For example, ultrasound, computer tomography and magnetic resonance imaging have all been assessed as more direct measures of body fat [10]. Unfortunately, these techniques all require invasive and expensive procedures, which have prevented them from being applied at the population level. For this reason, these methods will not be discussed in further sections.

1.3 Definition of body size

Based on these different anthropometric measures, obesity has been defined several different ways (see Table 1.1). For BMI, the World Health Organization (WHO) defines overweight as 25.0-29.9 kg/m^2, obesity as 30.0-39.9 kg/m^2, and morbid obesity as greater

than 39.9 kg/m^2 [5]. While BMI cut-off values have been well-established, the classification system for WHR, WC and body fat % are much less clear. For example, although the International Agency for Research on Cancer (IARC) endorsed the critical values developed by Han *et al.* [11] for WHR in their 2002 publication [4], researchers have criticized this scale because of its reliance on BMI thresholds for determining WC cut-off values. As an alternative, Zhu *et al.* used data from National Health and Nutrition Examination Survey (NHANES) III to derive WC risk thresholds without referral to BMI, recommending that 83 cm and 93 cm define the overweight and obese border respectively for women [12]. Lastly, the National Heart Lung and Blood Institute (NHLBI) has developed guidelines to combine BMI and WC for cardiovascular disease (CVD). They suggest that women with a WC greater than 0.88 cm and have a BMI between 25 and 30 are at high risk of disease; women with the same WC but have a BMI over 30 are considered at "very high risk" for CVD [13].

TABLE 1.1.Obesity classifications based on 4 different measures of body composition

	Underweight	**Normal Weight**	**Overweight**	**Obese**
BMI (kg/m2)*	< 18.50	18.50-24.99	25.00-29.99	≥ 30.00
WHR†	--	≤ 0.85	> 0.85	--
WC (cm)‡	--	< 80	80-87	≥ 88
Body fat (%)§	< 23	23-33	34-39	≥ 40

*BMI=body mass index, based on values cited by the World Health Organization [5]
†WHR = Waist to hip ratio, based on values provided by Han *et al.* [11] and cited by WHO [5] for women
‡Waist circumference values developed by Lean *et al.* [14] and recommended by IARC [4] for women
§Numbers based on women ages 40-59, developed by Gallagher *et al.* [15] and recommended by IARC [4]

1.4 DETERMINANTS OF OBESITY

Obesity is the direct result of excess energy intake relative to energy expenditure although the distal causes of this imbalance are more intangible. Decreased physical activity and diets high in saturated fat are typically blamed although genetics and macro-level

factors, such as proximity to grocery stores and median neighborhood household income, have also been implicated [4, 16]. Other risk factors for obesity include age, which will be discussed in the following section, ethnicity, educational level, and alcohol consumption [4]. Heredity is estimated to explain approximately 25-40% of obesity based on twin, adoption and family studies [4]. Several genes such as leptin, PPAR-gamma and alpha, have been suggested to play a role in obesity development.

1.5 Timing of Weight Change

In women, three life events have been suggested to be of particular importance in the development of obesity: menarche, pregnancy and menopause. These events mark critical time windows of hormonal fluctuation that may play an important role in disease development (see Figure 1.2). Gradual gains in weight are typically found in women ages 25-40, which have been attributed to pregnancy and a more sedentary lifestyle [4]. Menopause also represents a critical time period for obesity as metabolism typically slows and physical activity levels decrease [17]. Post-menopausal weight gain has also been linked to abdominal adiposity in women (more so than pre-menopausal weight gain), suggesting this time window might be particularly important in breast cancer development [17, 18].

1.5.1 Menarche and obesity

The menstrual cycle can be divided into two phases, the follicular phase and the luteal phase, both of which have implications on hormone secretion. The first stage, or follicular phase, is characterized by increasing estradiol and low progesterone levels as the uterine endometrium cells proliferate and the follicle reaches the ovarian surface [19]. In contrast, the second phase, or luteal stage, covers the formation and break-down of the corpus luteum, primarily dominated by a rise and fall in progesterone while estradiol levels

consistently decline [19]. Because of the difference in secreted hormones, the different phases also have been suggested to influence food preferences with sweet cravings common in the follicle stage and fat preferred during the luteal stage [17]. Obesity has been linked to reduced progesterone levels due to anovulation and depressed progesterone levels during the luteal phase of the menstrual cycle [18]. The effect of obesity on progesterone levels has been used as a potential explanation for the inconsistent results found in studies examining the relationship between obesity and risk of pre-menopausal breast cancer, which will be discussed in the next section.

1.5.2 Pregnancy and obesity

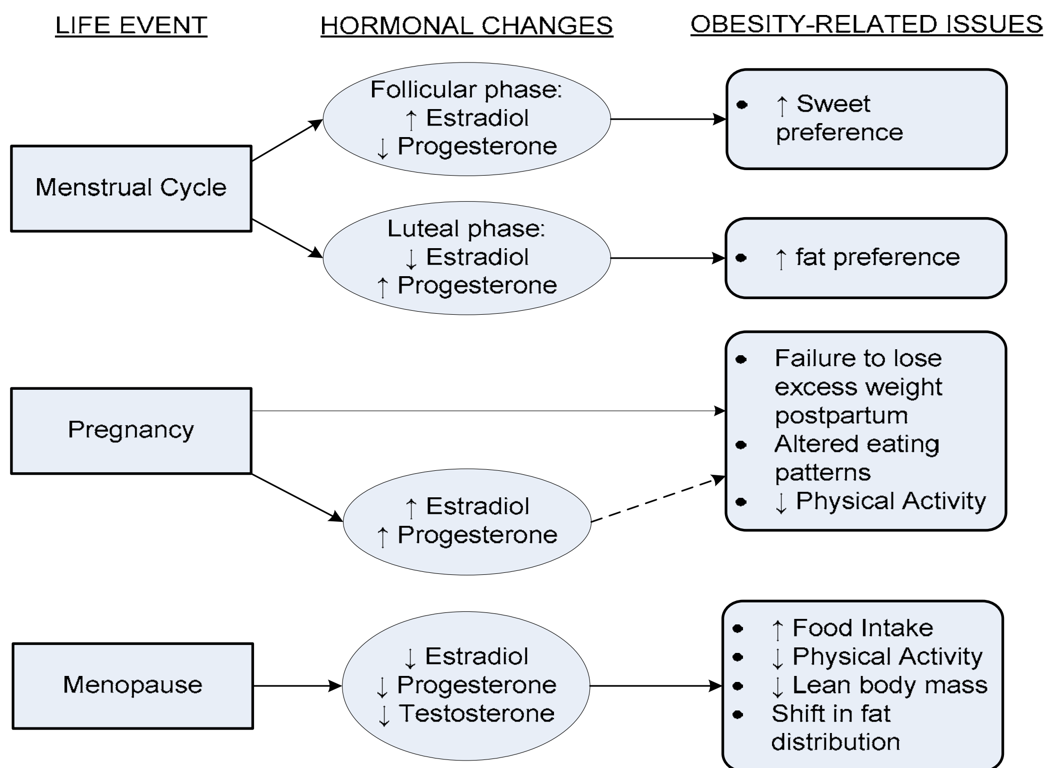
While the majority of women experience only small weight gain due to pregnancy (0.5-2.4 kg), approximately 10-20% of women will experience over 15 kg in postpartum weight gain [17]. As might be expected, pre-pregnancy weight, parity and gestational weight gain play large roles in influencing weight retention following pregnancy. Beyond weight retention, however, hormonal levels also change dramatically during pregnancy, which could point to another potential critical time period for breast cancer development. Once fertilization has occurred, estradiol and progesterone levels remain high for the duration of the pregnancy [19]. This hormone change coupled with possible weight retention may represent an understudied critical time window for disease development.

1.5.3 Menopause and obesity

Women are classified as post-menopausal if they experience “the absence of menstruation for one year” [19]. Menopause is marked by several distinct hormonal changes, which make women in this group more vulnerable to central adiposity. Because the ovaries are no longer developing follicles, estrogen levels fall dramatically. Similarly, since the corpus luteum no longer forms, progesterone levels fall dramatically as well. In

fact, after menopause, the primary source of estrogen for women becomes the aromatization of androgens to estrogen in fat tissue [19]. The decline in sex hormones and decrease in lipoprotein lipase activity in adipocytes also leads to a shift in fat accumulation from the periphery (buttocks and thighs) to the abdomen [17]. Thus, post-menopausal weight gain is most likely to result in abdominal adiposity than weight gain earlier in life.

FIGURE 1.2. Hormonal factors involved in the critical windows for obesity development in women, adapted from Lovejoy [17]



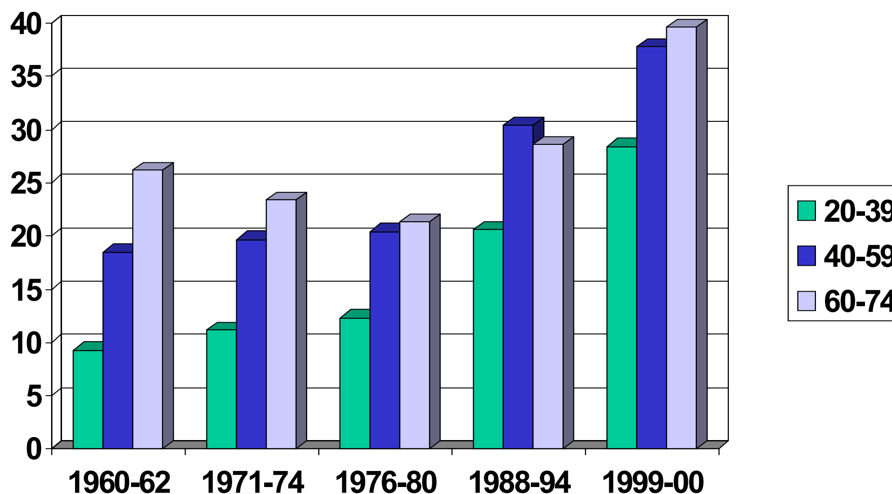
1.6 OBESITY PREVALENCE

The prevalence of obesity in women has increased dramatically over the last decade in the United States (see Figure 1.3). Recent data from NHANES indicate a 7.4% increase in the proportion of women aged 40-59 who are obese (body mass index ≥ 30) from NHANES III to NHANES 1999-2000 [20]. In particular, women ages 40-49 showed an 8.5% increase in obesity over this time period while women ages 50-59 saw a 5.6% increase, and

women ages 60-69 saw a 12.7% increase [20]. More recently, Hedley *et al.* estimated that 64.9% of women aged 40-59 and 68.4% of women over age 60 are overweight or obese [21]. These increases in obesity are occurring in the age group of women who are most at risk of developing breast cancer.

Similarly, abdominal adiposity in women has increased dramatically in the last three decades. Using data from the National Health Examination Survey (NHES I), NHANES III and NHANES 1999-2000, Okosun *et al.* found that the mean waist circumference was 77 cm for 1960-1962, 92 cm for 1988-1994, and 94 cm for 1999-2000 in women ages 20-79 years [22]. Parallel increases were seen in the prevalence of women with abdominal obesity, defined as a waist circumference of ≥ 88cm, over these same time periods (19.4%, 38.8% and 59.9% respectively) [22]. These increases in abdominal adiposity are troubling because of its strong association with many adverse health outcomes, including cancer, Type II diabetes mellitus and cardiovascular disease [23].

FIGURE 1.3. Trends in obesity prevalence in women, using data from the National Health Examination Survey 1960-1962, National Health and Nutrition Examination Survey, 1971-1974, 1976-1980, 1988-1994, 1999-2000.



1.7 METABOLIC CONSEQUENCES OF OBESITY

1.7.1 Overview

Obesity is associated with a broad range of morbidities, and excess mortality attributed to obesity is estimated to be second only to tobacco in the United States in the year 2000 (435,000 versus 385,000 deaths) [24]. Adult weight gain is associated with increased abdominal adiposity, which is an indicator of insulin resistance [6], and has been linked to Type 2 diabetes mellitus [3] and breast cancer in postmenopausal women [18]. Central adiposity is also associated with increased levels of small dense low density lipoproteins (LDL), which have been linked to coronary heart disease independently of cholesterol level, and lower levels of high density lipoprotein cholesterol (HDL) [3].

1.7.2 Endocrine disturbances

As mentioned earlier in this section, abdominal adiposity is associated with a variety of endocrine responses. Visceral adipose tissue has more cortisol and androgen receptors than subcutaneous adipose tissue, which makes it more sensitive to hormone stimulation and lipid metabolism [5]. This sensitivity results in several hormonal changes, highlighted in Table 1.2, including insulin resistance, increased insulin secretion and decreased sex hormone binding globulin (SHBG) for women [5]. Insulin resistance is defined as the impaired response of muscles and other cells to insulin [2] and can lead to elevated blood glucose levels with reduced utilization of glucose in the muscles [4]. Hyperinsulinemia, or increased insulin secretion, is the consequence of reduced glucose removal [3] and has been suggested to be the body's response to obesity and insulin resistance [5]. Additionally, central adiposity is also associated with high triglyceride levels, low HDL levels, and increased levels of small, dense LDL particles, which may explain its connection to cardiovascular disease [3, 5]. The impact of obesity on estrogen levels will be discussed

more in the next section, but it is important to note that central adiposity is associated with increased bioavailable estrogen in post-menopausal women but not in pre-menopausal women [4].

TABLE 1.2. Common hormonal abnormalities associated with abdominal adiposity according to the World Health Organization [5]

- Insulin resistance and increased insulin secretion
- Increased free testosterone and free androstenedione levels associated with decreased sex hormone binding globulin (SHBG) in women
- Decreased progesterone levels in women
- Increased cortisol production
- Decreased growth hormone levels

1.7.3 Obesity Associated disease

As Table 1.3 indicates, obesity is associated with a broad range of health problems, including Type II diabetes, coronary heart disease and some cancers. Not surprisingly given the strong association between insulin resistance and obesity, Type II Diabetes Mellitus (DM) has been consistently linked to obesity in both males and females. In fact, in the Nurses' Health Study, a risk ratio (RR) of 49.0 (95% confidence interval (CI): 34.0-71.0) was found for Type II DM for women with a BMI > 35 kg/m^2 compared to women with a BMI < 22 kg/m^2 [25]. This relationship holds even when weight gain is examined rather than BMI as a measure of body size, adjusting for BMI at age 18 [25]. For cancer, the evidence is growing, especially for post-menopausal breast cancer and endometrial and colon cancer, although the RRs observed are much smaller than those found in the diabetes literature. In a recent meta-analysis, Bergstrom *et al.* found a RR of 2.52 (CI not provided) for endometrial cancer, a RR of 1.52 for colon cancer and a RR of 1.25 for post-menopausal breast cancer in obese vs. normal weight individuals [26]. While the WHO might label these associations as "slight

increases," they are important because of the high prevalence of obesity, giving these associations large public health significance at the population level.

TABLE 1.3. Relative risk (RR) of health problems associated with obesity according to the World Health Organization [5] *

Greatly Increased (RR > 3.0)	**Moderately increased (RR 2.0 – 3.0)**	**Slightly increased (RR 1.0 – 2.0)**
Type II Diabetes Mellitus	Coronary Heart Disease • Metabolic Syndrome • Stroke	Some Cancers: • Post-menopausal breast cancer • Endometrial cancer • Colon Cancer
Gallbladder disease	Hypertension	Reproductive hormone abnormalities
Dyslipidemia	Osteoarthritis (knees)	Polycystic ovary syndrome
Insulin resistance	Hyperuricemia and gout	Impaired fertility
Breathlessness		Low back pain from obesity
Sleep apnea		Fetal defects associated with maternal obesity

*All RRs are approximate values

1.8 CONCLUSION

Body size is typically measured using BMI, but recent research indicates that weight gain may better capture the endocrine changes that result with increased central adiposity. The Long Island Breast Cancer Study Project is well-equipped to address obesity-related mechanisms because data on weight change by decade, BMI by decade, and potential critical windows for disease development are available for investigation. Given the epidemic proportion of obesity in the United States, particularly in post-menopausal women, clarifying the role of obesity in breast cancer development and survival has become increasingly important.

CHAPTER 2 BREAST CANCER INCIDENCE

2.1 INTRODUCTION

Much of what is known about breast cancer has evolved from its risk factor epidemiology. For example, it has become apparent that hormones, specifically estrogen, play an important role in disease development. Demographic, behavioral and dietary factors that influence hormone levels have been consistently associated with risk of breast cancer. This section outlines the trends in incidence, possible disease mechanisms and established and suspected risk factors for the disease.

2.2 TRENDS IN INCIDENCE

With 240,510 expected cases in 2007, breast cancer represents the most commonly diagnosed cancer in women, composing 26% of all female cancer diagnoses. It is also the second most common cause of cancer-related deaths in women with 40,460 estimated attributable deaths in 2007 comprising 15% of all female cancer-related mortalities [27]. Breast cancer also tends to affect older women with 94% of incident cases and 96% of breast cancer mortality occurring in women over the age of 40 [28]. Unlike many other cancers, Caucasians show a consistently higher rate of breast cancer compared to other races. From 2000-2004 in the United States, the annual incidence rate of breast cancer was 132.5 per 100,000 person-years in Caucasian women compared to 118.3 per 100,000 in African American women [28].

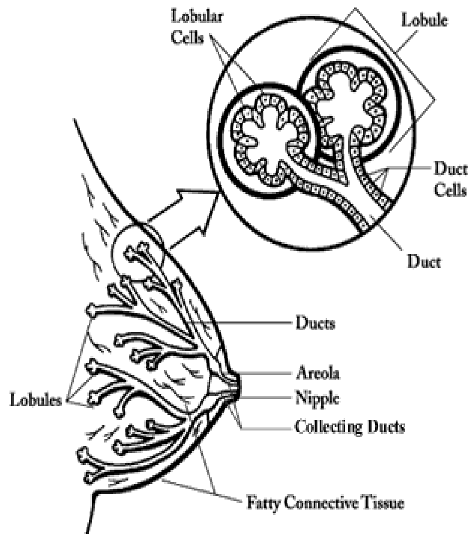


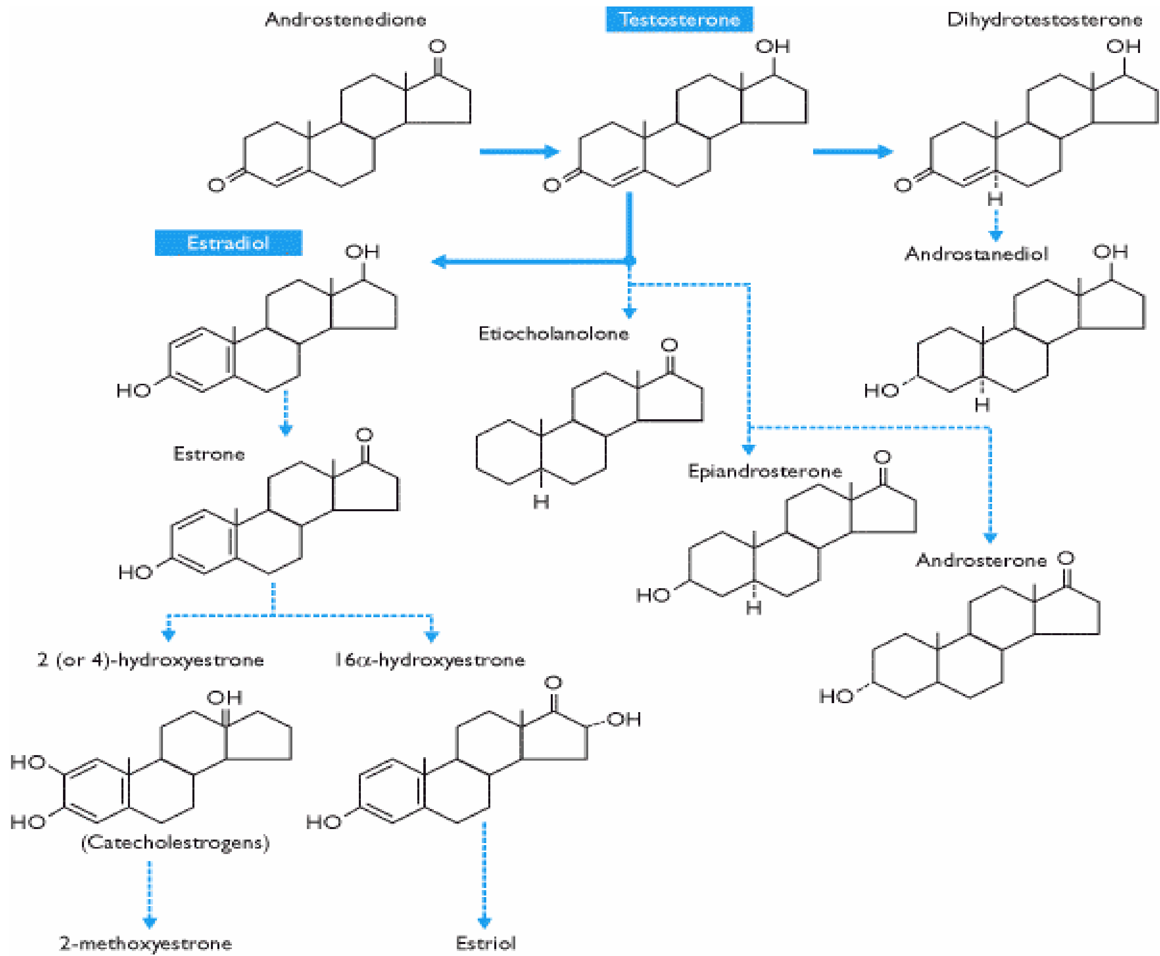
FIGURE 2.1. Diagram of the Breast [28]

Breast cancer incidence also varies by staging with invasive cancer consistently occurring at higher rates than *in situ* breast cancer. According to the American Cancer Society, invasive breast cancer can be defined as cancers that have penetrated the breast tissue from the lobules or ducts while *in situ* cancers have not yet spread beyond their point of origin in the duct (for ductual carcinoma in situ) or in the lobules (lobular carcinoma *in situ*) (see Figure 2.1) [28]. The incidence of DCIS has been steadily increasing in the last ten years, possibly due to the advent of mammography, while invasive breast cancer rates have been more constant [28]. Women over the age of 55 years experience the highest rate of both breast cancer types although invasive cancer is the most common with 124,300 estimated cases in 2007 (77.0%) compared to 37,110 expected *in situ* cases (23.0%) [28]. Across all ages, 211,240 invasive (74.2%) and 62,030 in situ (25.8%) cases were estimated in 2007 [28]. These percentages are consistent with the numbers observed in the Long Island Breast Cancer Project, where 1273 invasive and 235 in situ cases were found (84.4% vs. 15.6% respectively). Further distinction between the different kinds of breast cancer will be discussed in more depth in Chapter 3 on breast cancer survival.

2.3 MECHANISM OF ACTION

Known risk factors for breast cancer have long pointed to increased endogenous estrogen levels as the primary underlying mechanism of action in breast cancer carcinogenesis and progression. Estrogen has been suggested to elevate risk of breast cancer by promoting tumor development through increased cell proliferation. Rapidly dividing cells may be more prone to DNA replication errors, leading to potentially deleterious mutations [29]. Although ovaries produce very little estrogen following menopause, post-menopausal women may be exposed to high endogenous estrogen levels through the aromatization of androgens in fat tissue, such as testosterone [30]. Post-menopausal obesity elevates estrogen levels further since adipose cells secrete aromatase, leading to increased conversion of androstenedione to estradiol [31]. In contrast, pre-menopausal women experience consistently high levels of estrogen, particularly during the follicle phase of the menstrual cycle, regardless of weight, so that any additional estrogen from the aromatization of androgens due to obesity will not have a significant impact on overall estrogen levels [4]. Additionally, estrogen metabolism has two primary products: 16α–hydroxyestrone (16HE) and 2-hydroxyestrone (2HE). Obesity is associated with decreased levels of 2HE, which is less bioactive, while maintaining higher levels of 16HE, which remains bioactive (see Figure 2.2) [32]. Thus, obesity may impact risk of breast cancer by influencing estrogen metabolism.

FIGURE 2.2. Major metabolic pathways of estrogen synthesis from *Endocrinology: An integrated approach* [19]



More controversially, progesterone has been suggested to work synergistically with estrogen to increase risk of breast cancer. This hypothesis, referred to as the "estrogen augmented by progesterone" theory, was originally proposed by Key and Pike in 1988 [33], but has been met with some opposition. In vivo studies have shown that cell proliferation is greatest in the luteal phase of the menstrual cycle, when progesterone levels are increasing [33, 34], but it has not clearly been established that progesterone is responsible for this increased cell division. As an alternative, this increased cell proliferation has been suggested to be a delayed response to peak estrogen levels, occurring 4-5 days earlier in the cycle [34]. While some epidemiologic evidence, discussed in section 2.4, suggests that

progesterone may play a role in breast cancer, this is not supported by in vivo and histological studies [34].

Insulin and insulin-like growth factor (IGF)-1 have also been suggested to work with estrogen to influence breast cancer risk, particularly in pre-menopausal women [35]. These findings run contrary to the expected since central adiposity is associated with post-menopause rather than pre-menopause status. Like estrogen, insulin and IGF-1 have been shown to increase cell proliferation while preventing apoptosis [30]. Estrogen has been linked to the regulation of IGF signaling with high estrogen levels resulting in an increase in IGF activators and a decrease in IGF inhibitors although this remains controversial [36]. Insulin and possibly IGF-1 have also been implicated in increased estrogen bioactivity through the decreased synthesis of sex-hormone-binding globulin (SHBG) in the liver [30]. SHBG is a common transporter of estradiol, and decreased SHBG levels will result in increased levels of bioavailable free estradiol [18]. Identifying genetic polymorphisms that influence obesity and estrogen bioavailability could play an important role in advancing knowledge of breast cancer etiology through these two important growth pathways.

2.4 Risk Factor Epidemiology

3.4.1 Reproductive factors

While several risk factors have been found to influence the risk of breast cancer, cumulative exposure to estrogen appears to be an important component of breast cancer development. Parity, age at first birth, lactation, age at menarche and age at menopause are considered established risk factors for breast cancer and have been hypothesized to influence circulating levels of estrogen and breast tissue's lifetime exposure to estrogen [29, 30]. Breast-feeding, for example, has been suggested to have an inverse relationship with breast cancer in both pre- and post-menopausal women by inhibiting ovulation. Ovulation,

as mentioned earlier, increases estradiol levels and influences progesterone regulation in the luteal phase of the menstrual cycle. High parity and early age at first birth are believed to influence risk in a similar manner by reducing the number of ovulatory cycles while early age of menarche and late age of menopause may be an indicator of high levels of cumulative estrogen exposure. Consistent with Long Island Breast Cancer Study Project (LIBCSP) findings [37], this research indicates that the reproductive behaviors of women have an important impact on their risk of breast cancer.

3.4.2 Exogenous hormones

In premenopausal women, oral contraceptive use is a common source of exposure to exogenous hormones. Birth control pills can contain estrogen, typically ethinyl estradiol, alone or in combination with progesterone [19, 38] and doses have ranged between 100 μg in 1960 to 20 μg currently [38]. In a pooled analysis of 53,297 breast cancer cases and 100,239 controls, representing 54 study populations, a slightly increased risk of breast cancer was observed among current users (RR=1.24; 95% CI: 1.15-1.33) compared to never users [39]. This risk decreased after stopping pill use (RR=1.16 for 1-4 years after stopping and 1.07 for 5-9 years after stopping) and no elevation in risk was found 10 years following the end of pill use (RR=1.01; 95% CI: 0.96-1.05) [39]. The authors also noted that tumors in oral contraceptive users are more likely to be early stage disease and localized than the tumors in non-users [39].

Like oral contraceptives, hormone replacement therapy (HRT) can contain estrogen alone, typically estradiol or estradiol valerate, or may include progesterone [19]. It can be used for short-term alleviation of menopausal symptoms or be used prophylactically to delay bone density loss [19]. In a pooled analysis of 52,705 breast cancer cases and 108,411 controls, representing 51 study populations, an elevated risk of breast cancer was found in current versus never users (RR=1.21, SE=0.05, CI not reported) although this risk

disappeared 5 years after HRT cessation (RR=1.07, SE=0.04, CI not reported) [40]. Similarly, the Million Woman Study, which consisted of 1,084,110 British women aged 50–64 years, found an elevated risk of breast cancer in current HRT users versus never users (RR=1.66; 95% CI: 1.58-1.75) with estrogen-progesterone combinations showing the greatest risk (RR=2.00; 95% CI: 1.88-2.12) in comparison to estrogen alone or tibolone, which contains neither estrogen or progesterone [41]. Based on these large studies, HRT is now regarded as a breast cancer risk factor. These results are consistent with those found in the LIBCSP, where the odds of breast cancer were elevated for ever OC use, both OC and HRT use, and long-term HRT use [42].

Interestingly, non-steroidal anti-inflammatory drug (NSAID) use has emerged as a possible chemopreventive breast cancer agent although the evidence remains inconclusive [43-45]. Biologically, estrogen plays a role in the inflammatory pathway and NSAIDs may act through an estrogen-mediated pathway by targeting cyclooxygenase (COX), which inhibits prostaglandin production. NSAID use may reduce breast cancer risk through decreased estrogen production from progesterone. Inflammation also results in cell proliferation and may promote angiogenesis [46]; thus, NSAID use may influence breast cancer development in a non-estrogen related pathway as well.

In the LIBCSP, ever aspirin use was associated with decreased odds of breast cancer than never use (OR=0.80; 95% CI: 0.66-0.97) [45]. These results are consistent with a meta-analysis published in 2003 of 9 studies examining the relationship between NSAID use and breast cancer (RR=0.77; 95% CI: 0.66-0.88) [47]. In the LIBCSP, NSAID use also implicated hormone receptor positive tumor development, where aspirin was associated with hormone receptor positive breast cancer. Tumors with one or more positive hormone receptor showed decreased odds of breast cancer in ever versus never users compared to tumors with no positive hormone receptors (OR=0.74, 95% CI: 0.60-0.93 for ≥ one positive hormone receptor; OR=0.97, 95% CI: 0.67-1.40 for no positive hormone receptors) [45]. In

contrast, Marshall *et al.* found an increased risk of breast cancer in hormone receptor positive tumors compared to hormone receptor negative tumors in daily versus non-regular NSAID users using data from the California Teachers cohort (RR= 1.03, 95% CI: 0.90-1.17 for ER/PR-positive tumors; RR=1.48, 95% CI: 1.01-1.89 for ER/PR-negative tumors) [48]. Thus, while there is a suggestion of differential risk for hormone receptor positive and negative tumors with NSAID use, the direction of the effect is not consistent across studies.

3.4.3 Diet

Alcohol consumption, the dietary factor for which the evidence of an association with breast cancer is strongest [29], has been shown to increase estrogen levels, which may explain the elevated association with breast cancer observed in habitual drinkers. An international pooled analysis found an elevated risk of 1.32 (standard error (SE)=0.059) in women who consumed 35-44 grams of alcohol daily and an RR of 1.46 (SE=0.06) in women who consumed over 44 grams daily compared to women who never drink alcohol [49].

The evidence for an association between breast cancer and fat intake is much less convincing. Using a multivariate nutrient density model approach, a pooled analysis of eight cohort studies found no association with saturated (RR=1.09; 95% CI: 1.00-1.19), monosaturated (RR=0.93; 95% CI: 0.84-1.03) or polysaturated fats (RR=1.05; 95% CI: 0.96-1.16) for an increment of 5% of energy [50]. This same group also investigated fruit and vegetable intake in connection to breast cancer and found similar findings. The relative risk for a 100 g/day increment in fruit consumption reflected no association with breast cancer risk (RR=0.99, 95% CI: 0.98-1.00 for total fruits; RR=0.99, 95% CI: 0.98-1.00 for total fruits and vegetables) in this analysis of eight cohort studies [51] although case-control studies show a consistent reduction in breast cancer risk with high fruit and vegetable intake [52]. Dietary fiber has also been suggested to reduce risk of breast cancer by preventing estrogen absorption in the intestinal track. A meta-analysis of 12 case-control studies in

1990, however, found that contrary to this hypothesis, fiber consumption was positively associated with breast cancer (RR= 1.46, comparing highest quintile to lowest; CI not presented) [53]. Thus, while the association between heavy alcohol intake has been consistently observed in epidemiologic studies, the association between fat, fiber and fruit and vegetable consumption is much less convincing.

3.4.4 Environmental Factors

As with most cancers, ionizing radiation (IR) has also been found to associated with an increased risk of breast cancer with odds ratios and relative risks reaching over 9.0 [38]. IR is capable of breaking DNA bonds and, consequently, will lead to very high rates of mutation [38]. Information from Hiroshima survivors have played an important role in defining the role of radiation in breast cancer development and critical windows for disease development. In fact, an RR of 2.42 per Sievert was found for women ages 10-19 at the time of the atomic bomb while no elevation in risk of breast cancer was observed for women over the age of 40 [54], suggesting that puberty may represent a susceptible period for disease development.

Active smoking has not been found to be associated with breast cancer consistently. In fact, despite the increased exposure to carcinogens with smoking, it has been suggested to reduce risk of breast cancer by minimizing weight gain through appetite suppression and through its anti-estrogenic effects, including early initiation of menopause and alteration of hormone metabolism [38]. Smoking and alcohol use are strongly correlated. To tease out the effects of smoking from alcohol consumption on breast cancer development, a collaboration of 53 epidemiologic studies examined breast cancer risk by alcohol consumption status. No association was found between active smoking, measured as never/ever smoked, and breast cancer risk among individuals who reported consuming alcohol (RR= 1.09; SE: 0.018) and among non-drinkers (RR=1.03; SE: 0.023) [49]. This

finding is consistent with the hypothesis that the dueling biologic mechanisms of smoking will cancel out any evidence of adverse or protective effects. LIBCSP data did indicate, however, a positive association with breast cancer for women who resided with a smoking spouse for over 27 years (OR=2.10; 95% CI: 1.47-3.02) [55], which is consistent with most previous studies that have reported increased breast cancer risk with passive smoke exposure [56].

Polycyclic aromatic hydrocarbon (PAH)-DNA adducts are one of the most consistently reported environmental risk factors associated with breast cancer. PAHs represent widespread environmental contaminants, most commonly found in cigarette smoke as well as in grilled and smoked meats, and are typically measured by the formation of DNA adducts. They have been implicated as mammary carcinogens in rodents [57] and have been strongly linked to breast cancer in three studies [58-60]. Within the Long Island cohort, increased odds of breast cancer were observed comparing individuals in the highest PAH quintile to the lowest (OR=1.51; 95% CI: 1.04-2.20) [59].

3.4.5 Obesity and physical activity

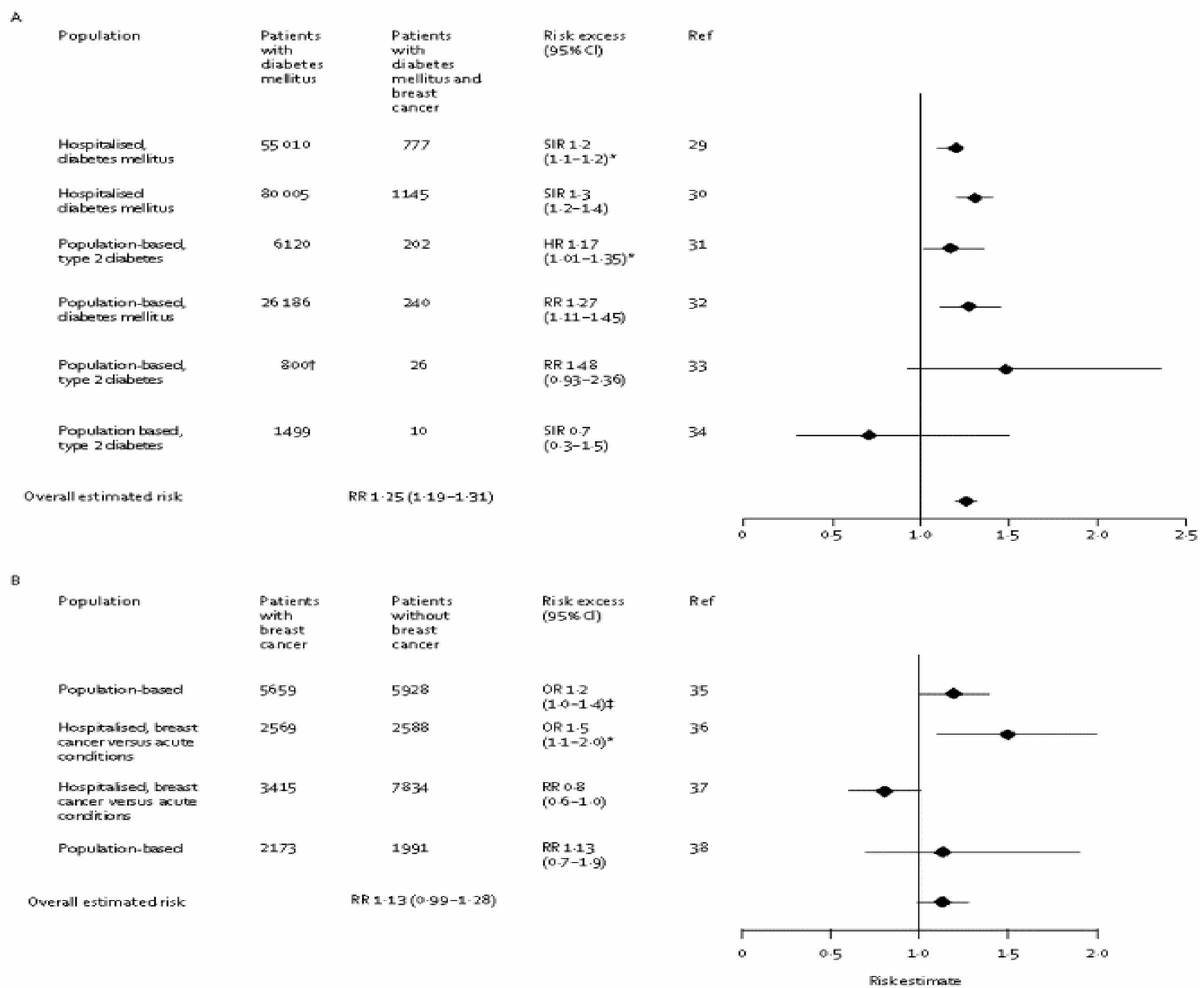
Large body size appears to reduce the risk of breast cancer in pre-menopausal women, possibly due to its role in decreasing ovulation, but has consistently been found to elevate risk in post-menopausal women. In a pooled analysis of prospective cohort studies, a pooled RR of 1.26 (95% CI: 1.09, 1.46) was found for post-menopausal breast cancer comparing women with a BMI less than 21 kg/m^2 to women with a BMI greater than 31 kg/m^2 [61]. In LIBCSP, post-menopausal women who gained more than 15 kg since age 20 had elevated odds of breast cancer compared to those who gained less than 3 kg (OR= 1.58, 95% CI: 1.11, 2.26) [9]. In pre-menopausal women, the relationship between obesity and breast cancer is much less consistent. A decreased risk was observed in the same analysis of prospective cohort studies with a pooled RR of 0.58 (95% CI: 0.34, 1.00)

comparing women with a BMI less than 21 kg/m^2 to women with a BMI greater than 31 kg/m^2 [61]. This pooled RR fits the hypothesis that obesity may decrease risk of pre-menopausal breast cancer by reducing the number of ovulatory cycles.

As mentioned earlier, the observed relationship between obesity and breast cancer in post-menopausal women is consistent with research indicating that obesity is associated with higher levels of estrogen. In post-menopausal women, obesity may be a significant source of bioavailable estrogen because the ovaries have stopped estrogen production. In contrast, this correlation is not found in pre-menopausal women, who already have high levels of estrogen from menstruation. This suggests that the additional estrogen exposure from obesity in pre-menopausal women may be of little consequence when compared to the naturally higher levels found at this point in life and by its inhibitory effects on ovulation [62].

The relationship between insulin and IGF-1 and breast cancer is further supported by recent literature suggesting an association between breast cancer and Type II diabetes mellitus. An estimated 16% of breast cancer cases in women over the age of 65 have also been found to have coincident diabetes mellitus [63]. In comparison, according to the Behavioral Risk Factor Surveillance System, the median nationwide prevalence of diabetes was 7.5% in 2006 [64]. In a recent meta-analysis, presented in Figure 2.3, cohort studies conducted in populations of type II diabetics found a slightly elevated but statistically significance increase in risk of breast cancer (RR = 1.25; 95% CI: 1.19-1.31) while case-control studies of breast cancer have found a smaller increase in risk (RR = 1.13; 95% CI: 0.99-1.28) with diabetes mellitus [63].

FIGURE 2.3. Type II Diabetes Mellitus and Breast Cancer from Wolf *et al.* [63]



There is convincing evidence that physical activity reduces risk of breast cancer, although this decreased risk is most pronounced in postmenopausal women [65]. In post-menopausal women, physical activity produces two positive effects, which may account for this difference: (1) physical activity reduces obesity, which is the primary source of endogenous estrogen in post-menopausal women, and (2) physical activity alone also reduces insulin as well as insulin-like growth factor levels [65]. In pre-menopausal women, the reduction in obesity is not necessarily an advantage as a means to prevent breast cancer development since obesity has been shown to have a protective effect in this group,

but may help in weight maintenance and prevention of post-menopausal obesity and promote anovulation.

3.4.6 Genetic Factors

Perhaps the most consistent risk factor for breast cancer is diagnosis of the disease in a first degree relative. Family history of breast cancer does not necessarily imply genetic susceptibility, however, but could also indicate shared environment and learned behaviors. Among women with a family history of breast cancer, a proportion will carry mutant alleles in *BRCA1* or *BRCA2*, which confer a high lifetime risk of breast cancer. In particular, the Ashkenazi Jewish population has been estimated to have the highest prevalence of these mutations at 2.2% [38]. Only a small proportion of cases carry the *BRCA* mutations (2-5%), but the two genes are believed to be responsible for most hereditary breast cancers, particularly early-onset breast cancer [38].

Research on common, low penetrance polymorphisms, or genetic variation that is common in the population and contribute only a slight increase in risk, has primarily focused on genes involved in inflammation, oxidative stress, estrogen metabolism and DNA repair. This dissertation will focus on the pathways most relevant to PPAR-alpha and obesity, specifically inflammation and insulin resistance. The following sub-sections will highlight several genes that have been investigated in connection to breast cancer along the inflammatory and insulin resistance pathways.

3.4.6.1 Inflammatory pathway. Inflammation has been suggested to increase risk of breast cancer two different ways. First, inflammation leads to increased cell proliferation, which increases cancer risk and may promote angiogenesis [46]. Second, inflammation may operate through an estrogen-mediated response to increase prostaglandin production, which is involved in cell differentiation and growth [46]. Cytokines and their receptors are

involved in inflammatory response, and polymorphisms in the cytokine genes have been hypothesized to be associated with breast cancer although study findings have been inconsistent.

TNFA. Tumor necrosis factor-alpha (*TNFA*) is well-established in the cancer literature and has been the focus of anticancer drug development [66]. Recently, eight TNFA SNPs were evaluated in connection to breast cancer development using data from USA and Poland Breast Cancer Study pooled together. Gaudet *et al.* found that heterozygotes of rs36152, an intronic polymorphism, had elevated odds of breast cancer development compared to common homozygotes (OR = 1.20; 95% CI: 1.05–1.37) [67]. Other studies that have examined a common polymorphism resulting in a guanine to adenine substitution at position -308 from the *TNFA* start site [68] have also found positive associations. Smith *et al.* reported differences in common homozygote genotype frequency for this polymorphism (79.7 vs. 68.2%, P = 0.03) [69] while Giordani *et al.*, in a small case-control study, found elevated odds of breast cancer in rare homozygotes compared to common homozygotes (OR=1.62; 95% CI: 0.08–96.42) [70].

Interleukins. Research on the interleukin (IL) family has focused primarily on *IL6* and *IL10*. In a study of five *IL6* polymorphisms, Slattery *et al.* found elevated odds of breast cancer for rs1800796 (OR=1.53; 0.99-2.37) for heterozygotes and rare homozygotes combined compared to common homozygotes using data from the 4-Corner's Study [71]. The ORs for the other SNPs were closer to the null or subject to imprecision. *IL10* has been studied in four case-controls studies, but sample sizes were small [72]. Overall, the studies indicate slightly elevated odds of breast cancer but examination in a larger study population is needed.

3.4.6.1 Insulin resistance pathway. Several genes that lie on the growth pathway in addition to promoting cell proliferation have also been hypothesized to have an inflammatory effect. Included in this set of growth factor genes is insulin-like growth factor (IGF)-I, which has been investigated in connection to breast cancer.

IGF-1. The IGF-I gene encodes for growth hormones that promote cell division in breast tissue. Research has focused on a simple tandem CA repeat that lies in the 5' region of the gene near the transcription start site. The common allele has 19 CA repeats while the 'at risk' variant is typically defined as all other number of repeats. Most studies have compared 19 / 19 homozygotes to 19 allele carriers (19 / -) and non-carriers (- / -) separately, but results have been inconsistent [73-76]. One study found reduced odds of breast cancer with the - / - genotype compared to 19 / 19 individuals (OR=0.50; 95% CI: 0.23-1.05) [73] while another found elevated odds (OR=1.34; 95% CI: 0.89-2.04) with this same comparison [75]. All studies found no association between 19 / - and 19 / 19 homozygotes [73-76]. In contrast, the LIBCSP data indicate slightly elevated odds of post-menopausal breast cancer (OR=1.19; 95% CI: 0.94-1.52) and decreased odds of pre-menopausal breast cancer (OR=0.87; 95% CI: 0.62-1.22) in the 19 / - versus 19 / 19 genotypes [77]. An elevated risk of post-menopausal breast cancer was also observed comparing the - / - to 19 / 19 genotype (OR=1.27; 95% CI: 0.89-1.81) while no difference in risk of pre-menopausal breast cancer was found (OR=1.05; 95% CI: 0.63-1.75) [77].

2.5 Conclusion

Breast cancer affects many families and represents an important public health concern. Estrogen, specifically its most bioactive form estradiol, seems to be a key component in the breast carcinogenesis pathway although its exact role remains uncertain. Because of its implications on the endocrine system, obesity appears to play an important role in breast cancer development in post-menopausal women. Given the rising prevalence

of obesity, particularly in post-menopausal women, identifying factors that may increase susceptibility through this mechanism has become increasingly important.

CHAPTER 3 BREAST CANCER SURVIVAL

3.1 INTRODUCTION

While breast cancer survival has increased dramatically in the last 30 years due to improved therapy options and early detection, survival remains extremely low in patients with advanced disease. Additionally, while several tumor markers are used to identify patients with poor prognosis or to aid in treatment decision-making, it is uncertain whether factors that influence breast cancer risk, such as obesity, also influence survival. This section outlines the different types of breast cancer, diagnostic procedures, key pathological tumor characteristics, trends in breast cancer survival, and the association between survival and obesity.

3.2 TYPES OF BREAST CANCER

As discussed in Chapter 2, breast cancers fall under two categories, carcinoma in situ, composing approximately 26% of all breast cancers, and invasive carcinoma, representing an estimated 74% of all breast cancers. In situ carcinomas are believed to be a part of the carcinogenesis continuum and may be an indicator of increased risk of invasive breast cancer. Ductal carcinoma in situ (DCIS) is the most common in situ cancer and originates in the ductal cells of the breast while lobular carcinoma in situ is much less common and originates in the lobule walls [78]. Invasive cancers have spread from their origin site into the fatty tissues of the breast, referred to as stromal invasion, and may have spread into regional lymph nodes or to other more distant locations through the blood

stream. Infiltrating ductal carcinoma makes up the bulk of the invasive cancers while infiltrating lobular, medullary, colloid, and tubular carcinomas and inflammatory breast cancer compose a small percent of the invasive cancers.

3.3 DIAGNOSIS

Because DCIS by definition is localized, patients are usually asymptomatic and detection occurs during screening mammography [79]. For this reason, the incidence of diagnosed DCIS has increased rapidly over the last decade as mammography has become more widely available and awareness of its importance has grown. It is important to note, however, that DCIS is often difficult to diagnose. First, it is a challenge to differentiate DCIS from atypical ductal hyperplasia, a DCIS precursor [80], and, second, approximately 15% of DCIS cases are later found to be invasive cancer once additional testing is done [80]. For invasive cancers, initial detection may occur during a routine mammogram or a physician examination, particularly if the patient is symptomatic or a lump was noted [31]. In both situations, a battery of follow-up procedures, including a diagnostic mammogram, ultrasonography, and biopsy, may be necessary to further characterize the mass.

3.3.1 Breast cancer imaging

Following an abnormal screening mammogram, a diagnostic mammogram is often necessary to further define irregularities. It will also be conducted in symptomatic patients to aid in diagnosis and assess disease spread [31]. While mammography can be particularly useful in patients where physician examination yields uncertain results, such as women with heavy, lumpy breasts [31], the imaging quality will be diminished in women with dense breasts [79]. In this situation, ultrasound may be the most appropriate tool for detecting small non-palpable lesions [79] and allows distinction between tumors and cysts [31, 79]. Magnetic resonance imaging (MRI) has also been used following mammography to evaluate

newly detected lesions although the expense and time required for this procedure limits its utility in most clinical settings [31]. Typically, it is only used in women at a high risk for breast cancer, such as known or suspected *BRCA* carriers [79]. Lastly, positron emission tomography (PET) scans may be conducted when metastatic disease is suspected [79].

3.3.2 Breast cancer biopsies

After imaging has detected a breast tumor, a biopsy will be performed to examine tumor characteristics and pathology. In particular, fine needle and core aspiration biopsies are often conducted since they are inexpensive, easy to perform and minimize patient discomfort [31, 80]. Unfortunately, the accuracy of results for fine needle aspiration is highly dependent on the skill of the physician obtaining the specimen [80] while core aspiration may not be able to find the tissue of small breast tumors [31]. To combat this limitation, core aspiration may be performed with x-ray technology to help penetrate small tumors [31]. Lastly, surgical biopsy may be necessary if the tumor is non-palpable[80].

3.4 PATHOLOGY

3.4.1 DCIS

DCIS tumors are categorized into one of three grading categories: high, intermediate and low. These classifications have implications on recurrence rates and the probability of progression to invasive cancer. For example, high grade is associated with a high rate of recurrence with frequent progression to invasive breast cancer while low grade has low probability of recurrence and tends to remain localized [80]. Histological architecture is also an important feature of the tumor and plays a role in grading assignment and evaluating future recurrences. The comedo type is typically a high grade tumor with a high degree of necrosis and low cell differentiation [81]. In contrast, cribriform, papillary and micropapillary

subtypes tend to be low grade tumors with infrequent mitoses and varying cell sizes [81]. Table 3.1 displays the principle pathologic features of interest in *in situ* breast cancers. It is important to note that hormone receptor status can also be ascertained in these tumors, but is not routinely performed because of the high treatment success rate [80].

TABLE 3.1. In situ breast cancer pathology features of interest from Bilous *et al.* [80]

Nuclear Grade
- High grade
- Intermediate grade
- Low grade

Lesion size

Architecture
- Comedo
- Solid
- Cribriform
- Micropapillary
- Papillary

Margin status
- Margins involved with tumor
- Margins clear of tumor cells

3.4.2 Invasive breast cancer

Using surgical biopsy samples, important pathological characteristics of invasive tumor can be investigated. These characteristics help define key prognostic and predictive markers that may influence treatment choices and allow for grading of the tumor. Table 3.2 provides a concise description of select pathological tumor characteristics and their implications on prognosis and treatment response.

TABLE 3.2. Invasive breast cancer pathology features of interest from Bilous *et al.* [80]

Histological Grade
- Grades 1-3 based on nuclear grade, tubule formation and mitotic rate

Tumor size

Lymph node involvement
- The number of nodes dissected and the number involved with metastatic carcinoma

Lymphovascular invasion
- Evidence of tumor cells invading into vessels

Tumor receptors
- ER and PR receptor status
- HER2/neu status

3.4.2.1 Prognostic and predictive factors. Prognostic factors can be described as markers that "provide information on the clinical outcome at the time of diagnosis, independent of therapy" [82]. Common prognostic factors for invasive breast cancer include tumor size, lymph node status, lymphovascular invasion, and histological grade [80, 82]. In contrast, predictive factors detail the probability of a positive treatment response to a specific therapy [82] and include hormone receptor status. Interestingly, HER-2 is believed to be both a prognostic and predictive factor since it indicates a poor prognosis and also suggests a specific line of treatment [80, 82]. The following bullets will highlight the two most widely accepted prognostic and predictive factors and their implication on either disease prognosis or treatment response.

- Estrogen (ER) and progesterone (PR) receptor status

Hormone receptor status plays an important role in treatment planning and may also indicate tumor etiology. For example, obesity has been linked to hormone receptor positive tumors, which is consistent with research indicating estrogen levels are a key component of breast cancer development [83]. Perhaps most importantly, however, hormone receptor status is an indicator of response to endocrine therapy for metastatic disease or as an

adjuvant therapy option [82]. Patients with hormone receptive positive tumors, which are the most frequently diagnosed tumor in U.S. women, show increased survival compared to patients with receptor negative tumors, and hormone receptor status has been linked to disease stability or the likelihood of metastasis [84]. Receptor positive tumors are treated with hormone therapy, which has been shown to be clinically effective even in metastatic breast cancers and has less negative side effects than chemotherapy [84]. The purpose of hormone therapy (*e.g.* Tamoxifen and Fareston) is to block the production of estrogen and/or progesterone to minimize the proliferation of the cancer cells. Receptor negative tumors do not respond to hormone therapy and, for those patients, systemic chemotherapy remains the most appropriate treatment approach [84].

- Human epidermal growth factor receptor (HER-2/Neu, also called c-erbB-2)

According to the National Comprehensive Cancer Network, approximately 33% of breast cancers have highly overexpressed HER-2/neu genes or have multiple copies of the gene, leading to high HER-2/neu protein levels [78]. HER-2/neu is a member of the epithelial growth factor receptor family and is involved in the signaling cascade that ultimately results in cell growth and differentiation [85]. Given this, it is not surprising that HER-2/neu tumors are associated with poor prognosis. Recent research has indicated that these tumors will not respond to endocrine therapy (although this remains controversial) [85]. Trastuzumab (brand name: Herceptin), a HER-2/neu agonist, is typically advocated at treatment in this tumor type [78, 85].

3.4.2.2 Grading and Staging. Unlike DCIS tumors, invasive breast cancers are graded not only on nuclear grade, but also on tubule formation and mitotic rate, which is an indicator of cell proliferation [78, 80]. Based on these three factors, tumors are assigned a

grade (I-III) to better characterize the observed histological pattern and give a quantitative measure of the tumor's aggressiveness.

While both breast cancer stage and grade assist in the treatment decision process and in the prognostic evaluation, stage incorporates histological tumor characteristics, such as disease spread, and clinical features to provide a more detailed categorization of disease. For this reason, integrating cancer stage is a key component in any survival analysis. Currently, the TNM staging system [86], developed by the International Union Against Cancer, is upheld as the universal standard for breast cancer staging [31]. This staging was updated in 2002 to adapt to the widespread use of screening mammography, leading to early detection of small tumors, and changes in treatment standards, including greater use of sentinel lymph node dissection [87]. The most current TNM staging criteria are described below while Appendix C describes in greater detail the differences between the 5th and 6th edition of the TNM classifications.

Briefly, the TNM classification system incorporates information on three characteristics: (1) T for tumor size and spread, (2) N for lymph node involvement, which has both a clinical and pathological component, and (3) M for metastases to distant organs. These three assessments are then combined to assign an overall staging classification of 0, I, II (A or B), III (A, B, or C) or IV with 0 indicating in situ cancers and IV representing extremely advanced, metastatic disease [31, 78]. The T category will range from Tis, which is used to indicate non-invasive cancers, to T4, which indicates a cancer that has spread to the chest wall (regardless of tumor size). T3 notes large tumors (> 5cm) that have not spread to nearby tissue. For the N category, N0 is the most favorable, where cancer has not spread to the lymph nodes, while N2 describes the most diffuse cancer (more than 10 lymph nodes involved). Lastly, M0 indicates non-metastatic cancer while M1 notes cancer that has spread to distant organs. Table 3.3, adapted from the National Comprehensive Cancer Network [78], displays the TNM staging system and the different categories for evaluation.

TABLE 3.3. TNM staging system from the National Comprehensive Cancer Network [78]

Overall stage	T Category	N Category	M category
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	Any N	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

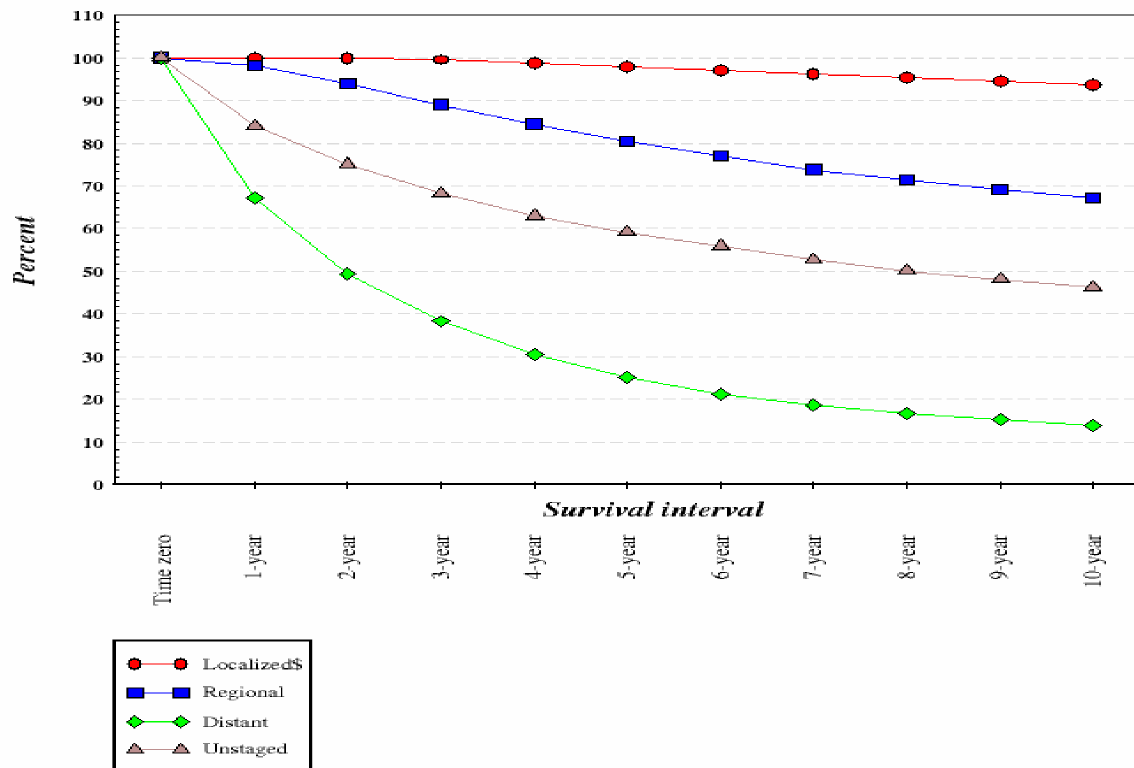
3.5 TRENDS IN BREAST CANCER SURVIVAL

Before survival trends can be discussed, it is important to define survival. Typical statistical analyses will use two different survival outcomes, observed survival and corrected survival. Observed survival is simply the proportion of patients alive beginning at date of diagnosis and ending at a specific time point without regard to cause of death while corrected survival only includes deaths attributable to the breast cancer diagnosis [38]. Most published statistics examining trends in survival, however, use relative survival rates, defined as the ratio of the observed survival in the cohort of cancer patients and the expected survival [38]. Expected survival will depend on the context and relative survival rates may compare survival in specific groups, such as women with ER/PR positive tumors to women with ER/PR negative tumors.

As mentioned earlier, stage plays an important role in prognosis and survival. As illustrated in Figure 3.1, localized breast cancer has the best prognosis with a 5-year relative survival of 98.5% while distant, metastatic breast cancer patients fare the worst with only a

27.7% relative survival. Unstaged cancers are also important to consider since a variety of relevant factors may influence whether a tumor is staged, including access to care, patient health, and metastatic status. Given this, it is not completely surprising to see that unstaged cancer patients have a poor prognosis compared to patients with localized and regional tumors.

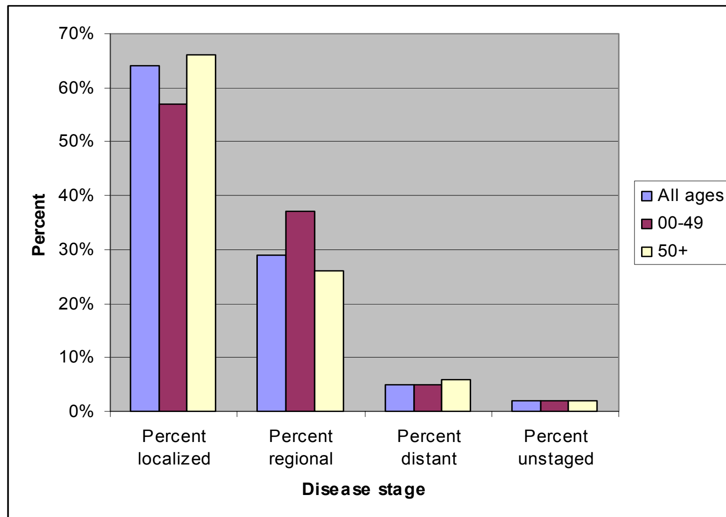
FIGURE 3.1. SEER Relative Survival Rates by Stage at Diagnosis in Whites using data from the SEER 9 Registries for 1988-2001 [88]



Age at diagnosis has also been shown to influence survival. Pre-menopausal women are less likely to have localized cancer and more likely to be diagnosed with regional disease (see Figure 3.2). This trend is possibly due to the different etiology of pre-menopausal breast cancer, leading to more aggressive cancers, or the low prevalence of screening in the population [89]. According to Surveillance, Epidemiology and End Results

(SEER), the 5-year relative survival is 82% in women under the age of 40, but is approximately 88% in women 40 years of age or older independent of disease stage [88, 89]. Stage at diagnosis in these women could, in part, explain these observed differences in survival (Figure 3.2).

FIGURE 3.2. SEER Stage Distribution Age Group in Whites using data from the SEER 9 Registries for 1988-2001



3.6 Epidemiology of Breast Cancer Survival

Many of the same themes from the breast cancer risk factor epidemiology have emerged as important predictors of breast cancer prognosis and survival. For example, parity, exogenous hormone use, diet, and obesity have all been extensively studied to determine their influence on survival although the mechanism of action remains uncertain for many of these factors. The following section will discuss known and suspected prognostic factors found in the breast cancer survival literature.

3.6.1 Reproductive factors

Parity and time between last birth and diagnosis are emerging as important reproductive factors influencing mortality from breast cancer in premenopausal women. For

example, several studies have found that the hazards of death are highest in women who had given birth 2 years prior to their breast cancer diagnosis (Hazard Ratio (HR)~2.0) [90-92]. This risk quickly declines with a larger time interval between last birth and diagnosis, so that no differences in mortality are observed in women with 5 or more years between last birth and diagnosis compared to nulliparous cases [90-92]. These findings are consistent with the hypothesis that pregnancy induces cell proliferation, increasing malignancy in the short-term, but decreases carcinogenesis in the long term. Several studies have also found that parity was associated with decreased survival compared to nulliparity women [90-92], suggesting that pregnancy alone may have detrimental effects on prognosis in pre-menopausal women.

3.6.2 Exogenous hormones

As will be discussed in Chapter 5 on research methods, the survival portion of this study will use covariate data obtained at the baseline interview; therefore, this literature review will focus on cases who were taking exogenous hormones at the time of diagnosis. There is a growing body of research detailing the impact of taking HRT following a breast cancer diagnosis on survival, but this literature is not relevant to the current study and will not be detailed. Instead, this literature review will focus on HRT use before diagnosis.

Several studies have investigated the impact of HRT use at time of diagnosis in connection to breast cancer survival. In a Swedish cohort of 948 cases, ever use of HRT was associated with higher overall survival compared to never users even after adjusting for tumor stage (HR=0.78; 95% CI: 0.65-0.93) [93]. This finding is consistent with several other studies that have examined this association, all of which found improved survival with HRT use prior to the breast cancer diagnosis [94-96]. In addition, HRT use has been investigated as potentially influencing histological and prognostic tumor characteristics. The Women's Health Initiative, a clinical trial of 16,608 post-menopausal women randomized to receive

either placebo or a specific estrogen-progesterone HRT regimen, found HRT use was associated with larger tumor size and more advanced disease compared to placebo [97]. These results, however, oppose the majority of observational study findings that have indicated HRT use is associated with decreased tumor size, better tumor differentiation and more localized disease. The authors suggest that these differences could suggest surveillance bias in the observational studies since HRT users may be more likely to see a physician regularly and be screened more frequently or it could be a function of an older WHI study population than those typically found in the observational studies (approximately 66% of the WHI population was over the age of 60). An estimated 25% of WHI participants in both the active and placebo groups were either current or past HRT users at the start of the trial, so selection bias by HRT use is not likely to explain this difference [97].

The relationship between oral contraceptive (OC) use and breast cancer survival remains controversial. The three studies that have investigated ever versus never use of OCs and breast cancer survival have found different results: one found null results [98], one found a slightly elevated hazard ratio [99] and one improved survival [100]. Further, among the studies examining duration of OC use, differing results were also found with one finding no difference with duration of use [98], one finding increased survival with short term use [101] and another with long term use [100]. Thus, while there is a suggestion of a prognostic impact with OC use, the direction of the associations is not consistent and the importance of duration of use has not been clearly established.

3.6.3 Diet

Research on diet and breast cancer survival has primarily focused on dietary factors that have been found to influence risk of breast cancer, but have been met with very limited success. Differences in exposure categorizations and range of dietary values observed in the cohort (particularly relevant when percentiles are used as cut points) make comparison

difficult. For example, several studies have investigated the association between alcohol intake and breast cancer survival [102-105], but the magnitude of effects differed dramatically. Two studies found null results [103, 105] for hazard ratios per two drinks a week and comparing individuals who drank 5.0-14.9 grams per day to non-drinkers respectively. In contrast, Zhang *et al.* found an inverse hazard of death comparing women who drank more than 4 grams/day to non-drinkers (HR=0.7; 95% CI: 0.3, 1.5) [102] while Ewertz *et al.* reported decreased survival with consumption of over 121 grams of alcohol per week to no alcohol consumption (HR=1.26; 95% CI: 0.90, 1.74) [104].

Similarly, several studies have examined dietary fiber intake and breast cancer survival [106] although considerable variation in the point estimates is evident. In particular, the Nurses' Health Study found a HR of 0.69 (95% CI: 0.50, 0.97) comparing cases who had eaten over 20 grams of fiber per day to those who had consumed less than 12.5 grams of fiber per day [105]. Another study found increased hazard of death with fiber consumption (HR=1.17; 95% CI: 0.83, 1.66 per 20 grams/day).

While the evidence of an association between fat intake and risk of breast cancer was not convincing, it may influence survival. Among studies that investigated total fat intake and breast cancer survival [102-104, 107, 108], hazard ratios ranged from 0.96 (95% CI: 0.75, 1.22) comparing fourth to first total fat quartiles (values not presented) [104] to 2.1 (95% CI: 0.11, 4.3) comparing women who consumed 56-76 grams/day to 19-56 grams/day [102]. When only saturated fat consumption is examined, a consistent trend towards decreased survival with high saturated fat intake is seen. Among the five studies that examined saturated fat in connection to breast cancer survival [102, 103, 105, 107, 108], HRs ranged from 1.13 (95% CI: 1.05-1.22) for a twofold increase in percent energy [108] to 1.72 (95% CI: 1.00, 2.96) per 20 grams/day [103]. It is important to note, however, that residual confounding may be present. Fat intake is correlated with energy consumption and obesity [106]. Because of this, it may be necessary to adjust for these possible confounders

in the analysis although some consider energy consumption and obesity to be causal intermediates, and therefore, should not be adjusted for in analyses.

3.6.4 Smoking

Although only a small number of studies have investigated smoking and breast cancer prognosis, they consistently show reduced survival in smokers compared to non-smokers. In a study of 792 breast cancer cases in Malmo, Sweden, a breast cancer-specific HR of 2.14 (95% CI: 1.47-3.10) comparing current smokers to never smokers, adjusting for stage and age [109]. This finding aligns with two other studies which both found decreased survival in smokers but no histological, size or stage differences in the cancers [110, 111]. Several explanations have been presented to explain this observation including compromised immune systems in smokers as well as dietary and socio-economic status differences between smokers and non-smokers that may confound the association [109].

3.6.5 Obesity and Physical Activity

Many epidemiologic studies have examined the association between obesity and breast cancer survival. Appendix B highlights recent research examining this association along with study characteristics and important findings. Hazard ratios and relative risks range between 0.60 to 3.3 and most studies adjusted for ER/PR status, age at diagnosis and hormone therapy use. Like breast cancer incidence, a growing body of literature suggests an association with obesity and negative prognosis in post-menopausal women. Recent research has also implicated obesity and decreased survival in pre-menopausal women, however.

Using data from the LIBCSP, Cleveland *et al.* found decreased overall survival with obesity (BMI ≥ 30 kg/m^2) compared to normal weight (BMI < 24.9 kg/m^2) in pre- and post-menopausal women (HR=2.62, 95% CI: 1.26, 5.45; HR=1.63, 95% CI: 1.08, 2.45

respectively) [112]. Consistent with these results, Abrahamson *et al.* observed increased all cause mortality with obesity (BMI ≥ 30 kg/m^2 versus BMI between 18.5 and 24.9) in a population of primarily pre-menopausal women (HR=1.65; 95% CI: 1.23, 2.21) [113]. Similar results for all cause mortality were noted in a cohort of pre- and post-menopausal women comparing obese patients (BMI ≥ 30 kg/m^2) to normal weight cases (BMI < 25.0 kg/m^2) (HR=1.53; 95% CI: 1.37, 1.72) [114]. In contrast, while the Million Women Study in the United Kingdom found an association with increased mortality in post-menopausal women (HR=1.49; 95% CI: 1.27, 1.75) an inverse association with mortality was noted in pre-menopausal women comparing obese to normal weight women (HR=0.64; 95% CI: 0.34, 1.21) [115]. Thus, while obesity is consistently associated with decreased survival in post-menopausal women, the association remains controversial in pre-menopausal women. In the LIBCSP, however, obesity clearly implicates mortality in both pre- and post-menopausal women.

In addition to its association with poorer survival in both pre- and post-menopausal women [112-115], high BMI has also been linked to later stage at diagnosis [116, 117], larger tumor size [116, 118], and positive estrogen receptor status [83]. This difference in prognosis has been suggested to be the result of increased breast density with obesity, which hinders the diagnostic ability of mammograms [116], but this would not explain the observed negative impact of obesity on survival in early stage breast cancer [119-121]. Alternative explanations have ranged from obesity leading to more aggressive forms of breast cancer due to increased levels of insulin and leptin in these patients to differences in adjuvant therapy use in obese versus normal weight patients [121]. An article by Colleoni *et al.* suggest that obese patients with hormone receptor negative breast cancer are often given reduced chemotherapy doses and that this reduction has negative implications on prognosis [122].

Very few studies have investigated physical activity prior to breast cancer diagnosis in connection to breast cancer survival [123-125]. Two studies found no associations with any kind of physical activity at diagnosis and breast cancer mortality or consistency when women were stratified by menopause status [123, 124]. These studies may have been hampered by small sample sizes, however, since they only included 412 and 603 patients respectively. In contrast, Abrahamson *et al.* using data from a cohort of 1,264 primarily pre-menopausal women found physical activity one year prior to diagnosis was associated with decreased mortality (HR=0.78; 95% CI: 0.56, 1.08), comparing highest quartile to lowest quartile of activity [125]. The effect of PA was modified by obesity with overweight or obese women showing a greater survival benefit than women of normal weight.

3.6.6 Genetic Factors

3.6.6.1 BRCA1 and BRCA2. *BRCA1*-related breast cancers have been suggested to have a negative impact on survival. These cancers tend to be hormone receptor negative, poorly differentiated and HER-2/neu negative [126], factors which are known to have an adverse effect on prognosis. Studies examining this relationship range from linkage studies and smaller clinic-based explorations to larger cohorts of Ashkenazi Jews, who have much higher *BRCA1* prevalence than those found in the general population. While previous studies have demonstrated negative impact on survival among carriers of the *BRCA1* mutation [126], recent research has been much less conclusive [127-129]. In a national cohort of 1,794 Israeli women, Rennert *et al.* found no evidence of a 10-year overall survival difference in *BRCA1* carriers (n=76) compared to non-carriers (HR=1.09; 95% CI: 0.79, 1.51) [127]. It is important to note that no prognostic differences were found with tumor stage and nodal status in this cohort, casting doubt on the validity of their findings [130]. This finding is consistent with El-Tamer *et al.* who found no evidence of a survival difference comparing Kaplan-Meier curves [129], but not with Brekelmans *et al.* who found elevated

hazard of death in *BRCA1* carriers compared to sporadic breast cancer cases (HR=1.30; 95% CI: 0.91, 1.85) [128].

Like *BRCA1* cancers, *BRCA2* breast cancers tend to be of higher grade than sporadic breast cancers although *BRCA1* and *2* cancers differ histologically from one another [131]. *BRCA2* breast cancers tend to be more tubular and have a larger intraductal component than cancers associated with *BRCA1* [131]. Based on eight studies, survival appears to be slightly worse for *BRCA2* carriers compared to non-carriers although the evidence is much less convincing than the *BRCA1* literature [127-129, 131]. In general, sample sizes were small and studies were based primarily on convenience samples.

3.6.6.2 Low penetrance polymorphisms. Very few studies have investigated low penetrance genetic polymorphisms in relation to breast cancer survival, but they are becoming more common in the literature. Research has focused on mutations in xenobiotic metabolizing genes, such as the cytochrome P450 (CYP) family, and oxidative stress, like GST family and *MnSOD*, although small sample sizes have limited the ability of these studies to detect associations.

Xenobiotic metabolism. Several studies have investigated the association between *CYP2D6* and breast cancer survival, particularly among patients receiving tamoxifen. *CYP2D6* is a Phase I enzyme and genetic variants have implications on the rapidity of metabolism [132]. All four studies investigating *CYP2D6* in connection to breast cancer survival have suggested carriers of *CYP2D6* to show improved survival compared to non-carriers [133-136]. For example, Nowell *et al.* found carriers of the *CYP2D6**4 allele to have a HR of 0.77 (95% CI: 0.32, 1.81) compared to non-carriers [135]. Wegman *et al.*, examining recurrence-free survival, noted inverse hazards of death in *CYP2D6* heterozygotes and *4 homozygotes combined compared to *1 homozygotes (HR=0.33; 95% CI: 0.08, 1.43) [136].

Oxidative stress. In addition to being examined as a risk factor for breast cancer incidence, the *MnSOD* polymorphism leading to a valine to alanine amino acid substitution at codon 16 has also been investigated in connection to breast cancer survival. In a cohort of 251 women identified through the Arkansas Tumor Registry, individuals homozygous for the Ala allele showed improved survival compared to Val homozygotes (HR=0.66; 95% CI: 0.34-1.29) even after adjustment for stage, race, and ER and PR status [137]. Using this same study population, deletions in the *GSTM1* and *GSTT1* gene were also investigated in relation to breast cancer survival [138]. In both genes, the null genotype was associated with increased survival (HR=0.50, 95% CI: 0.36-0.97 for *GSTM1*; HR=0.51, 95% CI: 0.29-0.90 for *GSTT1*) with adjustment for age, race, stage and node status [138]. These results are comparable to those found by Syamala *et al.* who examined *GSTM1* and *GSTT1* deletions in a small cohort 222 sporadic breast cancer cases in South India [139]. More recently, Yang *et al.* found no differences in survival for cases with the null *GSTM1* and *GSTT1* genotype (HR=1.1, 95% CI: 0.8, 1.5 and HR=1.0, 95% 0.7, 1.4 respectively) in a cohort of 1,034 cases in Shanghai [140].

3.7 Conclusion

While breast cancer survival has improved in women with localized disease, prognosis is still poor for women with metastatic breast cancer. Advances in tumor pathology have helped to determine appropriate treatment options for various breast cancer subtypes, but options remain limited for women with systemic disease. Lastly, recent research has implicated BMI at diagnosis and increased mortality. *PPARA* has been linked to obesity (discussed in the next section) and examining this gene in connection with breast cancer survival could help elucidate this pathway.

CHAPTER 4 *PPARA*

4.1 The Biology of *PPARA*

4.1.1. Introduction

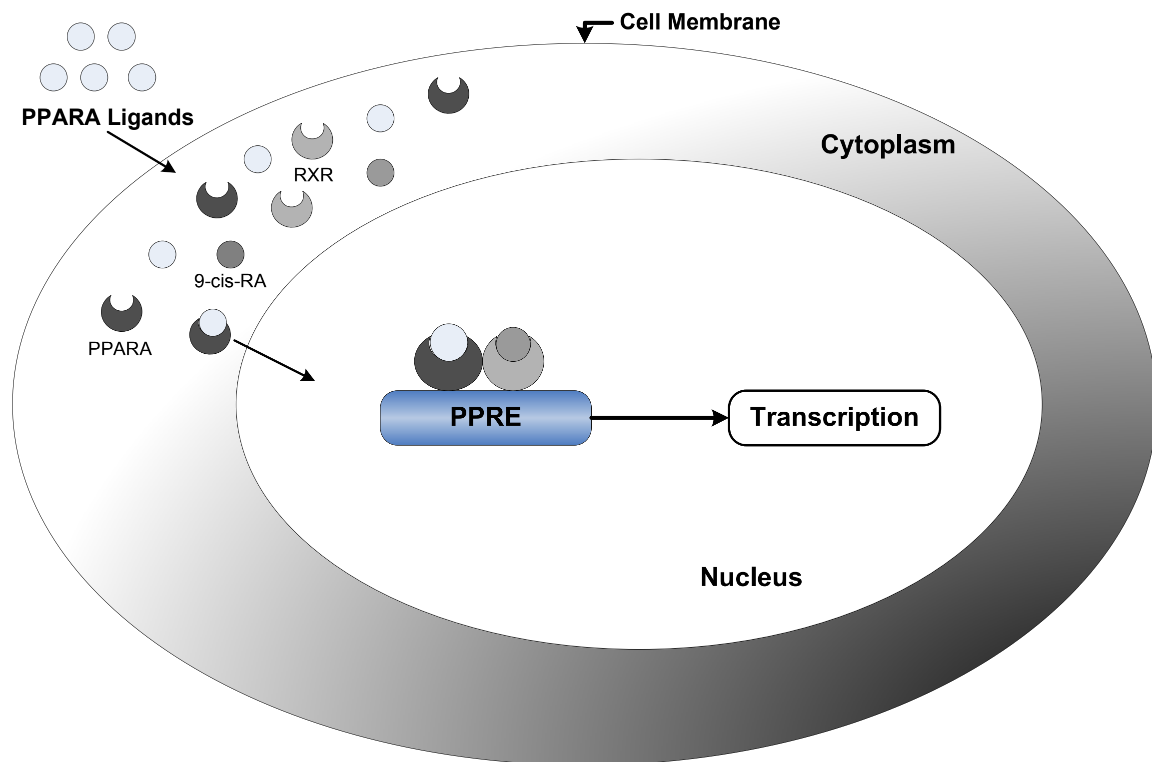
The peroxisome proliferator-activated receptor (PPAR) family is composed of three nuclear hormone receptor genes: PPAR-gamma (*PPARG*), PPAR-alpha (*PPARA*) and PPAR-delta (*PPARD*). In general, hormone nuclear receptor genes encode proteins that induce gene transcription by binding to the promoter region of a target gene [141]. The protein structure of nuclear hormone receptors has implications for its role in the regulation of gene expression and its potential involvement in breast cancer carcinogenesis. Please note that the PPAR proteins will be referred to using Greek letters (e.g. PPARα) while the gene names will be abbreviated using alphanumeric values (e.g. *PPARA*).

4.1.2 Nuclear Hormone Receptors (NHR)

4.1.2.1 Mechanism. NHRs are activated when small, lipophilic molecules (e.g. hormones for *PPARA*), called ligands, travel across the membrane of the target cell and bind to the ligand-specific NHR [141]. PPARα ligands include palmitic acid, arachidonic acid and stearic acid although exogenous compounds, like fenofibrate, bezafibrate and non-steroidal anti-inflammatory drugs, have also been shown to target the PPARα receptor [142, 143]. After the ligand has bound to the receptor, PPARα will heterodimerize, which allows it to bind with a specific regulatory DNA sequence (*i.e.* the peroxisome proliferator response element (PPRE)), in the target gene. By binding to the PPRE, PPARα induces transcription of the

target gene. The PPRE is typically located in the 5' flanking region of the gene [144] and, for the PPAR family, consists of direct repeats separated by one base pair (DR1) [142, 144]. Retinoid X receptor (RXR) has also been shown to work synergistically with PPARα to activate expression of the target gene through the ligand-dependent activation function 2 (AF-2) [142, 144]. Figure 4.1 illustrates the mechanism of action of PPARα, including ligand diffusion into the cytoplasm, the heterodimeric complex traveling into the nucleus and RXR acting as a co-factor to increase gene expression.

FIGURE 4.1. PPARα mechanism of action to induce gene expression, adapted from Kiec-Wilk *et al.* [145] and Kota *et al.* [142]

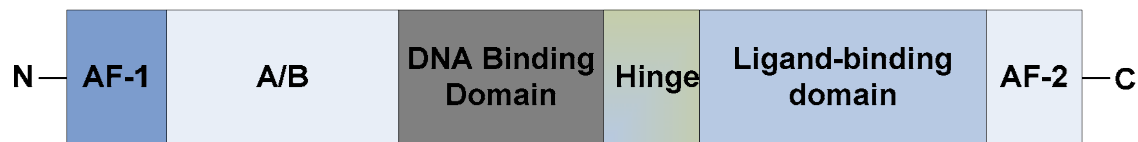


4.1.2.2 Structure. As depicted in Figure 4.2, NHRs are composed of 4 different regions: (1) the A/B domain, (2) the DNA binding domain (DBD), (3) the hinge region, and (4) the ligand binding domain (LBD) [142, 144]. The A/B domain is located at the N terminus and is highly variable within the family of NHR genes; in fact, spliced isoforms will often occur in this region [144]. In PPARα, this region also contains a transcriptional activation function

(AF-1) that allows for transcription activation without the presence of the ligand. Through the AF-1, the phosphorylation of mitogen-activated protein kinase can lead to PPARα activation [144]. For example, insulin has been shown to activate PPARα through this mechanism [143]. In contrast, the DNA binding domain is well conserved and is responsible for recognizing and binding to the PPRE. The hinge region is not well characterized and varies by NHR. Research suggests that it is the link between the DBD and the LBD [144] and the binding domain for cofactors [142]. Lastly, the LBD contains the AF-2 site for ligand-dependent gene transcription and contains the binding "pocket" for the ligand. Once the ligand has bound to the receptor, the LBD provides the dimeritization surface for travel across the nucleus [144].

Interestingly, two *PPARA* isoforms have been characterized (*PPARA*1 and *PPARA*2). *PPARA*1 encodes the entire gene while *PPARA*2 is truncated at exon 6. This truncation results in the absence of the ligand-binding domain in the gene's protein and, consequently, prevents activation by the ligand. Therefore, this study is interested in *PPARA*1 because of its protein's known activity.

FIGURE 4.2. Schematic depiction of the functional domains of PPARα, adapted from Aranda and Pascual [144] and Kota *et al.* [142]

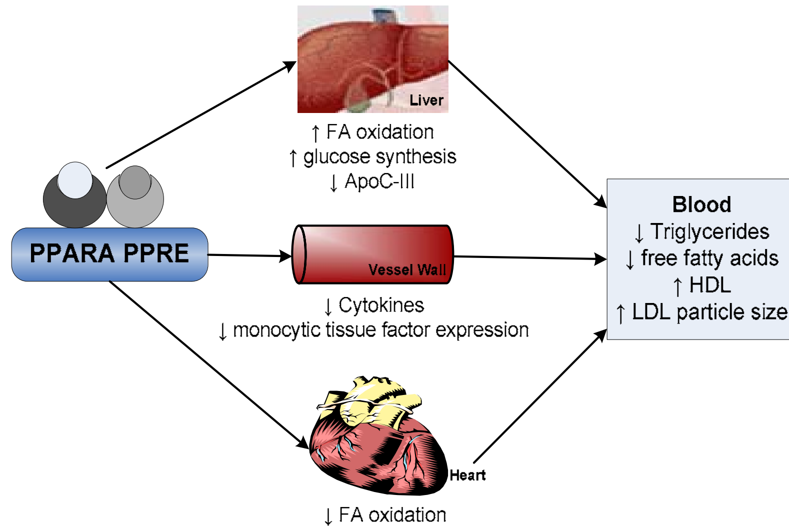


4.1.3 Biologic Mechanisms of PPARα

PPARα has been implicated in fatty acid and lipoprotein metabolism, particularly triglyceride metabolism, as well as atherosclerosis but has not been studied in cancer. Figure 4.3 displays current knowledge about the biologic effects of PPARα activation in

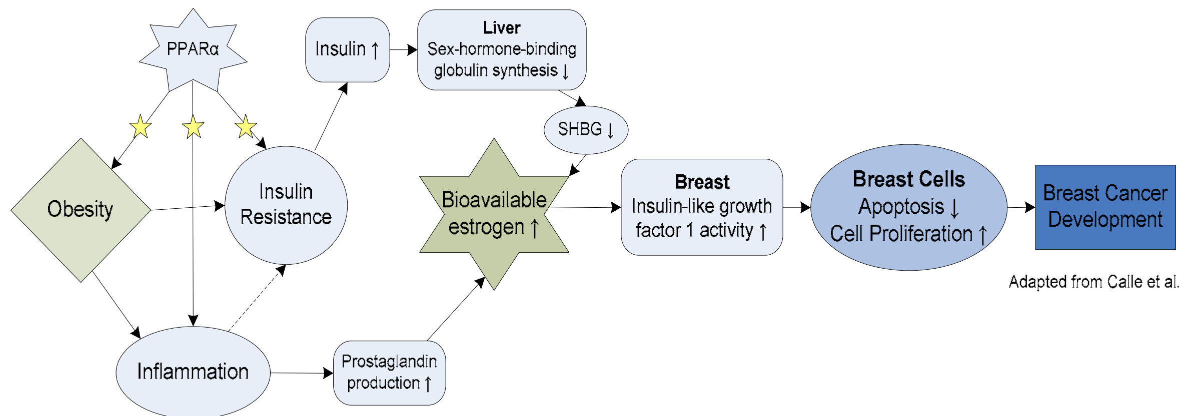
different organ systems while Figure 4.4 uses this information to demonstrate the potential role of PPARα in the obesity-breast cancer relationship.

FIGURE 4.3. Biologic effects of PPARα activation in different organ systems, adapted from Kota *et al.* [142]



As Figure 4.3 suggests, PPARα is primarily expressed in organs with increased fatty acid oxidation rates [143, 146]. In particular, PPARα is largely found in the liver, kidney, heart, skeletal muscle, brown adipose tissue, vascular endothelial cells, vascular smooth muscle cells and macrophages and in small quantities in white adipose tissue [142]. It has also been implicated in multiple pathways, including obesity, inflammation and insulin sensitivity, which are key components of carcinogenesis and illustrate PPARα's potential role in breast cancer development (Figure 4.4). The next section will detail research investigating these three pathways to further describe the associations starred in Figure 4.4.

FIGURE 4.4. A conceptual diagram of PPARα and obesity-mediated breast cancer development, adapted from Calle and Thun [147]



4.1.3.1 Obesity. PPARα has been shown to regulate lipid metabolism by controlling the uptake and oxidation of fatty acids [143]. This regulation influences the availability of free fatty acids, which contribute to insulin resistance [148]. Given this, it is not surprising that PPARα has been shown to influence fat storage in the liver and reduce steatosis although this remains controversial [146]. PPARα also increases glucose levels through its involvement in gluconeogenesis and plays a role in the synthesis of apolipoproteins, including apoAI and apoAII [146]. Fibrates, which are *PPARA* agonists, have also been shown to reduce the expression of multiple cytokines, including interleukin-6, fibrinogen, and C reactive protein in humans [143].

4.1.3.2 Inflammation. PPARα has also been suggested to reduce the expression of multiple cytokines, including interleukin (IL)-1 and IL-6, and inhibit COX-2 activity [143]. For example, fibrates have been shown to reduce IL-6, fibrinogen, C reactive protein, and tumor necrosis factor (TNF)-α [143]. Interestingly, TNF-α suppresses insulin signaling and has been linked to insulin resistance, suggesting a role of PPARα in Type 2 diabetes and insulin sensitivity. Additionally, PPARα decreases monocytic tissue factor expression, found on the surface of monocytes and macrophages in human atherosclerotic lesions [143], and may

induce apoptosis in cytokine-activated macrophages [149]. PPAR's involvement in the inflammatory pathway is consistent with the growing body of literature linking PPARα to atherogenesis and could suggest a possible role in breast cancer development.

4.1.3.3 Insulin Resistance. While the evidence for a relationship between PPARα and insulin resistance is not as convincing as that for inflammation, research suggests that the two could be related. As mentioned earlier, PPARα influences the expression of TNF-α, a cytokine involved in insulin resistance. Additionally, recent research has suggested that PPARα may lead to insulin sensitivity by increasing insulin production in beta cells, possibly due to its involvement in the regulation of free fatty acids [150]. These facts point to a link between PPARα and insulin resistance and sensitivity although the exact mechanism remains uncertain. Given the association between breast cancer incidence and Type II Diabetes Mellitus, described in Chapter 2, these facts point to *PPARA*'s possible involvement in breast cancer development through an insulin resistance pathway.

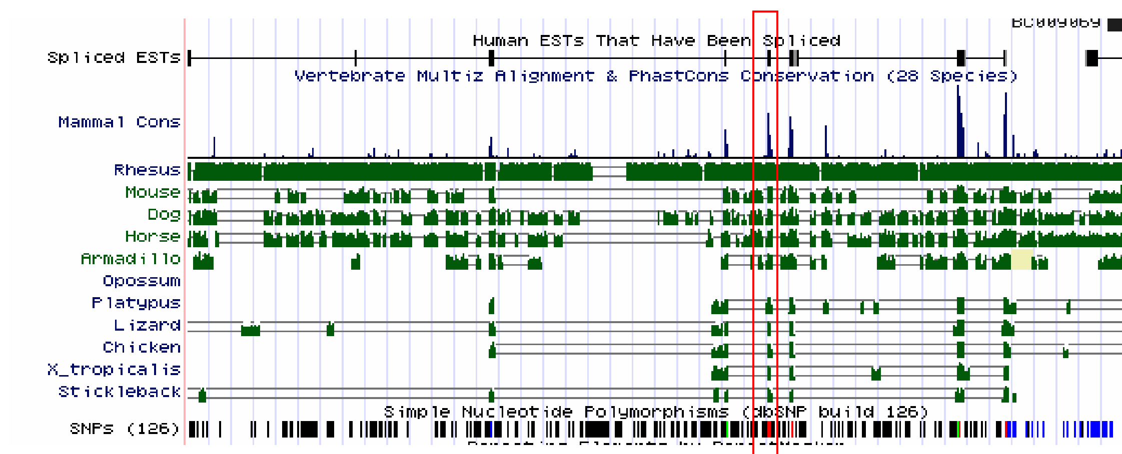
4.2 EPIDEMIOLOGY OF *PPARA*

4.2.1 Introduction

Given its role in energy homeostasis, it seems feasible that genetic variation in *PPARA* could have implications on disease development. In particular, the epidemiologic literature has focused on one functional polymorphism, representing a cytosine to guanine base change. This base change occurs in the DNA binding domain region of the *PPARA* protein and results in a leucine to valine substitution at codon 162 of exon 5 (L162V). This missense polymorphism has been shown to have functional impact on the receptor. Using co-transfection assays, the V162 allele showed elevated ligand-dependent transcription activity compared to the L162 while no differences in activity were seen in the absence of

PPARA ligands.[151] Figure 4.5 details the sequence alignments across 11 mammalian species to assess the evolutionary conservation of PPARA across species using the genome browser from the University of California, Santa Cruz. It is interesting to note that exon 5, which contains the L162V SNP, (highlighted with a red box) is not well-conserved across most species. As indicated by the bar labeled “Mammalian Conservation”, the fifth exon of PPARA varies from species to species, suggesting some evolutionary pressure. Other studied single nucleotide polymorphisms (SNP) include an adenine to cytosine base change in intron 1 of *PPARA*, a guanine to cytosine base change in the intron 7 region and a guanine to adenine base change in the second intronic region. These intronic SNPs have been evaluated in several studies [152-159] although most *PPARA* research has centered on the L162V polymorphism.

FIGURE 4.5. Conservation of L162V across 11 mammalian species from the University of California, Santa Cruz genome browser



The frequencies of these SNPs vary tremendously by study population. For example, in a cohort study of 292 hypertriglyceridemic patients in Quebec, Canada, 27% of the population was determined to be carriers of the V162 allele [160]. In contrast, 2.8% of 395 Finnish men who participated in the LOCAT trial were estimated to carry the V162 allele

[158]. Genetic databases, such as the Program for Genetic Applications (PGA) and HapMap, have estimated the frequency of the V162 allele to be 2.2% and 4.2% respectively in Caucasian populations, pointing to the variability surrounding these allele frequencies. The intronic SNPs are much more prevalent with frequencies ranging from 31.2% [154] to 13.4% [158] for the C Intron 7 allele.

A total of 29 epidemiologic articles have investigated genetic variants of *PPARA* in association with multiple health outcomes, including lipoprotein levels, obesity, fibrate treatment response, Type 2 diabetes and atherosclerosis (see Appendix A) [151-179]. Of these 24 articles, seven publications used clinical trial populations [153, 154, 158, 159, 166, 177, 179], seven used data from cohort studies [152, 158, 160-162, 165, 167, 176], including the Framingham Offspring study and the Quebec Family study, five represent case-control studies [155, 157, 163, 170, 174], primarily convenience samples from hospital clinics (cases) and the general population (controls), and five were cross-sectional studies using convenience samples [164, 168, 169, 175, 178]. One study [158] published data using two different study populations, LOCAT, a clinical trial population, and the Second Northwick Park Heart Study, a cohort study. While most studies have not found an association, there are many possible explanations for this. Small sample sizes, particularly in the case-control studies and cross-sectional studies, and a failure to account for genetic variation across the entire gene could play an important role in limiting the power of these studies for finding an association. Similarly, many of these studies used convenience sampling, which will dramatically influence the quality and accuracy of their findings. The following summary of the literature has been categorized by outcome to help clarify the state of the literature for *PPARA* and a summary of findings is presented in Table 4.1.

4.2.2 Lipoprotein levels

Fifteen studies have examined the association between L162V (rs1800206) and lipoprotein levels [152, 157, 160-164, 166, 167, 174-179]. The evidence of an association is most convincing for apolipoprotein (apo) B. Carriers of the V162 allele showed elevated levels of apoB compared to non-carriers in healthy controls [157, 164, 174] and women [162], although lower levels were noted in Familial Combined Hyperlipidemia patients [157]. Similarly, in a study of 3,012 men in the Second Northwick Park Heart Study, however, apoB levels were slightly lower in V162 carriers compared to L162 homozygotes (0.84±0.28 and 0.88±0.28 respectively) [158].

For LDL and triglycerides (TG), the relationships with *PPARA* are much less consistent. Recent studies have not observed an association between LDL and L162V [158, 175-179] while older studies have noted higher LDL levels in V162 carriers in the Framingham Offspring Study [161], women [162], and healthy controls [163, 174]. More recent studies were typically larger in size and better designed (*e.g.* did not use convenience samples). Higher LDL levels appear more common in L162 homozygotes in diseased cohorts, including Type II Diabetes Mellitus [163] and obese [166] patients. Similarly, V162 carriers appear to have higher TG levels than L162 homozygotes in a wide range of study populations, including clinical trial cohorts [159, 177, 179], young adults [175], men [164]. It is important to note, however, that no differences in LDL levels were noted for several studies, including a large cross-sectional study of healthy adults [174, 178], a Type II DM cohort [174, 176], and the Second Northwick Park Heart and LOCAT studies [158].

In summary, there is evidence of an association between *PPARA* genetic variants and ApoB and limited evidence of an association with LDL and triglyceride levels. HDL does not appear to be associated with *PPARA*. In general, the V162 allele appears to influence

lipoprotein levels in diseased and non-diseased populations although the direction of V162 activity varies by disease status.

4.2.3 Obesity

In general, *PPARA* does appear to be associated with obesity, measured as mean BMI in all studies. The direction of the effects is inconsistent and differences in effect by genotype were minimal. Of the 10 studies examining this association [157, 163-165, 168, 174-176, 178, 179], BMI was slightly higher in V162 homozygotes and heterozygotes compared to L162 homozygotes in four studies [157, 175, 178, 179]. No differences in mean BMI were evident for three studies [163, 164, 174, 176], and V162 carriers had slightly lower BMI values than non-carriers in two studies of Type II DM patients [163, 168] and in a cohort of 393 women [165]. Among studies that examined obesity in more detail, Nieters *et al.* found increased odds of obesity, defined as BMI > 35.0 compared to BMI between 18.5 and 25.0, in V162 allele carriers versus L162 homozygotes (OR=1.35, p=0.45; no CIs reported) [170]. Consistent with that finding, Bosse *et al.* noted an obesity (BMI > 30 vs. BMI ≤ 30) OR of 1.77 (p=0.04, no CIs reported) for V162 carriers compared to non-carriers, adjusting for age, gender and alcohol consumption [165]. While multiple studies have investigated PPARA in connection to obesity, the literature is primarily limited to studies using very rudimentary statistics to compare BMI means in carriers versus non-carriers. The two studies that have used multivariable methods are of dubious quality and failed to report valuable information (*e.g.* confidence intervals) that would have aided the interpretation of their results.

4.2.4 Coronary Heart Disease (CHD)

Among Type II Diabetes Mellitus patients, no difference in V162 allele frequency was observed between patients with or without CHD (6.6% vs. 6.4% respectively) [173]. In

contrast, another study examining the L162V polymorphism and CHD found the variant had no influence on risk of CHD in the non-diseased control group (OR=1.03, 95% CI: 0.12-8.7) but may play a role in CHD development in diabetics (OR=0.44, 95% CI: 0.18-1.11) [163]. Using survival analysis, Flavell *et al.* [158] reported a hazard ratio of 0.75 (95% CI: 0.45-1.26) for the risk of ischemic heart disease in V162 carriers vs. L162 homozygotes, suggesting that the V162 allele may reduce risk of IHD.

Studies focusing on the common L162V polymorphism and atherosclerosis have been more consistent with one study suggesting lower odds of atherosclerosis among individuals with Type II diabetes (OR=0.49, 95% CI: 0.21-1.15) and in controls (OR=0.68, 95% CI: 0.20-2.33) [163]. Another implicated the polymorphism in atherosclerotic progression, measured as change in average diameter of coronary segments and change in minimum luminal diameter [158]. This research provides limited evidence that *PPARA* influences risk of atherosclerosis (although more research is needed), but the results are less convincing for CHD.

4.2.5 Type II Diabetes Mellitus

Only four studies have examined *PPARA* in connection to Type II DM [152, 176, 178, 179]. Comparing the percentage of individuals with diabetes in V162 carriers versus L162 homozygotes yielded inconsistent results with one study finding no differences (13.0%, n=144 in L162 homozygotes vs. 12.6%, n=12 in V162 carriers) [178] and one indicating slightly higher prevalence of Type II DM in V162 carriers (25.0% in L162 homozygotes vs. 37.0% in V162 carriers) [179]. Interestingly, Tai *et al.* found increased hazards of stroke, non-fatal myocardial infarction and CHD death (combined) in V162 carriers among individuals with no history of insulin resistance or diabetes mellitus (HR=1.63; 95% CI: 0.82, 3.24) but an inverse association for the same outcome among those with diabetes mellitus or insulin resistance (HR=0.40; 95% CI: 0.17, 0.92). Thus, Type II DM status may modify

associations with *PPARA* and cardiovascular disease outcomes. Another study found V162 carriers to have an earlier age of diagnosis with Type II DM compared to non-carriers although effect estimates were imprecise (beta coefficient=2.6, 95% CI: 0.2, 5.1) [176].

Related to Type II DM, fasting glucose levels have also been investigated in connection to L162V. While Robitaille *et al*. found no difference in glucose levels in V162 carriers versus L162 homozygotes (5.39±0.75 mmol/L, 5.36±0.72 mmol/L respectively) [164], V162 carriers had lower levels in three other studies [167, 174, 175]. Less studied, fasting insulin levels were slightly lower in V162 carriers compared to non-carriers in two studies [167, 174], suggesting more research is needed to confirm these findings.

4.2.6 Fibrate Response

Interest in *PPARA*'s potential role in predicting reduced efficacy of fibrate treatment was spurred initially by Flavell *et al.* who showed larger changes in total cholesterol and non-HDL cholesterol in V162 allele carriers (6.31 mmol/L, 1.09 mmol/L respectively) compared to L162 homozygotes (5.77 mmol/L, 0.99 mmol/L respectively) and a slight difference in change in ApoB levels (1.48 mmol/L in V162 carriers, 1.34 in L162 homozygotes) in 2000 [151]. Since then, an additional seven studies [153, 154, 158-160, 166, 172] have been published on this subject with mixed results. Most recently, Chen *et al.* found no association between *PPARA* haplotypes and response to fluvastatin [153]. Two studies that examined changes in lipoprotein levels following gemfibrozil therapy found no difference in triglyceride levels or LDL [158, 166] although a slight difference in HDL levels was suggested [166]. Similarly, Brisson *et al.* [160] and Foucher *et al.* [154] both investigated response to fenofibrate treatment by L162V genotype and found no association with residual hypertriglyceridemia or triglyceride level reduction respectively. These results are more consistent with Puckey *et al.* [172] who reported that the L162V polymorphism was not associated with changes in HDL, total cholesterol or triacylglycerol in a group of patients

treated with fibrates. Interestingly, among individuals who received bezafibrate, the V162 allele was associated with a decrease in triglyceride levels but not fibrinogen [159]. These results suggest that *PPARA* variants do not influence response to fibrate treatment even though fibrates are known *PPARA* agonists.

4.2.7 Other diseases

Lastly, the one case-control study that has examined the association between the L162V polymorphism and *PPARA* found increased odds of Alzheimer's disease with the V162 allele (OR=2.24, 95% CI: 1.12-4.50) after adjusting for gender and BMI, suggesting that *PPARA* could be involved in Alzheimer's disease development [155]. Additionally, Ishiguro *et al.* investigated the Val227Ala polymorphism in connection to schizophrenia and alcohol dependence but found no differences in minor allele frequency between the two diseases and the control group [171].

TABLE 4.1. Summary of evidence based on epidemiology studies of *PPARA* and various disease outcomes

	Association			
Outcome	**Probable**	**Possible**	**Not Likely**	**Insufficient data**
Lipoproteins				
apoA1				X
apoB	X			
apoC				X
HDL			X	
LDL		X		
Triglyceride		X		
Obesity		X		
Coronary Heart Disease		X		
Type II Diabetes Mellitus	X			
Insulin		X		
Glucose		X		
Atherosclerosis		X		
Metabolic Syndrome		X		
Fibrate Response			X	
Alzheimers				X
Alcohol Dependence				X
Schizophrenia				X

4.3 CONCLUSION

Although no research to date has investigated the possible link between *PPARA* and breast cancer, the biology of PPARα and the epidemiology of *PPARA* suggest it could play a role in breast cancer development. PPARα has been shown to increase fatty acid oxidation and decrease cytokine levels and may be involved in insulin production, supporting its involvement in three critical pathways for breast cancer development. Additionally, genetic variants of *PPARA* have been linked to several lipoprotein levels, cardiovascular disease, obesity and type II diabetes, diseases that share many mechanistic similarities with breast cancer.

CHAPTER 5 RESEARCH DESIGN AND METHODS

5.1 Study Overview

This study proposed to assess whether interindividual variability in *PPARA* influences breast cancer risk and survival. *PPARA* has been hypothesized to affect breast cancer development and survival through two different but related mechanisms: (1) insulin sensitivity, and (2) anti-inflammatory effects. This study explored whether genetic polymorphisms in *PPARA* were associated with breast cancer incidence and survival, and whether they interacted with body mass index (BMI), weight gain and NSAID use to influence disease development and prognosis.

To determine whether *PPARA* was associated with breast cancer or interacted with body mass index or weight gain to influence risk or breast cancer survival, data from the Long Island Breast Cancer Study Project (LIBCSP), a large population-based case-control study, was used. First, genotyping for *PPARA* was conducted using available, banked DNA from 1,052 breast cancer cases and 1,098 controls. The prevalence of the *PPARA* alleles was determined in this population-based sample of breast cancer cases and women without breast cancer. Second, maximum likelihood estimates of the haplotype frequencies were determined. Third, statistical analyses were conducted to determine whether *PPARA* haplotypes were associated with breast cancer risk and survival. Fourth, genotype data was then coupled with baseline questionnaire data (that was already collected as part of the parent case-control study) to explore gene-environment interactions with BMI, post-menopausal weight gain (prior to diagnosis) and NSAID use. Corresponding survival

analyses were also conducted.

The following details the LIBCSP and follow-up study population and design, followed by a description of the laboratory assays, statistical analyses, and study power, ending with a discussion on the strengths and limitations of this ancillary study.

5.2 LONG ISLAND BREAST CANCER STUDY PROJECT

5.2.1 Subject Eligibility

Women newly diagnosed with a primary in situ or invasive breast cancer between August 1, 1996, and July 31, 1997, and who were 20 years of age or older at diagnosis and were residents of Nassau or Suffolk counties in Long Island, New York, were eligible as subjects for the case-control study. Controls were selected from among female residents of the same two Long Island counties and were frequency matched to the expected age-distribution of case subjects by 5-year age group. At the request of the National Cancer Institute, there were no upper age restrictions on study subjects. All subjects were required to speak English and census data indicates that over 97% of all residents on Long Island are English-speaking. Figure 5.1 displays a map of Long Island, including Suffolk and Nassau counties as well as the surrounding area.

FIGURE 5.1. Map of Long Island

5.2.2 Case Identification

For this case-control study, cases were identified using a "super" rapid reporting system established by the study investigators for the LIBCSP, resulting in a 96 day period between diagnosis and interview for cases on average [37]. A concerted effort was made to contact eligible cases prior to commencement of chemotherapy. For the case-control study, the 33 hospitals that were known to treat or diagnose Long Island women with breast cancer were considered part of the rapid reporting network. This network was established in the first year of the ongoing study, in an area where a population-based case-control study of cancer had not been conducted in over a decade. During case identification, most hospitals were contacted on a daily basis, but a few hospitals that were expected to identify only a few cases per year were contacted weekly. Study personnel contacted the hospital pathologist, or his/her designee in the pathology department, to identify any patients who were newly diagnosed with in situ or invasive breast cancer. The patient's name, birthday, date of diagnosis, preliminary diagnosis, and the diagnosing physician's name were obtained.

The potentially eligible case's physician was then contacted to confirm eligibility for the study and to seek permission to contact the patient for possible participation in the case-control study. To speed this process, the study team established contact in the first year of the study, prior to field activities, with nearly 450 primary care physicians, surgeons, and oncologists who could possibly treat or diagnose Long Island breast cancer patients. The reason for the study and the need for timely approval was explained to the Long Island physicians, and they were asked to fax back a signed form indicating their willingness to participate in the study. Prior to commencing the field activities, all physicians who were contacted returned such a form or indicted their willingness to participate through a telephone conversation.

Using these case identification methods, 2,271 women were identified, of whom

2,030 were ultimately eligible. Of these, physician approval to contact cases was obtained for 1,837 (90.3%) cases. No physician systematically refused permission. Physician refusal was most common for older patients with co-morbid conditions. Prognosis is poor for these women, and physicians perceived them as too fragile for study participation.

5.2.3 Control Identification

Potentially eligible control women were identified by Waksberg's method of random digit dialing (RDD) [180] for those under 65 years of age, and by Health Care Finance Administration (HCFA) rosters for those 65 years of age and older. HCFA selection occurred twice during the 12-month identification period that coincided with the 12 months of case ascertainment. RDD selection began July 1, 1996, and continued in eight waves over the following twelve months. The response rate to the RDD telephone screener was 77.9% although this response rate is only applicable to the control respondents who are under age 65 years (approximately 57.9% of all control respondents).

5.2.4 Subject Recruitment

Eligible case women, with physician permission, and potentially eligible control women, were sent a "recruitment" package by overnight mail that included a letter inviting them to participate in the study along with a descriptive brochure. Within days of receiving the package, a skilled, trained recruiter telephoned the subject to answer questions, describe the study, invite the subject to participate, and, if possible, schedule an appointment for the interview. The main questionnaire was completed by 1,508 (82.1%) of eligible case women and 1,556 (62.7%) of eligible control women. The reasons for non-response to the interview among cases and controls included subject refusal (n = 218 (12.4%) and 573 (21.6%), respectively); too ill, cognitively impaired, or deceased (76 (4.1%) and 193 (7.8%)), and non-locatable, moved out of area, or other (26 (1.4%) and 195

(7.9%)). Study subjects ranged in age from 24 to 98 years and response to the interview varied by respondents' age with 88.9% of cases and 76.1% of controls under age 65 years participating versus 71.6% of cases and 43.3% and controls over the age of 65 participating.

5.2.5 Case-Control Study Interview

The interview included four components, which were administered in the following order: (1) signed informed consent; (2) the interview-administered main questionnaire; (3) collection of biologic samples (blood and urine) and administration of a specimen checklist; and (4) a self-administered food frequency questionnaire (FFQ). These components required between 2 and 3 hours to complete. Completion of all components of the interview was done in the respondent's home. All interviewers, who were certified phlebotomists in the state of New York, received a one-week, standardized, intensive training course in all aspects of interview administration. Among case and control respondents who completed the interviewer-administered questionnaire, 98.2% and 97.6%, respectively, self-completed the food frequency questionnaire, 73.0% and 73.3% donated a blood sample, and 93.0% and 83.3% donated a urine sample. Not surprisingly, women who donated blood tended to be younger than non-donating women, but no differences in donation proportions by case control status were evident [37]. Figure 5.2 diagrams the participation rates, including subject recruitment and case identification, for the LIBCSP.

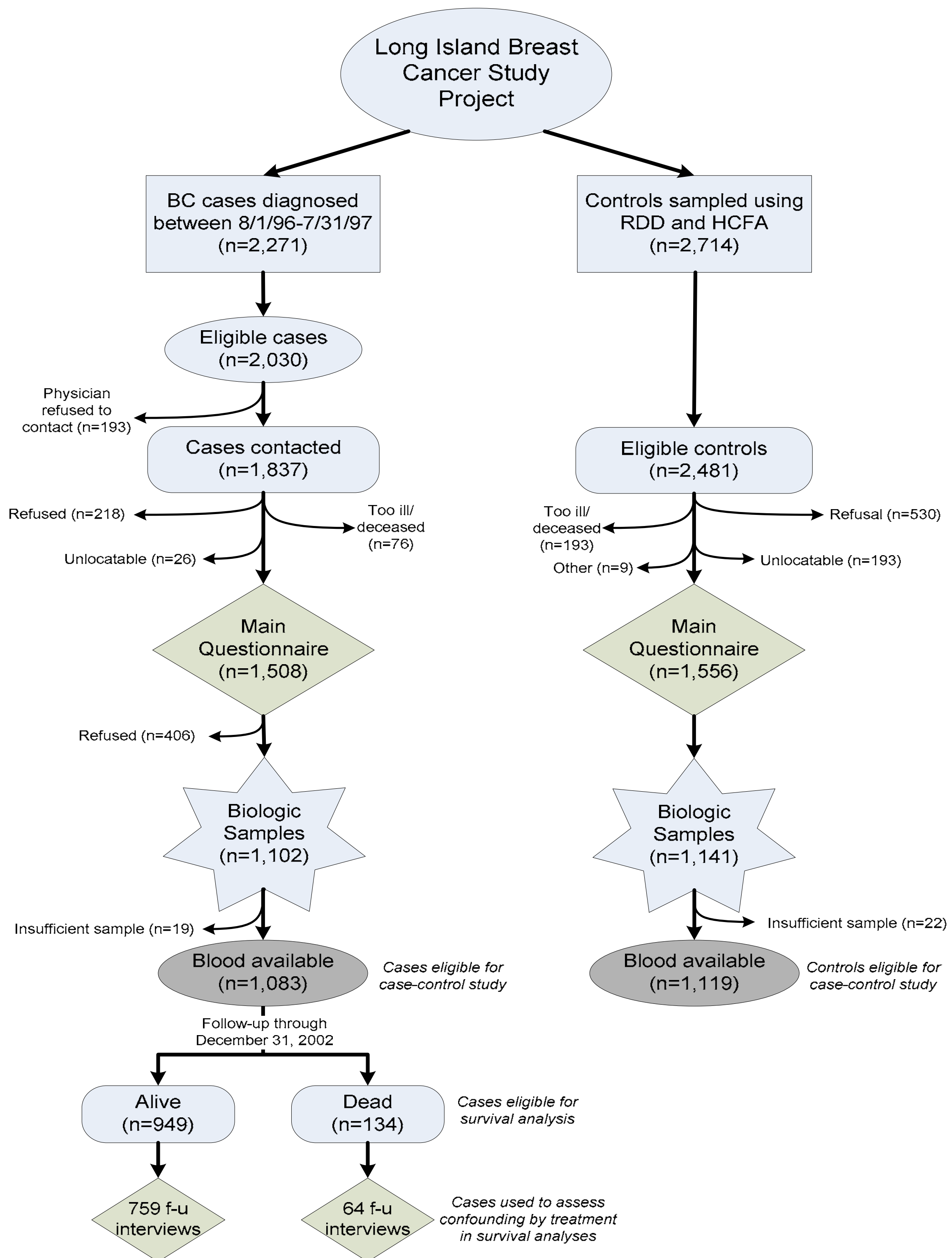


FIGURE 5.2. Participation in the Long Island Breast Cancer Study Project and Follow-up Study

5.2.5.1 Informed consent. Prior to commencing the interview, the interviewer explained the contents of the informed consent form to each eligible subject, and if the participant wished to continue, she was asked to sign the informed consent forms.

TABLE 5.1. Distribution of select demographic characteristics by cases and control status, Long Island Breast Cancer Study Project, 1996-1997 [37]

	Cases (n=1,508)		**Controls (n=1,556)**	
Characteristic	**No.**	**%**	**No.**	**%**
*Age at interview**				
< 35 years	39	2.6	45	2.9
35-44 years	181	12.0	245	15.7
45-54 years	397	26.3	423	27.2
55-64 years	372	24.7	403	25.9
65-74 years	365	24.2	310	19.9
75-84 years	134	8.9	112	7.2
85+ years	20	1.3	18	1.2
Missing	0		0	
Race				
White	1411	93.8	1429	91.8
Black	69	4.6	85	5.5
Other	25	1.7	42	2.7
Missing	3		7	
Education				
< High School	183	12.2	150	9.7
High school graduate	538	35.8	526	33.9
Some college	360	24.0	415	26.7
College graduate	191	12.7	236	15.2
Post college	230	15.3	225	14.5
Missing	6		4	
Income				
<$15,000	115	8.9	84	6.4
$15,000-19,999	70	5.4	83	6.3
$20,000-24,999	78	6.0	98	7.5
$25,000-34,999	176	13.6	139	10.6
$35,000-49,999	192	14.8	203	15.5
$50,000-69,999	214	16.5	245	18.7
$70,000-89,999	169	21.7	177	13.5
$90,000+	281	13.1	281	21.5
Missing	213		246	
Parity status				
Nulliparous	198	13.1	171	11.0
Parous	1310	86.9	1385	89.0
1 child	166	12.7	148	10.7
2 children	508	38.8	518	37.4
3 children	358	27.3	379	27.4
4+ children	278	21.2	340	24.5
Missing	0		0	
Menopausal Status				
Pre-menopausal	468	31.0	500	32.1
Post-menopausal	1010	67.0	993	63.8
Missing	30	2.0	63	4.1

*frequency matched variable

5.2.5.2 Main questionnaire. The comprehensive questionnaire took an average of 101 minutes to complete. The instrument included assessment of known and suspected risk factors for breast cancer in the following order: reproductive history; occupational history; residential history and history of exposure to environmental pollutants; history of living or working on a farm; medical history; family history of cancer; body size and physical activity; alcohol consumption; smoking history; menstruation and menopause history; contraceptive history; hormone medication use; pregnancy related problems; background information; and conclusion. Table 5.1 highlights select demographic and population characteristics.

To assess weight and weight gain, subjects were asked about their weight in pounds by decade from age 20 to age 70 and the year prior to the reference date. Height at age 20 and one year prior to the reference data was also obtained to calculate BMI.

5.2.5.3 Biologic Samples. After completing an additional informed consent form, participants were asked to donate blood and urine samples. From each participant, about 40 mL of blood were donated (5 EDTA-treated lavender-top tubes) as well as urine. At the time of sample donation, a specimen checklist was also administered by the interviewer. The checklist queried subjects about drugs, foods, and behavior they may have engaged in the few days prior to the sample donation, which may influence interpretation of the blood and urine assay findings (e.g., hormone use, intake of cruciferous vegetables, etc.). After collection of the biologic samples at the participants' homes, the samples were shipped overnight to Dr. Regina Santella's laboratory at Columbia University for processing. The blood samples were shipped at ambient temperatures, and the urine was shipped on ice. Aliquots of plasma from the 40 mL of blood and aliquots of urine were stored at -80° C.

Of the 1,102 cases who donated blood, 19 samples (1.7%) were later found to be insufficient for a total of 1,083 cases with available blood for genotyping. Similarly, of the 1,141 control blood donors, 22 samples (1.9%) were found to be insufficient for a total of

1,119 controls. We also anticipated that genotypes for an additional 2% of the samples would be unable to be determined based on previous genotype analyses using the LIBCSP data [181, 182]. Therefore, we anticipated a final sample size of 1,061 cases and 1,097 controls for the case-control study. The 1,061 cases would also be included in the survival analysis. In fact, a slightly higher number of cases and controls were successfully genotyped (n=1,073 and 1,112, respectively).

5.2.5.4 Food frequency questionnaire (FFQ). Participants were asked to self-complete a food frequency questionnaire after collection of the biologic samples, or if no samples were donated, after completion of the main questionnaire. The instrument used to assess diet in the past year was a modification of the NCI-Block questionnaire that was adapted by Dr. Nancy Potischman [183] to assess intake of fat and micronutrients more completely. The instrument was further modified by the study team to allow adequate assessment of isoflavonoids [184]. The interviewer remained at the respondent's home until the FFQ was completed.

5.2.6 Case Subjects' Medical Record Retrieval and Abstracting

As part of the case-control study, signed medical record release forms were obtained from study participants. For cases, the relevant medical records and pathology records for the initial primary breast cancer were obtained from the appropriate hospital. Data abstracted from the case medical records included stage, nodal involvement, hormone receptor status (estrogen and progesterone), and treatment.

5.2.7 Interview Data Management and Statistical Analyses

All subjects were tracked by computer. All data collected by questionnaire has been entered onto computer, and thoroughly checked for consistency and accuracy. Selected

relevant results of LIBCSP case-control distributions are presented in Table 5.1.

5.2.8 Summary

Detailed statistical analyses of the LIBCSP data have linked post-menopausal breast cancer to obesity and weight gain, particularly in women over the age of 50 [9]. Because of its large sample size, including 1,083 cases and 1,119 controls that have donated a blood sample and completed an interview, the LIBCSP represented an efficient and appropriate population to address this study question.

5.3 Long Island Breast Cancer Study Project Follow-up

5.3.1 Subject Eligibility for Follow-up

Women with newly diagnosed in situ or invasive breast cancer between August 1, 1996 and July 31, 1997, who were 20 years of age or older at diagnosis and a resident of Nassau or Suffolk counties in Long Island, NY, and who completed the main questionnaire for LIBCSP were eligible to participate in the proposed follow-up study.

5.3.2 Subject Recruitment for Follow-up

Ninety-four women who refused re-contact at the end of the case-control study were not recruited for a total of 1,414 potential cases in the follow-up study. Eligible case participants were re-contacted initially by mail and then by telephone to invite them to participate in the Follow-up Study. In the event that a case could not be located using this approach, the next of kin were contacted based on information provided at the end of the case-control study interview. Lastly, patient physicians were contacted if a subject could not be traced. Interviews were conducted using computer assisted telephone interviewing (CATI).

Of the 1508 women with in situ or invasive breast cancer identified through LIBCSP, 1,098 were successfully contacted and re-interviewed in 2002-2004 (72.8%). For 1,033 women, complete treatment and demographic follow-up interview data were available while 65 cases only completed a short questionnaire detailing their treatment therapies. For the 410 women who were not successfully interviewed, failure to obtain follow-up information was due to refusal (n=237), passive refusal (n=22), inability to locate (n=55) and death with no identifiable proxy (n=96).

For the 1,098 patients who were successfully interviewed, 177 (16.1%) had in situ breast cancer while 921 (83.9%) were invasive cases based on information provided in the initial case-control study. Of these 1,098 patients who were successfully interviewed, 823 have given blood as part of the original case-control study. This subset of the LIBCSP cases were used to assess for modification by treatment since complete treatment information and blood were available for these women (discussed in section 5.5 and highlighted in Figure 5.2).

5.3.3 Follow-up Study Interview

The interview included two components: (1) signed informed consent and (2) the follow-up questionnaire administered by telephone. Completion of all components of the interview was done using CATI by trained interviewers. For the survival analyses, information from the follow-up questionnaire was only used to obtain complete treatment information. This treatment information was used to evaluate confounding by treatment in the subset of 823 cases that have both donated blood and completed a follow-up interview.

Three different interviews – subject, proxy and critical – were used to maximize the number of participants. Of the 1,098 women with follow-up data, 784 completed the subject interview (71.4%). Case subjects were asked to recall all treatments and procedures undergone since the diagnosis date of the initial breast cancers, such as needle biopsies,

tumor biopsies, modified mastectomy, radiation, chemotherapy and hormone treatments. For each type of procedure, subjects were asked to recall the number of times the procedure was performed, the frequency, data and location of institution. Breast cancer recurrences and any second primary cancer diagnoses were also queried.

The proxy interview was a full length questionnaire, identical to the subject interview except for minor wording changes to reflect the questions referred to the index subject rather than the proxy. A brief section asking about the proxy's relationship to the subject was also added. Of the 1,098 cases, 84 proxy interviews were conducted (7.7%), which is consistent with the expected mortality rate of breast cancer according the Surveillance, Epidemiology, and End Results Program (SEER) [88].

The critical interview was designed to be less detailed than the subject interview for subjects or proxies that did not have the time or were too ill to complete the full length questionnaire. Of the 1,098 subjects with follow-up information, 165 respondents completed the critical interview (15.0%). Participants were still questioned on various treatments and outcomes as well as changes in their behaviors following their diagnosis, but only the most necessary (or "critical") questions were included in the questionnaire.

Lastly, 65 respondents (56 subject women, 9 proxies) were unwilling to complete a questionnaire, but did volunteer the most basic treatment information (5.9%). For this, a series of three questions were developed, oncologist, treatment institution and hormonal medicine use (yes/no), designed to take less than five minutes to complete but maximize the treatment information provided. Therefore, for these subjects, no demographic follow-up information will be available but treatment data could be included in all analyses.

5.3.4 Case Subjects' Medical Records and Abstracting

As part of the study, subjects were asked to authorize the release of their medical records for abstracting by study personnel. Of the 1,414 potential cases, 600

women signed the release form (42.3%). One record of the 600 authorized could not be located. These medical records were used to evaluate the reliability of subject reported outcomes and treatments.

Of the 600 women who signed a medical record release form as part of the follow-up study, 587 completed a follow-up interview. There was a high degree of concordance between treatments reported by subject and those reported in the medical record, with less than 1% discordance for radiation (Kappa (κ) = 0.99) and chemotherapy (κ=0.98), and 3.2% discordance for hormone therapy (κ=0.93). Women who completed a proxy interview showed less concordance with 10.0% discordance for radiation therapy (κ=0.78) and 11.1% discordance for chemotherapy (κ =0.78) although the hormone therapy data was 100% concordant (κ=1.00). Concordance between subject reported recurrences and/or second primary cancer diagnoses and the medical record were much lower than those found for treatment. For example, concordance of the medical record abstraction and interview for whether a woman had a recurrence was 8.6% discordant but concordance for whether a woman had a second primary breast cancer was 33.3% and an additional 25% referred to the second primary in the medical records as a recurrence in the interview. Thus, the data on recurrence was not used as part of this investigation.

5.3.5 Outcome Ascertainment

Of the 1508 women, 198 were identified as deceased by the National Death Index (NDI) while the rest of the cohort (n=1,310) were presumed alive through December 31, 2002, the last date that the NDI considered the death data to be accurate. Thus, for the survival analyses, 198 women were deceased and 1,310 alive through the end of the follow-up period. For the 1,083 cases who donated blood and are, therefore, eligible for the survival analysis, 949 were alive (87.6%) and 134 dead (12.4%) at the end of the follow-up period. Among the subset of 823 cases that completed a follow-up interview and donated

blood, 759 were alive (92.2%) and 64 deceased (7.8%) at the end of the follow-up study (see Figure 5.2).

Death certificate data, provided by the NDI, were then used to determine underlying cause of death, primary cause of death, and breast cancer related deaths as well as date of death. A primary cause represents the most immediate cause of death while an underlying cause is the condition that gave rise to the primary cause. Table 5.2 highlights the NDI codes used to create the underlying cause of death and primary cause of death variables while Table 5.3 and Table 5.4 depicts the frequencies of these causes of death for the 1,083 women eligible to be included in the survival analysis and the 823 women who donated blood and completed a follow-up interview respectively.

TABLE 5.2. NDI death codes used to create the broad categorical variables for underlying, primary and breast cancer-related death

Variable and Coding	Broad Categories of Death	NDI codes
NDI_under		
0	Alive	
1	Breast cancer	1749, C509
2	Lung cancer	1629, C349
3	Other cancer	1919, 1991, 2050, B49, C169, C259, C482, C541, C56, C80, C920, C97
4	CVD	I269, I38, I499, I509, I619, I639, I711, I509, 3949, 4029, 410, 4140, 4275, 436, I10, I119, I219, I250, I251
5	Other	B146, A219, 1619, 5609, 856, G309, G419, J189, J449, J841, K440, K572, K801, N19, N390, R99
NDI_primary		
0	Alive	
1	Breast cancer	1749, C509
2	Lung cancer	1629, C349
3	Other cancer	C56, C80, C920, 1509, 1619, 1919, 1970, 1988, 2050, B49, C169, C189, C259, C482, C541
4	CVD	3949, 4029, 410, 4140, 4275, I10, I119, I219, I251, I254, I469
5	Other	0389, A402, A415, A419, D649, E039, E149, G20, G309, G934, J449

TABLE 5.3. Frequencies of causes of death for the 1,083 women who donated blood and completed the baseline case-control interview

NDI Variable	Alive	Breast Cancer	Lung Cancer	Other Cancer	CVD	Other
NDI_under	949	74 (55.2%)	6 (4.5%)	13 (9.7%)	27 (20.1%)	14 (10.4%)
NDI_primary	949	85 (63.4%)	6 (4.5%)	12 (9.0%)	17 (12.7%)	14 (10.4%)

TABLE 5.4. Frequencies of causes of death for the 823 women who donated blood and completed a follow-up interview

NDI Variable	Alive	Breast Cancer	Lung Cancer	Other Cancer	CVD	Other
NDI_under	759	41 (64.1%)	2 (3.1%)	6 (9.4%)	10 (15.6%)	5 (7.8%)
NDI_primary	759	44 (68.8%)	2 (3.1%)	6 (9.4%)	8 (12.5%)	4 (6.3%)

5.3.6 Follow-up Interview Data Management and Statistical Analyses

All subjects were tracked by computer. Because data were collected using CATI, programming edits and consistency checks were built into the questionnaire as well as automatic question skips for complex questions. Therefore, data management efforts for the follow-up study were less than those required for the case-control study. Table 5.5 highlights preliminary distributions of reported treatment therapies by vital status for the 823 individuals who donated blood and completed a follow-up interview.

TABLE 5.5. Distribution of treatment by vital status for the 823 women who donated blood and completed a follow-up interview

	Alive (n=759)		Dead (n=64)		Total (n=823)	
Treatment	**No.**	**%**	**No.**	**%**	**No.**	**%**
Radiation therapy						
Yes	431	60.4	38	63.3	469	60.6
No	283	39.6	22	36.7	305	39.4
Missing	45		4		49	
Hormone therapy						
Yes	433	61.0	34	61.8	467	61.0
No	277	39.0	21	38.2	298	39.0
Missing	49		9		58	
Chemotherapy						
Yes	262	36.8	37	61.7	299	38.7
No	450	63.2	23	38.3	473	61.3
Missing	47		4		51	

5.3.7 Summary

Of the 1,508 potential cases, 1,098 subjects or their proxies completed a follow-up interview (72.8%). Comparisons between the medical record and self-report indicate that treatments can be accurately recalled; therefore, we used this self-reported treatment information with little concern of misclassification. Among the 1,083 cases that donated blood, vital status, assessed by the NDI through December 31, 2002, indicated that 134 women were dead and 949 were alive for the survival component of the study. Complete treatment data are available on 823 of these cases. Of these 823 cases, the NDI has indicated that 759 were alive and 64 dead at the end of the follow-up period.

5.4 GENOTYPING

5.4.1 SNP selection and Haplotypes

Interindividual variation in the human genome is the product of DNA changes that arose on ancestral segments of chromosomes [185]. Genetic recombination disrupted tightly linked regions (i.e., regions in linkage disequilibrium) of the genome that contain polymorphic loci [186]. The contribution of individual SNPs to disease susceptibility may be difficult to determine because of SNPs in linkage disequilibrium (LD) and cis-acting effects between SNPs on the same allele [186, 187]. Therefore, a broader multilocus haplotype approach, where multiple markers on the same chromosome are used, has been advocated to provide greater power [188, 189] and is considered more powerful to detect susceptibility alleles than single locus SNPs, especially when the causal SNPs are unknown or when the interaction of multiple mutations on the same chromosome has a particularly large effect on the disease phenotype. Methods for selecting haplotype tagging SNPs within regions of high linkage disequilibrium are in their infancy. Initial approaches focused on selecting SNPs based on known or hypothesized functions, usually within coding exons and regulatory

regions. However, potentially important areas of genetic variation exist throughout the gene and should be considered comprehensively.

To study associations between inter-individual variation of PPAR and breast cancer, we examined polymorphisms of *PPARA* using a multi-locus haplotype approach (Table 5.6). This gene was chosen because it plays a critical role in the inflammation process and insulin sensitivity. For *PPARA*, sequencing data was available for 23 European-Americans and 24 African-American Coriell samples on the University of Washington-Fred Hutchinson Cancer Research Center Variation Discovery Resource website, http://pga.gs.washington.edu/ (or Program for Genomic Applications (PGA)). Given the racial homogeneity of the LIBCSP study population (92.8% Caucasian and 5.0% African American), we identified haplotype tagging SNPs using the 23 European-Americans only. The haplotype tagging SNPs were produced using the PGA LDSelect Program [190] run only for European Americans. This program has been shown to select a maximally informative set of common SNPs that distinguishes 80% of common haplotypes and is based on the r^2 LD statistic [191].

The r^2 statistic represents a measure of correlation between two SNPs. For example, an r^2=1 implies “perfect LD”, which happens only when the minor allele frequency (MAF) for two SNPs are the same. In contrast, when D´, another measure of LD, is equal to one, “complete LD” is implied and the MAF for the two SNPs do not have to be the same. This characteristic of the D´ can create arbitrary block boundaries, making it a less desirable measure of LD than r^2 [190]. Because of power concerns for rare alleles, we focused on alleles with a MAF of 10% or more, which capture common variation while minimizing expense.

The LDSelect algorithm begins with all SNPs above the MAF threshold (in this case 10%) and calculates the pairwise r^2 between one SNP and all other SNPs. SNPs in which the r^2 is greater than 0.80 are then put together into a bin or the minimum set of informative SNPs. This iterative process continues until no unbinned SNPs remain. SNPs that do not

meet the correlation threshold value with any other SNPs are placed into their own unique, singleton bin. The algorithm then determines the tag SNP for each bin, or the SNP that can act as a marker for all other SNPs in that bin. Collectively, this group of tag SNPs represents the maximally informative set of SNPs for the gene. When multiple possible tag SNPs were identified for a bin, non-synonymous SNPs were given priority. SNPs in repetitive regions were avoided due to the difficulty in genotyping these polymorphisms successfully. Based on this program, fourteen *PPARA* SNPs were identified for sequencing and are presented in Table 6.6.

In addition to LDSelect, other software packages, including Tagger [192], were considered for tag SNP identification. Tagger uses aggressive tagging to identify tag SNPs and was developed by Paul de Bakker (http://www.broad.mit.edu/mpg/tagger/). For this software, the CEU population (30 Utah trios with ancestry from northern and western Europe) from the International HapMap Project was used as the reference panel [193]. Like ldSelect, Tagger identified fourteen tag SNPs although there was very little overlap in the SNPs selected with the two programs. These differences could be due to random variation since the programs are expected to perform similarly when the same reference population is used. LDSelect was ultimately chosen as the preferable method for tag SNP selection because it uses resequencing data from the PGA website rather than the density SNPs from HapMap. In general, resequencing data permits a more comprehensive assessment of the genetic variation compared to density SNPs and is, therefore, considered the gold standard.

5.4.2 Genotyping

For this study, genotyping was conducted by Dr. Santella's laboratory at Columbia University, New York, NY. All LIBCSP DNA samples are available on 96 well master plates. Plates have a 10% duplication rate with laboratory personnel blinded to case control status and duplication. Genotyping was carried out using iPLEX technology (Sequenom,

SanDiego, CA) on a MassARRAY Compact Analyzer. This multiplex method uses the mass of the incorporated nucleotide for identification of genotype. For SNPs that could not be mulitplexed (rs4253623 and rs4253699), Taqman (Applied Biosystems, Foster City, CA) assays were used. For Taqman, samples were run on an ABI 7500 Real Time PCR system. For Taqman, the rs#s for the SNPs of interest were given to Applied Biosystems for the preparation of the specific kits.

TABLE 5.6. Haplotype tagging SNPs of *PPARA* to be evaluated identified using ldSelect and the PGA European-American population as the reference panel

rs#	**SNP Location**	**Base Pair Change (major>rare)**	**Minor Allele Frequency**		**Reference**
			African-Americans	**European-Americans**	
rs4253730	Intron 3	A>G	0.479	0.182	LDSelect
rs4253760	Intron 6	T>G	0.667	0.196	LDSelect
rs4253705	Intron 2	T>C	0.375	0.190	LDSelect
rs135543	Intron 2	C>T	0.391	0.283	LDSelect
rs135542	Intron 2	T>C	0.348	0.205	LDSelect
rs4253649	Not Validated	C>G	0.783	0.370	LDSelect
rs4253758	Intron 6	T>C	0.810	0.217	LDSelect
rs4253699	Intron 2	T>C	0.312	0.182	LDSelect
rs4253655	Intron 2	G>A	0.022	0.143	LDSelect
rs4253681	Intron 2	T>C	0.167	0.136	LDSelect
rs4253755	Intron 5	G>A	0.042	0.130	LDSelect
rs4253706	Intron 2	G>A	0.043	0.119	LDSelect
rs4253623	Intron 2	A>G	0.021	0.109	LDSelect
rs1800206	Exon 5	C>G	0.021	0.032	SNP500
	Leu>Val			0.022	PGA

5.5 COVARIATES

5.5.1 Overview

This section highlights the selection and coding of covariates, including weight gain and BMI as effect modifiers and all potential confounders, in the case-control and survival analyses. Outcomes of interest for both analyses are also presented. Confounders were chosen *a priori* using directed acyclic graphs (DAG) and modeling strategies for these variables are discussed in section 5.7.2 as part of the data analysis. The previous section

(Section 5.4) detailed the SNP selection process and genotyping for *PPARA*, the main exposure of interest, and the construction of the *PPARA* haplotypes from these SNPs are discussed in section 5.7.2.

For the case-control study, exposure information was derived from the baseline questionnaire (http://epi.grants.cancer.gov/LIBCSP/projects/Questionnaire.html) [37], which was interviewer-administered to participants in the their home within a few months of identification. Respondents were asked about their pregnancy, occupational, residential and environmental histories; family history of breast cancer; medical history including use of aspirin and other NSAID; self-reported weight and height by decade of life; recreational physical activities since menarche; active and passive cigarette smoking; alcohol use; menstrual history; use of exogenous hormones; and demographic characteristics.

For the survival analysis, information from the baseline case-control study interview was used for most confounders and effect modifiers, including obesity and weight gain. Vital status through the end of 2002 was determined through the National Death Index (NDI). Although treatment was not believed to be a confounder, it was investigated as a possible effect modifier. Treatment information was available from four different sources: (1) the baseline interview (n=1,508 cases), (2) the baseline medical record (n=1,508 cases), (3) the follow-up questionnaire (n=1,098 cases), and (4) the abstracted medical records at follow-up (n=599 cases). Treatment information obtained at the initial case-control study interview and case-control medical abstraction were incomplete since the average time of interview occurred within three months of diagnosis [37]. Therefore, it was important to do sub-group analyses using the more complete treatment information. Treatment data from the follow-up interview was used to determine the influence of treatment because of the larger number of cases who completed the follow-up interview compared to those with follow-up abstracted medical records.

Appendix D Table D.1. presents the table of variables that were included in both the case-control and survival analyses, highlighting their definition and coding. Discussions in the text are below.

5.5.2 Outcome of interest

For the case-control study, the outcome of interest was defined by case/control status while two outcomes were investigated in the survival analysis. As mentioned earlier, vital status was ascertained using the NDI and is accurate through December 31, 2002. Based on the information available through the NDI and the death certificates, both vital status based on all cause mortality and breast cancer-related death, where ICD codes 174.9 and C-50.9 are cited as either primary or secondary causes of death, were evaluated as potential outcomes.

5.5.3 Obesity as an effect modifier

Obesity could act as an effect modifier for *PPARA* based on the gene's involvement in the insulin and inflammatory pathways. Increased levels of estrogen, insulin and cytokines from obesity could act synergistically with the "at risk" *PPARA* genotypes to increase breast cancer risk by a greater quantity than their separate effects would suggest.

In the case-control study questionnaire, participants were asked about their weight by decade of life, beginning at age 20 as well as height at age 20 and at the date of interview. Using these variables, BMI by decade and weight change throughout life were ascertained. In particular, based on the results of Eng *et al.* using LIBCSP data, weight change from age 20 to one year prior to the interview date and post-menopausal weight gain were associated with an increased risk of post-menopausal breast cancer even after adjusting for BMI at age 20 and 50 respectively [9].

For the survival analysis, both pre- and post-menopausal cases were included. As highlighted in Section 3.6 and Appendix B, obesity has an impact on survival in both pre- and post-menopausal women. This is consistent with LIBCSP, which found decreased survival with obesity in both pre- and post-menopausal women [112]. Therefore, in this study, we combined pre- and post-menopausal cases for the survival analyses. Stratification by menopausal status was done to confirm that menopausal status did not modify the association. Weight change from age 20 to reference and BMI at reference were considered as potential effect modifiers. Table 6.7 presents the distribution of BMI at the reference date by case-control status.

TABLE 5.7. Distribution of Body Mass Index among Cases and Controls, Long Island Breast Cancer Study Project, 1996-1997 [37]

	Cases (n=1,508)		**Controls (n=1,556)**	
	No.	**%**	**No.**	**%**
BMI at reference				
< 25.00	700	46.4	778	50.0
25.00-29.99	476	31.6	455	29.2
≥ 30.00	332	22.0	323	20.8
Missing	0		0	

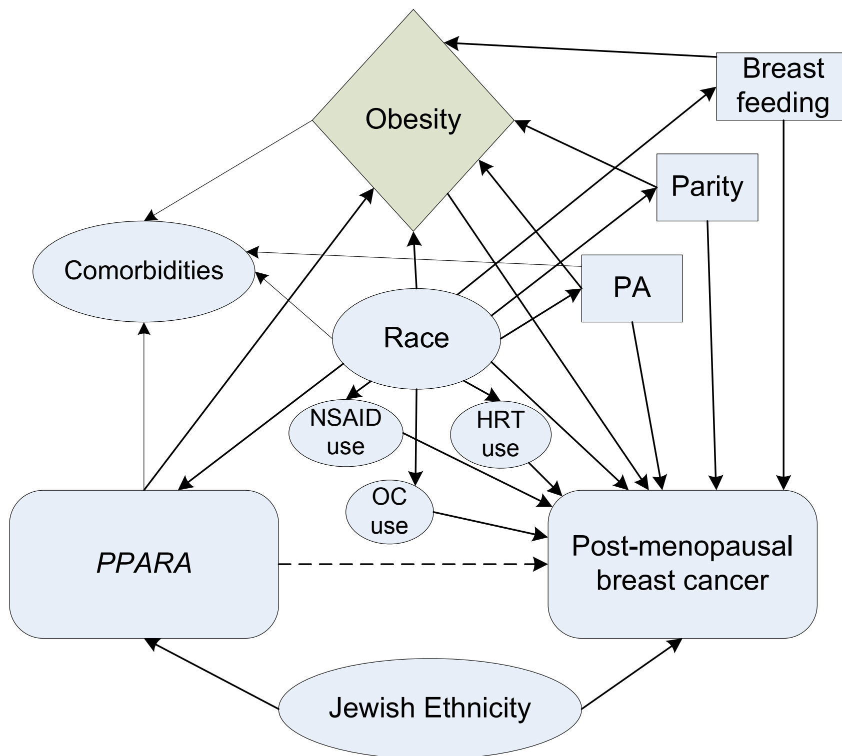
5.5.4 Confounders for the case-control study

For the association between post-menopausal breast cancer and *PPARA*, minimal confounding was expected. As the DAG indicates (Figure 5.3), race and religion appear to be the only factors that might confound this relationship. Given the small percentage of African-Americans in the LIBCSP population (5%), it was unlikely that race would play an important role in these analyses, however. To address the potential impact of Judaism, which is essentially acting as a surrogate for increased prevalence of the *BRCA1* mutation (see section 3.4.6), adjustment for religion was important, particularly since 17.2% of case and 15.4% of controls are Jewish in this population [37]. Initial analyses in the LIBCSP

population did not find an association between the Jewish ethnicity and breast cancer incidence, suggesting that it would not be a confounder (OR=1.16; 95% CI: 0.95-1.42) [37]. Consistent with the DAG and previous analyses, neither race nor Jewish ethnicity was found to be a confounder in the case-control study analyses (see Appendix D, Tables D.8-D.21 model building procedures for each polymorphism individually).

It was also possible that family history of breast cancer could be important to include in the analysis if it was associated with *PPARA*. Since no studies have investigated breast cancer and *PPARA*, the relationship between family history and *PPARA* was unknown. This study found that family history of breast cancer was not a confounder and, therefore, it was not included in analyses (see Appendix D, Tables D.8-D.21).

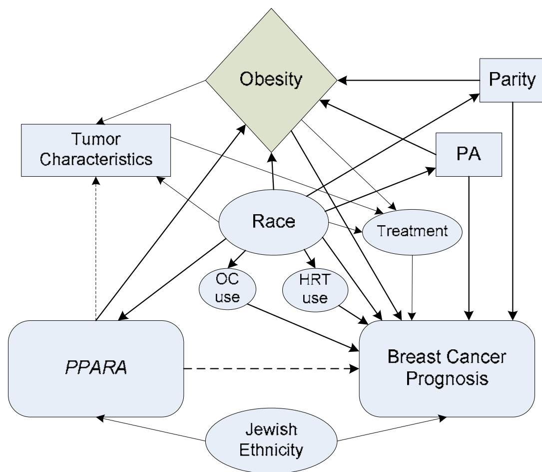
FIGURE 5.3. Directed acyclic graph for the association between post-menopausal breast cancer risk and *PPARA*.



5.5.5 Confounders for the survival analysis

For the association between breast cancer prognosis and *PPARA*, minimal confounding was expected. As the DAG indicates (Figure 5.4), race and Jewish ethnicity appear to be the only factors that might confound this relationship. Once again, because of the small number of African-Americans in the LIBCSP, race was unlikely to play a role in this population. This research did not find any evidence of confounding by Jewish ethnicity or race (see Appendix D, Tables D.23-D.28 for model building procedures for each SNP individually). Treatment could modify the association if *PPARA* affects tumor characteristics. Consequently, treatment and several tumor characteristics were evaluated as potential effect modifiers. Stage, ER and PR status were available for all genotyped cases while treatment data was considered using the 823 cases that completed the follow-up interview and donated blood. Stratification by select tumor characteristics and treatment information did not yield any evidence of effect modification (see Table D.29, Appendix D). It is important to note that stratification by stage and treatment data led to small cell sizes and very imprecise effect measures; therefore, it was difficult to assess modification by these variables. A comparison of survival in all cases versus invasive breast cancer cases only yielded similar results (see Table D.39, Appendix D).

FIGURE 5.4. Directed acyclic graph for the association between breast cancer prognosis and *PPARA*.



5.5.6 Summary

No variables confounded the association between breast cancer risk (see Appendix D, Tables D.8-D.21) and survival (Appendix D, D.23-D.28) and *PPARA*. For the case-control study, race, Jewish ethnicity and family history of breast cancer were included as potential confounders while race and Jewish ethnicity were considered confounders in the survival analysis. As discussed in Section 3.4, breast cancer treatment will vary by tumor characteristics with certain tumors more likely to respond to treatment than others. *PPARA* could influence tumor type. Therefore, treatment and select tumor characteristics were examined as potential effect modifiers (see Appendix D, Table D.29). Sample size limitations prevented these stratified analyses from being conclusive, however. These covariates along with their coding and variable descriptions are included in Appendix D.

5.6 Data Management for Case-Control and Follow-up Studies

Baseline data from LIBCSP have been published [37]. Standard data cleaning procedures, which include range checks and distributions on all variables to detect outliers or implausible values, have already been performed on LIBCSP. To confirm data cleaning, univariate analyses for all potential confounders, PPAR genotype frequencies and measures

of obesity were conducted to ensure that the range of values were consistent with the questionnaire and were biologically feasible.

Standardized procedure manuals and extensive interviewer training including probing procedures and questionnaire administration exist for all studies. Epidemiologic field staff were trained to implement protocols for subject identification and recruitment, questionnaire administration, biological specimen collection and handling in a standardized fashion to avoid introduction of bias. All data were double keyed and discrepancies reconciled.

For this study, genotype data were entered into a database and then merged with questionnaire and survival data by participant ID number. Individuals with missing genotype and covariate data were excluded from analyses. To address the impact of missing data on this study, a comparison of select tumor and demographic characteristics, including age, stage and tumor size, was performed. This analysis revealed no differences between genotyped cases (n=1,073) and all cases (n=1,508) by vital status (see Table D.30 in Appendix D). Additionally, in analyses in which risk factors for breast cancer among all women in the study were compared to those among women who donated blood, similar results were found.

Haplotypes were determined for individuals with genotyping data available for more than 50% of the included SNPs. Data were imported into a SAS data file for analysis. Quality control procedures for laboratory assays included a blinded 10% repeat sample embedded throughout the case-control plates. In addition, positive and negative controls were included in each assay. Laboratory personnel did not know case-control information. The genotyping core was unaware of the exposure status of the study participants.

5.7 Data Analysis

5.7.1 Overview

Important contributions of this project are to determine the prevalence of *PPARA* polymorphisms and to evaluate *PPARA* SNPs and haplotypes to breast cancer risk and survival among a population-based sample of women in the United States. There were four major goals of the study analyses: **(AIM 1)** to estimate the association between breast cancer risk and *PPARA* polymorphisms in post-menopausal women, **(AIM 2)** to explore possible interactions between PPAR polymorphisms and obesity (measured as BMI$\geq$ 30 and weight gain), and their influence on post-menopausal breast cancer risk, **(AIM 3)** to estimate the association between PPAR variants (SNPs and haplotypes) and breast cancer survival in pre- and post-menopausal women, and **(AIM 4)** to explore interactions between PPAR polymorphisms and obesity and their influence on breast cancer survival in pre- and post-menopausal women.

5.7.2 Haplotype Analysis

Using the selected *PPARA* SNPs (discussed in Section 5.4) as building blocks, haplotypes were constructed and analyzed in both case-control and survival analyses. These tag SNP haplotypes captured most common genetic variation over the entire gene to evaluate cis-cis interactions (discussed in Chapter 8). The statistical methods involved in this analysis of genetic data can be divided into five main components. These components contain elements of Hardy-Weinberg Equilibrium (HWE) calculation, haplotype construction, and statistical modeling.

- Calculation of Genotype and Allele Frequencies

As mentioned earlier, no studies have investigated the association between *PPARA* and breast cancer. To the best of our knowledge, this is the first large population-based study to determine the allele and genotype frequencies for each SNP and is an important contribution to the *PPARA* literature. These frequencies were calculated using SAS version 9.1.

- Testing Hardy-Weinberg Equilibrium

HWE was tested among the controls to ensure assumptions of parametric statistical tests were met [194] using SAS / Genetics. After the allele frequencies were calculated, the HWE equation, $p^2+2pq+q^2 = 1$, where p equals the frequency of the major allele and q equals the frequency of the minor allele, were applied to determine the expected genotypic proportions if the conditions of HWE were met. The observed and expected genotypic proportions were then compared using a chi-square test with one degree of freedom. In this case, the null hypothesis was that the population is in HWE; therefore, a $\chi^2 \geq 3.85$ led to the rejection of the null hypothesis that the population is in HWE. Due to stringent quality control measures in the genotyping procedures, violations in HWE were not anticipated. In these analyses, no SNPs were out of HWE in the control group. SNPs not in HWE would have been removed from the analysis to avoid possible errors in the estimation process. Additionally, HWE was also calculated among the cases to confirm that no major deviations were present in this group as well. While two polymorphisms were out of HWE in the cases, these deviations could be explained by evolutionary change and most likely do not reflect genotype error [195].

- Characterization of Linkage Disequilibrium between SNPs

Although the tag SNPs and functional SNP (rs1800206) were initially identified to determine a set of the minimally sufficient number of SNPs necessary to characterize the

gene in the Caucasian reference population, it was still important to describe the linkage disequilibrium between the tag SNPs in our study population. Therefore, we computed pairwise LD [196] for each loci combination [197] using Haploview. This allowed for comparison with the SeattleSNP population that was used as the reference panel of tag SNP selection. Because the tag SNPs were selected to have an $r^2 < 0.80$, we expected our polymorphisms to show little LD. We found LD patterns consistent with this expectation with little to no LD evident in pair-wise comparisons.

- Haplotype Reconstruction

Unknown phase is a problem in a sample of unrelated individuals, where no genetic information is available from related family members. Several statistical methods such as Clark's algorithm [198] and expectation-maximization (EM) algorithm [199] have emerged as effective and cost-efficient for inferring and reconstructing phase of alleles. The EM algorithm is inefficient with a large number of polymorphisms due to its iterative process in haplotype construction, but is more than adequate for handling this study's six SNPs. Additionally, EM is able to use partially missing genotypes, where some but not all the alleles are missing, by iteratively updating the frequency of a partially missing genotype and then updating all existing haplotypes that have alleles identical to the non-missing alleles of the partially missing haplotypes. The EM algorithm is the default for SAS / Genetics and was used in the determination of this study's haplotypes. Based on this program, 12 haplotypes in all women and 14 in post-menopausal women only were identified with a frequency greater than 5% (see Table D.22 in Appendix D).

- Modeling associations of interest with haplotypes

Currently, there is no consensus on the best method to analyze haplotype data from association studies. As part of SAS / Genetics, SAS has added the HAPLOTYPE procedure,

which first constructs the haplotypes and then allows for easy transformation to a data set that can be used to perform regression analyses. This unique feature of SAS / Genetics eliminates the need for other programs, including Haplo Stat, which requires transforming a SAS data set into one compatible with R programming [200, 201]. Because of its facility, it has been used widely since its conception [202, 203]. For haplotype-disease associations, each haplotype will be examined for associations with breast cancer relative to all other haplotypes.

5.7.2.1 AIM 1. The polymorphisms were examined individually and as haplotypes using unconditional logistic regression [204] to estimate the odds ratios (OR) of post-menopausal breast cancer risk and corresponding 95% confidence intervals. Confounding was investigated using a backwards elimination strategy, where the full model was initially specified and covariates were removed from the model using a 10% change in estimate approach (*i.e.* covariates that resulted in a 10% or greater change in a genotype effect estimate were considered confounders). Covariates included in the full model were family history of breast cancer, Jewish ethnicity, and race based on the DAG in Section 6.5.4. This study did not find evidence of confounding by these three variables in each SNP evaluated individually (see Appendix D, Tables D.8-D.21).

5.7.2.2 AIM 2. To evaluate the role of genetic polymorphisms as modifiers of the associations between obesity and breast cancer risk, several statistical methods were employed, including likelihood ratio tests [204], the Breslow-Day test [205], and the interaction contrast ratio [206]. Interaction terms were included in the full model and the likelihood ratio test comparing a model with and without the interaction terms was used to assess model fit [204]. The Breslow-Day (BD) test was also be used to evaluate OR effect measure modification for modeling the individual SNPs [205]. Additive interaction

was assessed using the interaction contrast ratio (ICR). ICRs were calculated using the following formula, ICR = OR_{11} - OR_{10} - OR_{01} + 1. ICRs greater than zero imply superadditivity while ICRs of zero suggest no interaction on the additive scale. ICRs less than zero indicate less than additive effects (but interaction is still present on the additive scale) [206]. The results of the BD tests by individual SNP as well as the effect modification tables for covariates with BD p-value of less than 0.15 are found in Appendix D, Tables D.2-D.7.

5.7.2.3 AIMS 3 & 4. Kaplan-Meier survival curves were first created for *PPARA* haplotypes to portray the cumulative incidence for survival. Cox regression was then implemented to estimate the crude and adjusted hazard ratio (HR) and its 95% confidence intervals between haplotypes of *PPARA* and breast cancer mortality in both pre- and post-menopausal case women. Cox regression was also used to determine the crude and adjusted HR between breast cancer survival and BMI and weight gain by *PPARA* polymorphisms individually. As described above, effect modification of the obesity-breast cancer survival association was assessed using two methods, likelihood ratio tests and the interaction contrast ratios.

A log-rank test was conducted to evaluate the statistical significance of the difference between the cumulative incidence curves in common homozygotes versus heterozygotes and rare homozygotes combined for each SNP. There was no evidence of a difference under the dominant inheritance model for all investigated polymorphisms. The Kaplan-Meier curves and log rank tests can be found in Appendix D, Figures D.1-D.6 for all cause mortality.

To assess the assumption of proportional hazards, both statistical testing and graphical techniques were used. First, log-minus-log survival curves were examined under the dominant inheritance model. If the resulting curves appear non-parallel (such as

crossing hazards), the assumption of proportional hazards was deemed violated. Future models were then specified using time-varying interaction terms to relax this assumption. Formal tests of the proportional hazards assumption were also conducted by adding a time-dependent covariate into the model. If the interaction covariate was not statistically significant, the proportional hazards assumption was met for that covariate and the interaction term could be eliminated from the model. If the interaction was significant, however, the assumption was violated and the time-dependent variable was retained in the model. All results were then stratified by the time cut-off point specified by the interaction term. Only rs4253760 violated the proportional hazards assumption and a continuous time interaction term was necessary to include in the analysis to relax this assumption (see Appendix D, Tables D.23-D.28 and D.31-D.36 for model building procedures for each SNP individually, including the continuous time interaction term).

5.7.3 Hierarchical Modeling

Hierarchical regression involves multistage modeling, where the first stage can be expressed as ordinary logistic regression using maximum likelihood methods, and additional stages or "prior models" increase estimation accuracy by providing additional background information, such as physical distance between the SNPs [207, 208]. These additional stages use probabilistic methods, like shrinkage estimators, to weigh the data and the prior to give the posterior estimate greater stability [207, 208]. For example, empirical Bayes, a common shrinkage estimator, is based on the assumption of exchangeability, which states that the target parameters are indistinguishable in the absence of data [209]. Under the assumption of exchangeability, a summary parameter can be determined by weighing target parameters using the observed data, such as study sample size [209]. Initial semi-Bayesian analyses included SNP location (intron vs. coding) to better group the polymorphisms. Since functionality is not well understood and is not necessarily correlated with SNP location,

however, this additional classification was removed from the analyses. Thus, all semi-Bayesian analyses assumed the SNPs were exchangeable and no Z-matrix was specified.

Despite the additional computation complexity, hierarchical modeling offers two advantages over the more conventional logistic regression: (1) shrinkage estimation methods reduces Type II error rate, so that multiple comparisons are no longer a concern [208, 210], and (2) it reduces instability in the effect estimates due to multiple correlated exposures, such as multiple SNPs in the same model [208]. With multiple SNPs included in the model, over-parameterization with highly correlated exposure data was a concern. Hierarchical modeling neatly addressed this problem and provided a solution for multiple comparisons.

5.7.4 Sample Size and Power

5.6.4.1 Aims 1 and 2. Because these aims were restricted to post-menopausal women, genotyping data were estimated to be available for 870 cases and approximately the same number of controls. For the case-control analyses, the estimates for study power varied with the prevalence of the at-risk genotype or inferred haplotype (range 10% to 30%) and the expected ORs for the association between the genotypes and risk of breast cancer. The study had more than adequate power to detect modest associations. For the main gene or inferred haplotype effects, power was estimated at greater than 80% for ORs ≥ 2.0 for the postmenopausal women alone. Table 5.8 shows the estimates for the expected study power for varying prevalences of the at-risk genotype and varying odds ratios. These power calculations were done using Episheet.xls and the methods are described in detail in Rothman *et al* [211].

TABLE 5.8. Study power for main gene effects assuming an alpha of 0.05 and a 1:1 control/case ratio

Genotype Prevalence	Odds Ratio	Study Power
30%	3.0	>99%
	2.5	>99%
	2.0	>99%
	1.5	98%
20%	3.0	>99%
	2.5	>99%
	2.0	>99%
	1.5	95%
10%	3.0	>99%
	2.5	>99%
	2.0	>99%
	1.5	78%

For effect modification, the power was 80% or greater to detect substantial interactions of an OR ≥ 3.2. This interaction OR represents the OR for the doubly exposed (genotype and exposure positive) compared to the doubly non-exposed (genotype and exposure negative), and is calculated by the software program. These power calculations were conducted using POWER version 3.0 software available through the National Cancer Institute. This software is described in detail in Garcia-Closas *et al* [212]. The following assumptions were made to determine study power:

- The prevalence of obesity at the time of diagnosis for cases and reference date for controls, defined as BMI greater than 30.0, is 25% in this sample. [37]
- The prevalence of breast cancer in normal weight individuals is 1%.
- An OR of 1.80 was assumed for the effect of obesity on breast cancer risk based on preliminary analyses of the Long Island Breast Cancer Study Project.

Table 5.9 shows the estimates for study power for varying prevalences of the at-risk genotype and varying odds ratios for the genotype-breast cancer relationship.

5.6.4.2 Aims 3 and 4. For the survival analysis, power estimates were calculated assuming different underlying hazard ratios using the NQuery Advisor (v. 5.0) program. The survival rate in the wildtype group was assumed to be 0.124, which is consistent with the number of deaths observed in the LIBCSP follow-up study. The time period was set at five years and an alpha of 0.05 was used. The smallest detectable hazard ratios ranged from 1.5-2.0 assuming 80% study power and varied by genotype prevalence (10%-30%) and expected hazard ratio for the association between the genotype and risk of death in this cohort (see Table 6.10).

TABLE 5.9. Study power for gene-environment interactions assuming an alpha of 0.05 and a 1:1 case/control ratio

Prevalence		OR	OR	OR*	Study
Genotype	Exposure	Genotype	Exposure	Interaction	Power
30%	25%	3.0	1.8	10.8	88%
		2.5		9	88%
		2.0		7.2	89%
		1.5		5.4	89%
20%	25%	3.0	1.8	10.8	83%
		2.5		9	83%
		2.0		7.2	83%
		1.5		5.4	83%
10%	25%	3.0	1.8	10.8	63%
		2.5		9	62%
		2.0		7.2	61%
		1.5		5.4	60%

*OR (Odds Ratio) for genotype positive/exposure positive as compared with genotype negative/exposure negative subjects.

TABLE 5.10. Study power for main gene effects assuming an alpha of 0.05 for survival analysis (n=1,061)

Genotype Prevalence	Risk Ratio	Study Power
30%	3.0	>99%
	2.5	>99%
	2.0	>99%
	1.5	96.7%
20%	3.0	>99%
	2.5	>99%
	2.0	>99%
	1.5	91.3%
10%	3.0	>99%
	2.5	>99%
	2.0	99%
	1.5	72.0%

For effect modification in the survival analysis portion of the study, this study was slightly under powered (see Table 5.11). Given the lack of literature on this subject, however, this question was important to address despite the sample size limitations. These power calculations were conducted using POWER version 3.0 software available through the National Cancer Institute. This software is described in detail in Garcia-Closas *et al* [212]. The following assumptions were made to determine study power:

- The prevalence of obesity at the time of diagnosis for cases, defined as BMI greater than 30.0, is 25% in this sample [37].
- The prevalence of death in normal weight individuals is 12.4%, the prevalence observed across the entire cohort.

An HR of 1.80 was assumed for the effect of obesity on breast cancer survival based on hazard ratios and risk ratios presented in the literature (see Appendix B).

TABLE 5.11. Study power for survival analysis gene-environment interactions assuming an alpha of 0.05 (n=1,061)

Prevalence		HR	HR	HR*	Study
Genotype	Exposure	Genotype	Exposure	Interaction	Power
30%	25%	3.0	1.8	10.8	51.1%
		2.5		9	51.0%
		2.0		7.2	50.4%
		1.5		5.4	48.7%
20%	25%	3.0	1.8	10.8	42.8%
		2.5		9	42.8%
		2.0		7.2	42.2%
		1.5		5.4	40.5%
10%	25%	3.0	1.8	10.8	27.8%
		2.5		9	27.9%
		2.0		7.2	27.6%
		1.5		5.4	26.4%

*OR (Odds Ratio) for genotype positive/exposure positive as compared with genotype negative/exposure negative subjects.

Study power to examine associations were mostly adequate (see Tables 5.8, 5.9, 5.10 and 5.11), particularly for the main effects of the various genotypes. For the gene-environment interaction component, however, power was reduced for the less prevalent genotypes. For example, if the gene frequency is 10% in the case-control study, we will only have 60% power to detect an association. In the survival analysis, study power is even lower. Because of the novelty of this research question, however, the question of potential gene-environment interaction in the survival analysis was explored despite these limitations.

5.8 Study Advantages and Limitations

A major advantage of the proposed project is that the cases and controls were drawn from a population-based sample, which will increase the generalizability of the prevalence and estimates of risk associated with the *PPARA* polymorphisms. Another advantage of this investigation is the large sample size, which allowed examination of gene-environment interactions with sufficient statistical power. This project was also cost-effective since the biological specimens and exposure information had already been collected and were available to this study.

Despite the large overall sample size, a limitation of the study is the relatively restricted power to evaluate modification by certain breast cancer subtypes, such as estrogen receptor positive tumors. Small sample size is an inherent limitation of most investigations in molecular epidemiology due to the rarity of some minor alleles and the expense of both obtaining the blood samples needed and performing the assays. Despite these limitations, the project offered adequate power to detect interactions of substantial public health significance.

The detailed and extensive exposure information from LIBCSP enabled us to explore two different measures of obesity and their potential interaction with *PPARA* to influence breast cancer risk and survival. To the best of our knowledge, LIBCSP is the only case-

control study of breast cancer to inquire about weight by decade beginning at the age of 20, which afforded the unique opportunity to explore weight gain by decade and BMI by decade to determine critical time periods of risk for obesity mediated breast cancer. This proposal expands upon this research by investigating variants of *PPARA* that might play a role in breast cancer development and survival.

Reported weight at age 20 and other past exposures are subject to recall bias and nondifferential misclassification since disease diagnosis has occurred before exposure ascertainment. Several studies have found a high degree of recall in weight over time [213-216] although it has been suggested that underweight women may overestimate their past weight while overweight women may underestimate their past weight [216]. It is important to note, however, that the main gene effect relationship will not be subject to recall bias since participants are unlikely to know their *PPARA* status. Additionally, if abdominal adiposity is the most biologically relevant obesity measure, then BMI and weight gain may not fully capture the relationship between breast cancer, obesity and *PPARA*.

It has become increasingly apparent that examination of a single polymorphism within a given gene has not been particularly fruitful in identifying subgroups of women who may be genetically susceptible to breast cancer. A haplotype approach comprehensively assesses common variation over the entire gene to pinpoint haplotypes of potential functional importance without assuming any knowledge of functionality. Additionally, identifying tag SNPs reduces genotyping costs while maximizing the ability to capture genetic variation in a gene region [213]. Contributions by rare variants are missing using this approach, however.

Differences in participation rates between the cases and controls (82.1% vs. 62.7% respectively) suggest that participation bias may be present. In particular, response rates differences are most apparent in women over the age of 65, where cases showed a 71.2% response rate and controls demonstrated a 43.3% response. In comparison, the evidence of

participation bias is much less in women under the age of 65 with 88.9% and 76.1% response rate in cases and controls respectively. It is difficult to quantify the impact of participation bias on the frequency of the PPAR genotypes.

Approximately 73.0% of cases donated blood, suggesting missing data may bias our results. Additionally, complete treatment data is only available for a small subset of patients who participated in the follow-up interview. To address the impact of missing data on this study, a comparison of select tumor and demographic characteristics, including age, stage and tumor size, was performed. This analysis revealed no differences between genotyped cases (n=1,073) and all cases (n=1,508) by vital status (see Table D.30 in Appendix D).

Participants of this study are a population-based sample of women who reside on Long Island, of which 92.7% are whites, 5.0% are African-American and 2.3% are Asian-Americans or other. Because the ethnic distribution of our Long Island subjects differs from that of the American population as a whole, results from this project may not be readily applied to the U.S. population in general. Although the underlying prevalence of specific alleles and exposures may vary with ethnicity, it seems unlikely that the biological relations with breast cancer among participants in this study will differ from women in general.

Although *PPARA* has been studied in cardiovascular disease, no studies have examined *PPARA* in connection to breast cancer despite its involvement in insulin sensitivity and the inflammatory process. This study is the first to evaluate this possibility in a large population-based sample.

5.9 Summary

Over the past 30 years, considerable progress has been made in describing the epidemiology of breast cancer. Obesity appears to reduce risk of breast cancer in pre-menopausal women, but has consistently been found to elevate risk in post-menopausal women. These differences suggest that post-menopausal women may be most vulnerable

to obesity-related breast cancer. Obesity is also one of the only non-clinical factors considered an indicator of prognosis for breast cancer survival in both pre- and post-menopausal women. Weight gain, particularly in adulthood, leads to many endocrine changes, including increases in cytokines, insulin and progesterone. Peroxisome proliferators-activated receptor alpha (*PPARA*) has been suggested to promote insulin sensitivity by decreasing cytokine activity, and, thus may play an important role in breast cancer development in post-menopausal women.

We conducted an ancillary study using interview data and blood samples from the Long Island Breast Cancer Study Project. LIBCSP was a large population-based case-control study of breast cancer in Nassau and Suffolk counties in New York and was conducted from August 1996 to July 1997. Controls were frequency matched to cases on age and identified using random digit dialing for individuals under the age of 65 and Health Care Finance Administration rosters for controls aged 65 and older. 1,508 eligible cases and 1,556 eligible controls consented to participate in the study and approximately 73% of participants donated blood samples. Participants were questioned on their weight by decade, beginning at age 20, and on the reference date and case medical records were abstracted for tumor clinical and histological characteristics. We examined the association between genetic variations in *PPARA* and risk of breast cancer in post-menopausal women; evaluated interaction between *PPARA* polymorphisms, obesity and risk of breast cancer in post-menopausal women; examined the association between genetic variations in *PPARA* and breast cancer survival in pre- and post-menopausal women; and assessed interaction between *PPARA*, obesity and survival from breast cancer in pre- and post-menopausal women. This research helped advance knowledge of breast cancer etiology and prognosis through obesity mediated mechanisms.

CHAPTER 6 CASE CONTROL STUDY MANUSCRIPT

6.1 ABSTRACT

Peroxisome proliferator-activated receptor-alpha (PPARA) has been shown to increase fatty acid oxidation and decrease cytokine levels, and has been implicated in insulin production. Genetic variants of *PPARA* have been associated with cardiovascular disease, obesity and type II diabetes mellitus. Although no research to date has investigated the possible link between *PPARA* and breast cancer, the function of this gene suggests that it could play a role in breast cancer development. Six *PPARA* polymorphisms were evaluated in association with incident breast cancer in a population-based case-control study (n=1073 cases, n=1112 controls), using unconditional logistic and multilevel regression, and haplotype-based analyses. The odds of breast cancer were doubled among women with *PPARA* polymorphism rs4253760 (OR=1.97 for rare vs. common homozygote alleles; 95% CI: 1.14, 3.43). This association remained constant with the inclusion of all interrogated polymorphisms studied in hierarchical models. No additive interactions with body mass index or weight gain were present, but there was some evidence of interaction with aspirin use, defined as use at least once per week for six months or longer, for rs135542 and rs4253699. Fourteen haplotypes were imputed with frequencies greater than 1% among post-menopausal women, but no statistically significant differences in haplotype frequencies between cases and controls were evident. Our results are the first to evaluate the relationship between *PPARA* and breast cancer incidence and suggest that replication in an independent cohort is warranted.

6.2 INTRODUCTION

The peroxisome proliferator-activated receptor (PPAR) family is composed of three nuclear hormone receptor (NHR) genes: PPAR-gamma, PPAR-alpha (*PPARA*) and PPAR-delta. In general, NHRs encode proteins that induce gene transcription by binding to the promoter region of a target gene. PPARs are activated when small lipophilic hormones (ligands) bind to a ligand-specific NHR [141]. *PPARA* ligands include palmitic acid, arachidonic acid and stearic acid in addition to compounds such as fenofibrate, bezafibrate and non-steroidal anti-inflammatory drugs (NSAIDs) [142, 143].

PPARA is primarily expressed in organs with high fatty acid oxidation rates [143, 146], such as the liver, kidney, heart, brown adipose tissue, and, in small quantities, white adipose tissue [142]. *PPARA* has also been found in human breast cancer cell lines, where its activation has been associated with increased proliferation [217]. *PPARA* has been shown to regulate lipid metabolism by controlling the uptake and oxidation of fatty acids [143]. This regulation can lead to an excess of free fatty acids, which may contribute to insulin resistance [148]. Fibrates, which are *PPARA* agonists, have also been shown to reduce the expression of multiple cytokines, including interleukin-6, fibrinogen, and C reactive protein in humans [143].

Two *PPARA* isoforms have been characterized (*PPARA*1 and *PPARA*2). Both isoforms are expressed in human tissue [218, 219]. *PPARA*1 encodes the entire gene while *PPARA*2 is truncated at exon 6. This truncation results in the absence of the ligand-binding domain in the gene's protein and, consequently, prevents activation by the ligand. Therefore, all study inferences pertain to *PPARA*1 because of its protein's known activity.

Given its role in energy homeostasis, it seems feasible that genetic variation in *PPARA* could influence disease incidence. The epidemiologic literature has focused on a functional polymorphism that results in a leucine to valine substitution at codon 162 of exon

5 (L162V, rs1800206) and a subsequent cytosine to guanine base change in the DNA binding domain region of PPARA protein. This missense polymorphism has been shown to have a functional impact based on co-transfection assays; specifically, the V162 allele showed elevated ligand-dependent transcription activity compared to the L162 allele [151].

Although no research to date has investigated the possible link between *PPARA* and breast cancer, the biology and epidemiology of the gene suggest it could play a role in breast cancer incidence. Genetic variants of *PPARA* have been linked to lipoprotein levels [162, 220, 221], cardiovascular disease [176, 179, 222], obesity [165, 169] and type II diabetes [152, 167, 223]. These conditions operate on the same pathways as breast carcinogenesis, including inflammation and insulin resistance. Therefore, the goals of this study were to examine the association between *PPARA* genetic polymorphisms and breast cancer development using single polymorphism and haplotype-based approaches and semi-Bayesian techniques. Interactions with body mass index (BMI), weight gain, aspirin use and menopausal status were also explored.

6.3 Materials and Methods

To evaluate the study aims, we utilized data and samples from the Long Island Breast Cancer Study Project (LIBCSP), a large population-based case-control study; details of the parent study population and data collection methods have been previously published [37]. The study protocol was approved by the institutional review boards of participating institutions.

6.3.1 Study population

English speaking women newly diagnosed with a primary *in situ* or invasive breast cancer between August 1, 1996, and July 31, 1997 were eligible to be study cases if they were 20 years of age or older at diagnosis and were residents of Nassau or Suffolk counties

on Long Island, New York. Cases were identified through daily contact with the 33 Long Island and New York City hospitals that served women with breast cancer in these two counties. Physician permission was obtained prior to case contact.

Controls were randomly selected from among English-speaking female residents of the same two Long Island counties, and were frequency matched to the expected age-distribution of case subjects by 5-year age group. Potentially eligible controls were identified by Waksberg's method of random digit dialing [180] for women under 65 years of age, and by Health Care Finance Administration rosters for women 65 years of age and older.

Participants in the LIBCSP included 1,508 (82.1%) of eligible case women and 1,556 (62.7%) of eligible control women. Study subjects ranged in age from 24 to 98 years, and 93.8% of cases and 91.8% of controls were Caucasian while 4.6% and 5.5% were African American, respectively. Approximately 68% of cases and 67% of controls were post-menopausal (n=1,010 and 993 respectively).

6.3.2 Exposure assessment

6.3.2.1 Questionnaire. Case-control interviews were administered by trained interviewers in respondents' homes. Interviews took an average of 101 minutes to complete, and included assessment of known and suspected risk factors for breast cancer. In previous analyses, an increase in weight and weight gain, particularly after age 50, was positively associated with postmenopausal breast cancer [9], and aspirin use was inversely associated with breast cancer among women of all ages [45]. Other factors found to be associated with breast cancer in the LIBCSP have also been previously described [37].

6.3.2.2 Biologic specimens. Among respondents who completed the interview, 73.0% of cases and 73.3% of controls donated a blood sample. DNA was isolated using methods previously described [59]. Of the 1,102 cases and 1,141 controls who donated blood, 19

(1.7%) and 22 (1.9%), respectively, were later found to have insufficient DNA. Thus, there were a total of 1,083 case and 1,119 control samples available for genotyping. For analyses restricted to post-menopausal women, genotyping was available for 708 cases and 692 controls.

6.3.3 Genetic polymorphisms

6.3.3.1 Selection of tagging SNPs. Tag single nucleotide polymorphisms (SNP) were selected to represent comprehensive coverage of the *PPARA* gene by binning SNPs with a minor allele frequency greater than 0.10 and an estimated minimum pairwise correlation of 0.80. For *PPARA*, sequence data were available for 23 European-Americans and 24 African-American Coriell samples on the University of Washington-Fred Hutchinson Cancer Research Center Variation Discovery Resource website, http://pga.gs.washington.edu/ (or Program for Genomic Applications (PGA)). Haplotype tagging SNPs were identified using the PGA LDSelect Program [190] run for European Americans only (given the relative racial homogeneity of the LIBCSP population). This program has been shown to select a maximally informative set of common SNPs that distinguish 80% of common haplotypes, and is based on the r^2 linkage disequilibrium (LD) statistic [191]. Because of its low prevalence but functional importance, L162V was forced into the program. Based on this program, fourteen *PPARA* SNPs were identified for genotyping (Table 6.1).

6.3.3.2 Genotyping**.** Genotyping was conducted at Columbia University, New York, NY. All LIBCSP DNA samples are available on 96 well master plates. Approximately 10% of the samples on each plate were duplicates, and laboratory personnel were blinded to case-control and duplicate status. Genotyping was carried out using iPLEX technology (Sequenom, San Diego, CA) on a MassARRAY Compact Analyzer. This multiplex method uses the mass of the incorporated nucleotide for identification of genotype. For SNPs that

could not be mulitplexed (rs4253623 and rs4253699), Taqman (Applied Biosystems, Foster City, CA) assays were developed and were run on an ABI 7500 Real Time PCR system.

Kappa statistics were estimated to determine concordance between blinded repeat samples on each plate, and only those SNPs with a minimum kappa statistic of 0.90 were included in analyses. Six of the 14 identified tag SNPs for *PPARA* met this criterion (Table II).

6.3.4 Statistical methods

Hardy-Weinberg equilibrium (HWE) was tested among controls to ensure assumptions of parametric statistical tests were met using a permuted version of the exact test in SAS/Genetics version 9.1 (Cary, NC) [194, 224]. Pairwise LD [196, 197] for the six assayed SNPs was determined using Haploview 4.0 [225]. Unconditional logistic regressions including individual *PPARA* SNPs and all SNPs together were conducted using SAS. Main gene effects were modeled by using the full genotype model and by combining heterozygotes and rare homozygotes. All models were adjusted for age, the frequency matched variable, and common homozygotes were the reference group for all analyses. Linear trend tests for allelic effects were also performed by coding each genotype as 0, 1, or 2 based on the number of risk alleles. Separate genetic models were also used to estimate effects among post-menopausal women only. We hypothesized *a priori* that the gene's effect would be most pronounced in post-menopausal women because of the relationship between obesity and breast cancer incidence in these women [4, 61].

In addition to conventional unconditional logistic regression modeling, hierarchical modeling using a semi-Bayesian approach was performed among all women and among post-menopausal women only. SAS IML commands developed by Witte et al. [226] were used to fit the multilevel models. Hierarchical models assumed that all SNPs were exchangeable. The first hierarchical model specified a τ^2, or prior residual variance, of 0.169

while the second hierarchical model assumed a τ^2 of 0.345. A residual variance equal to 0.345 specifies that the odds ratio will fall within a 10-fold range with 95% confidence, while a τ^2 of 0.169 specifies a five-fold range.

Confounders were chosen *a priori* using directed acyclic graphs (DAG). Race (White/non-White), family history of breast cancer (yes/no), and Jewish ethnicity (Jewish/not Jewish) were examined as potential confounders based on the DAG. Covariates that resulted in a 10% or greater change in the beta coefficient of the genotype effect estimate were considered confounders. Using this criterion, no confounders were identified.

Product interaction terms were added to conventional logistic and Bayesian models for gene variants and aspirin use, defined as use at least once per week for six months or longer, (among all women), and for BMI at reference and weight gain since age 50 (among post-menopausal women). Due to small sample sizes, heterozygotes and rare homozygotes were combined for all interaction models. Interactions were considered for each SNP separately with adjustment for all other SNPs and age. Weight gain models were also adjusted for BMI at age 50 to account for the potential influence of body mass on weight gain. Interaction contrast ratios and 95% confidence intervals (CI) were calculated to assess departures from additive risks [227]. Departures from additivity were defined as having the 95% confidence interval for the interaction contrast ratio excluding the null value of zero.

Haplotype reconstruction was performed using an expectation maximization (EM) algorithm [199]. EM haplotype inference uses an individual's genotype data to impute the probability of having a certain haplotype pair [187]. Haplotype specific odds ratios (OR) and 95% CIs were estimated relative to all other haplotypes for all women and for post-menopausal women only. All haplotype analyses were conducted using unconditional logistic regression in SAS/Genetics.

6.4 RESULTS

Six *PPARA* SNPs were included in analyses: rs135542, rs1800206, rs4253623, rs4253699, rs4253755, and rs4253760 (as ordered in Figure 1). All six SNPs were in HWE, and MAFs ranged from 5% to 22.4% (Table 6.2). We found very low correlation between the *PPARA* polymorphisms (Figure 6.1), suggesting the PGA population is an appropriate reference for tag SNP selection in the LIBCSP.

As shown in Table 6.2, *PPARA* polymorphism rs4253760 was associated with nearly a 100% increase in the risk of postmenopausal breast cancer (OR=1.97 for rare versus common homozygotes; 95% CI: 1.14, 3.43) and showed evidence of linear trend (p=0.02). Consistent findings were noted when all SNPs were assessed in one model and also for the hierarchical models (Table 6.3).

In general, hierarchical models produced more precise estimates compared to conventional analyses. This effect is most evident for the most unstable estimates, such as SNPs with a low prevalence in this population. For example, the OR for the association comparing rare to common homozygotes in rs1800206 was 4.14 (95% CI: 0.43, 39.79) among post-menopausal women, adjusting for all other *PPARA* SNPs (model labeled Conventional[b] in Table 6.3); using multi-level modeling, this OR decreased to 1.29 (95% CI: 0.58, 2.87; Hierarchical[d]).

Statistically significant interaction by aspirin use was present for rs135542 and rs4253699 in analyses including all women although comparable departures from additivity were evident for all SNPs except rs1800206 (Table 6.4). Among post-menopausal women, there was no evidence of interaction by BMI at reference or by weight gain since age 50, both measured as three level categorical variables, with no consistent pattern of elevated additive effects with increasing levels of obesity (results not shown).

The haplotype reconstruction and analysis created 12 haplotypes with frequencies greater than 1% from the six SNPs analyzed in all women, and 14 haplotypes in the post-menopausal women. Haplotype distributions were similar between cases and controls among all women (data not shown) and among post-menopausal women only (Table 6.5). Due to the low prevalence of select haplotypes in this population, effect estimates were imprecise, particularly those for associations among postmenopausal women only. For example, although the OR for haplotype 10 among post-menopausal women (n=30 cases and 11 controls) was elevated relative to all other haplotypes, the 5-fold width of the confidence interval indicated substantial imprecision (OR=5.02; 95% CI: 1.45, 17.39).

6.5 Discussion

We found that the *PPARA* genetic polymorphism rs4253760 was associated with a two-fold increase in the odds of post-menopausal breast cancer. This association persisted in the hierarchical models which were adjusted for the other five *PPARA* SNPs. This finding is consistent with our prior expectation of a more pronounced effect in post-menopausal women. rs4253760 is located in intron six and tags ten SNPs (MAF>10%) based on LDSelect and PGA data. This polymorphism may be correlated with the causal SNP in the *PPARA* gene although it does not tag any non-synonymous coding polymorphisms and the haplotype analyses with the rare variant of rs4253760 are not supportive of a causal effect.

Our results also implicate aspirin use as a possible effect modifier, particularly for rs135542 and rs4253699. In the LIBCSP population, aspirin use was associated with a 20% reduction in the odds of breast cancer [45]. In this study, NSAID users with the reference genotype showed a slight inverse association with breast cancer. Non-users with the variant genotype demonstrated a weak positive association while a strong inverse association is evident among users with the variant genotype. These findings suggest that aspirin use may modify the PPARA gene effect.

Despite the additional computational complexity, hierarchical modeling offers two advantages over conventional logistic regression: (1) the shrinkage estimation method reduces Type II error rate [208, 210], and (2) it reduces instability in the effect estimates due to multiple correlated exposures, such as multiple SNPs in the same model [208]. For this study, two different τ^2 values were considered: 0.169 and 0.345. The difference between the hierarchical and conventional models was most apparent for the fully specified conventional model, where each genotype was modeled simultaneously. Here the τ^2 0.169 models consistently produced estimates that were closer to the null and more precise than either the τ^2 0.345 hierarchical models or the conventional logistical models.

It has become increasingly apparent that studies of a single polymorphism are not necessarily the best approach to identifying deleterious variants [228]. Haplotypes that use tag SNPs selected from bins comprehensively assess variation over the entire gene to identify *cis*-*cis* interactions, where tag SNPs are interacting on the same chromosome to increase disease risk. Our study implicates haplotype 10 in breast cancer incidence, although the low precision of the estimate makes interpretation difficult.

One limitation of this data is the low concordance of eight tag SNPs with their blinded repeats, which prevented us from including them in our analyses. Five of the requested SNPs were found in repeat regions of the *PPARA* gene and, therefore, would have been difficult to genotype successfully. The low concordances found for the remaining three polymorphisms (rs4253730, rs4253655 and rs135543) with their blinded repeats are not easily explained and could be the result of genotyping error.

Although these omissions reduced gene coverage, this study still provides more coverage than previous studies and analyses a well-characterized population. Only two studies have examined *PPARA* haplotypes to date, none of which examined cancer outcomes or included more than three SNPs [152, 176]. Flavell et al. [152] investigated three SNPs in connection to age of onset and progression of Type II diabetes mellitus while

Doney et al. [176] explored the link between myocardial infarction risk among individuals with Type II diabetes mellitus and two *PPARA* polymorphisms. Thus, even with reduced gene coverage, this study advances scientific knowledge of *PPARA* and its role in breast cancer incidence.

Reported weight at age 50 and other past exposures may be subject to recall bias and non-differential misclassification since disease diagnosis has occurred before exposure ascertainment. Several studies have found a high degree of recall in weight over time [213-216] although it has been suggested that underweight women may overestimate their past weight while overweight women may underestimate their past weight [216]. This study examined body size as a potential modifier of the gene's effect; therefore, for recall bias to be present, cases and controls would have to recall their weight differentially by *PPARA* genotype. Since women are unlikely to know their PPARA status, recall bias is unlikely to play a role in these analyses.

Although a benefit of our study is that it is population-based, only 73% of cases and controls donated blood. However, the distributions of risk factors for breast cancer among all women in the study were comparable to those of women who donated blood (data not shown). Lastly, the ethnic distribution of our Long Island subjects differs from that of the American population as a whole, with 92.7% of our study population being Caucasian; thus, results from this study may not be readily applied to the U.S. population in general. Although the underlying prevalence of specific alleles and exposures may vary with ethnicity, it seems unlikely that the biological relations with breast cancer among participants in this study will differ from women in general. In fact, as expected, the minor allele frequencies differed only slightly between the PGA European-American population and our Long Island women, with the greatest difference noted for rs4253755 (18.2% versus 22.0%, respectively). This difference is most likely due to PGA's small sample (n=23) for determining the allele frequencies in European-Americans.

Lack of reproducibility among genetic studies has called into question the utility of association studies in genetic epidemiology [229]. This is the first study to examine the association between *PPARA* polymorphisms and breast cancer development. While our use of semi-Bayesian techniques minimizes many issues inherent with small cell sizes, such as large but imprecise effects and low p-values, further replication is needed to confirm our findings.

In summary, although *PPARA* has been studied in cardiovascular disease, no studies have examined *PPARA* in connection to breast cancer. This study is the first to investigate its relationship to breast cancer risk and interaction by aspirin use and obesity in a large population-based sample. This study suggests that variants of *PPARA* modify the association between breast cancer and aspirin use. Research investigating *PPARA*'s possible involvement in the inflammatory pathway is needed. We found that among postmenopausal women carrying the homozygous alleles for rs4253760, the odds of breast cancer was nearly doubled. These findings warrant further investigation in an independent cohort.

6.6 ACKNOWLEDGEMENTS

This research was funded in part by grants from the National Cancer Institute and the National Institute of Environmental Health Sciences (5T32CA72319, UO1CA/ES66572, P30ES09089, and P30ES10126); and the Breast Cancer Research Foundation

.

6.7 TABLES

TABLE 6.1. PPARA haplotype tagging SNPs identified using LDSelect and the PGA European-American population as the reference panel

rs#	SNP Location	Base Pair Change (major>minor)	Minor Allele Frequency[a]	Reference
rs4253730	Intron	A>G	0.182	LDSelect[a]
rs4253760[b]	Intron	T>G	0.196	LDSelect
rs4253705	Intron	T>C	0.190	LDSelect
rs135543	Intron	G>A	0.283	LDSelect
rs135542[b]	Intron	A>G	0.205	LDSelect
rs4253649	Intron	C>G	0.370	LDSelect
rs4253758	Intron	T>C	0.217	LDSelect
rs4253699[b]	Intron	T>C	0.182	LDSelect
rs4253655	Intron	G>A	0.143	LDSelect
rs4253681	Intron	T>C	0.136	LDSelect
rs4253755[b]	Intron	G>A	0.130	LDSelect
rs4253706	Intron	G>A	0.119	LDSelect
rs4253623[b]	Intron	A>G	0.109	LDSelect
rs1800206[b]	Exon Leu>Val	C>G	0.022	PGA[a]

[a]Sequence data on 23 European-Americans Coriell samples are available on the University of Washington-Fred Hutchinson Center Research Center Variation Discovery Resource (PGA) website (http://pga.gs.washington.edu/).
[b]SNPs included in analyses

TABLE 6.2. Summary table of ORs for association between six PPARA polymorphisms and breast cancer risk by menopausal status in LIBCSP

Genotype	Genotype[a]	MAF[b]	Cases		Controls		P for trend[d]	OR[e]	95% CI
			N	%[c]	N	%[c]			
All women									
rs135542	AA	0.224	602	59.4	634	59.3		1.00	
	AG		370	36.5	392	36.7		1.01	0.84, 1.21
	GG		41	4.1	43	4.0	0.90	1.01	0.65, 1.57
	AG+GG		411	40.6	435	40.7		1.01	0.85, 1.21
	Total		*1013*		*1069*				
rs1800206 (L162V)	CC (L/L)	0.054	927	89.7	973	89.7		1.00	
	CG (L/V)		100	9.7	109	10.1		0.97	0.73, 1.30
	GG (V/V)		7	0.7	3	0.3	0.72	2.44	0.63, 9.50
	L/V+V/V		107	10.4	112	10.3		1.01	0.76, 1.34
	Total		*1034*		*1085*				
rs4253623	AA	0.123	811	77.5	849	77.0		1.00	
	AG		218	20.8	236	21.4		0.97	0.79, 1.19
	GG		17	1.6	17	1.5	0.77	0.98	0.49, 1.93
	AG+GG		235	22.5	253	23.0		0.97	0.79, 1.19
	Total		*1046*		*1102*				
rs4253699	TT	0.220	624	60.0	671	61.5		1.00	
	CT		358	34.4	362	33.2		1.06	0.88, 1.27
	CC		58	5.6	59	5.4	0.52	1.07	0.73, 1.57
	CT+CC		416	40.0	421	38.6		1.06	0.89, 1.27
	Total		*1048*		*1092*				
rs4253755	GG	0.120	803	76.5	845	77.4		1.00	
	AG		231	22.0	231	21.2		1.05	0.85, 1.29
	AA		16	1.5	16	1.5	0.64	1.05	0.52, 2.12
	AG+AA		247	23.5	247	22.6		1.05	0.86, 1.29
	Total		*1050*		*1092*				
rs4253760	TT	0.186	675	66.2	713	67.2		1.00	
	GT		293	28.7	302	28.5		1.02	0.84, 1.24
	GG		52	5.1	46	4.3	0.41	1.25	0.83, 1.87
	GT+GG		345	33.8	348	32.8		1.05	0.88, 1.26
	Total		*1020*		*1061*				
Post-menopausal									
rs135542	AA	0.220	395	59.0	397	60.0		1.00	
	AG		247	36.9	239	36.1		1.07	0.85, 1.35
	GG		27	4.0	26	3.9	0.61	1.04	0.59, 1.81
	AG+GG		274	41.0	265	40.0		1.07	0.86, 1.33
	Total		*669*		*662*				
rs1800206 (L162V)	CC (L/L)	0.052	610	89.3	609	89.7		1.00	
	CG (L/V)		68	10.0	69	10.2		1.01	0.71, 1.44
	GG (V/V)		5	0.7	1	0.2	0.51	5.07	0.59,

									43.71
	L/V+V/V		73	10.7	70	10.3		1.06	0.75, 1.51
	Total		*673*		*679*				
rs4253623	AA		533	77.4	522	75.9		1.00	
	AG		145	21.0	150	21.8		0.96	0.74, 1.24
	GG	0.132	11	1.6	16	2.3	0.43	0.66	0.30, 1.45
	AG+GG		156	22.6	166	24.1		0.93	0.72, 1.20
	Total		*689*		*688*				
rs4253699	TT		411	59.8	425	62.2		1.00	
	CT	0.217	243	35.4	220	32.2		1.13	0.90, 1.42
	CC		33	4.8	38	5.6	0.62	0.92	0.56, 1.49
	CT+CC		276	40.2	258	37.8		1.10	0.88, 1.37
	Total		*687*		*683*				
rs4253755	GG		532	76.3	535	78.8		1.00	
	AG	0.113	153	22.0	135	19.9		1.12	0.86, 1.46
	AA		12	1.7	9	1.3	0.29	1.36	0.57, 3.27
	AG+AA		165	23.7	144	21.2		1.14	0.88, 1.47
	Total		*697*		*679*				
rs4253760	TT		441	65.0	454	69.6		1.00	
	GT		199	29.4	177	27.2		1.14	0.90, 1.46
	GG	0.168	38	5.6	21	3.2	0.02	1.97	1.14, 3.43
	GT+GG		237	35.0	198	30.4		1.23	0.98, 1.55
	Total		*678*		*652*				

[a]The combined heterozygotes and rare homozygotes were modeled separately and compared to common homozygotes
[b]Minor allele frequency calculated among controls
[c]May not add up to 100.0 due to rounding
[d]P-value for trend was calculated by coding each genotype as 0, 1, or 2 based on the number of risk alleles
[e]Adjusted for age, measured in five year intervals

TABLE 6.3. Odds ratios and 95% CIs for six PPARA polymorphisms and breast cancer among post-menopausal women in the LIBCSP for conventional and hierarchical models with varying priors

SNP	Alleles[a]	Conventional[b]		Conventional[c]		Hierarchical[d]		Hierarchical[e]	
		OR[f]	95% CI	OR[f]	95% CI	OR[f]	95% CI	OR[f]	95% CI
rs135542	A**G**	1.07	0.85, 1.35	1.05	0.82, 1.34	1.06	0.84, 1.34	1.06	0.83, 1.34
	GG	1.04	0.59, 1.81	1.13	0.63, 2.02	1.13	0.69, 1.83	1.13	0.67, 1.91
rs1800206	C**G**	1.01	0.71, 1.44	0.93	0.60, 1.46	0.98	0.67, 1.43	0.96	0.64, 1.43
	GG	5.07	0.59, 43.71	4.14	0.43, 39.79	1.29	0.58, 2.87	1.47	0.51, 4.27
rs4253623	A**G**	0.96	0.74, 1.24	1.15	0.86, 1.52	1.14	0.87, 1.49	1.14	0.86, 1.51
	GG	0.66	0.30, 1.45	0.61	0.26, 1.43	0.83	0.46, 1.53	0.76	0.38, 1.52
rs4253699	**C**T	1.13	0.90, 1.42	0.97	0.73, 1.28	1.00	0.77, 1.29	0.99	0.76, 1.29
	CC	0.92	0.56, 1.49	0.89	0.49, 1.62	1.02	0.64, 1.62	0.98	0.59, 1.64
rs4253755	**A**G	1.12	0.86, 1.46	1.06	0.74, 1.52	1.07	0.78, 1.47	1.06	0.76, 1.49
	AA	1.36	0.57, 3.27	1.29	0.44, 3.76	1.20	0.62, 2.30	1.22	0.56, 2.69
rs4253760	**G**T	1.14	0.90, 1.46	1.17	0.85, 1.61	1.13	0.85, 1.49	1.15	0.85, 1.54
	GG	1.97	1.14, 3.43	2.19	1.17, 4.13	1.71	1.05, 2.80	1.90	1.10, 3.28

[a]Minor alleles in bold
[b]Models included one SNP and age only
[c]All SNPs plus age in one model
[d]All SNPs exchangeable and prior $\tau^2 = 0.169$
[e]All SNPs exchangeable and prior $\tau^2 = 0.345$

TABLE 6.4. Odds ratios, interaction contrast ratios and 95% CIs for six *PPARA* polymorphisms and breast cancer by aspirin use in all LIBCSP women under the dominant inheritance model

Interaction[b]	**Conventional**[a]		**Interaction**[a]	
	OR[c]	**95% CI**	**ICR**	**95% CI**
rs135542				
AA*(Aspirin Use)[d]	0.92	0.69, 1.23		
(A**G**+**GG**)*(Non-Use)	1.09	0.88, 1.36		
(A**G**+**GG**)* (Aspirin Use)	0.56	0.39, 0.81	-0.45	-0.87, -0.04
rs1800206				
CC*(Aspirin Use)	0.73	0.57, 0.92		
(C**G**+**GG**)*(Non-Use)	1.05	0.71, 1.56		
(C**G**+**GG**)*(Aspirin Use)	0.94	0.46, 1.89	0.16	-0.59, 0.91
rs4253623				
AA*(Aspirin Use)	0.64	0.49, 0.84		
(A**G**+**GG**)*(Non-Use)	0.94	0.73, 1.22		
(A**G**+**GG**)*(Aspirin Use)	1.07	0.71, 1.61	0.48	-0.01, 0.98
rs4263699				
TT*(Aspirin Use)	0.60	0.45, 0.80		
(**C**T+**CC**)*(Non-Use)	0.85	0.66, 1.09		
(**C**T+**CC**)*(Aspirin Use)	0.91	0.62, 1.32	0.46	0.08, 0.84
rs4253755				
GG*(Aspirin Use)	0.65	0.50, 0.84		
(**A**G+**AA**)*(Non-Use)	0.93	0.68, 1.27		
(**A**G+**AA**)*(Aspirin Use)	1.08	0.67, 1.75	0.50	-0.05, 1.05
rs4253760				
TT*(Aspirin Use)	0.62	0.47, 0.82		
(**G**T+**GG**)*(Non-Use)	0.92	0.70, 1.21		
(**G**T+**GG**)*(Aspirin Use)	0.97	0.65, 1.44	0.42	-0.01, 0.85

[a]Interaction by each SNP was modeled separately although all models were adjusted for all SNPs
[b]Minor allele in bold
[c]Adjusted for age, measured in five year age intervals
[d]Aspirin use was defined as use at least once per week for six months or longer

TABLE 6.5. Odds ratios and frequencies by case-control status for 14 PPARA haplotypes relative to all other haplotypes among post-menopausal women in LIBCSP

Haplotype Number[a]	rs135542	rs1800206	rs4253623	rs4253699	rs4253755	rs4253760	Control N (%)	Case N (%)	OR (95% CI)[b]
1	A	C	A	**C**	**A**	**G**	55 (5.5)	53 (5.3)	1.01 (0.49, 2.09)
2	A	C	A	**C**	**A**	T	18 (1.9)	32 (3.2)	2.03 (0.71, 5.82)
3	A	C	A	**C**	G	**G**	18 (1.8)	14 (1.4)	0.97 (0.23, 4.18)
4	A	C	A	**C**	G	T	55 (5.5)	53 (5.2)	0.89 (0.40, 1.99)
5	A	C	A	T	**A**	**G**	25 (2.5)	32 (3.2)	1.64 (0.61, 4.46)
6	A	C	A	T	G	**G**	19 (1.9)	36 (3.6)	2.83 (0.93, 8.65)
7	A	C	A	T	G	T	414 (41.7)	396 (39.2)	0.92 (0.70, 1.20)
8	A	C	**G**	**C**	G	T	13 (1.5)	10 (1.0)	0.77 (0.12, 4.91)
9	A	C	**G**	T	G	T	104 (10.5)	103 (10.2)	0.97 (0.59, 1.59)
10	A	**G**	A	**C**	G	**G**	11 (1.1)	30 (3.0)	5.02 (1.45, 17.39)
11	A	**G**	A	T	G	**G**	13 (1.3)	8 (0.8)	0.47 (0.08, 2.83)
12	**G**	C	A	**C**	G	T	15 (1.5)	8 (0.8)	0.42 (0.05, 3.55)
13	**G**	C	A	T	G	**G**	11 (1.1)	11 (1.1)	1.07 (0.13, 8.67)
14	**G**	C	A	T	G	T	173 (17.4)	182 (18.0)	1.19 (0.79, 1.80)

[a]Haplotypes with frequency ≥ 0.01, minor alleles in bold
[b]Adjusted for age, measured in five year age intervals

6.8 LEGENDS TO FIGURES

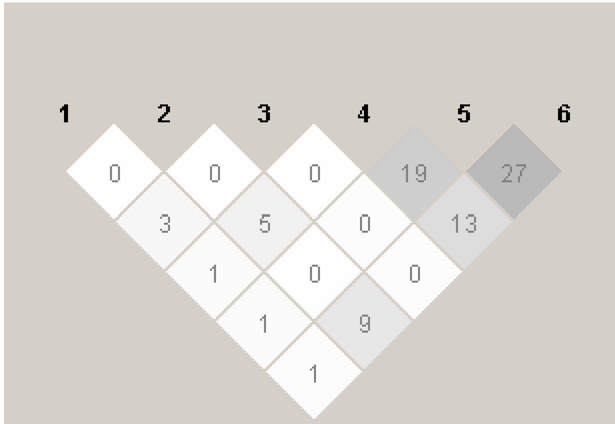


FIGURE 6.1. Linkage disequilibrium (r^2) between six PPARA tag SNPs among LIBCSP controls

CHAPTER 7 SURVIVAL STUDY MANUSCRIPT

7.1 ABSTRACT

Peroxisome proliferator-activated receptor-alpha (*PPARA*) is part of the nuclear hormone receptor family that encodes proteins that induce gene transcription by binding to the promoter region of a target gene. *PPARA* has been linked to obesity and Diabetes Mellitus, but has not yet been studied in relation to breast cancer survival. Because breast cancer survival is believed to operate through similar pathways, this gene could contribute to breast cancer survival. Six *PPARA* polymorphisms were evaluated for an association with survival among a population-based cohort of women diagnosed with a first primary breast cancer in 1996-1997 (n=1073). Interviews and blood samples were collected shortly after diagnosis. The National Death Index was used to determine vital status through December 31, 2002. Cox regression and haplotype-based analyses were performed for all-cause and breast cancer-specific mortality (n = 132 (12.3%) and 88 (8.2%), respectively). *PPARA* polymorphism rs4253760 was associated with a two-fold increase in all-cause mortality with inclusion of a continuous time interaction term (HR=2.25 for rare vs. common homozyote alleles; 95% CI: 1.00, 5.08 at baseline). This interaction term implies that survival improves over time; thus, caution is necessary when interpreting this HR. Haplotype reconstruction created 12 haplotypes with frequencies greater than 1%, but no clear differences in haplotype frequencies by all cause or breast cancer-specific mortality

were evident. Although the confidence intervals for the effect estimates are wide, these results suggest *PPARA* may be involved in survival among women with breast cancer and attempts to replicate in an independent cohort are warranted.

7.2 INTRODUCTION

The peroxisome proliferator-activated receptor (PPAR) family is composed of three nuclear hormone receptor genes: PPAR-gamma, PPAR-alpha (*PPARA*) and PPAR-delta. In general, hormone nuclear receptor genes encode proteins that induce gene transcription by binding to the promoter region of a target gene [141]. *PPARA* is activated when compounds such as fenofibrate, bezafibrate and non-steroidal anti-inflammatory drugs, bind to its receptor [142, 143].

The *PPARA* epidemiologic literature has focused on one functional polymorphism, rs1800206 or L162V. A cytosine to guanine base change occurs at codon 162 of exon 5, resulting in a leucine to valine substitution in the binding domain region of the PPARA protein. This missense polymorphism has been shown to have a functional impact on the receptor, increasing ligand-dependent transcription activity in rare homozygotes [151].

L162V has been linked to atherosclerosis [158, 163, 179], lipoprotein levels [162, 174, 178], obesity [165, 169], and Type II Diabetes Mellitus [152, 167, 223], but has not yet been studied in relation to breast cancer survival. Because breast cancer survival is believed to operate through similar pathways, including inflammation and insulin resistance, this research could elucidate a genetic contributor to breast cancer survival.

7.3 Materials And Methods

7.3.1 *Study Population*

The 1,508 cases who participated in the Long Island Study Project (LIBCSP) were eligible for this follow-up study. Women were eligible for the current study if they had newly diagnosed in situ or invasive breast cancer between August 1, 1996 and July 31, 1997, were 20 years of age or older at diagnosis, were residents of Nassau or Suffolk counties in Long Island, NY, and completed the baseline LIBCSP questionnaire.

7.3.2 *Data and Biological Sample Collection*

The baseline interview was administered by a trained interviewer in the respondents' home and took an average of 101 minutes to complete. Details of the parent study population and data collection methods have been previously published [37]. Among respondents who completed the baseline interview, 73.0% of cases and 73.3% of controls donated a blood sample. DNA was isolated using methods previously described [59].

Of the 1508 LIPCSP case participants, 198 (13.1%) died during follow-up according to the National Death Index (NDI), while the rest (n=1,310) were alive through December 31, 2002, the last date that the NDI considered mortality data to be accurate. For the 1,073 cases that were successfully genotyped, 941 were alive (87.7%) and 132 dead (12.3%) at the end of the follow-up period. Eighty-eight deaths (8.2%) were breast-cancer related while 44 (4.1%) were from other causes, such as lung cancer and cardiovascular disease.

7.3.3 *Selection of tagging SNPs*

A minimal number of tag single nucleotide polymorphisms (SNP) were selected to represent comprehensive coverage of the *PPARA* gene using an estimated minimum pairwise correlation of 0.80 between SNPs. The haplotype tagging SNPs were identified

using the Program for Genomic Applications' LDSelect Program [190] run for 23 European-Americans Coriell samples (given the relative racial homogeneity of the LIBCSP population). This program has been shown to select a maximally informative set of common SNPs, and is based on the r^2 LD statistic [191]. Based on this program, 14 SNPs were identified for genotyping.

7.3.4 Laboratory Analysis

Genotyping was conducted at Columbia University, New York, NY. All LIBCSP DNA samples are available on 96 well master plates. Plates include an approximately 10% duplication rate, and laboratory personnel were blinded to case-control and duplication status. Genotyping was carried out using iPLEX technology (Sequenom, San Diego, CA) on a MassARRAY Compact Analyzer. For SNPs that could not be mulitplexed (rs4253623 and rs4253699), Taqman (Applied Biosystems, Foster City, CA). SNP assays were used. Six of the 14 identified tag SNPs for *PPARA* met an *a priori* minimum kappa statistic of 0.90 for concordance among the blinded repeat samples, and were included in subsequent analyses.

7.3.5 Statistical Analysis

Hardy-Weinberg Equilibrium (HWE) was calculated using exact p-values due to the low expected prevalence of rare homozygotes. Pairwise linkage disequilibrium (LD) [196] was determined using Haploview 4.0 [225]. Kaplan-Meier survival curves were created for all SNPs individually and *PPARA* haplotypes (data not shown). Cox regression was used to estimate crude and adjusted hazard ratios (HR) and 95% confidence intervals (CI) for all cause and breast cancer specific mortality. All analyses were conducted in SAS version 9.1.3 (Cary, N.C.).

Haplotype reconstruction was performed using an expectation maximization (EM) algorithm [199]. Haplotype frequencies were compared by vital status, and haplotype specific HRs and 95% CIs, using all other haplotypes as the reference, were calculated using Cox regression for all cause and breast cancer specific mortality. All haplotype analyses were conducted using SAS/Genetics version 9.1.3.

To assess the assumption of proportional hazards, both statistical testing and graphical techniques were used. First, log-minus-log survival curves were examined for each SNP to see if the resulting curves appear non-parallel. Second, categorical and continuous time interaction variables were tested using the likelihood ratio test (LRT). If the LRT yielded a p-value less than 0.05, the proportional hazards assumption was deemed violated and the time-dependent variable was retained in the model.

Race (White/non-White) and Jewish ethnicity (Jewish/not Jewish) were identified as potential confounders based on *a priori* knowledge and directed acyclic graphs (DAG) [230]; however, neither variable met a pre-specified 10% or greater change in effect estimate criterion for inclusion as confounders in final models.

7.4 RESULTS

Two SNPs, rs1800206 and rs4253760, deviated from HWE (p=0.03, p=0.01, respectively) when examined in the cases. For both deviations, excesses of common and rare homozygotes were observed compared to expected numbers under HWE. We found very low correlation between the *PPARA* polymorphisms (Figure 7.1), suggesting the PGA population is an appropriate reference for tag SNP selection in the LIBCSP.

No differences in survival were evident for all *PPARA* SNPs evaluated individually (Table 7.1). There is a suggestion of decreased survival for rs4253760 at baseline, the only SNP that required the inclusion of a continuous time interaction term (HR=2.25; 95% CI: 1.00, 5.08 for overall mortality at time=0, or at diagnosis) in both the all cause and breast

cancer-specific mortality analyses. In this model, however, the time interaction beta-coefficient is negative, implying survival improves with each day after baseline. Therefore, five years following diagnosis, the rs4253760 HR will decrease from 2.25 to 0.69 (95% CI: 0.36, 1.29). Thus, caution is necessary when interpreting the results for this polymorphism.

The haplotype reconstruction created 12 haplotypes with a frequency greater than 0.01. No haplotypes were clearly associated with survival in this cohort (Table 7.2). Due to the low prevalence of select haplotypes, several estimates lacked precision. For example, haplotype 10, occurring in 1% of alive and <1% of dead cases, had a confidence interval width of 721, measured by dividing the upper by the lower interval. Results were similar for breast cancer-specific mortality although the imprecision of the effect estimates makes interpretation difficult (data not shown).

7.5 Discussion

We found *PPARA* polymorphism rs4253760 was associated with all cause and breast cancer specific mortality at baseline when the hazard model was correctly specified by including a continuous time interaction variable. The inclusion of the interaction term also implies that survival is improving over time for carriers of the rare allele; therefore, this HR should be interpreted with caution. This association is consistent with other LIBCSP results implicating rs4253760 in breast cancer development [231]. Additionally, the stronger association found with breast cancer-related deaths compared to all cause mortality suggest that the gene's known connection to cardiovascular disease risk factors is not driving our results. This SNP also tags 10 other SNPs with a frequency greater than 10% according to the PGA and may act as a marker for the true causative polymorphism although the haplotype analyses are not supportive of a causal effect.

Caution is necessary in interpreting our findings, given that estimates were based on small numbers of observations. Further, the SNP was not in HWE although recent research

has suggested that deviations from HWE in the cases may reflect evolutionary survival mechanisms rather than laboratory error; therefore, these deviations may not bias the resulting effect estimate [195].

Approximately 73.0% of cases donated blood, suggesting that missing data may be concern. However, in a comparison of select tumor and demographic characteristics, including age, stage and tumor size, no differences between genotyped cases (n=1,073) and all cases (n=1,508) were found (data not shown).

The number of SNPs included in our study was limited because of the low concordance of some of the genotyped SNPs with their blinded repeats. Out of the 14 tag SNPs identified, only six polymorphisms that achieved the *a priori* kappa statistic cut-off of 0.90 were included in our analyses; however, no studies to date have investigated the link between breast cancer survival and any *PPARA* SNPs.

In conclusion, while no current research has examined the association between breast cancer survival and *PPARA* polymorphisms, the biology and epidemiology of the gene suggest it could play a role in disease prognosis. PPARA has been linked to obesity, inflammation and insulin resistance, pathways which have also been implicated in breast cancer survival. One PPARA SNP, rs4253760, was inversely associated with survival in this unique cohort of 1,073 breast cancer cases from the LIBCSP.

7.6 Tables and Figures

TABLE 7.1. Hazard ratios and frequencies by vital status for six SNPs in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n=1073)

	Overall Mortality			Breast Cancer-Specific Mortality		
Genotype	**N Cases/Cohort**	**HR (95% CI)**	**HR* (95% CI)**	**N Cases/Cohort**	**HR (95% CI)**	**HR* (95% CI)**
rs135542						
AA	78/602	1.00	1.00	45/569	1.00	1.00
AG+GG	44/411	0.82	0.80	36/403	1.14	0.67
		(0.57, 1.19)	(0.34, 1.86)		(0.73, 1.76)	(0.25, 1.82)
rs1800206						
CC	115/927	1.00	1.00	76/888	1.00	1.00
CG+GG	10/107	0.76	1.17	8/105	0.91	1.54
		(0.40, 1.46)	(0.30, 4.58)		(0.44, 1.89)	(0.33, 7.06)
rs4263623						
AA	97/811	1.00	1.00	65/779	1.00	1.00
AG+GG	28/235	1.00	1.54	19/226	1.01	0.88
		(0.66, 1.52)	(0.65, 3.68)		(0.60, 1.68)	(0.28, 2.78)
rs4253699						
TT	75/624	1.00	1.00	51/600	1.00	1.00
CT+CC	52/416	1.05	1.16	35/399	1.04	1.06
		(0.74, 1.50)	(0.45, 2.96)		(0.68, 1.60)	(0.41, 2.76)
rs4253755						
GG	97/803	1.00	1.00	68/774	1.00	1.00
AG+AA	32/247	1.08	1.18	18/233	0.87	0.99
		(0.72, 1.61)	(0.54, 2.60)		(0.52, 1.47)	(0.32, 3.07)
rs4253760						
TT	80/675	1.00	1.00	57/652	1.00	1.00
GT+GG	44/345	1.10	2.25	25/326	0.89	2.50
		(0.76, 1.59)	(1.00, 5.08)		(0.55, 1.42)	(0.90, 6.96)

*Adjusted for a continuous time interaction term; interaction term necessary only for rs4253760 based on likelihood ratio test

TABLE 7.2. Hazard ratios and frequencies by vital status for haplotypes in relation to all-cause mortality (n=132) among a cohort of women with breast cancer (n=1073)

							All-cause Mortality		
Haplotype Number*	**rs135542**	**rs1800206**	**rs4253623**	**rs4253699**	**rs4253755**	**rs4253760**	**Alive N (%)**	**Dead N (%)**	**HR (95% CI)**
1	A	C	A	C	A	G	74 (5.6%)	14 (6.8%)	1.01 (0.36, 2.86)
2	A	C	A	C	A	T	41 (3.0%)	6 (3.1%)	0.93 (0.21, 4.14)
3	A	C	A	C	G	G	23 (1.7%)	0 (<1%)	0.31 (0.03, 3.76)
4	A	C	A	C	G	T	60 (4.6%)	16 (8.3%)	2.11 (0.67, 6.58)
5	A	C	A	T	A	G	37 (2.8%)	5 (2.6%)	0.90 (0.18, 4.53)
6	A	C	A	T	G	G	43 (3.3%)	11 (5.7%)	1.77 (0.57, 5.46)
7	A	C	A	T	G	T	520 (39.7%)	76 (38.3%)	0.79 (0.51, 1.23)
8	A	C	G	T	G	T	134 (10.2%)	24 (12.2%)	0.98 (0.45, 2.10)
9	A	G	A	C	G	G	32 (2.4%)	4 (2.3%)	0.77 (0.11, 5.21)
10	A	G	A	C	G	T	15 (1.2%)	0 (<1%)	0.19 (0.01, 7.21)
11	G	C	A	C	G	T	17 (1.3%)	2 (1.0%)	0.12 (0.00, 6.45)
12	G	C	A	T	G	T	241 (18.4%)	33 (16.5%)	0.69 (0.35, 1.36)

*haplotypes with frequency ≥ 0.01

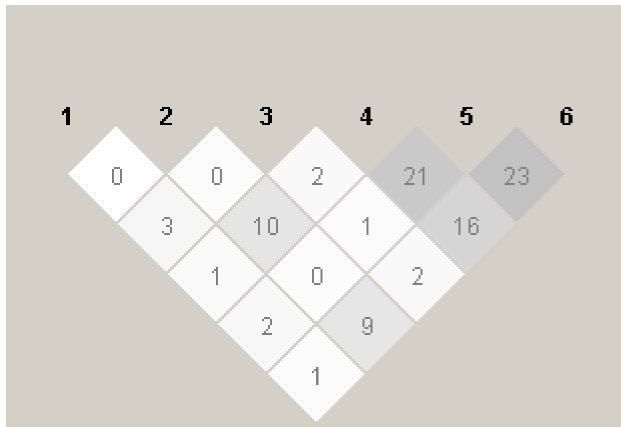


FIGURE 7.1. Linkage disequilibrium (r^2) between the six *PPARA* polymorphisms

CHAPTER 8 DISCUSSION

8.1 Specific Aims

The overarching goal of this doctoral research was to determine whether genetic variation in *PPARA* was associated with breast cancer risk and survival, utilizing a haplotype-based approach. To address this goal, four specific study aims were completed, as discussed below, utilizing population-based data from the Long Island Breast Cancer Study Project. Genotyping was completed for 1,073 women with breast cancer and 1,112 controls, for whom demographic, risk factor, and clinical data was available from personal interviews and medical record abstraction.

The first specific study aim was to examine the association between genetic variations in *PPARA* and risk of breast cancer in all women and post-menopausal women only. Three different analytic strategies were employed to meet this objective: (1) conventional logistic models were constructed for all SNPs separately and collectively under dominant inheritance and full genotype specification; (2) semi-Bayesian methods were used to model all polymorphisms jointly, and (3) haplotypes were created and analyzed. To be included in all three analytic strategies, SNPs were required to have a minimum kappa statistic of 0.90 with their replicates. Based on this *a priori* cut-off, six of the fourteen genotyped polymorphisms were included in all analyses.

Few differences in precision and estimate effects were observed between the conventional and semi-Bayesian results under dominant inheritance, although the fully specified genotype models showed substantial pull towards the null value of one in the

Bayesian models. These findings are consistent with expected results as the specified residual prior variance (τ^2) will place more weight on the null value than the data. rs4253760 was associated breast cancer incidence among post-menopausal women in conventional and semi-Bayesian models. This polymorphism is located in the fifth intronic region of *PPARA* and “binned” ten other SNPs with MAFs of 0.10 or greater. The haplotype analyses were constrained by small cell sizes and their imprecision makes interpretation difficult.

The second specific aim was to evaluate interaction by *PPARA* polymorphisms with the association between obesity, measured using body mass index (BMI), and weight gain and risk of breast cancer in all women and post-menopausal women only. Using conventional logistic regression and semi-Bayesian techniques, this research examined additive interaction by three different variables: (1) BMI at reference, (2) weight gain since age 50, and (3) aspirin use. BMI was analyzed as a three level categorical variable (normal weight, overweight and obese) based on the definitions from the World Health Organization [5]. Weight gain was also analyzed as a three level variable and classification was based on levels used in Eng *et al.* [9] which used this population. All weight gain models were adjusted for BMI at age 50 to better distinguish the impact of post-menopausal weight gain on disease risk apart from weight at baseline. Additive interaction was deemed present if the 95% confidence intervals for the interaction contrast ratios did not included the null value of zero.

Based on this criterion, no interaction was found with BMI and weight gain, but some evidence of interaction was found with aspirin use in two polymorphisms (rs135542 and rs4253699) in analyses including all women. The magnitude of interaction was consistent across all six polymorphisms (with the exception of rs1800206); however, there was a suggestion of differing functionality between rs135542 and the other four SNPs. For rs135542, there appears to be a weak inverse association between NSAID use and breast cancer among common homozygotes. Non-users with the “at risk” allele show a weak

positive association while users with the “at risk” allele have a strong inverse association with breast cancer. Thus, the “at risk” allele and NSAID use may work synergistically to exacerbate the protective effect of NSAIDs for rs135542.

For rs4253623, rs4253699, rs4253755 and rs4253760, the relationship between breast cancer and NSAID use is more clear. Among common homozygotes, NSAID use has a protective effect, but there is no association with NSAID use among those with the “at risk” allele. These results are consistent with NSAIDs operating through mechanisms involving *PPARA* in part.

The third specific aim was to examine the association between genetic variations in *PPARA* and breast cancer survival in pre- and post-menopausal cases. To meet this aim, two different analytic techniques were conducted under the dominant inheritance models: (1) Cox proportional hazard models were performed, including time interactions when necessary, and (2) haplotypes were constructed and analyzed. Both all cause and breast cancer-specific mortality were explored. Like the case-control analysis, rs4253760 showed an increase in both all cause and breast cancer-specific mortality. This suggests that cardiovascular disease risk factors, which have been shown to be associated with *PPARA*, are not driving the poor prognosis found for rs4253760. The haplotype analyses did not further elucidate survival relationships.

Lastly, the study aimed to assess interaction by polymorphic variation in *PPARA* with the association between obesity and weight gain and survival from breast cancer in pre- and post-menopausal women. This objective was underpowered due to the high survival rates found in this population (87.7% of cases were alive as of 2002) and the low minor allele frequencies (MAF) for several polymorphisms; therefore, this aim was exploratory only. The survival rates in LIBCSP are consistent with those found in Surveillance Epidemiology and End Results (SEER), which reported five year relative survival rate of 88.6% from 1996-2003 [232].

In addition to obesity and weight gain, additive and multiplicative interactions with several tumor and demographic characteristics were also examined, including patient age, stage, hormone receptor status and treatment. Stage acted an additive and multiplicative effect modifier for rs4253699 while the effect of rs4253755 was modified by radiation treatment. It is unclear whether these findings reflect a true association, or if instead they are due to biases associated with restricted power.

8.2 STRENGTHS AND LIMITATIONS

8.2.1 Tag SNP selection and genotyping

Tag SNPs were selected using the PGA IdSelect Program [190] run for European Americans only because of the racial homogeneity of the LIBCSP population. Therefore, the fourteen tag SNPs identified to provide maximum coverage of *PPARA* are applicable only to Caucasians. The number of polymorphisms necessary to capture the gene for African Americans (AA) is much larger (45 SNPs) due to the greater diversity in their population history compared to Caucasians. In this study, both races were included in analyses since AAs were not substantial enough to influence results yet added additional statistical power. For example, Table 8.1 presents the observed MAFs for Caucasians, AAs and the combined group compared to the PGA individuals of European descent. As this table shows, the combined population frequencies reflect very little AA influence with the exception of rs4253760, where the increased frequency of common homozygotes drives up the MAF slightly in the combined group.

The statistical analyses were limited to six of the fourteen SNPs. The eight other polymorphisms were purposely omitted because of the poor concordance between the blinded repeats and the replicates (kappa statistic < 0.90). Possible explanations for this low concordance include poor DNA quality, genotyping error and the location of the SNPs. Poor

DNA quality seems unlikely given the success of the Taqman assays in genotyping three of the SNPs. Genotyping error is a distinct possibility considering initial concerns with the laser intensity. Lastly, the location of the polymorphisms could have contributed to the low concordance. Five of the requested SNPs were found in repeat regions of the *PPARA* gene and, therefore, would have been difficult to genotype successfully.

TABLE 8.1. Minor allele frequencies by race for LIBCSP versus PGA

rs#	**LIBCSP Combined**	**LIBCSP Caucasian**	**PGA Caucasian**	**LIBCSP AA**	**PGA AA**
All cases					
rs135542	0.224 (n=1069)	0.226 (n=988)	0.205	0.240 (n=50)	0.348
rs1800206	0.054 (n=1085)	0.054 (n=1004)	0.022	0.010 (n=48)	0.021
rs4253623	0.123(n=1102)	0.123 (n=1018)	0.109	0.127 (n=51)	0.021
rs4253699	0.220 (n=1092)	0.220 (n=1008)	0.182	0.294 (n=51)	0.312
rs4253755	0.120 (n=1092)	0.124 (n=1009)	0.130	0.069 (n=51)	0.042
rs4253760	0.186 (n=1061)	0.167 (n=982)	0.196	0.585 (n=47)	0.667
Post-menopausal women					
rs135542	0.220 (n=662)	0.221 (n=621)		0.269 (n=26)	
rs1800206	0.052 (n=679)	0.053 (n=638)		0.020 (n=25)	
rs4253623	0.132 (n=688)	0.132 (n=646)		0.115 (n=26)	
rs4253699	0.217 (n=683)	0.218 (n=641)		0.288 (n=26)	
rs4253755	0.113 (n=679)	0.118 (n=638)		0.038 (n=26)	
rs4253760	0.168 (n=652)	0.153 (n=614)		0.565 (n=23)	

Whether inclusion of only six *PPARA* polymorphisms in the analyses is sufficient to represent variability within the gene is an important consideration. A comparison with the linkage disequilibrium (LD) from the HapMap version B36 shows that five of the analyzed SNPs capture three out of the four haplotype blocks (Figure 8.1) among individuals of European descent. The excluded haplotype block (Block 1) is less than one kilobase in length and only includes two polymorphisms; thus, its exclusion only minimally impacts gene coverage. rs4253699 is not included in HapMap, but its location would be between rs7364220 and rs4253701. Based on HapMap data, therefore, the six SNPs included in this study appear to adequately cover the majority of PPARA despite the loss of eight tag SNPs.

By definition, tag SNPs should show minimal LD in the LIBCSP since they were selected to be maximally informative ($r^2 < 0.80$) using PGA data (with the exception of

rs1800206, which was forced into the program because of its functionality). Any polymorphisms with an r^2 greater than 0.80 would, therefore, reflect deviations from the PGA population and would create doubt on the applicability of using the PGA population as the reference population for tag SNP selection. LD in the LIBCSP is consistent with expected results with all polymorphisms showing low pair-wise correlation in both cases and controls. Therefore, the PGA population is an appropriate reference population for this study.

Among controls, all polymorphisms were in Hardy-Weinberg Equilibrium ($p>0.05$). For the survival analyses, however, two SNPs (rs1800206 and rs4253760) showed deviations from HWE with p-values of 0.03 and 0.01 using a permuted version of the exact test, respectively, in the cases. These deviations will be present in regions of association and suggest that underlying evolutionary selection factors could be influencing the distribution of these genotypes in this population [195, 233].

Lastly, approximately 71.2% and 71.5% of cases and controls respectively had DNA available for these analyses. Thus, it is possible our results may not be applicable to all women. However, in analyses in which we compared risk factors for breast cancer among all women in the study as compared with those among women who donated blood, similar results were found. Similarly, in a comparison of tumor characteristics and other clinical indicators for survival, little difference was noted between all cases and cases who donated blood. Thus, restricting analyses to those who donated blood does not appear to bias results but does influence the precision of the effect estimates.

8.2.2 Recall bias

Reported weight at age 50 and other past exposures may be subject to recall bias and non-differential misclassification since disease diagnosis has occurred before exposure ascertainment. Several studies have found a high degree of recall in weight over time [214-216] although it has been suggested that underweight women may overestimate their past

weight while overweight women may underestimate their past weight [216]. Additionally, aspirin use may be subject to misclassification if subjects were taking the drug periodically or used more than one NSAID [234]. Misclassification in these exposures is unlikely to bias the interaction models, however, since misclassification will not differ by genotype status. Additionally, the main gene effect analyses will not be affected by recall or diagnostic bias since laboratory personnel were blinded to case-control status and participants were unaware of their *PPARA* genotype. Thus, recall bias will minimally impact the results of this dissertation.

8.2.3 Study Power

For the case-control analyses, study power was sufficient to detect an association for both the main gene effect and the interaction models. Power was also adequate for the common haplotypes but limited for the haplotypes that occurred in less than 5% of the population. Similarly, interaction models for the functional SNP, rs1800206, were also limited because of its low prevalence in this population.

For the main gene effect survival analyses, study power was more than sufficient to detect an association. Study power was limited for the haplotype and effect modification survival analyses, however, and the hazard ratios for these analyses were extremely imprecise. Therefore, we consider the haplotype and interaction models in the survival analyses to be exploratory and it would be difficult to draw any conclusions based on these findings alone. Our results do suggest, however, that further investigation of *PPARA* in another larger cohort of breast cancer survivors is warranted.

The power constraints in the interaction survival models are a function of three different factors: (1) the low MAFs for these polymorphisms, (2) high breast cancer survival rates, and (3) the limited availability of select clinical features from the follow-up interview. The MAFs for the six polymorphisms ranged from 5.4% to 22.4%. When these

polymorphisms are then stratified by characteristics that occur in less than 25% of the population, several small cell sizes result. This imbalance is further compounded by the high overall survival rates observed in this cohort and in SEER [232]. As mentioned earlier, 132 cases were deceased in these analyses. When these cases are stratified by genotype and then by possible modifiers, cells for select variables, including treatment and stage, contain less than ten observations. Treatment data is only available for approximately 750 genotyped cases of which only 56 were deceased. While all genotyped cases have stage information, only seven *in situ* cases were deceased. Consequently, the *in situ* stratum was severely imbalanced and estimates were very imprecise. In summary, based on these three factors, only 43% power is obtained when the modifier is found in 25% of the population for the even the most common polymorphism in this cohort.

8.3 Future Directions

This dissertation implicated rs4253760 in both breast cancer development and survival. Although this SNP is located in the fifth intron of the gene, it is highly correlated with ten other polymorphisms, none of which have known functionality. Therefore, the biological relevance of this particular polymorphism has not been established. The consistency of these results for both the survival and case-control analyses is very provocative and deserves further investigation.

8.3.1 Replication

This dissertation is the first to examine the associations between *PPARA* and breast cancer incidence and survival. While our use of semi-Bayesian techniques reduces the probability of false positives, further replication is needed to confirm our findings. Genotyping error and limited power (particularly for the survival analyses) could be responsible for these results. Therefore, it is important to replicate these associations in a

different cohort with a larger sample size to ensure our findings reflect a true biological relationship. Further, given the poor performance of mass spectrometry in this population, it may be important to use different genotyping methods to improve concordance with blinded replicates, including Illumina genotyping. Lastly, this study only investigated single nucleotide polymorphisms; however, insertions or deletions could also play a role in disease development and may be important to include in future analyses.

8.3.2 Expansion to the PPAR family

Other members of the PPAR family may be involved in breast cancer incidence and survival. In particular, among the three PPAR genes, research has focused primarily on PPAR-gamma (*PPARG*), which is involved in adipocyte differentiation. *PPARG* agonists have demonstrated strong anti-diabetic effects, suggesting it is involved in insulin regulation [149]. In a meta-analysis of 16 family-based studies that examined the association between a *PPARG* polymorphism that results in a proline to alanine substitution at codon 12 ($Pro^{12}Ala$) and Type 2 diabetes, a genotype relative risk of 0.79 ($p<0.001$, no confidence intervals provided) for the Ala allele was observed [235]. *PPARG* may also hinder estrogen biosynthesis in human breast tissue by inhibiting expression of the gene encoding aromatase [236] and have anti-inflammatory effects although this is remains controversial [149].

Genetic variants of *PPARG* have been examined in association with thyroid, endometrial, prostate and colorectal cancers [237-242], but only one study has investigated its association with body mass and breast cancer incidence [243]. This case-control study was nested within the Nurses Health Study to examine the relationship between the $Pro^{12}Ala$ polymorphism, breast cancer and body weight in 725 cases and 953 controls. This polymorphism has been suggested to have a functional effect, leading to a decrease in DNA binding and transactivation activity. Weight gain since the age of 18 appeared to be the

most strongly associated with *PPARG* genotype, but no other weight gain intervals were presented. The authors report an odds ratio of 1.08 (95% CI: 0.85-1.38) for Ala allele carriers compared to non-carriers using conditional logistic regression and adjusting for matched factors (age, menopausal status, postmenopausal hormone use, date of blood draw, time of blood draw and fasting status).

*PPAR*G has also been implicated in breast cancer survival. In fact, a pilot study was recently published examining the impact of rosiglitazone, a drug therapy used in the treatment of diabetes that has been shown to target *PPARG* gene expression, in women with breast cancer [244]. Although the PPAR gamma agonist did not reduce tumor cell proliferation, the study highlights the interest in *PPARG* and breast cancer survival. Several other clinical trials have also been published examining PPAR gamma agonists in connection to prostate, breast and colorectal cancer.

As with many functional SNPs, the Pro12Ala polymorphism is relatively rare with minor allele frequencies of 7.5% and 6.8% in the HapMap CEU and ED Seattle SNP populations respectively. Therefore, statistical power is a concern. Given the interest in this gene and the fact it has already been genotyped successfully in the Long Island population, however, investigating this SNP in connection to breast cancer incidence and survival is the natural follow-up to this dissertation. As a second component to this analysis, gene-gene interaction could also be explored, looking at the functional *PPARA* SNP (rs1800206, L162V) in combination with the *PPARG* Pro12Ala SNP. Sample size would definitely be an issue for this interaction, but it is an intriguing exploratory analysis to consider.

Lastly, it would be interesting to explore all three PPAR genes (PPAR delta, gamma and alpha) together in a pathway-based multigenic approach. Recently, classification and regression tree (CART) analysis [245] has emerged as a tool for identifying individuals at high risk for disease development. This approach could be applied on a much larger scale to

include several other genes, such as leptin or insulin-like growth factor-1, on the inflammatory or insulin resistance pathways.

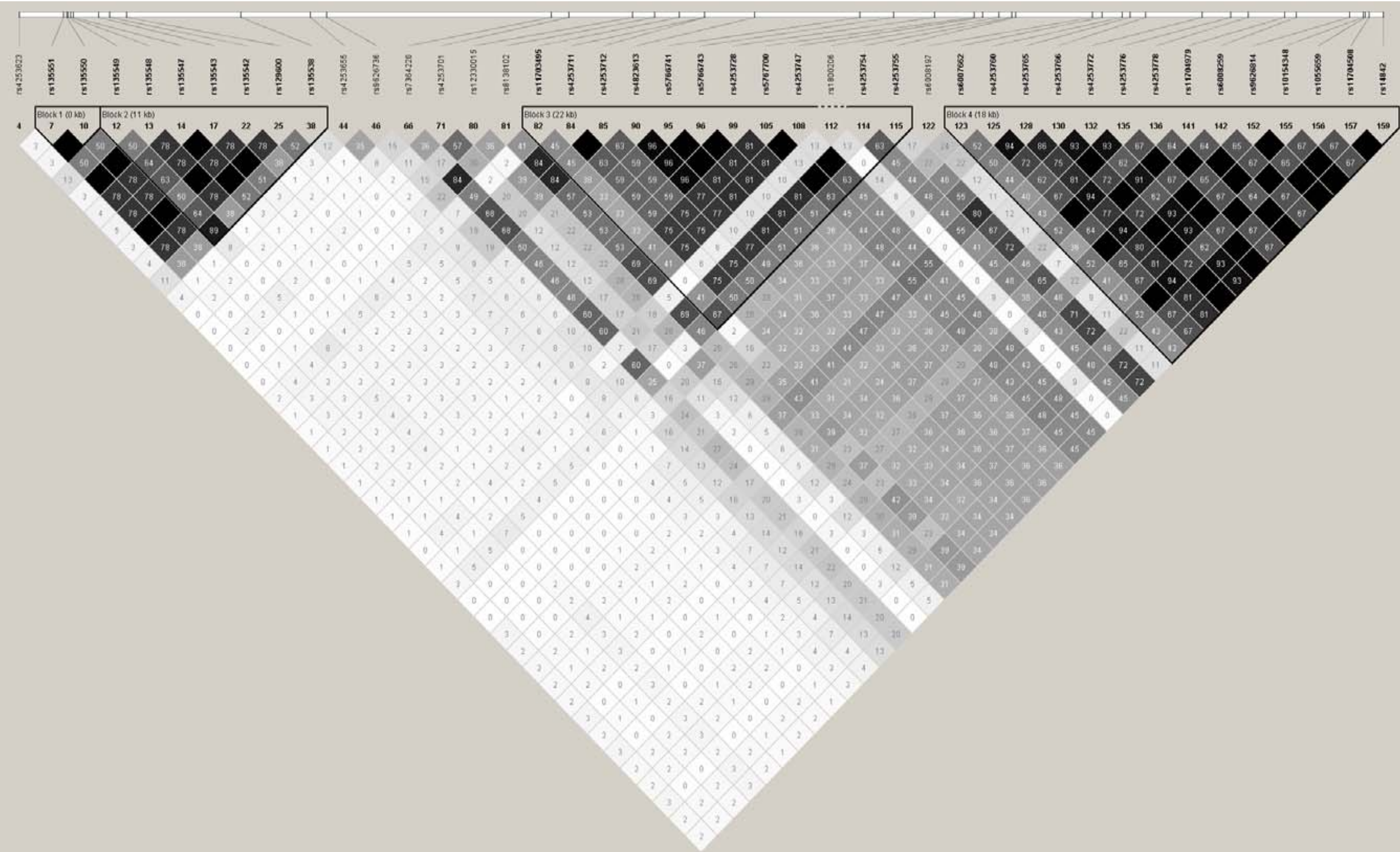
8.3.3 Survival

As discussed above, the major shortcoming of the survival analyses, particularly for the interaction models, was our limited study power. This limitation could perhaps be minimized by extending the follow-up time from five years to ten. Vital status for this dissertation was ascertained through December 31, 2002; however, additional follow-up could be undertaken through linkages with the National Death Index (NDI) to determine the long-term vital status of the cohort. The results of this additional follow-up time could potentially boost the statistical power of the survival analyses by increasing the number of deaths observed in the study. In fact, vital status information through the end of 2005 has already been requested from the NDI for this cohort, so that this direction could be explored.

8.4 SUMMARY

This study is the first to examine *PPARA* polymorphisms in connection to breast cancer incidence and survival. Interactions with BMI, NSAID use and post-menopausal weight gain were also explored. Our results suggest that *PPARA* does influence breast cancer development and prognosis. Replication in another cohort is warranted to confirm these findings. In particular, rs4253760 should be further investigated to better elucidate its role in disease development and progression.

FIGURE 8.1. *PPARA* linkage disequilibrium from HapMap version B36 in the CEU population



APPENDICES

APPENDIX A: TABLE OF PPAR-ALPHA STUDIES

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Shin MJ (2007) [177]	trial	609 Caucasians and 335 African Americans who participated in the Cholesterol and Atherosclerosis Pharmacogenetics Study, simuvastatin trial	▪ L162V ▪ rs4253728	▪ apo-CIII (μg/mL) ▪ Triglycerides (mg/dL) ▪ LDL (mg/dL) ▪ HDL (mg/dL)	▪ MAF=1.5% in AAs and 6.1% in Caucasians for L162V	▪ TG: 120.9±6.7 CC 139.7±13.2 GC/GG ▪ LDL: 124.5±3.8 CC 122.5±7.4 GC/GG ▪ HDL: 55.0±1.5 CC 59.6±2.9 GC/GG ▪ apoCIII: 12.7±0.5 CC 15.1±0.9 GC/GG	▪ age ▪ sex ▪ BMI ▪ Waist ▪ Smoking ▪ Alcohol ▪ Exercise ▪ Estrogen use ▪ Hyprtension ▪ Type II DM
Uthurralt J (2007) [175]	Cross-section	601 young Caucasian adults (mean age=24 years)	▪ L162V	▪ Triglycerides ▪ HDL ▪ LDL ▪ Fasting glucose ▪ Fasting insulin ▪ BMI	▪ MAF for L162V = 7.5%	▪ TG: 102.92±4.49 CC 146.80±11.84 CG ▪ HDL: 48.13±0.56 CC 44.67±1.48 CG ▪ LDL: 96 CC 98 CG ▪ BMI: 25.1 CC 29.5 CG in men ▪ Glu: 85.78±0.41 CC 83.65±1.03 CG in women	▪ age

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Manresa JM (2006) [178]	Cross-section	1748 participants (910 women, 838 men) between 9/94-3/96 in Spain	▪ L162V	▪ HDL (mg/dL) ▪ LDL (mg/dL) ▪ Triglycerides (mg/dL) ▪ BMI ▪ Type II DM	▪ MAF for L162V = 4.1% ▪ Mean (SD) ▪ HDL: 53.0 (15.1) CC 49.0 (15.7) GC/GG ▪ BMI: 26.4 (4.3) 27.1 (4.6) ▪ LDL: 150 (39) CC 150 (43) GC/GG ▪ TG: 93 (68, 127) CC 93 (73, 128) GC/GG ▪ DM: 144 (13.0%) CC 12 (12.6%) GC/GG		
Tai ES (2006) [179]	Trial	827 men from the Veterans Affairs HDL Intervention Trial, randomized to receive gemfibrozil (n=413, placebo; n=414, gemfibrozil)	▪ L162V	▪ LDL (mg/dL) ▪ HDL (mg/dL) ▪ Triglycerides (mg/dL) ▪ BMI ▪ Diabetes ▪ apoAI (mg/dL) ▪ apoB (mg/dL) ▪ Combined CVD endpts (stroke, non-fatal MI, CHD death)	▪ MAF for L162V = 6.8% ▪ Mean (SD) ▪ BMI: 29.1 (4.5) CC 29.8 (5.0) GC/GG ▪ LDL: 114 (22) CC 113 (23) GC/GG ▪ HDL: 31.8 (5.2) CC 30.7 (4.9) GC/GG ▪ apoAI: 107 (17) CC 104 (16) GC/GG ▪ apoB: 97 (20) CC 96 (20) GC/GG ▪ TG: 152 (86) CC 159 (89) GC/GG ▪ DM: 25% CC 37% GC/GG	▪ HR=1.63 (0.82, 3.24) for combined end points for V162 carriers vs. non-carriers among individuals with no DM or insulin resistance ▪ HR=0.40 (0.17, 0.92) for combined end points for V162 carriers vs. non-carriers among individuals with DM or insulin resistance	▪ Age ▪ Hypertension ▪ Smoking ▪ BMI ▪ Lipids ▪ Tmt group

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Doney AS (2005) [176]	Cohort	1810 Caucasian subjects with Type II DM from the Go-DARTS Study	▪ L162V ▪ G2528C	▪ BMI ▪ LDL (mmol/L) ▪ HDL (mmol/L) ▪ Triglycerides ▪ Non-fatal MI ▪ Age at diagnosis with Type II DM	▪ MAF=6.9% L162V ▪ BMI: 30.5 (30.2, 30.7) CC 30.7 (30.0, 31.4) CG ▪ LDL: 2.89 (2.84, 2.92) CC 2.89 (2.78, 3.00) CG ▪ HDL: 1.22 (1.20, 1.24) CC 1.23 (1.20, 1.28) CG ▪ Tri: 2.7 (2.6, 2.8) CC 2.8 (2.6, 3.1) CG	▪ HR=0.31 (0.10, 0.93) for non-fatal MI in V162 carriers vs. non-carriers ▪ $B_{insulin}$ = 2.6 (0.2, 5.1) for age at diagnosis with Type II DM in V162 carriers vs. non-carriers	▪ Smoking ▪ Gender ▪ Age ▪ Insulin tmt ▪ Prevalent angina ▪ Prevalent cerebro-vascular disease ▪ Prevalent MI
Flavell DM (2005) [152]	Cohort	912 Caucasian Type II diabetics, who are participating in the UDACS and EDSC studies in Great Britain. Study period not stated.	▪ L162V ▪ A/C Intron 1 ▪ G/C Intron 7	▪ Age of Onset ▪ Progression of DM2	▪ Individual variants not associated with outcomes ▪ 12.5% of the participants were carriers of the V162 allele ▪ age at dx: 52.7±0.5 CC 54.7±1.2 GC/GG	For C-L-C haplotype compared to A-L-G: ▪ -5.86±2.57 (p=0.02) effect on age at dx ▪ OR=3.75 (1.65, 8.56) for early age at dx	▪ Family history of diabetes ▪ Sex ▪ Smoking
Tai ES (2005) [161]	Cohort	2106 men and women who participated in the Framingham Offspring Study and went to their 5th exam from 1992-1995	▪ L162V	▪ Interaction between PUFA intake and L162V ▪ LDL, HDL, Triglycerides	▪ 13.7% of men and 13.1% of women were carriers of the V162 allele	▪ PPARA and PUFA intake interacted with plasma triglyceride levels (p=0.048) and apoC-III (p<0.01) ▪ No interactions were found for LDL-C, HDL-C, or total cholesterol	▪ Gender ▪ Age ▪ Family ▪ BMI ▪ Smoking ▪ Alcohol consumption ▪ Drug use ▪ Energy ▪ Total fat intake

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Chen S (2004) [153]	Trial	372 participants in the LCAS trial and received fluvastatin. Study location and time period were not provided.	▪ L162V ▪ A/C Intron 1	▪ Total cholesterol ▪ HDL-C ▪ LDL-C ▪ Triglyceride	▪ No associations found between PPARA and lipoprotein levels (data not presented)		
Foucher C (2004) [154]	Trial	155 participants in DAIS, a multinational study, who were randomized to receive fenofibrate and had the SNPs for the genes of interest	▪ L162V ▪ G/C Intron 7	▪ Reduction of triglyceride levels > 30% vs. ≤30%	▪ 92.9% vs. 91.4% of high vs. low responders to fenofibrate with the L/L allele (p>0.05) ▪ 84.7% vs. 68.8% of high vs. low responders with the G/G allele (p<0.05)	▪ OR=3.19 (1.28-7.98), comparing Intron 7 G/G vs. G/C+C/C ▪ No associations found with L162V (data not shown)	▪ Baseline TG level ▪ CEPT, apoE, LPL and LIPC SNPs ▪ Gender ▪ Age ▪ BMI ▪ Smoking ▪ Baseline HbA1c
Gouni-Berthold I (2004) [163]	Case-control	842 subjects recruited between 1999 and 2002 in Cologne; 404 subjects with DM-2 were matched with 438 non-diseased on sex and age.	▪ L162V	▪ Atherosclerosis ▪ Lipoprotein levels ▪ Obesity	▪ 9.41% vs. 11.42% with the V162 allele in cases vs. controls (p=0.34) ▪ *In DM patients:* ▪ BMI: 28.8±5.1 CC 27.8±4.9 GC/GG ▪ LDL: 140±46 CC 135±44 GC/GG ▪ HDL: 50±17 CC 51±18 GC/GG ▪ *In non-DM controls:* ▪ BMI: 26.5±3.7 CC 26.7±3.8 GC/GG ▪ LDL: 171±69 CC 175±51 GC/GG ▪ HDL: 60±17 CC 62±14 GC/GG	▪ OR=0.49 (0.21-1.15) for atherosclerosis in V162 carriers vs. non-carriers among patients with DM ▪ OR=0.44 (0.18-1.11) for CHD in V162 carriers vs. non-carriers among patients with DM ▪ OR=0.68 (0.20, 2.33) for atherosclerosis in V162 carriers vs. non-carriers among healthy controls ▪ OR=1.03 (0.12, 8.70) for CHD in V162 carriers vs. non-carriers among healthy controls	▪ CHD family history ▪ Age ▪ LDL/HDL ratio ▪ Hyprtension ▪ BMI ▪ Sex

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Robitaille J (2004) [164]	Cross-section	632 men recruited through the Chicoutimi Hospital Lipid Clinic in Canada. Study period not specified.	▪ L162V	▪ Metabolic Syndrome ▪ BMI ▪ LDL (mmol/L) ▪ HDL (mmol/L) ▪ Triglycerides (mmol/L) ▪ apoB (g/L) ▪ fasting glucose (mmol/L)	▪ Frequency of V162 allele was similar in diseased (n=281) vs. not diseased (n=351) ▪ 10.6% of subjects were V162 allele carriers ▪ BMI: 27.0±4.3 CC 26.8±4.1 GC/GG ▪ LDL: 3.84±1.01 CC 3.72±0.99 GC/GG ▪ HDL: 1.01±0.34 CC 0.97±0.35 GC/GG ▪ TG: 2.63±2.72 CC 3.26±3.02 GC/GG ▪ apoB: 1.15±0.25 CC 1.20±0.25 GC/GG ▪ glucose: 5.36±0.72 CC 5.39±0.75 GC/GG	▪ Carriers of V162 allele had higher triglyceride levels (p<0.0001) ▪ V162 had higher freq among men with abdominal obesity, hypertriglyceridemia, and low HDL-C simultaneously	▪ Age ▪ BMI
Bosse Y (2003) [165]	Cohort	663 subjects enrolled in the Quebec Family Study (241 nuclear families). Time period not stated.	▪ L162V	▪ Fasting glucose (mmol/L) ▪ Fasting insulin (pmol/L)	▪ 15% of the cohort are carriers of the V162 allele	▪ Geo mean (95% CI) ▪ Glucose: 5.19 (5.06, 5.31) CC 5.06 (4.86, 5.26) GC/GG ▪ Insulin: 63.1 (56.2, 70.9) CC 61.9 (50.2, 76.3) GC/GG	▪ Age ▪ Gender ▪ BMI

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Bosse Y (2003) [167]	Cohort	698 subjects enrolled in the Quebec Family Study (253 nuclear families). Time period not stated.	▪ L162V	▪ Obesity, BMI > 30 vs. BMI ≤ 30 ▪ % body fat ▪ BMI ▪ Weight (kg)	▪ 15% of the cohort are carriers of the V162 alleles ▪ OR=1.46 (p=0.15; no CI) of obesity (BMI > 30 vs. BMI ≤ 30) in V162 carriers vs. non-carriers ▪ *In women:* ▪ BMI: 28.1±8.6 CC 25.9±6.3 GC/GG ▪ Weight: 71.4±22.1 CC 65.6±15.3 GC/GG	▪ OR=1.77 (p=0.041, no CI) of obesity (BMI > 30 vs. BMI ≤ 30) in V162 carriers vs. non-carriers	▪ Age ▪ Gender ▪ Alcohol consumption
Brune S (2003) [155]	Case-control	104 AD cases were recruited from the University of Bonn hospital; 123 healthy controls were identified from the general population. Time period not stated.	▪ L162V ▪ G/C Intron 7	▪ Alzheimer Disease	▪ 14% vs. 7% with the V162 allele in cases vs. controls (p=0.03) ▪ 17% vs. 19% with the Intron 7 C allele in cases vs. controls (p=0.82)	▪ OR=2.24 (1.12-4.50) of AD, comparing presence vs. absence of V162 allele	▪ Sex ▪ Age
Bosse Y (2002) [166]	Trial	63 obese participants were randomized to receive placebo (n=31) or gemfibrozil (600 mg) (n=32) in Quebec, Canada. Time period not stated.	▪ L162V	▪ Triglycerides (mmol/L) ▪ LDL (mmol/L) ▪ HDL (mmol/L) At baseline	▪ 17% V162 allele frequency in all subjects ▪ TG: 2.66±0.80 CC 2.34±0.59 GC/GG ▪ LDL: 0.39±0.10 CC 0.34±0.04 GC/GG ▪ HDL: 0.23±0.05 CC 0.21±0.03 GC/GG		

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Brisson D (2002) [160]	Cohort	292 hypertriglyceridemic individuals who had underwent a 3-month therapy with fenofibrate in Quebec, Canada. No time period given.	▪ L162V	▪ post-tmt TG>2.0 mmol/l ▪ Plasma TC/HDL-C ratio > 5	▪ 27% of individuals in the cohort were carriers of the V162 allele	▪ OR=1.02 (0.55-1.90) of residual hypertrigly for carriers vs. non-carriers ▪ OR=1.19 (0.61-2.33) of TC/HDL-C>5 for carriers vs. non-carriers following fenofibrate treatment	▪ Age ▪ Gender ▪ TC/HDL-C before tmt
Eurlings P (2002) [157]	Case-control	102 FCHL probands and 124 spouses (controls) were recruited from two lipid clinics in the Netherlands. Study period not provided.	▪ L162V ▪ G/A Intron 2 ▪ G/C Intron 7	▪ Familial combined hyperlipidemia ▪ BMI ▪ HDL (mmol/L) ▪ Triglycerides (mmol/L) ▪ apoB (g/L)	▪ The V162 allele was present in 2% of cases and 4.8% of controls (p=0.10). No further analyses were conducted due to small numbers. ▪ *In FCHL patients:* ▪ BMI: 27.3±3.2 CC 30.1±3.5 GC/GG ▪ HDL: 0.91±0.25 CC 1.14±0.30 GC/GG ▪ TG: 4.35±8.27 CC 1.74±0.27 GC/GG ▪ apoB: 1.40±0.32 CC 1.23±0.61 GC/GG ▪ *In healthy spouses:* ▪ BMI: 25.2±3.9 CC 26.6±3.4 GC/GG ▪ HDL: 1.23±0.38 CC 1.12±0.42 GC/GG ▪ TG: 1.31±0.59 CC 1.47±0.73 GC/GG ▪ apoB: 0.99±0.25 CC 1.01±0.30 GC/GG		

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Flavell D (2002) [158]	Trial & Cohort	▪ 395 Finnish men who participated in LOCAT and were randomized to receive gemfibrozil to investigate its effect on atherosclerosis progression ▪ 3012 men who participated in the Second Northwick Park Heart Study in the United Kingdom (7 year study period) and examined risk of ischemic heart disease	▪ L162V ▪ G/C Intron 7	▪ Plasma lipid concentrations ▪ Atherosclerosis progression, measured as the change in average diameter of coronary segments (ΔADS) and change in minimum luminal diameter (ΔMLD) ▪ IHD risk ▪ Triglyceride (mmol/L) ▪ LDL (mmol/L) ▪ HDL (mmol/L) ▪ apoB	▪ For LOCAT: freq of V162 = 2.8% and Intron 7 C allele = 13.4% ▪ For NPHS2, freq of V162 = 6.3% and C allele = 17.4% ▪ No associations with lipid concentrations at baseline or following tmt with gemfibrozil ▪ For ΔADS, V162 carriers showed higher, positive diameter changes (p=0.022) while C allele carriers showed reduced diameter changes (p=0.095) ▪ For ΔMLD, V162 were protective (p=0.064) while C allele carriers showed greater decreases in MLD (p=0.003) ▪ *In LOCAT* ▪ TG: 1.82±0.73 CC 1.52±0.71 GC/GG ▪ LDL: 3.42±0.60 CC 3.41±0.54 GC/GG ▪ HDL: 1.02±0.17 CC 1.05±0.17 GC/GG ▪ *In NPHS2* ▪ TG: 1.80±0.98 CC 1.79±0.91 GC/GG ▪ HDL: 0.80±0.24 CC 0.80±0.25 GC/GG ▪ apoB: 0.88±0.28 CC 0.84±0.28 GC/GG	▪ HR=0.75 (0.45, 1.26) of IHD risk comparing V162 allele carriers to non-carriers ▪ HR=1.83 (0.96, 3.51) of IHD risk comparing Intron 7 C allele carriers vs. non-carriers	▪ Age ▪ BMI ▪ Cholesterol ▪ Fibrinogen ▪ Smoking ▪ Systolic BP

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Jamshidi Y (2002) [159]	Trial	A subset of 908 participants in the LEADER trial in Great Britain were genotyped. 608 received bezafibrate while 300 were a random subset from the placebo group. No time period specified.	▪ L162V ▪ G/A Intron 2 ▪ G/C Intron 7	▪ Triglycerides (mmol/L) ▪ Decrease in TG ▪ Decrease in Fibrinogen (among those who received active drug)	▪ MAF for V162 = 7% ▪ *In DM patients—mean (SD)* ▪ TG: 2.63 (1.35) CC 2.48 (1.07) GC/GG ▪ *In non-DM individuals* ▪ TG: 2.33 (1.14) CC 2.55 (1.15) GC/GG ▪ Decrease in TG among those who received bezafibrate: 0.58 (0.79) CC 0.82 (0.80) GC/GG ▪ Decrease in fibrinogen (g/L) among those who received bezafibrate: 0.40 (0.61) CC 0.40 (0.55)	▪ In the diabetic group (n=158), Intron 7 C allele carriers had lower baseline TG levels than non-carriers (2.26 vs. 2.83, p=0.048) ▪ No differences in baseline TG levels were evident between V162 allele carriers vs. non-carriers (2.48 vs. 2.63, p=0.77) ▪ In the non-diabetic group (n=654), no differences in baseline TG levels were evident in Intron 7 C allele carriers vs. non-carriers (2.42 vs. 2.33, p=0.20) ▪ V162 allele carriers were had higher TG levels (2.55 vs. 2.33, p=0.022)	▪ Age ▪ BMI
Nieters A (2002) [170]	Nested CCS	154 cases and 154 controls nested within the EPIC-Heidelberg cohort of 25,544 participants. Participants were recruited in 1998.	▪ L162V	▪ Obesity, measured as BMI $\geq$ 35 kg/m^2	▪ OR=1.35, no CI presented ▪ χ^2 p=0.45 for freq of L162 homozy vs. V162 carriers (11.1% in cases; 8.0% in controls)	▪ OR=1.45 (p=0.36, no CI presented)	▪ Sports activity ▪ Occ activity ▪ Daily hours of TV ▪ Energy intake

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Tai ES (2002) [162]	Cohort	2372 men and women who participated in the Framingham Offspring Study and went to their 5th exam from 1992-1995	▪ L162V	▪ LDL (mg/dL) ▪ HDL (mg/dL) ▪ apoB (g/L)	▪ V162 allele frequency in cohort was 6.9% ▪ *In women—mean (SD)* ▪ LDL:124 (1.07) CC 128 (3.01) GC/GG ▪ HDL: 58.5 (0.50) CC 57.1 (1.18) GC/GG ▪ apoB: 108 (0.95) CC 112 (2.88) GC/GG	▪ In men, the V162 allele was associated with higher levels of LDL (p<0.001), apoB (p<0.01) and apoC-III (p<0.01) but no differences in triglyceride levels were noted (p=0.19) ▪ In women, only apoB levels achieved statistical significance (p=0.03) although a trend of higher lipoprotein levels in carriers was evident.	▪ Familial relationships ▪ Age ▪ BMI ▪ Smoking ▪ Drug use
Evans D (2001) [168]	Cross-section	▪ 381 patients from a medical clinic in Hamburg, Germany ▪ 369 morbidly obese patients who are undergoing gastric banding surgery in Dinslaken ▪ 199 Blood Donors in Hamburg	▪ L162V	▪ Obese (BMI>30) ▪ DM2	▪ No significant differences in frequencies were noted for either obesity or DM2 ▪ 7% vs. 6% V162 allele freq in obese vs. non-obese individuals		

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Evans D (2001) [169]	Cross-section	▪ 370 morbidly obese patients who are undergoing gastric banding surgery in Dinslaken ▪ 199 Blood Donors in Hamburg ▪ 112 patients with mixed hyperlipoproteinamia ▪ 76 patients with hypercholesterolemia ▪ 154 patients with DM2 (all from Hamburg)	▪ L162V	▪ DM2 ▪ Morbid obesity ▪ Glucose < 126 mg/dl	▪ The allele freq was 6% in the cohort combined ▪ No difference in freq between DM2 patients and non-cases ▪ In women with DM ▪ BMI: 29±7 CC 24±3 GC/GG		

TABLE A.1. Table of epidemiologic literature on PPAR-alpha for all health outcomes

Author (Year)	Study Design	Study Population	SNPs	Outcome(s)	Unadjusted Association	Adjusted Association	Adjusted for confounders
Vohl MC (2000) [174]	Case-control	▪ 121 Caucasians with DM2 were age and sex-matched with non-diseased controls in Quebec, Canada in 1998 ▪ 193 healthy Caucasian men from Quebec were also investigated. No time period was specified.	▪ L162V	▪ BMI ▪ Fasting glucose (mmol/L) ▪ Fasting insulin (pmol/L) ▪ LDL (mmol/L) ▪ HDL (mmol/L) ▪ Triglycerides (mmol/L) ▪ apoB	▪ In the second sample, the V162 allele frequency was 6.6%. ▪ *Sample 1—DM&Controls* ▪ BMI: 28.9±4.1 CC 28.8±4.5 GC/GG ▪ Glucose: 5.89±1.61 CC 5.73±1.11 GC/GG ▪ Insulin: 134.5±115.7 CC 133.2±100.4 GC/GG ▪ LDL: 4.17±1.18 CC 4.55±1.19 GC/GG ▪ HDL: 0.98±0.35 CC 1.00±0.33 GC/GG ▪ TG: 3.58±3.23 CC 3.30±2.65 GC/GG ▪ apoB: 1.19±0.26 CC 1.28±0.27 GC/GG ▪ *Sample 2—DM* ▪ BMI: 29.9±4.3 CC 29.1±2.9 GC/GG ▪ Glucose: 5.41±0.56 CC 5.33±0.33 GC/GG ▪ Insulin: 109.9±83.7 CC 85.8±42.5 GC/GG ▪ LDL: 3.49±0.74 CC 3.86±0.56 CG/GG ▪ HDL: 0,93±0.20 CC 0.91±0.17 GC/GG ▪ TG: 2.17±1.16 CC 1.98±0.80 GC/GG ▪ apoB: 1.07±0.24 CC 1.18±0.16 GC/GG		

APPENDIX B: BREAST CANCER SURVIVAL AND OBESITY LITERATURE REVIEW

TABLE B.1. Breast cancer survival and obesity literature (2003-Present)

Author (Year)	Study Population	Follow-up	Obesity timepoint	Pre-menopausal Associations	Post-menopausal Associations	Adjusted for confounders*
Cleveland R (2007) [112]	1,508 cases in the Long Island Breast Cancer Study Project	Mean follow-up 66.7 years	BMI one year prior to diagnosis (BMI < 24.9 (ref), 25.0-29.9, ≥ 30)	▪ HR=2.62 (1.26, 5.45)* for overall survival in women with a BMI ≥ 30 compared to those with a BMI < 24.9	▪ HR=1.63 (1.08, 2.45)* for overall survival in pre- and post-menopausal women with a BMI ≥ 30 compared to those with a BMI < 24.9	▪ Age ▪ History of hypertension
Majed B (2007) [114]	Patients at Curie Institute 1981-99 for a total of 14,709 patients	Maximum follow-up of 20 years; median follow-up of 8 years	BMI at diagnosis (BMI < 24.9 (ref), 25.0-30, ≥ 30)	▪ HR=1.32 (1.22, 1.42)* for overall survival in pre- and post- menopausal women with a BMI between 25.0 and 29.9 compared to those with a BMI < 24.9 ▪ HR=1.53 (1.37, 1.72)* for overall survival in pre- and post- menopausal women with a BMI ≥ 30 compared to those with a BMI < 24.9		▪ Age ▪ Tumor dimension ▪ Node involvement ▪ Menopausal status ▪ Year diagnosed ▪ ER/PR status ▪ Tumor extension ▪ Scarf-Bloom Richardson Grade
Reeves GK (2007) [115]	1.2 million UK women recruited into the Million Women Study (1,179 pre-menopausal, 5,629 post-menopausal)	Median 7.0 years follow-up	BMI at baseline (BMI < 22.5, 22.5-24.9 (ref), 25.0-27.4, 27.5-29, ≥ 30)	▪ HR=0.91 (0.49, 1.70)* for overall survival in individuals with a BMI between 27.5 and 29.0 compared to those with a BMI between 18.5 and 24.9 ▪ HR=0.64 (0.34, 1.21)* for overall survival in individuals with a BMI ≥ 30 compared to those with a BMI between 18.5 and 24.9	▪ HR=1.22 (0.99, 1.49)* for overall survival in individuals with a BMI between 27.5 and 29.0 compared to those with a BMI between 18.5 and 24.9 ▪ HR=1.49 (1.27, 1.75)* for overall survival in individuals with a BMI ≥ 30 compared to those with a BMI between 18.5 and 24.9	▪ Age ▪ Geographic region ▪ SES ▪ Reproductive history ▪ Smoking ▪ Alcohol ▪ Physical activity

Author (Year)	Study Population	Follow-up	Obesity timepoint	Pre-menopausal Associations	Post-menopausal Associations	Adjusted for confounders*
Abrahamson PE (2006) [113]	1,254 women ages 20-54 diagnosed with invasive breast cancer between 1990-92 in Atlanta or New Jersey	8-10 years of follow-up	BMI at diagnosis (BMI <18.5, 18.5-24.9 (ref), 25.0-29.9, ≥ 30)	▪ HR=1.38 (1.04, 1.83)* for overall survival in individuals with a BMI between 25.0 and 29.9 compared to those with a BMI between 18.5 and 24.9 ▪ HR=1.65 (1.23, 2.21)* for overall survival in individuals with a BMI ≥ 30 versus those with a BMI between 18.5 and 24.9		▪ Income ▪ Stage
Kroenke CM (2005) [246]	5,204 participants (1062 pre-menopausal, 4073 post-menopausal) in the Nurses Health Study who were diagnosed with non-metastatic breast cancer from 1996-2000	Ranged from 2-26 years (median: 9 years)	BMI at baseline --Height in 1976 and weight reported closest but prior to diagnosis date	▪ RR=2.02 (1.13-3.61)* for breast cancer death in individuals with BMI ≥ 30 compared to BMI 21-22 (ref) ▪ RR=1.47 (0.94-2.32)* for breast cancer death in individuals with BMI between 25-29 compared to BMI 21-22	▪ RR=0.88 (0.61-1.28)* for breast cancer death in individuals with BMI ≥ 30 compared to BMI 21-22 (ref) ▪ RR=0.99 (0.74-1.33)* for breast cancer death in individuals with BMI between 25-29 compared to BMI 21-22	▪ Age ▪ Oral Contraceptive Use ▪ Birth index ▪ Age @ meno ▪ HRT use ▪ Protein intake ▪ Tumor size ▪ Nodal status ▪ Chemotherapy ▪ Tamoxifen use
Loi S (2005) [247]	1,101 participants (813 pre-menopausal, 288 post-menopausal) in the ABCFS study who were diagnosed with incident, primary breast cancer and lived in Sydney or	Ranged from 0.2-10.8 years (median: 5 years)	BMI 1 year prior to diagnosis	▪ HR=1.50 (1.00-2.26)* of distant recurrence in obese (BMI>30) compared to non-obese (BMI ≤ 30) individuals ▪ HR=1.71 (1.05-2.77)* of mortality from any cause in obese versus non-obese individuals	▪ HR=2.03 (0.99-4.21)* of distant recurrence in obese (BMI>30) compared to non-obese (BMI ≤ 30) individuals ▪ HR=0.84 (0.28-2.56)* of mortality from any cause in obese versus non-obese individuals	▪ Age ▪ Tumor grade ▪ Nodal Status ▪ Progesterone receptor status

Author (Year)	Study Population	Follow-up	Obesity timepoint	Pre-menopausal Associations	Post-menopausal Associations	Adjusted for confounders*
	Melbourne Australia.					
Whiteman MK (2005) [248]	3,924 participants (2551 pre-menopausal, 1373 post-menopausal) in the multicenter CASH study who were diagnosed with primary breast cancer from 12/1/80-12/31/82	Followed until 12/31/97 (median: 14.6 years)	Adult BMI --Usual weight as an adult	▪ HR=1.38 (1.05-1.80)* of breast cancer mortality in individuals with a BMI ≥ 30 compared to BMI < 23 ▪ HR=1.27 (1.05-1.52)* of breast cancer mortality in individuals with a BMI between 25-30 compared to BMI < 23	▪ HR=1.32 (0.94-1.83)* of breast cancer mortality in individuals with a BMI ≥ 30 compared to BMI < 23 ▪ HR=1.23 (0.97-1.57)* of breast cancer mortality in individuals with a BMI between 25-30 compared to BMI < 23	▪ Age @ diagnosis ▪ Race ▪ Radiation therapy ▪ History of benign breast disease ▪ Education ▪ Cancer stage
Berclaz G (2004) [249]	6,370 patients (3494 pre-menopausal, 2876 post-menopausal) who participated in the randomized clinical trials of the International breast cancer Study Group	Median: ~14 years	Not stated (most likely BMI at diagnosis)	▪ HR=1.16 (1.02-1.33)* of disease free survival in obese (BMI ≥ 30) versus normal weight (BMI < 25) individuals ▪ HR=1.06 (0.96-1.17)* of disease free survival in overweight (BMI between 25-30) versus normal weight (BMI< 25) individuals ▪ HR=1.22 (1.05-1.42)* of overall survival in obese versus normal weight individuals ▪ HR=1.11 (0.97-1.24)* of overall survival in overweight versus normal weight individuals	▪ HR=1.06 (0.94-1.20)* of disease free survival in obese (BMI ≥ 30) versus normal weight (BMI < 25) individuals ▪ HR=1.04 (0.94-1.16)* of disease free survival in overweight (BMI between 25-30) versus normal weight (BMI< 25) individuals ▪ HR=1.10 (0.96-1.26)* of overall survival in obese versus normal weight individuals ▪ HR=1.06 (0.94-1.19)* of overall survival in overweight versus normal weight individuals	▪ Nodal status ▪ Tumor size ▪ Vessel invasion ▪ ER status ▪ Progesterone receptor status ▪ Tumor grade ▪ Treatment Regimens

Author (Year)	Study Population	Follow-up	Obesity timepoint	Pre-menopausal Associations	Post-menopausal Associations	Adjusted for confounders*
Enger SM (2004) [121]	1,376 patients who were diagnosed with invasive breast cancer from 1/1/88 through 12/31/95 at the Kaiser Permanente Hospital in San Diego	Median: 6.8 years	Weight at diagnosis	▪ HR=1.60 (0.99-2.56) of breast cancer mortality in individuals who weigh over 175 lbs compared to those who weigh under 133 lbs. ▪ HR=1.41 (0.86-2.29) of breast cancer mortality in individuals weighing between 151-174 lbs compared to those who weigh < 133 lbs. ▪ Linear trend of death from breast cancer was evident among individuals with early stage disease with higher weight increasing the risk of death (p=0.03).		▪ Age ▪ Grade ▪ Stage ▪ Tumor size ▪ Lymph node status ▪ ER status
Enger SM (2004) [250]	717 pre-menopausal breast cancer patients who were diagnosed between 7/1/83 and 12/31/89 in Los Angeles County	Ranged from 0.3-15.1 years (median:10.4 years)	BMI and weight one year prior to diagnosis --weight change from age 18 to one year prior to diagnosis	▪ HR=0.76 (0.53-1.07)* of breast cancer death in women with BMI ≥ 24.9 compared to BMI < 20.4 ▪ HR=0.86 (0.60-1.23)* of breast cancer death in women who weighed 68.2 kg or greater compared to those who weighed less than 54.1 kg ▪ HR=0.93 (0.61-1.42)* of breast cancer death in women who gained 10 kg or more compared to women who gained no weight.		▪ Age ▪ Stage at diagnosis ▪ Physical activity ▪ Height (weight model only)

Author (Year)	Study Population	Follow-up	Obesity timepoint	Pre-menopausal Associations	Post-menopausal Associations	Adjusted for confounders*
Borugian MJ (2003) [251]	603 women (229 pre-menopausal, 357 post-menopausal) who were diagnosed with breast cancer in Vancouver between 7/1991 and 12/1992	Followed for up to 10 years	Waist-to-hip ratio at time of diagnosis	▪ RR=1.2 (0.4-3.4)* of breast cancer mortality in women in the 4th WHR quartile (WHR>0.848) compared to women in the lowest quartile (WHR<0.756) ▪ RR=0.6 (0.2-1.8)* of breast cancer mortality in women in the third WHR quartile (WHR 0.801-0.848) compared to women in the lowest quartile (WHR<0.756)	▪ RR=3.3 (1.1-10.4)* of breast cancer mortality in women in the 4th WHR quartile (WHR>0.848) compared to women in the lowest quartile (WHR<0.756) ▪ RR=2.3 (0.7-7.1)* of breast cancer mortality in women in the third WHR quartile (WHR 0.801-0.848) compared to women in the lowest quartile (WHR<0.756)	▪ Age ▪ BMI ▪ Family history of breast cancer ▪ ER status ▪ Stage at diagnosis ▪ Systemic treatment

APPENDIX C: COMPARISON OF TNM STAGING EDITIONS [252]

TABLE C.1. Comparison of TNM Staging Editions [252]

TNM 5th Edition	**TNM 6th EDITION**
Tis • Carcinoma in situ, includes intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor	Tis – carcinoma in situ • Tis (DCIS) • Tis (LCIS) • Tis (Paget's)
N2 • Metastasis to ipsilateral axillary lymph node(s) fixed to one another or to other structures	N2 • Metastasis in ipsilateral axillary lymph nodes fixed or matted, or in clinically apparent* ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis
N2a – excluded	N2a • Metastasis in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures
N2b – excluded	N2b • Metastasis only in clinically apparent* ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis
N3 • Metastasis to ipsilateral internal mammary node(s)	N3 • Metastasis in ipsilateral infraclavicular lymph node(s), with or without axillary lymph node involvement, or in clinically apparent* ipsilateral internal mammary lymph nodes in the presence of clinically evident axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s), with or without axillary or internal mammary lymph node involvement
N3a – excluded	N3a • Metastasis in ipsilateral infraclavicular lymph node(s)
N3b – excluded	N3b • Metastasis in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c – excluded	N3c • Metastasis in ipsilateral supraclavicular lymph node(s)
M1 • Distant metastasis, includes metastasis to ipsilateral supraclavicular lymph node(s)	M1 • Distant metastasis
Overall classification	*Overall classification*

TNM 5th Edition		TNM 6th EDITION	
Stage 0	Tis N0 M0	Stage 0	Tis N0 M0
Stage I	T1 N0 M0	Stage I	T1 N0 M0
Stage IIA	T0 N1 M0	Stage IIA	T0 N1 M0
	T1 N1 M0		T1 N1 M0
	T2 N0 M0		T2 N0 M0
Stage IIB	T2 N1 M0	Stage IIB	T2 N1 M0
	T3 N0 M0		T3 N0 M0
Stage IIIA	T0 N2 M0	Stage IIIA	T0 N2 M0
	T1 N2 M0		T1 N2 M0
	T2 N2 M0		T2 N2 M0
	T3 N1 M0		T3 N1 M0
	T3 N2 M0		T3 N2 M0
Stage IIIB	T4 Any N M0	Stage IIIB	T4 N0 M0
	Any T N3 M0		T4 N1 M0
Stage IV	Any T Any N M1		T4 N2 M0
		Stage IIIC	Any T N3 M0
		Stage IV	Any T Any N M1

* Clinically apparent is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination or grossly visible pathologically.

APPENDIX D: METHODS AND DATA ANALYSIS TABLES

TABLE D.1. Variables and variable definitions included in case-control and survival analyses

Analysis			Variable	
CCS	Survival	Topic	Name	Labels and Values
		Body Size		
	X		BMI20H20	BMI at age 20 using height at age 20
X	X		BMI50H20	BMI at age 50 using height at age 20
X	X		BMI50HRF	BMI at age 50 using height at reference
X	X		WTCH20CAT	Weight change from age 20 to reference age. 0= if weight loss is within 3.00 kg (reference) 1= if –44.91 to –3.01 kg weight loss 2= if 3.01 to 7.71 kg weight gain 3= if 7.72 to 8.15 kg weight gain 4= if 8.16 to 14.96 kg weight gain 5= if 14.97 to 87.09 kg weight gain
X	X		WTCH20	Continuous weight change from age 20 to reference age.
X	X		WTCH50_CAT	Weight change from age 50 to reference. 0= if no change in weight (reference) 1= if –68.04 to –0.04 kg weight loss 2= if 0.01 to 2.71 kg weight gain 3= if 2.72 to 4.98 kg weight gain 4= if 4.99 to 11.33 kg weight gain 5= if 11.34 to 62.14 kg weight gain
X	X		WTCH50	Continuous weight change from age 50 to reference.
		Reproductive		
X			AGEFB	Age at first live birth
	X		MENPSTAT	Menopausal status (derived from data on last menstrual period, oophorectomies/hysterectomies and other surgical information, pregnancy status, lactation status, and HRT use) 1 = Premenopausal 2 = Postmenopausal
		Demographic		
X	X		RACE	Self-identified race 1 = White 2 = Black 3 = Other (Asian, Native American, etc.)
X			RELIGION	Religion in which subject was raised 1 = None 2 = Protestant 3 = Catholic 4 = Jewish 5 = Other
		Medical history / Treatment		

Analysis		Topic	Variable Name	Labels and Values
CCS	Survival			
X			ASPNEW	Aspirin usage 0=non-users 1=users
	X		ERPRSTAT	Combined estrogen receptor and progesterone receptor status among cases 0= ER-/PR- 1= ER-/PR+ 2= ER+/PR- 3= ER+/PR+
	X		CHEMO_INT	Chemotherapy at F/U based on interview data 1=yes 2= no
X			FAMHX1	mother, sister, or daughter with breast cancer 0 = No FamHx of BrCa in Mother, Sister, or Daughter 1 = FamHx of BrCa in Mother, Sister, or Daughter
	X		HT_INT	Hormone Therapy at F/U based on interview data 1=yes 2= no
	X		RAD_INT	Radiation Therapy at F/U based on interview data 1=yes 2= no
	X		STAGE	stage of breast cancer 0 = Control 1 = In situ 2 = Invasive
		Matched variables		
X	X		AGEGRP	Five-year diagnosis age group for subject 1 = 20-24 2 = 25-29 3 = 30-34 4 = 35-39 5 = 40-44 6 = 45-49 7 = 50-54 8 = 55-59 9 = 60-64 10 = 65-69 11 = 70-74 12 = 75-79 13 = 80-84 14 = 85-89 15 = 90-94 16 = 95-100
		Outcome		

Analysis		Topic	Variable	Labels and Values
CCS	Survival		Name	
X			CASECONT	Case/control status 1 = Case 2 = Control
	X		DUR	Time-to-event variable
	X		NDI_2002	Vital status (based on all cause mortality) for all 1508 LIBCSP cases 1 = deceased 0 = alive
	X		NDI_BRCA	Indicator variable for breast cancer related death 0=alive 1=BRCA-related death 9=other causes of death
	X		NDI_PRIMARY	Vital status determined from the primary cause of death listed on the NDI records 0=alive 1= BRCA death 2=Lung CA death 3=Other CA death 4=CVD death 5=Other cause of death
	X		NDI_UNDER	Vital status determined from the underlying cause of death listed on the NDI records 0=alive 1= BRCA death 2=Lung CA death 3=Other CA death 4=CVD death 5=Other cause of death

TABLE D.2. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs135542 (AA vs. AG+GG) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	1.07 (0.50, 2.31)		
18.50-24.99	1.01 (0.75, 1.36)	0.95 (0.77, 1.19)	1.69 (3), 0.64
25.00-29.99	0.76 (0.50, 1.16)		
≥ 30.00	1.18 (0.60, 2.29)		
BMI at age 50 (kg/m^2)			
< 25.00	1.03 (0.78, 1.36)	0.96 (0.77, 1.20)	0.60 (1), 0.44
≥ 25.00	0.86 (0.61, 1.23)		
BMI at reference (kg/m^2)			
< 18.50	1.14 (0.30, 4.37)		
18.50-24.99	0.80 (0.57, 1.22)	0.95 (0.76, 1.18)	5.39 (3), 0.15
25.00-29.99	0.83 (0.56, 1.23)		
≥ 30.00	1.48 (0.95, 2.32)		
BMI at reference (kg/m^2)			
< 25.00	0.82 (0.59, 1.14)	0.95 (0.76, 1.18)	1.42 (1), 0.23
≥ 25.00	1.07 (0.80, 1.43)		
Weight change from 50 to reference date (6 levels)			
no change	0.80 (0.48, 1.34)		
-68.04 to -0.04 kg	0.65 (0.38, 1.13)		
0.01 to 2.71 kg	0.61 (0.33, 1.14)	0.93 (0.74, 1.17)	7.97 (5), 0.16
2.72 to 4.98 kg	1.11 (0.62, 1.37)		
4.99 to 11.33 kg	1.23 (0.70, 2.16)		
11.34 to 62.14 kg	1.54 (0.86, 2.74)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	0.73 (0.50, 1.05)		
0.01-4.98 kg	0.86 (0.56, 1.30)	0.94 (0.75, 1.18)	5.38 (2), 0.07
≥ 4.99 kg	1.37 (0.91, 2.04)		
Physical activity (hrs/wk)			
0	1.09 (0.72, 1.65)		
< 0.69	0.95 (0.60, 1.51)	0.93 (0.74, 1.17)	1.43 (3), 0.70
0.7-2.6	0.92 (0.57, 1.50)		
≥ 2.7	0.75 (0.46, 1.20)		
Jewish Ethnicity			

Jewish	1.30 (0.72, 2.36)	0.97 (0.78, 1.21)	1.11 (1), 0.29
Non-Jewish	0.92 (0.73, 1.17)		
Race			
White	0.96 (0.76, 1.20)	0.96 (0.77, 1.20)	0.08 (1), 0.77
Non-white	1.10 (0.44, 2.73)		
Aspirin Use			
Aspirin user	1.42 (0.90, 2.26)	0.99 (0.79, 1.24)	3.14 (1), 0.08
Non-user	0.88 (0.68, 1.14)		
Diabetes Diagnosis			
Yes	1.20 (0.59, 2.46)	0.96 (0.77, 1.19)	0.43 (1), 0.51
No	0.93 (0.74, 1.18)		

TABLE D.3. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs1800206 (CC vs. CG+GG) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	1.21 (0.31, 4.81)		
18.50-24.99	1.13 (0.71, 1.81)	1.04 (0.73, 1.47)	0.48 (3), 0.92
25.00-29.99	0.87 (0.44, 1.72)		
≥ 30.00	0.93 (0.35, 2.48)		
BMI at age 50 (kg/m^2)			
< 25.00	1.15 (0.74, 1.79)	1.04 (0.74, 1.47)	0.49 (1), 0.48
≥ 25.00	0.89 (0.51, 1.56)		
BMI at reference (kg/m^2)			
< 18.50	---		
18.50-24.99	0.81 (0.45, 1.46)	1.03 (0.72, 1.45)	3.15 (3), 0.37
25.00-29.99	1.37 (0.78, 2.40)		
≥ 30.00	1.00 (0.48, 2.09)		
BMI at reference (kg/m^2)			
< 25.00	0.77 (0.43, 1.36)	1.02 (0.72, 1.44)	1.49 (1), 0.22
≥ 25.00	1.21 (0.77, 1.88)		
Weight change from 50 to reference date (6 levels)			
no change	0.66 (0.27, 1.66)		
-68.04 to -0.04 kg	1.44 (0.57, 3.64)		
0.01 to 2.71 kg	0.63 (0.27, 1.48)		
2.72 to 4.98 kg	2.59 (0.63, 10.68)	0.99 (0.69, 1.43)	5.81 (5), 0.33
4.99 to 11.33 kg	0.80 (0.39, 1.65)		
11.34 to 62.14 kg	1.68 (0.62, 4.54)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	0.97 (0.51, 1.84)		
0.01-4.98 kg	0.97 (0.48, 1.95)	0.99 (0.69, 1.43)	0.02 (2), 0.99
≥ 4.99 kg	1.02 (0.58, 1.81)		
Physical activity (hrs/wk)			
0	1.61 (0.81, 3.22)		
< 0.69	0.70 (0.30, 1.61)	1.01 (0.71, 1.45)	5.05 (3), 0.17
0.7-2.6	1.38 (0.65, 2.93)		
≥ 2.7	0.63 (0.31, 1.26)		
Jewish Ethnicity			

Jewish	2.24 (0.74, 6.79)	1.04 (0.73, 1.47)	2.16 (1), 0.14
Non-Jewish	0.95 (0.66, 1.37)		
Race			
White	1.04 (0.73, 1.49)	1.04 (0.74, 1.47)	0.00 (1), 0.99
Non-white	1.03 (0.19, 5.42)		
Aspirin Use			
Aspirin user	0.97 (0.47, 1.99)	1.12 (0.78, 1.61)	0.22 (1), 0.64
Non-user	1.18 (0.77, 1.80)		
Diabetes Diagnosis			
Yes	1.33 (0.36, 4.96)	1.03 (0.73, 1.46)	0.16 (1), 0.69
No	1.01 (0.70, 1.45)		

TABLE D.4. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253623 (AA vs. AG+GG) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	1.28 (0.56, 2.94)		
18.50-24.99	0.88 (0.63, 1.24)	0.93 (0.72, 1.20)	0.98 (3), 0.81
25.00-29.99	0.86 (0.52, 1.41)		
≥ 30.00	1.11 (0.54, 2.28)		
BMI at age 50 (kg/m^2)			
< 25.00	0.92 (0.67, 1.26)	0.92 (0.72, 1.19)	0.00 (1), 0.98
≥ 25.00	0.93 (0.61, 1.40)		
BMI at reference (kg/m^2)			
< 18.50	1.78 (0.33, 9.48)		
18.50-24.99	0.93 (0.62, 1.38)	0.91 (0.71, 1.17)	0.71 (3), 0.87
25.00-29.99	0.89 (0.58, 1.38)		
≥ 30.00	0.85 (0.52, 1.40)		
BMI at reference (kg/m^2)			
< 25.00	0.96 (0.65, 1.42)	0.91 (0.71, 1.17)	0.12 (1), 0.73
≥ 25.00	0.88 (0.63, 1.22)		
Weight change from 50 to reference date (6 levels)			
no change	0.71 (0.38, 1.32)		
-68.04 to -0.04 kg	1.32 (0.72, 2.44)		
0.01 to 2.71 kg	1.14 (0.60, 2.15)	0.90 (0.70, 1.17)	4.20 (5), 0.52
2.72 to 4.98 kg	0.66 (0.33, 1.37)		
4.99 to 11.33 kg	0.70 (0.38, 1.28)		
11.34 to 62.14 kg	1.06 (0.53, 2.10)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	0.97 (0.63, 1.50)		
0.01-4.98 kg	0.90 (0.56, 1.45)	0.90 (0.70, 1.17)	0.24 (2), 0.89
≥ 4.99 kg	0.83 (0.53, 1.31)		
Physical activity (hrs/wk)			
0	1.04 (0.66, 1.66)		
< 0.69	0.92 (0.57, 1.50)	0.98 (0.76, 1.27)	0.95 (3), 0.81
0.7-2.6	1.20 (0.66, 2.19)		
≥ 2.7	0.82 (0.48, 1.42)		
Jewish Ethnicity			

Jewish	0.71 (0.37, 1.39)	0.92 (0.72, 1.18)	0.66 (1), 0.42
Non-Jewish	0.96 (0.73, 1.26)		
Race			
White	0.94 (0.73, 1.22)	0.92 (0.72, 1.19)	0.35 (1), 0.55
Non-white	0.68 (0.24, 1.92)		
Aspirin Use			
Aspirin user	1.24 (0.76, 2.03)	0.94 (0.72, 1.21)	1.72 (2), 0.19
Non-user	0.84 (0.62, 1.14)		
Diabetes Diagnosis			
Yes	0.62 (0.29, 1.33)	0.92 (0.72, 1.19)	1.19 (1), 0.28
No	0.97 (0.74, 1.26)		

TABLE D.5. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253699 (TT vs. CT+TT) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	0.75 (0.33, 1.69)		
18.50-24.99	1.20 (0.89, 1.62)	1.10 (0.89, 1.37)	2.21 (3), 0.53
25.00-29.99	0.95 (0.63, 1.42)		
≥ 30.00	1.39 (0.72, 2.70)		
BMI at age 50 (kg/m^2)			
< 25.00	1.14 (0.86, 1.51)	1.10 (0.89, 1.37)	0.12 (1), 0.73
≥ 25.00	1.05 (0.75, 1.49)		
BMI at reference (kg/m^2)			
< 18.50	1.63 (0.39, 6.82)		
18.50-24.99	1.09 (0.77, 1.53)	1.10 (0.88, 1.37)	0.89 (3), 0.83
25.00-29.99	0.99 (0.68, 1.45)		
≥ 30.00	1.24 (0.80, 1.94)		
BMI at reference (kg/m^2)			
< 25.00	1.11 (0.80, 1.55)	1.10 (0.88, 1.36)	0.01 (1), 0.92
≥ 25.00	1.09 (0.81, 1.45)		
Weight change from 50 to reference date (6 levels)			
no change	0.89 (0.54, 1.48)		
-68.04 to -0.04 kg	1.01 (0.58, 1.76)		
0.01 to 2.71 kg	0.85 (0.48, 1.50)		
2.72 to 4.98 kg	1.31 (0.73, 2.34)	1.13 (0.90, 1.42)	4.44 (5), 0.49
4.99 to 11.33 kg	1.38 (0.80, 2.40)		
11.34 to 62.14 kg	1.67 (0.93, 2.99)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	0.94 (0.65, 1.37)		
0.01-4.98 kg	1.05 (0.70, 1.58)	1.14 (0.91, 1.42)	3.07 (2), 0.22
≥ 4.99 kg	1.51 (1.01, 2.26)		
Physical activity (hrs/wk)			
0	1.64 (1.09, 2.46)		
< 0.69	0.75 (0.47, 1.19)	1.08 (0.86, 1.35)	7.88 (3), 0.05
0.7-2.6	1.18 (0.73, 1.91)		
≥ 2.7	0.82 (0.51, 1.32)		
Jewish Ethnicity			

Jewish	0.97 (0.57, 1.67)	1.11 (0.89, 1.38)	0.26 (1), 0.61
Non-Jewish	1.14 (0.90, 1.44)		
Race			
White	1.10 (0.88, 1.38)	1.10 (0.89, 1.37)	0.00 (1), 0.96
Non-white	1.08 (0.44, 2.65)		
Aspirin Use			
Aspirin user	1.56 (1.00, 2.45)	1.06 (0.85, 1.32)	3.84 (1), 0.05
Non-user	0.93 (0.72, 1.21)		
Diabetes Diagnosis			
Yes	1.05 (0.51, 2.13)	1.11 (0.89, 1.37)	0.02 (1), 0.88
No	1.11 (0.88, 1.40)		

TABLE D.6. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253755 (GG vs. AG+AA) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	1.50 (0.63, 3.56)		
18.50-24.99	1.19 (0.84, 1.69)	1.16 (0.90, 1.50)	1.73 (3), 0.63
25.00-29.99	1.23 (0.75, 2.02)		
≥ 30.00	0.75 (0.36, 1.58)		
BMI at age 50 (kg/m^2)			
< 25.00	1.22 (0.88, 1.68)	1.15 (0.90, 1.49)	0.29 (1), 0.59
≥ 25.00	1.06 (0.70, 1.59)		
BMI at reference (kg/m^2)			
< 18.50	2.42 (0.48, 12.30)		
18.50-24.99	1.16 (0.79, 1.73)	1.15 (0.89, 1.49)	1.09 (3), 0.78
25.00-29.99	1.19 (0.75, 1.89)		
≥ 30.00	1.02 (0.61, 1.68)		
BMI at reference (kg/m^2)			
< 25.00	1.21 (0.83, 1.78)	1.16 (0.90, 1.49)	0.11 (1), 0.74
≥ 25.00	1.11 (0.79, 1.57)		
Weight change from 50 to reference date (6 levels)			
no change	1.25 (0.67, 2.35)		
-68.04 to -0.04 kg	1.47 (0.77, 2.79)		
0.01 to 2.71 kg	0.74 (0.38, 1.45)	1.15 (0.88, 1.51)	2.58 (5), 0.76
2.72 to 4.98 kg	1.32 (0.68, 2.59)		
4.99 to 11.33 kg	1.21 (0.65, 2.28)		
11.34 to 62.14 kg	1.03 (0.51, 2.07)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	1.35 (0.87, 2.12)		
0.01-4.98 kg	0.99 (0.62, 1.59)	1.15 (0.88, 1.50)	0.92 (2), 0.63
≥ 4.99 kg	1.12 (0.70, 1.78)		
Physical activity (hrs/wk)			
0	1.34 (0.83, 2.16)		
< 0.69	0.57 (0.33, 0.98)	1.06 (0.82, 1.38)	7.12 3, 0.07
0.7-2.6	1.42 (0.80, 2.54)		
≥ 2.7	1.15 (0.67, 1.97)		
Jewish Ethnicity			

Jewish	1.16 (0.58, 2.35)	1.15 (0.89, 1.49)	0.00 (1), 0.98
Non-Jewish	1.15 (0.88, 1.51)		
Race			
White	1.15 (0.89, 1.49)	1.15 (0.89, 1.48)	0.02 (1), 0.90
Non-white	1.03 (0.19, 5.42)		
Aspirin Use			
Aspirin user	1.46 (0.88, 2.45)	1.12 (0.86, 1.46)	1.39 (1), 0.24
Non-user	1.02 (0.75, 1.39)		
Diabetes Diagnosis			
Yes	0.94 (0.40, 2.19)	1.15 (0.89, 1.49)	0.25 (1), 0.62
No	1.18 (0.90, 1.54)		

TABLE D.7. ORs across levels for select covariates with adjusted OR and test for homogeneity for rs4253760 (TT vs. GT+GG) in post-menopausal women

Covariate	Stratum-specific OR (95% CI)	Mantel-Haenszel OR (95% CI)	X^2 Breslow-Day test for homogeneity of the OR (df), p-value
BMI at age 50 (kg/m^2)			
< 18.50	1.34 (0.58, 3.09)		
18.50-24.99	1.27 (0.93, 1.75)	1.23 (0.98, 1.55)	0.34 (3), 0.95
25.00-29.99	1.21 (0.78, 1.87)		
≥ 30.00	1.04 (0.53, 2.04)		
BMI at age 50 (kg/m^2)			
< 25.00	1.29 (0.96, 1.73)	1.23 (0.98, 1.55)	0.21 (1), 0.65
≥ 25.00	1.15 (0.80, 1.66)		
BMI at reference (kg/m^2)			
< 18.50	0.72 (0.16, 3.20)		
18.50-24.99	1.26 (0.88, 1.81)	1.22 (0.97, 1.54)	2.16 (3), 0.54
25.00-29.99	1.46 (0.97, 2.19)		
≥ 30.00	0.97 (0.61, 1.55)		
BMI at reference (kg/m^2)			
< 25.00	1.22 (0.86, 1.74)	1.22 (0.97, 1.54)	0.00 (1), 0.99
≥ 25.00	1.22 (0.90, 1.66)		
Weight change from 50 to reference date (6 levels)			
no change	1.38 (0.81, 2.34)		
-68.04 to -0.04 kg	1.29 (0.73, 2.29)		
0.01 to 2.71 kg	0.78 (0.42, 1.43)		
2.72 to 4.98 kg	1.45 (0.76, 2.76)	1.22 (0.96, 1.55)	2.71 (5), 0.74
4.99 to 11.33 kg	1.34 (0.75, 2.38)		
11.34 to 62.14 kg	1.20 (0.65, 2.24)		
Weight gain from 50 to reference date (3 levels)			
No change or loss	1.34 (0.91, 1.97)		
0.01-4.98 kg	1.05 (0.68, 1.63)	1.22 (0.96, 1.56)	0.68 (2), 0.71
≥ 4.99 kg	1.27 (0.83, 1.93)		
Physical activity (hrs/wk)			
0	1.39 (0.90, 2.15)		
< 0.69	0.95 (0.59, 1.55)	1.20 (0.94, 1.52)	1.89 (3), 0.60
0.7-2.6	1.07 (0.64, 1.79)		
≥ 2.7	1.40 (0.85, 2.29)		

Jewish Ethnicity			
Jewish	1.56 (0.82, 2.97)	1.23 (0.98, 1.55)	0.60 (1), 0.44
Non-Jewish	1.19 (0.93, 1.53)		
Race			
White	1.27 (0.99, 1.61)	1.24 (0.98, 1.56)	0.60 (1), 0.44
Non-white	0.87 (0.34, 2.19)		
Aspirin Use			
Aspirin user	1.29 (0.81, 2.04)	1.24 (0.97, 1.57)	0.04 (1), 0.85
Non-user	1.22 (0.92, 1.61)		
Diabetes Diagnosis			
Yes	1.49 (0.70, 3.16)	1.23 (0.98, 1.55)	0.27 (1), 0.60
No	1.21 (0.95, 1.54)		

TABLE D.8. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, ΔW interaction & BMI

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1,196	1,196	1,196	1,198	1,239	1,239
Main Exposure: **rs135542**	0.2788 (0.1964)	0.0810 (0.1205)	0.0828 (0.1200)	0.0850 (0.1199)	0.0933 (0.1176)	0.0901 (0.1174)
INTERCEPT	-1.4217 (0.3619)	-1.3530 (0.3554)	-1.3606 (0.3520)	-1.3823 (0.3489)	-1.3192 (0.3427)	-1.1875 (0.3329)
WTCH_1	0.00910 (0.1873)	-0.0270 (0.1439)	-0.0274 (0.1438)	-0.0269 (0.1436)	-0.0390 (0.1414)	-0.0291 (0.1411)
WTCH_2	0.4699 (0.2038)	0.2541 (0.1661)	0.2556 (0.1658)	0.2486 (0.1653)	0.2671 (0.1616)	0.4055 (0.1395)
RACE_DIC	-0.0427 (0.2550)	-0.0580 (0.2545)	-0.0540 (0.2532)	REMOVED	---	---
JEWISH	-0.0150 (0.1636)	-0.0253 (0.1632)	REMOVED	---	---	---
FAMHX1	0.1122 (0.1582)	0.1097 (0.1579)	0.1090 (0.1578)	0.1173 (0.1575)	REMOVED	---
BMI50	0.1600 (0.0885)	0.1597 (0.0883)	0.1598 (0.0883)	0.1613 (0.0876)	0.1451 (0.0856)	REMOVED
AGEGRP	0.1141 (0.0350)	0.1160 (0.0349)	0.1162 (0.0349)	0.1184 (0.0347)	0.1165 (0.0340)	0.1100 (0.0337)
Interactions:						
WTCH1SNP1	-0.0799 (0.2919)	REMOVED	---	---	---	---
WTCH2SNP1	-0.5416 (0.2883)	REMOVED	---	---	---	---
-2 log L	1624.204	1628.158	1628.182	1630.803	1688.051	1690.924
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1	Removed BMI50
Test of interaction term		χ^2 (2): 1624.204–1628.158= 3.954; p=0.14	---	---	---	---
Change in estimate			|0.0810-0.0828| = 0.0018	|0.0810-0.0850| = 0.0040	|0.0810-0.0933| = 0.0123	|0.0810 - 0.0901| = 0.0091

Compare to:	Full Model	Model 2	Model 2	Model 2	Model 2
Conclusion:	Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. BMI is not a confounder and is removed from the model

TABLE D.9. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, BMI interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1,211	1,211	1,211	1,211	1,253	1,331
Main Exposure: **rs135542**	0.1697 (0.1802)	0.0437 (0.1201)	0.1694 (0.1793)	0.1706 (0.1791)	0.1981 (0.1765)	0.2199 (0.1704)
INTERCEPT	-1.3218 (0.3175)	-1.2462 (0.3114)	-1.3209 (0.3125)	-1.3282 (0.3102)	-1.3547 (0.3061)	-1.3479 (0.2964)
BMIREF_1	0.2607 (0.1745)	0.2784 (0.1358)	0.2606 (0.1744)	0.2605 (0.1744)	0.2446 (0.1727)	0.2445 (0.1671)
BMIREF_2	0.6447 (0.1966)	0.3958 (0.1490)	0.6445 (0.1961)	0.6511 (0.1958)	0.6296 (0.1913)	0.6903 (0.1839)
RACE_DIC	-0.0349 (0.2666)	-0.0467 (0.2654)	-0.0353 (0.2654)	REMOVED	---	---
JEWISH	0.00254 (0.1622)	-0.0146 (0.1616)	REMOVED	---	---	---
FAMHX1	0.1287 (0.1567)	0.1178 (0.1562)	0.1288 (0.1566)	0.1358 (0.1563)	REMOVED	
ASPNEW	-0.3698 (0.1352)	-0.3825 (0.1348)	-0.3698 (0.1351)	-0.3712 (0.1351)	-0.3555 (0.1326)	REMOVED
AGEGRP	0.1258 (0.0310)	0.1244 (0.0310)	0.1257 (0.0310)	0.1262 (0.0309)	0.1339 (0.0303)	0.1150 (0.0292)
Interactions:						
BMI1SNP1	0.0706 (0.2786)	REMOVED	0.0706 (0.2786)	0.0696 (0.2784)	0.00421 (0.2728)	0.0182 (0.2639)
BMI2SNP1	-0.5954 (0.3009)	REMOVED	-0.5951 (0.3005)	-0.6046 (0.3002)	-0.5935 (0.2946)	-0.6016 (0.2848)
-2 log L	1640.226	1645.351	1640.226	1641.041	1697.881	1813.388
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1	Removed ASPNEW
Test of interaction term		χ^2 (2): = 1645.351-1640.226=5.125; p=0.08	---	---	---	---
Change in estimate			|0.1697-0.1694| = 0.0003	|0.1697-0.1706| = 0.0009	|0.1697-0.1981| = 0.0284	|0.1697-0.2199| = 0.0502
Compare to:		Full Model	Full Model	Full Model	Full Model	Full Model

Conclusion:	Interaction term is contributing to the model and will be kept in the model. Results should be presented as stratified.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. Aspirin use is not a confounder and is removed from the model

TABLE D.10. Model building procedure for SNP rs135542 (AA vs. AG+GG) in post-menopausal women, NSAID interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1,126	1,126	1,127	1,127	1,164	1,253
Main Exposure: **rs135542**	0.2193 (0.1439)	0.0585 (0.1249)	0.2154 (0.1439)	0.2195 (0.1433)	0.2508 (0.1412)	0.1768 (0.1347)
INTERCEPT	-1.4200 (0.3629)	-1.3022 (0.3581)	-1.4194 (0.3584)	-1.4344 (0.3559)	-1.4286 (0.3518)	-1.1935 (0.2965)
WTCH_1	0.0151 (0.1485)	0.00478 (0.1481)	0.0200 (0.1484)	0.0187 (0.1483)	0.00556 (0.1465)	REMOVED
WTCH_2	0.4042 (0.1496)	0.4012 (0.1492)	0.4030 (0.1496)	0.4070 (0.1491)	0.4013 (0.1460)	REMOVED
RACE_DIC	-0.00012 (0.2714)	0.0299 (0.2711)	REMOVED	---	---	---
JEWISH	-0.0551 (0.1697)	-0.0468 (0.1694)	-0.0586 (0.1688)	REMOVED	---	---
FAMHX1	0.1184 (0.1631)	0.1223 (0.1629)	0.1286 (0.1627)	0.1261 (0.1626)	REMOVED	---
ASPNEW	-0.1955 (0.1755)	-0.4354 (0.1407)	-0.1995 (0.1754)	-0.1990 (0.1754)	-0.1543 (0.1729)	-0.1667 (0.1651)
AGEGRP	0.1410 (0.0364)	0.1356 (0.0362)	0.1411 (0.0362)	0.1414 (0.0362)	0.1448 (0.0357)	0.1364 (0.0302)
Interactions						
ASPSNP1	-0.6634 (0.2933)	REMOVED	-0.6599 (0.2929)	-0.6582 (0.2929)	-0.7470 (0.2871)	-0.5505 (0.2736)
-2 log L	1518.470	1523.644	1519.729	1519.849	1568.494	1706.293
Notes		Evaluate EMM with the likelihood ratio test*	Removed RACE_DIC	Removed JEWISH	Removed FAMHX1	Removed WTCH_1, WTCH_2
Test of interaction term		χ^2 (1): = 1523.644-1518.470=5.174; p=0.02	---	---	---	---
Change in estimate			|0.2193-0.2154| = 0.0039	|0.2193-0.2195| = 0.0002	|0.2193-0.2508| = 0.0315	|0.2193-0.1768| = 0.0425
Compare to:		Full Model	Full Model	Full Model	Full Model	Full Model

Conclusion:	Interaction term is contributing to the model and will be kept. Results should be presented as stratified.	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. Weight change is not a confounder and is removed from the model

TABLE D.11. Summary of odds ratios for rs135542 stratified by three effect modifiers						
Exposure	**Status**	**Genotypes**	**Controls (N)**	**Cases (N)**	**Stratum OR (95% CI)**	**Adjusted OR (95% CI)***
Weight gain from age 50 (kg)	No change or loss	AA	149	114	1.00	1.00
		AG+AA	94	99	1.38 (0.95, 2.00)	1.40 (0.96, 2.04)
	0.01-4.98	AA	123	101	1.00	1.05 (0.73, 1.51)
		AG+AA	75	72	1.17 (0.77, 1.77)	1.23 (0.82, 1.85)
	≥ 4.99	AA	98	156	1.00	1.92 (1.34, 2.74)
		AG+AA	73	85	0.73 (0.49, 1.09)	1.47 (0.99, 2.19)
BMI at reference (kg/m^2)	< 24.99	AA	187	143	1.00	1.00
		AG+AA	127	119	1.23 (0.88, 1.71)	1.25 (0.89, 1.74)
	25.00-29.99	AA	131	133	1.00	1.28 (0.92, 1.77)
		AG+AA	73	89	1.20 (0.81, 1.78)	1.62 (1.11, 2.37)
	≥ 30.00	AA	79	119	1.00	1.99 (1.39, 2.86)
		AG+AA	65	66	0.67 (0.43, 1.05)	1.36 (0.91, 2.05)
Aspirin use	Non-user	AA	256	279	1.00	1.00
		AG+AA	177	219	1.14 (0.87, 1.47)	1.19 (0.92, 1.55)
	User	AA	106	102	1.00	0.85 (0.61, 1.17)
		AG+AA	68	46	0.70 (0.44, 1.12)	0.58 (0.38, 0.88)

*Adjusted for age

TABLE D.12. Model building procedure for SNP rs1800206 (CC vs. CG+GG) in post-menopausal women, Jewish interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1,316	1,316	1,317	1,317	1,362	1,362
Main Exposure: **rs1800206**	-0.1048 (0.1927)	-0.0171 (0.1821)	-0.0188 (0.1820)	-0.0174 (0.1818)	0.0386 (0.1785)	0.0624 (0.1778)
INTERCEPT	-1.1147 (0.2924)	-1.1214 (0.2920)	-1.1220 (0.2892)	-1.1281 (0.2863)	-1.1362 (0.2809)	-0.9833 (0.2744)
BMIREF_DIC	0.3375 (0.1126)	0.3348 (0.1125)	0.3370 (0.1124)	0.3379 (0.1122)	0.3153 (0.1103)	REMOVED
RACE_DIC	0.00342 (0.2437)	0.00558 (0.2436)	REMOVED	---	---	---
JEWISH	-0.0816 (0.1608)	-0.0213 (0.1549)	-0.0233 (0.1541)	REMOVED	---	---
FAMHX1	0.1185 (0.1499)	0.1202 (0.1497)	0.1278 (0.1494)	0.1271 (0.1493)	REMOVED	---
AGEGRP	0.0972 (0.0295)	0.0970 (0.0295)	0.0970 (0.0294)	0.0972 (0.0294)	0.1026 (0.0287)	0.1050 (0.0286)
Interactions:						
JEWSNP2	0.8447 (0.6123)	REMOVED	---	---	---	---
-2 log L	1800.186	1802.167	1803.269	1803.292	1866.184	1874.379
Notes		Evaluate EMM with the likelihood ratio test*	Removed RACE_DIC	Removed JEWISH	Removed FAMHX1	Removed BMIREF_DIC
Test of interaction term		χ^2 (1): = 1802.167-1800.186=1.981; p=0.16	---	---	---	---
Change in estimate			|-0.0171--0.0188| = 0.0017	|-0.0171--0.0174| = 0.0003	|-0.0171-0.0386| = 0.0557	|-0.0171-0.0624| = 0.0795
Compare to:		Full Model	Model 2	Model 2	Model 2	Model 2

Conclusion:	Interaction term is not contributing to the model (p>0.10) and can be removed.	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. BMI is not a confounder and is removed from the model

TABLE D.13. Model building procedure for SNP rs1800206 (CC vs. CG+GG) in post-menopausal women, PA interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1,233	1,233	1,233	1,234	1,275	1,362
Main Exposure: **rs1800206**	0.4547 (0.3616)	-0.0127 (0.1885)	-0.0101 (0.1882)	-0.0130 (0.1881)	0.0318 (0.1852)	0.0624 (0.1778)
INTERCEPT	-0.9554 (0.3192)	-0.9002 (0.3164)	-0.9087 (0.3142)	-0.8924 (0.3116)	-0.9491 (0.3078)	-0.9833 (0.2744)
PA_1	0.1252 (0.1622)	0.0552 (0.1553)	0.0535 (0.1551)	0.0480 (0.1550)	0.0927 (0.1527)	REMOVED
PA_2	-0.1608 (0.1716)	-0.1605 (0.1619)	-0.1612 (0.1619)	-0.1657 (0.1618)	-0.1346 (0.1592)	REMOVED
PA_3	-0.1502 (0.1692)	-0.2679 (0.1592)	-0.2688 (0.1591)	-0.2752 (0.1590)	-0.2799 (0.1567)	REMOVED
RACE_DIC	0.0810 (0.2535)	0.0777 (0.2531)	0.0833 (0.2519)	REMOVED	---	---
JEWISH	-0.0304 (0.1608)	-0.0365 (0.1605)	REMOVED	---	---	---
FAMHX1	0.1275 (0.1545)	0.1123 (0.1540)	0.1111 (0.1539)	0.1178 (0.1535)	REMOVED	---
AGEGRP	0.1025 (0.0307)	0.1017 (0.0306)	0.1020 (0.0306)	0.1013 (0.0305)	0.1090 (0.0300)	0.1050 (0.0286)
Interactions:						
PA1SNP2	-0.8180 (0.5702)	REMOVED	---	---	---	---
PA2SNP2	-0.0895 (0.5281)	REMOVED	---	---	---	---
PA3SNP2	-1.0248 (0.5136)	REMOVED	---	---	---	---
-2 log L	1685.107	1690.851	1690.903	1692.180		
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1	Removed PA_1, PA_2, PA_3
Test of interaction term		χ^2 (3): = 1690.851-1685.107=5.744; p=0.12	---	---	---	---
Change in estimate			|-0.0127--0.0101| = 0.0026	|-0.0127--0.0130| = 0.0003	|-0.0127-0.0318| = 0.0445	|-0.0127--0.0624| = 0.0751

Compare to:	Full Model	Model 2	Model 2	Model 2	Model 2
Conclusion:	Interaction term is not contributing to the model (p>0.10) and can be removed.	Change in estimate < 0.10. Jewish is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. PA is not a confounder and is removed from the model

TABLE D.14. Model building procedure for SNP rs4253623 (AA vs. AG+GG) in post-menopausal women, NSAID interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1253	1253	1253	1254	1295	1377
Main Exposure: **rs4253623**	-0.1747 (0.1583)	-0.0734 (0.1350)	-0.0729 (0.1350)	-0.0741 (0.1349)	-0.0532 (0.1330)	-0.0716 (0.1281)
INTERCEPT	-0.9549 (0.2936)	-0.9813 (0.2927)	-0.9895 (0.2895)	-0.9801 (0.2871)	-1.0153 (0.2828)	-0.9837 (0.2740)
ASPNEW	-0.4644 (0.1511)	-0.3737 (0.1313)	-0.3734 (0.1313)	-0.3761 (0.1313)	-0.3606 (0.1291)	REMOVED
RACE_DIC	0.0642 (0.2551)	0.0648 (0.2549)	0.0694 (0.2538)	REMOVED	REMOVED	REMOVED
JEWISH	-0.0258 (0.1586)	-0.0295 (0.1585)	REMOVED	REMOVED	REMOVED	REMOVED
FAMHX1	0.1076 (0.1549)	0.1161 (0.1547)	0.1151 (0.1546)	0.1220 (0.1542)	REMOVED	REMOVED
AGEGRP	0.1187 (0.0304)	0.1190 (0.0304)	0.1193 (0.0303)	0.1188 (0.0302)	0.1262 (0.0297)	0.1071 (0.0285)
Interactions:						
ASPSNP3	0.3702 (0.3024)	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1711.883	1713.379	1713.414	1714.576	1770.055	1894.097
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1	Removed ASPNEW
Test of interaction term		χ^2 (1): = 1713.379-1711.883=1.496; p=0.22	---	---	---	---
Change in estimate			|-0.0734 - -0.0729| = 0.0005	|-0.0734- -0.0741| = 0.0007	|-0.0734- -0.0532| = 0.0202	|-0.0734- -0.0716| = 0.0018
Compare to:		Full Model	Model 2	Model 2	Model 2	Model 2

Conclusion:	Interaction term is not contributing to the model (p>0.10) and can be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. Aspirin is not a confounder and is removed from the model

TABLE D.15. ORs, ICRs and multiplicative p-values for rs4253623 by aspirin use

Exposure	Status	Genotypes	Controls (N)	Cases (N)	$OR_{adjusted}$ (95% CI)*	ICR (95% CI)	P-value†
Aspirin Use	Non-User	AA	339	403	1.00		
		AG+GG	108	108	0.86 (0.63, 1.16)		
	User	AA	140	113	0.64 (0.48, 0.85)		
		AG+GG	42	42	0.79 (0.50, 1.25)	0.96 (-0.91, 2.82)	0.22

*adjusted for age
†multiplicative p-value

TABLE D.16. Model building procedure for SNP rs4253699 (TT vs. CT+CC) in post-menopausal women, PA interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Sample Size	1239	1239	1239	1240	1281	1370
Main Exposure: **rs4253699**	0.4709 (0.2135)	0.0889 (0.1182)	0.0891 (0.1180)	0.0919 (0.1179)	0.0685 (0.1160)	0.0954 (0.1115)
INTERCEPT	-1.1443 (0.3237)	-1.0222 (0.3170)	-1.0210 (0.3149)	-1.0104 (0.3123)	-1.0479 (0.3085)	-1.0703 (0.2753)
PA_1	0.3334 (0.1960)	0.0716 (0.1552)	0.0718 (0.1550)	0.0671 (0.1550)	0.1099 (0.1527)	REMOVED
PA_2	0.00480 (0.2051)	-0.1213 (0.1607)	-0.1212 (0.1607)	-0.1257 (0.1606)	-0.0970 (0.1581)	REMOVED
PA_3	-0.0506 (0.2053)	-0.2945 (0.1595)	-0.2944 (0.1595)	-0.3003 (0.1593)	-0.3019 (0.1570)	REMOVED
RACE_DIC	0.0642 (0.2524)	0.0580 (0.2515)	0.0572 (0.2503)	REMOVED	REMOVED	REMOVED
JEWISH	-0.0198 (0.1603)	0.00538 (0.1595)	REMOVED	REMOVED	REMOVED	REMOVED
FAMHX1	0.1105 (0.1545)	0.1075 (0.1542)	0.1076 (0.1541)	0.1146 (0.1537)	REMOVED	REMOVED
AGEGRP	0.1073 (0.0305)	0.1098 (0.0305)	0.1097 (0.0305)	0.1092 (0.0304)	0.1163 (0.0299)	0.1111 (0.0285)
Interactions:						
PA1SNP4	-0.7006 (0.3223)	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
PA2SNP4	-0.3260 (0.3294)	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
PA3SNP4	-0.6200 (0.3275)	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1690.326	1696.261	1696.262	1697.439	1752.145	1882.945
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1	Removed PA_1, PA_2, PA_3
Test of interaction term		χ^2 (3): = 1696.261-1690.326=5.935; p=0.11	---	---	---	---
Change in estimate			|0.0889-0.0891| = 0.0002	|0.0889-0.0919| = 0.003	|0.0889-0.0685| = 0.0204	|0.0889-0.0954| = 0.0065

Compare to:	Full Model	Model 2	Model 2	Model 2	Model 2
Conclusion:	Interaction term is not contributing to the model (p>0.10) and is removed from the model	Change in estimate < 0.10. Jewish is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. PA is not a confounder and is removed from the model

TABLE D.17. Model building procedure for SNP rs4253699 (TT vs. CT+CC) in post-menopausal women, NSAID interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	1247	1247	1247	1248	1288
Main Exposure: **rs4253699**	-0.0372 (0.1358)	0.0764 (0.1178)	-0.0378 (0.1357)	-0.0332 (0.1356)	-0.0738 (0.1334)
INTERCEPT	-1.0044 (0.2960)	-1.0628 (0.2939)	-1.0085 (0.2931)	-1.0050 (0.2907)	-1.0152 (0.2867)
ASPNEW	-0.5631 (0.1664)	-0.3959 (0.1321)	-0.5629 (0.1664)	-0.5632 (0.1664)	-0.5521 (0.1635)
RACE_DIC	0.0421 (0.2576)	0.0289 (0.2570)	0.0445 (0.2564)	REMOVED	REMOVED
JEWISH	-0.0156 (0.1590)	-0.0182 (0.1589)	REMOVED	REMOVED	REMOVED
FAMHX1	0.1355 (0.1543)	0.1369 (0.1542)	0.1349 (0.1542)	0.1423 (0.1539)	REMOVED
AGEGRP	0.1220 (0.0305)	0.1236 (0.0305)	0.1222 (0.0304)	0.1220 (0.0304)	0.1290 (0.0298)
Interactions:					
ASPSNP4	0.4546 (0.2721)	REMOVED	0.4549 (0.2720)	0.4487 (0.2719)	0.4687 (0.2673)
-2 log L	1699.775	1702.574	1699.785	1700.908	1755.090
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1
Test of interaction term		χ^2 (1): = 1702.574-1699.775=2.799; p=0.09	---	---	---
Change in estimate			|-0.0372--0.0378| = 0.0006	|-0.0372--0.0332| = 0.004	|-0.0372- -0.0738| = 0.0366
Compare to:		Full Model	Full Model	Full Model	Full Model

Conclusion:	Interaction term is contributing to the model (p<0.10) and will be kept in the model. Results will be stratified.	Change in estimate < 0.10. Jewish is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model

TABLE D.18. ORs, ICRs and multiplicative p-values for rs4253699 by physical activity level and aspirin use

Exposure	**Status**	**Genotypes**	**Controls (N)**	**Cases (N)**	**$OR_{adjusted}$ (95% CI)***	**ICR (95% CI)**	**P-value†**
Physical Activity (hrs/wk)	0	TT	124	113	1.00		
		CT+CC	63	94	1.61 (1.07, 2.43)		
	< 0.69	TT	89	117	1.48 (1.01, 2.16)		
		CT+CC	56	55	1.12 (0.71, 1.77)	-1.62 (-2.89, -0.36)	
	0.7-2.6	TT	91	80	1.03 (0.69, 1.53)		
		CT+CC	54	56	1.20 (0.76, 1.90)	-0.92 (-2.21, 0.37)	
	≥ 2.7	TT	93	79	0.97 (0.65, 1.45)		
		CT+CC	69	48	0.79 (0.50, 1.24)	-1.08 (-2.23, 0.06)	0.06
Aspirin Use	Non-User	TT	261	310	1.00		
		CT+CC	182	201	0.93 (0.72, 1.21)		
	User	TT	125	90	0.58 (0.42, 0.79)		
		CT+CC	56	63	0.85 (0.57, 1.28)	1.09 (-0.83, 3.01)	0.08

*adjusted for age
†multiplicative p-value

TABLE D.19. Model building procedure for SNP rs4253755 (GG vs. AG+AA) in post-menopausal women, PA interaction

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	1,242	1,242	1,242	1,243	1,285
Main Exposure: **rs4253755**	0.2551 (0.2531)	0.0913 (0.1396)	0.2552 (0.2531)	0.2684 (0.2516)	0.2431 (0.2462)
INTERCEPT	-0.9480 (0.3165)	-0.9351 (0.3143)	-0.9497 (0.3143)	-0.9367 (0.3114)	-0.9825 (0.3078)
PA_1	0.1976 (0.1748)	0.0169 (0.1545)	0.1973 (0.1746)	0.1962 (0.1746)	0.2243 (0.1722)
PA_2	-0.2238 (0.1804)	-0.1823 (0.1608)	-0.2238 (0.1804)	-0.2242 (0.1804)	-0.1843 (0.1777)
PA_3	-0.3585 (0.1823)	-0.3526 (0.1597)	-0.3586 (0.1823)	-0.3613 (0.1821)	-0.3362 (0.1798)
RACE_DIC	0.0783 (0.2552)	0.0800 (0.2543)	0.0794 (0.2539)	REMOVED	---
JEWISH	-0.00704 (0.1594)	0.00152 (0.1589)	REMOVED	---	---
FAMHX1	0.0876 (0.1549)	0.0816 (0.1544)	0.0874 (0.1549)	0.0933 (0.1546)	REMOVED
AGEGRP	0.1062 (0.0305)	0.1084 (0.0304)	0.1063 (0.0305)	0.1054 (0.0304)	0.1129 (0.0299)
Interactions:					
PA1SNP3	-0.8364 (0.3783)	REMOVED	-0.8361 (0.3782)	-0.8536 (0.3776)	-0.7958 (0.3700)
PA2SNP3	0.2061 (0.3970)	REMOVED	0.2061 (0.3970)	0.1891 (0.3964)	0.1253 (0.3867)
PA3SNP3	-0.00054 (0.3794)	REMOVED	-0.00039 (0.3793)	-0.0162 (0.3787)	-0.1371 (0.3701)
-2 log L	1691.449	1699.463	1691.451	1692.522	1749.938
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC	Removed FAMHX1
Test of interaction term		χ^2 (3): = 1699.463-1691.449=8.014; p=0.05	---	---	---
Change in estimate			|0.2551-0.2552| = 0.0001	|0.2551-0.2684| = 0.0133	|0.2551-0.2431| = 0.0120

Compare to:	Full Model	Full Model	Full Model	Full Model
Conclusion:	Interaction term is contributing to the model and will be kept. Results should be presented as stratified..	Change in estimate < 0.10. Jewish is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model

TABLE D.20. Summary of odds ratios for rs4253755 stratified by PA

Exposure	Status	Genotypes	Controls (N)	Cases (N)	Stratum OR (95% CI)	Adjusted OR (95% CI)*
Physical Activity (hrs/wk)	0	AA	147	162	1.00	1.00
		AG+AA	53	36	1.34 (0.83, 2.16)	1.28 (0.79, 2.07)
	< 0.69	AA	107	144	1.00	1.25 (0.89, 1.75)
		AG+AA	39	30	0.57 (0.33, 0.98)	0.72 (0.42, 1.22)
	0.7-2.6	AA	123	93	1.00	0.83 (0.59, 1.18)
		AG+AA	38	33	1.42 (0.80, 2.54)	1.20 (0.68, 2.12)
	≥ 2.7	AA	119	103	1.00	0.71 (0.50, 1.02)
		AG+AA	26	32	1.15 (0.67, 1.97)	0.79 (0.47, 1.34)

TABLE D.21. Model building procedure for SNP rs4253760 (TT vs. GT+GG) in post-menopausal women, no interactions

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1,282	1,283	1,284	1,330
Main Exposure: **rs4253760**	0.1891 (0.1221)	0.1957 (0.1206)	0.1918 (0.1202)	0.2056 (0.1178)
INTERCEPT	-1.0274 (0.2890)	-1.0247 (0.2868)	-1.0211 (0.2841)	-1.0330 (0.2789)
RACE_DIC	0.0340 (0.2507)	REMOVED	---	---
JEWISH	0.0335 (0.1558)	0.0304 (0.1552)	REMOVED	---
FAMHX1	0.0872 (0.1522)	0.0944 (0.1519)	0.0929 (0.1518)	REMOVED
AGEGRP	0.1032 (0.0298)	0.1030 (0.0297)	0.1034 (0.0296)	0.1076 (0.0290)
Interactions:				
-2 log L	1761.540	1762.675	1763.988	1826.088
Notes		Removed RACE_DIC	Removed JEWISH	Removed FAMHX1
Test of interaction term	---	---	---	---
Change in estimate		|0.1891-0.1957| = 0.0066	|0.1891-0.1918| = 0.0027	|0.1891-0.2056| = 0.0165
Compare to:		Full Model	Full Model	Full Model
Conclusion:		Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Jewish is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model

TABLE D.22. Haplotypes, frequencies and Odds Ratios by menopausal status for six PPARA SNPs

Haplotype Number*	rs135542	rs1800206	rs4253623	rs4253699	rs4253755	rs4253760	Control Freq	Case Freq	OR (95% CI)†
All women									
1	A	C	A	C	A	G	0.06	0.06	0.99 (0.58, 1.70)
2	A	C	A	C	A	T	0.02	0.03	1.68 (0.73, 3.88)
3	A	C	A	C	G	G	0.02	0.02	0.98 (0.33, 2.89)
4	A	C	A	C	G	T	0.05	0.05	0.79 (0.41, 1.52)
5	A	C	A	T	A	G	0.03	0.03	1.07 (0.47, 2.44)
6	A	C	A	T	G	G	0.03	0.03	1.60 (0.74, 3.44)
7	A	C	A	T	G	T	0.41	0.40	0.98 (0.79, 1.22)
8	A	C	G	T	G	T	0.10	0.10	1.09 (0.72, 1.63)
9	A	G	A	C	G	G	0.01	0.02	2.49 (0.90, 6.86)
10	A	G	A	C	G	T	0.01	0.01	1.18 (0.32, 4.42)
11	G	C	A	C	G	T	0.01	0.01	1.37 (0.29, 6.55)
12	G	C	A	T	G	T	0.18	0.18	1.03 (0.74, 1.42)
Post-menopausal women only									
1	A	C	A	C	A	G	0.06	0.05	1.01 (0.49, 2.09)
2	A	C	A	C	A	T	0.02	0.03	2.03 (0.71, 5.82)
3	A	C	A	C	G	G	0.02	0.01	0.97 (0.23, 4.18)
4	A	C	A	C	G	T	0.06	0.05	0.89 (0.40, 1.99)
5	A	C	A	T	A	G	0.03	0.03	1.64 (0.61, 4.46)
6	A	C	A	T	G	G	0.02	0.04	2.83 (0.93, 8.65)
7	A	C	A	T	G	T	0.42	0.39	0.92 (0.70, 1.20)
8	A	C	G	C	G	T	0.01	0.01	0.77 (0.12, 4.91)
9	A	C	G	T	G	T	0.10	0.10	0.97 (0.59, 1.59)
10	A	G	A	C	G	G	0.01	0.03	5.02 (1.45, 17.39)
11	A	G	A	T	G	G	0.01	0.01	0.47 (0.08, 2.83)
12	G	C	A	C	G	T	0.01	0.01	0.42 (0.05, 3.55)
13	G	C	A	T	G	G	0.01	0.01	1.07 (0.13, 8.67)
14	G	C	A	T	G	T	0.17	0.18	1.19 (0.79, 1.80)

*haplotypes with frequency ≥ 0.01
†adjusted for age, frequency matched variable

FIGURE D.1. Kaplan-Meier curves for rs135542 (AA vs. AG+GG) for overall survival

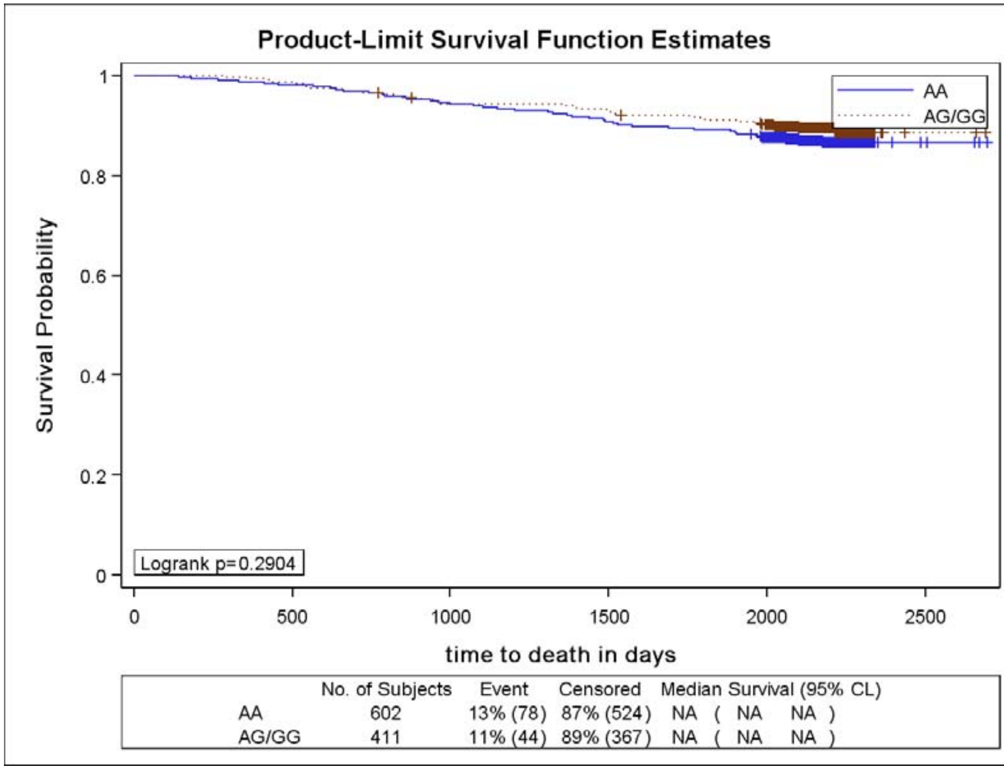


FIGURE D.2. Kaplan-Meier curve for rs1800206 (CC vs. CG+GG) for overall survival

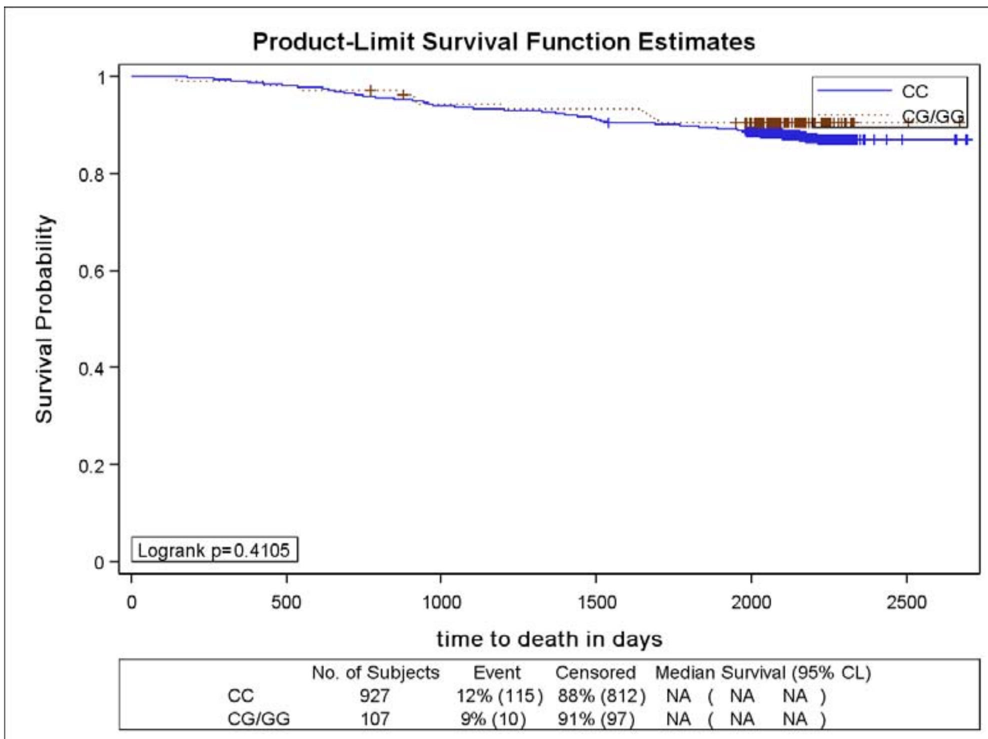


FIGURE D.3. Kaplan-Meier curves for rs4253623 (AA vs. AG+GG) for overall survival

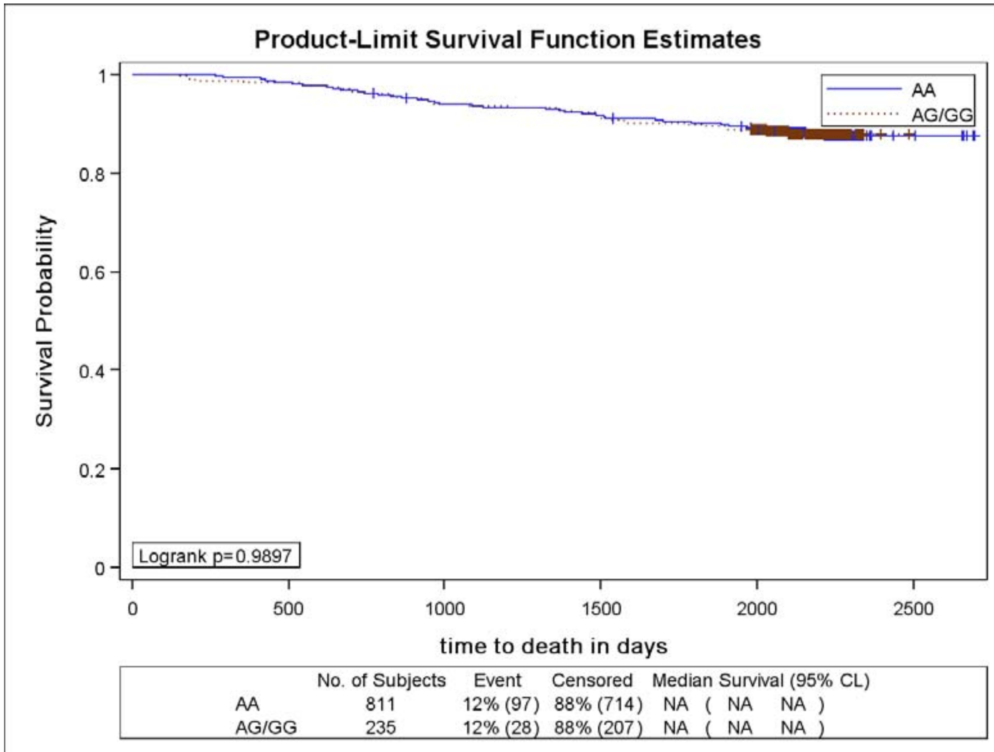


FIGURE D.4. Kaplan-Meier curve for rs4253699 (TT vs. CT+CC) for overall survival

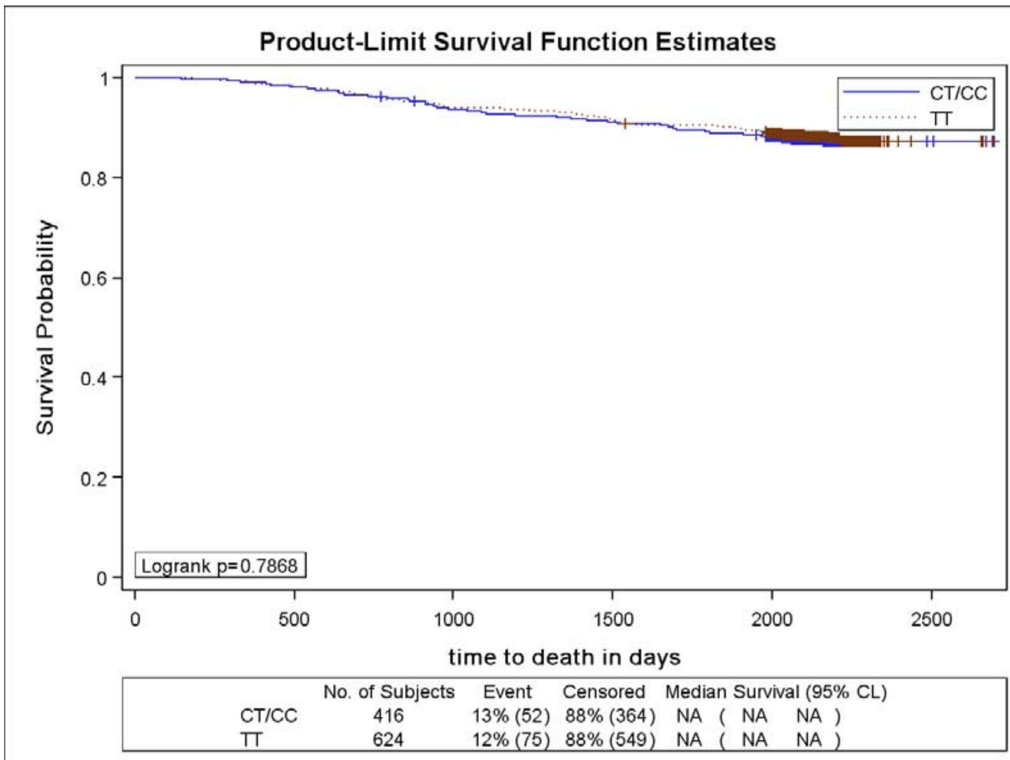


FIGURE D.5. Kaplan-Meier curve for rs4253755 (GG vs. AG+AA) for overall survival

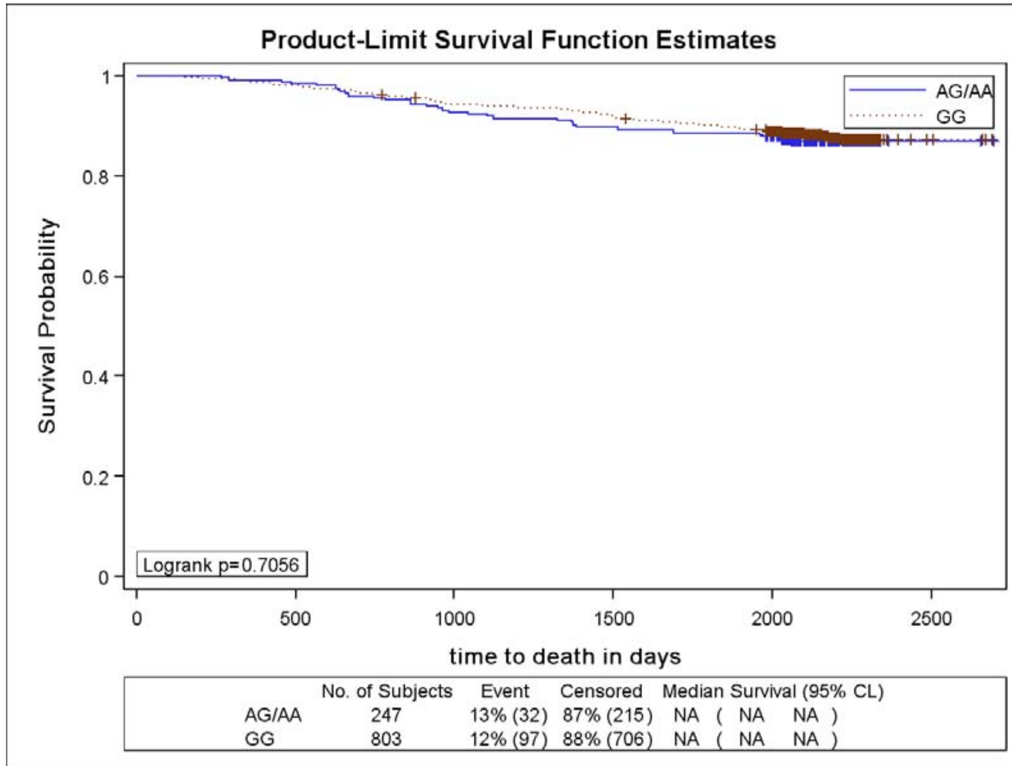


FIGURE D.6. Kaplan-Meier curve for rs4253760 (TT vs. GT+GG) for overall survival

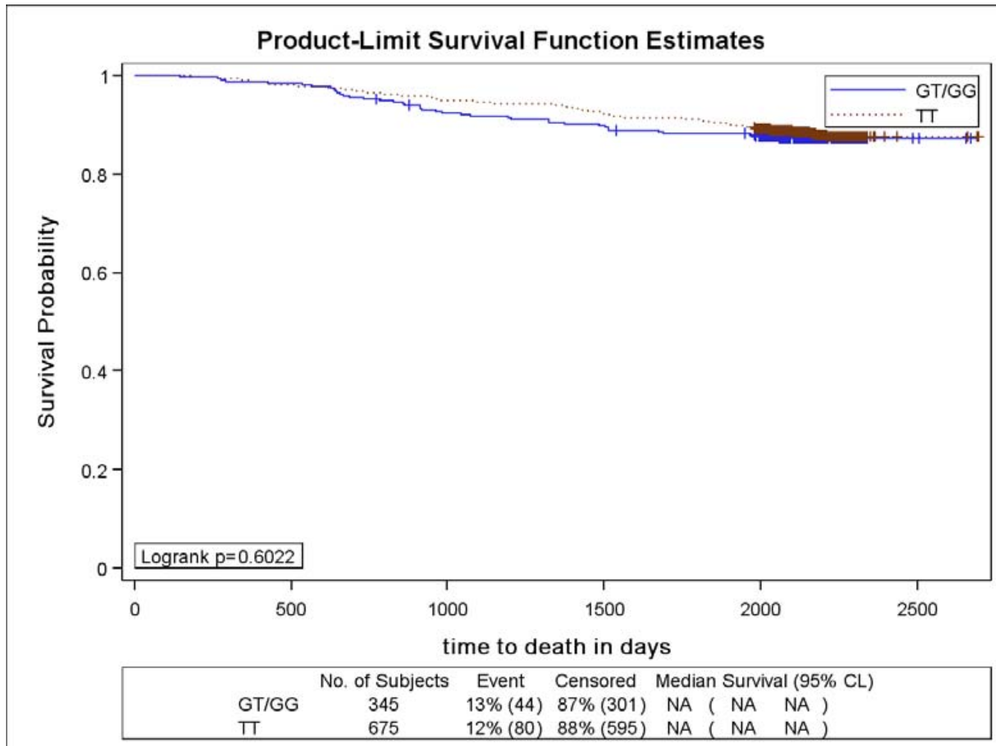


TABLE D.23. Model building procedure for SNP rs135542 (AA vs. AG+GG), looking at overall survival with continuous interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1010	1010	1011	1013
Main Exposure: **rs135542**	-0.24419 (0.43248)	-0.22084 (0.18959)	-0.18859 (0.18866)	-0.19903 (0.18854)
JEWISH	-0.40175 (0.27825)	-0.40183 (0.27825)	REMOVED	---
RACE_DIC	0.52756 (0.31793)	0.52756 (0.31793)	0.58208 (0.31631)	REMOVED
Interactions:				
SNP1_DUR	0.0000202 (0.0003356)	REMOVED	---	---
-2 log L	1659.942	1659.946	1662.497	1665.901
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1659.946– 1659.942= 0.004; p=0.95	---	---
Change in estimate			|-0.22084--0.18859| = 0.0323	|-0.22084- -0.19903| = 0.0218
Compare to:		Full Model	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.24. Model building procedure for SNP rs1800206 (CC vs. CG+GG), looking at overall survival with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1032	1032	1033	1034
Main Exposure: **rs1800206**	0.15791 (0.69445)	-0.28570 (0.32995)	-0.29635 (0.32985)	-0.27023 (0.32966)
JEWISH	-0.55079 (0.29430)	-0.55031 (0.29430)	-0.58661 (0.29314)	REMOVED
RACE_DIC	0.56211 (0.30492)	0.56025 (0.30490)	REMOVED	---
Interactions:				
SNP2_DUR	-0.0004162 (0.0006035)	REMOVED	---	---
-2 log L	1702.728	1703.216	1706.424	1711.347
Notes		Evaluate EMM with the likelihood ratio test*	Removed RACE_DIC	Removed JEWISH
Test of interaction term		χ^2 (1): 1703.216–1702.728=0.488; p=0.48	---	---
Change in estimate			|-0.28570- -0.29635| = 0.0107	|-0.28570- -0.27023| = 0.0155
Compare to:		Full Model	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model

TABLE D.25. Model building procedure for SNP rs4253623 (AA vs. AG+GG), looking at overall survival with continuous interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1,043	1,043	1,044	1,046
Main Exposure: **rs4253623**	0.11838 (0.47793)	-0.03073 (0.21483)	-0.00347 (0.21454)	-0.00277 (0.21453)
JEWISH	-0.55396 (0.29459)	-0.55430 (0.29458)	REMOVED	REMOVED
RACE_DIC	0.54084 (0.30487)	0.54045 (0.30486)	0.60686 (0.30364)	REMOVED
Interactions:				
SNP3_DUR	-0.0001324 (0.0003828)	REMOVED	REMOVED	REMOVED
-2 log L	1706.738	1706.858	1711.172	1715.084
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1706.738–1706.858= 0.12; p=0.73	---	---
Change in estimate			|-0.03073--0.00347| = 0.02726	|-0.03073--0.00277| = 0.0296
Compare to:		Full Model	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.26. Model building procedure for SNP rs4253699 (TT vs. CT+CC), looking at overall survival with continuous interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1,038	1,038	1,039	1,040
Main Exposure: **rs4253699**	0.17208 (0.40302)	0.04990 (0.18048)	0.04900 (0.18047)	0.04902 (0.18046)
JEWISH	-0.49808 (0.28453)	-0.49764 (0.28452)	REMOVED	REMOVED
RACE_DIC	0.58091 (0.30466)	0.58060 (0.30465)	0.64145 (0.30339)	REMOVED
Interactions:				
SNP4_DUR	-0.0001083 (0.0003202)	REMOVED	REMOVED	REMOVED
-2 log L	1732.883	1732.998	1736.701	
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1732.883– 1732.998= 0.115; p=0.73	---	---
Change in estimate			|0.04990-0.04900| = 0.009	|0.04990-0.04902| = 0.0008
Compare to:		Full Model	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.27. Model building procedure for SNP rs4253755 (GG vs. AG+AA), looking at overall survival with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1047	1047	1048	1050
Main Exposure: **rs4253755**	0.45333 (0.44432)	0.10377 (0.20539)	0.11206 (0.20534)	0.07699 (0.20387)
JEWISH	-0.34978 (0.26833)	-0.34957 (0.26832)	REMOVED	---
RACE_DIC	0.46547 (0.31890)	0.46856 (0.31894)	0.51489 (0.31755)	REMOVED
Interactions:				
SNP5_DUR	-0.0003158 (0.0003641)	REMOVED	---	---
-2 log L	1764.960	1765.722	1767.806	1770.639
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1764.960–1765.722=0.762; p=0.38	---	---
Change in estimate			|0.10377-0.11206| = 0.0083	|0.10377-0.07699| = 0.0268
Compare to:		Full Model	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.28. Model building procedure for SNP rs4253760 (TT vs. GT+GG), looking at overall survival with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)
Sample Size	1017	1017	1018	1020
Main Exposure: **rs4253760**	0.76088 (0.41558)	0.04700 (0.19100)	0.79878 (0.41410)	0.81284 (0.41404)
JEWISH	-0.38738 (0.27671)	-0.38712 (0.27669)	-0.41388 (0.27558)	REMOVED
RACE_DIC	0.44255 (0.32260)	0.44358 (0.32255)	REMOVED	---
Interactions:				
SNP6_DUR	-0.0006499 (0.0003442)	REMOVED	-0.0006512 (0.0003443)	-0.0006510 (0.0003443)
-2 log L	1686.081	1689.785	1688.057	1691.044
Notes		Evaluate EMM with the likelihood ratio test*	Removed RACE_DIC	Removed JEWISH
Test of interaction term		χ^2 (1): 1686.081–1689.785=3.704; p=0.05	---	---
Change in estimate			|0.76088-0.79878| = 0.0379	|0.76088-0.81284| = 0.05196
Compare to:		Full Model	Full Model	Full Model
Conclusion:		Interaction term is contributing to the model and will be kept.	Change in estimate < 0.10. Race is not a confounder and is removed from the model	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
rs135542								
Age at diagnosis (years)	< 45	AA	85	6	1.00	1.00		
		AG+GG	53	7	1.82 (0.61, 5.41)	1.84 (0.62, 5.47)		
	45-54	AA	136	16	1.00	1.66 (0.65, 4.24)		
		AG+GG	104	10	0.83 (0.38, 1.83)	1.37 (0.50, 3.77)	-1.13 (-3.60, 1.35)	
	55-64	AA	124	11	1.00	1.27 (0.47, 3.43)		
		AG+GG	106	11	1.16 (0.50, 2.67)	1.46 (0.54, 3.95)	-0.65 (-2.86, 1.57)	0.48
	≥ 65	AA	179	45	1.00	3.31 (1.41, 7.75)		
		AG+GG	104	16	0.65 (0.37, 1.14)	2.14 (0.84, 5.48)	-2.00 (-5.11, 1.11)	
Menopausal Status	Pre-menopausal	AA	175	16	1.00	1.00		
		AG+GG	117	14	1.28 (0.62, 2.62)	1.26 (0.63, 2.54)		
	Post-menopausal	AA	335	60	1.00	1.85 (1.09, 3.13)		
		AG+GG	244	30	0.71 (0.46, 1.10)	1.31 (0.73, 2.35)	-0.80 (-2.05, 0.45)	0.17
Stage	In-situ	AA	106	4	1.00	1.00		
		AG+GG	60	3	1.32 (0.30, 5.92)	1.33 (0.30, 5.94)		
	Invasive	AA	418	74	1.00	4.45 (1.63, 12.18)		
		AG+GG	307	41	0.77 (0.53, 1.13)	3.44 (1.23, 9.59)	-1.35 (-4.46, 1.77)	0.50
ER status	ER-	AA	71	25	1.00	1.00		
		AG+GG	53	11	0.64 (0.32, 1.31)	0.64 (0.31, 1.29)		
	ER+	AA	268	34	1.00	0.38 (0.23, 0.64)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		AG+GG	173	16	0.75 (0.41, 1.36)	0.29 (0.15, 0.54)	0.27 (-0.24, 0.78)	0.72
PR status	PR-	AA	101	31	1.00	1.00		
		AG+GG	70	18	0.85 (0.48, 1.53)	0.85 (0.48, 1.52)		
	PR+	AA	238	28	1.00	0.41 (0.24, 0.68)		
		AG+GG	156	9	0.52 (0.24, 1.09)	0.21 (0.10, 0.44)	-0.05 (-0.62, 0.53)	0.30
Received chemotherapy	No	AA	244	17	1.00	1.00		
		AG+GG	180	5	0.41 (0.15, 1.11)	0.41 (0.15, 1.11)		
	Yes	AA	136	19	1.00	1.95 (1.01, 3.75)		
		AG+GG	108	13	0.87 (0.43, 1.76)	1.70 (0.83, 3.50)	0.34 (-1.02, 1.70)	0.21
Received radiation	No	AA	149	14	1.00	1.00		
		AG+GG	121	6	0.55 (0.21, 1.42)	0.54 (0.21, 1.42)		
	Yes	AA	232	22	1.00	1.02 (0.52, 1.99)		
		AG+GG	168	12	0.76 (0.38, 1.54)	0.78 (0.36, 1.68)	0.21 (-0.64, 1.07)	0.58
Received hormone therapy	No	AA	144	13	1.00	1.00		
		AG+GG	117	7	0.67 (0.27, 1.69)	0.67 (0.27, 1.68)		
	Yes	AA	236	19	1.00	0.88 (0.43, 1.78)		
		AG+GG	169	11	0.82 (0.39, 1.73)	0.72 (0.32, 1.61)	0.17 (-0.71, 1.06)	0.73
NSAID Use	No	AA	380	54	1.00	1.00		
		AG+GG	300	35	0.84 (0.55, 1.28)	0.83 (0.55, 1.28)		
	Yes	AA	127	20	1.00	1.10 (0.66, 1.84)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		AG+GG	54	8	0.95 (0.42, 2.15)	1.04 (0.50, 2.19)	0.11 (-0.84, 1.05)	0.79
Weight gain from age 20 (kg)*	No change or loss	AA	41	7	1.00	1.00		
		AG+GG	37	6	0.97 (0.33, 2.89)	0.96 (0.32, 2.87)		
	3.01-8.15	AA	126	17	1.00	0.84 (0.35, 2.05)		
		AG+GG	95	10	0.78 (0.36, 1.71)	0.66 (0.25, 1.74)	-0.15 (-1.43, 1.14)	
	≥ 8.16	AA	161	33	1.00	1.22 (0.54, 2.76)		0.75
		AG+GG	106	13	0.64 (0.34, 1.21)	0.78 (0.31, 1.95)	-0.40 (-1.83, 1.02)	
BMI at reference (kg/m^2)	< 25.00	AA	239	26	1.00	1.00		
		AG+GG	177	18	0.93 (0.51, 1.70)	0.93 (0.51, 1.70)		
	25.00-29.99	AA	166	24	1.00	1.30 (0.75, 2.26)		
		AG+GG	119	12	0.72 (0.36, 1.44)	0.93 (0.47, 1.85)	-0.30 (-1.29, 0.70)	
	≥ 30.00	AA	119	28	1.00	2.02 (1.19, 3.45)		0.98
		AG+GG	71	14	0.88 (0.46, 1.66)	1.76 (0.92, 3.37)	-0.19 (-1.54, 1.15)	
rs1800206								
Age at diagnosis (years)	< 45	CC	123	9	1.00	1.00		
		CG+GG	16	2	1.67 (0.36, 7.75)	1.72 (0.37, 7.96)		
	45-54	CC	229	27	1.00	1.60 (0.75, 3.39)		
		CG+GG	22	2	0.81 (0.19, 3.39)	1.29 (0.28, 5.98)	-1.03 (-4.34, 2.29)	
	55-64	CC	207	20	1.00	1.33 (0.60, 2.91)		
		CG+GG	27	1	0.40 (0.05, 2.96)	0.52 (0.07, 4.13)	-1.52 (-4.65, 1.61)	0.90

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	≥ 65	CC	253	59	1.00	2.98 (1.48, 6.01)		
		CG+GG	32	5	0.73 (0.29, 1.82)	2.17 (0.73, 6.47)	-1.53 (-5.08, 2.02)	
Menopausal Status	Pre-menopausal	CC	271	26	1.00	1.00		
		CG+GG	27	4	1.53 (0.53, 4.39)	1.48 (0.52, 4.21)		
	Post-menopausal	CC	524	86	1.00	1.60 (1.05, 2.44)		
		CG+GG	67	6	0.58 (0.25, 1.32)	0.93 (0.38, 2.23)	-1.15 (-2.99, 0.68)	0.17
Stage	In-situ	CC	147	7	1.00	1.00		
		CG+GG	25	0	---	---		
	Invasive	CC	665	108	1.00	3.25 (1.51, 6.99)		
		CG+GG	72	10	0.90 (0.47, 1.72)	2.92 (1.11, 7.67)	0.67 (-1.26, 2.60)	0.17
ER status	ER-	CC	114	29	1.00	1.00		
		CG+GG	12	7	2.07 (0.91, 4.73)	2.15 (0.94, 4.91)		
	ER+	CC	404	49	1.00	0.50 (0.31, 0.78)		
		CG+GG	49	2	0.37 (0.09, 1.50)	0.18 (0.04, 0.76)	-1.46 (-3.31, 0.39)	0.02
PR status	PR-	CC	154	44	1.00	1.00		
		CG+GG	19	7	1.28 (0.58, 2.85)	1.29 (0.58, 2.86)		
	PR+	CC	364	34	1.00	0.35 (0.22, 0.55)		
		CG+GG	42	2	0.55 (0.13, 2.30)	0.19 (0.05, 0.79)	-0.44 (-1.53, 0.64)	0.28
Received chemotherapy	No	CC	393	20	1.00	1.00		
		CG+GG	38	2	1.09 (0.25, 4.65)	1.08 (0.25, 4.64)		
	Yes	CC	215	30	1.00	2.63 (1.50, 4.64)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		CG+GG	37	4	0.81 (0.28, 2.29)	2.14 (0.73, 6.27)	-0.58 (-3.41, 2.26)	0.76
Received radiation	No	CC	246	19	1.00	1.00		
		CG+GG	29	2	0.96 (0.21, 3.93)	0.92 (0.22, 3.97)		
	Yes	CC	364	31	1.00	1.10 (0.62, 1.94)		
		CG+GG	46	4	1.08 (0.38, 3.07)	1.18 (0.40, 3.46)	0.15 (-1.65, 1.96)	0.87
Received hormone therapy	No	CC	238	18	1.00	1.00		
		CG+GG	28	3	1.41 (0.42, 4.79)	1.45 (0.43, 4.93)		
	Yes	CC	368	29	1.00	1.02 (0.57, 1.84)		
		CG+GG	47	3	0.87 (0.26, 2.85)	0.87 (0.26, 2.97)	-0.60 (-2.69, 1.49)	0.54
NSAID Use	No	CC	616	83	1.00	1.00		
		CG+GG	75	7	0.74 (0.34, 1.59)	0.74 (0.34, 1.59)		
	Yes	CC	168	27	1.00	1.19 (0.77, 1.83)		
		CG+GG	18	3	1.03 (0.31, 3.39)	1.22 (0.38, 3.86)	0.30 (-1.26, 1.85)	0.65
Weight gain from age 20 (kg)*	No change or loss	CC	74	12	1.00	1.00		
		CG+GG	4	1	1.69 (0.22, 13.05)	1.66 (0.22, 12.88)		
	3.01-8.15	CC	200	26	1.00	0.86 (0.43, 1.72)		
		CG+GG	28	3	0.86 (0.26, 2.84)	0.72 (0.20, 2.58)	-0.79 (-4.34, 2.75)	
	≥ 8.16	CC	240	44	1.00	1.17 (0.61, 2.23)		0.45
		CG+GG	34	2	0.34 (0.08, 1.41)	0.41 (0.09, 1.82)	-1.43 (-5.04, 2.19)	
BMI at reference (kg/m^2)	< 25.00	CC	391	41	1.00	1.00		
		CG+GG	36	3	0.82 (0.25, 2.65)	0.82 (0.25, 2.63)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	25.00-29.99	CC	249	35	1.00	1.32 (0.84, 2.07)		
		CG+GG	41	4	0.73 (0.26, 2.04)	0.96 (0.34, 2.67)	-0.18 (-1.61, 1.26)	
	≥ 30.00	CC	172	39	1.00	2.04 (1.32, 3.16)		0.99
		CG+GG	20	3	0.74 (0.23, 2.38)	1.49 (0.46, 4.80)	-0.37 (-2.43, 1.70)	
rs4253623								
Age at diagnosis (years)	< 45	AA	113	10	1.00	1.00		
		AG+GG	27	4	1.65 (0.52, 5.27)	1.63 (0.51, 5.20)		
	45-54	AA	197	22	1.00	1.27 (0.60, 2.68)		
		AG+GG	56	7	1.10 (0.47, 2.58)	1.39 (0.53, 3.64)	-0.51 (-2.70, 1.67)	
	55-64	AA	182	15	1.00	0.94 (0.42, 2.09)		
		AG+GG	55	5	1.10 (0.40, 3.03)	1.05 (0.36, 3.06)	-0.52 (-2.65, 1.60)	0.75
	≥ 65	AA	222	50	1.00	2.41 (1.22, 4.74)		
		AG+GG	69	12	0.79 (0.42, 1.49)	1.91 (0.83, 4.42)	-1.13 (-3.56, 1.30)	
Menopausal Status	Pre-menopausal	AA	235	24	1.00	1.00		
		AG+GG	65	8	1.18 (0.53, 2.63)	1.14 (0.52, 2.51)		
	Post-menopausal	AA	461	72	1.00	1.44 (0.93, 2.25)		
		AG+GG	138	18	0.85 (0.51, 1.43)	1.23 (0.68, 2.23)	-0.36 (-1.50, 0.79)	0.51
Stage	In-situ	AA	135	6	1.00	1.00		
		AG+GG	42	1	0.54 (0.07, 4.50)	0.54 (0.07, 4.48)		
	Invasive	AA	579	91	1.00	3.38 (1.48, 7.73)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		AG+GG	165	27	1.04 (0.68, 1.60)	3.52 (1.45, 8.52)	0.60 (-1.27, 2.46)	0.52
ER status	ER-	AA	98	27	1.00	1.00		
		AG+GG	30	8	0.97 (0.44, 2.13)	0.96 (0.44, 2.12)		
	ER+	AA	343	41	1.00	0.45 (0.28, 0.74)		
		AG+GG	112	12	0.89 (0.47, 1.70)	0.41 (0.21, 0.80)	-0.01 (-0.83, 0.80)	0.89
PR status	PR-	AA	133	40	1.00	1.00		
		AG+GG	42	10	0.84 (0.42, 1.67)	0.83 (0.42, 1.67)		
	PR+	AA	308	28	1.00	0.33 (0.20, 0.54)		
		AG+GG	100	10	1.07 (0.52, 2.20)	0.35 (0.18, 0.71)	0.19 (-0.44, 0.82)	0.62
Received chemotherapy	No	AA	332	16	1.00	1.00		
		AG+GG	106	5	0.97 (0.36, 2.65)	0.97 (0.36, 2.66)		
	Yes	AA	198	24	1.00	2.43 (1.29, 4.58)		
		AG+GG	54	11	1.62 (0.79, 3.31)	3.94 (1.83, 8.49)	1.53 (-1.15, 4.21)	0.41
Received radiation	No	AA	205	16	1.00	1.00		
		AG+GG	68	5	0.94 (0.34, 2.56)	0.94 (0.34, 2.56)		
	Yes	AA	327	24	1.00	0.94 (0.50, 1.77)		
		AG+GG	92	11	1.61 (0.79, 3.29)	1.50 (0.70, 3.24)	0.63 (-0.66, 1.91)	0.39
Received hormone therapy	No	AA	211	15	1.00	1.00		
		AG+GG	58	5	1.21 (0.44, 3.33)	1.22 (0.44, 3.35)		
	Yes	AA	318	23	1.00	1.00 (0.52, 1.92)		
		AG+GG	101	9	1.22 (0.57, 2.65)	1.22 (0.54, 2.79)	0.00 (-1.48, 1.49)	>0.99

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
NSAID Use	No	AA	551	73	1.00	1.00		
		AG+GG	146	17	0.89 (0.53, 1.51)	0.89 (0.53, 1.51)		
	Yes	AA	137	21	1.00	1.15 (0.71, 1.88)		
		AG+GG	52	10	1.21 (0.57, 2.57)	1.40 (0.72, 2.72)	0.36 (-0.74, 1.45)	0.51
Weight gain from age 20 (kg)*	No change or loss	AA	63	10	1.00	1.00		
		AG+GG	16	3	1.29 (0.36, 4.70)	1.30 (0.35, 4.75)		
	3.01-8.15	AA	184	24	1.00	0.89 (0.42, 1.90)		
		AG+GG	51	4	0.62 (0.21, 1.78)	0.55 (0.17, 1.78)	-0.64 (-2.55, 1.27)	
	≥ 8.16	AA	206	34	1.00	1.10 (0.54, 2.25)		>0.99
		AG+GG	67	11	1.00 (0.51, 1.98)	1.10 (0.46, 2.60)	-0.30 (-2.15, 1.54)	
BMI at reference (kg/m^2)	< 25.00	AA	339	31	1.00	1.00		
		AG+GG	94	13	1.51 (0.79, 2.88)	1.50 (0.79, 2.87)		
	25.00-29.99	AA	228	32	1.00	1.52 (0.93, 2.49)		
		AG+GG	69	7	0.71 (0.32, 1.62)	1.09 (0.48, 2.47)	-0.94 (-2.37, 0.50)	
	≥ 30.00	AA	147	34	1.00	2.37 (1.46, 3.85)		0.45
		AG+GG	44	8	0.82 (0.38, 1.77)	1.94 (0.89, 4.21)	-0.93 (-2.82, 0.95)	
rs4253699								
Age at diagnosis (years)	< 45	TT	83	8	1.00	1.00		
		CT+CC	58	6	1.07 (0.37, 3.07)	1.08 (0.37, 3.11)		
	45-54	TT	152	19	1.00	1.30 (0.57, 2.96)		
		CT+CC	96	10	0.84 (0.39, 1.82)	1.10 (0.43, 2.78)	-0.28 (-1.81, 1.25)	

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	55-64	TT	151	12	1.00	0.84 (0.34, 2.06)		
		CT+CC	83	8	1.21 (0.49, 2.95)	1.02 (0.38, 2.71)	0.09 (-1.27, 1.46)	>0.99
	≥ 65	TT	163	36	1.00	2.18 (1.01, 4.69)		
		CT+CC	127	28	1.01 (0.62, 1.65)	2.20 (1.00, 4.83)	-0.06 (-1.63, 1.51)	
Menopausal Status	Pre-menopausal	TT	178	21	1.00	1.00		
		CT+CC	120	11	0.80 (0.38, 1.65)	0.77 (0.38, 1.57)		
	Post-menopausal	TT	360	51	1.00	1.16 (0.71, 1.88)		
		CT+CC	235	41	1.22 (0.81, 1.84)	1.41 (0.85, 2.34)	0.49 (-0.24, 1.22)	0.31
Stage	In-situ	TT	114	7	1.00	1.00		
		CT+CC	60	0	---	---		
	Invasive	TT	435	68	1.00	2.45 (1.13, 5.34)		
		CT+CC	304	52	1.10 (0.76, 1.57)	2.69 (1.22, 5.92)	1.23 (0.28, 2.18)	0.01
ER status	ER-	TT	82	19	1.00	1.00		
		CT+CC	45	18	1.59 (0.83, 3.03)	1.63 (0.85, 3.10)		
	ER+	TT	270	31	1.00	0.51 (0.29, 0.90)		
		CT+CC	182	22	1.05 (0.61, 1.82)	0.54 (0.29, 0.99)	-0.60 (-1.68, 0.48)	0.31
PR status	PR-	TT	112	29	1.00	1.00		
		CT+CC	61	23	1.38 (0.80, 2.39)	1.39 (0.80, 2.41)		
	PR+	TT	240	21	1.00	0.36 (0.20, 0.63)		
		CT+CC	166	17	1.17 (0.62, 2.22)	0.42 (0.23, 0.76)	-0.33 (-1.12, 0.46)	0.69
Received	No	TT	275	13	1.00	1.00		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
chemotherapy		CT+CC	164	9	1.17 (0.50, 2.74)	1.17 (0.50, 2.73)		
	Yes	TT	138	20	1.00	2.93 (1.46, 5.88)		
		CT+CC	114	14	0.86 (0.44, 1.71)	2.54 (1.19, 5.40)	-0.56 (-2.73, 1.61)	0.59
Received radiation	No	TT	170	10	1.00	1.00		
		CT+CC	105	11	1.76 (0.75, 4.13)	1.76 (0.75, 4.15)		
	Yes	TT	244	23	1.00	1.58 (0.75, 3.31)		
		CT+CC	174	12	0.75 (0.37, 1.51)	1.18 (0.51, 2.73)	-1.16 (-3.09, 0.77)	0.13
Received hormone therapy	No	TT	165	11	1.00	1.00		
		CT+CC	104	10	1.42 (0.61, 3.35)	1.44 (0.61, 3.38)		
	Yes	TT	246	19	1.00	1.13 (0.54, 2.37)		
		CT+CC	174	12	0.91 (0.44, 1.88)	1.03 (0.45, 2.33)	-0.54 (-2.02, 0.95)	0.43
NSAID Use	No	TT	422	57	1.00	1.00		
		CT+CC	271	35	0.97 (0.63, 1.47)	0.96 (0.63, 1.47)		
	Yes	TT	109	17	1.00	1.14 (0.66, 1.96)		
		CT+CC	78	13	1.07 (0.52, 2.21)	1.22 (0.67, 2.23)	0.12 (-0.83, 1.06)	0.81
Weight gain from age 20 (kg)*	No change or loss	TT	44	9	1.00	1.00		
		CT+CC	34	4	0.63 (0.19, 2.05)	0.60 (0.19, 1.97)		
	3.01-8.15	TT	144	13	1.00	0.49 (0.21, 1.15)		
		CT+CC	89	15	1.83 (0.87, 3.84)	0.88 (0.39, 2.03)	0.79 (-0.01, 1.59)	
	≥ 8.16	TT	164	29	1.00	0.92 (0.43, 1.94)		>0.99

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		CT+CC	108	18	0.93 (0.52, 1.68)	0.86 (0.39, 1.93)	0.34 (-0.56, 1.24)	
BMI at reference (kg/m^2)	< 25.00	TT	260	28	1.00	1.00		
		CT+CC	170	18	0.99 (0.55, 1.78)	0.99 (0.55, 1.78)		
	25.00-29.99	TT	174	24	1.00	1.26 (0.73, 2.17)		
		CT+CC	119	14	0.87 (0.45, 1.69)	1.10 (0.58, 2.08)	-0.15 (-1.14, 0.84)	
	≥ 30.00	TT	115	23	1.00	1.77 (1.02, 3.08)		0.80
		CT+CC	75	20	1.30 (0.72, 2.37)	2.32 (1.31, 4.12)	0.56 (-0.78, 1.90)	
rs4253755								
Age at diagnosis (years)	< 45	GG	108	11	1.00	1.00		
		AG+AA	32	3	0.91 (0.25, 3.25)	0.90 (0.25, 3.23)		
	45-54	GG	195	21	1.00	1.06 (0.51, 2.20)		
		AG+AA	57	7	1.13 (0.48, 2.66)	1.20 (0.47, 3.10)	0.24 (-1.26, 1.74)	
	55-64	GG	180	19	1.00	1.04 (0.49, 2.18)		
		AG+AA	54	2	0.37 (0.09, 1.57)	0.38 (0.08, 1.71)	-0.56 (-2.11, 0.98)	0.92
	≥ 65	GG	223	46	1.00	1.93 (1.00, 3.72)		
		AG+AA	72	20	1.32 (0.78, 2.23)	2.55 (1.22, 5.31)	0.72 (-0.90, 2.34)	
Menopausal Status	Pre-menopausal	GG	226	27	1.00	1.00		
		AG+AA	72	5	0.60 (0.23, 1.55)	0.60 (0.23, 1.56)		
	Post-menopausal	GG	464	68	1.00	1.23 (0.80, 1.90)		
		AG+AA	138	27	1.31 (0.84, 2.05)	1.62 (0.96, 2.73)	0.78 (-0.08, 1.65)	0.12
Stage	In-situ	GG	136	5	1.00	1.00		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		AG+AA	38	2	1.45 (0.28, 7.45)	1.43 (0.28, 7.35)		
	Invasive	GG	570	92	1.00	4.18 (1.70, 10.29)		
		AG+AA	177	30	1.05 (0.70, 1.58)	4.38 (1.70, 11.30)	-0.22 (-3.09, 2.65)	0.73
ER status	ER-	GG	107	31	1.00	1.00		
		AG+AA	23	7	1.03 (0.45, 2.33)	1.03 (0.45, 2.34)		
	ER+	GG	349	40	1.00	0.42 (0.26, 0.67)		
		AG+AA	109	13	1.05 (0.56, 1.96)	0.44 (0.23, 0.83)	-0.01 (-0.90, 0.87)	>0.99
PR status	PR-	GG	145	45	1.00	1.00		
		AG+AA	34	9	0.86 (0.42, 1.77)	0.87 (0.42, 1.77)		
	PR+	GG	311	26	1.00	0.29 (0.18, 0.48)		
		AG+AA	98	11	1.33 (0.66, 2.70)	0.39 (0.20, 0.76)	0.23 (-0.43, 0.89)	0.40
Received chemotherapy	No	GG	323	15	1.00	1.00		
		AG+AA	108	7	1.39 (0.57, 3.42)	1.39 (0.57, 3.42)		
	Yes	GG	201	33	1.00	3.36 (1.83, 6.19)		
		AG+AA	54	3	0.36 (0.11, 1.17)	1.20 (0.35, 4.14)	-2.56 (-5.29, 0.17)	0.06
Received radiation	No	GG	211	13	1.00	1.00		
		AG+AA	63	8	2.00 (0.83, 4.84)	1.99 (0.83, 4.81)		
	Yes	GG	314	35	1.00	1.76 (0.93, 3.33)		
		AG+AA	100	2	0.19 (0.05, 0.79)	0.33 (0.08, 1.48)	-2.42 (-4.85, 0.01)	<0.01
Received hormone therapy	No	GG	205	15	1.00	1.00		
		AG+AA	58	5	1.18 (0.43, 3.24)	1.17 (0.42, 3.21)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	Yes	GG	317	30	1.00	1.25 (0.67, 2.33)		
		AG+AA	104	4	0.42 (0.15, 1.19)	0.53 (0.18, 1.59)	-0.89 (-2.45, 0.67)	0.17
NSAID Use	No	GG	539	71	1.00	1.00		
		AG+AA	159	21	1.00 (0.62, 1.63)	1.00 (0.62, 1.63)		
	Yes	GG	141	23	1.00	1.22 (0.76, 1.95)		
		AG+AA	47	9	1.17 (0.54, 2.53)	1.43 (0.71, 2.85)	0.21 (-0.95, 1.36)	0.74
Weight gain from age 20 (kg)*	No change or loss	GG	63	9	1.00	1.00		
		AG+AA	18	5	1.79 (0.60, 5.33)	1.80 (0.60, 5.36)		
	3.01-8.15	GG	180	18	1.00	0.73 (0.33, 1.64)		
		AG+AA	52	11	2.06 (0.97, 4.35)	1.51 (0.62, 3.66)	-0.02 (-1.98, 1.94)	
	≥ 8.16	GG	210	40	1.00	1.34 (0.65, 2.78)		0.14
		AG+AA	66	8	0.64 (0.30, 1.37)	0.88 (0.34, 2.28)	-1.26 (-3.55, 1.02)	
BMI at reference (kg/m^2)	< 25.00	GG	333	31	1.00	1.00		
		AG+AA	100	15	1.57 (0.85, 2.91)	1.57 (0.85, 2.91)		
	25.00-29.99	GG	228	33	1.00	1.52 (0.93, 2.48)		
		AG+AA	67	7	0.75 (0.33, 1.68)	1.13 (0.50, 2.56)	-0.96 (-2.41, 0.48)	
	≥ 30.00	GG	145	33	1.00	2.30 (1.41, 3.76)		0.49
		AG+AA	48	10	0.91 (0.45, 1.84)	2.11 (1.04, 4.31)	-0.76 (-2.59, 1.07)	
rs4253760								
Age at diagnosis (years)	< 45	TT	91	8	1.00	1.00		
		GT+GG	43	5	1.29 (0.42, 3.96)	1.29 (0.42, 3.94)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	45-54	TT	169	17	1.00	1.13 (0.49, 2.62)		
		GT+GG	78	10	1.31 (0.60, 2.86)	1.48 (0.58, 3.75)	0.06 (-1.62, 1.75)	
	55-64	TT	159	15	1.00	1.07 (0.46, 2.53)		
		GT+GG	72	5	0.75 (0.27, 2.07)	0.81 (0.26, 2.47)	-0.56 (-2.35, 1.24)	0.96
	≥ 65	TT	176	40	1.00	2.41 (1.13, 5.15)		
		GT+GG	108	24	1.01 (0.61, 1.67)	2.42 (1.09, 5.39)	-0.28 (-2.17, 1.61)	
Menopausal Status	Pre-menopausal	TT	199	21	1.00	1.00		
		GT+GG	90	9	0.97 (0.44, 2.11)	0.96 (0.45, 2.08)		
	Post-menopausal	TT	383	58	1.00	1.41 (0.87, 2.29)		
		GT+GG	203	34	1.12 (0.73, 1.71)	1.58 (0.93, 2.69)	0.21 (-0.75, 1.17)	0.73
Stage	In-situ	TT	111	5	1.00	1.00		
		GT+GG	60	2	0.75 (0.15, 3.88)	0.74 (0.14, 3.83)		
	Invasive	TT	484	75	1.00	3.27 (1.32, 8.08)		
		GT+GG	241	42	1.14 (0.78, 1.67)	3.73 (1.48, 9.44)	0.72 (-0.99, 2.44)	0.61
ER status	ER-	TT	83	20	1.00	1.00		
		GT+GG	43	17	1.57 (0.82, 3.00)	1.59 (0.83, 3.03)		
	ER+	TT	304	35	1.00	0.50 (0.29, 0.86)		
		GT+GG	141	16	1.57 (0.82, 3.00)	0.50 (0.26, 0.97)	-0.58 (-1.65, 0.49)	0.32
PR status	PR-	TT	114	32	1.00	1.00		
		GT+GG	59	19	1.17 (0.66, 2.06)	1.17 (0.66, 2.06)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
	PR+	TT	273	23	1.00	0.33 (0.19, 0.56)		
		GT+GG	125	14	1.35 (0.69, 2.62)	0.44 (0.23, 0.82)	-0.06 (-0.75, 0.64)	0.75
Received chemotherapy	No	TT	287	14	1.00	1.00		
		GT+GG	134	8	1.25 (0.53, 2.99)	1.25 (0.52, 2.97)		
	Yes	TT	160	26	1.00	3.16 (1.65, 6.05)		
		GT+GG	83	7	0.55 (0.24, 1.28)	1.74 (0.70, 4.32)	-1.66 (-4.08, 0.75)	0.18
Received radiation	No	TT	181	12	1.00	1.00		
		GT+GG	82	8	1.48 (0.60, 3.61)	1.46 (0.60, 3.58)		
	Yes	TT	267	28	1.00	1.55 (0.79, 3.04)		
		GT+GG	136	7	0.52 (0.23, 1.19)	0.80 (0.32, 2.04)	-1.21 (-3.02, 0.60)	0.09
Received hormone therapy	No	TT	170	12	1.00	1.00		
		GT+GG	90	7	1.11 (0.44, 2.82)	1.10 (0.43, 2.80)		
	Yes	TT	275	26	1.00	1.29 (0.65, 2.56)		
		GT+GG	127	6	0.53 (0.22, 1.28)	0.68 (0.26, 1.82)	-0.71 (-2.18, 0.75)	0.26
NSAID Use	No	TT	463	59	1.00	1.00		
		GT+GG	219	29	1.06 (0.68, 1.65)	1.06 (0.68, 1.65)		
	Yes	TT	115	19	1.00	1.27 (0.76, 2.14)		
		GT+GG	65	12	1.14 (0.55, 2.35)	1.45 (0.78, 2.69)	0.12 (-0.98, 1.21)	0.87
Weight gain from age 20 (kg)*	No change or loss	TT	55	9	1.00	1.00		
		GT+GG	24	4	1.02 (0.31, 3.30)	0.99 (0.31, 3.24)		
	3.01-8.15	TT	143	16	1.00	0.72 (0.32, 1.64)		

TABLE D.29. Stratified and common referent hazard ratios and interaction contrast ratios and multiplicative p-values for select characteristics

Characteristic	Status	Genotype	Alive (N)	Dead (N)	Stratum HR (95% CI)	Referent HR (95% CI)	ICR (95% CI)	P-Value†
		GT+GG	80	12	1.35 (0.64, 2.86)	0.97 (0.41, 2.30)	0.25 (-1.02, 1.52)	
	≥ 8.16	TT	175	33	1.00	1.16 (0.56, 2.44)		0.88
		GT+GG	96	15	0.84 (0.46, 1.55)	1.00 (0.43, 2.28)	-0.16 (-1.56, 1.23)	
BMI at reference (kg/m^2)	< 25.00	TT	285	29	1.00	1.00		
		GT+GG	136	14	1.02 (0.54, 1.93)	1.02 (0.54, 1.93)		
	25.00-29.99	TT	186	25	1.00	1.29 (0.75, 2.20)		
		GT+GG	102	15	1.13 (0.60, 2.14)	1.45 (0.78, 2.70)	0.14 (-0.94, 1.23)	
	≥ 30.00	TT	124	26	1.00	1.94 (1.14, 3.30)		0.96
		GT+GG	63	15	1.14 (0.61, 2.16)	2.23 (1.20, 4.17)	0.27 (-1.21, 1.75)	

*weight gain models are adjusted for BMI at age 20
†multiplicative p-value, calculated using likelihood ratio tests to compare models with interaction term(s) to those without

TABLE D.30. Distribution of select survival covariates by genotyping information

Variable	All Cases		Alive		Dead	
	N_T (%)*	N_G (%)*	N_T (%)	N_G (%)	N_T (%)	N_G (%)
	1508	1073	1310	941	198	132
Age						
< 45 years old	220 (15)	160 (15)	196 (15)	146 (16)	24 (12)	14 (11)
45-54 years old	397 (26)	285 (27)	359 (27)	256 (27)	38 (19)	29 (22)
55-64 years old	372 (25)	264 (25)	336 (26)	241 (26)	36 (18)	23 (17)
≥ 65 years old	519 (34)	364 (34)	419 (32)	298 (32)	100 (51)	66 (50)
ER/PR status						
ER-/PR-	212 (14)	138 (13)	160 (12)	103 (11)	52 (26)	35 (27)
ER-/PR+	52 (3)	33 (3)	44 (3)	29 (3)	8 (4)	4 (3)
ER+/PR-	143 (9)	99 (9)	115 (9)	79 (8)	28 (14)	20 (15)
ER+/PR+	583 (39)	421 (39)	522 (40)	387 (41)	61 (31)	34 (26)
Missing	518 (34)	382 (36)	469 (36)	343 (36)	49 (25)	39 (30)
Stage						
In situ	235 (16)	185 (17)	22 (17)	178 (19)	10 (5)	7 (5)
Invasive	1273 (84)	888 (83)	1085 (83)	763 (81)	188 (95)	125 (95)
Tumor Size						
≤ 2 cm	466 (31)	352 (33)	444 (34)	336 (36)	22 (11)	16 (12)
2-5 cm	102 (7)	79 (7)	94 (7)	74 (8)	8 (4)	5 (4)
>5 cm	11 (1)	6 (1)	10 (1)	5 (1)	1 (1)	1 (1)
Missing	929 (62)	636 (59)	762 (58)	526 (56)	167 (84)	110 (83)
Nodal involvement						
No positive nodes	462 (31)	355 (33)	440 (34)	337 (36)	23 (12)	18 (14)
Positive ipsilateral nodes or positive regional nodes	134 (9)	98 (9)	125 (10)	93 (10)	9 (5)	5 (4)
Positive distant nodes	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Missing	911 (60)	620 (58)	745 (57)	511 (54)	166 (84)	109 (83)
Chemotherapy at follow-up						
Yes	423 (28)	296 (28)	372 (28)	259 (28)	51 (26)	37 (28)
No	599 (40)	469 (44)	565 (43)	446 (47)	34 (17)	23 (17)
Missing	486 (32)	308 (29)	373 (28)	236 (25)	113 (57)	72 (55)
Hormone Therapy at follow-up						
Yes	616 (41)	463 (43)	567 (43)	429 (46)	49 (25)	34 (26)
No	393 (26)	295 (27)	365 (28)	274 (29)	28 (14)	21 (16)
Missing	499 (33)	315 (29)	378 (29)	238 (25)	121 (61)	77 (58)
Radiation Therapy at follow-up						
Yes	625 (41)	463 (43)	575 (44)	425 (45)	50 (25)	38 (29)
No	401 (27)	304 (28)	365 (28)	282 (30)	36 (18)	22 (17)
Missing	482 (32)	306 (29)	370 (28)	234 (25)	112 (57)	72 (55)
BMI at reference (kg/m^2)						
< 24.99	700 (46)	490 (46)	623 (48)	442 (47)	77 (39)	48 (36)
25.00-29.99	476 (32)	341 (32)	416 (32)	301 (32)	60 (30)	40 (30)
≥ 30.00	332 (22)	242 (23)	271 (21)	198 (21)	61 (32)	44 (33)
Weight gain from age 20 (kg)						
Maintenance or loss	139 (9)	96 (9)	117 (9)	82 (9)	22 (11)	14 (11)

3.01-8.15 gain	382 (25)	266 (25)	338 (26)	236 (25)	44 (22)	30 (23)
≥ 8.16 gain	465 (31)	329 (31)	390 (30)	280 (30)	75 (38)	49 (37)
Missing	522 (35)	382 (36)	465 (36)	343 (36)	57 (29)	39 (30)

*N_T = all cases, N_G=cases for whom genotyping data is available

TABLE D.31. Model building procedure for SNP rs135542 (AA vs. AG+GG), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	939	939	940	970	972
Main Exposure: **rs135542**	-0.27543 (0.51975)	0.19764 (0.23170)	0.21298 (0.23029)	0.13841 (0.22375)	0.12633 (0.22361)
FAMHX	0.36668 (0.27154)	0.36679 (0.27155)	0.36127 (0.27127)	REMOVED	REMOVED
JEWISH	-0.16823 (0.33039)	-0.16972 (0.33040)	REMOVED	REMOVED	REMOVED
RACE_DIC	0.71142 (0.40158)	0.71031 (0.40158)	0.73617 (0.39903)	0.70723 (0.37265)	REMOVED
Interactions:					
SNP1_DUR	0.0004421 (0.0004326)	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1024.382	1025.439	1025.845	1100.060	1103.398
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed FAMHX	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1024.382–1025.439= 1.057; p=0.30	---	---	---
Change in estimate			|0.19764-0.21298| = 0.0153	|0.19764-0.13841| = 0.05923	|0.19764-0.12633| = 0.07131
Compare to:		Full Model	Model 2	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Fam history is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.32. Model building procedure for SNP rs1800206 (CC vs. CG+GG), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	962	962	991	992	993
Main Exposure: **rs135542**	0.47436 (0.78157)	-0.01875 (0.37327)	-0.10467 (0.37205)	-0.08129 (0.37182)	-0.09386 (0.37173)
FAMHX	0.34479 (0.27027)	0.34340 (0.27025)	REMOVED	REMOVED	REMOVED
JEWISH	-0.46077 (0.35674)	-0.45996 (0.35674)	-0.50418 (0.35500)	REMOVED	REMOVED
RACE_DIC	0.73319 (0.37739)	0.73023 (0.37736)	0.68922 (0.35490)	0.75766 (0.35284)	REMOVED
Interactions:					
SNP2_DUR	-0.0004962 (0.0007306)	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1066.969	1067.451	1140.507	1142.961	1146.928
Notes		Evaluate EMM with the likelihood ratio test*	Removed FAMHX	Removed JEWISH	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1067.451– 1066.969= 0.482; p=0.49	---	---	---
Change in estimate			|-0.01875- -0.10467| = 0.08592	|-0.01875- -0.08129| = 0.06254	|-0.01875- -0.09386| = 0.07511
Compare to:		Full Model	Model 2	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Fam history is not a confounder and is removed from the model	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.33. Model building procedure for SNP rs4253623 (AA vs. AG+GG), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	972	972	973	1003	1005
Main Exposure: **rs135542**	0.16276 (0.60127)	-0.08152 (0.27405)	-0.06111 (0.27379)	0.00433 (0.26080)	0.00644 (0.26080)
FAMHX	0.36710 (0.27029)	0.36782 (0.27030)	0.34958 (0.27007)	REMOVED	REMOVED
JEWISH	-0.47746 (0.35688)	-0.47814 (0.35687)	REMOVED	REMOVED	REMOVED
RACE_DIC	0.70352 (0.37732)	0.70341 (0.37733)	0.76741 (0.37549)	0.73462 (0.35275)	REMOVED
Interactions:					
SNP3_DUR	-0.0002315 (0.0005164)	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1068.739	1068.942	1071.107	1145.133	1149.062
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed FAMHX	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1068.739–1068.942= 0.203; p=0.65	---	---	---
Change in estimate			|-0.08152--0.06111| = 0.02041	|-0.08152-0.00433| = 0.08585	|-0.08152-0.00644| = 0.087961
Compare to:		Full Model	Model 2	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.34. Model building procedure for SNP rs4253699 (TT vs. CT+CC), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	968	968	969	998	999
Main Exposure: **rs135542**	0.01237 (0.49998)	0.04439 (0.22618)	0.04491 (0.22616)	0.03811 (0.21951)	0.03812 (0.21950)
FAMHX	0.36929 (0.26431)	0.36952 (0.26430)	0.35475 (0.26404)	REMOVED	REMOVED
JEWISH	-0.39154 (0.34020)	-0.39170 (0.34019)	REMOVED	REMOVED	REMOVED
RACE_DIC	0.74518 (0.37693)	0.74544 (0.37692)	0.80005 (0.37493)	0.76219 (0.35226)	REMOVED
Interactions:					
SNP4_DUR	0.0000299 (0.0004167)	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1095.610	1095.615	1097.216	1171.229	1175.252
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed FAMHX	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1095.615–1095.610= 0.005; p=0.94	---	---	---
Change in estimate			|0.04439-0.04491| = 0.00052	|0.04439-0.03811| = 0.00628	|0.04439-0.03812| = 0.006271
Compare to:		Full Model	Model 2	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate < 0.10. Family history is not a confounder and is removed from the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.35. Model building procedure for SNP rs4253755 (GG vs. AG+AA), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	974	974	975	1005	977
Main Exposure: **rs4253755**	-0.35848 (0.63062)	-0.29349 (0.28837)	-0.28580 (0.28828)	-0.09506 (0.26701)	-0.32441 (0.28651)
FAMHX	0.32536 (0.26927)	0.32559 (0.26926)	0.31578 (0.26901)	REMOVED	0.28404 (0.26772)
JEWISH	-0.27586 (0.32665)	-0.27597 (0.32664)	REMOVED	REMOVED	REMOVED
RACE_DIC	0.51131 (0.40219)	0.51126 (0.40218)	0.55271 (0.40000)	0.57711 (0.37397)	REMOVED
Interactions:					
SNP4_DUR	0.0000610 (0.0005238)	REMOVED	REMOVED	REMOVED	REMOVED
-2 log L	1098.460	1098.473	1099.395	1174.094	1101.388
Notes		Evaluate EMM with the likelihood ratio test*	Removed JEWISH	Removed FAMHX	Removed RACE_DIC
Test of interaction term		χ^2 (1): 1098.473–1098.460= 0.013; p=0.91	---	---	---
Change in estimate			|-0.29349- -0.28580| = 0.00769	|-0.29349- -0.09506| = 0.19843	|-0.29349- -0.32441| = 0.03092
Compare to:		Full Model	Model 2	Model 2	Model 2
Conclusion:		Interaction term is not contributing to the model and will be removed.	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model	Change in estimate > 0.10. Family history is a confounder and will remain in the model	Change in estimate < 0.10. Race is not a confounder and is removed from the model

TABLE D.36. Model building procedure for SNP rs4253760 (TT vs. GT+GG), breast cancer specific mortality with continuous time interaction term

	Full Model β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)
Sample Size	945	945	975	946	948
Main Exposure: **rs4253760**	0.67005 (0.54552)	-0.28124 (0.25736)	0.84292 (0.52474)	0.72983 (0.54340)	0.74355 (0.54327)
FAMHX	0.24078 (0.28239)	0.23796 (0.28241)	REMOVED	0.20292 (0.28130)	0.18903 (0.28100)
JEWISH	-0.33446 (0.34136)	-0.33259 (0.34135)	-0.37835 (0.33952)	-0.37253 (0.33964)	REMOVED
RACE_DIC	0.65411 (0.40792)	0.65438 (0.40781)	0.61834 (0.38177)	REMOVED	REMOVED
Interactions:					
SNP4_DUR	-0.0009544 (0.0005104)	REMOVED	-0.00102 (0.0004849)	-0.0009557 (0.0005107)	-0.0009543 (0.0005106)
-2 log L	1035.496	1039.340	1107.41	1037.870	1039.520
Notes		Evaluate EMM with the likelihood ratio test*	Removed FAMHX	Removed RACE_DIC	Removed JEWISH
Test of interaction term		χ^2 (1): 1039.340–1035.496= 3.844; p=0.05	---	---	---
Change in estimate			|0.67005-0.84292| = 0.17287	|0.67005-0.72983| = 0.05978	|0.67005-0.74355| = 0.07355
Compare to:		Full Model	Full Model	Full Model	Full Model
Conclusion:		Interaction term is contributing to the model and will remain in the model.	Change in estimate >0.10. Family history is a confounder and will remain in the model	Change in estimate > 0.10. Race is not a confounder and is removed in the model	Change in estimate < 0.10. Jewish ethnicity is not a confounder and is removed from the model

TABLE D.37. Hazard ratios and frequencies by vital status for six SNPs in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n = 1073)

	Overall Mortality					**Breast Cancer-Specific Mortality**				
Genotype	**N†**	**MST* (days)**	**Log Rank P-value**	**HR (95% CI)**	**HR† (95% CI)**	**N†**	**MST* (days)**	**Log Rank P-value**	**HR (95% CI)**	**HR‡ (95% CI)**
rs135542										
AA	78/602	2044.9	0.29	1.00		45/569	2086.3	0.57	1.00	
AG+GG	44/411	2101.9		0.82		36/403	2120.3		1.14	
				(0.57, 1.19)					(0.73, 1.76)	
rs1800206										
CC	115/927	2080.8	0.41	1.00		76/888	2118.2	0.80	1.00	
CG+GG	10/107	1634.3		0.76		8/105	1616.7		0.91	
				(0.40, 1.46)					(0.44, 1.89)	
rs4263623										
AA	97/811	2087.0	0.99	1.00		65/779	2120.2	0.98	1.00	
AG+GG	28/235	1981.5		1.00		19/226	1953.3		1.01	
				(0.66, 1.52)					(0.60, 1.68)	
rs4253699										
TT	75/624	2086.4	0.79	1.00		51/600	2118.9	0.86	1.00	
CT+CC	52/416	2025.7		1.05		35/399	2061.9		1.04	
				(0.74, 1.50)					(0.68, 1.60)	
rs4253755										
GG	97/803	2087.1	0.71	1.00		68/774	2114.9	0.60	1.00	
AG+AA	32/247	1930.0		1.08		18/233	1900.3		0.87	
				(0.72, 1.61)					(0.52, 1.47)	
rs4253760										
TT	80/675	2095.1	0.60	1.00	1.00	57/652	2122.8	0.62	1.00	1.00
GT+GG	44/345	1923.8		1.10	2.25	25/326	1887.3		0.89	2.50
				(0.76, 1.59)	(1.00, 5.08)				(0.55, 1.42)	(0.90, 6.96)

*The mean survival time will be underestimated because largest survival time is censored.
†Sample size for cases/cohort
‡Adjusted for a continuous time interaction.

TABLE D.38. Hazard ratios and frequencies by vital status for haplotypes in relation to all-cause (n=132) and breast cancer specific mortality (n=88) mortality among a cohort of women with breast cancer (n = 1073)

							Overall Mortality			Breast Cancer Mortality		
Haplotype Number*	**rs135542**	**rs1800206**	**rs4253623**	**rs4253699**	**rs4253755**	**rs4253760**	**Alive Freq**	**Dead Freq**	**HR (95% CI)**	**Alive Freq**	**Dead Freq**	**HR (95% CI)**
1	A	C	A	C	A	G	0.06	0.07	1.01 (0.36, 2.86)	0.06	0.06	0.83 (0.21, 3.29)
2	A	C	A	C	A	T	0.03	0.03	0.93 (0.21, 4.14)	0.03	0.03	1.11 (0.19, 6.50)
3	A	C	A	C	G	G	0.02	0.00	0.31 (0.03, 3.76)	0.02	0.00	0.09 (0.00, 6.27)
4	A	C	A	C	G	T	0.05	0.08	2.11 (0.67, 6.58)	0.05	0.07	1.50 (0.30, 7.39)
5	A	C	A	T	A	G	0.03	0.03	0.90 (0.18, 4.53)	0.03	0.02	0.40 (0.03, 4.59)
6	A	C	A	T	G	G	0.03	0.06	1.77 (0.57, 5.46)	0.03	0.03	1.09 (0.20, 5.91)
7	A	C	A	T	G	T	0.40	0.38	0.79 (0.51, 1.23)	0.40	0.38	0.83 (0.48, 1.44)
8	A	C	G	T	G	T	0.10	0.12	0.98 (0.45, 2.10)	0.10	0.12	0.96 (0.37, 2.50)
9	A	G	A	C	G	G	0.02	0.02	0.77 (0.11, 5.21)	0.02	0.03	1.15 (0.12, 10.76)
10	A	G	A	C	G	T	0.01	0.00	0.19 (0.01, 7.21)	0.01	0.00	0.44 (0.01, 15.91)
11	G	C	A	C	G	T	0.01	0.01	0.12 (0.00, 6.45)	0.01	0.01	0.53 (0.01, 27.13)
12	G	C	A	T	G	T	0.18	0.17	0.69 (0.35, 1.36)	0.18	0.21	1.11 (0.52, 2.36)

*haplotypes with frequency ≥ 0.01

TABLE D.39. Hazard ratios and frequencies for all cause mortality by vital status for six SNPs in all cases verus invasive cases only

	All Cases			Invasive Cases Only		
Genotype	**N Cases/Cohort**	**HR (95% CI)**	**HR* (95% CI)**	**N Cases/Cohort**	**HR (95% CI)**	**HR* (95% CI)**
rs135542						
AA	78/602	1.00	1.00	74/492	1.00	1.00
AG+GG	44/411	0.82	0.80	41/348	0.77	0.66
		(0.57, 1.19)	(0.34, 1.86)		(0.53, 1.13)	(0.28, 1.59)
rs1800206						
CC	115/927	1.00	1.00	108/773	1.00	1.00
CG+GG	10/107	0.76	1.17	10/82	0.90	1.25
		(0.40, 1.46)	(0.30, 4.58)		(0.47, 1.72)	(0.32, 4.95)
rs4263623						
AA	97/811	1.00	1.00	91/670	1.00	1.00
AG+GG	28/235	1.00	1.54	27/192	1.04	1.20
		(0.66, 1.52)	(0.65, 3.68)		(0.68, 1.60)	(0.46, 3.10)
rs4253699						
TT	75/624	1.00	1.00	68/503	1.00	1.00
CT+CC	52/416	1.05	1.16	52/356	1.10	1.09
		(0.74, 1.50)	(0.45, 2.96)		(0.76, 1.57)	(0.49, 2.44)
rs4253755						
GG	97/803	1.00	1.00	92/662	1.00	1.00
AG+AA	32/247	1.08	1.18	30/207	1.05	1.42
		(0.72, 1.61)	(0.54, 2.60)		(0.70, 1.58)	(0.58, 3.47)
rs4253760						
TT	80/675	1.00	1.00	75/559	1.00	1.00
GT+GG	44/345	1.10	2.25	42/283	1.14	2.17
		(0.76, 1.59)	(1.00, 5.08)		(0.78, 1.67)	(0.95, 4.96)

*Adjusted for a continuous time interaction term; interaction term necessary only for rs4253760 based on likelihood ratio test

WORKS CITED

1. *The American Heritage dictionary of the English language*. 4th ed. 2000, Boston: Houghton Mifflin. xxxvii, 2074.

2. Stedman, T.L., *Stedman's medical dictionary*. 27th ed. 2000, Philadelphia: Lippincott Williams & Wilkins. xxxvi, 5, 122, 2098, 4 , 59 of plates.

3. Bray, G.A., *Medical consequences of obesity.* J Clin Endocrinol Metab, 2004. **89**(6): p. 2583-9.

4. Vainio, H., F. Bianchini, and Sadie Jenkins Harmon Collection., *Weight control and physical activity*. IARC handbooks of cancer prevention ; v. 6. 2002, Lyon: IARC Press. xii, 315.

5. *Obesity: preventing and managing the global epidemic. Report of a WHO consultation.* World Health Organ Tech Rep Ser, 2000. **894**: p. i-xii, 1-253.

6. Bray, G.A., *The underlying basis for obesity: relationship to cancer.* J Nutr, 2002. **132**(11 Suppl): p. 3451S-3455S.

7. Prentice, A.M. and S.A. Jebb, *Beyond body mass index.* Obes Rev, 2001. **2**(3): p. 141-7.

8. National Institute of Diabetes and Digestive and Kidney Diseases, *Weight Cycling*, in *NIH Publication No. 01-3901*, W.c.a.i. network, Editor. 2001, National Institutes of Health: Bethesda.

9. Eng, S., et al., *Body size changes in relation to postmenopausal breast cancer among women on Long Island, New York.* Am J Epidemiol, 2004.

10. Bray, G.A., *Determining body composition in adults*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2004.

11. Han, T.S., et al., *The influences of height and age on waist circumference as an index of adiposity in adults.* Int J Obes Relat Metab Disord, 1997. **21**(1): p. 83-9.

12. Zhu, S., et al., *Waist circumference and obesity-associated risk factors among whites in the third National Health and Nutrition Examination Survey: clinical action thresholds.* Am J Clin Nutr, 2002. **76**(4): p. 743-9.

13. Institute", N.H.L.B., http://www.nhlbi.nih.gov/health/public/heart/obesity/lose_wt/index.htm. Obesity Education Initiative.

14. Lean, M.E., T.S. Han, and C.E. Morrison, *Waist circumference as a measure for indicating need for weight management.* Bmj, 1995. **311**(6998): p. 158-61.

15. Gallagher, D., et al., *Weight loss in postmenopausal obesity: no adverse alterations in body composition and protein metabolism.* Am J Physiol Endocrinol Metab, 2000. **279**(1): p. E124-31.

16. Bray, G.A. and C.M. Champagne, *Beyond energy balance: there is more to obesity than kilocalories.* J Am Diet Assoc, 2005. **105**(5 Suppl 1): p. S17-23.

17. Lovejoy, J.C., *The influence of sex hormones on obesity across the female life span.* J Womens Health, 1998. **7**(10): p. 1247-56.

18. Ballard-Barbash, R., *Anthropometry and breast cancer. Body size--a moving target.* Cancer, 1994. **74**(3 Suppl): p. 1090-100.

19. Nussey, S. and S.A. Whitehead, *Endocrinology : an integrated approach*. 2001, Oxford, UK: Bios. xi, 358.

20. Flegal, K.M., et al., *Prevalence and trends in obesity among US adults, 1999-2000.* Jama, 2002. **288**(14): p. 1723-7.

21. Hedley, A.A., et al., *Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002.* Jama, 2004. **291**(23): p. 2847-50.

22. Okosun, I.S., et al., *Abdominal adiposity in U.S. adults: prevalence and trends, 1960-2000.* Prev Med, 2004. **39**(1): p. 197-206.

23. Bray, G.A., *Health hazards associated with obesity in adults*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2005.

24. Mokdad, A.H., et al., *Actual causes of death in the United States, 2000.* Jama, 2004. **291**(10): p. 1238-45.

25. Colditz, G.A., et al., *Weight gain as a risk factor for clinical diabetes mellitus in women.* Ann Intern Med, 1995. **122**(7): p. 481-6.

26. Bergstrom, A., et al., *Overweight as an avoidable cause of cancer in Europe.* Int J Cancer, 2001. **91**(3): p. 421-30.

27. American Cancer Society, *Cancer Facts & Figures 2007*. 2006, American Cancer Society: Atlanta, GA.

28. American Cancer Society, *Breast Cancer Facts and Figures 2007-2008*. 2007, American Cancer Society: Atlanta, GA.

29. Hankinson, S.E., G.A. Colditz, and W.C. Willett, *Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones.* Breast Cancer Res, 2004. **6**(5): p. 213-8.

30. Muti, P., *The role of endogenous hormones in the etiology and prevention of breast cancer: the epidemiological evidence.* Ann N Y Acad Sci, 2004. **1028**: p. 273-82.

31. Kufe, D.W., et al., *Cancer medicine 6*. 6th ed. 2003, Hamilton, Ont.: BC Decker. 2 v.

32. Matthews, C.E., et al., *Physical activity, body size, and estrogen metabolism in women.* Cancer Causes Control, 2004. **15**(5): p. 473-81.

33. Key, T.J. and M.C. Pike, *The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer.* Eur J Cancer Clin Oncol, 1988. **24**(1): p. 29-43.

34. Campagnoli, C., et al., *Progestins and progesterone in hormone replacement therapy and the risk of breast cancer.* J Steroid Biochem Mol Biol, 2005. **96**(2): p. 95-108.

35. Eliassen, A.H., et al., *Circulating insulin and c-peptide levels and risk of breast cancer among predominately premenopausal women.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(1): p. 161-4.

36. Yee, D. and A.V. Lee, *Crosstalk between the insulin-like growth factors and estrogens in breast cancer.* J Mammary Gland Biol Neoplasia, 2000. **5**(1): p. 107-15.

37. Gammon, M.D., et al., *The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer.* Breast Cancer Res Treat, 2002. **74**(3): p. 235-54.

38. Adami, H.-O., et al., *Textbook of cancer epidemiology*. 2002, Oxford ; New York, N.Y.: Oxford University Press. xxiv, 599.

39. Collaborative Group on Hormonal Factors in Breast Cancer, *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.

40. Collaborative Group on Hormonal Factors in Breast Cancer, *Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer.* Lancet, 1997. **350**(9084): p. 1047-59.

41. Beral, V., *Breast cancer and hormone-replacement therapy in the Million Women Study.* Lancet, 2003. **362**(9382): p. 419-27.

42. Shantakumar, S., et al., *Age and menopausal effects of hormonal birth control and hormone replacement therapy in relation to breast cancer risk.* Am J Epidemiol, 2007. **165**(10): p. 1187-98.

43. Egan, K.M., et al., *Prospective study of regular aspirin use and the risk of breast cancer.* J Natl Cancer Inst, 1996. **88**(14): p. 988-93.

44. Moorman, P.G., et al., *Association between non-steroidal anti-inflammatory drugs (NSAIDs) and invasive breast cancer and carcinoma in situ of the breast.* Cancer Causes Control, 2003. **14**(10): p. 915-22.

45. Terry, M.B., et al., *Association of frequency and duration of aspirin use and hormone receptor status with breast cancer risk.* Jama, 2004. **291**(20): p. 2433-40.

46. Mazhar, D., R. Ang, and J. Waxman, *COX inhibitors and breast cancer.* Br J Cancer, 2006. **94**(3): p. 346-50.

47. Gonzalez-Perez, A., L.A. Garcia Rodriguez, and R. Lopez-Ridaura, *Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis.* BMC Cancer, 2003. **3**: p. 28.

48. Marshall, S.F., et al., *Nonsteroidal anti-inflammatory drug use and breast cancer risk by stage and hormone receptor status.* J Natl Cancer Inst, 2005. **97**(11): p. 805-12.

49. Hamajima, N., et al., *Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease.* Br J Cancer, 2002. **87**(11): p. 1234-45.

50. Smith-Warner, S.A., et al., *Types of dietary fat and breast cancer: a pooled analysis of cohort studies.* Int J Cancer, 2001. **92**(5): p. 767-74.

51. Smith-Warner, S.A., et al., *Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies.* Jama, 2001. **285**(6): p. 769-76.

52. Gandini, S., et al., *Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients.* Eur J Cancer, 2000. **36**(5): p. 636-46.

53. Howe, G.R., et al., *Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies.* J Natl Cancer Inst, 1990. **82**(7): p. 561-9.

54. *Risk of ionizing radiation exposure to children: a subject review. American Academy of Pediatrics. Committee on Environmental Health.* Pediatrics, 1998. **101**(4 Pt 1): p. 717-9.

55. Gammon, M.D., et al., *Environmental tobacco smoke and breast cancer incidence.* Environ Res, 2004. **96**(2): p. 176-85.

56. Terry, P.D. and T.E. Rohan, *Cigarette smoking and the risk of breast cancer in women: a review of the literature.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(10 Pt 1): p. 953-71.

57. International Agency for Research on Cancer. and World Health Organization., *Polynuclear aromatic compounds. chemical, environmental and experimental data*. 1983, Lyon, France: International Agency for Research on Cancer. 477.

58. Rundle, A., et al., *The relationship between genetic damage from polycyclic aromatic hydrocarbons in breast tissue and breast cancer.* Carcinogenesis, 2000. **21**(7): p. 1281-9.

59. Gammon, M.D., et al., *Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(8): p. 677-85.

60. Li, D., et al., *Aromatic DNA adducts in adjacent tissues of breast cancer patients: clues to breast cancer etiology.* Cancer Res, 1996. **56**(2): p. 287-93.

61. van den Brandt, P.A., et al., *Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk.* Am J Epidemiol, 2000. **152**(6): p. 514-27.

62. Lukanova, A., et al., *Body mass index, circulating levels of sex-steroid hormones, IGF-I and IGF-binding protein-3: a cross-sectional study in healthy women.* Eur J Endocrinol, 2004. **150**(2): p. 161-71.

63. Wolf, I., et al., *Diabetes mellitus and breast cancer.* Lancet Oncol, 2005. **6**(2): p. 103-11.

64. Centers for Disease Control and Prevention, *Behavioral Risk Factor Surveillance System Survey Data*, C.f.D.C.a.P. U.S. Department of Health and Human Services, Editor. 2008: Atlanta, GA.

65. Friedenreich, C.M., *Physical activity and breast cancer risk: the effect of menopausal status.* Exerc Sport Sci Rev, 2004. **32**(4): p. 180-4.

66. Park, K.S., et al., *Polymorphisms of tumour necrosis factors A and B in breast cancer.* Eur J Immunogenet, 2002. **29**(1): p. 7-10.

67. Gaudet, M.M., et al., *Genetic variation in tumor necrosis factor and lymphotoxin-alpha (TNF-LTA) and breast cancer risk.* Hum Genet, 2007. **121**(3-4): p. 483-90.

68. Wilson, A.G., et al., *An allelic polymorphism within the human tumor necrosis factor alpha promoter region is strongly associated with HLA A1, B8, and DR3 alleles.* J Exp Med, 1993. **177**(2): p. 557-60.

69. Smith, K.C., et al., *Cytokine gene polymorphisms and breast cancer susceptibility and prognosis.* Eur J Immunogenet, 2004. **31**(4): p. 167-73.

70. Giordani, L., et al., *Association of breast cancer and polymorphisms of interleukin-10 and tumor necrosis factor-alpha genes.* Clin Chem, 2003. **49**(10): p. 1664-7.

71. Slattery, M.L., et al., *IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(4): p. 747-55.

72. Howell, W.M. and M.J. Rose-Zerilli, *Cytokine gene polymorphisms, cancer susceptibility, and prognosis.* J Nutr, 2007. **137**(1 Suppl): p. 194S-199S.

73. Figer, A., et al., *Insulin-like growth factor I polymorphism and breast cancer risk in Jewish women.* Isr Med Assoc J, 2002. **4**(10): p. 759-62.

74. Missmer, S.A., et al., *A sequence repeat in the insulin-like growth factor-1 gene and risk of breast cancer.* Int J Cancer, 2002. **100**(3): p. 332-6.

75. DeLellis, K., et al., *IGF1 genotype, mean plasma level and breast cancer risk in the Hawaii/Los Angeles multiethnic cohort.* Br J Cancer, 2003. **88**(2): p. 277-82.

76. Wagner, K., et al., *Polymorphisms in the IGF-1 and IGFBP 3 promoter and the risk of breast cancer.* Breast Cancer Res Treat, 2005. **92**(2): p. 133-40.

77. Cleveland, R.J., et al., *IGF1 CA repeat polymorphisms, lifestyle factors and breast cancer risk in the Long Island Breast Cancer Study Project.* Carcinogenesis, 2005.

78. National Comprehensive Cancer Network, *Breast Cancer Treatment Guidelines for Patients, version VII*, A.C. Society and N.C.C. Network, Editors. 2005: Atlanta.

79. Veronesi, U., et al., *Breast cancer.* Lancet, 2005. **365**(9472): p. 1727-41.

80. Bilous, M., et al., *Making sense of breast pathology.* Aust Fam Physician, 2005. **34**(7): p. 581-6.

81. Bleiweiss, I.J., *Pathology of breast cancer: The in situ carcinomas*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2005.

82. Hayes, D.F., *Measurement of prognostic factors in breast cancer*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2005.

83. Althuis, M.D., et al., *Etiology of hormone receptor-defined breast cancer: a systematic review of the literature.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(10): p. 1558-68.

84. Elledge, R., *Hormone receptors in breast cancer: Measurement and clinical implications*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2005.

85. Yamauchi, H. and D.F. Hayes, *HER-2/neu (c-erbB-2) and predicting response to therapy in breast cancer*, in *UptoDate Online 13.3*, B.D. Rose and J.M. Rush, Editors. 2005.

86. Greene, F.L., et al., *AJCC cancer staging manual*. 6th / ed. 2002, New York: Springer-Verlag. xiv, 421.

87. Singletary, S.E., et al., *Revision of the American Joint Committee on Cancer staging system for breast cancer.* J Clin Oncol, 2002. **20**(17): p. 3628-36.

88. Surveillance, E., and End Results Program (www.seer.cancer.gov/),, *SEER*Stat Database: Incidence - SEER 9 Regs Public-Use, Nov 2004 Sub (1973-2002)*, in *National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch*. released April 2005, based on the November 2004 submission.

89. American Cancer Society, *Breast Cancer Facts and Figures 2005-2006*. 2005, American Cancer Society: Atlanta, GA.

90. Trivers, K.F., et al., *Association between reproductive factors and breast cancer survival in younger women.* Breast Cancer Res Treat, 2007. **103**(1): p. 93-102.

91. Phillips, K.A., et al., *Prognosis of premenopausal breast cancer and childbirth prior to diagnosis.* J Clin Oncol, 2004. **22**(4): p. 699-705.

92. Daling, J.R., et al., *The relation of reproductive factors to mortality from breast cancer.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(3): p. 235-41.

93. Jernstrom, H., et al., *Hormone replacement therapy before breast cancer diagnosis significantly reduces the overall death rate compared with never-use among 984 breast cancer patients.* Br J Cancer, 1999. **80**(9): p. 1453-8.

94. Schairer, C., et al., *Estrogen replacement therapy and breast cancer survival in a large screening study.* J Natl Cancer Inst, 1999. **91**(3): p. 264-70.

95. Bergkvist, L., et al., *Prognosis after breast cancer diagnosis in women exposed to estrogen and estrogen-progestogen replacement therapy.* Am J Epidemiol, 1989. **130**(2): p. 221-8.

96. Bonnier, P., et al., *Clinical and biologic prognostic factors in breast cancer diagnosed during postmenopausal hormone replacement therapy.* Obstet Gynecol, 1995. **85**(1): p. 11-7.

97. Antoine, C., et al., *Influence of HRT on prognostic factors for breast cancer: a systematic review after the Women's Health Initiative trial.* Hum Reprod, 2004. **19**(3): p. 741-56.

98. Trivers, K.F., et al., *Oral contraceptives and survival in breast cancer patients aged 20 to 54 years.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(9): p. 1822-7.

99. Sauerbrei, W., et al., *The effect of oral contraceptive use on the prognosis of node positive breast cancer patients. German Breast Cancer Study Group.* Eur J Cancer, 1998. **34**(9): p. 1348-51.

100. Schonborn, I., P. Nischan, and K. Ebeling, *Oral contraceptive use and the prognosis of breast cancer.* Breast Cancer Res Treat, 1994. **30**(3): p. 283-92.

101. Holmberg, L., et al., *Oral contraceptives and prognosis in breast cancer: effects of duration, latency, recency, age at first use and relation to parity and body mass index in young women with breast cancer.* Eur J Cancer, 1994. **30A**(3): p. 351-4.

102. Zhang, S., et al., *Better breast cancer survival for postmenopausal women who are less overweight and eat less fat. The Iowa Women's Health Study.* Cancer, 1995. **76**(2): p. 275-83.

103. Saxe, G.A., et al., *Diet and risk for breast cancer recurrence and survival.* Breast Cancer Res Treat, 1999. **53**(3): p. 241-53.

104. Ewertz, M., et al., *Survival of breast cancer patients in relation to factors which affect the risk of developing breast cancer.* Int J Cancer, 1991. **49**(4): p. 526-30.

105. Holmes, M.D., et al., *Dietary factors and the survival of women with breast carcinoma.* Cancer, 1999. **86**(5): p. 826-35.

106. Rock, C.L. and W. Demark-Wahnefried, *Nutrition and survival after the diagnosis of breast cancer: a review of the evidence.* J Clin Oncol, 2002. **20**(15): p. 3302-16.

107. Jain, M., A.B. Miller, and T. To, *Premorbid diet and the prognosis of women with breast cancer.* J Natl Cancer Inst, 1994. **86**(18): p. 1390-7.

108. Thiebaut, A.C., et al., *Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort.* J Natl Cancer Inst, 2007. **99**(6): p. 451-62.

109. Manjer, J., et al., *Survival of women with breast cancer in relation to smoking.* Eur J Surg, 2000. **166**(11): p. 852-8.

110. Yu, G.P., et al., *Smoking history and cancer patient survival: a hospital cancer registry study.* Cancer Detect Prev, 1997. **21**(6): p. 497-509.

111. Fentiman, I.S., D.S. Allen, and H. Hamed, *Smoking and prognosis in women with breast cancer.* Int J Clin Pract, 2005. **59**(9): p. 1051-4.

112. Cleveland, R.J., et al., *Weight gain prior to diagnosis and survival from breast cancer.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(9): p. 1803-11.

113. Abrahamson, P.E., et al., *General and abdominal obesity and survival among young women with breast cancer.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(10): p. 1871-7.

114. Majed, B., et al., *Is obesity an independent prognosis factor in woman breast cancer?* Breast Cancer Res Treat, 2007.

115. Reeves, G.K., et al., *Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study.* Bmj, 2007. **335**(7630): p. 1134.

116. Cui, Y., et al., *Body mass and stage of breast cancer at diagnosis.* Int J Cancer, 2002. **98**(2): p. 279-83.

117. Petrelli, J.M., et al., *Body mass index, height, and postmenopausal breast cancer mortality in a prospective cohort of US women.* Cancer Causes Control, 2002. **13**(4): p. 325-32.

118. Carmichael, A.R., et al., *Does obesity compromise survival in women with breast cancer?* Breast, 2004. **13**(2): p. 93-6.

119. Tretli, S., T. Haldorsen, and L. Ottestad, *The effect of pre-morbid height and weight on the survival of breast cancer patients.* Br J Cancer, 1990. **62**(2): p. 299-303.

120. Newman, S.C., A.W. Lees, and H.J. Jenkins, *The effect of body mass index and oestrogen receptor level on survival of breast cancer patients.* Int J Epidemiol, 1997. **26**(3): p. 484-90.

121. Enger, S.M., et al., *Body weight correlates with mortality in early-stage breast cancer.* Arch Surg, 2004. **139**(9): p. 954-58; discussion 958-60.

122. Colleoni, M., et al., *Relation between chemotherapy dose, oestrogen receptor expression, and body-mass index.* Lancet, 2005. **366**(9491): p. 1108-10.

123. Rohan, T.E., W. Fu, and J.E. Hiller, *Physical activity and survival from breast cancer.* Eur J Cancer Prev, 1995. **4**(5): p. 419-24.

124. Borugian, M.J., et al., *Insulin, macronutrient intake, and physical activity: are potential indicators of insulin resistance associated with mortality from breast cancer?* Cancer Epidemiol Biomarkers Prev, 2004. **13**(7): p. 1163-72.

125. Abrahamson, P.E., et al., *Recreational physical activity and survival among young women with breast cancer.* Cancer, 2006. **107**(8): p. 1777-85.

126. Robson, M., *Are BRCA1- and BRCA2-associated breast cancers different? Prognosis of BRCA1-associated breast cancer.* J Clin Oncol, 2000. **18**(21 Suppl): p. 113S-8S.

127. Rennert, G., et al., *Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations.* N Engl J Med, 2007. **357**(2): p. 115-23.

128. Brekelmans, C.T., et al., *Tumour characteristics, survival and prognostic factors of hereditary breast cancer from BRCA2-, BRCA1- and non-BRCA1/2 families as compared to sporadic breast cancer cases.* Eur J Cancer, 2007. **43**(5): p. 867-76.

129. El-Tamer, M., et al., *Survival and recurrence after breast cancer in BRCA1/2 mutation carriers.* Ann Surg Oncol, 2004. **11**(2): p. 157-64.

130. Goodwin, P.J., K.A. Phillips, and D.W. West, *Prognosis of breast cancer in carriers of BRCA1 and BRCA2 mutations.* N Engl J Med, 2007. **357**(15): p. 1555; author reply 1556.

131. Verhoog, L.C., et al., *Prognostic significance of germline BRCA2 mutations in hereditary breast cancer patients.* J Clin Oncol, 2000. **18**(21 Suppl): p. 119S-24S.

132. Goetz, M.P., A. Kamal, and M.M. Ames, *Tamoxifen Pharmacogenomics: The Role of CYP2D6 as a Predictor of Drug Response.* Clin Pharmacol Ther, 2007.

133. Schroth, W., et al., *Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes.* J Clin Oncol, 2007. **25**(33): p. 5187-93.

134. Goetz, M.P., et al., *The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen.* Breast Cancer Res Treat, 2007. **101**(1): p. 113-21.

135. Nowell, S.A., et al., *Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients.* Breast Cancer Res Treat, 2005. **91**(3): p. 249-58.

136. Wegman, P., et al., *Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer.* Breast Cancer Res, 2007. **9**(1): p. R7.

137. Ambrosone, C.B., et al., *Polymorphisms in genes related to oxidative stress (MPO, MnSOD, CAT) and survival after treatment for breast cancer.* Cancer Res, 2005. **65**(3): p. 1105-11.

138. Ambrosone, C.B., et al., *Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer.* Cancer Res, 2001. **61**(19): p. 7130-5.

139. Syamala, V.S., et al., *Influence of germline polymorphisms of GSTT1, GSTM1, and GSTP1 in familial versus sporadic breast cancer susceptibility and survival.* Fam Cancer, 2007.

140. Yang, G., et al., *Genetic polymorphisms in glutathione-S-transferase genes (GSTM1, GSTT1, GSTP1) and survival after chemotherapy for invasive breast carcinoma.* Cancer, 2005. **103**(1): p. 52-8.

141. Strachan, T. and A.P. Read, *Human Molecular Genetics 3*. 3rd ed. 2004, London: Garland Science.

142. Kota, B.P., T.H. Huang, and B.D. Roufogalis, *An overview on biological mechanisms of PPARs.* Pharmacol Res, 2005. **51**(2): p. 85-94.

143. van Raalte, D.H., et al., *Peroxisome proliferator-activated receptor (PPAR)-alpha: a pharmacological target with a promising future.* Pharm Res, 2004. **21**(9): p. 1531-8.

144. Aranda, A. and A. Pascual, *Nuclear hormone receptors and gene expression.* Physiol Rev, 2001. **81**(3): p. 1269-304.

145. Kiec-Wilk, B., et al., *The selected pathophysiological aspects of PPARs activation.* J Physiol Pharmacol, 2005. **56**(2): p. 149-62.

146. Kersten, S., *Peroxisome proliferator activated receptors and obesity.* Eur J Pharmacol, 2002. **440**(2-3): p. 223-34.

147. Calle, E.E. and M.J. Thun, *Obesity and cancer.* Oncogene, 2004. **23**(38): p. 6365-78.

148. Evans, R.M., G.D. Barish, and Y.X. Wang, *PPARs and the complex journey to obesity.* Nat Med, 2004. **10**(4): p. 355-61.

149. Moller, D.E. and J.P. Berger, *Role of PPARs in the regulation of obesity-related insulin sensitivity and inflammation.* Int J Obes Relat Metab Disord, 2003. **27 Suppl 3**: p. S17-21.

150. Terauchi, Y. and T. Kadowaki, *Peroxisome proliferator-activated receptors and insulin secretion.* Endocrinology, 2005. **146**(8): p. 3263-5.

151. Flavell, D.M., et al., *Variation in the PPARalpha gene is associated with altered function in vitro and plasma lipid concentrations in Type II diabetic subjects.* Diabetologia, 2000. **43**(5): p. 673-80.

152. Flavell, D.M., et al., *Peroxisome proliferator-activated receptor alpha gene variation influences age of onset and progression of type 2 diabetes.* Diabetes, 2005. **54**(2): p. 582-6.

153. Chen, S., et al., *Effects of PPARalpha, gamma and delta haplotypes on plasma levels of lipids, severity and progression of coronary atherosclerosis and response to statin therapy in the lipoprotein coronary atherosclerosis study.* Pharmacogenetics, 2004. **14**(1): p. 61-71.

154. Foucher, C., et al., *Response to micronized fenofibrate treatment is associated with the peroxisome-proliferator-activated receptors alpha G/C intron7 polymorphism in subjects with type 2 diabetes.* Pharmacogenetics, 2004. **14**(12): p. 823-9.

155. Brune, S., et al., *Polymorphism in the peroxisome proliferator-activated receptor alpha gene influences the risk for Alzheimer's disease.* J Neural Transm, 2003. **110**(9): p. 1041-50.

156. Nielsen, E.M., et al., *Evidence for an association between the Leu162Val polymorphism of the PPARalpha gene and decreased fasting serum triglyceride levels in glucose tolerant subjects.* Pharmacogenetics, 2003. **13**(7): p. 417-23.

157. Eurlings, P.M., et al., *Identification of the PPARA locus on chromosome 22q13.3 as a modifier gene in familial combined hyperlipidemia.* Mol Genet Metab, 2002. **77**(4): p. 274-81.

158. Flavell, D.M., et al., *Peroxisome proliferator-activated receptor alpha gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease.* Circulation, 2002. **105**(12): p. 1440-5.

159. Jamshidi, Y., et al., *Genetic determinants of the response to bezafibrate treatment in the lower extremity arterial disease event reduction (LEADER) trial.* Atherosclerosis, 2002. **163**(1): p. 183-92.

160. Brisson, D., et al., *Effect of apolipoprotein E, peroxisome proliferator-activated receptor alpha and lipoprotein lipase gene mutations on the ability of fenofibrate to improve lipid profiles and reach clinical guideline targets among hypertriglyceridemic patients.* Pharmacogenetics, 2002. **12**(4): p. 313-20.

161. Tai, E.S., et al., *Polyunsaturated fatty acids interact with the PPARA-L162V polymorphism to affect plasma triglyceride and apolipoprotein C-III concentrations in the Framingham Heart Study.* J Nutr, 2005. **135**(3): p. 397-403.

162. Tai, E.S., et al., *Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study.* Arterioscler Thromb Vasc Biol, 2002. **22**(5): p. 805-10.

163. Gouni-Berthold, I., et al., *Association between the PPARalpha L162V polymorphism, plasma lipoprotein levels, and atherosclerotic disease in patients with diabetes mellitus type 2 and in nondiabetic controls.* Am Heart J, 2004. **147**(6): p. 1117-24.

164. Robitaille, J., et al., *Association between the PPARalpha-L162V polymorphism and components of the metabolic syndrome.* J Hum Genet, 2004. **49**(9): p. 482-9.

165. Bosse, Y., et al., *The peroxisome proliferator-activated receptor alpha L162V mutation is associated with reduced adiposity.* Obes Res, 2003. **11**(7): p. 809-16.

166. Bosse, Y., et al., *Influences of the PPAR alpha-L162V polymorphism on plasma HDL(2)-cholesterol response of abdominally obese men treated with gemfibrozil.* Genet Med, 2002. **4**(4): p. 311-5.

167. Bosse, Y., et al., *Combined effects of PPARgamma2 P12A and PPARalpha L162V polymorphisms on glucose and insulin homeostasis: the Quebec Family Study.* J Hum Genet, 2003. **48**(12): p. 614-21.

168. Evans, D., et al., *Association between polymorphisms in candidate genes and morbid obesity.* Int J Obes Relat Metab Disord, 2001. **25 Suppl 1**: p. S19-21.

169. Evans, D., et al., *A polymorphism, L162V, in the peroxisome proliferator-activated receptor alpha (PPARalpha) gene is associated with lower body mass index in patients with non-insulin-dependent diabetes mellitus.* J Mol Med, 2001. **79**(4): p. 198-204.

170. Nieters, A., N. Becker, and J. Linseisen, *Polymorphisms in candidate obesity genes and their interaction with dietary intake of n-6 polyunsaturated fatty acids affect obesity risk in a sub-sample of the EPIC-Heidelberg cohort.* Eur J Nutr, 2002. **41**(5): p. 210-21.

171. Ishiguro, H., et al., *Mutation analysis of the retinoid X receptor beta, nuclear-related receptor 1, and peroxisome proliferator-activated receptor alpha genes in schizophrenia and alcohol dependence: possible haplotype association of nuclear-related receptor 1 gene to alcohol dependence.* Am J Med Genet, 2002. **114**(1): p. 15-23.

172. Puckey, L.H. and B.L. Knight, *Variation at position 162 of peroxisome proliferator-activated receptor alpha does not influence the effect of fibrates on cholesterol or triacylglycerol concentrations in hyperlipidaemic subjects.* Pharmacogenetics, 2001. **11**(7): p. 619-24.

173. Lacquemant, C., et al., *Mutation screening of the PPARalpha gene in type 2 diabetes associated with coronary heart disease.* Diabetes Metab, 2000. **26**(5): p. 393-401.

174. Vohl, M.C., et al., *Molecular scanning of the human PPARa gene: association of the L162v mutation with hyperapobetalipoproteinemia.* J Lipid Res, 2000. **41**(6): p. 945-52.

175. Uthurralt, J., et al., *PPARalpha L162V underlies variation in serum triglycerides and subcutaneous fat volume in young males.* BMC Med Genet, 2007. **8**: p. 55.

176. Doney, A.S., et al., *Association of common variation in the PPARA gene with incident myocardial infarction in individuals with type 2 diabetes: a Go-DARTS study.* Nucl Recept, 2005. **3**: p. 4.

177. Shin, M.J., A.M. Kanaya, and R.M. Krauss, *Polymorphisms in the peroxisome proliferator activated receptor alpha gene are associated with levels of apolipoprotein CIII and triglyceride in African-Americans but not Caucasians.* Atherosclerosis, 2007.

178. Manresa, J.M., et al., *Relationship of classical and non-classical risk factors with genetic variants relevant to coronary heart disease.* Eur J Cardiovasc Prev Rehabil, 2006. **13**(5): p. 738-44.

179. Tai, E.S., et al., *The L162V polymorphism at the peroxisome proliferator activated receptor alpha locus modulates the risk of cardiovascular events associated with insulin resistance and diabetes mellitus: the Veterans Affairs HDL Intervention Trial (VA-HIT).* Atherosclerosis, 2006. **187**(1): p. 153-60.

180. Waksberg, J., *Sampling methods for random digit dialing.* J Amer Statistic Assoc, 1978. **73**: p. 40-46.

181. Shen, J., et al., *Polymorphisms in XRCC1 modify the association between polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, dietary antioxidants, and breast cancer risk.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(2): p. 336-42.

182. Terry, M.B., et al., *Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(12): p. 2053-8.

183. Potischman, N., et al., *Comparison of the 60- and 100-item NCI-block questionnaires with validation data.* Nutr Cancer, 1999. **34**(1): p. 70-5.

184. Fink, B.N., et al., *Construction of a flavonoid database for assessing intake in a population-based sample of women on Long Island, New York.* Nutr Cancer, 2006. **56**(1): p. 57-66.

185. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.

186. Fallin, D., et al., *Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease.* Genome Res, 2001. **11**(1): p. 143-51.

187. Zaykin, D.V., et al., *Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals.* Hum Hered, 2002. **53**(2): p. 79-91.

188. Botstein, D. and N. Risch, *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* Nat Genet, 2003. **33 Suppl**: p. 228-37.

189. Collins, F.S., M.S. Guyer, and A. Charkravarti, *Variations on a theme: cataloging human DNA sequence variation.* Science, 1997. **278**(5343): p. 1580-1.

190. Carlson, C.S., et al., *Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium.* Am J Hum Genet, 2004. **74**(1): p. 106-20.

191. Carlson, C.S., et al., *Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans.* Nat Genet, 2003. **33**(4): p. 518-21.

192. de Bakker, P., *Tagger,* http://www.broad.mit.edu/mpg/tagger.

193. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.

194. Schaid, D.J. and S.J. Jacobsen, *Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions.* Am J Epidemiol, 1999. **149**(8): p. 706-11.

195. Wittke-Thompson, J.K., A. Pluzhnikov, and N.J. Cox, *Rational inferences about departures from Hardy-Weinberg equilibrium.* Am J Hum Genet, 2005. **76**(6): p. 967-86.

196. Lewontin, R.C., *On measures of gametic disequilibrium.* Genetics, 1988. **120**(3): p. 849-52.

197. Weir, B., *Genetic data analysis II*. 1996, Sunderland MA: Sinauer associations Inc.

198. Clark, A.G., *Inference of haplotypes from PCR-amplified samples of diploid populations.* Mol Biol Evol, 1990. **7**(2): p. 111-22.

199. Excoffier, L. and M. Slatkin, *Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.* Mol Biol Evol, 1995. **12**(5): p. 921-7.

200. Schaid, D.J., *Evaluating associations of haplotypes with traits.* Genet Epidemiol, 2004. **27**(4): p. 348-64.

201. Schaid, D.J., et al., *Score tests for association between traits and haplotypes when linkage phase is ambiguous.* Am J Hum Genet, 2002. **70**(2): p. 425-34.

202. Terry, K.L., et al., *Genetic variation in the progesterone receptor gene and ovarian cancer risk.* Am J Epidemiol, 2005. **161**(5): p. 442-51.

203. Hooker, S., et al., *NAT2 and NER genetic variants and sporadic prostate cancer susceptibility in African Americans.* Prostate Cancer Prostatic Dis, 2007.

204. Kleinbaum, D.G. and M. Klein, *Logistic Regression: A self-learning text*. 2nd edition ed. Statistics for Biology and Health, ed. K. Dietz, et al. 2002, New York: Springer.

205. Stokes, M.E., C.S. Davis, and G.G. Koch, *Categorical data analysis using the SAS® system*. 2nd ed. 2001, Cary, NC: SAS Institute. 626.

206. Thomas, D.C., D.S. Ward, and B.H. Kaplan, *Statistical methods in genetic epidemiology*. 2004, Oxford ; New York: Oxford University Press. xxi, 435.

207. Rothman, K.J. and S. Greenland, *Modern epidemiology*. 2nd ed. 1998, Philadelphia, Pa.: Lippincott-Raven. xiii, 737.

208. Witte, J.S., *Genetic analysis with hierarchical models.* Genet Epidemiol, 1997. **14**(6): p. 1137-42.

209. Greenland, S., *Principles of multilevel modelling.* Int J Epidemiol, 2000. **29**(1): p. 158-67.

210. Hung, R.J., et al., *Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(6): p. 1013-21.

211. Rothman, K. and J. Boice, *Epidemiologic Analysis with a Programmable Calculator*, in *NIH Publication No. 79-1649*. 1979.

212. Garcia-Closas, M. and J.H. Lubin, *Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches.* Am J Epidemiol, 1999. **149**(8): p. 689-92.

213. Johnson, G.C., et al., *Haplotype tagging for the identification of common disease genes.* Nat Genet, 2001. **29**(2): p. 233-7.

214. Klipstein-Grobusch, K., A. Kroke, and H. Boeing, *Reproducibility of self-reported past body weight.* Eur J Clin Nutr, 1998. **52**(7): p. 525-8.

215. Olivarius, N.F., A.H. Andreasen, and J. Loken, *Accuracy of 1-, 5- and 10-year body weight recall given in a standard questionnaire.* Int J Obes Relat Metab Disord, 1997. **21**(1): p. 67-71.

216. Stevens, J., et al., *Accuracy of current, 4-year, and 28-year self-reported body weight in an elderly population.* Am J Epidemiol, 1990. **132**(6): p. 1156-63.

217. Suchanek, K.M., et al., *Peroxisome proliferator-activated receptor alpha in the human breast cancer cell lines MCF-7 and MDA-MB-231.* Mol Carcinog, 2002. **34**(4): p. 165-71.

218. Gervois, P., et al., *A truncated human peroxisome proliferator-activated receptor alpha splice variant with dominant negative activity.* Mol Endocrinol, 1999. **13**(9): p. 1535-49.

219. Hanselman, J.C., et al., *Expression of the mRNA encoding truncated PPAR alpha does not correlate with hepatic insensitivity to peroxisome proliferators.* Mol Cell Biochem, 2001. **217**(1-2): p. 91-7.

220. Klos, K.L., et al., *Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants.* Arterioscler Thromb Vasc Biol, 2006. **26**(8): p. 1828-36.

221. Chan, E., et al., *The V227A polymorphism at the PPARA locus is associated with serum lipid concentrations and modulates the association between dietary polyunsaturated fatty acid intake and serum high density lipoprotein concentrations in Chinese women.* Atherosclerosis, 2006. **187**(2): p. 309-15.

222. Balcerzyk, A., I. Zak, and J. Krauze, *Synergistic effect between polymorphisms of PPARA and ABCA1 genes on the premature coronary artery disease.* Acta Cardiol, 2007. **62**(3): p. 233-8.

223. Andrulionyte, L., et al., *Single nucleotide polymorphisms of the peroxisome proliferator-activated receptor-alpha gene (PPARA) influence the conversion from impaired glucose tolerance to type 2 diabetes: the STOP-NIDDM trial.* Diabetes, 2007. **56**(4): p. 1181-6.

224. Guo, S.W. and E.A. Thompson, *Performing the exact test of Hardy-Weinberg proportion for multiple alleles.* Biometrics, 1992. **48**(2): p. 361-72.

225. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.

226. Witte, J.S., et al., *Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer.* Epidemiology, 1994. **5**(6): p. 612-21.

227. Hosmer, D.W. and S. Lemeshow, *Confidence interval estimation of interaction.* Epidemiology, 1992. **3**(5): p. 452-6.

228. Daly, M.J., et al., *High-resolution haplotype structure in the human genome.* Nat Genet, 2001. **29**(2): p. 229-32.

229. Redden, D.T. and D.B. Allison, *Nonreplication in genetic association studies of obesity and diabetes research.* J Nutr, 2003. **133**(11): p. 3323-6.

230. Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research.* Epidemiology, 1999. **10**(1): p. 37-48.

231. Golembesky, A.K., et al., *Peroxisome proliferator-activated receptor-alpha (PPARA) genetic polymorphisms and breast cancer incidence: a LIBCSP ancillary study.* In progress, 2007.

232. Ries, L., et al., eds. *SEER Cancer Statistics Review, 1975-2004.* http://seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER web site, ed. N.C. Institute. 2007: Bethesda, MD.

233. Nielsen, D.M., M.G. Ehm, and B.S. Weir, *Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus.* Am J Hum Genet, 1998. **63**(5): p. 1531-40.

234. Baron, J.A., *Aspirin and cancer.* Prev Med, 1995. **24**(2): p. 121-4.

235. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.* Nat Genet, 2000. **26**(1): p. 76-80.

236. Rubin, G.L., et al., *Peroxisome proliferator-activated receptor gamma ligands inhibit estrogen biosynthesis in human breast adipose tissue: possible implications for breast cancer therapy.* Cancer Res, 2000. **60**(6): p. 1604-8.

237. Gong, Z., et al., *The PPAR(gamma) Pro12Ala polymorphism and risk for incident sporadic colorectal adenomas.* Carcinogenesis, 2005. **26**(3): p. 579-85.

238. Smith, M.R. and P.W. Kantoff, *Peroxisome proliferator-activated receptor gamma (PPargamma) as a novel target for prostate cancer.* Invest New Drugs, 2002. **20**(2): p. 195-200.

239. Paynter, R.A., et al., *No evidence of a role for PPARgamma Pro12Ala polymorphism in endometrial cancer susceptibility.* Pharmacogenetics, 2004. **14**(12): p. 851-6.

240. Paltoo, D., et al., *Pro12Ala polymorphism in the peroxisome proliferator-activated receptor-gamma (PPAR-gamma) gene and risk of prostate cancer among men in a large cancer prevention study.* Cancer Lett, 2003. **191**(1): p. 67-74.

241. Slattery, M.L., et al., *PPARgamma, energy balance, and associations with colon and rectal cancer.* Nutr Cancer, 2005. **51**(2): p. 155-61.

242. Zmuda, J.M., et al., *Peroxisome proliferator-activated receptor-gamma polymorphism, body mass and prostate cancer risk: evidence for gene-environment interaction.* Oncology, 2006. **70**(3): p. 185-9.

243. Memisoglu, A., et al., *Lack of association of the codon 12 polymorphism of the peroxisome proliferator-activated receptor gamma gene with breast cancer and body mass.* Pharmacogenetics, 2002. **12**(8): p. 597-603.

244. Yee, L.D., et al., *Pilot study of rosiglitazone therapy in women with breast cancer: effects of short-term therapy on tumor tissue and serum markers.* Clin Cancer Res, 2007. **13**(1): p. 246-52.

245. Xu, X., et al., *Polymorphisms of one-carbon-metabolizing genes and risk of breast cancer in a population-based study.* Carcinogenesis, 2007. **28**(7): p. 1504-9.

246. Kroenke, C.H., et al., *Weight, weight gain, and survival after breast cancer diagnosis.* J Clin Oncol, 2005. **23**(7): p. 1370-8.

247. Loi, S., et al., *Obesity and outcomes in premenopausal and postmenopausal breast cancer.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(7): p. 1686-91.

248. Whiteman, M.K., et al., *Body mass and mortality after breast cancer diagnosis.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(8): p. 2009-14.

249. Berclaz, G., et al., *Body mass index as a prognostic feature in operable breast cancer: the International Breast Cancer Study Group experience.* Ann Oncol, 2004. **15**(6): p. 875-84.

250. Enger, S.M. and L. Bernstein, *Exercise activity, body size and premenopausal breast cancer survival.* Br J Cancer, 2004. **90**(11): p. 2138-41.

251. Borugian, M.J., et al., *Waist-to-hip ratio and breast cancer mortality.* Am J Epidemiol, 2003. **158**(10): p. 963-8.

252. American Joint Committee on Cancer, *AJCC Comparison Guide: Fifth Versus Sixth Edition (http://www.cancerstaging.net)*. 2002.