GENETIC VARIATION AT THE INSULIN-LIKE GROWTH FACTOR 1 GENE AND ASSOCIATION WITH BREAST CANCER, BREAST DENSITY AND ANTHROPOMETRIC MEASURES

by

Gordon Markus Fehringer

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy (Epidemiology)
Graduate Department of Public Health Sciences
University of Toronto

© Gordon Markus Fehringer 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-59010-2
Our file Notre référence
ISBN: 978-0-494-59010-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Title:	Genetic variation at the insulin-like growth factor 1 gene and association with breast cancer, breast density and anthropometric measures
Degree:	Doctor of Philosophy
Year:	2008
Name:	Gordon Markus Fehringer
Department:	Department of Public Health Sciences, University of Toronto

Background and objectives

Evidence suggests that circulating IGF-I levels increase mammographic density (a breast cancer risk factor) and breast cancer risk in premenopausal women. The objective of this thesis was to examine the association of genetic variation at the *IGF1* gene with IGF-I concentration, mammographic density, breast cancer risk, and related anthropometric measures in premenopausal women.

Methods

Three *IGF1* CA repeat polymorphisms (at the 5′ and 3′ ends, and in intron 2) were genotyped. A cross-sectional design was used to investigate their associations with IGF-I levels, mammographic density, BMI, weight, and height. Families from registries in Ontario and Australia were used to investigate associations with breast cancer risk and also BMI, weight and height.

Results

In the cross-sectional study, greater number of copies of the 5′ 19 allele were associated with lower circulating IGF-I levels. Greater number of 3′ 185 alleles were associated with greater percentage breast density, smaller amount of non-dense tissue, and lower BMI. Including BMI in regression models removed the association of the 3′ 185 allele with percentage breast density.

In the family based study, nominally significant associations (5′ 21 allele, intron 2 212 allele, intron 2 216 allele) with breast cancer risk were observed, but significance was lost after multiple comparison adjustment. There was a stronger association between the intron 2 216 allele and risk under a recessive model, and 5′ allele groupings of length 18 to 20 and 20 or more repeats produced significant positive and negative associations respectively. These associations were not strongly supported in analyses stratified by registry. Results from the family based study did not support an association between genetic variation at *IGF1* with BMI, weight or height.

Conclusions

No specific *IGF1* variant influenced each of circulating IGF-I levels, mammographic density, and breast cancer risk. The failure to replicate the association of the 3′ 185 allele with BMI in the family based study suggests that the association of the 3′ 185 allele with percentage breast density is spurious, since this association was mediated through the relationship with BMI (suggesting IGF-I action on body fat). Evidence for an association between *IGF1* and breast cancer risk was limited.

Acknowledgements

I would like to thank my thesis supervisor Dr. Norman Boyd for the advice and guidance he provided throughout this project. His patience is greatly appreciated.

I would also like to extend my thanks to members of my committee. The efforts of Dr. Hilmi Ozcelik who welcomed me into his laboratory and provided guidance throughout this project are greatly appreciated. I am also extremely grateful to Dr. Andrew Paterson for his mentorship during the course of this project and the constructive and insightful comments provided to me by Dr. Julia Knight.

Special thanks to Dr. John Hopper who granted me access to DNA samples and data from the Australian Breast Cancer Family Registry. Thanks also to Dr. Gillian Dite who facilitated my access to these data and addressed numerous questions. The efforts of Dr. Melissa Southey in sending me the Australian DNA samples are also appreciated.

Dr. Hamdi Jarjanazi contributed directly to this project by assisting with allele calls and provided guidance while I worked in the laboratory. Keith Wong provided assistance while he worked for me as a summer student. Priscilla Chan ensured that I received DNA samples from the Ontario Family Breast Cancer Registry in a timely fashion. Xiu Hong Shi deserves special mention for the patience she demonstrated while providing me with my initial laboratory training. All of these individuals are, or were at one time, affiliated with Dr. Ozcelik's laboratory and they and the rest of the "Ozcelik lab" have my sincere gratitude.

Thanks also to members of Dr. Bharati Bapat's laboratory who kindly tolerated my presence in their work space and provided assistance on many occasions. I am also grateful to Ian Craine for reading and providing comments on this manuscript.

Table of Contents

List of Tables ..... **ix**

List of Figures ..... **xi**

Chapter 1. Introduction **1**

1.1. Hypothesis and main objectives **2**

Chapter 2. Background **4**

2.1. Breast Cancer **4**

2.1.1. Background and risk factors 4

2.1.1.1. Established risk factors 4

2.1.1.2. Genetic factors 7

2.2. Mammographic density **9**

2.2.1. Background and risk factors 9

2.2.1.1. Mammographic density and breast cancer risk factors 10

2.2.1.2. Genetic factors 13

2.3. IGF-I **14**

2.3.1. IGF-I Biology 14

2.3.2. Anthropometric and lifestyle factors related to circulating IGF-I levels 15

2.3.3. Genetic factors related to circulating IGF-I levels 16

2.4 IGF-I, IGFBP-3 and association with breast cancer risk and mammographic density **16**

2.4.1. IGF-I, IGFBP-3 and breast cancer risk 16

2.4.2. IGF-I, IGFBP-3 and mammographic density 21

2.4.3. Summary of association of IGF-I and IGFBP-3 with breast cancer and mammographic density 21

2.5. *IGF1* gene **22**

2.5.1. Linkage disequilibrium at *IGF1* 22

2.6. *IGF1* gene circulating concentrations of IGF-I, mammographic density and breast cancer risk 24

2.6.1. *IGF1* gene and circulating concentrations of IGF-I 25

2.6.1.1. Summary of *IGF1* gene and circulating concentrations of IGF-I 27

2.6.2. *IGF1* gene and mammographic density 31

2.6.3. *IGF1* gene and breast cancer risk 31

2.6.3.1. 5′ polymorphism and breast cancer 31

2.6.3.2. 19 allele and breast cancer 32

2.6.3.3. 19 allele and premenopausal breast cancer 36

2.6.3.4. 19 allele, family history, and breast cancer 36

2.6.3.5. Other 5′ polymorphism alleles and breast cancer 37

2.6.3.6. Other *IGF1* polymorphisms and breast cancer 38

2.6.3.7. Summary *IGF1* and breast cancer risk 39

2.7. Summary of literature 41

2.8. Description of study 43

Chapter 3. Methodologic issues 46

3.1. Design 46

3.1.1. Family based and cross-sectional studies 46

3.1.2. Selection of polymorphisms 49

3.2. Measurement 50

3.2.1. Breast Cancer 50

3.2.2. Mammographic density 50

3.2.3. IGF-I and IGFBP-3 levels 51

3.2.4. Height, weight, and BMI 52

3.2.5. Genotyping 53

3.3. Analysis 54

3.3.1. Family based association tests 54

3.3.2. Multiple testing 56

3.3.2.1. Cross sectional study 56

3.3.2.2. Family based study 57

3.3.2.3. Interpretation of results 58

Chapter 4. Association of variant alleles of an *IGF1* CA repeat polymorphism with mammographic density and anthropometric measures in premenopausal Caucasian women 59

Chapter 5. Family based study examining the association of *IGF1* polymorphisms and breast cancer risk in premenopausal women 101

Chapter 6. Family based study examining the association of *IGF1* polymorphisms with body mass index, weight and height 148

Chapter 7. Discussion 181

7.1. Summary of purpose 181

7.2. Description of study 183

7.3. Summary of results 183

7.3.1. *IGF1*, IGF-I concentration, mammographic density, and breast cancer risk 185

7.3.1.1. Hypothesized model 185

7.3.1.2. Interpretation of results in context of the model 186

7.4. Potential sources of error 188

7.4.1. Sampling 188

7.4.1.1. Limitations of cross-sectional studies 188

7.4.1.2. Recruitment in cross-sectional study 190

7.4.1.3. Limitations of family based studies 190

7.4.1.4. Recruitment in family based study 192

7.4.1.5. Breast cancer rates in source populations 196

7.4.1.6. Discrepancies in results between cross-sectional and family based study 196

7.4.2. Measurement error 198

7.4.2.1. Genotyping 198

7.4.2.2. IGF-I and IGFBP-3 200

7.4.2.3. Mammographic density 200

7.4.2.4. Anthropometric measures 201

7.4.2.5. Breast cancer 202

7.4.2.6. Density of polymorphisms 202

7.5. Power 203

7.6. Final results interpreted in context with other studies 205

7.6.1. *IGF1*, IGF-I concentration, mammographic density, and breast cancer risk 207

7.7. Conclusions and future research 210

References 213

Appendix 1. Contribution of the author to the design and implementation of the project 235

List of Tables

Chapter 2

Table 1. Relative risk for selected breast cancer risk factors. 5

Table 2. Quantitative studies of breast density and breast cancer risks: summary of methods and results. 12

Table 3. Association of circulating IGF-I and breast cancer risk in premenopausal women. Cohort studies. 20

Table 4. Comparison of mean circulating IGF-I concentrations (ng/mL) in relation to 19 repeat allele. 29

Chapter 4

Table 1. Association of IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio with percentage breast density in premenopausal women. 63

Table 2. Selected characteristics for study subjects. 72

Table 3. Counts and percent frequency (in brackets) for alleles and genotypes of three *IGF1* polymorphisms. 75

Table 4. Pearson correlation among variables in this study 76

Table 5. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and IGF-I, IGFBP-3 levels, IGF-I/IGFBP-3 ratio and GH levels. 77

Table 6. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and mammographic features. 80

Table 7. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and anthropometric measures. ... 81

Table 8. Linkage disequilibrium between specific alleles. 83

Chapter 5

Table 1. Criteria for defining moderate/high risk families used by the OFBCR. 107

Table 2. Criteria for removing individuals or families when a Mendelian inconsistency was identified. 115

Table 3. Descriptive characteristics of ABCFR and OFBCR samples. 120

Table 4. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR. 124

Table 5. Linkage disequilibrium between specific alleles. 127

Table 6. Association between *IGF1* haplotypes and breast cancer risk in the ABCFR and OFBCR 130

Table7. Association between 5′allele groupings of *IGF1* and breast cancer risk in the ABCFR and OFBCR. 132

Table 8. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR in families with and without a first degree relative with breast cancer - family history vs. no family history. 135

Chapter 6

Table 1. Descriptive characteristics of ABCFR and OFBCR samples 164

Table 2. Association between alleles of *IGF1* polymorphisms and BMI in the ABCFR and OFBCR. 167

Table 3. Association between alleles of *IGF1* polymorphisms and weight in the ABCFR and OFBCR. 168

Table 4. Association between alleles of *IGF1* polymorphisms and height in the ABCFR and OFBCR. 169

Table 5. Association between *IGF1* haplotypes and BMI in the ABCFR and OFBCR. 172

Table 6. Association between *IGF1* haplotypes and weight in the ABCFR and OFBCR.... 174

Table 7. Association between *IGF1* haplotypes and height in the ABCFR and OFBCR. ... 176

Chapter 7

Table 1. Summary of main findings. 184

Table 2. Descriptive characteristics for ABCFR and OFBCR families and probands. 195

Table 3. Summary of main findings compared to those in the published literature. 206

List of Figures

Chapter 1

Figure 1. Proposed model for relationship of *IGF1* with IGF-I levels, mammographic density and breast cancer risk.. 2

Chapter 2

Figure 1. Linkage disequilibrium across *IGF1* in Caucasians. 24

Figure 2. *IGF1* genotype and breast cancer risk, odds ratios and 95% confidence limits. 35

Figure 3. The proposed model for relationship of *IGF1* with IGF-I levels, mammographic density and breast cancer risk. 45

Chapter 3

Figure 1. Computer assisted method for measuring mammographic density. 51

Chapter 4

Figure 1. Presence and absence of the 3′ 185 allele and mammographic features, for women in the upper and lower halves of circulating IGFBP-3 concentration.. 88

Figure 2. Presence and absence of the 3′ 185 allele and mammographic features, for women in the upper and lower halves of circulating IGF-I concentration. 89

Figure 3. Presence and absence of the 3′ 185 allele and mammographic features for women in upper and lower halves for the IGF-I/IGFBP-3 ratio. 90

Chapter 5

Figure 1. Recruitment of OFBCR study subjects and selection of OFBCR families for this study. 108

Figure 2. Recruitment of ABCFR study subjects and selection of ABCFR families for this study. 109

Chapter 6

Figure 1. Recruitment of OFBCR study subjects and selection of OFBCR families for this study. 154

Figure 2. Recruitment of ABCFR study subjects and selection of ABCFR families for this study. 155

Chapter 1. Introduction

A number of studies indicate that greater circulating concentration of insulin-like growth factor-I (IGF-I), a breast mitogen, is associated with increased breast cancer risk in premenopausal women, although results have not been entirely consistent (*1-3*). Furthermore, several (but not all) studies indicate that mammographic density, which reflects the relative areas of fat to epithelial and stromal tissue and is positively associated with breast cancer risk (*4*), is greater in premenopausal women with higher circulating IGF-I concentrations (*5-8*). A possible relationship of circulating IGF-I levels with both breast cancer and mammographic density in premenopausal women, suggests a mechanism of action where breast cancer risk is related to the increased proliferative activity and quantity of stromal and epithelial tissue in the breast resulting from greater circulating IGF-I concentration (*4*).

Twin studies have reported that 27% of the variation in breast cancers, 63% of the variation in mammographic density and 38-63% of the variation in IGF-I levels can be explained by additive genetic factors (*9-12*). Although these estimates should not be directly compared to one another, particularly since breast cancer is a dichotomous state while mammographic density and IGF-I levels are continuous variables, they do suggest that genetic factors play an important role in influencing breast cancer risk, mammographic density, and circulating IGF-I levels.

Given the evidence that IGF-I levels, mammographic density and breast cancer risk are in part influenced by genetic factors, a model can be proposed where specific genes modify circulating IGF-I levels, which as outlined above promote the proliferative activity and quantity of stromal and epithelial tissue in the breast, increasing mammographic density

and the risk of developing breast cancer. In this study, *IGF1* was chosen as a candidate gene that might influence IGF-I levels and its role in the proposed model is shown in Figure 1.

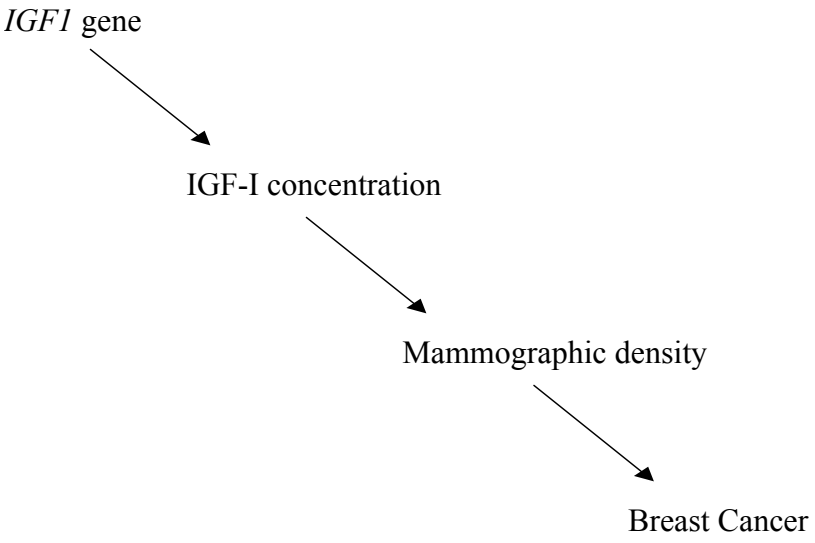

Figure 1. Proposed model for relationship of *IGF1* with IGF-I levels, mammographic density and breast cancer risk. The *IGF1* gene is hypothesized to alter circulating IGF-I concentrations. Greater circulating IGF-I concentrations stimulates proliferation of breast stromal and epithelial cells increasing both mammographic density and the risk of developing breast cancer.

1.1. Hypothesis and main objectives

In this study, it was hypothesized that consistent with the proposed model, genotypic variation at the *IGF1* locus is associated with variation in IGF-I concentration, mammographic density and the risk of breast cancer. Three cytosine-adenosine (CA) repeat polymorphisms, one a frequently investigated polymorphism at the 5′ end of the gene and

CA repeat polymorphisms in intron 2 and at the 3′ end of the gene, were examined for association with these three outcomes. The association of genetic variation at *IGF1* with IGF-I levels, mammographic density and breast cancer risk was also examined by constructing haplotypes using these three polymorphisms. The specific objectives are as follows:

1. Examine the association of each of the polymorphisms with circulating IGF-I levels.
2. Examine the association of each of the polymorphisms with percentage breast density.
3. Examine the association of each of the polymorphisms with breast cancer risk.
4. Determine linkage disequilibrium between the markers. Construct haplotypes and examine their association with circulating IGF-I levels, percentage breast density and breast cancer risk.

In addition, the association of *IGF1* with height, weight, and body mass index was examined. These are all risk factors for breast cancer and the latter two are also associated with mammographic density.

This thesis is written in journal format with background information for the study given in Chapter 2. A discussion of design and analysis issues is presented in Chapter 3. Individual findings are presented and discussed in chapters 4-6 and a general discussion of results is provided in Chapter 7.

Chapter 2. Background

This chapter provides the background information outlining factors associated with breast cancer, mammographic density and circulating IGF-I concentration (sections 2.1 to 2.3). Evidence relating circulating IGF-I concentrations to mammographic density and breast cancer risk (i.e. the final three components of the model presented in Chapter 1, Fig. 1) is presented in section 2.4. The first component of the model, the *IGF1* gene, is introduced in section 2.5. Association of the *IGF1* gene with circulating IGF-I concentration, mammographic density and breast cancer risk is then presented in section 2.6, followed by a critique and summary of the evidence supporting these associations.

2.1. Breast Cancer

2.1.1. Background and risk factors

Breast cancer is the most commonly diagnosed cancer in women in Canada with 18,824 new cases being reported in 2003. In 2001, the last year for which complete mortality data is available, there were 4,968 deaths due to breast cancer, exceeded only by lung cancer among cancer causes of death (*13*).

2.1.1.1. Established risk factors

Table 1, which is based on a table from Veronesi et al. (*14*), provides risk estimates for most of the established risk factors for breast cancer. Of these, age is the most strongly associated with risk. Although the risk of developing breast cancer increases with age, risk increases more slowly after about age 50, around the time of the menopause (*15*).

Table 1. Relative risk for selected breast cancer risk factors.

	Relative Risk	High-risk group
Age	>10	Elderly individuals
Age at menarche	3	Before age 11 years
Age at menopause	2	After age 54 years
Age at first full term pregnancy	3	First child after age 40 years
Breastfeeding and parity	Relative risk falls by 4.3% for every 12 months of breastfeeding in addition to a 7% reduction for every birth	Women who do not breast feed
Oral contraceptives	1.2	Current users
Hormone replacement therapy	1.7	Current users
BMI		
Premenopausal	0.7	High BMI
Postmenopausal	2	High BMI
Alcohol consumption	1.07	7% increase with every daily drink
Ionizing radiation	3	Abnormal exposure to young girls after age 10 years
Socioeconomic group	2[†]	High socioeconomic status
Previous benign breast disease	4-5	Atypical hyperplasia
Mammographic density	2.8-6.0[‡]	> 75% densities
Family history	≥ 2	Breast cancer in first-degree relative

From Veronesi et al. 2005 (*14*) except:
[†]Kelsey and Bernstein, 1996 (*16*).
[‡]Boyd et al. 2005 (*4*).

An association of breast cancer risk with exposure to ovarian hormones is supported by the observed association between several reproductive factors and risk. Early age at menarche, late age at menopause and late age at first pregnancy are associated with modest increases in risk (*17*). Greater number of full term pregnancies (*18*) and longer periods of time spent breast feeding, which results in a substantial delay in reestablishing ovulation (*19*), are associated with a small reduction in risk (Table 1) (*18*). Further support for a role for ovarian hormones comes from the observation that early removal of the ovaries reduces risk (*16*). Finally, greater circulating concentrations of endogenous estrogen are associated with increased risk, but only in postmenopausal women (*20*).

Exogenous hormones have been implicated in increasing the risk of breast cancer. Both hormone replacement therapy (HRT) and oral contraceptive use are associated with a slight increase in risk (Table 1) (*21,22*). The increased risk for combined estrogen and progestin HRT was established in a recent clinical trial (*22*). Results from observational studies suggest that HRT with estrogen alone also increases risk, but not to the extent as combined estrogen and progestin (*23-25*).

Obesity is associated with a modest increase in risk in postmenopausal women, but appears to be associated with a slightly decreased risk prior to menopause (*16*). Rapid childhood growth and greater attained adult height likely increases risk (*26*). These observations, in part, suggest a role for diet or energy balance in breast cancer etiology. A large number of studies have examined dietary factors and risk, but alcohol consumption is the only established dietary risk factor for breast cancer (*27*), with risk increasing slightly with amount consumed (Table 1) (*28*).

Ionizing radiation and high socioeconomic status are associated with modest increases in risk (*14*). Confirmed benign proliferative breast disease is a strong risk factor for breast cancer (Table 1) (*29*). A previous history of breast, ovarian or endometrial cancer, is associated with increased risk (*27*).

Mammographic density, reviewed below, is one of the strongest risk factors for breast cancer. Women with dense tissue in greater than 75% of the breast have a three to six times greater risk of breast cancer than those with little or no dense tissue in the breast (Table 1) (*4*).

Women with a family history of breast cancer are at increased risk with risk estimates of approximately 2 for women who have a first degree relative with breast cancer (Table 1).

The risk increases with increasing number of first degree relatives with breast cancer and is greater when relatives are affected at an early age (*30*). The association with family history is consistent with a role for genetic factors in the aetiology of breast cancer.

2.1.1.2. Genetic factors

In addition to the increased risk reported in women with a family history of breast cancer, results from a recent twin study provides support for a role for genetic factors in the aetiology of breast cancer. Analysis of combined data from Swedish, Finnish and Danish twin registries indicates an estimated 27% of the variance in breast cancer is attributable to genetic factors (*9*).

There are a number of identified gene mutations where associations with increased risk are well established. Most notable are mutations in the *BRCA1* and *BRCA2* genes which greatly increase the risk of developing breast cancer (*31*). Pooled estimates using data from patients unselected for family history indicate an average cumulative risk of 65% by age 70 years in *BRCA1* mutation carriers, and a cumulative risk of 45% by age 70 for *BRCA2* mutation carriers (*32*).

Carriers of homozygous or compound heterozygous mutations of the *ATM* gene suffer from the rare recessive degenerative disorder, ataxia-telangiectasia, characterized by progressive cerebellar ataxia, immunological dysfunction, hypersensitivity to ionizing radiation, oculocutaneous telangiectasia and a greatly increased susceptibility to cancer. Studies based on relatives of ataxia-telangiectasia cases indicate that female heterozygous carriers have a two to fourfold increased risk of breast cancer (*31,33*).

An increased risk of breast cancer is relatively well established for carriers of the 1100delC truncating mutation in *CHEK2*, a G2 checkpoint kinase that plays a critical role in

DNA repair. An approximately 2-fold increase in non-carriers of *BRCA1/2* mutations has been confirmed in a recent collaborative analysis (*34*). This mutation does not appear to increase the risk in *BRCA1* or *BRCA2* mutation carriers (*31*).

Li-Fraumeni syndrome and Cowden disease are very rare autosomal dominant disorders that predispose to breast and other cancers (*31,35*). Li-Fraumeni syndrome results from mutations in the *TP53* gene in the majority of Li-Fraumeni families, while in other families mutations in *CHEK2* have been described (*36*). Cowden disease results from mutations in *PTEN* (*35*). Estimate of lifetime risk of developing breast cancer is approximately 18 times that of the general population for Li-Fraumeni Syndrome (*35*). The excess risk for Cowden disease is estimated at approximately 30-50% (*31*).

Mutations in genes leading to ataxia-telangiectasia, Li-Fraumeni syndrome and Cowden disease and those at *CHEK2* explain only a small proportion of excess familial risk of breast cancer. A much larger proportion of excess familial risk, approximately 15-20%, is explained by mutations in *BRCA1* and *BRCA2* (*31*). Still the genes so far identified explain only a small proportion of excess familial breast cancer, approximately 20% (*31*), and an even smaller proportion of all breast cancer cases.

In recent years, candidate gene studies have investigated common variants that are suspected to have low penetrance, but because of higher frequencies could explain a large proportion of cancers. Genes in a number of different pathways have been investigated such as those involved in estrogen metabolism, DNA repair, and metabolism of carcinogens. Definitive associations between these common variants and breast cancer risk have been difficult to demonstrate, although a combined analysis provides convincing evidence that a common coding variant of the *CASP8* gene (the protein of which plays a role in apoptosis) is

associated with breast cancer risk (*31,37*). SNPs in or near several other genes *FGFR2*, *TNRC9*, *MAP3K1*, and *LSP1* have been identified as being associated with risk in large studies that included several replicate samples (*38,39*).

2.2. Mammographic density

2.2.1. Background and risk factors

Differences among women in the radiological appearance of the breast reflect differences in tissue composition. Fat is radiologically lucent and appears dark on a mammogram, while epithelial and stromal tissues are radiologically dense and appear light.

An association between mammographic pattern and breast cancer risk was first reported by Wolfe in 1976 (*40,41*). In more than 25 subsequent studies, the large majority of those conducted, mammographic patterns that reflected greater dense areas of the breast according to Wolfe's original classification scheme were reported to be associated with a greater risk of breast cancer (*42*). A criticism of this original classification scheme was that it was subjective, leading to great heterogeneity in risk estimates (*42*). Many studies have since used quantitative methods in assessing mammographic density (*43-57*). The results of these studies were summarized in a recent review (*4*) and are shown in Table 2. All of these studies show an increased risk of breast cancer with greater percentage breast density (amount of dense tissue divided by total area of the breast expressed as a percentage). Risk estimates for categories of percentage breast density of greater than 75% relative to less than 1-10%, range from 2.8 to 6.0 (four studies in Table 2). Consistent with these estimates are the results of a recent large nested case-control study, where an increased risk of 4.7 was reported for a comparison of these same two categories of density (*58*). Greater risk of breast cancer is also associated with greater amount of dense area, but of the three studies shown in Table 2 that

permit comparison, percentage breast density was associated with larger gradients in risk (*4*). Attributable risk estimates for density in more than 50% of the breast was estimated at 16%, 28% and 33% in three studies (*42,43,58*).

2.2.1.1. Mammographic density and breast cancer risk factors

Mammographic density is associated with a number of the known risk factors for breast cancer. Mammographic density is inversely associated with age and parity (*43,59-70*), increased by hormone replacement therapy (*71,72*), and decreased by Tamoxifen (*73-75*). Body weight and body mass index show a strong inverse correlation with mammographic density (*43,49,59-64,67-70,76-81*). Menopause is reported to be correlated with a reduction in mammographic density, although this reduction does not account fully for the effects of age on mammographic density (*82*).

The inverse association between mammographic density with age appears at first to be inconsistent with the relationship between age and breast cancer incidence. The prevalence of radiological dense tissue does, however, appear to be related to the rate at which breast cancer incidence changes in the population as shown by the slope of the age incidence curve, which is reduced following the menopause (*42*).

The inverse association between measures of body size and mammographic density is also inconsistent with the observed association of body size with breast cancer risk in postmenopausal women. It has been suggested, however, that mammographic density and body size operate through different pathways (*83*). In fact, results from a recent study suggest that not only are mammographic density and body mass index independent risk factors for breast cancer, but mammographic density may confound the relationship of body mass index with breast cancer risk. An inverse association of body mass index with breast cancer risk in

premenopausal women was reversed when percentage breast density was introduced into the model, while in postmenopausal women a positive association became stronger when percentage breast density was included in the model (*83*). (A confounding role for body mass index on the association of mammographic density with breast cancer risk was previously recognized (*4*)).

Several dietary risk factors examined in association with breast cancer risk have also been investigated in relation to mammographic density. There is some evidence for a positive association between alcohol intake and mammographic density, although statistically significant results have been reported in only two of seven studies (*50,68,77,84-87*). Observational studies have not provided consistent results with regards to the relationship between dietary fat intake and mammographic density (*50,85,88-90*). A reduction in mammographic density (specifically the amount of dense tissue) resulting from a low fat high carbohydrate diet was found in a dietary intervention study (*91*).

The known risk factors for breast cancer only explain about 20-30% of the variation in mammographic density (*4*). Much of the residual variation however, may be explained by genetic factors, and in fact, some breast cancer risk factors that are associated with mammographic density, such as BMI, are also strongly influenced by genetic factors (*92*).

Table 2. Quantitative studies of breast density and breast cancer risks: summary of methods and results.

	Age (years)	n	Type of measurement	Partition*	Odds ratio (95% CI)	Trend†	Adjustments
Nested case-control in cohort	NR	1880 patients 2152 controls	Planimetry	0 vs ≥75%	4·3 (3·1–6·1)	Yes	Weight, age at birth of first child, family history, years of education, alcohol use, previous benign biopsy sample, and number of reproductive years
Case-control	35–64	622 patients 443 controls	Computer assisted	<1% vs >75%	5·2 (1·7–16·1)	NR	Age, body-mass index, age at menarche, family history, number of full-term pregnancies, menopausal status, hormone use, and age at first full-term pregnancy
Cohort	40–80	111 patients 3100 controls	Computer assisted	0·5% vs >46%	3·49 (1·4–5·2)	Yes	Age, education, parity, height, and body-mass index
Nested case-control in cohort	40–59	354 pairs	Estimation by observer and computer assisted	0 vs ≥75%	6·0‡ (2·8–13·0) 4·0§ (2·1–7·7)	Yes Yes	Age, parity, age at birth of first child, weight, height, number of births, age at menarche, and family history
Case-control	40–65	183 pairs	Estimation by three observers	<10% vs ≥75%	6·0‖ (2·5–14·1) 2·8‖ (1·4–5·6) 3·7‖ (1·7–4·1)	Yes Yes Yes	Age at birth of first child, parity, and family history
Case-control	20–69	408 patients 1021 controls	Estimation by observer	0 vs ≥60%	5·4¶ (2·5–11·4) 3·8** (1·6–8·7)	Yes Yes	Parity, age at birth of first child, family history, age at menopause, and hormone use
Case-control	NR	362 patients 686 controls	Estimation by observer	0 vs ≥60%	4·4 (2·5–7·9)	Yes	Weight and height
Case-control	40–67	290 patients 645 controls	Estimation by observer	0 vs ≥60%	4·6 (2·4–8·5)¶ 3·2 (1·6–6·5)** 5·5 (2·3–13·2)††	Yes	Age, parity, education, weight, and height
Case-control	60 (mean)	647 pairs	Computer assisted	<10% vs >50%	1·8 (1·1–3·0)	No	Age at menarche, menopausal, status parity, age at birth of first child, family history, hormone use, and breast problems
Case-control	<50	547 patients 472 controls	Planimetry	<26·7% vs >70·3%	4·4 (3·0–6·7)	NR	Age and study
Case-control	>35	108 patients 400 controls	Computerised	<5% vs >25%	3·3 (1·5–7·2)	NR	Age, year of screening, menopausal status, and body-mass index
Case-control	>35	139 patients 553 controls	Computerised	<5% vs >25%	2·9 (1·6–5·6)	NR	Age and parity
Case-control	30–85	160 pairs	Planimetry	<20% vs ≥70%	4·3 (1·8–10·4)	No	Parity
Nested case-control in cohort	35–65	197 patients 521 controls	Planimetry	Upper vs lower‡‡	3·6‡‡ (1·7–7·9) 2·1§§ (1·1–3·8)	Yes Yes	Body-mass index, parity, and menopause
Nested case-control in cohort	35–74	266 patients 301 controls	Planimetry	<5% vs ≥65%	4·3 (2·1–8·8)	Yes	Age, weight, and parity

NR=Not reported. *Categories of least and most widespread density from which odds ratios were calculate. †Significantly increased risk of breast cancer across all categories of density analysed in study. ‡Area of density estimated by radiologist. §Areas of density calculated by computer-assisted measurement. ‖Results from individual observers. ¶Data for homogeneous density. **Data for nodular density. ††Data for total density. ‡‡Data for premenopausal patients. §§Data for postmenopausal participants.

References for table are from top to bottom 43-57.

Reprinted from the Lancet Oncology, 6, Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, et al., Mammographic breast density as an intermediate phenotype for breast cancer, 798-808, Copyright (2005), with permission from Elsevier.

2.2.1.2. Genetic factors

Genetic factors appear to have an important influence on mammographic density. A recent classical twin study has estimated the heritability of mammographic density in a population of Australian and North American twins to be 63%, after adjusting for age and other covariates (*10*).

Only a few studies have investigated the relationship between gene polymorphisms and mammographic density. Many of the genes investigated play a role in the estrogen pathway. Significant associations have been reported for polymorphisms of the estrogen receptor α gene (*93*), *AIB1* in postmenopausal women only (*94*), *3HSDB1*, although with opposing effects for the same allele in African-American and Caucasian women, (*95*), and *UGT1A1,* with opposing effects on percentage density in postmenopausal and premenopausal women (*94*). A significant association of *COMT* with mammographic density in premenopausal women has been reported in two studies, but this result was not supported in a third study (*94,96,97*). *CYP1A2* was reported to be associated with mammographic density in premenopausal women in one study, but an opposing effect for the same allele was found in a sample of pre- and postmenopausal women (*97*).

Variant alleles of two polymorphisms in the pituitary growth hormone gene were reported to be associated with percentage density (*98*). Variant alleles of a promoter region polymorphism of the *IGFBP3* gene was found to be associated with mammographic density in premenopausal women in one study (*99*). Two other studies, one which reported results stratified by menopausal status and the other that reported on a combined sample of pre- and postmenopausal women, did not provide support for this association (*5,100*).

Only the relationship of the *COMT, CYP1A2* and *IGFBP3* genes with percent density have been investigated in more than one study and results have been inconsistent. Results for the other genes require verification in future studies, particularly since some false positive results are likely, as many of the significant associations reported here came from the examination of sub-groups, or from studies that examined several genes or gene polymorphisms.

2.3. IGF-I

2.3.1. IGF-I Biology

IGF-I has characteristics of both a circulating hormone and a tissue growth factor. Most circulating IGF-I originates from the liver and growth hormone plays a dominant role in up-regulating hepatic production of IGF-I. IGF-I is also synthesized in peripheral tissues, including the breast, and can exert its effect through autocrine or paracrine mechanisms (*101,102*). IGF-I affects the proliferative behaviour of breast cancer cells (*103,104*) and plays an essential role in the normal development of breast tissue (*105*).

IGF-I is a ligand for the IGF-I receptor (IGF-IR), which is situated on cell surfaces. Binding to IGF-IR activates intracellular signaling pathways that promote cell proliferation and survival. Bioavailability of IGF-I is influenced by binding proteins (IGFBPs), six of which have been characterized. IGFBP-3 binds most of the circulating IGF-I and greatly prolongs IGF-I half life. In extracellular fluid, interactions between IGF-I and IGF-IR are modulated by IGFBPs, which can either increase or decrease IGF-I signaling. This complexity is poorly understood, but may be related to IGFBPs increasing IGF-I half life but also competing with cell surface receptors. Certain IGFBPs, including IGFBP-3, have a direct inhibitory effect on cell growth (*101*).

2.3.2. Anthropometric and lifestyle factors related to circulating IGF-I levels

In both sexes, IGF-I blood levels peak at puberty and then decrease gradually with age (*106*). There is some evidence that parity is inversely associated with circulating IGF-I levels (98). Hormone replacement therapy HRT (specifically oral estrogens) (*107-114*) and oral contraceptive use, both lower IGF-I levels (*115-117*). This may be due to a hepatic first pass effect of oral estrogen intake on IGF-I production by the liver (*116*). The inclusion of progestins in HRT formulations appears to attenuate or abrogate the effect of oral estrogens on IGF-I (*118*). Tamoxifen also reduces circulating IGF-I concentrations (*101*).

Studies that involve fasting male and female subjects indicate that restriction of caloric or protein intake results in lower IGF-I concentrations (*119*). Cross-sectional studies performed on largely well nourished populations generally do not support an association with caloric intake (*120*), although two studies using large samples from the Nurses Health Study and the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts suggest an association (*121,122*). These and other studies also suggest an association with total protein intake and circulating IGF-I levels (*120-122*).

A number of studies have examined the relationship between circulating IGF-I concentrations and measures of body composition in men and women. Studies that have measured body mass index have observed mainly null associations (*123-128*), but inverse (*118*), and positive associations (*129*) have been reported. A non-linear association between IGF-I concentrations and body mass index (BMI) is suggested by the results of some studies with lower IGF-I concentrations reported in individuals with relatively low or high BMI (*130-135*). There is some evidence for an inverse association between IGF-I levels and visceral adipose tissue in obese subjects (*136-139*) and a consistent inverse relationship

between IGFBP-I levels and BMI provides support for a relationship between the IGF-I system and body fat stores (*125,129,132,133,140-143*).

Although some studies report an association of circulating IGF-I with height in adults (*144-146*) many report no association (*107,118,123,127,132,133,147*). IGF-I levels do, however, appear to be related to height in childhood (*148,149*).

2.3.3. Genetic factors related to circulating IGF-I levels

Two studies of middle aged to elderly male and female twins indicate that 38-63% of variation in circulating IGF-I levels can be attributed to genetic factors (*11,12*). One of these studies also reported the heritability of IGFBP-3 levels to be 60% (*11*). These results indicate that genetic factors play an important role in determining circulating concentrations of IGF-I and IGFBP-3. The association of the *IGF1* gene with circulating IGF-I levels is examined in section 2.5.

2.4 IGF-I, IGFBP-3 and association with breast cancer risk and mammographic density

This section examines the evidence for the association of circulating IGF-I (and IGFBP-3) with the main outcomes of the model, breast cancer and mammographic density. An assessment of the evidence supporting these associations is also presented.

2.4.1. IGF-I, IGFBP-3 and breast cancer risk

A number of studies have examined the association between circulating IGF-I and IGFBP-3 concentrations and breast cancer risk, and the results have been summarized in two recent meta-analyses. Both indicated an increased risk of breast cancer in premenopausal women with increasing IGF-I and IGFBP-3 concentrations. One study, using stringent inclusion criteria, reported the results of four cohort studies (one of which did not include

IGFBP-3 measurement) and two case-control studies (*2*). Summary odds ratios were 1.65 (95% confidence interval: 1.14-1.95) for the 75th verses the 25th percentile of IGF-I concentration and 1.51 (95% confidence interval: 1.01-2.27) for the 75th verses the 25th percentile of IGFBP-3 concentration. Dose response analysis demonstrated a relationship between both increasing IGF-I and IGFBP-3 levels and breast cancer risk. Results did not provide support for an association of IGF-I or IGFBP-3 levels and postmenopausal breast cancer risk. Inclusion of data from a large cohort study (*150*), which was not used in primary analyses because of the lack of a clinical definition for menopausal status, did not alter the results appreciably (*2*). The second meta-analysis used less stringent inclusion criteria and examined 18 studies, five of which were cohort studies. Similar results were obtained with odds ratios of 1.39 (95% confidence interval: 1.16-1.66) for IGF-I and 1.42 (95% confidence interval: 1.15-1.74) for IGFBP-3, for comparisons based on circulating concentrations above and below the median in premenopausal women. Again, no relationship was observed in postmenopausal women (*1*).

Since the publication of these two meta-analyses, there have been seven additional studies reporting on IGF-I and IGFBP-3 concentrations and breast cancer risk with measurement of IGF-I and IGFBP-3 levels prior to breast cancer diagnosis. Three found no relationship between IGF-I levels and breast cancer risk in postmenopausal women (*151-153*), while two studies reported greater circulating IGF-I levels associated with breast cancer risk in women diagnosed after the ages of 50 and 60 respectively (*154,155*). IGFBP-3 levels were also reported to be associated with greater risk in postmenopausal women, women over 50, and women over 60 (*153-155*). Two of the three studies that examined risk in premenopausal women (one an update of an earlier study using the Nurses Health Study

cohort) found greater circulating levels of IGF-I to be associated with a 60% (top vs. bottom quintile) to 70% (top vs. bottom tertile) increase in risk (*151,152*). However in one of these, the result (*152*) was not statistically significant and the risk estimate was lower (i.e. 1.2), prior to the adjustment for IGFBP-3 levels. The third investigation using the Nurses Health Study II cohort found no association with risk (*156*). A lack of an association in women where cases were diagnosed under the age of 50 was also observed in two studies (although in one of these a positive association was observed when analyses were restricted to cases whose blood was drawn for IGF-I testing two years or more prior to diagnosis) (*154,155*). An update of a previous study by Toniolo et al. (*157*), found positive associations with IGF-I in premenopausal women. This was consistent with their previous results. However, when adjusted for IGFBP-3, the strength of the association was found to be strongly dependent on IGFBP-3 assay type, although at least an increased risk of 60% was observed (top vs. bottom quartile) (*158*). This study also reported an increased risk associated with IGFBP-3, although this was dependent on the assay used (*158*). Among other recent studies none reported a significant association between IGFBP-3 levels and breast cancer risk in premenopausal women (three studies) or women under 50 (two studies) (*151,152,154-156*).

The lack of consistency among recent studies of IGF-I and IGFBP-3 with conclusions from previous meta-analyses, has prompted a new meta-analysis of cohort studies examining circulating IGF-I levels and breast cancer risk in premenopausal women. Again, this analysis excludes results from two European cohort studies with a lack of a clinical definition of menopause (both of which reported null results for the association of IGF-I concentrations with risk in women under 50). A risk estimate of 1.69 (95% confidence interval: 1.17-2.45) comparing women in the 75th to the 25th percentile of circulating IGF-I concentration was

reported for the eligible cohort studies (*3*). The studies included in this analysis are shown in Table 3. Odds ratios for all but one are consistent with an association of greater circulating IGF-I levels with increased risk of breast cancer. Odds ratios for IGFBP-3 levels were not associated with breast cancer risk. Variation between studies in time between measurement of circulating IGF-I levels and diagnosis of cancer, lack of standardization of the assays used, and variability within and across populations studied (e.g. genetic variability) are possible reasons for inconsistency among the results (*3,156*).

Table 3. Association of circulating IGF-I and breast cancer risk in premenopausal women. Cohort studies.

Year	Cases/ controls	Category	Odds ratios (95% CI)	Variables matched or adjusted for[†]
1998 Hankinson (*159*)	76/105	Tertiles	2.88 (1.21-6.85)	Matching factors (Age, month/time/fasting status of blood draw, HRT) and adjusted for IGFBP-3
2000 Toniolo (*157*)	172/486	Quartiles	1.60 (0.91-2.81)	Matching factors (age at enrollment, time/day/phase of menstrual cycle at blood draw) and adjusted for history of benign breast disease, family history of breast cancer, parity
2002 Muti (*160*)	69/265	Quartiles	3.12 (1.13-8.60)	Adjusted for age, BMI, SES, age at menarche age at first child, parity, current HRT users excluded
2002 Krajcic (*161*)	66/66	Quartiles	2.01 (0.33-12.4)	Matching factors (Age, date of study health exam, duration of follow-up) and adjusted for insulin, glucose, BMI and IGFBP-3
2005 Allen (*152*)	70/209	Tertiles	1.71 (0.74-3.95)	Adjusted for BMI, age at first birth, age at menarche, and IGFBP-3
2005 Schernhammer* (*151*)	218/281	Tertiles	1.6 (1.0-2.6)	Matching factors (Age, month/time/fasting status at blood draw, HRT) and adjusted for IGFBP-3
2006 Schernhammer (*156*)	239/478	Quartiles	0.94 (0.63-1.42)	Matching factors (Age, month/time/fasting status and luteal date at blood draw, HRT, ethnicity) and adjusted for IGFBP-3

[†] Risk factors for breast cancer were generally not adjusted for as they did not appreciably alter results.
[*] Update of previous Nurses Health Study report (*159*).

2.4.2. IGF-I, IGFBP-3 and mammographic density

Several studies have investigated the association of insulin-like growth factor-I (IGF-I) with mammographic density, with five of seven providing at least some support for a relationship in premenopausal women (*5-8,99*). Mammographic density has been shown to be positively associated with IGF-I concentrations in blood samples of premenopausal women in three studies (*5,7,8*). Two of these studies also found a strong inverse correlation between premenopausal mammographic density and insulin-like growth factor binding protein-3 (IGFBP-3) levels and a positive association with the IGF-I/IGFBP-3 ratio (*5,8*). A fourth study observed an inverse association between mammographic density and circulating IGFBP-3 concentrations and a positive association with the IGF-I/IGFBP-3 ratio, but no association with IGF-I levels was found (*6*). A fifth study reported an association with IGF-I and IGFBP-3 levels and the amount of dense tissue (*162*). An association of IGF-I levels with amount of dense tissue was also reported by Boyd et al. (*7*). Eight studies have investigated the association of IGF-I and IGFBP-3 with mammographic density in postmenopausal women. One reported a significant association with IGF-I and the IGF-I/IGFBP-3 ratio (*163*). Results from the other studies did not support these observed associations (*5,7,8,99,162,164,165*) and none provide support for an association between mammographic density and IGFBP-3 levels (*5,7,8,99,162-165*)

2.4.3. Summary of association of IGF-I and IGFBP-3 with breast cancer and mammographic density

Overall, evidence from both case-control and cohort studies support an association between greater circulating concentrations of IGF-I and premenopausal breast cancer risk. There is, however, some inconsistency among results which could reflect difficulty in

measuring circulating IGF-I concentrations or effect modification by other factors. Current evidence does not support an association between IGFBP-3 concentrations and breast cancer risk. Most studies provide some support for an association of circulating IGF-I levels with mammographic density in premenopausal women. Therefore, there is at least some evidence to support the model of greater circulating IGF-I concentration increasing the proliferative activity and quantity of stromal and epithelial tissue in the breast, resulting in greater breast density and increased risk of breast cancer in premenopausal women. Still, the lack of consistency observed in the results underscores the need for better understanding of the various factors, including genetic factors, related to IGF-I concentration, mammographic density, and breast cancer risk.

2.5. *IGF1* gene

The *IGF1* gene has been localized to 12q22.1-q24.1 (*166*). The genomic sequence is approximately 85 kb in length (*167*) and has 6 exons (*168*). About 200 SNPs have been identified through ongoing genotyping programs: dbSNP, Seattle SNPs, HUGO and the HapMap project.

2.5.1. Linkage disequilibrium at *IGF1*

Linkage disequilibrium is a central concept in genetic epidemiology. It is defined as the (nonrandom) association between two or more alleles such that certain combinations of alleles are more likely to occur together on a chromosome than other combinations of alleles. Strong linkage disequilibrium across all or parts of a gene facilitates the study of the association between common genetic variants and outcome, since the actual functional variant may not necessarily need to be genotyped to show an association. Instead, the

genotyping of polymorphisms that may be in linkage disequilibrium with an ungenotyped functional variant (these are often referred to as markers or when appropriate tagging SNPs) can be used to test for an association with disease. The international HapMap project has identified many SNPs throughout the human genome (including *IGF1* SNPs) that can be utilized as markers for this purpose. DNA samples from 90 individuals comprised of 30 families with two parents and a child were obtained from a U.S. Utah population with Northern and Western European ancestry, and the Yoruba people from Ibaden, Nigeria. As well, 45 DNA samples were obtained from unrelated Japanese in Tokyo, Japan, and also from Han Chinese in Beijing China (*169*). Phase I of the HapMap produced genotype data for 1.3 million SNPs and a further 2.1 million SNPs were genotyped on the same sample in phase II (170).

Figure 1 shows HapMap data for the *IGF1* gene (*170*). Strong linkage disequilibrium does not extend the length of the gene. However, typical of the human genome in general there are blocks of strong linkage disequilibrium, four in all according to the criteria of Gabriel et al. (171). This block like structure is advantageous as it reduces the number of SNPs that are required to be genotyped in order to capture genetic variation at other ungenotyped loci. A commonly used criteria for selecting tagging SNPs is to ensure that a minimal set of SNPs are genotyped that ensure that all other identified SNPs show an R^2 of 0.8 or greater with at least one member of the minimal set. For *IGF1* this would require genotyping of 26 tagging SNPs (in order to capture variation in all other SNPs with a minimum minor allele frequency of 0.1% or greater, using the Haploview program Tagger and the Northern and Western European HapMap sample from release 22).

The HapMap data were not available when this study began. However, the concepts of using markers to identify unknown variants was central to the design of this study (see below) and the interpretation of other studies that are discussed in this thesis.

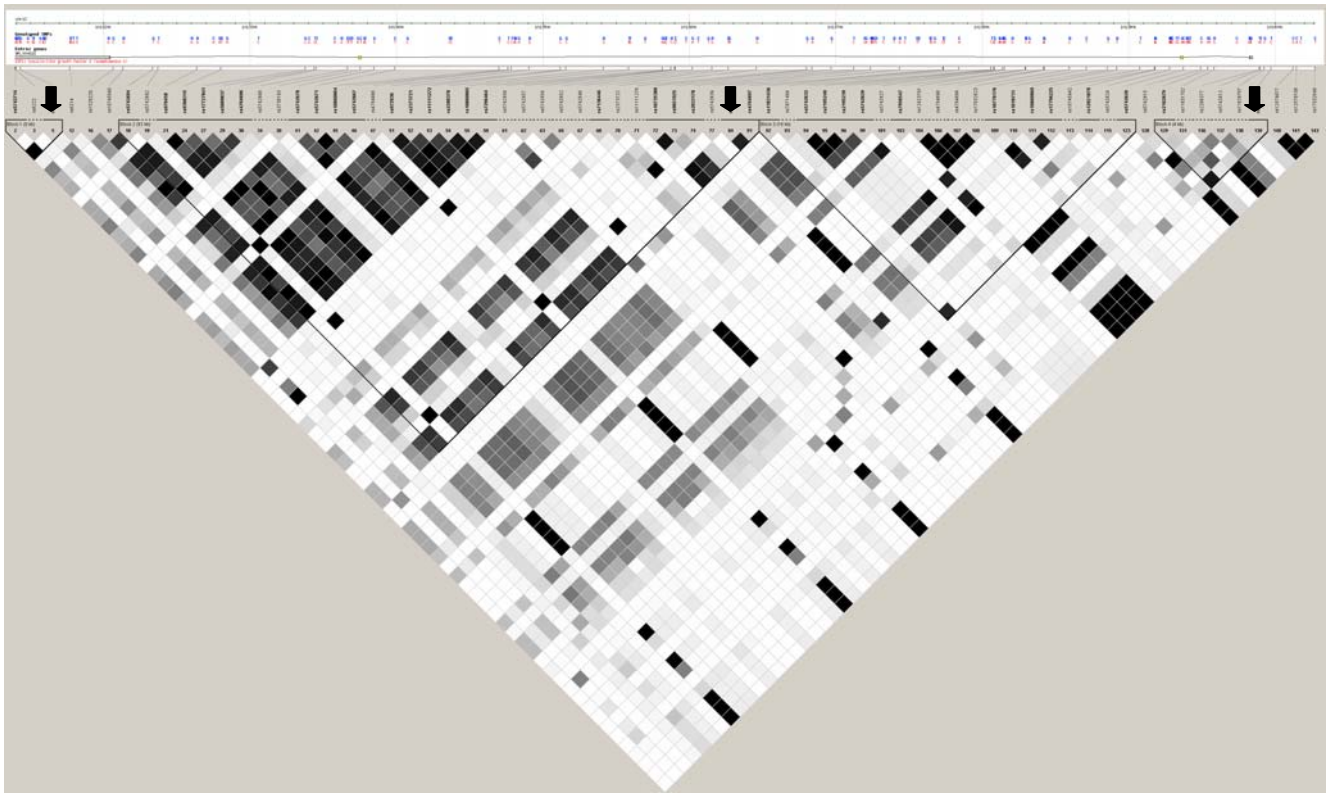

Figure 1. Linkage disequilibrium across *IGF1* in Caucasians. The 5′ end of the gene is on the right side. Blocks (triangular regions) were determined according to Gabriel et al. 2002 (*171*). Arrows show positions of 5′, intron 2 and 3′ polymorphisms (from right to left). Grey scale colours represent strength of linkage disequilibrium between markers according to R^2, (white: $R^2 = 0$, black: $R^2 = 1$, shades of grey $0 < R^2 < 1$). The SNP data set source was HapMap data release 22, phase II, April 2007, Northern and Western European ancestry sample.

2.6. *IGF1* gene circulating concentrations of IGF-I, mammographic density and breast cancer risk

This section presents the evidence for a relationship between the *IGF1* gene (the first component in the model) and each of circulating IGF-I concentration, mammographic

density, and breast cancer risk. The strength of the evidence for each of these associations is also evaluated.

2.6.1. *IGF1* gene and circulating concentrations of IGF-I

The association of a 5′-CA repeat polymorphism with circulating IGF-I concentrations has been examined in a number of studies. This polymorphism has a number of alleles and the most common has 19 CA repeats and a frequency of 60% or greater in Caucasian populations (*172-175*). This repeat lies in the promoter region of the *IGF1* gene. There are no experimental studies to indicate that it is functionally relevant, but association with phenotype could still arise if it is in linkage disequilibrium with another functional variant in the gene.

Most of the studies that have reported on the association between alleles of this polymorphism and IGF-I levels, have used a categorization scheme that is based on either the number of 19 repeat alleles, or the presence of absence of this allele. Most study samples are either restricted to one race or ethnic group or results are reported by race/ethnicity. Table 4 shows results for all studies that included women. Some studies did not provide results for males and females separately, for these combined results are presented. The population studied in the Netherlands was assumed to be Caucasian, although this information was not provided by the study authors (*176-178*). Covariates matched or adjusted for are shown in the table.

The majority of studies have reported on Caucasian or primarily Caucasian populations. Of the five studies that examined Caucasian women, only one (using a sample from the Nurses Health Study), reported a significant association, in which homozygotes for the 19 allele had greater circulating concentrations of IGF-I than women who did not carry

this allele (*179*). A second study using a sample from the Nurses Health Study did not observe this association, although different groupings for 19 and non-19 alleles were used while the ethnic mix of this population was not reported (*180*). Two studies (conducted in the U.S. and the U.K) which reported combined results for Caucasian men and women observed significantly lower IGF-I levels in homozygotes for the 19 allele, although allele groupings for the comparison group differed (*175,181*). A third study conducted in the Netherlands (presented as one study in Table 4 as they each used largely the same sample) reported greater IGF-I levels in males and females homozygous for the 19 allele relative to non-carriers (*176-178*). A lack of consistency between association of the number of 19 alleles and IGF-I levels was also observed among U.S. and U.K male populations (*173,175*). It therefore does not appear likely that the association of the 19 allele with circulating IGF-I levels would be clarified had all studies provided results for females only. It is possible that the contradictory results can be explained by different patterns of linkage disequilibrium, although this would require important differences in linkage disequilibrium patterns between populations in the Netherlands and those in both the U.S. and U.K.

Significant associations based on genotype categorizations using the 19 allele and IGF-I levels have not been found among African American, Japanese American, Latino White, Chinese or Korean populations (*172,182-185*). Homozygotes for the 20 allele in a Korean sample were found to have significantly increased IGF-I levels (*185*).

Three studies have examined *IGF1* genotype and IGF-I levels in premenopausal women, the group that is the focus of this dissertation. Two did not find a significant association with the number of 19 alleles and IGF-I levels (*99,179*). The other study, (not

shown in Table 4) reported that presence of the 19 allele was associated with decreased IGF-I levels in oral contraceptive users only (*115*).

Rietveld et al., have suggested that inconsistencies in the association between the 19 allele and IGF-I levels may have resulted from a more complicated pattern of association between alleles of this polymorphism and circulating IGF-I concentrations (*186*). Their analysis of a sample from the Netherlands indicated that alleles shorter than the 19 allele and those longer than the 20 repeat allele, were associated with lower IGF-I levels (*186*). Few other studies have looked at alternate alleles. A sample of women from the Nurses Health Study found a non-significant trend of lower IGF-I levels with increasing CA repeat length genotype (*179*), although allele categorizations do not permit a meaningful comparison with the results of Rietveld et al. Among Singapore Chinese, IGF-I levels were not associated with the number of 21 alleles (*183*), and no alleles shorter than the 19 repeat allele were found to be significantly associated with IGF-I levels in a Korean population (*185*). These latter two studies appear to contradict the Rietveld study, although it is possible that the disagreement could be explained by different patterns of linkage disequilibrium among Chinese, Korean and Caucasian populations.

Two studies have investigated other *IGF1* gene polymorphisms in relation to IGF-I levels. A promoter region SNP was not found to be associated with IGF-I concentrations in a Chinese population (*183*), but significant associations were found for several tagging SNPs spanning the *IGF1* gene in a largely Caucasian population in Great Britain (*187*).

2.6.1.1. Summary of *IGF1* gene and circulating concentrations of IGF-I

In conclusion, there is little evidence to support an association between alleles at the 5′ polymorphism and IGF-I levels. Some of the significant results observed for the 5′ 19

allele may be explained by confounding due to population stratification or chance. Associations with alleles other than 5′ 19 cannot be ruled out, but without replication, these results should be viewed with caution since spurious associations may have resulted due to multiple testing. Recent results from a large study in Great Britain that genotyped several SNPs provide some support for an association with allelic variants of *IGF1* and circulating IGF-I concentrations (*187*).

Table 4. Comparison of mean circulating IGF-I concentrations (ng/mL) in relation to 19 repeat allele.

Sample	Sex (n)	Homozygote	Heterozygote	Homozygote or Heterozygote	Non-carriers	Heterozygote or non-carrier	Variables matched or adjusted for	P-value
Caucasian, * U.S.A, All women Premenopausal Missmer (*179*)	F (418) F(70)	173 193			146 155		Age, fasting status, date and time at blood draw, and excluding HRT users	0.005 0.10 0.78[a]
Netherlands, Vaessen, Schut Rietveld (*176-178*)	M+F (150) (192)	21[†] 19[†]			17[†] 17[†]		Age, sex, BMI	0.003 0.01
Non-Hispanic white, U.S.A., Slattery (*182*)*1*)	F (233)	124	118		116		Age, menopausal status	0.76
Caucasian, Canada, Premenopausal Postmenopausal Lai (*99*)	F (174) F (183)	174 144	177 149		176 163		BMI, parity/ age, BMI, family history of breast cancer, alcohol, coffee	0.7 0.07
Caucasian, U.K., Frayling (*181*)	M+F (640)	133	144		143		Age, sex	0.01
Non-Latino White, U.S.A., Dellelis (*172*)	F (58)	146	141		157		Age	>0.05
Latino American, U.S.A., Dellelis (*172*)	F (154)	117	126		148		Age	>0.05
Hispanic, U.S.A., Slattery (*182*)	F (169)	107	116		120		Age, menopausal status	0.37

Continued on next page.

Table 4 Comparison of mean circulating IGF-I concentrations (ng/mL) in relation to 19 repeat allele (continued from previous page).

Sample	Sex (n)	Homozygote	Heterozygote	Homozygote or Heterozygote	Non-carriers	Heterozygote or non-carrier	Variables matched or adjusted for	P-value
African American, U.S.A., Dellelis (*172*)	F (123)	143	168		149		Age	>0.05
Japanese American, U.S.A., Dellelis (*172*)	F (71)	143	147		144		Age	>0.05
Chinese, Singapore, Wong (*183*)	M+F (628)	125‡	132‡		127‡		Age, sex, BMI, dialect group but reported crude	0.56
Caucasian, U.S.A. Rosen (*175*)	M+F (116)	129				154	None	0.03
U.S.A., Giovannucci (*180*)	F (404)	169				168	Age, month and fasting status at blood draw, year and indication of endoscopy	0.88
U.S.A., Morimoto (*188*)	M+F (323)	125				128	Age	>0.05
Caucasian, U.S.A., Kato (*184*)	F (23)			88	92		HRT, age < 40	0.79
African, U.S.A., Kato (*184*)	F (30)			111	96		HRT, age, smoking, height	0.35
Chinese, China, Wen (*189*)	F (351)			139	139		Age	0.95
Korean, Korea, Kim (*185*)	F (229)	NA	NA	NA	NA	NA	Age, BMI, years since menopause	>0.05[b]

[*] 8 non-Caucasian subjects
[†] nmol/L
[‡] Median
[a] Two 19 alleles vs. 1 19 allele vs. no 19 alleles
[b] Two 19 CA repeats vs. no 19 CA repeat
NA (not available)

2.6.2. *IGF1* gene and mammographic density

Two studies examined the association between genetic variation at *IGF1* and mammographic density. One examined the association of the 5′ 19 allele with density. Study subjects were healthy pre- and postmenopausal Caucasian subjects (206 pre- and 206 postmenopausal women) from hospitals in Toronto, Canada. Analyses were adjusted for age and additional confounders. No association between the number of 19 alleles and percentage breast density was found in either pre- or postmenopausal women, although there was some suggestion for an inverse association in postmenopausal women (*99*). A second study examined the association of *IGF1* haplotypes (particular combinations of alleles from two or more polymorphisms) and SNPs using a sample of primarily postmenopausal women from the Nurses Health Study cohort (n=1121). Age and other potential confounders were adjusted for (*100*). An association with mammographic density was observed for *IGF1* haplotypes and several tagging SNPs. Sample size did not permit a meaningful comparison by menopausal status.

2.6.3. *IGF1* gene and breast cancer risk

2.6.3.1. 5′ polymorphism and breast cancer

Seven case-control or nested case-control studies have reported on the *IGF1* gene promoter region CA repeat polymorphism and breast cancer risk. Three studies used either hospital based or other convenience samples for controls (*190-192*), the others used samples from defined populations or cohorts (*172,179,189,193*). In all but one study (*191*), investigators reduced the potential for confounding due to population stratification by either examining populations comprised either solely or largely of one racial or ethnic group (*179,189,190,193*), stratifying by race or ethnicity (*172*), or matching on race (*192*). Data on

potential confounding factors was collected in five studies, and reported odds ratios were adjusted for age or age and additional factors (*172,179,189,192,193*).

Four studies have reported a significant association with breast cancer risk, with two finding the 19 allele to be associated with increased risk of disease (*189,192*). Another study reported a protective effect of 21 and 22 repeat alleles (*191*), and the fourth found an increase in risk for carriers of alleles smaller than 19 repeats compared to 19 allele carriers, with significant results restricted to premenopausal women (*193*). Three studies did not find a significant association with breast cancer risk (*172,179,190*).

2.6.3.2. 19 allele and breast cancer

Figure 2 gives a summary of study results showing breast cancer risk as related to the number of 19 alleles, which again was the most commonly examined allele in relation to risk. Crude odds ratios are estimated for Missmer et al., as this study did not analyze risk in relation to the 19 allele (*179*). Crude odds ratios were estimated for Cleveland et al., for all women combined (since they reported results by menopausal status) and for premenopausal women (since they used homozygotes for the 19 allele as the reference group) (*193*). The use of crude odds ratios for these two studies should provide a reasonable estimate of adjusted odds ratios. Crude odds ratios did not differ from multivariate adjusted odds ratios by more than 10% for Cleveland et al. In the study by Missmer et al. there was a greater than 10% change between crude and multivariate adjusted odds ratios for four of seven different genotype categories where the 19 allele was at least included in the referent category, but these ranged from 16-26% and represented only 40% and 43% of cases and controls respectively. Odds ratios for Yu et al. (*192*) were adjusted for alcohol consumption, menopausal status (where appropriate) and the matching variables age and race. Both

Dellelis et al. (*172*) and Wen et al. (*189*) adjusted for age only, with Wen et al. reporting that they found no appreciable differences when other breast cancer risk factors were adjusted for. Wagner et al. reported crude odds ratios (*190*).

The results of Figer et al. (*191*) are excluded, since results based on the 19 allele were not presented and data were not available to assess the association between this allele and breast cancer risk. Results for different ethnic or racial groups and women with family history of disease are shown when they comprised distinct samples in the study and were presented as part of the main analysis. Results for premenopausal women are shown where available.

Among Caucasian populations (*172,179,190,193*) there is little evidence to support an association between the 19 allele and breast cancer risk, and the Cochrane-Mantel-Haensel odds ratio calculated from data available in these publications is close to unity (Fig. 2). A statistically significant odds ratio of 2.9 (*192*) was reported in a study examining a mainly African American population in Louisiana. The number of 19 repeat alleles also corresponded to relatively high odds ratios for African Americans in Hawaii and Los Angeles (Dellelis et al.) that were heterozygous for the 19 allele, but the odds ratio for the 19 allele homozygotes was close to one and for both homozygotes and heterozygotes confidence limits included one (*172*). Latino White populations in Hawaii and Los Angeles, also showed elevated odds ratios but again confidence limits included one (*172*). Among people of Japanese ethnicity, the odds ratio was slightly less than one for women heterozygous for the 19 allele and greater than one for those homozygous for this allele (*172*). The evidence for an association in other ethnic groups, at least in the U.S.A., appears limited and the relatively high odds ratios seen for some of these is probably due to small sample size. However, some

additional support for an association of the 19 allele with breast cancer risk comes from a large study conducted in Shanghai China, where a modest but statistically significant increase in risk (OR=1.2 for 19 allele carriers) was reported (*189*).

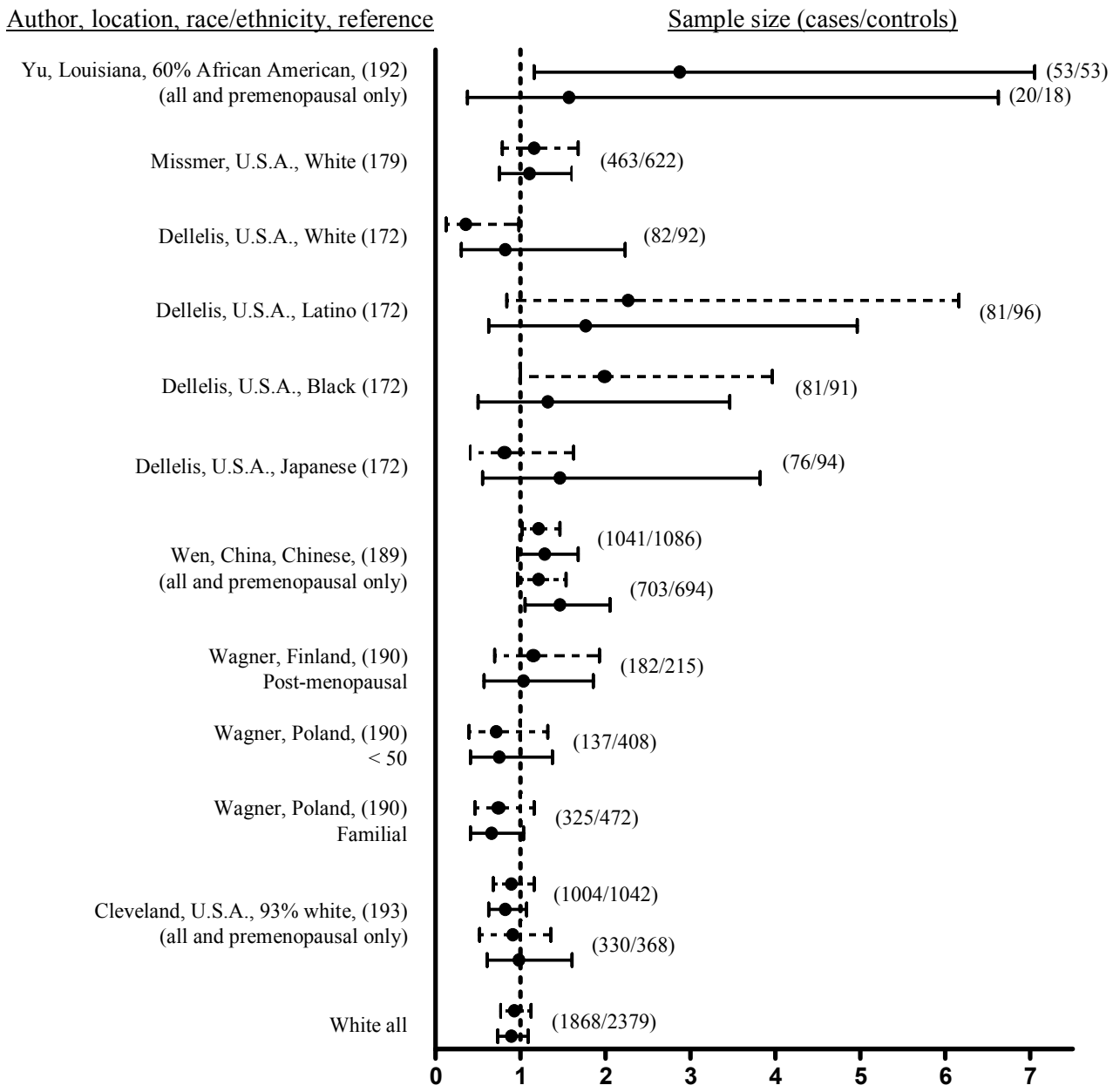

Figure 2. *IGF1* 5' 19 genotype and breast cancer risk. Odds ratios and 95% confidence limits shown for 19 allele heterozgotes (dashed lines) and homozygotes for each study with non-carriers of 19 allele as the referent group (except for Yu, where 19 allele carriers vs. referent are shown). Odds ratios and 95% CL for premenopausal sub-group are also shown in addition to pre-and postmenopausal combined for Yu, Wen and Cleveland. These are always the second or second pair of odds ratios shown for a study. Odds ratio and confidence limits for Cleveland (pre- and postmenopausal combined) and Missmer are crude odds ratios calculated by the author. Cochran-Mantel-Haensel odds ratio and confidence limits were calculated for Caucasians 'White All', which includes data from Missmer (>99% Caucasian), DeLellis (White only), Wagner (Polish and therefore assumed to be Caucasian), and Cleveland (93% Caucasian).

2.6.3.3. 19 allele and premenopausal breast cancer

Evidence for an association of circulating IGF-I concentrations with breast cancer risk is largely restricted to premenopausal women. However, only a few studies examined the association of IGF-I genotype and breast cancer risk stratified by menopausal status. As described above, a U.S. study (Long Island, New York) reported a significant increase in risk among premenopausal women whose genotype included alleles smaller than 19 repeats when compared to 19 allele carriers, while this same comparison showed a non-significant increase in postmenopausal women (*193*). A study conducted on women in China, found that the number of 19 alleles was significantly associated with breast cancer risk in premenopausal but not postmenopausal women (*189*). A Louisiana study reported a slightly lower but still positive association in premenopausal women with the 19 allele, but the sample size was very small (20 cases, 18 controls) and results were not statistically significant (*192*). The study using the Nurses Health Study sample did not find evidence to support an association for various 19 allele genotypes and breast cancer risk in either pre- or postmenopausal women, and risk estimates were not appreciably different in either group, although sample size for premenopausal women was again small (65 cases, 70 controls) (*179*). Results from a sample of Polish women under 50, which likely was comprised primarily of premenopausal women, suggested a slight protective effect for the 19 allele, although confidence limits included one (*190*).

2.6.3.4. 19 allele, family history, and breast cancer

Three studies examined association of the 19 allele with breast cancer risk in women with a family history of disease (*179,190*). Two reported a decrease in risk for 19 allele carriers (OR=1.51, 95% confidence limits: 0.96-2.39, when comparing non-carriers to those

homozygous for the 19 allele (*190*), and OR=0.49, 95% confidence limits: 0.20-1.23 for comparison of homozygotes for the 19 allele to all others (*179*)). A third reported a protective effect observed mainly in premenopausal women (OR=2.70, 95% confidence limits: 0.70-10.4 for comparison of non-carriers to 19 allele carriers (*193*)). None of these results were statistically significant.

2.6.3.5. Other 5′ polymorphism alleles and breast cancer

Some studies examined other alleles, or grouped alleles of the 5′ polymorphism in relation to breast cancer risk. As reported above, a study conducted on an Israeli population, found a significant protective effect for women whose genotype was composed of longer alleles (21 and 22 vs. all others). This study did not find a significant association for genotypes comprised of shorter alleles (11, 16 or 17) (*191*). In the Nurses Health study, homozygotes for the 19 allele were compared to six other common genotypes but no significant associations were found (*179*). No significant results were found with categorization based on average allele length, ((allele 1 length + allele 2 length)/2) or genotypes comprised of fewer than 17 CA repeats verses all others, or greater than 20 repeats verses all others, in the Louisiana study (*192*). The study of women from Shanghai examined several different genotypes in addition to women heterozygous or homozygous for the 19 allele. A significantly decreased risk was observed for women who carried the 17 allele (OR = 0.80; 95% CI = 1.04-1.47). This study also reported a positive but non-significant association (OR = 1.92; 95% CI = 0.92-4.02) with breast cancer risk for women who carried a copy of one of the rare alleles (< 1%) in this population (11, 13, 16 or 23 repeats). In addition to the comparison of women with genotypes comprised of alleles smaller than 19 repeats verses 19 repeat carriers, the Long Island study also compared women whose

genotype included alleles greater than 19 repeats to 19 allele carriers, and also created comparison groups based on cumulative allele length (greater that or less than 38 repeats (e.g. 19 repeats * 2)). A significant relationship with risk was not observed for either of these categorizations (*193*).

2.6.3.6. Other *IGF1* polymorphisms and breast cancer

Two studies that specifically reported on multiple SNPs across *IGF1* reported no significant association of *IGF1* polymorphisms with breast cancer risk after adjustment for multiple comparisons. Each, however, found nominally significant associations with different pairs of SNPs at the 5′ end, (*194,195*) three of four of which are in strong linkage disequilibrium with each other (as determined by using data from the HapMap project (*196*)). A third study, conducted in Great Britain, reported significant associations with 5 tagging SNPs across the *IGF1* gene (*187*), although there was inconsistency with results from another study that examined one of the same tagging SNPs (*194*). In addition, some caution is necessary in interpreting the British study as 27% of cases were prevalent cases.

None of the studies reported associations stratified by menopausal status and two of the three had only a moderate sized sample of premenopausal women. The British study included an approximately 70% sample of premenopausal women in a study with over 4500 cases and controls. However, recruitment of prevalent cases among largely younger women probably resulted in even a greater proportion of prevalent cases among this group than the 27% reported for the entire sample (*187*).

In addition to the studies described above, three genome wide scans have recently reported their findings (*38,39,197*). These studies used initial screens on a smaller samples of individuals (using approximately 250,000 to 500,000 SNPs per study) and then tested the

most significant SNPs on larger samples. No *IGF1* SNPs were reported to be associated with risk. Only one of these studies provided results related to *IGF1* SNPS with no significant associations (at the P<0.05 level) found in the sample of postmenopausal women examined. One of the studies included a large sample of younger and therefore likely premenopausal women in the initial screening sample (*39*). However, this sample was much smaller than studies the have specifically investigated *IGF1* SNPs (approximately 400 cases and 400 controls verses 1000-4000 cases and similar numbers of controls) and it is possible that associations with *IGF1* were therefore missed. As well, true associations may have been missed because of the need to use adjust for the many comparisons made.

2.6.3.7. Summary *IGF1* and breast cancer risk

Although some studies have reported a significant relationship between variant alleles of the 5′ polymorphism and breast cancer risk, overall the data are inconsistent. In Caucasian populations the association of the 19 allele with breast cancer risk appears to be close to unity (*172,179,190,191,193*). Evidence for an association between the number of 19 alleles and breast cancer risk appears strongest in African-American and Chinese populations (*189,192*). Examination of different alleles or allele groupings have produced significant associations but results are not consistent across studies (*179,189,191,192,196*). The association of variant alleles of the 5′ polymorphism with breast cancer risk in premenopausal women (who are the subject of this investigation) is also inconsistent. Among studies with adequate sample size, either no significant association was observed or there is disagreement about which alleles have been found to be associated with disease (*189,190,193*).

The inconsistent results could be the result of different patterns of linkage disequilibrium across the *IGF1* gene (particularly among different ethnic groups) or effect

modification by other genetic or environmental factors. Variation in risk estimates may also be due to bias resulting from population stratification. Inconsistency in results may also be explained by false positive associations resulting from multiple testing.

Another limitation that may explain the inconsistent results is that strong linkage disequilibrium does not extend across *IGF1* (Fig. 1). Therefore the 5′ polymorphism may not predict the presence of causal variants. It is also possible that several alleles are in weak linkage disequilibrium with another causal variant, which could explain the sporadic associations with risk that have been observed. Some recent studies have examined several polymorphisms to better account for genetic variation across the *IGF1* gene. Two studies have provided suggestive results for SNPs at the 5′ end (*194,195*), although different SNPs were tested and results were not statistically significant after adjustment for multiple testing. A third study found several tagging SNPs to be associated with risk (*187*). The interpretation of the results are however complicated by the presence of prevalent cases in this study and results are inconsistent with a previous study that genotyped one of the same tagging SNPs. Studies examining *IGF1* tagging SNPs then, have not provided convincing evidence for an association with risk.

Only one of the studies that used tagging SNPs included a large sample of premenopausal women, the same study that included prevalent cases (*187*). None of the studies that used tagging SNPs reported on *IGF1* and breast cancer stratified by menopausal status.

Three recent genome wide scans have failed to show an association with *IGF1*. However, these have not used large samples of premenopausal women at the initial screening

stage and this and the need to adjust for the many comparisons made may have resulted in true associations being missed.

2.7. Summary of literature

As outlined above, although not entirely consistent, evidence from epidemiologic studies suggests that greater circulating IGF-I levels increase both mammographic density and the risk of breast cancer in premenopausal women. The relationship of circulating IGF-I levels with both breast cancer and mammographic density in premenopausal women suggests that IGF-I action affects the proliferative activity and quantity of stromal and epithelial tissue in the breast, resulting in increased risk of breast cancer (*4*).

Circulating IGF-I levels, breast cancer risk, and mammographic density, all have a heritable component suggesting that understanding the influence of genetic factors may be important in relating breast cancer risk to these two phenotypes. Above a model was proposed where *IGF1* gene variants modify circulating IGF-I levels, which in turn promote the proliferative activity and quantity of stromal and epithelial tissue in the breast, increasing mammographic density and the risk of developing breast cancer.

Most research into the relationship between *IGF1* and circulating IGF-I concentrations has examined a single (5′) CA-repeat polymorphism. However, there is little evidence to support an association between alleles at the 5′ polymorphism and IGF-I levels. Recent results from a large study in Great Britain that genotyped several SNPs provide support for an association with allelic variants of *IGF1* and circulating IGF-I concentrations (*187*). This result underscores a need to use more markers to capture linkage disequilibrium patterns across *IGF1* in studies that look at genotypic variation and outcome (e.g. IGF-I concentrations or breast cancer risk).

A critical outcome in the proposed model is mammographic density. Two studies have examined the association of an *IGF1* gene polymorphism and mammographic density. One found no association with the 5′ 19 allele. The other using a large sample from the Nurses Health Study cohort found evidence that common genetic variation in *IGF1* was associated with mammographic density in a sample of mainly postmenopausal women. It is of interest to note that the SNP most strongly associated with mammographic density in this study was also found to be associated with IGF-I levels and breast cancer risk in Great Britain (*187*), but in the opposite direction to what would be anticipated if the same variant allele was responsible for increasing IGF-I levels, mammographic density, and breast cancer risk. Furthermore, evidence from a second study did not support the association of the same genetic variants with breast cancer risk as the Nurses Health Study sample (*195*). Further research into the association of *IGF1* with mammographic density is needed in order to clarify the relationship.

Most studies examining the relationship between *IGF1* and breast cancer risk have also only examined the 5′ polymorphism. Results from these studies have been inconsistent. Three studies have used tagging SNPs to better capture linkage disequilibrium across the *IGF1* gene. As described above, they have not produced entirely consistent results. Discrepancies may be explained by variation in other risk factors that modify the relationship between *IGF1* and breast cancer risk, confounding due to population stratification or different patterns of and/or weak linkage disequilibrium between this polymorphism and a causal variant at another *IGF1* locus. For the 5′ polymorphism the possibility of spurious associations must also be considered, as this polymorphism has a number of alleles which has resulted in some studies producing multiple tests of association.

Two of the important limitations of studies that have been described, not genotyping a sufficient number of markers to capture and genetic variation across the *IGF1* gene and possible confounding due to population stratification, can be viewed as methodologic issues. These can be addressed by examining more polymorphisms at *IGF1* and through study design (e.g. family studies). An interesting observation is that considering the number of studies performed, few reports have attempted to examine the association of genetic variation at *IGF1* with IGF-I concentrations or breast cancer risk in premenopausal women. It is this group in which the association of circulating IGF-I levels with breast cancer risk has been reported.

In addition several risk factors are related to mammographic density and or breast cancer (e.g., height, weight and body mass index). There is currently only a small amount of literature dealing with the association of *IGF1* and anthropometric measures and more attention should be given to this area.

2.8. Description of study

In Chapter 1, a model was presented, where *IGF1* gene variants modify circulating IGF-I levels, which promotes the proliferative activity and quantity of stromal and epithelial tissue in the breast, increasing the risk of developing breast cancer (Fig. 3). This model was investigated here by determining the association of variant alleles of the 5′ polymorphism and two other CA repeat polymorphisms with circulating IGF-I levels, mammographic density and breast cancer risk in premenopausal women. In addition, the association of *IGF1* with height, weight, and body mass index was examined (Fig. 3). All three of these variables are breast cancer risk factors, while the latter two are associated with mammographic density.

A family based design using families from the Ontario Familial Breast Cancer Registry and the Australian Family Breast Cancer Registry was employed to examine the relationship between genetic variation at *IGF1* and breast cancer risk. The advantage of a family based study is that this design is not subject to bias resulting from population stratification. The association of IGF-I level and mammographic density with *IGF1* was examined in a cross-sectional study of women sampled from local Toronto hospitals. The association of *IGF1* with anthropometric measures was examined in both of these samples.

Testing polymorphisms other than the 5′ polymorphism is of interest since the *IGF1* gene is large (approximately 85kb) and strong linkage disequilibrium does not extend across the whole gene. In this study, three polymorphisms that span the *IGF1* gene were tested.

In addition, linkage disequilibrium between polymorphisms is presented and haplotypes constructed and their relationship to breast cancer risk examined. The purpose of constructing haplotypes is to increase power to detect an ungenotyped causal variant on the gene. Some studies have shown stronger associations between genotype and disease by constructing haplotypes as opposed to examining single polymorphisms (*198*).

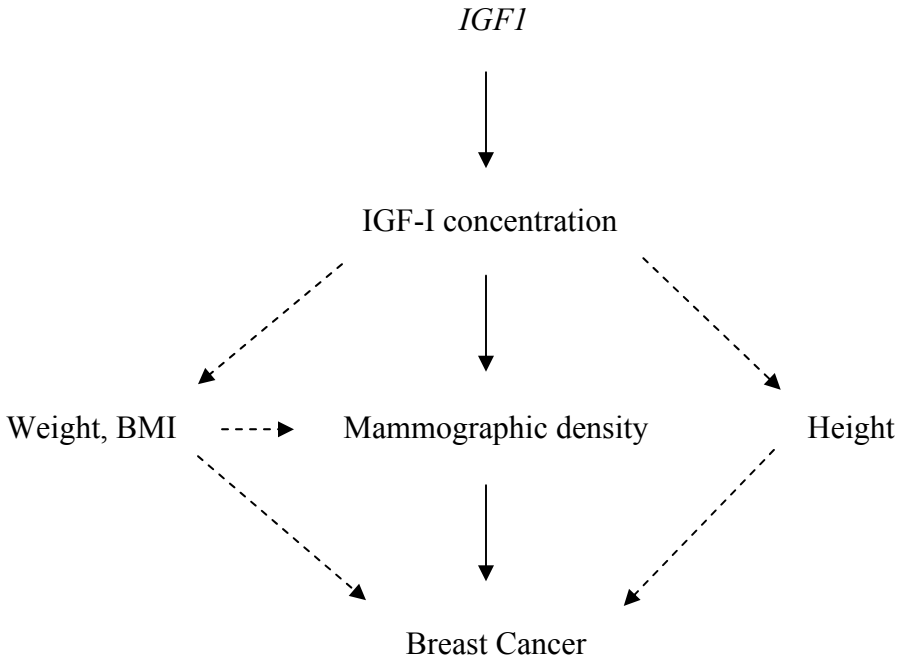

Figure 3. The proposed model for relationship of *IGF1* with IGF-I levels, mammographic density and breast cancer risk is shown with solid lines. The *IGF1* gene is hypothesized to alter circulating IGF-I concentrations. Greater circulating IGF-I concentration stimulates proliferation of breast stromal and epithelial cells increasing both mammographic density and the risk of developing breast cancer. In addition to the investigation of this main model, the association of *IGF1* with BMI, weight and height (anthropometric measures) was also examined. All three are associated with breast cancer risk and both BMI and weight are associated with mammographic density. These potential pathways are shown with dashed lines, although not all are examined (e.g., *IGF1*, IGF-I concentrations, anthropometric measures (i.e., BMI, weight or height) and breast cancer risk could not be examined since all were not available in a single sample).

Chapter 3. Methodologic issues

This section describes the most important methodologic issues related to the design of the studies presented, the measurement of variables, and statistical analyses. Specific details of methods can be found through chapters 4-6. The final chapter returns to a number of these issues and discusses them in the context of the results.

3.1. Design

3.1.1. Family based and cross-sectional studies

As described in the previous chapter, two separate designs were employed for the overall study, a cross-sectional design using unrelated individuals and a family based design. The family based design was employed to examine the association of *IGF1* genotype with breast cancer risk. The availability of self reported height, weight and therefore BMI, also permitted the examination of the association of *IGF1* genotype with these breast cancer risk factors. The advantage of family based studies relative to those that use unrelated individuals, is that they are not susceptible to bias from confounding due to population stratification. This is a particular form of confounding which can occur when allele frequencies vary with population sub-group, generally thought of as an ethnic group. If ethnicity is associated with disease because of differences in either genetic or non-genetic risk factors among ethnic groups, confounding of the relationship between genotype and disease can result.

The practical importance of this form of confounding has been questioned, as it has been argued that bias in general should be small and self reported ethnicity should be sufficient as a control, unless disease and allele frequencies vary widely (*199*). It has also been argued that restricting analyses to subjects of European ancestry, a common approach in

examining the association of genes with disease in case-control studies, should generally result in negligible bias when studying North American populations (*200*).

In contrast to these reports, a recent study by Campbell et al. (*201*) indicates that population stratification can result in strongly biased risk estimates in North American studies that use subjects of European background. Campbell et al. created a case-control panel with over 1000 individuals ranked in each of the 90th through 95th percentiles and the 5th through 10 percentiles for adult height, which is known to vary across Europe. The association between height and a *LCT* (lactase gene) polymorphism, known to show a strong gradient in allele frequencies across Europe, was tested. An association between genotype and height was observed, and two tests for population stratification using different panels of genetic markers (one with 67 markers (*202*) and the other with 111 markers (*203*)) failed to detect the presence of population stratification. Matching samples on parental ethnicity reduced the strength of the association, although it was still statistically significant. However, when analyses were restricted to Polish or Scandinavian individuals, there was no association between *LCT* and height. The latter result suggests that the observed association was the result of confounding by undetected population stratification that could not be controlled for using conventional methods.

As stated above, family based studies are not susceptible to bias from population stratification (*204*), and this as well as the availability of family based samples motivated the choice of a family based design. Disadvantages of family based studies include difficulties in recruitment (particularly since parents of older cases may no longer be alive), lower statistical power, bias resulting from genotyping error and potential difficulties obtaining representative samples resulting from the lack of availability of suitable relative controls. The

latter two issues are dealt with in chapter 7, where they are discussed in context of the final results. Study size and power, which of course are related to recruitment, are discussed here.

The family based association tests used in the analyses performed here can use either parents or siblings as controls (or any combination of these). These tests are based on reconstructing the parental genotypes to generate an expected genotype distribution. Therefore, families with both parents are generally the most informative, although when there are several siblings, parental genotypes can also be unambiguously reconstructed. Power in studies that use case-parent trios compares well to those that use cases and unrelated controls (*205*). However, if genotype data is only available for a single sibling, power is considerably reduced (*206*).

The studies presented in chapters 5 and 6 used probands (cases) and first degree relatives (parents or siblings) from two sites of the Breast Cancer Family Registry, the Ontario Familial Breast Cancer Registry (OFBCR) and the Australian Breast Cancer Family Registry (ABCFR). Since this study examined premenopausal probands, who were seldom older than age 50, there were a number of families where two parents were available. There were, however, many families with one parent and siblings, or siblings only (see Chapter 5 Table 3). A relatively small sample with only a modest number of two parent families was available from the OFBCR. This prompted the inclusion of families from the ABCFR so that combined analyses with suitable power could be performed. The overall sample size was 808 to 823 families (the number of families varied as genotyping failures and removal of data after the identification of potential genotyping errors varied for each polymorphism). The power to detect a relative risk in these families of 2 for homozygotes and 1.5 for heterozygotes of the most common allele of the previously investigated 5′ polymorphism

was 0.79 (calculated using PBAT version 2.5 (*207*), a sample size of 823 women with an allele frequency of 0.62 at α=0.05, and assuming an additive model and complete linkage disequilibrium between this allele and the disease allele). In comparison to other currently published studies that examined this polymorphism in relation to breast cancer risk in premenopausal women (*179,189,192,193*), only one, an investigation of a Chinese population (*189*), had greater power.

A cross-sectional design (chapter 4) was chosen for the investigation into the association of *IGF1* with IGF-I levels and mammographic density, since data were readily available from a previous investigation examining circulating IGF-I concentrations and breast mitogens in relation to breast cancer risk (*7*). The possibility of confounding due to population stratification of course exists, and the approach of restricting analyses to Caucasian subjects was used to minimize this potential source of bias.

This sample had 163 women and in the planning of the overall study design for this thesis, the sample size was considered fixed. However, as an example of study power, a 15.3% difference in mammographic density between women homozygous for, or non-carriers of, the most common allele of the 5′ polymorphism, would be detected with a power of 80%, with α set at 0.05 and assuming an additive model (using NCSS PASS (2006) (*208*)).

3.1.2. Selection of polymorphisms

Three repeat polymorphisms (microsatellites) were selected for analysis. These were not selected because they were hypothesized to have functional significance. Instead as previously discussed, the goal was to take advantage of linkage disequilibrium across *IGF1*. An association observed with one or more of these markers (or their haplotypes) would indicate the presence of an ungenotyped functional variant that was also associated with

outcome. Repeat polymorphisms were chosen since initial examination of the NCBI DNA sequence for *IGF1* revealed several CA repeats distributed across the gene and for practical reasons (cost and genotyping effort), three polymorphisms was a reasonable number for investigation.

3.2. Measurement

A brief discussion of methods used for measurement of outcome variables and genotype is given here. This includes some discussion of accuracy and reliability, although these are discussed in detail in the individual chapters presenting the results (chapters 4 to 6), and in the final discussion (chapter 7).

3.2.1. Breast Cancer

All breast cancer diagnoses in the family based study were primary cases and verified by pathology report. Therefore, accuracy of diagnosis should not be an issue here.

3.2.2. Mammographic density

There are no generally accepted standard methods for classification of variation in radiological appearance of breast tissue (*4*). Mammographic density measurements in this study were performed using a previously developed computer assisted method described in Figure 1 (*209*). This method has been shown to produce similar gradients in risk estimates for breast cancer as those produced when categorization of mammographic density is performed by radiologists. The reliability within and between observers using this method is quite high (0.90 or greater) (*46,209*).

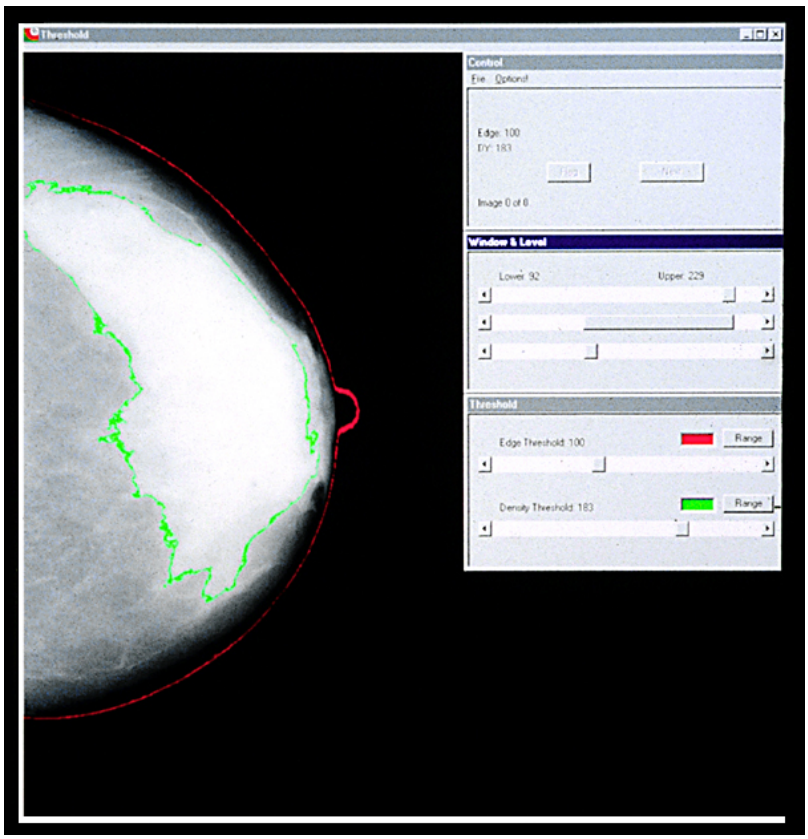

Figure 1. Computer assisted method for measuring mammographic density. Digitized film of mammograms are projected onto a screen and a grey level threshold is first selected to separate the breast from the background and then to separate the dense tissue from the non-dense (fat) tissue. Sums of the dense and total areas are automatically generated and % breast density can be calculated

3.2.3. IGF-I and IGFBP-3 levels

IGF-I and IGFBP-3 were measured using a competitive binding radioimmunoassay (RIA). There has been some discussion recently about whether lack of correlation of different assays may have contributed to inconsistent results between studies examining the association of IGFBP-3 levels and breast cancer risk (*3,156,158*). Studies that have directly compared RIA methods to those using another common method of measuring IGFBP-3

(enzyme-linked immunoadsorbent assay or ELISA) have produced conflicting results regarding the correlation of these two methods (*156,158*). The RIA method has been presented as the gold standard for IGFBP-3 measurement (*156*), although it has been stated that at least with respect to cancer risk this has not been thoroughly tested (*3*). Currently though, RIA measurement is at least among the most appropriate methods of measuring IGFBP-3.

In many studies, a single measurement of IGF-I and IGFBP-3 is implicitly assumed to represent circulating levels for all or at least long periods of a persons life. Few studies have examined how a single measurement reflects long term circulating levels, although strong correlations have been found over both one and three year periods for both IGF-I (one year: 0.81, three year: 0.83) and IGFBP-3 (one year: 0.60, three year: 0.76) (*160,210*).

3.2.4. Height, weight, and BMI

The cross-sectional study described in chapter 4, used trained staff to measure height and weight. Height and weight were however, self-reported in the family based studies described in chapter 5. The correlation between self-reported height and weight and actual measures of these same variables is high ($r > 0.9$) (*211*). However, many studies report that women generally overestimate height and underestimate weight (*212-215*). This results in an underestimate of BMI (*214*). Several studies that examined the relationship between weight and reporting error in women, found that greater weight was associated with greater error in reporting both height and weight. Therefore measurement error of height, weight and BMI will often be greatest in heavier women (*214,215*). Still, given the strong overall correlation between self-reported height and weight and actual measures, the use of self reported data

should provide reasonable accuracy. Further discussion of this issue in context of the results is provided in chapter 7.

3.2.5. Genotyping

Genotyping was performed by The Centre for Applied Genomics (TCAG) (*216*). Alleles were separated according to their size on a performance Optimized Polymer 6 (POP6) gel capillary electrophoresis system (ABI 3700 and 3100 systems). Studies using similar methods have produced very high concordance rates (95% to just under 100%) for microsatellite markers (*217,218*). Therefore, for primary analyses examining the association of variant alleles of these polymorphisms with outcome, genotyping error should have little impact on study results. Genotyping errors will have a greater influence on the results of haplotype analyses, since haplotypes are constructed from two or more polymorphisms, increasing the chance of error. The potential influence of these errors on the results of the studies presented here is discussed in the last chapter.

Quality of genotyping was assessed in this study by having the TCAG genotype duplicate samples to which they were blinded. Comparison of the assignment of genotype by the TCAG to those assigned by the author and a third individual experienced in assigning genotypes using TCAG methodology, was also used to assess quality of genotyping. In addition to these approaches, checking for Mendelian errors (incongruent genotype assignments among family members), assessing Hardy-Weinberg equilibrium, and the comparison of allele frequencies for a specific polymorphism between studies can all help assess the extent of genotyping error in a sample. Therefore, along with concordance and comparison of genotype assignments, these approaches were also used where appropriate, to determine the quality of genotyping (see chapters 4-6).

3.3. Analysis

In the cross-sectional study, multiple linear regression, using the method of least square means was used for primary analyses. In the family based study, family based association tests were used for primary analyses. This is a relatively new statistical approach and is described in more detail here.

3.3.1. Family based association tests

Family based association tests incorporated into the programs FBAT (Family Based Association Testing, version 1.5) (*219*) and PBAT (Pedigree Based Association Testing, version 2.5) (*207*) were used for main analyses (e.g. association with breast cancer and BMI). These test for association under a null hypothesis of no linkage or linkage disequilibrium between the marker(s) and trait alleles. These tests are similar to the transmission disequilibrium test (TDT) (*220*). The TDT uses a family unit of a diseased case (proband) and two parents. Given Mendelian inheritance patterns all alleles are equally likely to be transmitted to the case or proband and deviation from this expected distribution is evidence for association of the marker with disease. The family based association tests employed in FBAT and PBAT also use parental genotypes to determine the expected distribution, but have the advantage that they can utilize information from siblings of the proband to reconstruct missing parental genotype data. Since the expected distribution is based on the parental genotype data, the resulting statistical test is not susceptible to confounding due to population stratification. Furthermore, no assumptions need to be made regarding assortative mating (deviation from Hardy-Weinberg equilibrium) and the ascertainment condition (*219*). Covariate information can be included in models to improve the power of the statistical test (*221*).

Both FBAT and PBAT allow for the coding of traits that include an offset. For example, the trait (Y_{ij}) can be adjusted so that $T_{ij} = Y_{ij} - \mu$, where μ is an offset which minimizes variance (FBAT) or maximizes power (PBAT), and T_{ij} is the adjusted trait value. For a dichotomous trait, the offset in FBAT approximates the sample prevalence where offspring are weighted by the number of heterozygous parents ($\mu = n_{Aff}/(n_{Aff} + n_{Unaff})$ where n_{Aff} and n_{Unaff} represents the number of affected and unaffected individuals) (*222*). For a continuous trait the value of μ that minimizes the variance is given by the sample average of $Y_{ij,}$ weighting each offspring by the number of heterozygous parents (*222*). The purpose of using PBAT was to compare associations using the more recently implemented offset for optimal power to those in FBAT and also to explore the effect of covariates (which is greatly simplified using the PBAT program). However, neither choice of software nor the inclusion of covariates had a substantial influence on the results, and results from FBAT were generally reported, as unlike PBAT the output provides a clear indication of the magnitude and direction of effect. The FBAT program (FBAT 1.5) was also used in analyzing association of haplotypes with outcome, as PBAT cannot perform haplotype analysis with polymorphisms that are multi-allelic.

Each of the polymorphisms genotyped had rare alleles. In order to ensure reliable results, alleles were included in analyses only if at least ten of the families were informative for a given allele. An informative family is one where parental genotypes (or the reconstructed parental genotypes) include at least one parent that is heterozygous for a given allele.

In general, additive models were used in analyses. Correct false positive rates are not dependent on proper specification of the genetic model. Furthermore, results from a

simulation study indicated only a small decrease in power when the true genetic model was recessive and an additive model was chosen, while coding for an additive model was preferable to coding for a dominant model, even when the true genetic model was dominant (*223*).

3.3.2. Multiple testing

Each of the individual studies described in chapter 4-6, included three polymorphisms with multi-allelic markers and at times multiple outcomes. This resulted in several statistical tests being performed. Approaches for dealing with multiple testing in the cross-sectional and family based studies are discussed below.

3.3.2.1. Cross sectional study

In the cross-sectional study, primary analysis was based on an allele classification where for each polymorphism, each individual was classified as having two, one or no copies of the most common allele (homozygotes, heterozygotes and non-carriers). The choice of grouping alleles in this manner was appropriate for this study as sample size was relatively small and power to detect an association for most of the other alleles was relatively low. The number of copies of the most common allele was entered into models as a continuous variable. This coding implies an additive model. If the true genetic model was not additive an additive marker coding would be less likely to detect an association. However, further examination of least square mean values for each category (2,1 or 0 copies of the most common allele) did not indicate that model misspecification was an issue. No correction for multiple comparisons was employed in this study, since for each outcome only three

statistical tests were performed. However, consistency of results across outcomes is discussed in chapter 4, and in the final discussion (chapter 7).

3.3.2.2. Family based study

In the family based study there was greater power to detect an association with outcome for some of the less common alleles relative to the cross-sectional study. This resulted in the testing of many alleles and therefore a formal approach to adjusting for multiple comparisons, the Benjamini-Hochberg correction, was employed (*224*).

The Benjamini-Hochberg correction controls for the false discovery rate; the expected proportion of errors among the rejected null hypotheses. A major reason for selecting this method is that unlike other methods that were primarily designed for genome wide studies (*225,226*), the Benjamini-Hochberg correction does not require the assumption of a continuous distribution of P-values to provide an accurate estimate of the false discovery rate. This method also maintains its control properties under certain forms of dependency and its use is still appropriate when there is mild correlation structure among P-values (*225*).

The Benjamini-Hochberg correction is straightforward to apply. First a value is chosen to control the false discovery rates at q^* (for example 0.05). Then for m tests, the P-values are ranked in ascending order $P_{(1)}, \leq P_{(2)}, \leq \ldots \leq P_{(m)}$. Letting k be the largest i for which

$$P_{(i)} \leq \frac{i}{m} q^*$$

and denoting by $H_{(i)}$ the null hypotheses corresponding to $P_{(i)}$, all null hypotheses from $H_{(1)}$ to $H_{(k)}$ are rejected. In other words, starting with the largest P-value, each P-value is compared against the result for the above equation, and for the first one that satisfies this equation and for all smaller P-values, the null hypothesis is rejected.

3.3.2.3. Interpretation of results

A significant result after correction for the false discovery rate was only one of the criteria used to indicate the presence of an association. In the family based studies, two samples, one from the ABCFR and one from the OFBCR were available. Consistency across samples was examined to assess evidence for an association. In addition, association of height, weight, and BMI were investigated in both the cross-sectional and family based studies and results of these studies were compared. Finally, comparison of results obtained here were interpreted in the context of those found in the published literature, in order to assess the relevance of associations observed in this study.

Chapter 4. Association of variant alleles of an *IGF1* CA repeat polymorphism with mammographic density and anthropometric measures in premenopausal Caucasian women

ABSTRACT

Background and objectives

Results from several studies indicate that mammographic density, a strong risk factor for breast cancer, is increased by greater circulating IGF-I concentration in premenopausal women. Both mammographic density and circulating IGF-I concentration appear to be partly heritable traits. Here it was hypothesized that *IGF1* gene variants modify circulating IGF-I levels and therefore variation in breast density. This study examined the association of genetic variation at *IGF1* with circulating IGF-I concentrations and mammographic density in premenopausal women. The association of *IGF1* with body mass index (BMI), weight, and height was also explored.

Methods

A cross-sectional design (n=163) was used to investigate the association between IGF-I levels, mammographic density, anthropometric measures and three CA repeat polymorphisms at *IGF1*, including a previously investigated 5′ repeat.

Results

A greater number of copies of the 5′ 19 allele were associated with lower circulating levels of IGF-I (P=0.02), while a greater number of copies of the 3′ 185 allele were associated with greater percentage breast density (P=0.03) and a smaller amount of non-dense (fat) tissue (P=0.02). The number of 3′ 185 alleles was positively associated with

height (P=0.01) and inversely associated with BMI (P=0.01). Including BMI in regression models resulted in a loss of significance and substantial reduction in strength of effect of the 3′ 185 allele on percentage breast density.

Conclusions

Results from this study suggest an association between the number of copies of the 3′ 185 allele and mammographic density. This association appears to be mediated through an influence on body fat but not through circulating IGF-I levels.

INTRODUCTION

Differences among women in the radiological appearance of the breast reflect differences in tissue composition. Fat is radiologically lucent and appears dark on a mammogram while epithelial and stromal tissues are radiologically dense and appear light. Quantitative assessment of the extent of dense areas in the breast have reported estimates of about 3 to 6 for risk of breast cancer in women with the most extensive areas of density, compared to those with little or no density (*4*).

Mammographic density is influenced by a number of risk factors for breast cancer. Age, parity (*43,59-70*), body weight and body mass index (*43,49,59-64,67-70,76-81*), are associated with reduced density in the breast. Reduction in mammographic density has been observed following menopause (*4*), while increased density in the breast is associated with hormone replacement therapy (*71,72*).

Risk factors for breast cancer only explain about 20-30% of the variation in mammographic density (*4*). However, a large proportion of the variation appears to be explained by genetic factors. A recent classical twin study has estimated the heritability of

mammographic density in a population of Australian and North American twins to be 63% (*10*).

There is considerable interest in the influence of the growth hormone/IGF-I axis on mammographic density, as a number of reports suggest greater circulating levels of IGF-I increases the risk of breast cancer (*3*). Seven studies have investigated the association of serum or plasma insulin-like growth factor-I (IGF-I) levels or its main binding protein IGFBP-3 with mammographic density in premenopausal women (*5-8,99,162,164*). Mammographic density expressed as percentage dense area of the breast has been shown to be positively associated with IGF-I concentrations in blood samples in three studies (*5,7,8*). A positive association of borderline significance was also reported (Table 1) (*6*). Of these four studies, three reported either significant or borderline significant inverse associations with percentage breast density and IGFBP-3 levels and significant or borderline significant positive associations with the IGF-I/IGFBP-3 ratio (*5,6,8*). Most studies do not support an association between IGF-I or IGFBP-3 and mammographic density in postmenopausal women (*5,7,8 ,99,162,165*).

Circulating IGF-I concentration appears to be partially determined by genetic factors. Twin studies estimate the heritability of circulating IGF-I levels to be 38%-63% in middle aged and elderly men and women (*11,12*). Evidence indicating that mammographic density and IGF-I concentrations are partly heritable, and that circulating IGF-I and IGFBP-3 concentrations are related to mammographic density, suggests that genes in the growth hormone/IGF-I axis are suitable candidates to investigate for association with mammographic density. Here it is hypothesized that *IGF1* gene variants modify circulating

IGF-I levels, which influence the proliferative activity and quantity of stromal and epithelial tissue in the breast, and therefore variation in breast density.

This study investigated the association of genetic variants in the *IGF1* gene with circulating IGF-I levels and with mammographic density in premenopausal women. The relationship between genetic variants of this gene and other anthropometric variables was also examined. Three cytosine-adenosine (CA) repeat polymorphisms were chosen for this investigation: a promoter region repeat (5′ polymorphism) previously examined in relation to breast cancer risk (*172,179,189-193*) and repeats in intron 2 and at the 3′ end the gene.

Table 1. Association of IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio with percentage breast density in premenopausal women.

Study/sample size	IGF-I Coefficient† (P-values)	IGFBP-3 Coefficient† (P-values)	IGF-I/IGFBP-3 Coefficient† (P-values)	Variables matched or adjusted for
Byrne 2000 n=65 (*8*)	0.36 (0.007)	-0.24 (0.07)	0.39 (0.004)	Age, alcohol, assay batch, BMI, and where appropriate IGF-I, IGFBP-3
Boyd 2002 n=193 (*7*)	0.01 (0.03)	-0.02 (0.10)		Age, waist, IGFBP-3 for IGF-I, Age and waist for IGFBP-3
Maskarinec 2003 n=263 (*6*)	0.11 (0.06)	-0.15 (0.02)	0.13 (0.03)	Age, ethnicity, body mass index, year of lab analysis, family history of breast cancer, reproductive variables and IGF-I or IGFBP-3 where appropriate
Lai 2004 n=206 (*99*)	-0.01 (0.9)	-0.07 (0.4)	-0.02 (0.8)	IGFBP-3 for IGF-I, IGF-I for IGFBP-3, adjusted for BMI, parity and coffee for ratio
Diorio 2005 n=783 (*5*)	0.08 (0.02)	-0.12 (0.0005)	0.07 (0.06)	Age, BMI, IGF-I or IGFBP-3 where appropriate
Dos Santos Silva 2006 n=215 (*162*)	(0.36)	(0.27)	(0.24)	Age, time, since blood collection, age at first birth, BMI, waist circumference smoking habits, past oral contraceptive use, and IGFBP-3 if outcome variable was IGF-I, and IGF-I and IGF-II if outcome variable was IGFBP-3
Maskarinec 2007 n=525 (*164*)	(0.83)	(0.91)	(0.67)	Age, ethnic group/location, BMI, quadratic BMI term, digital vs. regular mammogram, parity, age at menarche, age at first live birth, IGF-I or IGFBP-3 where appropriate.

†Correlation coefficients shown for all except Boyd (regression coefficients) and Dos Santos Silva and Maskarinec 2007 (not provided).

METHODS

The following describes all methods relevant to this study, part of a larger investigation into genetic, hormonal and other factors that influence mammographic density. Full details of these methods are provided elsewhere (*7*).

Recruitment

Potential subjects, both pre- and postmenopausal women, were identified from mammographic units in Mount Sinai, Women's College, and St. Michael's Hospitals in Toronto between 1994 and 1997. Women were referred to these units for a variety of reasons, including suspicion of breast disease, the presence of risk factors such as family history of cancer, or for routine examination. Initially, radiologists in each hospital visually estimated the extent of mammographic density in mammograms of potential study subjects, and expressed the results as percentage of dense area on a five-point scale. The purpose of this initial categorization was to recruit approximately equal numbers of women without breast cancer in each of five categories of percent breast density. The distribution of breast density was therefore anticipated to be different than that in the general population. The number of subjects recruited in each of the five radiological categories were as follows: <10% n=101, 10 to <25% n=62, 25 to < 50% n=60, 50 to <75% n=60, and ≥ 75% n=99.

Subjects identified in the manner described above were sent a letter followed by a phone call, during which their eligibility was determined. All women who had taken any type of exogenous hormone preparation during the previous six months, had breast augmentation or reduction, a previous history of breast cancer, or were being investigated for breast cancer, were excluded from the study. Women who were menstruating regularly, not pregnant or

breast feeding, and had not had a hysterectomy or oophorectomy, were eligible for recruitment as premenopausal subjects, the focus of this investigation.

Initially, 1874 letters were sent in an effort to recruit both pre- and postmenopausal women. Forty percent (n=740) of potential subjects could not be contacted, 28% (n=526) were not eligible for the study, 9% (n=174) stated they were not interested before eligibility could be determined, and 3% (n=52) were eligible but not interested in participating. A total of 382 women, representing 88% of all subjects contacted and found to be eligible (434) agreed to participate, 193 of which were premenopausal.

Obtaining consent for genetic study

After study subjects participated in an initial study on determinants of breast density, they were sent a letter that described the new genetic component to the study that requested their consent to analyze their DNA for genetic polymorphisms that might affect breast density. Of the 193 premenopausal women, 11 could not be contacted for consent to use their DNA. DNA was not available for one of the remaining subjects, and ambiguous labelling of DNA samples resulted in 4 additional samples being removed. Therefore, there were 177 premenopausal subjects available for analysis. Only data for Caucasians (see below for definition) were used in the analysis, leaving a sample size of 163.

Measurements

Blood analytes

Blood samples were collected in the morning after a 12 hour overnight fast during the luteal phase of the menstrual cycle (between days 20-24). Blood was refrigerated

immediately after collection, spun and serum separated within 1-2 hours of collection and stored at –70^{o}C.

IGF-I, IGFBP-3 and growth hormone (GH) were measured by Esoterix, California, USA. IGF-I and IGFBP-3 were measured using a competitive binding radioimmunoassay, and GH with a two site immunometric assay. The percent coefficient of variation was less than 7% within assays, and less than 10% between assays for these analytes.

Anthropometric measures

Each subject was weighed on a balance scale and measured for height by a research assistant trained and certified by the Department of Athletics and Recreation, University of Toronto.

Epidemiologic data

A questionnaire was used to obtain information about other epidemiologic risk factors, such as age, menstrual and reproductive history. Subjects were also asked for their country of birth and countries of birth for parents and grandparents and "What is your ethnic and cultural background", and given instructions to mark all appropriate categories. Women who either indicated they were white (e.g., British, French, Latin/South American or of European background) or Jewish were classified as Caucasians.

Mammographic density

A randomly selected craniocaudal mammographic view of one breast from each subject was used for breast density measurements. Images were analysed using a computer assisted method (*209*). Randomly ordered mammograms were digitized using a Lumisys model 85 and were presented to the observer (N.F. Boyd) for analysis as an array of 675 X

925 pixels (0.0676 mm^2/pixel). The observer selected two threshold grey backgrounds: one to separate the image of the breast from the background, and the second to identify regions representative of radiographically dense tissue. A software program automatically summed the pixels within the two areas. The percentage of radiographic density was calculated by dividing the projected area of dense tissue by the total projected area of the breast and multiplying by 100.

A subset (10%) of duplicate images were included as a check on reliability in the original study (*7*). Reliability of the measurements was high, with a test-retest correlation of at least 0.9.

Genotyping

DNA was purified from buffy coats of blood samples using a modified chaotropic method (*227*). Three polymorphisms were selected along the *IGF1* gene for analysis. One was a previously identified promoter region 5′ CA repeat polymorphism (*175,228*). In addition, two other CA repeats, located in intron 2 and at the 3′ end of the gene were identified using Tandem Repeat Finder software version 2.02 (*229*). The oligonucleotide primers used for PCR were as follows: 5′-GCTAGCCAGCTGGTGTTATT-3′, 5′-ACCACTCTGGGAGAAGGGTA-3′ for the 5′ polymorphism, 5′ CATACTTCTTAGCTCCTCAGG-3′, 5′-CCCTCACAGAAAGCAGAA-3′ for the intron 2 polymorphism, and 5′-CTTTTTAAGATGAGGCAGTTCC-3′, 5′GATTTCTTTTCAGTATTCCATTGG for the 3′ polymorphism. Position of amplified regions of DNA corresponded to positions 26,357,243-26,357,436, 26,332,059-26,332,274 and 26,275,038-26,275,220 on contig NT_019546.15 of build 35.1 of the NCBI's genome

annotation for the 5′, intron 2 and 3′ CA repeats respectively. Primers were obtained from Invitrogen Life Technologies. The forward primer was labeled with a 5′ fluorescent dye (HEX) and the reverse primer with an additional sequence (GTTTCTT) at the 5′ site. PCR was performed with 5 nanograms of DNA, 10X PCR buffer (Invitrogen Life Technologies), 0.8mM DNTP, 0.18mM of each primer, 4mM $MgCl_2$ and 5 U (0.3µl and 0.2µl) Platinum Taq for each 50µl reaction. All reagents were supplied by Invitrogen Life Technologies. PCR amplification conditions were 94ºC for 3 minutes, and 35 cycles of 94ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 1 minute. Genotyping was performed by The Centre for Applied Genomics (TCAG) (*216*). Alleles were separated according to their size on a performance Optimized Polymer 6 (POP6) gel capillary electrophoresis system (ABI 3700 and 3100 systems). Quality control samples representing 15% of the total were genotyped with TCAG staff being blind to the status of these samples. Concordance between original and replicate samples was 100%. TCAG allele calls were verified by two readers (G. Fehringer and H. Jarjanazi). All discrepancies were resolved between readers before comparing allele calls to those made by TCAG. On one occasion, the readers altered a genotype assigned by TCAG.

Ethics approval for this study was received from the University of Toronto and Mt. Sinai Hospital Research Ethics Boards.

Analysis

Hardy-Weinberg equilibrium for each polymorphism was examined using a permutation version of the exact test (*230*) (implemented in Power Marker 3.25 (*231*)).

Linear regression analysis (GLM procedure, SAS (version 8.02) statistical software (*232*)) was used to examine the relationship of IGF-I, IGFBP-3, the IGF-I IGFBP-3 ratio,

growth hormone (GH), mammographic features and anthropometric variables with genotype. Age was included as a covariate for all models unless otherwise indicated. Data were inspected for departures from normality and where necessary, variables were transformed to approximate a normal distribution. Square root transforms were applied to percentage breast density and area of dense breast tissue. Natural log transforms were applied to area of non-dense breast tissue, IGFBP-3, and GH concentration. Inverse transforms were used for weight and BMI. In addition to age, the influence of other covariates were examined in models where mammographic features (covariates examined were BMI, IGF-I and IGFBP-3 concentration, the IGF-I/IGFBP-3 ratio and parity), anthropometric measures (covariates examined were IGF-I and IGFBP-3 concentration, and the IGF-I/IGFBP-3 ratio) or measures of IGF-I and IGFBP-3 (IGF-I or IGFBP-3 concentration where appropriate) were included as dependent variables. Results of analyses with these covariates are only reported when they had an appreciable influence on main effects. Exploratory analyses included examining interactions between genotype and BMI, IGF-I, IGFBP-3, and the IGF-I/IGFBP-3 ratio, with either percentage density, amount of dense tissue, amount of non-dense tissue and IGF-I as the outcome variable (for the latter an interaction between IGF-I and *IGF1* genotype was not investigated). These same relationships were also examined using stratified analyses. For the presentation of primary results in tables and for some exploratory analyses age adjusted least square means were calculated and where appropriate back transformed (e.g., BMI was back transformed from the inverse to indicate mean BMI (kg/m^2), which also permitted presentation of the direction of effect). As well, when an inverse transform was employed (e.g. BMI), direction of effect in the text is always reported as the direction of effect after back transforming (i.e. the true direction of effect).

Values for GH were missing for 3 women and undetectable for 47 others. Undetectable values were assumed to be caused by the pulsatile nature of GH release, resulting in considerable variability in basal hormone levels. Analyses relating *IGF1* genotype to GH were performed by assigning nondeterminate samples a value of 0.2 ng/litre, the lower limit of sensitivity for the assay used, and with all nondeterminate samples excluded from the analysis. The results did not differ appreciably and therefore are presented here with the assigned values for all subjects included (with the exception of those where the GH sample was missing).

Different genotype coding schemes were used to examine the association between genotype and outcome. Ideally, each individual in the population should be assigned their observed genotype in order to use all of the genotype data available. This however, results in an appreciable loss in power because of the large number of genotype categories, some of which are rare, that must be incorporated into the model. In order to ensure adequate power in the primary analyses, genotype for each individual at each polymorphism was classified according to the presence of two, one or no copies of the most common allele at each loci. The number of copies of the most common allele was then treated as a continuous variable in regression analyses. This coding implies an additive model. In regression analyses exploring interaction between *IGF1* genotype and either BMI, concentrations of IGF-I, IGFBP-3 or the IGF-I/IGFBP-3 ratio, genotype coding reflected presence or absence of the most common allele (consistent with a dominant effect). Examination of least square mean values of outcome against genotype category did not suggest that model misspecification was an issue.

When results from the primary analyses indicated an association with the most common allele, additional exploratory analyses investigated the association with other

alleles. Individual regression analysis examined the number of alleles verses mammographic features and anthropometric variables unless the number of homozygotes fell below 15, in which case a binary allele classification was used to ensure suitable numbers. Rare alleles (5 alleles or less in the sample) were not tested.

Linkage disequilibrium was assessed between loci (program PowerMarker vers. 3.25) (*231*) and between specific alleles (haplotypes) of each loci. Haplotypes were constructed when tests for linkage disequilibrium between loci was statistically significant. The association of haplotypes of pairs of polymorphisms with outcome was examined using haplotype trend regression, which constructs haplotypes using the EM algorithm (program PowerMarker vers. 3.25) (*233*).

RESULTS

Characteristics of subjects and completeness of data

Table 2 shows characteristics of study subjects. There were 163 premenopausal women in the sample, 153 (94%) Caucasian women and 10 (6%) Jewish women. Mean percent density was 29%. The number of subjects in each of the five categories of density were as follows: <10% n=52, 10 to <25% n=27, 25 to < 50% n=49, 50 to <75% n=32 and ≥ 75% n=3. Measurement of IGF-I, IGFBP-3 was missing in one subject and measures of GH were missing in three subjects.

Table 2. Selected characteristics for study subjects.

	Number	Mean (SD)
Risk factors		
Age, yr	163	44.8 (4.6)
BMI, kg/m^2	163	25.3 (5.9)
Height, cm	163	163.6 (6.1)
Weight, kg	163	67.6 (16.1)
Age at first birth, yr	113	28.2 (5.6)
Age at menarche, yr	163	12.7 (1.4)
Number of live births	163	1.5 (1.2)
1st degree breast cancer relative (% yes)	160	25% (NA)
Mammographic density, %[a]	163	28.8 (22.9)
Growth factors		
IGF-I, ng/ml	162	155.2 (36.1)
IGFBP3, mg/litre	162	2.7 (0.5)
GH, μg/litre	160	1.8 (2.4)

[a] Percentage of breast area occupied by dense tissue.

Genotype frequencies

Table 3 shows the observed allele and genotype frequencies for each polymorphism. The intron 2 polymorphism was the most highly polymorphic of the three and the most common allele had an allele frequency of 34%. Allele frequencies for the most common allele of the 5′ and 3′ polymorphism were 63% and 43% respectively. Significant deviation from Hardy-Weinberg equilibrium was found for the 5′ polymorphism, but not for the intron 2 or 3′ polymorphism (Table 3). The one borderline significant result among the three comparisons made does not provide strong evidence for extensive admixture in the population. It is also unlikely that extensive genotyping error explains this departure from

Hardy-Weinberg equilibrium, since reliability tests indicate that genotyping error should be minimal.

Relationships among measures

Table 4 shows correlations between outcome measures. Mammographic features were strongly correlated with weight and BMI. Significant correlations between mammographic features and IGF-I, IGFBP-3 levels and the IGF-I/IGFBP-3 ratio were observed for all but dense tissue with IGF-I concentration and non-dense tissue with IGFBP-3 concentration. GH was significantly correlated with the amount of non-dense tissue, but not percentage density or amount of dense tissue.

IGF-I was significantly associated with percentage density in regression analyses after adjustment for age and IGFBP-3 levels (F=16.70, P=0.0001). Unlike the original investigation from which this sample was derived (see Table 1) (*7*), there was no significant association with percentage density after adjustment for waist measurement (F=2.38, P= 0.13), although the direction of effect (positive) was the same. This may have been due to small sample size, as not all women from the original study were included here (i.e. some women did not give consent for the genetic component of the study and others were not Caucasian and therefore excluded here).

***IGF1* genotype and IGF-I and IGFBP-3 levels**

Table 5 shows least square mean values and 95% confidence limits for levels of IGF-I, IGFBP-3 the IGF-I/IGFBP-3 ratio and GH according to genotype category (homozygotes, heterozygotes and non-carriers of the most common allele for each polymorphism), adjusted for age. The number of 19 alleles of the 5′ polymorphism was associated with decreasing

levels of circulating IGF-I (F=5.51, P=0.02) and the decrease appeared to be consistent with an additive model (166.9, 156.9 and 149.0 μg/litre for 0, 1 and 2 copies of the 19 allele). Number of 19 alleles explained 5% of the variance in IGF-I levels. IGFBP-3 levels increased with greater number of 5′ 19 alleles but this increase was not statistically significant. Results did not indicate that the number of 5′ 19 alleles was associated with the IGF-I/IGFB-3 ratio or with GH levels. The number of 216 alleles of the intron 2 polymorphism and the number of 185 alleles of the 3′ polymorphism were not associated with increased circulating GH, IGF-I, or IGFBP-3 concentrations or the IGFBP-3 ratio (Table 5).

Further analyses explored the association of other alleles (17, 18, 20, 21) of the 5′ polymorphism with IGF-I concentrations. No significant associations were observed.

Table 3. Counts and percent frequency (in brackets) for alleles[†] and genotypes of three *IGF1* polymorphisms.

5′ polymorphism		Intron 2		3′ polymorphism	
Allele	Counts	Allele	Counts	Allele	Counts
11	1 (0.3)	204	7 (2.2)	181	10 (3.1)
17	6 (1.8)	206	2 (0.6)	183	3 (0.9)
18	19 (5.8)	208	2 (0.6)	185	141 (43.3)
19	205 (62.9)	212	6 (1.8)	187	93 (28.5)
20	68 (20.9)	214	21 (6.4)	189	65 (19.9)
21	24 (7.4)	216	110 (33.7)	191	9 (2.8)
22	3 (0.9)	218	61 (18.7)	193	5 (1.5)
		220	55 (16.9)		
		222	39 (12.0)		
		224	20 (6.1)		
		226	2 (0.6)		
		230	1 (0.3)		
Genotype frequencies		Genotype frequencies		Genotype frequencies	
19/19	67 (41.1)	216/216	24 (14.7)	185/185	30 (18.4)
19/other	71 (43.6)	216/other	62 (38.0)	185/other	81 (49.7)
other/other	25 (15.3)	other/other	77 (47.2)	other/other	52 (31.9)
HW-test[‡], P=0.05		HW-test[‡], P=0.60		HW-test[‡], P=0.99	

[†]Intron 2 and 3′ polymorphisms allele size reflects mobility units of PCR product in gel electrophoresis. 5′ allele size reflects number of CA repeats.
[‡]Hardy-Weinberg includes all alleles.

Table 4. Pearson correlation among variables in this study (second row gives P-values).

	Percentage density	Dense tissue	Non-dense tissue	Height	Weight	BMI	IGF-I	IGFBP-3	IGF-I/IGFBP-3 ratio	GH
Percentage density	-	0.90 <.0001	-0.85 <.0001	0.02 0.85	-0.67 <.0001	-0.69 <.0001	0.20 0.009	-0.16 0.04	0.33 <.0001	0.11 0.17
Dense tissue	-	-	-0.57 <.0001	0.00 0.98	-0.53 <.0001	-0.55 <.0001	0.13 0.10	-0.21 0.01	0.28 0.0003	0.03 0.68
Non-dense tissue	-	-	-	-0.03 0.66	0.70 <.0001	0.73 <.0001	-0.26 0.001	0.08 0.33	-0.33 <.0001	-0.17 0.03
Height	-	-	-	-	0.20 0.01	-0.11 0.17	0.07 0.41	0.00 1.00	0.06 0.42	0.01 0.94
Weight	-	-	-	-	-	0.95 <.0001	-0.22 0.004	0.16 0.04	-0.34 <.0001	-0.15 0.06
BMI	-	-	-	-	-	-	-0.25 0.002	0.17 0.03	-0.37 <.0001	-0.16 0.05
IGF-I	-	-	-	-	-	-	-	0.45 <.0001	0.69 <.0001	-0.09 0.28
IGFBP-3	-	-	-	-	-	-	-	-	-0.31 <.0001	-0.01 0.93
IGF-I/IGFBP-3 ratio	-	-	-	-	-	-	-	-	-	-0.11 0.18

All sample sizes are equal to 163 except for comparisons that include IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio (n=162) and GH (n=160) (with the exception of specific comparisons that include either IGF-I, IGFBP-3 or the IGF-I/IGFBP-3 ratio verses GH (n=159)). Data transformed as indicated in methods section with the exception of weight and BMI, where the inverse transform was not used in order to show direction of effect.

Table 5. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and IGF-I, IGFBP-3 levels, IGF-I/IGFBP-3 ratio and GH levels.[†]

	Least square mean (95% CI) for number of 5′ 19, intron 2 216, or 3′ 185 alleles				
	Two copies	One copy	Non-carrier	F	P
IGF-I (ng/ml)					
5′ 19	149.0 (141.0-157.0), *n=67*	156.9 (149.1-164.7), *n=70*	166.9 (153.8-179.9), *n=25*	5.51	0.02
Intron 2 216	166.8 (153.4-180.2), *n=24*	152.6 (144.3-161.0), *n=62*	153.6 (146.1-161.1), *n=76*	1.69	0.20
3′ 185	152.6 (140.5-164.7), *n=30*	158.1 (150.7-165.5), *n=80*	152.2 (143.0-161.4), *n=52*	0.07	0.80
IGFBP-3 (mg/litre)					
5′ 19	2.6 (2.5-2.7), *n=67*	2.7 (2.6-2.8), *n=70*	2.8 (2.6-2.9), *n=25*	3.05	0.08
Intron 2 216	2.8 (2.6-3.0), *n=24*	2.6 (2.5-2.7), *n=62*	2.7 (2.6-2.8), *n=76*	0.01	0.93
3′ 185	2.7 (2.5-2.8), *n=30*	2.7 (2.6-2.8), *n=80*	2.6 (2.5-2.7), *n=52*	0.76	0.38
IGF-I/IGFBP-3 ratio					
5′19	57.6 (54.8-60.5), *n=67*	58.1 (55.3-60.9), *n=70*	60.6 (55.9-65.2), *n=25*	0.88	0.35
Intron 2 216	60.2 (55.4-64.9), *n=24*	58.7 (55.7-61.7), *n=62*	57.3 (54.6-60.0), *n=76*	1.20	0.27
3′ 185	56.7 (52.5-61.0), *n=30*	58.7 (56.0-61.3), *n=80*	58.6 (55.3-61.8), *n=52*	0.34	0.56
GH (μg/litre)					
5′ 19	0.8 (0.6-1.1), *n=65*	1.0 (0.7-1.3), *n=70*	0.7 (0.4-1.1), *n=25*	0.00	0.95
Intron 2 216	1.0 (0.6-1.6), *n=23*	1.0 (0.6-1.6), *n=62*	1.0 (0.6-1.6), *n=75*	0.58	0.45
3′ 185	0.9 (0.6-1.4), *n=29*	0.9 (0.7-1.1), *n=80*	0.8 (0.6-1.1), *n=51*	0.13	0.71

[†]Test statistics based on the number of copies of the most common allele (5′19, intron 2 216, 3′185 allele) entered into models as a continuous variable. Mean values are back transformed where appropriate. All results are adjusted for age. Sample sizes for IGF-I, IGFBP-3 and IGF-I/IGFBP-3 ratio equal 162. Sample size for GH is 160.

IGF1 genotype and mammographic features

Regression analyses indicated that the number of 19 alleles at the 5′polymorphism or the number of 216 alleles of the intron 2 polymorphism were not associated with any of the three measures of mammographic features (Table 6). The number of 185 alleles at the 3′ polymorphism was significantly associated with increased percentage breast density (F=4.86, P=0.03) and amount of non-dense tissue (F=5.61, P=0.02) at the P≤0.05 level. Consistent with an additive model, least square mean estimates were 27.6%, 24.7% and 16.8% for percentage density and 66.7 cm^2, 78.2 cm^2 and 96.7 cm^2 for amount of non-dense tissue in women with 2, 1 and 0 copies of the 185 allele. The number of 185 alleles explained 3% of the variance in both percentage density and amount of non-dense tissue (prior to age adjustment). The amount of dense tissue increased with the number of 185 alleles, but the result was not significant (F=2.25, P=0.14) (Table 6).

Further analyses explored the association of other 3′polymorphism alleles and mammographic features. No significant associations between either the 181, 187, 189, or 191 alleles and mammographic features were observed.

IGF1 genotype and anthropometric measures

The number of 5′19 alleles and the number of intron 2 216 alleles were not associated with either BMI, weight or height (Table 7). Significant associations were, however, observed between the number of 3′185 alleles and height and BMI. Greater height was significantly associated with having a greater number of 185 alleles (F=8.06, P=0.01). Women with more copies of this allele had significantly lower mean BMI (F=7.33, P=0.01). The number of 185 alleles explained 5% and 4% of the variance in height and BMI respectively (prior to age adjustment).

Further analyses explored the association of other 3′polymorphism alleles and anthropometric measures. The number of 187 alleles was associated with height (F=6.8, P=0.01), but no other significant association between alleles at this polymorphism and anthropometric variables were observed.

Presence or absence of 3′ 185 allele, mammographic features and anthropometric measures

Analyses where genotype was coded for presence or absence of the 3′185 allele produced similar patterns of significance for mammographic features and anthropometric variables as analyses using number of 185 alleles. Significant associations were observed for percentage breast density (F=5.14, P=0.02), and non-dense tissue (F=4.54, P=0.03). A stronger association was observed for amount of dense tissue, but the result was still not significant at the P≤0.05 level (F=3.30, P=0.07). Significant results were also observed for height (F=10.4, P=0.002), and BMI (F=4.83, P=0.03), while results for weight (F= 0.92, P=0.34) were not significant.

Table 6. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and mammographic features.[†]

	Least square mean (95% CI) for number of 5′ 19, intron 2 216, or 3′ 185 alleles				
	Two copies	One copy	Non-carrier	F	P
Mammographic density, %					
5′ 19	21.1 (15.9-27.0), *n=67*	24.3 (18.9-30.5), *n=71*	21.3 (13.1-31.5), *n=25*	0.09	0.76
Intron 2 216	24.3 (15.4-35.3), *n=24*	24.7 (18.9-31.4), *n=62*	20.3 (15.5-25.6), *n=77*	1.02	0.31
3′ 185	27.6 (19.0-37.7), *n=30*	24.7 (19.6-30.4), *n=81*	16.8 (11.7-22.8), *n=52*	4.86	0.03
Amount of dense tissue (cm^2)					
5′ 19	23.8 (18.3-30.1), *n=67*	25.5 (19.9-31.8), *n=71*	25.5 (16.5-36.5), *n=25*	0.13	0.72
Intron 2 216	30.2 (20.2-42.3), *n=24*	25.6 (19.6-32.3), *n=62*	22.6 (17.6-28.3), *n=77*	1.66	0.20
3′ 185	27.0 (18.5-37.1), *n=30*	27.4 (22.0-33.4), *n=81*	19.9 (14.3-26.4), *n=52*	2.25	0.14
Amount of non-dense tissue (cm^2)					
5′ 19	86.5 (72.8-102.9), *n=67*	74.9 (63.3-88.5), *n=71*	86.7 (65.3-115.0), *n=25*	0.16	0.69
Intron 2 216	87.7 (65.8-116.8), *n=24*	72.6 (60.7-86.8), *n=62*	87.0 (74.1-102.1), *n=77*	0.27	0.61
3′ 185	66.7 (51.7-86.1), *n=30*	78.2 (67.0-91.4), *n=81*	96.7 (79.6-117.4), *n=52*	5.61	0.02

[†]Test statistics based on the number of copies of the most common allele (5′19, intron 2 216, 3′185 allele) entered into models as a continuous variable. Mean values are back transformed. All results are adjusted for age. Sample sizes for each outcome are equal to 163.

Table 7. *IGF1* 5′, intron 2 and 3′ polymorphism genotype and anthropometric measures.[†]

	Least square mean (95% CI) for number of 5′ 19, intron 2 216, or 3′185 alleles				
	Two copies	One copy	Non-carrier	F	P
Height (cm)					
5′ 19	163.1 (161.6-164.5), *n=67*	164.3 (162.9-165.7), *n=71*	163.1 (160.7-165.5), *n=25*	0.17	0.68
Intron 2 216	163.9 (162.5-165.2), *n=24*	164.0 (162.4-165.5), *n=62*	161.9 (159.5-164.4), *n=77*	1.13	0.29
3′ 185	164.8 (162.7-167.0), *n=30*	164.6 (163.3-165.9), *n=81*	161.4 (159.8-163.0), *n=52*	8.06	0.01
Weight (kg)					
5′ 19	64.7 (61.7-68.1), *n=67*	64.5 (61.6-67.8), *n=71*	64.0 (59.2-69.5), *n=25*	0.06	0.81
Intron 2 216	65.2 (62.4-68.4), *n=24*	63.6 (60.6-67.0), *n=62*	64.7 (59.8-70.4), *n=77*	0.17	0.68
3′ 185	60.9 (57.0-65.4), *n=30*	65.0 (62.3-68.1), *n=81*	66.0 (62.5-69.9), *n=52*	2.65	0.11
BMI (kg/m^2)					
5′ 19	24.4 (23.3-25.7), *n=67*	23.9 (22.9-25.0), *n=71*	24.2 (22.4-26.2), *n=25*	0.18	0.67
Intron 2 216	24.4 (23.3-25.5), *n=24*	23.7 (22.6-24.9), *n=62*	24.7 (22.9-26.9), *n=77*	0.00	0.96
3′ 185	22.6 (21.2-24.1), *n=30*	24.0 (23.1-25.1), *n=81*	25.4 (24.1-26.9), *n=52*	7.33	0.01

[†]Test statistics based on the number of copies of the most common allele (5′19, intron 2 216, 3′185 allele) entered into models as a continuous variable. Mean values are back transformed where appropriate. All results are adjusted for age. Sample sizes for each outcome are equal to 163.

Haplotype analyses

Linkage disequilibrium between specific allele pairs was low (Table 8). However, there was evidence for significant linkage disequilibrium for adjacent polymorphisms ($P < 0.0001$) and for polymorphisms at the 5′and 3′end ($P = 0.02$). Given these results and an interest in further exploring the associations observed above, global tests for association of haplotypes with IGF-I, IGFBP-3 levels, IGF-I/IGFBP-3 ratio, GH, mammographic features and anthropometric variables were performed with haplotypes constructed using these pairs of polymorphisms. There were no significant associations with any of the mammographic features (percent density: 5′and intron 2 haplotypes P=0.65, intron 2 and 3′ haplotypes P=0.30, 5′and 3′ haplotypes P=0.17; dense tissue: 5′and intron 2 haplotypes P=0.61, intron 2 and 3′ haplotypes P=0.34, 5′and 3′ haplotypes P=0.37; non-dense tissue: 5′ and intron 2 haplotypes P=0.40, intron 2 and 3′ haplotypes P=0.65, 5′ and 3′ haplotypes P=0.21). Among other measures, a significant result (at the $P \leq 0.05$ level) was observed for the overall association of haplotypes constructed from the 5′ and intron 2 polymorphism and circulating IGF-I levels (P=0.05), and for haplotypes constructed from the intron 2 and 3′ polymorphisms with height (P=0.01). None of the other comparisons produced statistically significant results.

Table 8. Linkage disequilibrium between specific alleles.

5′ – intron 2				Intron 2 – 3′				5′ – 3′			
haplotype	Frequency	D'	R^2	haplotype	Frequency	D'	R^2	haplotype	Frequency	D'	R^2
other/other	0.00	-0.29	<0.001	other/other	0.01	0.08	0.004	other/other	0.00	0.07	0.002
other/214	0.01	0.19	0.016	other/185	0.04	0.28	0.007	other/185	0.00	-1.00	0.024
other/216	0.01	0.07	<0.001	other/187	0.01	-0.29	0.002	other/187	0.01	0.02	<0.001
other/218	0.01	-0.03	<0.001	other/189	0.00	-0.72	0.008	other/189	0.02	0.44	0.025
other/220	0.00	-0.11	<0.001	214/other	0.00	-0.53	0.002	18/other	0.01	0.09	0.006
other/222	0.00	-1.00	0.004	214/185	0.01	-0.76	0.030	18/185	0.04	0.43	0.015
other/224	0.00	-1.00	0.002	214/187	0.05	0.62	0.066	18/187	0.01	-0.46	0.005
18/other	0.00	-0.59	0.001	214/189	0.01	-0.34	0.002	18/189	0.00	-0.99	0.015
18/214	0.01	0.14	0.018	216/other	0.01	-0.82	0.031	19/other	0.05	-0.01	<0.001
18/216	0.03	0.23	0.006	216/185	0.16	0.06	0.002	19/185	0.24	-0.16	0.021
18/218	0.02	0.13	0.004	216/187	0.16	0.31	0.076	19/187	0.18	0.01	<0.001
18/220	0.00	-1.00	0.013	216/189	0.02	-0.70	0.062	19/189	0.16	0.46	0.031
18/222	0.00	-1.00	0.008	218/other	0.02	0.12	0.006	20/other	0.01	-0.45	0.005
18/224	0.00	-1.00	0.004	218/185	0.13	0.42	0.052	20/185	0.10	0.08	0.002
19/other	0.04	-0.07	0.001	218/187	0.03	-0.51	0.024	20/187	0.09	0.18	0.022
19/214	0.04	-0.05	<0.001	218/189	0.01	-0.67	0.026	20/189	0.01	-0.70	0.032
19/216	0.24	0.20	0.012	220/other	0.01	-0.16	<0.001	21/other	0.01	0.02	<0.001
19/218	0.12	0.06	<0.001	220/185	0.10	0.25	0.017	21/185	0.06	0.57	0.034
19/220	0.04	-0.64	**0.139**	220/187	0.05	-0.06	<0.001	21/187	0.00	-0.99	0.031
19/222	0.10	0.46	0.017	220/189	0.01	-0.56	0.016	21/189	0.01	-0.28	0.002
19/224	0.06	1.00	0.038	222/other	0.02	0.11	0.008				
20/other	0.01	-0.25	0.001	222/185	0.01	-0.80	0.066				
20/214	0.01	-0.46	0.004	222/187	0.00	-1.00	0.054				
20/216	0.05	-0.26	0.009	222/189	0.09	0.70	**0.270**				
20/218	0.04	0.03	0.001	224/other	0.01	0.13	0.013				
20/220	0.08	0.36	**0.099**	224/185	0.00	-1.00	0.050				
20/222	0.01	-0.44	0.007	224/187	0.00	-1.00	0.026				
20/224	0.00	-1.00	0.017	224/189	0.05	0.74	**0.146**				
21/other	0.01	0.15	0.019								
21/214	0.00	-0.96	0.005								
21/216	0.01	-0.69	0.019								
21/218	0.00	-1.00	0.018								
21/220	0.04	0.49	**0.095**								
21/222	0.01	0.02	<0.001								
21/224	0.00	-1.00	0.005								

Bold indicates linkage disequilibrium values that are ≥ 0.10.

3′ polymorphism, BMI, and mammographic features

BMI, IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio were investigated as factors that might mediate the relationship between the 3′ polymorphism genotype and mammographic features. These variables were introduced into a regression model with genotype and age as independent variables, and either percentage density, amount of dense tissue or amount of non-dense tissue as the dependent variable. Inclusion of BMI in models with percentage density and non-dense area resulted in a loss of significance for the effect of genotype (percentage density: F=0.16, P=0.69, non-dense tissue: F=0.16, P=0.69). The relationship between amount of dense tissue and genotype was weakened with the P-value increasing to 0.91 from 0.14 after adjustment. Adjustment for IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio did not appreciably alter the strength of association between the number of copies of the 3′ 185 allele and either percentage dense tissue or the amount of non-dense tissue.

Interaction of 3′ polymorphism with IGF-I, IGFBP-3, and mammographic features

Interactions between presence or absence of the 3′ 185 allele and IGF-I, IGFBP-3 and the IGF-I/IGFBP-3 ratio were explored with percentage density, amount of dense tissue or amount of non-dense tissue as the dependent variable. A significant interaction was observed between genotype and IGFBP-3 concentration (entered in the model as a continuous variable) with either percentage density (F=4.43, P=0.04, β=4.89) or amount of dense tissue (F=4.24, P=0.04, β=4.76) as the outcome variable, but not for amount of non-dense tissue (F=2.11, P=0.15, β=-0.99). In stratified analyses strong and statistically significant inverse associations between IGFBP-3 and percentage density (F=10.36, P=0.002, β= -5.98) and amount of dense tissue (F=10.52, P=0.002, β= -6.30) and a more modest

positive association with amount of non-dense tissue (F=6.27, P=0.03, β=1.15) was observed when the 185 allele was absent. Non-significant and weak associations were observed when the 185 allele was present.

Figure 1 provides a visual representation of this relationship with differences in magnitude of mammographic features shown in relation to presence and absence of the 3′ 185 allele in women in the upper and lower halves of circulating IGFBP-3 levels. Significant inverse associations between presence of the 3′ 185 allele and percentage density (F=5.56, P=0.02) and amount of dense tissue (F=5.04, P=0.03) were observed for women with IGFBP-3 concentrations in the upper half of the distribution.

Entering IGF-I or the IGF-I/IGFBP-3 ratio into regression models as continuous variables did not produce significant interaction effects with genotype, but when dichotomized into IGF-I levels below or at and above the median, there was a significant interaction between genotype and IGF-I concentrations for percentage density (F=5.00, P=0.03, β=1.85) and amount of non-dense tissue (F=3.98, P=0.05, β=-0.46) and a borderline significant result for amount of dense tissue (F=3.60, P=0.06, β=1.59). The results of stratified analysis (depicted in Figure 2) revealed that presence of the 185 allele was associated with greater percentage density (F=9.94, P=0.003), greater amount of dense tissue (F=7.47, P=0.01) and lower amount of non-dense tissue (F=7.19, P=0.01), in women with circulating IGF-I concentrations in the upper half of the study group (greater than 152.5 ng/l), but not in women in the lower half of circulating IGF-I concentrations. Figure 3 shows this same relationship but with women stratified into upper and lower halves of the IGF-I/IGFBP-3 ratio. Tests for interaction did not indicate that being above or below the median in IGF-I/IGFBP-3 ratio modified the effect of genotype on mammographic features (percentage

density: F=0.79, P=0.38, β=0.74, dense tissue: F=1.16, P=0.28, β=0.90, non-dense tissue: F=0.12, P=0.73, β=-0.08). There was, however, some suggestion from stratified analyses that presence of the 3′ 185 allele was more strongly associated with greater percentage density and greater amount of dense tissue, and with smaller amount of non-dense tissue, in women who were above the median for the IGF-I/IGFBP-3 ratio.

Interaction of 3′ polymorphism with BMI, and IGF-I levels

A significant association between circulating levels of IGF-I and 3′ genotype was not observed. Since circulating IGF-I levels may be related to BMI and because an association between the 3′ 185 allele and BMI was observed, the interaction of the 3′ 185 allele with BMI in regression models with circulating IGF-I as the outcome variable was explored. Regression models relating IGF-I concentrations to genotype and BMI (entered into the model as a continuous variable) indicated a significant interaction between presence of the 3′ 185 allele and BMI (F=5.60, p=0.02). Stratified analysis indicated a significant positive association between presence of the 185 allele and circulating IGF-I concentrations in women with a BMI of less than 25 (F=4.17, p=0.04, β=15.8) and a non- significant but negative association between this allele and circulating IGF-I concentrations in women with BMI of 25 or greater (F=2.44, p=0.12, β=-12.5). There were, however, no significant interactions between BMI (coded as either a continuous or categorical variable (less than or greater or equal to 25 kg/m^2)) and presence of the 185 allele when either percentage density (continuous, F=0.76, P=0.38; categorical F=0.63, P=0.43), amount of dense tissue (continuous, F=1.37, P=0.24; categorical F=0.22, P=0.64), or amount of non-dense tissue (continuous, F=0.01, P=0.91; categorical F=1.52, P=0.22) was the dependent variable.

Furthermore, in women under 25 kg/m^2 the strength of association of the 3′ 185 allele with percentage density, amount of dense tissue and amount of non-dense tissue was not appreciably altered when IGF-I concentration was included as a covariate.

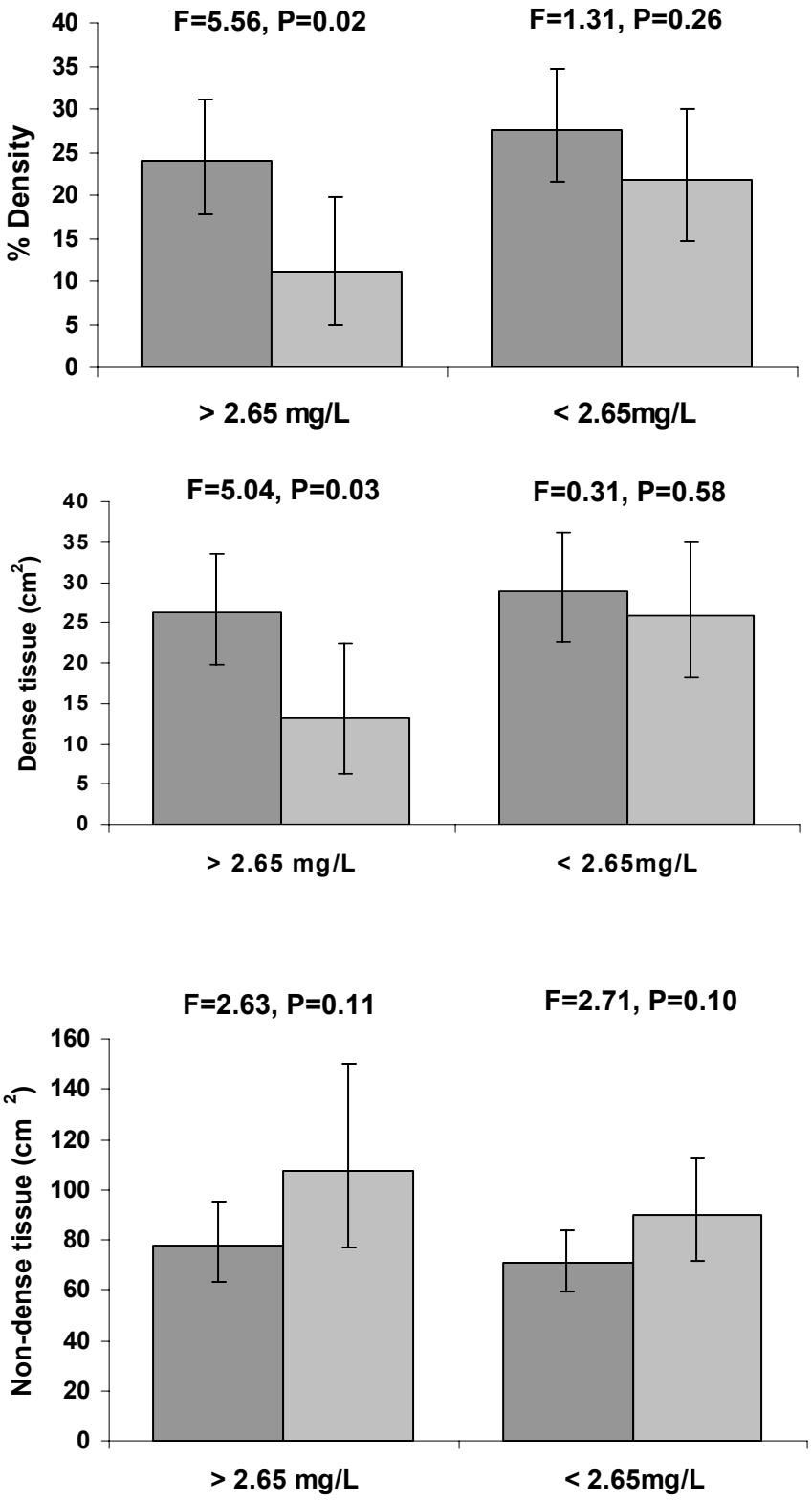

Figure 1. Presence (darker bar) and absence of the 3′ 185 allele and mammographic features, for women in the upper (n=81; 59 with 185 allele present) and lower (n=81; 51 with 185 allele present) halves of circulating IGFBP-3 concentration. Values are least square means from regression models adjusted for age. Error bars represent 95% confidence limits.

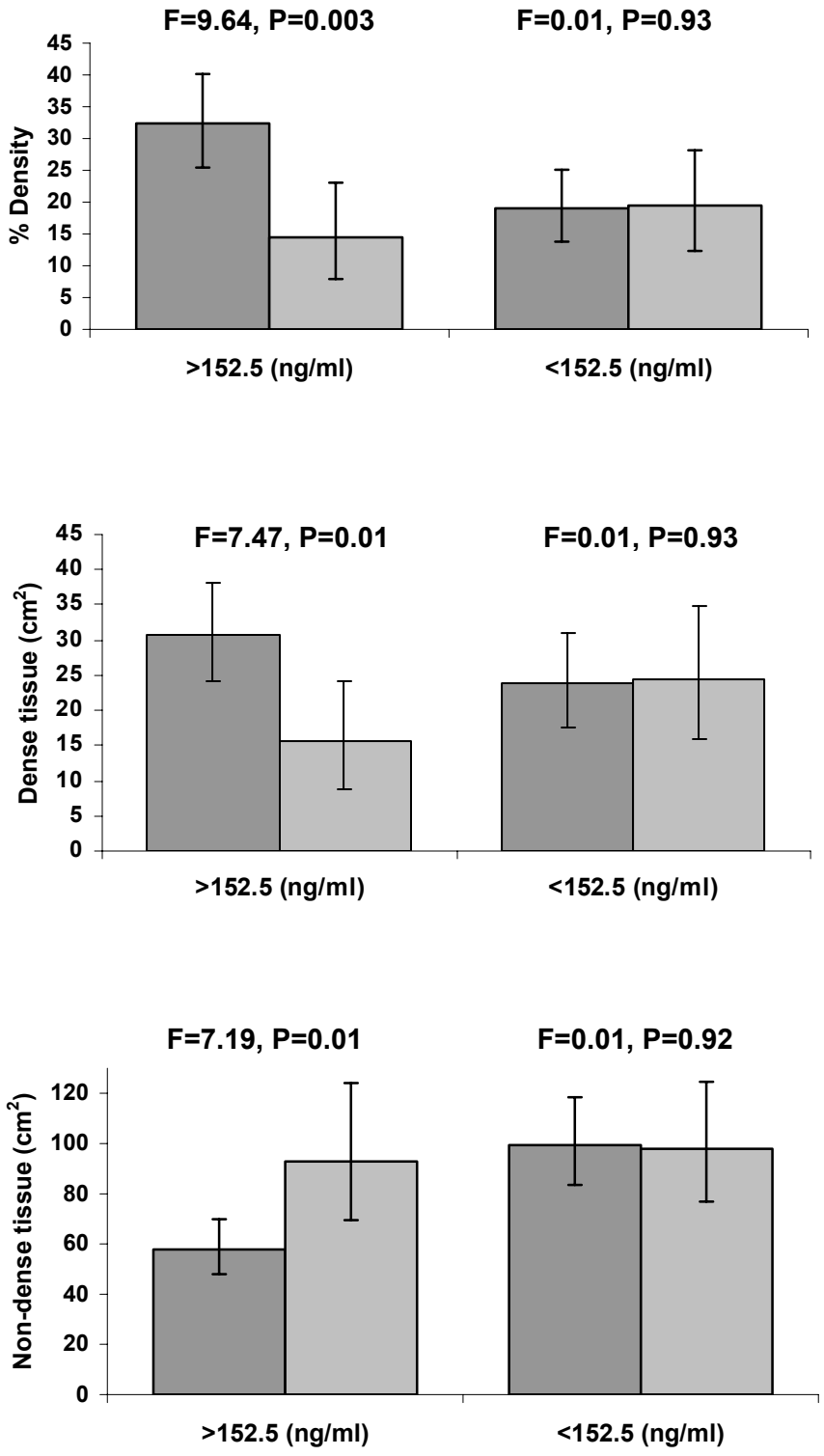

Figure 2. Presence (darker bar) and absence of the 3′ 185 allele and mammographic features, for women in the upper (n=81; 57 with 185 allele present) and lower halves (n=81; 53 with 185 allele present) of circulating IGF-I concentration. Values are least square means. Error bars represent 95% confidence limits.

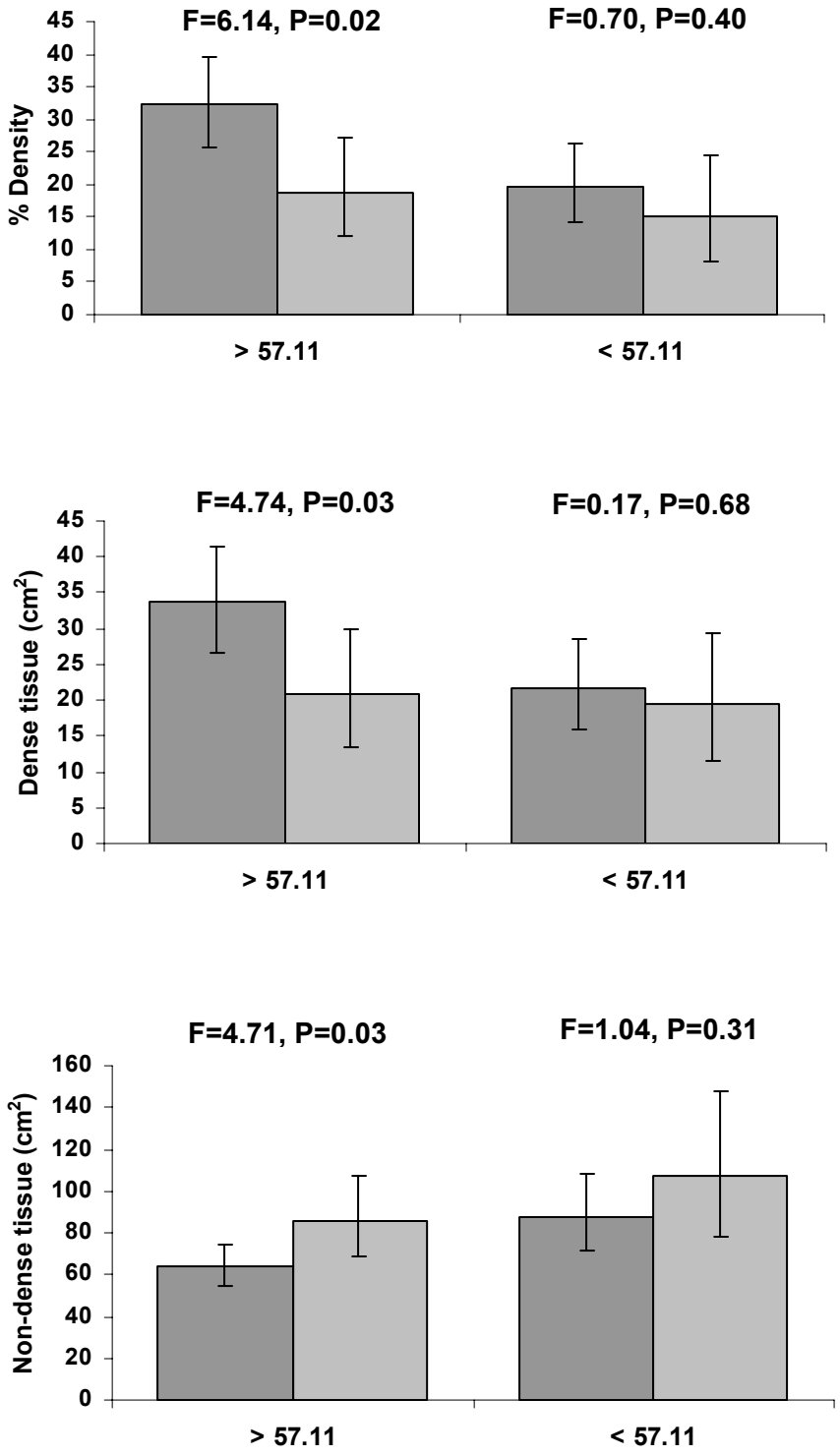

Figure 3. Presence (darker bar) and absence of the 3′ 185 allele and mammographic features for women in upper (n=81; 54 with 185 allele present) and lower (n=81; 56 with 185 allele present) halves for the IGF-I/IGFBP-3 ratio. Values are least square means. Error bars represent 95% confidence limits.

DISCUSSION

An inverse association of circulating IGF-I concentrations with the number of 19 alleles of a 5′ CA repeat polymorphism, previously examined in association with breast cancer risk with inconsistent results (*172,179,189-193*), was observed in this study. Additional analyses did not find significant associations with IGFBP-3 levels, the IGF-I/IGFBP-3 ratio or GH. Neither the intron 2 216 allele or the 3′ 185 allele was associated with levels of IGF-I, IGFBP-3, the IGF-I/IGFBP-3 ratio or GH.

The number of 5′ 19 alleles was not associated with mammographic features, which was inconsistent with the proposed model of variant *IGF1* alleles influencing mammographic density through an effect on circulating IGF-I levels. In fact, no specific allele of any of these polymorphisms was associated with both IGF-I levels and mammographic density (data not shown). This study did find an association between both percentage density and the area of non-dense (fat) tissue, and the number of 185 alleles of a polymorphism at the 3′ end of *IGF1*. The pattern of association was consistent with an additive model. The mean amount of dense area was greater in subjects with one or more copies of the 185 allele, but this result was not significant at the $P \leq 0.05$ level.

Haplotype analysis revealed an overall association between polymorphisms constructed from the 5′ polymorphism and intron 2 and IGF-I levels. The relationship was of borderline significance and did not provide additional information since a statistically significant association with the common 19 allele of the 5′ polymorphism was already observed. Inconsistent with the proposed model, however, haplotype analysis using any of the pairs of polymorphisms tested here did not provide evidence for an association of an *IGF1* variant with both mammographic density and IGF-I levels.

Further analyses focused on the 3′ repeat polymorphism. An inverse association was observed between the number of 185 alleles and BMI. Greater BMI is strongly associated with decreasing percentage breast density. In this study, including BMI in regression models with the 3′ genotype (number of 185 alleles) removed the significant association between genotype and percentage density and area of non-dense tissue. This result, in conjunction with the relationship between the 3′ 185 allele and amount of non-dense tissue, suggests that the association of genotype with percentage breast density is mediated through its relationship with body composition and in particular body fat.

There was evidence from exploratory analyses that suggested that the association of 3′ 185 allele with circulating IGF-I concentrations may be modified by BMI, as a positive association between the number of 185 alleles and circulating IGF-I concentration was observed only in leaner women (e.g. with a BMI of less than 25). There was however, no significant interaction between *IGF1* genotype and BMI in models where either percentage density, amount of dense tissue or amount of non-dense tissue were included as the dependent variable. Furthermore, including IGF-I concentration as a covariate in models examining the association of the number of 3′ 185 alleles with mammographic features in leaner women did not alter the results appreciably. Therefore, a more complicated model where number of 3′ 185 alleles influences circulating IGF-I levels and ultimately mammographic density only in women with lower BMI, is not supported. Instead of influencing IGF-I levels, allelic variants of *IGF1* might instead influence tissue specific expression of the IGF-I protein. This could occur through differential regulation of promoter activity or an influence on messenger RNA stability by alleles of a functional polymorphism that is also subject to regulation by tissue specific factors. Tissue specific expression of the

IGF-I protein could then in turn influence mammographic density and body fat stores through autocrine or paracrine mechanisms.

Although the 3′ 185 allele did not influence IGF-I levels, a modifying effect of IGF-I concentrations on the effect of the 3′ 185 allele on percentage density, the amount of dense tissue, and the amount of non-dense tissue was suggested by exploratory models, where IGF-I concentrations were dichotomized. The results indicated that presence of the 3′ 185 allele was more strongly associated with percentage density and amount of dense tissue, and absence of the allele was more strongly related with non-dense tissue, in women in the upper half of the distribution of IGF-I concentration. These results are consistent with the interpretation that the biological effect of variant *IGF1* alleles may be influenced by differences in IGF-I concentrations. The observation that percentage breast density and amount of dense tissue was reduced primarily in women who had high IGFBP-3 levels and did not carry the 3′ 185 allele suggests that IGFBP-3 concentration also plays a role in the effect of variant *IGF1* alleles. There are no common variants in *IGF1* coding regions that could account for these types of effects through an influence on IGF-I protein configuration. However, an allele that promotes greater tissue specific expression of IGF-I might act in concert with higher concentrations of circulating IGF-I to produce a stronger effect on local tissue. Similarly greater tissue specific expression of IGF-I might lead to a greater amount of free IGF-I in breast tissue, counteracting the putative protective effect of IGFBP-3 on percentage breast density.

A possible mechanism by which IGF-I may influence mammographic density and breast cancer risk is through proliferative and anti-apoptotic effects on epithelial cells. Dense tissue in the breast is comprised of epithelial cells and connective tissue. Although in this

study an association between *IGF1* genotype with breast density was observed, it appeared to be mediated through its influence on non-dense (fat) tissue in the breast. Still, a possible effect on dense breast tissue should not be ruled out, as the association of dense tissue with the presence of the 3′ 185 allele (a coding scheme which assumes a dominant effect of this allele) was close to significant. Furthermore, as discussed above, exploratory models suggest that the association of the 3′ 185 with the amount of dense tissue may be influenced by circulating IGF-I and IGFBP-3 levels.

There are several potential limitations to this study that should be considered. A potential source of bias in this study is confounding resulting from population stratification. Other genetic or environmental factors may be associated with both genotype and phenotype through population substructure (e.g. ethnicity) and could produce false associations between genotype, and the measured mammographic and anthropometric variables. Restricting the analysis to Caucasian women, however, should reduce the possibility of false positive associations that result from the relationship between population substructure, genotype, and phenotypic measures. Allele frequencies at the previously studied 5′ polymorphism in this sample were very similar to those in other Caucasian populations in the U.S., Israel and the northern parts of Europe (*175,177,191,234*). This suggests that within Caucasian populations, genetic variation at this locus is small, and if true for other *IGF1* loci, the chances of false positive or negative findings due to population stratification should be minimal. Still, the distribution of this and other alleles in Caucasian populations across Europe is not known, and if there is variation in allele frequencies, population stratification may be an issue since this sample should consist of a mixture of those populations (*201*).

This study was not population based; instead subjects were recruited from mammographic units of hospitals within the city of Toronto. Women were referred to these units for a variety of reasons, including suspicion of breast cancer, family history of breast cancer or for routine mammograms (although screening is not recommended in Ontario for most of the women in this sample, who are under 50 years of age). In addition, women were selected to produce a wide range of mammographic density. Therefore the women in this sample may not be representative of the population of Caucasian women in the surrounding area. A sample that is not representative could influence the magnitude of the relationship between genotype and phenotype. The presence of these relationships is unlikely to result from the recruitment approach, however, without the presence of significant confounding in the population due to population stratification. In addition, a previous study using all of the subjects from this investigation with some additional subjects (non-Caucasians and women who did not provide consent to have their DNA analysed) has produced associations between circulating IGF-I and mammographic density consistent with those observed in other studies (*5,8*). (The same association was observed in this sample, although perhaps due to smaller sample size, it was not significant after adjusting for IGFBP-3 levels). This suggests that the recruitment method did not have a major influence on study outcome.

In interpreting the results, it is important to consider that there was strong correlation between three of the outcomes found to be associated with the number of 3′ 185 alleles: percentage density, amount of non-dense tissue and BMI. It can be argued that the observed consistency in results may have been due to chance association with highly correlated variables. However, the number of copies of the 185 allele of the polymorphism at the 3′ end of the gene was also related to height. Height was not significantly correlated with

mammographic features or BMI and a relationship between *IGF1* genotype and height is consistent with the role of IGF-I in mediating the effects of growth hormone (*235*). This result provides some additional support for a possible relationship between the 3′ 185 allele and percentage density, amount of non-dense tissue and BMI.

Several other studies have examined the association of the 5′ 19 allele with circulating IGF-I levels with inconsistent results. Results from two studies conducted on Caucasian or mainly Caucasian populations give support to the results presented here, with an inverse relationship observed between the 19 allele and circulating IGF-I levels reported (*175,181*). Contradictory results in Caucasians have also been reported (*177,179*) and there are studies that report no association (*172,182*). There are three studies which reported on premenopausal women. In addition to the observed positive association of IGF-I levels with the 19 allele reported for both pre- and postmenopausal women in the Nurses Health Study (referred to above) a positive, although non-significant relationship of IGF-I levels with the 19 allele was also observed in premenopausal women (*179*). The other two studies recruited through institutions in the same city as hospitals used for recruitment for this investigation. Of these, one found the 5′ 19 allele to be inversely associated with circulating IGF-I concentrations in oral contraceptive users only (*174*), while the other found no association in premenopausal women but a borderline significant inverse association in postmenopausal women (*99*).

Several studies have found evidence for a relationship between mammographic density and circulating concentrations of IGF-I, IGFBP-3 or the IGF-I/IGFBP-3 ratio (*5-8,99*) in premenopausal women (Table 1). A previous study that included the same Caucasian sample as this one, with some additional subjects, reported significant positive associations

of IGF-I with percentage breast density and the amount of dense tissue (*7*). Two studies reported a positive association of IGF-I concentrations with percentage breast density in premenopausal women, and in addition reported an inverse association with circulating IGFBP-3 concentrations and a positive association with the IGF-I/IGFBP-3 ratio (*5,8*). A fourth study reported a positive association between IGF-I and percentage density of borderline significance, although results were not consistent across samples that were assayed at different times (Table 1) (*6*). The same study also reported an inverse association between IGFBP-3 levels and percentage breast density, and a positive association between the IGF-I/IGFBP-3 ratio and percentage breast density (*6*). All of these results are consistent with the association between percentage density and *IGF1* genotype reported here. Evidence from this study suggests that the association of IGFBP-3 levels with mammographic density may be modified by the 3′ polymorphism genotype. This might in part explain the inconsistently observed association between IGFBP-3 levels and mammographic density, across different studies (*5-8,99,162,164*) .

Two other studies have investigated the association between allelic variants of *IGF1* and mammographic density. Consistent with the results presented here, no association was found with the number of 19 alleles and percentage breast density (*99*). In a large sample comprised largely of postmenopausal women, a strong association between mammographic density and haplotypes from 3 of 4 haplotype blocks of the *IGF1* gene and 4 *IGF1* SNPs was observed (*100*). The SNPs associated with density were found in the introns, and consistent with the current investigation, at the 3′ end of the gene. Unlike the current investigation, the association was observed even though BMI was included as a covariate and therefore appears unlikely to be influenced by body fat, as suggested by the results here.

A number of studies have examined the relationship between circulating IGF-I concentrations and measures of body composition, but results are difficult to interpret. Studies that have measured body mass index have observed mainly null associations (*123-128*), but inverse (*118*), and positive associations (*129*), have been reported. A non-linear association between IGF-I concentrations and BMI is suggested by the results of some studies with lower IGF-I concentrations reported in individuals with relatively low or high BMI (*130-135*). There is some evidence for an inverse association between IGF-I levels and visceral adipose tissue in obese subjects (*136-139*) and a consistent inverse relationship between IGFBP-I levels and BMI provides support for a relationship between the IGF-I system and body fat stores (*125,129,132,133,140-143*). Consistent with our results, an *IGF1* CT repeat polymorphism was reported to be associated with fat mass, percentage fat and fat free mass (*236*).

The number of copies of the 3′ 185 allele had a strong influence on mammographic density. Average percentage breast density was about 11% greater in women with two copies of this allele verses non-carriers. This is a large difference when compared to most other factors related to mammographic density. Although BMI was found to be more strongly associated with percentage breast density in this study (a simple dichotomization of BMI into categories of less than or greater or equal to 25 kg/m^2 produced a greater than 25% gradient in percentage breast density) parity was not even significantly associated with percentage breast density. Furthermore, in a previous study, change in breast density attributable to change in menopausal status was observed to be only about 3% (over an average of 1.5 years during which menopausal status changed in study subjects) (82). The 11% change in percentage breast density can further be placed into context by considering the results of a

recent study conducted in Canada where an 80% increase in risk from the lowest to second lowest category of density was reported (< 10% density verses 10 < 25% density) (58). An 11% difference in percentage density is roughly comparable to the difference in the median of these two categories and therefore should result in a meaningful increase in breast cancer risk.

There appear to be no other studies that have reported on the repeat at the 3′ end of the gene and there is no evidence to suggest that this polymorphism, which lies in the 3′ untranslated region, has functional significance. However, a relationship between this polymorphism and mammographic and anthropometric features may be explained by linkage disequilibrium between this marker and another *IGF1* polymorphism that influences gene function. Although this study did investigate the association of haplotypes constructed from these polymorphisms, no significant associations with mammographic density were revealed. There were also no significant associations with anthropometric features other than height. Further studies that employ tightly spaced markers are needed to determine if there is a functional variant in the vicinity of the 3′ polymorphism.

In conclusion, results from this study suggest an association between the number of copies of the 3′ 185 allele and mammographic density. This association appears to be mediated through an influence on body fat. Associations with BMI and height were also observed. If, however, there is a relationship between this allele and these outcomes, it does not appear to be mediated through an effect on circulating IGF-I levels.

All of the outcomes for which an association of the 3′ 185 allele was reported are risk factors for breast cancer. Greater mammographic density is strongly related to increased breast cancer risk. In premenopausal women, obesity may reduce the risk of developing

breast cancer and height is suggested to increase risk (*17,26*). Several studies have examined *IGF1* polymorphisms and breast cancer risk. Four studies have reported a significant association of various alleles of the 5′ polymorphism with breast cancer risk, with two finding the 19 allele to be associated with disease (*189,191-193*). Three studies have reported significant associations with various SNPs and breast cancer risk (*187,194,195*). The results from this study suggest that further investigations into the relationship between breast cancer risk and *IGF1* are warranted and that these should include the 3′ polymorphism and other polymorphisms that may be in linkage disequilibrium with it.

Chapter 5. Family based study examining the association of *IGF1* polymorphisms and breast cancer risk in premenopausal women

ABSTRACT

Background and objectives

In general, studies indicate that greater circulating IGF-I concentration is associated with increased breast cancer risk in premenopausal women. Both breast cancer risk and circulating IGF-I concentrations appear to be partly heritable. Therefore, *IGF1* is a suitable candidate gene to examine in association with breast cancer risk.

Methods

DNA and risk factor data from 840 premenopausal probands with breast cancer and their first degree relatives were obtained from the Ontario Familial Breast Cancer Registry (OFBCR) and the Australian Breast Cancer Family Registry (ABCFR). The association between allelic variants of three CA repeat polymorphisms, including a previously investigated 5′repeat, with breast cancer risk in premenopausal women was examined.

Results

Some nominally significant associations (5′ 21 allele, P=0.03; intron 2 212 allele, P=0.04; intron 216 allele, P=0.04) were observed in the combined ABCFR OFBCR sample, but adjustment for multiple comparisons indicated that these had a high probability of being false discoveries. In additional analyses, a stronger association between the intron 2 216 allele with risk (P=0.01) was observed under a recessive model and a*d hoc* grouping of 5′ polymorphism alleles resulted in a significant positive association of with risk for alleles of length 18 to 20 (P=0.02) and a inverse association for alleles greater than 20 repeats in length

(P=0.01). These same associations were observed in the OFBCR (intron 2 216 allele recessive, P=0.02; both 5′ 18 to 20 allele grouping and >20 allele groupings, P=0.01) but were not strongly supported in the ABCFR (intron 2 216 recessive, P=0.14; 5′ 18 to 20 grouping, P=0.25; 5′ > 20 grouping P=0.20). Analysis of haplotypes resulted in few nominally significant associations that could have been due to chance.

Conclusions

Given the number of comparisons performed and the lack of strong consistency across samples, the results provide limited evidence for an association between genetic variants of *IGF1* and breast cancer risk.

INTRODUCTION

Insulin-like growth factor I (IGF-I) has both mitogenic and anti-apoptotic effects on mammary epithelial cells (*237,238*), suggesting a potential role in the development of breast cancer. A number of cohort and case-control studies have investigated the association of circulating IGF-I concentration and breast cancer risk. These studies indicate that the circulating concentration of IGF-I is associated with an increased risk of breast cancer in premenopausal women, although results are not entirely consistent. An association between circulating IGF-I levels and breast cancer risk in postmenopausal women is not supported (*1-3*).

It is estimated that 27% of breast cancers are attributable to genetic factors (*9*). Estimates of heritability of circulating IGF-I concentration range from 38-63% in middle aged and elderly men and women (*11,12*). Therefore, genetic factors may play a role in influencing IGF-I levels and ultimately breast cancer risk.

There is a cytosine-adenosine (CA) dinucleotide polymorphism in the promoter region of *IGF1* (5′ polymorphism). Seven studies have investigated the association between specific alleles of this polymorphism and breast cancer risk (*172,179,189-193*). Four have reported a significant association (*189,191-193*), although in only two studies has the same allele been reported to be associated with risk (*189,192*).

Only two of these studies (one in a Chinese sample, the other which examined Caucasians) investigated the association of this polymorphism in premenopausal women in samples of adequate size (*189,193*). Significant associations between variant alleles of the *IGF1* gene and breast cancer risk were reported, although with respect to the association of specific alleles with risk, inconsistent and contradictory results were reported (*189,193*).

Three studies have specifically examined *IGF1* tagging SNPs in relation to breast cancer risk. Two found nominally significant associations with SNPs at the 5′ end, although results were not significant after adjustment for multiple comparisons (*194,195*). A third study reported significant associations with five tagging SNPs across the *IGF1* gene with strongest support for an association of a tagging SNP in intron 3, based on its association with circulating IGF-I levels (*187*). None of the studies reported associations stratified by menopausal status, and two of three had only moderate sized samples of premenopausal women, while the third included prevalent cases (*187,194,195*).

Inconsistency among the reported relationships between *IGF1* and breast cancer risk may be due to the differences in linkage disequilibrium patterns among ethnic groups examined in different studies (e.g., Caucasian vs. Chinese (*189,193*)), confounding due to population stratification, or effect modification by other genetic or environmental factors. Multiple testing may also have played a role in creating spurious associations for the 5′

polymorphism, which has several alleles. The focus of most studies on the 5′ polymorphism may also contribute to inconsistency among studies since (if viewed as a marker) this one polymorphism alone is unlikely to capture genetic variation across the gene.

This study examined the association of CA repeat alleles of the *IGF1* gene and breast cancer risk in a large sample of premenopausal women from two population based registries. The study was family based, removing the potential for bias resulting from confounding due to population stratification. The 5′ polymorphism was examined with respect to breast cancer risk, as was the association of two other CA repeat polymorphisms, in intron 2 and at the 3′ end of the gene. Linkage disequilibrium between polymorphisms was examined, and the association of haplotypes constructed from these polymorphisms with breast cancer risk was investigated. Interactions between *IGF1* genotype and breast cancer risk factors (body mass index, family history of breast cancer and oral contraceptive use) were explored.

METHODS

This study used a family based design. Families had been previously recruited by the Breast Cancer Family Registry for the purpose of conducting studies on the genetic and molecular epidemiology of breast cancer (*239*). Families with affected premenopausal probands with either siblings (including some affected sisters), parents or combinations of siblings and parents were selected from this registry for the study presented here.

Recruitment

Two sites of the Breast Cancer Family Registry, the Ontario Familial Breast Cancer Registry (OFBCR) and the Australian Breast Cancer Family Registry (ABCFR) provided DNA samples and risk factor data for study subjects. During initial recruitment, cases were

identified using local population based cancer registries: the Ontario Cancer Registry in Ontario, Canada, and the Victorian and New South Wales Registries in Australia. Probands were either residents of Ontario, Melbourne or Sydney at the time of diagnosis.

Probands in OFBCR families were incident primary invasive breast cancer cases identified by pathology report beginning in 1996 and ending in 1998. Cases with prior diagnoses were reviewed to ensure they matched criteria established for a new primary cancer. During this time period, pathology reports were received for 90% of breast cancer cases that were recorded in the Ontario Cancer Registry.

Initial enrolment included female cases aged 20-54, a random sample of female cases aged 55-69, and male cases under age 80 (the latter comprised a small proportion of those enrolled as there were only about 100-120 incident male cases per year reported in Ontario). A family history questionnaire was used to classify families as moderate to high risk for developing breast cancer (Table 1). All moderate to high risk families were eligible for the study. A random sample (approximately 25%) of families designated low risk for breast cancer was also included in the sample (*240*).

Figure 1 shows steps in the recruitment of OFBCR families. Initially, 8,446 cases were eligible for contact for the purpose of classifying family history status. Physicians refused to give permission to contact 5.6% of cases. In addition 1.7% of cases were already deceased and an additional 2.1% of cases could not be contacted through physician records. Of the remaining 7,662 cases, 14.8% refused to participate and 20.4% either did not respond or could not be contacted. A completed family history questionnaire was returned by 4,962 cases (59% of the eligible 8,446 cases). After assessment of family history status, just over half of the cases, 2,585 (52%), met eligibility criteria. A risk factor questionnaire was

completed by 1,862 (72%) of these. Blood samples from a female proband and at least one first degree relative was obtained for 606 cases who completed the risk factor questionnaire. Probands and sisters with a diagnosis of breast cancer were considered premenopausal if their periods had not stopped for more than one year at the time of their diagnosis and they were either not using hormone replacement therapy, or stopped using hormone replacement therapy two or more years prior to diagnosis or a year prior to their periods ending after diagnosis. Of the 606 available probands, 400 were determined to be premenopausal at time of diagnosis, potentially providing 400 families for analysis (Fig. 1). However, in 62 of these families DNA from the Mount Sinai repository was unavailable for either the proband or relatives, leaving 338 families with DNA samples that potentially could be analysed. DNA of six probands was never tested because other ongoing studies indicated that these were of insufficient quality for analysis. This left 332 OFBCR families available for genotyping for this study.

Enrolment into the ABCFR required that cases were incident first primary female breast cancers aged 20-59 and diagnosed from 1992 to 2000 (Fig. 2). Enrolment from 1992 to 1995 pre-dated the formation of the ABCFRs and was restricted to women under the age of 40. Family history status was not a factor in determining eligibility for the ABCFR. Interviews were conducted for 1,578 of 2,303 eligible cases (68.5%). The patient's surgeon refused to allow contact for 8.5% of eligible cases. Refusal by the case to participate (16.4%), death prior to contact (1.8%), non-response by surgeon (1.3%) or case (1.2%), and failure to locate the case (2.3%), accounted for the remaining 23% that did not enroll. DNA samples were available from 1,453 cases (92% of those who completed an interview), 1,048

of whom were premenopausal. Of these, 564 (54%) had at least one first degree relative who provided a blood sample.

The same definition for premenopausal status that was used in the OFBCR, was employed for ABCFR probands recruited beginning in 1996. Prior to 1996, probands were considered premenopausal if their periods had not stopped for more than one year at the time of their diagnosis or interview, or if their periods had stopped due to pregnancy. No women with premenopausal breast cancer were diagnosed over the age of 56 in either the ABCFR or OFBCR.

Table 1. Criteria for defining moderate/high risk families used by the OFBCR.

- Proband + ≥ one 1st degree relative with breast or ovarian cancer.
- Proband + ≥ two 2nd degree relatives with breast or ovarian cancer.
- Proband diagnosed with breast cancer at age ≤ 35.
- Proband + ≥ one 2nd degree or ≥ one 3rd degree relative diagnosed with breast cancer at age ≤ 35, or ovarian cancer at age ≤ 60.
- Proband with ≥ one 2nd degree or ≥ one 3rd degree relative with male breast cancer.
- Proband with breast and ovarian or multiple breast primaries.
- Proband + ≥ one 2nd degree relative or ≥ one 3rd degree relative with breast and ovarian cancer.
- Proband + ≥ one 2nd degree relative or ≥ one 3rd degree relative with multiple breast cancer primaries.
- Family has three 1st degree relatives, each with any of the following cancers: breast, ovarian, colon, prostate, sarcoma, or pancreas, with at least one family member diagnosed ≤ 50 years of age.
- Proband is Ashkenazi Jewish.

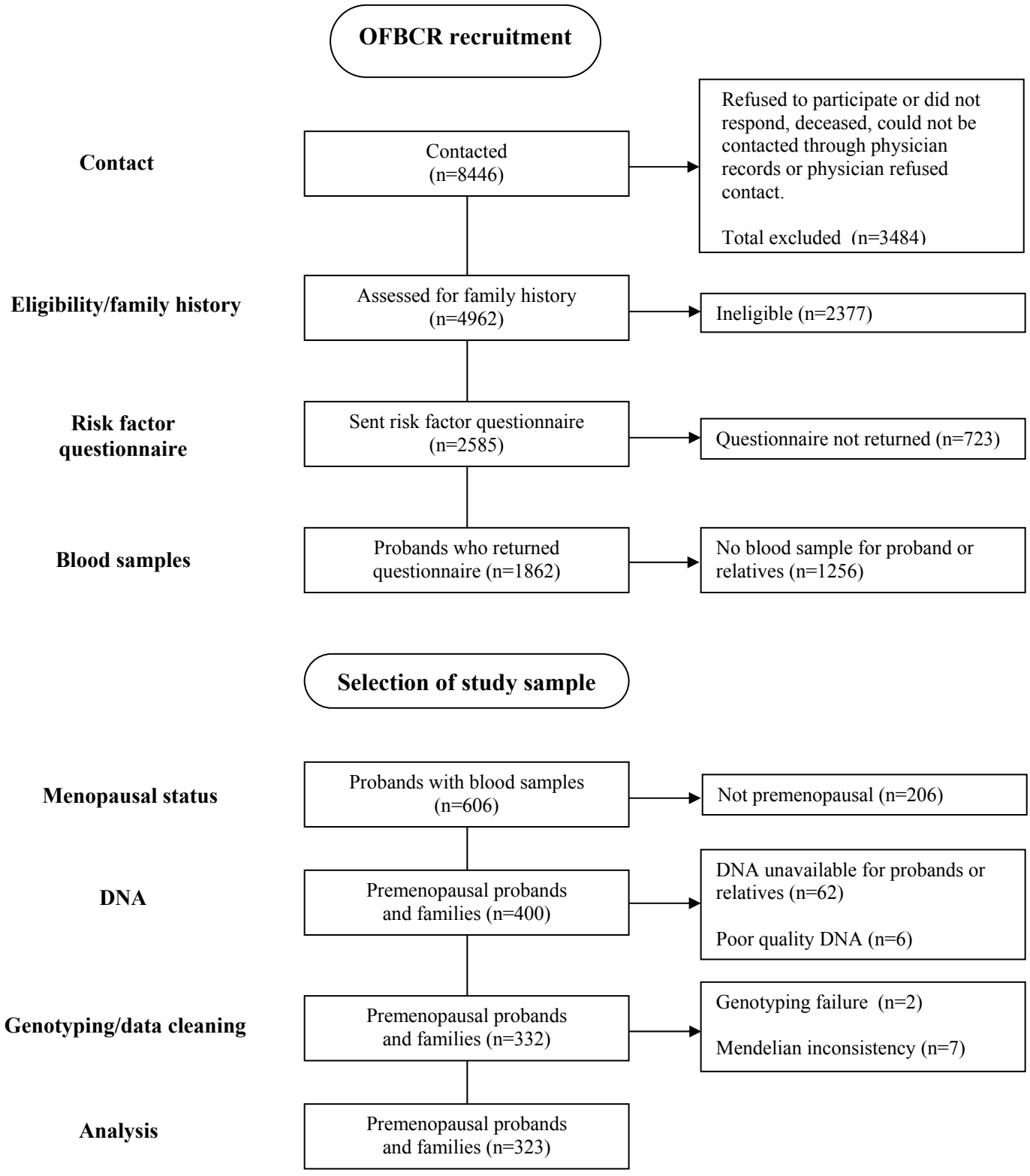

Figure 1. Recruitment of OFBCR study subjects and selection of OFBCR families for this study.

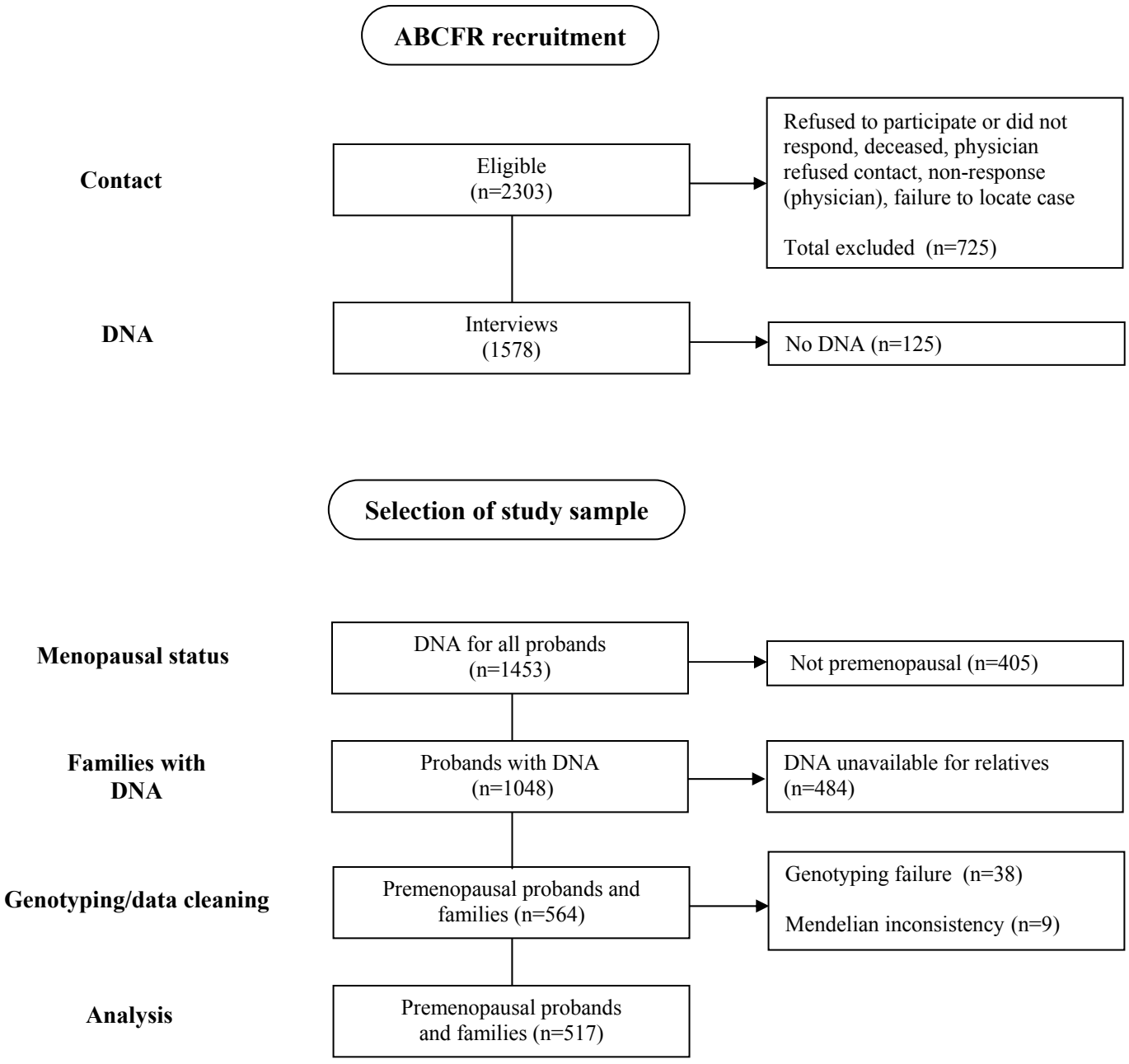

Figure 2. Recruitment of ABCFR study subjects and selection of ABCFR families for this study.

Measurements

Laboratory methods

DNA extraction

DNA samples were obtained from the Ontario Cancer Genetics Network repository in Ontario and from the ABCFR in Melbourne.

In Ontario, peripheral blood lymphocytes were pelleted and stored at –70ºC to –80ºC and DNA was later extracted from the pellets using either phenol/chloroform extraction or the Qiagen procedure (QIAamp DNA Mini Kit (250)™). The ABCFR stored dried blood samples on Guthrie cards. DNA was extracted from Guthrie card dried blood spots using a Chelex method. All samples were sent to the laboratory of Dr. H. Ozcelik where they were prepared for genotyping at The Centre for Applied Genomics (TCAG) (*216*).

Genotyping

Three polymorphisms were selected for analysis; the previously identified 5′ CA repeat polymorphism (*175,228*), and two other CA repeats, located in intron 2 and at the 3′ end of the gene, that were identified using Tandem Repeat Finder software version 2.02 (*229*). The oligonucleotide primers used for PCR were as follows: 5′-GCTAGCCAGCTGGTGTTATT-3′, 5′-ACCACTCTGGGAGAAGGGTA-3′ for the 5′ polymorphism, 5′ CATACTTCTTAGCTCCTCAGG-3′, 5′-CCCTCACAGAAAGCAGAA-3′ for the intron 2 polymorphism, and 5′-CTTTTTAAGATGAGGCAGTTCC-3′, 5′GATTTCTTTTCAGTATTCCATTGG-3′ for the 3′ polymorphism. Position of amplified regions of DNA corresponded to positions 26,357,243-26,357,436, 26,332,059-26,332,274

and 26,275,038-26,275,220 on contig NT_019546.15 of build 35.1 of the NCBI's genome annotation for the 5′, intron 2 and 3′ polymorphisms respectively.

For the OFBCR sample, PCR was performed with 5 ng of DNA, 5µl 10X PCR buffer, 0.8mM dNTPs, 0.18mM of each primer, and 5 U (0.3µl and 0.2µl) Platinum Taq in each 50ul reaction. $MgCl_2$ concentration was either 3 mM for the 5′ and 3′ polymorphism and 4mM for intron 2, or 2mM for all polymorphisms depending on $MgCl_2$ stock used. Amplification conditions were 94ºC for 3 minutes, and 35 cycles of 94ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 1 minute for the 5′ repeat. Amplification conditions were the same for the other polymorphisms with the exception that annealing temperatures were 60ºC or 55ºC for the intron 2 polymorphism (depending on $MgCl_2$ stock used) and 50ºC for the 3′ polymorphism. For ABCFR samples, PCR was performed with either 6mM or 4mM of magnesium depending on $MgCl_2$ stock used. Amplification conditions were identical to that used for the OFBCR sample with the exception that annealing temperature was either 55ºC (all 3 polymorphisms) or 55ºC for the 5′ and intron 2 repeat, and 50ºC for the 3′ repeat, depending on $MgCl_2$ stock used. All reagents were supplied by Invitrogen Life Technologies. The forward primer was labeled with a 5′ fluorescent dye (HEX) and the reverse primer with a 5′ sequence (GTTTCTT). Genotyping was performed by TCAG using performance Optimized Polymer 6 (POP6) gel capillary electrophoresis systems (ABI 3100 and ABI 3700 systems) (*216*). Computerized output used by TCAG to assign genotypes (electropherograms) was also examined by G. Fehringer to identify allele calls that were suspicious and needed to be repeated.

Retesting of Samples

A small percentage of samples were retested because they failed (8% in the ABCFR and 5% in the OFBCR) or resulted in irregular shaped electropherograms (3% in both the ABCFR and OFBCR) on the first genotyping attempt. OFBCR samples were tested up to two more times if at least one of three polymorphic loci originally tested for the same subject had been successfully genotyped. In cases where no genotype data was returned for all three loci for a particular subject, only one additional retest per polymorphism was performed. ABCFR samples were only retested once as a limited supply of DNA was available.

Reliability of genotyping

Reliability of TCAG allele calls was estimated by retesting a randomly selected sample representing a minimum of 10% of DNA specimens from each 96 well plate. Laboratory staff at TCAG were not informed which plates included samples for reliability tests.

Some of the reliability samples from the OFBCR and ABCFR did not run successfully (either TCAG did not provide an allele call or upon review of the electropherogram, the call was considered unreliable). If these samples were simply discarded, an overestimate of reliability might result if a greater proportion of the samples that failed when retested were of low quality and therefore would have been more likely to produce a call that disagreed with the original. Four OFBCR reliability test samples failed, one each for the 5′ and intron 2 polymorphisms and two for the 3′ polymorphism. These retested samples had been categorized as having good quality electropherograms upon previous inspection by G.F. This was considered evidence that indicated that the retest samples were of good quality and therefore need not be repeated. Instead, an additional random sample was substituted from the same plate of the failed sample (there were 2 additional random samples per 96 well

plate) and included in the calculation of percentage agreement between the original runs and the reliability sample. There were 16 failures for the ABCFR sample. There were, however, only 17 samples originally considered to have electropherograms of questionable quality and only 3 of these (all for the 3′ polymorphism) failed. The total number of samples used in calculation of reliability for the ABCFR for this polymorphism was 181. Therefore, the failed samples were not repeated since the number of failures was small, DNA was limited and the process of freezing and thawing of these samples each time they were retested was likely reducing the quality of the sample.

Concordance for the ABCFR sample was 97.2% (n=179), 94.1% (n=186) and 97.2% (n=181) for the 5′, intron 2 and 3′ polymorphisms respectively. Concordance for the OFBCR sample was 99.1%, 98.1%, and 100% for the 5′, intron 2 and 3′ polymorphisms respectively (n=108 for each polymorphism).

Assessment of Centre for Applied Genomics allele calls

Accuracy of TCAG allele calls was assessed by randomly selecting electropherograms representing at least 10% of allele calls for each 96 well plate and having readers assign genotype independently of TCAG. The ABCFR sample was inspected by two readers (G. Fehringer and H. Jaranazi) independently. Discrepancies were resolved and results compared to the TCAG allele calls. The OFBCR peaks were inspected by G. Fehringer alone. Agreement for the ABCFR sample was 98.4%, 98.9% and 100% for the 5′, intron 2 and 3′ polymorphisms respectively (n=190, 186 and 186 for each polymorphism respectively). Agreement for the OFBCR sample was 99.1% for all three polymorphisms (n=109, 108 and 107 for the 5′, intron 2 and 3′ polymorphisms respectively).

Mendelian inconsistency

Pedigree errors were examined using the program PEDAGREE (*241*). The extent of the Mendelian errors within a family determined whether the entire family was excluded from analyses or whether genotype data for a specific individual was removed for one or more polymorphisms. The criteria used are summarized in Table 2. Mendelian errors were identified in 15 of 330 OFBCR families resulting in the removal of 7 families prior to analysis. In the ABCFR sample, Mendelian errors were detected in 43 of 526 families resulting in the removal of 9 families.

Genotyping success rate

Removal of some families with Mendelian errors resulted in the number of families available in Ontario for analysis being reduced to 325. Genotyping success rate was 99.7%, 99.3% and 99.1% for the 5′, intron 2 and 3′ repeat respectively. Two families were discarded because of genotyping failures (for all three polymorphisms) in the probands, leaving 323 families for the analysis.

There were 555 families available for genotyping from the ABCFR sample (accounting for the removal of families with Mendelian errors). Genotyping success rate for the ABCFR samples was 90.1%, 90.3% and 90.0% for the 5′, intron 2 and 3′ polymorphisms respectively. Failures in genotyping resulted in 38 families being discarded leaving 517 families for analysis.

Table 2. Criteria for removing individuals or families when a Mendelian inconsistency was identified.

- Error in one polymorphism. Genotype data removed for a polymorphism unless:
    - Removal of one family member restored Mendelian consistency, while removal of no other single family member restored Mendelian consistency (genotype data of the one family member was then removed).
    - Alleles from mother were not observed among four offspring while alleles from father were observed (genotype data of mother removed).
- Error in two or more polymorphisms. All genotype data removed for family unless:
    - Removal of one family member restored Mendelian consistency for two polymorphisms while removal of no other single family member restored Mendelian consistency (genotype data for one family member was removed).
- All genotype data also removed for family in the following situations:
    - Genotype data for proband included allele not found in either parent for one polymorphism, and data missing for second polymorphism (one proband two parent family only).
    - Genotype data for proband did not agree with mother's genotype for one polymorphism in a family composed of a mother daughter pair.

Risk factor questionnaires

Extensive risk factor questionnaires were completed by both ABCFR and OFBCR family members. These included questions about oral contraceptive use, weight (prior to diagnosis for probands), and height. ABCFR questionnaires were completed through interview. OFBCR questionnaires were self administered.

Analysis

Hardy-Weinberg equilibrium for each polymorphism was examined using a permutation version of the exact test (*230*) (implemented in Power Marker 3.25 (*231*)).

Family based association tests incorporated into the programs FBAT (Family Based Association Testing, version 1.5) (*219*) and PBAT (Pedigree Based Association Testing, version 2.5) (*207*) were used for main analyses. These programs test for association under a null hypothesis of no linkage or linkage disequilibrium between the marker(s) and the disease susceptibility locus. The expected distribution is based on parental genotype data (using genotype data directly from both parents or reconstructed from a genotyped parent and/or siblings). Given Mendelian inheritance patterns all alleles are equally likely to be transmitted to the case or proband and deviation from this expected distribution is evidence for association of the gene with disease. Since the expected distribution is based on the parental genotype data, the resulting statistical test is not susceptible to confounding due to population stratification. Furthermore, no assumptions need to be made regarding assortative mating (deviation from Hardy-Weinberg equilibrium) and the ascertainment condition (*219*).

Each of the polymorphisms genotyped had rare alleles. In order to ensure reliable results, alleles were included in analyses only if at least ten of the families were informative for a given allele. An informative family is one where parental genotypes (or the reconstructed parental genotypes) include at least one parent that is heterozygous for a given allele.

All probands were premenopausal incident cases and were coded as diseased for all analyses. Sisters of probands who were diagnosed with premenopausal breast cancer, either during or prior to the recruitment period, were also coded as diseased. Sisters diagnosed with premenopausal cancer after the recruitment period were coded as not having disease, as were sisters who had no diagnosis of breast cancer or those diagnosed with breast cancer when they were no longer premenopausal. Two sisters with an unconfirmed report of breast cancer

(one each from the ABCFR and OFBCR) were coded as having unknown disease status. Family members other than sisters and probands were also coded as having unknown disease status. The FBAT program was used to analyze the association between allelic variants at individual loci and disease status, using an offset that minimizes variance. These analyses were repeated with PBAT (version 2.5) using an offset that maximizes power. Since there were no appreciable differences using FBAT and PBAT, results for FBAT are presented since the output provides a clear indicator of the direction of an effect. Additive models were used in analyses unless otherwise specified. Simulations have shown that the additive model has good power to detect an association, even when the true genetic model is not an additive one (*223*).

Association of allelic variants with breast cancer risk were examined for both the combined ABCFR OFBCR sample and for the two samples separately. Results between samples were compared with consistency (e.g. significance in both samples) being one criterion by which an association may indicate a true effect. In addition, the Benjamini-Hochberg correction was used to adjust P-values to account for multiple comparisons in the combined sample (*224*).

Linkage disequilibrium was assessed between loci using PowerMarker vers. 3.25 (*231*). A global test was performed and linkage disequilibrium between alleles from different loci was also examined. Haplotype analyses were performed using two loci when linkage disequilibrium between loci was statistically significant and at least one R^2 value between allele pairs in either the ABCFR or OFBCR was greater than 0.1. Analysis of the association of haplotypes with breast cancer risk used a weighted conditional approach incorporated into the program FBAT 1.5 (*242*).

Previous studies suggested that family history status may modify the association of the 19 allele with breast cancer risk (*179,190,193*). Analyses that stratified by family history and examined the relationship of specific alleles of the three polymorphisms with breast cancer risk were performed. Families were included in the family history stratum if one or more first degree relatives of the proband were reported to have breast cancer, otherwise they were considered not to have a family history of breast cancer.

Some evidence suggests that oral contraceptive use and BMI also may modify the relationship of the 19 allele with risk (*189,193*). Interactions between IGF-I genotype and BMI (analysed as a continuous variable) and oral contraceptive use (current use, ever use and duration of use) were investigated in relation to breast cancer risk using PBAT. Analyses examining genotype with risk was also performed, stratified by ever and never users of oral contraceptives. An offset was chosen to allow only individuals affected with breast cancer to contribute to the test statistic to avoid the problem of including families in the analysis with a mixture of affected or unaffected individuals with different histories of oral contraceptive use.

Interaction between the 185 allele of the 3′ polymorphism and BMI was also investigated in relation to breast cancer risk since results in chapter 4 indicate that the number of copies of this allele may interact with BMI to influence IGF-I levels. Interactions between both genotype and registry (ABCFR and OFBCR) and genotype and age in relation to risk were also investigated. For women diagnosed with premenopausal breast cancer during or (for some sisters of probands) prior to the study period, age was assigned as age at diagnosis. Women not diagnosed with breast cancer were assigned their age at interview or age at menopause, whichever was earlier, or age when hormone replacement therapy began

for women categorized as having unknown menopausal status due to hormone use. One woman diagnosed with breast cancer after the study period but prior to their interview date were assigned their age at the end of the study.

Ethics

Ethics approval for this study was received from the Mt. Sinai Hospital Research Ethics Board in Toronto.

RESULTS

Characteristics of subjects and completeness of data

Table 3 shows characteristics of the study subjects. The women in the ABCFR sample were somewhat younger than those in the OFBCR reflecting the restriction of enrollment to women under 40 from 1992 to 1995 in the ABCFR. In 93% of ABCFR families and 94% of OFBCR families, all family members who reported their race or ethnicity identified themselves as white. The vast majority of sisters in the ABCFR and the OFBCR did not have breast cancer.

The ABCFR sample had fewer one parent one proband families (Table 3), as these were not requested from the ABCFR since they add little power. The few that are present are the result of a failure to successfully genotype other family members. The OFBCR had relatively few samples with one or more siblings and both parents genotyped (Table 3) because for some families siblings were not genotyped when DNA from both parents was available. Adding siblings does not add information when parental genotype is already known and therefore does not improve study power. For each polymorphism, some families were discarded from analyses because of missing genotype data. Final counts of families by

polymorphisms were 321, 318 and 317 for the 5′, intron 2 and 3′ polymorphisms respectively in the OFBCR. In the ABCFR final counts were 502, 490 and 497 for the 5′, intron 2 and 3′ polymorphisms respectively.

The OFBCR over sampled probands with a family history of breast cancer. In this sample 241 of 323 families (75%), were classified as having a family history of breast cancer according to OFBCR criteria (shown in Table 1). Defining family history as having a mother or sister with breast cancer, 11% of ABCFR probands and 27% of OFBCR probands had a family history of breast cancer (Table 3).

Table 3. Descriptive characteristics of ABCFR and OFBCR samples.

	ABCFR	OFBCR
Number of Families	517	323
Age (S.D.) of diagnosis probands	37.3 (6.3)	42.2 (6.7)
Caucasian families	482 (93%)	303 (94%)
Affected sisters	5	5
Unaffected sisters	688	251
Family history[a]	58 (11%)	87 (27%)
Family configurations:		
Both parents no siblings	48 (9.3%)	53 (16.4%)
Both parents and siblings	110 (21.3%)	5 (1.5%)
One parent and siblings	125 (24.2%)	64 (19.8%)
One parent no siblings	4 (0.8%)	34 (10.5%)
Siblings only	230 (44.5%)	167 (51.7%)

[a] Probands with mother or sister with breast cancer

Allele frequencies

Frequencies of alleles examined in analyses (those common enough to permit analysis with 10 informative families or more) are shown for both ABCFR and OFBCR families,

combined and separately, in Table 4. These include all ethnic groups and individuals with and without breast cancer. Randomly selecting one family member (excluding women who ever had breast cancer) from each family where members reported themselves as being white, produced a sample with allele frequencies similar to those in Table 4. In the OFBCR allele frequencies for the 5′ polymorphism among unaffected Caucasian family members were 1.0%, 4.3% 62.8%, 20.3% 7.6%, 3.3% and 0.7% for alleles 17, 18, 19, 20, 21, 22 and all others combined respectively. In the ABCFR frequencies for these same alleles were 1.4%, 5.0%, 64.1%, 20.8%, 5.4%, 2.4% and 1.0% for unaffected Caucasian family members. Both sets of 5′ allele frequencies (those from randomly selected unaffected Caucasian family members and those presented in Table 4) are similar to those observed in other Caucasian (*175,177,234*) and Jewish populations (*191*).

Significant deviation from Hardy-Weinberg equilibrium was found for the 5′ polymorphism in the OBCFR (P=0.04) but not the ABCFR (P=0.92), or for intron 2 or 3′ polymorphism in either sample (intron 2: ABCFR P=0.61, OFBCR P=0.42, 3′: ABCFR 0.20, OFBCR 0.47). Although deviation from Hardy-Weinberg equilibrium as seen for the 5′ polymorphism can be an indication of genotyping error, there is unlikely to be significant error here since reliability tests indicated that genotyping error should be minimal.

IGF1 genotype and breast cancer risk

Preliminary analyses that examined interactions between registry and genotype in relationship to risk revealed two significant results (5′ 18 allele (P=0.03) and 5′ 19 allele (P=0.02)), from a total of 23 statistical tests. Since these might be due chance, data from both registries are combined, but analyses by registry are also presented to permit comparisons.

Association of specific alleles of the 5′, intron 2 and 3′ polymorphisms with breast cancer risk are shown in Table 4 (a minus sign (-) for the Z-score is shown for an inverse association otherwise associations are positive). Nominal significance ($P \leq 0.05$) was observed for the 21 allele of the 5′ polymorphism and the 212 and 216 alleles of the intron 2 polymorphism in the combined ABCFR OFBCR sample. When analyses were stratified by registry, there were no alleles for which a significant association was found in both the ABCFR and OFBCR. The 5′ 21 allele was associated with a decreased risk in both the ABCFR and OFBCR samples but nominal significance was observed only in the OFBCR sample. The intron 2 212 allele showed nominal significance in the ABCFR, but the number of informative families was too few to assess association in the OFBCR. The 216 allele was positively associated with risk in both samples, although neither was significant at the $P \leq 0.05$ level. Nominal significance also was observed for the 5′ 18 allele in the ABCFR sample and the 3′ 191 allele in the OFBCR sample. Results for the 5′ 18 allele in the OFBCR and the 3′ 191 allele in the ABCFR did not support these findings.

Benjamini-Hochberg correction of P-values was used with the false discovery rate (FDR) set at 0.05. This resulted in a corrected P-value of 0.007 for the largest nominally significant P-value of 0.04 in the set of P-values reported for the combined ABCFR OFBCR sample. A corrected P-value of 0.007 would be achieved only if the FDR was set at 0.31, indicating that significant results observed at $P \leq 0.05$ have a high probability of being false positives.

For the purpose of comparison to a previous study that reported an association between *IGF1* SNPs (*187*) using a recessive model, analyses were repeated using a recessive model in FBAT. In the combined ABCFR OFBCR sample, a nominally significant positive

association was observed for the 216 allele (P=0.01). A nominally significant positive association was also observed in the OFBCR (P=0.02), but although the direction of effect was positive in the ABCFR the results was not significant at P≤0.05 (P=0.14). In the ABCFR a nominally significant positive association of the 3′ 185 allele with risk was observed (P=0.02). However, in the OFBCR the association was not significant (P=0.58) and in the opposite direction. No other significant associations were observed using a recessive model.

Exploration of the interaction between age and genotype in relationship to breast cancer risk produced only a few nominally significant results for specific alleles, which could be attributed to chance. These were the intron 212 allele in the combined sample (P=0.05), and the 5′ 21 allele in the OFBCR (P=0.01). The interaction test for the intron 2 212 allele could not be compared across samples because there was only one informative family available in the OFBCR for this test. The test for interaction between 5′ 21 allele and age in the ABCFR (P=0.71) did not support the result observed for the OFBCR.

Table 4. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′												
11	0.5	11	0.24	0.81	-	<10	-	-	-	<10	-	-
17	1.6	33	0.03	0.97	1.5	21	0.23	0.82	1.8	12	-0.17	0.87
18	6.2	129	0.72	0.47	6.5	88	1.96	0.05*	5.7	41	-1.36	0.17
19	61.1	431	-0.31	0.76	60.6	274	-1.65	0.10	62.0	157	1.67	0.10
20	21.3	325	1.65	0.10	21.9	207	1.45	0.15	20.3	118	0.79	0.43
21	6.7	134	-2.17	0.03*	6.6	87	-1.19	0.23	6.9	47	-2.23	0.03*
22	2.1	46	-1.33	0.18	2.0	27	-0.23	0.82	2.4	19	-1.72	0.09
Intron 2												
204	2.3	51	-1.03	0.30	1.9	30	-0.26	0.79	2.9	21	-1.36	0.17
212	0.8	16	-2.03	0.04*	1.1	15	-2.28	0.02*	-	<10		-
214	3.0	61	-0.97	0.33	3.0	40	-0.31	0.76	3.0	21	-1.16	0.24
216	38.3	438	2.08	0.04*	38.3	266	1.35	0.18	38.3	172	1.71	0.09
218	21.9	343	0.11	0.91	21.3	218	0.10	0.92	22.8	125	0.36	0.72
220	16.8	266	-0.88	0.38	16.8	170	-0.31	0.76	16.8	96	-1.45	0.15
222	9.8	199	-0.85	0.40	10.4	134	-0.46	0.64	8.5	65	-0.98	0.33
224	6.2	144	-0.33	0.74	6.0	93	-0.33	0.74	6.4	51	0.06	0.95
226	0.6	15	0.44	0.66	0.6	10	-0.38	0.71	-	<10	-	-
3′												
181	0.9	23	-0.31	0.76	0.9	13	0.39	0.70	0.9	10	-0.83	0.41
183	1.2	26	0.99	0.32	1.1	18	0.47	0.64	-	<10		-
185	47.0	439	0.15	0.88	46.1	282	0.12	0.91	48.7	157	0.03	0.98
187	28.5	384	-0.19	0.85	29.2	238	-0.52	0.60	27.4	146	0.52	0.61
189	18.4	315	0.14	0.89	18.4	196	0.32	0.75	18.4	119	-0.25	0.80
191	2.9	64	-1.38	0.17	3.2	47	-0.52	0.60	2.3	17	-2.09	0.04*
193	1.0	20	0.40	0.69	1.1	13	0.09	0.93	-	<10	-	-

NIF: number of informative families

* Significant at $P \leq 0.05$.

- Results not reported when number of informative families is less than 10.

IGF1 haplotypes and breast cancer risk

Significant linkage disequilibrium was observed between adjacent polymorphisms and those at the 5′ and 3′ end (P=0.0002 for 5′ and 3′ polymorphisms in the OFBCR and P<0.00001 for all other pairs of polymorphisms in the ABCFR and OFBCR). Table 5 shows linkage disequilibrium between specific allele pairs for the three polymorphisms. Linkage disequilibrium between alleles was greater than 0.10 for one 5′-intron 2 haplotype and three intron 2-3′ haplotypes (5′–intron 2 22-214 haplotype (OFBCR only see Table 5 footnote), intron 2-3′ haplotypes 216-187, 222-189 224-189). Therefore, only associations with risk for haplotypes constructed using 5′-intron 2 and the intron 2-3′ polymorphism are reported (Table 6). Nominal significance was observed for three haplotypes (20-218: P=0.03; 21-220: P=0.03; 19-212: P=0.02) constructed from the 5′ and intron 2 polymorphism in the combined ABCFR OFBCR sample. Two of these (21-220 and 19-212) were constructed from alleles where nominal significance was observed in analyses of individual polymorphisms (allele 21 of the 5′ polymorphism and allele 212 of intron 2). No alleles were observed to show a significant association at the P≤0.05 level in both samples in analyses stratified by registry. Application of the Benjamini-Hochberg correction to the combined sample P-values in Table 6, resulted in no significant associations with the FDR set to 0.05. In analyses stratified by registry, a few nominally significant associations were observed in either the ABCFR or OFBCR that were not observed in the combined sample. However, comparisons of results across registries did not provide additional support for these findings.

The association of haplotypes constructed from the intron 2 216 allele with risk was also investigated under a recessive model. In the combined ABCFR OFBCR sample, a nominally significant positive association with risk was observed for the 19-216 allele

constructed from the 5′ and intron 2 polymorphisms (P=0.002). The association of this haplotype with disease risk was, however, inconsistent across the two registries as a strong association observed in the OFBCR (P=0.0004) was not supported in the ABCFR (P=0.40). This discrepancy in the results could not be explained by obvious differences among families where probands or sibling carried these haplotypes (e.g., differences in ethnicity, family history, *BRCA1* or *BRCA2* mutations) and therefore the results cannot be said to support an association of this haplotype with risk.

Table 5. Linkage disequilibrium between specific alleles.

	ABCFR			OFBCR		
Loci/haplotype	Frequency	D'	R^2	Frequency	D'	R^2
5′ – intron 2						
other-other	0.01	0.16	0.015	0.02	0.42	**0.124**[†]
other-216	0.02	0.01	<0.001	0.01	-0.66	0.014
other-218	0.01	-0.20	0.001	0.00	-0.56	0.005
other-220	0.01	-0.04	<0.001	0.01	-0.47	0.003
other-222	0.00	-1.00	0.006	0.01	0.01	<0.001
other-224	0.00	0.01	<0.001	0.01	0.06	0.002
18-other	0.00	-0.14	<0.001	0.01	0.14	0.011
18-216	0.04	0.49	0.021	0.02	0.20	0.003
18-218	0.01	-0.27	0.001	0.01	0.11	0.002
18-220	0.00	-0.71	0.005	0.00	-1.00	0.011
18-222	0.00	-1.00	0.007	0.00	-1.00	0.005
18-224	0.00	-0.19	<0.001	0.00	-1.00	0.003
19-other	0.04	-0.24	0.009	0.03	-0.29	0.011
19-216	0.22	-0.13	0.014	0.24	0.04	0.001
19-218	0.17	0.39	0.023	0.15	0.21	0.008
19-220	0.07	-0.27	0.024	0.08	-0.29	0.032
19-222	0.09	0.53	0.020	0.07	0.34	0.007
19-224	0.05	0.69	0.016	0.05	0.51	0.010
20-other	0.02	0.05	0.001	0.00	-0.98	0.019
20-216	0.10	0.16	0.011	0.09	0.09	0.003
20-218	0.02	-0.51	0.019	0.04	-0.05	<0.001
20-220	0.05	0.15	0.016	0.06	0.14	0.019
20-222	0.02	-0.33	0.004	0.01	-0.20	0.001
20-224	0.00	-0.91	0.014	0.00	-0.98	0.016
21-other	0.01	0.05	0.002	0.01	0.03	0.001
21-216	0.01	-0.45	0.007	0.02	-0.36	0.006
21-218	0.01	-0.18	0.001	0.00	-0.78	0.014
21-220	0.02	0.32	0.032	0.04	0.41	0.060
21-222	0.00	-0.44	0.001	0.00	-0.68	0.004
21-224	0.00	-1.00	0.003	0.01	0.01	>0.001

Continued on next page.

Table 5. Linkage disequilibrium between specific alleles (continued from previous page).

	ABCFR			OFBCR		
Loci/haplotype	Frequency	D'	R^2	Frequency	D'	R^2
Intron 2 – 3′						
other-other	0.01	0.04	0.001	0.00	-1.00	0.004
other-185	0.03	-0.06	<0.001	0.05	0.35	0.010
other-187	0.02	0.05	0.001	0.02	-0.02	<0.001
other-189	0.01	-0.23	0.001	0.00	-0.67	0.008
216-other	0.01	-0.50	0.011	0.01	-0.72	0.017
216-185	0.17	-0.02	<0.001	0.14	-0.23	0.031
216-187	0.18	0.40	**0.097**	0.19	0.53	**0.173**
216-189	0.03	-0.66	0.068	0.03	-0.49	0.031
218-other	0.01	-0.26	0.001	0.02	0.14	0.004
218-185	0.15	0.48	0.077	0.17	0.55	0.086
218-187	0.03	-0.55	0.032	0.02	-0.72	0.054
218-189	0.02	-0.45	0.014	0.02	-0.60	0.022
220-other	0.01	-0.03	<0.001	0.01	-0.31	0.001
220-185	0.08	0.11	0.003	0.11	0.22	0.011
220-187	0.05	0.07	0.002	0.04	-0.12	0.001
220-189	0.01	-0.52	0.013	0.02	-0.34	0.006
222-other	0.01	0.09	0.005	0.01	0.11	0.007
222-185	0.02	-0.68	0.048	0.02	-0.53	0.027
222-187	0.00	-0.91	0.041	0.00	-0.99	0.035
222-189	0.08	0.64	**0.205**	0.06	0.57	**0.144**
224-other	0.01	0.17	0.024	0.01	0.20	0.033
224-185	0.00	-0.95	0.047	0.00	-0.92	0.055
224-187	0.00	-0.99	0.024	0.00	-1.00	0.024
224-189	0.04	0.69	**0.119**	0.05	0.70	**0.145**

Continued on next page.

Table 5. Linkage disequilibrium between specific alleles (continued from previous page).

	ABCFR			OFBCR		
Loci/haplotype	Frequency	D'	R^2	Frequency	D'	R^2
5′ – 3′						
other-other	0.01	0.11	0.009	0.00	-1.00	0.003
other-185	0.01	-0.30	0.004	0.01	-0.57	0.017
other-187	0.02	0.20	0.005	0.03	0.42	0.026
other-189	0.00	-0.56	0.004	0.01	0.03	<0.001
18-other	0.00	-0.67	0.002	0.00	-1.00	0.003
18-185	0.04	0.65	0.026	0.03	0.42	0.008
18-187	0.01	-0.41	0.003	0.01	0.04	<0.001
18-189	0.00	-1.00	0.013	0.00	-1.00	0.010
19-other	0.04	-0.01	<0.001	0.05	0.66	0.014
19-185	0.30	0.04	0.001	0.31	0.00	<0.001
19-187	0.15	-0.19	0.025	0.14	-0.20	0.025
19-189	0.16	0.40	0.022	0.14	0.31	0.013
20-other	0.01	-0.39	0.003	0.00	-0.57	0.005
20-185	0.07	-0.24	0.013	0.09	-0.04	<0.001
20-187	0.10	0.25	0.044	0.08	0.16	0.017
20-189	0.03	-0.24	0.004	0.02	-0.36	0.007
21-other	0.01	0.07	0.003	0.00	-0.45	0.001
21-185	0.03	0.22	0.003	0.05	0.26	0.006
21-187	0.01	-0.40	0.003	0.01	-0.32	0.003
21-189	0.01	-0.34	0.002	0.01	-0.12	>0.001

Bold indicates linkage disequilibrium values that are ≥ 0.10.
† R^2 for 5′–intron 2 haplotype 22-214 (see Table 6) was 0.18.

Table 6. Association between *IGF1* haplotypes and breast cancer risk in the ABCFR and OFBCR .

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′ – intron 2												
19-216	24.2	366.9	1.71	0.09	23.9	225.6	0.31	0.75	25.0	138.5	2.22	0.03*
19-218	15.2	259.5	-0.86	0.39	14.8	166.9	-0.86	0.39	15.9	90.6	-0.17	0.87
20-216	9.0	159.9	0.51	0.61	9.1	94.8	0.32	0.75	8.5	62.0	0.49	0.63
19-222	8.2	165.0	-1.36	0.17	8.6	109.0	-0.91	0.36	7.5	51.9	-0.96	0.34
19-220	6.7	130.7	-0.07	0.95	6.5	78.0	-0.19	0.85	7.1	51.9	0.10	0.92
19-224	5.3	128.5	0.17	0.86	5.4	82.5	0.12	0.90	5.0	43.1	0.26	0.79
20-220	5.2	107.4	0.31	0.76	5.1	68.8	0.86	0.39	5.4	39.3	-0.84	0.40
20-218	3.9	93.2	2.15	0.03*	3.9	56.8	1.50	0.13	3.8	35.5	1.77	0.08
18-216	3.6	90.9	0.60	0.55	4.1	62.7	2.30	0.02*	3.0	28.0	-1.84	0.07
21-220	2.9	65.0	-2.24	0.03*	2.5	41.6	-1.62	0.11	3.4	20.1	-1.77	0.08
18-218	1.5	36.4	0.25	0.81	1.7	25.0	0.15	0.88	1.1	10.1	0.14	0.89
19-214	1.4	29.8	-0.63	0.53	1.2	19.1	-0.74	0.46	1.3	10.0	-0.11	0.91
20-222	1.2	29.5	0.71	0.48	1.3	23.6	0.83	0.41	-	<10	-	-
21-216	1.1	22.0	0.18	0.86	1.0	14.0	-0.55	0.58	-	<10	-	-
21-218	1.0	13.7	-1.25	0.21	0.9	10.5	0.03	0.98	-	<10	-	-
22-214	0.8	12.0	-0.49	0.63	-	<10	-	-	-	<10	-	-
19-204	0.8	20.4	1.13	0.26	0.7	13.1	0.66	0.51	-	<10	-	-
22-216	0.6	14.0	1.33	0.18	0.9	11.0	1.39	0.16	-	<10	-	-
17-220	0.6	12.9	0.61	0.54	-	<10	-	-	-	<10	-	-
19-212	0.6	13.3	-2.29	0.02*	0.8	12.4	-2.36	0.02*	-	<10	-	-
21-204	0.5	13.5	-0.04	0.97	-	<10	-	-	-	<10	-	-
17-218	0.5	12.9	-0.64	0.52	-	<10	-	-	-	<10	-	-

Continued on the following page.

Table 6. Association between *IGF1* haplotypes and breast cancer risk in the ABCFR and OFBCR (continued from previous page).

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
intron 2-3′												
216-187	18.0	305.2	1.15	0.25	18.2	185.4	0.54	0.59	17.6	117.8	1.15	0.25
216-185	17.3	264.9	0.79	0.43	17.6	166.0	0.96	0.34	16.9	97.9	0.44	0.66
218-185	16.8	284.2	1.31	0.19	16.4	180.0	0.83	0.41	17.7	100.5	1.33	0.18
220-185	9.3	173.3	-0.26	0.80	8.7	110.7	-0.15	0.88	10.0	57.8	-0.66	0.51
222-189	7.0	132.7	0.47	0.64	7.5	91.5	0.66	0.51	6.4	41.7	-0.10	0.92
224-189	4.1	95.7	-0.12	0.91	3.9	54.7	-0.02	0.99	4.4	41.0	-0.23	0.82
220-187	4.1	86.9	-0.81	0.42	4.2	54.2	-0.20	0.84	4.2	33.5	-0.90	0.37
216-189	3.0	69.4	1.66	0.10	3.0	41.4	0.90	0.37	2.7	26.1	1.17	0.24
218-187	2.7	62.7	-1.45	0.15	2.9	41.0	-1.14	0.26	2.2	20.6	-1.25	0.21
218-189	1.9	46.9	-1.31	0.19	2.1	28.2	-0.51	0.61	1.5	16.4	-1.66	0.10
204-185	1.8	47.4	-1.41	0.16	1.6	27.0	-0.41	0.68	2.1	21.0	-1.93	0.05*
220-189	1.7	31.0	-0.85	0.39	1.2	14.5	-0.21	0.84	2.3	14.9	-0.76	0.45
222-185	1.7	34.1	-0.98	0.33	1.7	23.0	-1.43	0.15	1.6	11.3	0.82	0.41
214-187	1.6	37.1	-0.74	0.46	1.7	26.0	-0.27	0.79	1.6	11.3	-0.50	0.61
224-191	1.2	24.7	-0.69	0.49	1.4	20.0	-0.14	0.89	-	<10	-	-
214-185	1.0	23.0	-0.29	0.77	0.8	13.0	0.08	0.94	-	<10	-	-
222-191	0.7	19.6	-1.71	0.09	0.7	10.9	-1.54	0.12	-	<10	-	-
216-183	0.4	12.0	0.45	0.66	-	<10	-	-	-	<10	-	-
218-181	0.3	10.8	0.95	0.34	-	<10	-	-	-	<10	-	-

NIF: number of informative families

* Significant at $P \leq 0.05$.

- Results not reported when number of informative families is less than 10.

IGF1 5′ polymorphism, allele length grouping and breast cancer risk

After inspection of the pattern of association between allelic variants of the 5′ polymorphism, alleles of this polymorphism were grouped into three categories according to allele length: 17 repeats and lower, 18 to 20 repeats, and 21 repeats and greater (Table 7). At the P=0.05 level, the category with intermediate allele lengths was positively associated with breast cancer risk (P=0.02), while longer alleles were inversely associated with risk (P=0.01). Similar P-values and the same direction of effect was observed in the OFBCR sample. Although there was qualitative agreement with this result in the ABFCR sample, results were not significant at the P≤0.05 level.

Table7. Association between 5′ allele groupings of *IGF1* and breast cancer risk in the ABCFR and OFBCR.

Study site/ allele grouping	Allele frequencies %	Number of informative families	Z	P-value
ABCFR and OFBCR				
<18	2.4	51	0.08	0.94
18-20	88.1	202	2.40	0.02*
>20	9.5	165	-2.74	0.01*
ABCFR				
<18	2.3	32	-0.11	0.91
18-20	88.6	131	1.14	0.25
>20	9.1	107	-1.30	0.20
OFBCR				
<18	2.5	19	0.07	0.94
18-20	87.3	71	2.50	0.01*
>20	10.1	58	-2.77	0.01*

* Significant at P≤0.05.

***IGF1*, family history of breast cancer and breast cancer risk**

Table 8 shows allele specific analyses for the 5′ intron 2 and 3′ polymorphism, stratified according to family history (one or more 1st degree relatives reported to have breast cancer). Consistent with other studies, the results from the combined ABCFR and OFBCR samples suggest a protective effect for the 5′ 19 allele in women with a family history of disease, although the results are not statistically significant (p=0.12). A significant inverse association between 5′ 19 alleles and risk was also observed among women with a family history of breast cancer at the P≤0.05 level in the ABCFR. This result was not, however, supported in the OFBCR sample. For the purpose of comparison to other studies that reported a protective effect of 5′ 19 homozygotes (*179,193*) in women with a family history of breast cancer, a recessive model was used to examine the relation of the this allele to breast cancer risk. Results were similar to those in Table 8, with P-values of 0.06 (combined ABCFR and OFBCR), 0.03 (ABCFR) and 0.37 (OFBCR) (*179,193*). A dominant model, which would be consistent with results describing a protective effect of the 19 allele in women with a family history in a third study (*190*), resulted in no nominally significant results in either the ABCFR, OFBCR or combined samples.

Among women with a family history of breast cancer, a nominally significant association was observed for the 5′ 20 allele in the combined sample. In both the ABCFR and OFBCR a positive relationship between the 20 allele and risk was observed, but neither was significant and the association was not particularly strong in the ABCFR.

Among families where there were no 1st degree relatives with breast cancer a nominally significant positive association with risk was observed for the 216 allele of the intron 2 polymorphism (P=0.05). Nominal significance was also observed for this allele in

the OBCFR. The direction of effect was consistent in the ABCFR but the result was not significant at the P≤0.05 level. This same allele was initially observed to be associated with breast cancer risk in the combined sample at the P≤0.05 level prior to stratification. There was no *a priori* hypothesis to suggest that women without 1st degree relatives should be at increased risk, and given the number of comparisons and the lack of a strong association in both the ABCFR and OFBCR a chance finding is possible.

Additional nominally significant associations in both the ABCFR and OFBCR were also observed in analyses stratified by registry (Table 8). Again, considering the number of comparisons made, and the lack of consistency of results across registries, these do not provide convincing evidence for an association of these alleles with disease.

Table 8. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR in families with and without a first degree relative with breast cancer - family history vs. no family history.

Loci/ family history/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′												
Family history												
18	5.0	14	1.40	0.16	-	<10	-	-	-	<10	-	-
19	61.9	65	-1.58	0.12	58.8	28	-2.50	0.01*	63.4	37	-0.06	0.95
20	21.5	50	1.99	0.05*	23.8	21	1.22	0.22	19.8	29	1.63	0.10
21	6.5	24	-0.97	0.33	6.2	12	0.55	0.58	7.1	12	-1.89	0.06
No family history												
11	0.5	10	0.24	0.81	-	<10	-	-	-	<10	-	-
17	1.6	27	-0.13	0.90	1.5	19	0.19	0.85	-	<10	-	-
18	6.5	115	0.27	0.79	6.8	81	1.47	0.14	6.1	34	-1.71	0.09
19	60.9	366	0.34	0.74	60.7	246	-0.89	0.37	61.3	120	1.97	0.05*
20	21.2	275	0.95	0.34	21.6	186	1.11	0.27	20.4	89	0.06	0.95
21	6.7	110	-1.86	0.06	6.6	75	-1.46	0.15	6.9	35	-1.37	0.17
22	1.9	40	-1.16	0.24	1.7	25	-0.49	0.62	2.5	15	-1.22	0.22

Continued on the following page.

Table 8. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR in families with and without a first degree relative with breast cancer - family history vs. no family history (continued from previous page).

Loci/ family history/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
Intron 2												
Family history												
214	3.5	11	-1.53	0.12	-	<10	-	-	-	<10	-	-
216	38.0	64	0.34	0.73	41.8	27	0.56	0.58	35.9	37	-0.17	0.86
218	24.2	62	0.56	0.57	24.2	27	0.31	0.75	24.1	35	0.55	0.59
220	15.7	44	0.73	0.47	13.7	15	1.47	0.14	16.7	29	-0.51	0.61
222	12.2	42	-1.75	0.08	11.7	19	-1.17	0.24	12.3	23	-1.07	0.29
224	3.7	18	1.62	0.11	-	<10	-	-	5.3	14	2.09	0.04*
Intron 2												
No family history												
204	2.5	46	-0.98	0.33	2.0	28	-0.21	0.84	3.4	18	-1.50	0.13
212	0.8	15	-1.65	0.10	1.1	14	-1.91	0.06	-	<10	-	-
214	2.9	50	-0.44	0.66	3.0	35	-0.03	0.98	2.8	15	-0.71	0.48
216	38.3	374	2.07	0.04*	37.9	239	1.24	0.22	39.3	135	1.98	0.05*
218	21.4	281	-0.14	0.89	20.9	191	-0.02	0.98	22.3	90	0.08	0.94
220	17.0	222	-1.34	0.18	17.2	155	-0.82	0.41	16.6	67	-1.45	0.15
222	9.2	157	0.05	0.96	10.2	115	-0.02	0.98	7.1	42	-0.32	0.75
224	6.7	126	-0.92	0.36	6.6	89	-0.20	0.84	6.8	37	-1.35	0.18
226	0.6	13	0.38	0.70	-	<10	-	-	-	<10	-	-

Continued on the following page.

Table 8. Association between alleles of *IGF1* polymorphisms and breast cancer risk in the ABCFR and OFBCR in families with and without a first degree relative with breast cancer - family history vs. no family history (continued from previous page).

Loci/ family history/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
3′												
Family history												
185	44.6	69	0.37	0.71	45.3	28	1.45	0.15	44.2	41	-0.85	0.40
187	30.0	69	-0.32	0.75	31.2	25	-1.72	0.08	29.0	44	0.93	0.35
189	20.1	51	0.04	0.97	17.8	20	0.15	0.88	21.6	31	-0.01	0.99
No family history												
181	0.9	19	-0.51	0.61	0.8	11	-0.02	0.99	-	<10	-	-
183	1.3	23	0.92	0.36	1.1	16	0.60	0.55	-	<10	-	-
185	47.5	370	-0.02	0.98	46.2	254	-0.41	0.69	50.1	116	0.64	0.52
187	28.2	315	-0.03	0.98	29.0	213	0.04	0.97	26.8	102	-0.01	1.00
189	18.1	264	0.16	0.88	18.4	176	0.29	0.78	17.3	88	-0.29	0.78
191	3.0	57	-0.79	0.43	3.4	44	-0.34	0.73	2.3	13	-1.38	0.17
193	1.0	15	-0.09	0.93	1.1	11	0.23	0.82	-	<10	-	-

NIF: number of informative families
* Significant at $P \leq 0.05$.
- Results not reported when number of informative families is less than 10.

IGF1, oral contraceptive use and breast cancer risk

Interactions between the 5′ 19 allele genotype and current, ever and duration of use of oral contraceptives were tested in relation to breast cancer risk. No significant interaction effects were observed for duration of use (P=0.21), current use (P=0.98) or ever use (P=0.57) for the 19 allele in the combined sample. In the ABCFR a borderline significant result was observed for duration of use (P=0.05) but not current (P=0.39) or ever use (P=0.27). Results in the OFBCR did not support the interaction with duration of oral contraceptive use seen in the ABCFR (P=0.38) and interaction effects for current use and (P=0.13) and ever use (P=0.43) were also not significant. Stratified analyses based on ever use of oral contraceptives were also examined for the purpose of comparing to previous research (*193*). There was no support for an association of the 19 allele with breast cancer risk among ever users of oral contraceptives (P=0.87). Among never users of oral contraceptives an additive model indicated a positive association between the 19 allele and risk, but this result was not statistically significant (P=0.07). In order to compare this association more closely to the previous study a dominant model was specified in analyses, resulting in a significant positive association with risk at the $P \leq 0.05$ level (P=0.04). This association was not consistent across samples (P=0.69 in the ABCFR and P=0.01 in the OFBCR). This comparison is based on small numbers (14 and 12 families in the ABCFR and OFBCR respectively) but it should also be considered that among all families in the OFBCR, a near nominally significant positive association (P=0.10) was observed. This would increase the probability of obtaining a nominally significant association due to chance in the OFBCR never user sub-group.

IGF1, BMI and breast cancer risk

No significant interaction effect was observed for the 19 allele of the 5′ polymorphism and BMI in relation to breast cancer risk in the combined sample (P=0.51) or in the ABCFR (P=0.16) or OFBCR (P=0.16). There was also no evidence for an interaction effect of the 3′ 185 allele and BMI (combined ABCFR and OFBCR: P=0.52; ABCFR P=0.68; OFBCR P=0.52).

DISCUSSION

The primary analysis examined the association of the three polymorphisms with breast cancer risk under an additive model. Some nominally significant results were observed (5′ 21 allele and intron 2 212 and 216 alleles) but adjustment for multiple comparisons indicated that these had a high probability of being false discoveries. None of the alleles where nominal significance was observed in the combined samples showed statistical significance in both populations. A recent large case-control study reported a significant association of several *IGF1* SNPs with breast cancer under a recessive model (*187*), and therefore a recessive model was also explored here. A stronger nominally significant association was observed for the combined sample for the intron 2 216 allele and in the OFBCR the association with this allele reached nominal significance. However, the observed association in the ABCFR, although in the same direction, did not provide strong support for the results in the OFBCR. Given the lack of clear support in the ABCFR and the many comparisons made, the observed association of the intron 2 216 allele with risk is not convincing. Analyses that examined the interaction between *IGF1* genotype and either registry (ABCFR and OFBCR) or age, or analyses stratified by registry, provided little evidence to suggest that effect modification by either of these variables should influence the

interpretation of results. Therefore, the results provide limited support for an association of specific alleles of these three polymorphisms with breast cancer risk.

Analyses that examined the association of haplotypes comprised of the 5′ and intron 2 polymorphism or the intron 2 and 3′ polymorphism with breast cancer risk, did not alter this interpretation. A strong association with risk was observed for the 5′ 19 intron 2 216 haplotype in the OFBCR under a recessive model, but results from the ABCFR did not support an association.

Grouping the alleles of the 5′ polymorphism into lengths of less than 18, 18 to 20 and greater than 20 repeats resulted in a significant positive association with risk for alleles of length 18 to 20 and a significant protective effect for alleles greater than 20 repeats in length. Caution should be used in interpreting this result as these were *ad hoc* groupings and although consistent with the direction of effect, results from the ABCFR did not approach statistical significance. In addition, there is no experimental evidence to suggest that this polymorphism is functionally relevant and therefore that allele length has biological meaning, although another explanation is that linkage disequilibrium patterns could explain a result such as the one observed here.

A number of studies have examined the association of the 5′ polymorphism with breast cancer risk (*172,179,189-193*). The association of the 19 allele with disease has been most often investigated with few studies finding significant associations. The 19 allele has been reported to be positively associated with breast cancer risk in a largely (60%) African American population (Louisiana) (*192*) and in a Chinese population from Shanghai (*189*), the latter reporting a strong positive association with risk in premenopausal women. Other studies, performed in Caucasian or largely Caucasian populations do not support an

association of breast cancer risk with this allele. In this study weak associations were observed in the ABCFR and OFBCR (at the P≤0.10 level), but these were in opposite directions. The results here appear to confirm a lack of association between this allele and breast cancer risk in Caucasian populations. The associations observed in the other populations are difficult to explain. Ethnicity may modify the association between this allele and breast cancer risk because of different patterns of linkage disequilibrium. Other explanations for these inconsistencies are the potential for confounding due to population stratification, possible effect modification by other genetic or environmental factors, or they may simply be spurious associations.

The results of this study did not support the inverse association observed between the 17 allele and breast cancer risk in the Shanghai population (*189*). There was, however, limited power in the current study to detect an association for this rare allele. The association with longer alleles reported here (21 repeats and greater) was reported previously in a small hospital based study conducted in Israel (*191*). In addition, the observed allele frequencies among cases and controls from a study conducted in Long Island (U.S.A), provided qualitative support for a protective effect of alleles of 21 repeats or greater with risk in premenopausal women. However, they did not support the association of alleles of intermediate allele length observed here (*193*). A positive association with alleles shorter than 19 repeats with risk in premenopausal women was also reported in the Long Island study. Inspection of the results presented here for specific alleles shorter than 19 repeats, did not support this relationship. Other studies examining pre- and postmenopausal women have presented sufficient data to at least allow for some comparison of results of the association of non 19 alleles (either in allele length groupings or separately) with breast cancer risk. These

provide little or no support for the pattern of association observed here, where intermediate length alleles were associated with increased risk and longer alleles associated with decreased risk (*179,189,192*).

The limited evidence for an association among these three polymorphisms or their haplotypes with breast cancer risk is in general agreement with two studies reporting primarily on postmenopausal women that have examined several SNPs in the *IGF1* gene in relation to breast cancer risk (*194,195*). Neither found evidence that genetic variation at *IGF1* was strongly related to risk. Both studies, however, reported nominally significant results for different pairs of 5′ SNPs. Given the results of these two studies, the evidence presented here, and the reports (although inconsistent) of associations observed between alleles of the 5′ polymorphism and breast cancer risk, an association with risk of some functional variant that is at the 5′ end of the gene or at least in linkage disequilibrium with 5′ *IGF1* polymorphisms, cannot be ruled out.

Another study conducted in Great Britain reported several SNPs along *IGF1* to be associated with disease risk under a recessive model (*187*). The interpretation of the results of this particular study are complicated, however, by the inclusion of prevalent cases. The results presented here did provide some support for an association under a recessive model for the intron 2 216 allele, although nominally significant associations were observed overall and in the OFBCR, but not in the ABCFR. It is not known if the intron 2 216 allele is in linkage disequilibrium with the SNPs examined in the British study, which were mainly in the vicinity of the 3′ polymorphism investigated here. As well, results from the British study are inconsistent with that of a previous study that examined one of the same tagging SNPs

(*194*). Still these results indicate that the association between breast cancer and genetic variants of *IGF1* that lie outside the 5′ region is also possible.

Genome wide association studies have not reported significant associations with *IGF1* (*38,39,197*). However, true associations could have been missed particularly in premenopausal women, a group in which large samples have not been investigated during initial screens. As well, stringent methods employed in adjusting for multiple comparisons may have resulted in missed associations.

Stratification of this sample into women with and without a family history of breast cancer suggested a slight protective effect of the 5′ 19 allele in women with a family history of disease. Results were not statistically significant in the combined ABCFR OFBCR sample, although a significant protective effect was observed in the ABCFR. This result is of some interest because it is consistent with three other studies which reported a non-significant protective effect of this allele among women with a family history of breast cancer (*179,190,193*). Two of these used the same definition of family history as this study (one or more first degree relatives affected with breast cancer) (*179,193*). These results combined with the non-significant inverse association reported here cannot, however, be considered convincing evidence for a protective effect of the 19 allele in women with a family history of disease. In two of three studies, the null value of one was well within the range of the confidence limits (OR=0.49, 95% CI=0.20-1.23 for 19 allele homozygotes vs. all others (*179*); OR=2.70, 95% CI=0.70-10.4 for 19 allele non-carriers vs. all others (*193*); OR = 1.51, 95% CI =0.96-2.39 for 19 allele non-carriers vs. all others (*190*)). In addition, there is some inconsistency among the three studies as two based their reports on samples of older women (mean age = 47, menopausal status not provided) or mainly postmenopausal women

(*179,190*), while the third found support for an association only among premenopausal women (OR=2.70, 95% CI=0.70-10.4 in premenopausal non-carriers of the 19 allele verses OR=1.11 95% CI 0.56-2.18 in postmenopausal non carriers of the 19 allele women) (*193*). Finally, a statistically significant association at the $P \leq 0.05$ level has only been observed in this study and only among ABCFR families, while in the OFBCR there was no evidence for an association.

A previous study reported that among current oral contraceptive users, circulating IGF-I concentrations were lower for carriers of the 19 repeat allele relative to non-carriers (*174*). In general, oral contraceptive use has been found to lower IGF-I levels (*115-117*) possibly due to a hepatic first pass effect of oral estrogen intake on IGF-I production by the liver (*116*). Although this would appear to indicate that oral contraceptive use should be associated with a reduced risk of breast cancer, it has been suggested that increased risk might be restricted to women who are genetically predisposed because of *IGF1* genotype (*174*). One previous study has investigated whether the association of the 19 repeat allele with breast cancer risk was modified by oral contraceptive use in premenopausal women. Ever use of oral contraceptives resulted in a modest but non-significant increase in breast cancer risk among non-19 alleles carriers, and a significant protective effect among non-carriers of the 19 allele among never users (*193*). In the current investigation, tests of interaction between the number of 19 alleles and current, duration or ever use of oral contraceptives did not indicate that oral contraceptive use modifies the relationship of this allele with breast cancer risk. Stratified analyses also did not support an association with greater risk among oral contraceptive users under an additive model. Support for a positive association of the 19 allele with risk under a dominant model among never users (the equivalent of a protective

effect among non-carriers as was reported previously (*193*)) may have been an artefact of choosing a sub-group from the overall OFBCR sample where the 19 allele already showed a weak positive association with risk.

Unlike some previous studies, the results presented here did not indicate that body mass index may modify the relationship of the 19 allele with breast cancer risk (*189,193*). There was also no evidence that BMI modified the association of the 18 allele of the 3′ polymorphism (which was found to be associated with BMI in chapter 4) with breast cancer risk.

There are several potential limitations to this study. In the OFBCR the selection of probands was weighted towards women with a family history of breast cancer, with 75% of probands classified as having a family history of disease, according to OFBCR criteria. Therefore, although population based, the study sample was not representative of the population that they were sampled from. Genetic effects, however, may be more likely to be observed among cases with a family history of disease, because they may be more likely to have variant alleles at gene loci that predispose to disease.

The sample of premenopausal probands here is younger than might be expected in a typical population based sample of premenopausal breast cancers cases, as prior to 1996 the ABCFR deliberately sampled women under 40. Therefore, effects that might be either restricted to, or largely observed in older premenopausal women, could be missed here. Although the possibility that an association of genetic variants of *IGF1* and breast cancer risk among older premenopausal women cannot be ruled out, analyses where the interaction of genotype with age was investigated found one significant result for the rare intron 2 212 allele, which could have been due to chance. Therefore, there is little evidence to suggest that

the age of the premenopausal women studied should have an important influence on outcome.

Although haplotypes were used to examine genetic variation of the *IGF1* gene and breast cancer risk, the density of markers may not have been great enough to effectively capture genetic variation across *IGF1*. Although there was significant linkage disequilibrium between polymorphisms, allele specific R^2 values never exceeded 0.21, which are indicative of weak linkage disequilibrium. Further investigation using haplotypes or tagging SNPs to assess genetic variation at the *IGF1* gene in relation to breast cancer risk is still desirable, particularly in premenopausal women, since it is this group for which the evidence suggests there is an association of circulating IGF-I concentration and breast cancer risk (*1-3*).

A major strength of this study is the family based design, which eliminates the possibility of confounding resulting from population stratification. In addition, a relatively large sample of premenopausal women was examined, a group for which the sample size in other studies is often fairly small. However, family based analyses such as the one conducted here, are also susceptible to bias. In the presence of genotyping errors, false positive associations are increased for both common alleles and rare alleles, with common alleles incorrectly associated with increased risk and rare alleles with decreased risk (*243*). However, given the efforts made to correct for multiple comparisons and the pattern of significant results, bias due to genotyping error should have little influence on interpretation of results in relation to breast cancer risk.

In conclusion, this study provides limited support for an association between genetic variants of the *IGF1* gene and breast cancer risk. Grouping alleles suggested a protective effect for longer alleles of the 5′ polymorphism and increased risk for alleles of intermediate

length, but chance findings based on multiple comparisons may have led to this association. Although other studies have not used this specific classification scheme, those that have used at least comparable groupings or presented genotype data that permits comparison, in general provide little support for this pattern of association with allele length. However, recent studies focusing on SNPs have provided some evidence for an association with the polymorphisms at the 5′ end of the gene. As well, showing consistency with the nominally significant association observed for the intron 2 216 allele, a recent study has reported a significant association with several *IGF1* tagging SNPs with risk under a recessive model. Therefore, further investigation of an association of *IGF1* variants with breast cancer risk is still warranted. These should, however, include dense SNP maps and functional variants (if detected) and compare results in pre-and postmenopausal women. In addition, future studies investigating either the 5′ repeat polymorphism or *IGF1* SNPs should include stratification by family history, as the inconsistent findings of an association of the 19 allele of the 5′ polymorphism with breast cancer risk in women with a family history of disease requires clarification.

Chapter 6. Family based study examining the association of *IGF1* polymorphisms with body mass index, weight and height

ABSTRACT

Background and objectives

A number of studies indicate that greater circulating concentrations of IGF-I increase risk of breast cancer in premenopausal women. Several anthropometric measures (body mass index (BMI), weight, and height) that appear to be linked to circulating IGF-I levels, have also been associated with breast cancer risk. The objective of this study was to examine the association of genetic variation at *IGF1*, with BMI, weight and height in premenopausal women.

Methods

A family based design was used to investigate the association between allelic variants of three *IGF1* repeat polymorphisms, including a previously investigated promoter region (5′) repeat, with these anthropometric measures. DNA specimens from 827 families, (2,569 subjects including 1,520 offspring with anthropometric measures), were obtained from the Ontario Familial Breast Cancer Registry and the Australian Breast Cancer Family Registry.

Results

Significant associations at the $P \leq 0.05$ level were observed for a rare allele (22) of the 3′ polymorphism and BMI (P=0.05), and the more common 3′ 19 allele and weight (P=0.05). However, given the number of comparisons, these few significant results could have been due to chance. Analysis of haplotypes also resulted in few nominally significant associations that could have been due to chance.

Conclusions

This study provides little support for an association between genetic variants of *IGF1* with BMI, weight, or height in premenopausal women.

INTRODUCTION

A number of cohort and case-control studies have investigated the association of circulating IGF-I concentration and breast cancer risk. Although not entirely consistent, these studies indicate an increased risk of breast cancer is associated with greater circulating concentrations of IGF-I in premenopausal women, but not in postmenopausal women (*1-3*).

Several anthropometric measures that appear to be linked to circulating IGF-I levels have also been associated with breast cancer risk. Greater adult height is reported to be associated with increased risk (*26*) and there is evidence to indicate that IGF-I levels are related to height in childhood (*148,149*). Greater BMI is associated with increased risk in postmenopausal women and appears to be associated with decreased risk in premenopausal women (*16,244*), (although the latter association may be explained by confounding by mammographic density, which is inversely associated with BMI, but positively associated with breast cancer risk (*83*)). Results of some studies suggest a non-linear association between IGF-I concentration and body mass index (BMI) (*130-135*), and a consistent inverse relationship between IGFBP-I levels and BMI provides support for a relationship between the IGF-I system and body composition (*125,129,132,133,140-143*).

Evidence from studies on middle aged and elderly male and female twins suggests that circulating IGF-I levels are partly heritable with estimates ranging from 38-63%. Twin studies also indicate a substantial heritable component for both height and BMI. Heritability

estimates range from about 70-90% for height (*92,245-247*). Estimates for BMI are more variable, but range from about 50-75% in larger studies (*92,245-247*).

Only a few studies have examined whether common variants of the *IGF1* gene influence height, or obesity related traits such as BMI and weight. Most have focused on the association of variant alleles of a 5′ (CA) cytosine-adenosine dinucleotide repeat polymorphism with one or more of these measures, but consistent results have not been observed (*179,186,189 ,248,249*). In chapter 4, results from a sample of unrelated premenopausal women indicated that having fewer copies of the most common allele (185) of a 3′ CA repeat was associated with increased BMI. Greater height was associated with having more copies of this same allele.

The purpose of this study was to investigate the association of genetic variants of the *IGF1* gene with BMI, height and weight in premenopausal women, using families from the Australian Breast Cancer Family Registry and the Ontario Family Breast Cancer Registry. The 5′ and 3′ repeat polymorphisms described above, and an intron 2 CA repeat were investigated in association with these anthropometric measures. Of particular interest is the relationship between genetic variants of the 3′ polymorphism with BMI, because of the association reported in chapter 4.

METHODS

This study used a family based design. Families had been previously recruited into the Breast Cancer Family Registry for the purpose of conducting studies on the genetic and molecular epidemiology of breast cancer (*239*). Families with affected premenopausal probands with either siblings (including some affected sisters), parents or combinations of siblings and parents were selected from the registry. Anthropometric variables for

premenopausal women (with or without a diagnosis of breast cancer) were used for the current investigation. The methods for this study are for the most part identical to those in chapter 5. They are described here, but in less detail.

Recruitment

Two sites of the Breast Cancer Family Registry, the Ontario Familial Breast Cancer Registry (OFBCR) and the Australian Breast Cancer Family Registry (ABCFR) provided DNA specimens and risk factor data for study subjects. During the initial recruitment, cases were identified using local population based cancer registries: the Ontario Cancer Registry in Ontario, Canada, and the Victorian and New South Wales Registries in Australia. Probands were either residents of Ontario, Melbourne or Sydney at the time of diagnosis.

Probands in OFBCR families were incident primary invasive breast cancer cases identified by pathology report beginning in 1996 and ending in 1998. Initial enrolment included female cases aged 20-54, a random sample of female cases aged 55-69, and male cases under age 80. A family history questionnaire was used to classify families as moderate to high risk for developing breast cancer. All moderate to high risk families were eligible for the study. A random sample, approximately 25%, of families designated low risk for breast cancer was also targeted for recruitment (*240*).

Figure 1 shows steps in the recruitment of OFBCR families. Initially, 8446 cases were eligible for contact in Ontario for the purpose of classifying family history status. Physicians refused to give permission to contact 5.6% of cases. In addition 1.7% of cases were already deceased and an additional 2.1% of cases could not be contacted through physician records. Of the remaining 7,662 cases, 14.8% refused to participate and 20.4% either did not respond or could not be contacted. A completed family history questionnaire was returned by 4,962

cases (59% of the eligible 8,446 cases). After assessment of family history status, just over half of the cases, 2,585 (52%), met eligibility criteria. A risk factor questionnaire was completed by 1,862 (72%) of these. Blood samples from a female proband and at least one first degree relative was obtained for 606 cases who completed the risk factor questionnaire. Probands and sisters with a diagnosis of breast cancer were considered premenopausal if their periods had not stopped for more than one year at the time of their diagnosis and they were either not using hormone replacement therapy, or stopped using hormone replacement therapy two or more years prior to diagnosis, or a year prior to their periods ending after diagnosis. All premenopausal breast cancer diagnoses occurred in women 56 years of age or under. Of the 606 available probands, 400 were determined to be premenopausal at time of diagnosis potentially providing 400 families for analysis (Fig. 1). However, in 62 of these families, DNA from the Mount Sinai repository was unavailable for either the proband or relatives, leaving 338 families with DNA samples that potentially could be analysed. DNA of six probands was never tested because other ongoing studies indicated that these were of insufficient quality for analysis. This left 332 OFBCR families available for genotyping for this study.

Enrolment into the ABCFR required that cases were incident first primary breast cancers aged 20-59 and diagnosed from 1992 to 2000 (Fig. 2). Enrolment from 1992 to 1995 pre-dated the formation of the Breast Cancer Family Registry and was restricted to women under the age of 40. Family history status was not a factor in determining eligibility for the ABCFR. Interviews were conducted for 1578 of 2303 eligible cases (68.5%). The patient's surgeon refused to allow contact for 8.5% of eligible cases. Refusal by the case to participate (16.4%), death prior to contact (1.8%), non-response by surgeon (1.3%) or case (1.2%), and

failure to locate the case (2.3%), accounted for the remaining 23% that did not enroll. DNA samples were available from 1453 cases (92% of those who completed an interview), 1048 of whom were premenopausal. Of these, 564 (54%) had at least one first degree relative with a DNA sample.

The same definition for premenopausal status for probands used in the OFBCR, was also employed for ABCFR probands recruited beginning in 1996. Prior to 1996, probands were considered premenopausal if their periods had not stopped for more than one year at the time of their diagnosis or interview, or if they indicated their periods had stopped due to pregnancy. One individual who indicated her periods had stopped due to pregnancy, but also indicated they had stopped more than twenty years previous, was considered to be postmenopausal and not included.

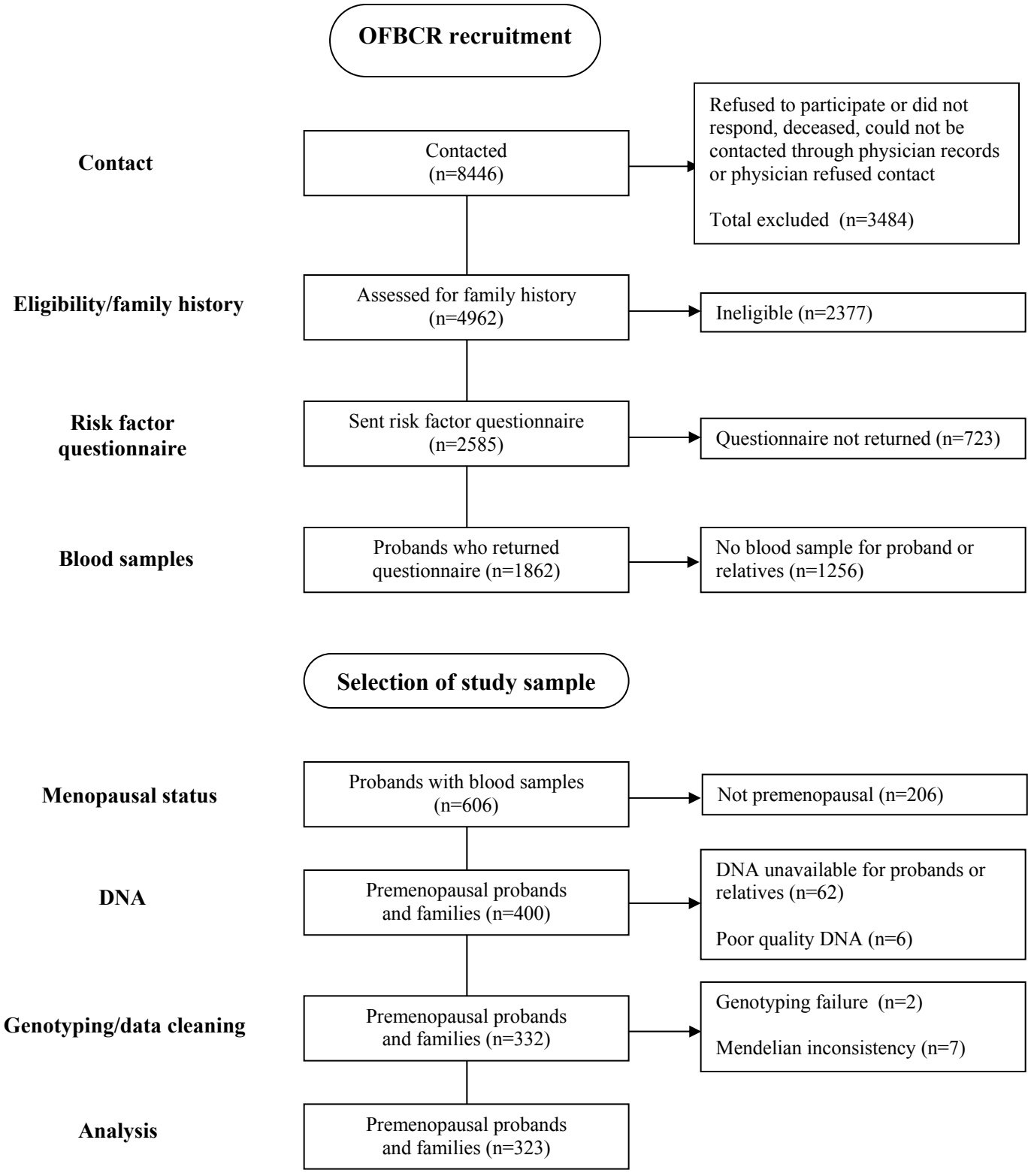

Figure 1. Recruitment of OFBCR study subjects and selection of OFBCR families for this study.

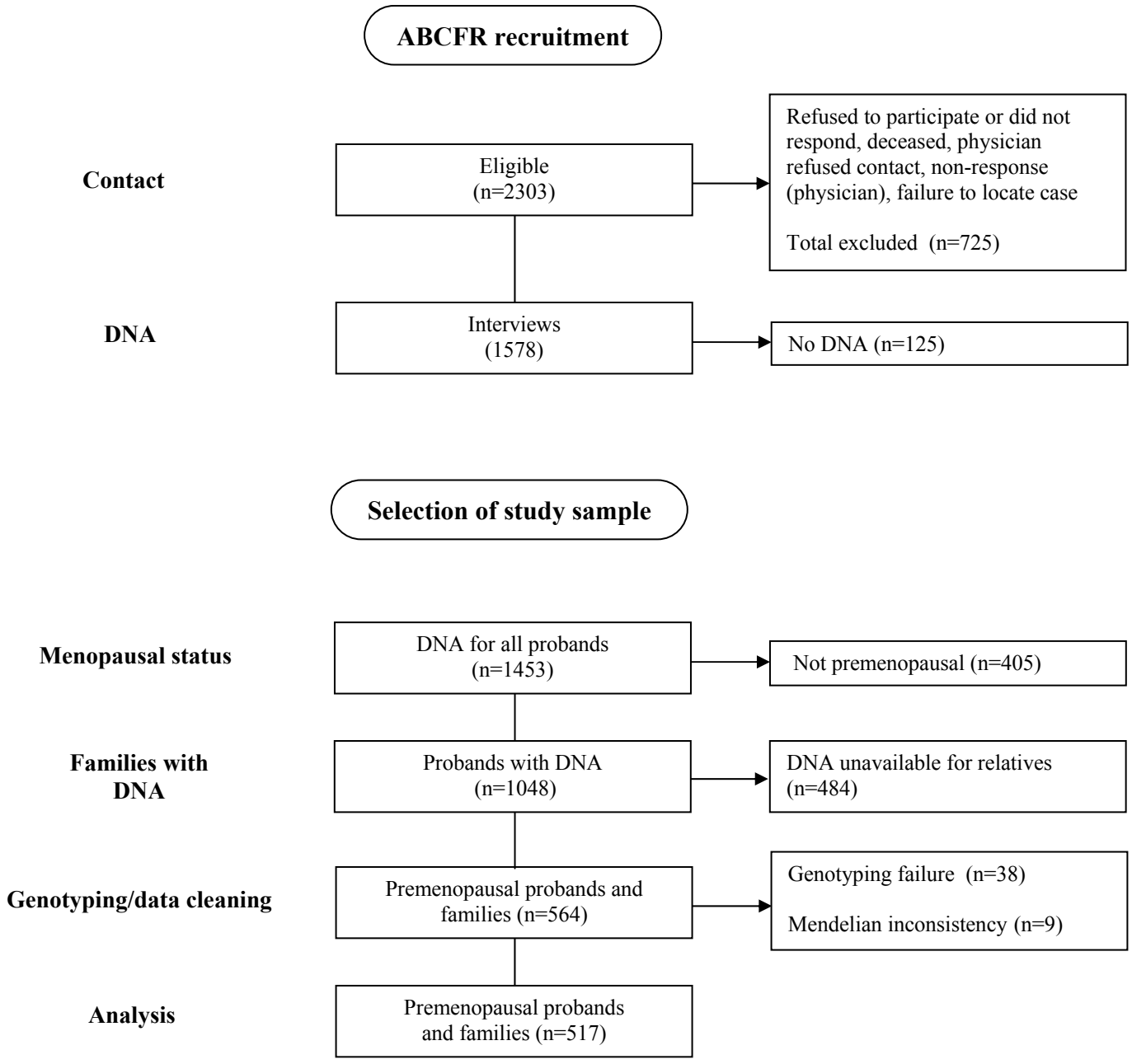

Figure 2. Recruitment of ABCFR study subjects and selection of ABCFR families for this study.

Measurements

Laboratory methods

DNA extraction

DNA samples were obtained from the Ontario Cancer Genetics Network repository in Ontario and from the ABCFR in Melbourne. In Ontario, peripheral blood lymphocytes were pelleted and stored at –70ºC to –80ºC and DNA was later extracted from the pellets using either phenol/chloroform extraction or the Qiagen procedure (QIAamp DNA Mini Kit (250)™). The ABCFR stored dried blood samples on Guthrie cards. DNA was extracted from Guthrie card dried blood spots using a Chelex method. All samples were sent to the laboratory of Dr. H. Ozcelik, where they were prepared for genotyping at The Centre for Applied Genomics (TCAG) (*216*).

Genotyping

Three polymorphisms were selected for analysis: the previously identified 5′ CA repeat polymorphism (*175,228*), and two other CA repeats, located in intron 2 and at the 3′ end of the gene that were identified using Tandem Repeat Finder software version 2.02 (*229*). The oligonucleotide primers used for PCR were as follows: 5′-GCTAGCCAGCTGGTGTTATT-3′, 5′-ACCACTCTGGGAGAAGGGTA-3′ for the 5′ polymorphism, 5′ CATACTTCTTAGCTCCTCAGG-3′, 5′-CCCTCACAGAAAGCAGAA-3′ for the intron 2 polymorphism, and 5′-CTTTTTAAGATGAGGCAGTTCC-3′, 5′GATTTCTTTTCAGTATTCCATTGG-3′ for the 3′ polymorphism. Position of amplified regions of DNA corresponded to positions 26,357,243-26,357,436, 26,332,059-26,332,274

and 26,275,038-26,275,220 on contig NT_019546.15 of build 35.1 of the NCBI's genome annotation for the 5′, intron 2 and 3′ polymorphisms respectively.

For the OFBCR sample, PCR was performed with 5ng of DNA, 5µl 10X PCR buffer, 0.8mM dNTPs, 0.18mM of each primer, and 5 U (0.3µl and 0.2µl) Platinum Taq in each 50ul reaction. $MgCl_2$ concentration was either 3mM for the 5′ and 3′ polymorphisms and 4mM for intron 2, or 2mM for all polymorphisms depending on $MgCl_2$ stock used. Amplification conditions were 94ºC for 3 minutes, and 35 cycles of 94ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 1 minute for the 5′ polymorphism. Amplification conditions were the same for the other polymorphisms with the exception that annealing temperatures were 60ºC or 55ºC for the intron 2 polymorphism (depending on $MgCl_2$ stock used) and 50ºC for the 3′ polymorphism. For ABCFR samples, PCR was performed with either 6mM or 4mM of magnesium depending on $MgCl_2$ stock used. Amplification conditions were identical to that used for the OFBCR sample with the exception that annealing temperature was either 55ºC (all 3 polymorphisms) or 55ºC for the 5′ and intron 2 polymorphism and 50ºC for the 3′ polymorphism, depending on $MgCl_2$ stock used. All reagents were supplied by Invitrogen Life Technologies. The forward primer was labeled with a 5′ fluorescent dye (HEX) and the reverse primer with a 5′ sequence (GTTTCTT). Genotyping was performed by TCAG using performance Optimized Polymer 6 (POP6) gel capillary electrophoresis systems (ABI 3100 and ABI 3700 systems) (*216*). Computerized output used by TCAG to assign genotypes (electropherograms) was also examined by G. Fehringer to identify allele calls that were suspicious and needed to be repeated.

Retesting of Samples

A small percentage of samples were retested because they failed (8% in the ABCFR and 5% in the OFBCR) or resulted in irregular shaped electropherograms (3% in both the ABCFR and OFBCR) on the first genotyping attempt. OFBCR samples were retested up to two more times if at least one of three polymorphic loci originally tested for the same subject had been successfully genotyped. In cases where no genotype data was returned for all three loci for a particular subject, only one additional retest per polymorphism was performed. ABCFR samples were only retested once as only a limited supply of DNA was available.

Reliability of genotyping

Reliability of TCAG allele calls was estimated by retesting a randomly selected sample representing a minimum of 10% of DNA specimens from each 96 well plate. Laboratory staff at TCAG were not informed which plates included samples for reliability tests.

Concordance for the ABCFR sample was 97.2% (n=179), 94.1% (n=186) and 97.2% (n=178) for the 5′, intron 2 and 3′ polymorphisms respectively. Concordance for the OFBCR sample was 99.1%, 98.1%, and 100% for the 5′, intron 2 and 3′ polymorphisms respectively (n=108 for each polymorphism).

Assessment of Centre for Applied Genomics allele calls

Accuracy of TCAG allele calls was assessed by randomly selecting electropherograms representing at least 10% of allele calls for each 96 well plate and having readers assign genotype independently of TCAG. The ABCFR sample was inspected by two readers (G. Fehringer and H. Jaranazi) independently. Discrepancies were resolved and results compared to the TCAG allele calls. The OFBCR peaks were inspected by G. Fehringer alone.

Agreement for the ABCFR sample was 98.4%, 98.9% and 100% for the 5′, intron 2 and 3′ polymorphisms respectively (n=190, 186 and 186 for each polymorphism respectively). Agreement for the OFBCR sample was 99.1% for all three polymorphisms (n=109, 108 and 107 for the 5′, intron 2 and 3′ polymorphisms respectively).

Mendelian inconsistency

Pedigree errors were examined using the program PEDAGREE (*241*). The extent of the Mendelian errors within a family determined whether the entire family was excluded from analyses or whether genotype data for a specific individual was removed for one or more polymorphisms. Mendelian errors were identified in 15 of 330 OFBCR families resulting in the removal of 7 of these prior to analysis (Fig. 1). In the ABCFR sample, Mendelian errors were detected in 43 of 526 families resulting in the removal of 9 families (Fig. 2).

Genotyping success rate

After removal of families with Mendelian inconsistencies, there were 325 families available for analysis in Ontario. Genotyping success rate was 99.7%, 99.3% and 99.1% for the 5′, intron 2 and 3′ polymorphisms respectively. Failures in genotyping resulted in two families being removed from the sample leaving 323 families for the analysis (Fig. 1).

There were 555 families available for genotyping from the ABCFR sample, after removal of families with Mendelian errors. Genotyping success rate for the ABCFR samples was 90.1%, 90.3% and 90.0% for the 5′, intron 2 and 3′ polymorphisms respectively. Failures in genotyping resulted in 38 families being discarded, leaving 517 families for analysis (Fig. 2).

Risk factor questionnaires

Extensive risk factor questionnaires were completed by both ABCFR and OFBCR family members. These included questions about weight, height, age and reproductive history. ABCFR questionnaires were completed through interview. OFBCR questionnaires were self administered.

Analysis

Hardy-Weinberg equilibrium for each polymorphism was examined using a permutation version of the exact test (*230*) (implemented in Power Marker 3.25 (*231*)).

Family based association tests incorporated into the programs FBAT (Family Based Association Testing, version 1.5) (*219*) and PBAT (Pedigree Based Association Testing, version 2.5) (*207*) were used for main analyses. These programs test for association under a null hypothesis of no linkage or linkage disequilibrium between the marker(s) and the locus associated with outcome. The expected distribution is based on parental genotype data (using genotype data directly from both parents or reconstructed from a genotyped parent and/or siblings). Given Mendelian inheritance patterns all alleles are equally likely to be transmitted to all siblings. Deviation from this expected distribution dependent on outcome is evidence for association. Since the expected distribution is based on the parental genotype data, the resulting statistical test is not susceptible to confounding due to population stratification. Furthermore, no assumptions need to be made regarding assortative mating (deviation from Hardy-Weinberg equilibrium) and the ascertainment condition (*219*). Additive models were used for all analyses as simulations have shown that the additive model has good power to detect an association, even when the true genetic model is not an additive one (*223*).

Inclusion/exclusion criteria

Height, weight and BMI were included for all premenopausal probands and sisters that had not had a previous diagnosis of breast cancer (malignant or *in situ*) and all premenopausal sisters without breast cancer. Premenopausal status for sisters without breast cancer was defined in the same way as premenopausal status of probands, although age at interview was used as a reference instead of age at diagnosis. For probands, a previous diagnosis of breast cancer (which only occurred in the OFBCR) was defined as one that preceded the diagnosis for which they were recruited. The diagnosis for which probands were recruited, and the subsequent treatment of breast cancer should not directly influence reported weight and the subsequent calculation of BMI, since the questionnaire asked for weight one year prior to the diagnosis for which probands had been recruited. However, diagnosis of breast cancer may influence recall. (Potential error resulting from including probands in the analysis is discussed below). Prior diagnosis of cancers other than breast cancers were not used as an exclusion criteria. There were 30 previous reported cancers identified in 28 individuals (7 cervical cancers, 5 thyroid cancers, 5 malignant skin cancers of uncertain type, 4 non-melanoma skin cancers, 3 malignant melanomas of the skin, 4 lymphomas, 1 colon cancer, 1 cancer of the anus and anal canal). In addition there were 3 other individuals with cancers with unspecified diagnosis date (1 colon cancer, 1 malignant melanoma of the skin, and 1 skin cancer). Weight and BMI are only known to be risk factors for one of these cancers (colon), and in 17 of these individuals cancer diagnoses preceded the reference date by at least three years and therefore treatment was less likely to have an effect on weight and BMI.

Analysis details

FBAT was used to analyze the association between allelic variants at individual loci and outcome, using an offset that minimizes variance. These analyses were repeated with PBAT using an offset choice of phenotypic residuals based on the mean model with outcome variables transformed to Z scores. Analyses in PBAT were also performed with age (all outcomes), current pregnancy, current or full term pregnancy in the last two years, number of full term pregnancies, or age with any one of the pregnancy variables included as covariates (BMI and weight only), using a multivariate extension of family based association tests based on generalized estimating equations incorporated into PBAT (*250*). An analysis restricted to women who never had breast cancer was also performed (BMI and weight only). Neither analyses with PBAT, with or without covariates, nor analyses restricted to non-breast cancer cases, altered the results appreciably. Therefore, results from analyses with FBAT (without the inclusion of covariates and including premenopausal women with and without breast cancer) are presented, since this program provides an indicator of the magnitude and direction of an effect.

Interactions between genotype and registry in association with outcome were examined using PBAT in order to assess the appropriateness of combining samples. In assessing the association of variant alleles with the investigated outcomes, results were examined for both the combined ABCFR and OFBCR sample as well as in the two samples separately. Results between samples were compared, with consistency (e.g. significance in both samples) being one criteria by which an association may indicate a true effect. In addition, the Benjamini-Hochberg correction was used to adjust P-values to account for multiple comparisons for

each outcome (*224*) in the combined sample. No correction was made for the testing of multiple outcomes.

Linkage disequilibrium was assessed between loci using PowerMarker vers. 3.25 (*231*). An overall test was performed and linkage disequilibrium between alleles from different loci was also examined. Haplotype analyses were performed using two loci when linkage disequilibrium between loci was statistically significant and at least one R^2 value between allele pairs in either the ABCFR or OFBCR was greater than 0.1. Analysis of the association of haplotypes with outcome used a weighted conditional approach incorporated into the program FBAT 1.5 (*242*).

Ethics

Ethics approval for this study was received from the Mt. Sinai Hospital Research Ethics Board in Toronto.

RESULTS

Characteristics of subjects and completeness of data

The women in the ABCFR sample were somewhat younger than those in the OFBCR reflecting the restriction of enrollment of women under 40 from 1992 to 1995 in the ABCFR sample. Women in the ABCFR had lower mean body weight and BMI. In 93% of ABCFR families and 95% of OFBCR families, all family members who reported their race or ethnicity identified themselves as white (Table 1). Defining family history as having a mother or sister with breast cancer, 11% of ABCFR probands and 27% of OFBCR probands had a family history of breast cancer (Table 3). The ABCFR sample had fewer one parent one proband families, as these were not requested from the ABCFR since they add little power to the analysis. The few that are present are the result of a failure to successfully

genotype other family members. The OFBCR had relatively few samples with siblings and both parents genotyped. For some families siblings were not genotyped when DNA was available for both parents. Adding siblings does not add information when parental genotype is already known and therefore does not improve study power. Since genotype was missing for a number of individuals, the final number of families used in analyses varied for each polymorphism. In the ABCFR, the final number of families were 504 for the 5′ and 3′ polymorphisms for all three outcomes and 501, 501 and 502 for BMI, weight and height respectively for the intron 2 polymorphism. In the OFBCR final number of families for the 5′, intron 2 and 3′ polymorphisms respectively were 295, 293 and 293 for BMI, 300, 298, and 298 for weight and 303, 302 and 302 for height.

Table 1. Descriptive characteristics of ABCFR and OFBCR samples

	ABCFR	OFBCR
Number of Families	517	310
Age (S.D.)[a]	37.4 (6.6)	41.9 (6.5)
Height, cm (S.D.)[b]	163.8 (7.3)	164.3 (6.7)
Weight, kg (S.D.)[c]	65.0 (13.8)	68.1 (14.1)
BMI, kg/m^2 (S.D.)[d]	24.2 (4.8)	25.3 (5.1)
Caucasian families	482 (93.2%)	295 (95.1%)
Family history[e]	58 (11%)	87 (27%)
Family configurations:		
Both parents no siblings	48 (9.3%)	49 (15.9%)
Both parents and siblings	110 (21.3%)	5 (1.6%)
One parent and siblings	125 (24.2%)	62 (20.1%)
One parent no siblings	4 (0.8%)	31 (10.0%)
Siblings only	230 (44.5%)	163 (52.6%)

[a]n=1078 (ABCFR), 442 (OFBCR) premenopausal probands and sisters
[b]n=1078 (ABCFR), 434 (OFBCR) premenopausal probands and sisters
[c]n=1067 (ABCFR), 425 (OFBCR) premenopausal probands and sisters
[d]n=1067 (ABCFR), 417 (OFBCR) premenopausal probands and sisters
[e]Probands with mother or sister with breast cancer

Allele frequencies

Frequencies of alleles examined in analyses (those common enough to permit analysis with 10 informative families or more) are shown for both ABCFR and OFBCR families, combined and separately, in Tables 2-4. These include all ethnic groups and individuals with and without breast cancer. However, these frequencies are similar to allele frequencies produced by randomly selecting one individual without breast cancer from each Caucasian family (see chapter 5), and for the 5′ polymorphism both of these sets of allele frequencies are similar to 5′ allele frequencies reported previously in Caucasian (*175,177,234*) or Jewish populations (*191*).

Significant deviation from Hardy-Weinberg equilibrium was only found for the 5′ polymorphism in the OBCFR (P=0.04). Although deviation from Hardy-Weinberg equilibrium can be an indication of genotyping error, there is unlikely to be significant error here since reliability tests indicated that genotyping error should be minimal.

***IGF1* genotype and anthropometric measures**

Preliminary analyses that examined interactions between registry and genotype in relationship to each outcome revealed few significant results (BMI: 3′ 181 allele (P=0.04), height: 5′ 19 allele (P=0.03) and 3′ 191 allele (P=0.05)), that might be due to chance. Therefore, families from both registries were combined for all outcomes, but analyses by individual registry are also presented to permit comparisons.

Association of specific alleles of the 5′, intron 2 and 3′ polymorphisms with BMI are shown in Table 2. Direction of effect is indicated by the sign preceding the Z score (minus sign (-) for an inverse association and no sign for a positive association). In the combined samples, nominal significance was observed for only the rare 193 allele of the 3′

polymorphism (P=0.05). Association of this allele with BMI could not be compared across samples because there were too few informative families in the OFBCR to produce a reliable test of association. Allele specific tests for the 5′, intron 2 and 3′ polymorphism did not produce significant results in either the ABCFR or OFBCR sample. Benjamini-Hochberg correction of P-values was used with the false discovery rate (FDR) set at 0.05. This resulted in a corrected P-value of 0.002 for the test of association of the 3′ 193 allele with BMI, therefore this result did not demonstrate a strong enough association to satisfy the criteria used for statistical significance.

In analyses where weight was set as the outcome variable, only one nominally significant association was observed in the combined ABCFR and OFBCR sample. The 3′ 187 allele was positively associated with increased weight (P=0.04) (Table 3). The same allele was also positively associated with weight in both the ABCFR and OFBCR but only in the ABCFR did the result approach nominal significance (ABCFR P=0.06, OFBCR P=0.43) (Table 3). In addition, the P-value required to achieve statistical significance was 0.002 after applying the Benjamini-Hochberg correction, and therefore this result cannot be viewed as statistically significant.

No significant associations were observed for alleles of any of the polymorphisms with height in the combined ABCFR OFBCR sample (Table 4). The 5′ 19 allele was positively associated with height in ABCFR families (P=0.03), but this result was not observed in the OFBCR, where results suggested an effect in the opposite direction (Table 4). The 3′ 189 allele showed a significant inverse association with height in the OFBCR (P=0.03), but this was not supported by the results from the ABCFR (P=0.39).

Table 2. Association between alleles of *IGF1* polymorphisms and BMI in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′												
11	0.5	11	0.54	0.59	-	<10	-	-	-	<10	-	-
17	1.6	33	0.07	0.94	1.5	21	0.72	0.47	1.8	12	-1.32	0.19
18	6.3	126	0.73	0.47	6.5	88	0.27	0.79	5.8	38	0.98	0.33
19	61.2	424	0.16	0.87	60.6	276	-0.37	0.72	62.1	148	0.69	0.49
20	21.2	319	0.24	0.81	21.9	209	0.37	0.71	19.9	110	-0.26	0.79
21	6.7	133	-1.36	0.17	6.6	87	-1.03	0.30	7.0	46	-0.86	0.39
22	2.2	46	-0.43	0.67	2.0	27	0.36	0.72	2.5	19	-1.77	0.08
Intron 2												
204	2.2	50	0.07	0.94	1.9	30	0.93	0.35	2.9	20	-1.21	0.23
212	0.8	16	-1.02	0.31	1.1	15	-0.86	0.39	-	<10	-	-
214	3.0	59	-0.12	0.90	3.0	40	-0.46	0.65	3.1	19	0.47	0.64
216	38.3	428	0.08	0.94	38.3	265	0.31	0.76	38.2	163	-0.43	0.66
218	21.9	333	0.18	0.85	21.3	217	-0.66	0.51	22.8	116	1.44	0.15
220	16.8	259	0.69	0.49	16.8	171	0.77	0.44	16.8	88	-0.03	0.98
222	9.7	193	-0.01	1.00	10.4	134	0.45	0.66	8.4	59	-0.69	0.49
224	6.2	141	-0.89	0.37	6.0	93	-0.75	0.45	6.4	48	-0.30	0.77
226	0.6	15	1.18	0.24	0.6	10	-0.17	0.87	-	<10-	-	-
3′ end												
181	0.9	21	-1.58	0.12	0.9	12	-1.48	0.14	-	<10	-	-
183	1.1	24	-0.19	0.85	1.1	18	-0.15	0.88	-	<10	-	-
185	46.9	430	-0.36	0.72	46.1	282	-0.20	0.85	48.5	148	-0.32	0.75
187	28.8	373	1.57	0.12	29.2	238	1.60	0.11	27.9	135	0.42	0.68
189	18.4	305	0.00	1.00	18.4	196	-0.32	0.75	18.3	109	0.52	0.61
191	2.9	63	0.48	0.63	3.2	47	0.72	0.47	2.4	16	-0.28	0.78
193	1.0	20	-1.96	0.05*	1.1	13	-1.74	0.08	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

Table 3. Association between alleles of *IGF1* polymorphisms and weight in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′												
11	0.5	11	0.68	0.50	-	<10	-	-	-	<10	-	-
17	1.6	33	-0.46	0.64	1.5	21	0.49	0.62	1.8	12	-1.65	0.10
18	6.3	125	0.35	0.73	6.5	87	-0.22	0.82	5.8	38	1.06	0.29
19	61.2	420	0.67	0.50	60.6	272	0.67	0.51	62.1	148	0.11	0.92
20	21.2	317	-0.11	0.91	21.9	207	-0.11	0.91	19.9	110	-0.18	0.86
21	6.7	132	-1.46	0.14	6.6	86	-1.51	0.13	7.0	46	-0.33	0.74
22	2.2	46	-0.06	0.95	2.0	27	0.54	0.59	2.5	19	-1.02	0.31
Intron 2												
204	2.2	50	0.09	0.92	1.9	30	0.73	0.46	2.9	20	-0.68	0.50
212	0.8	16	-0.78	0.43	1.1	15	-0.69	0.49	-	<10	-	-
214	3.0	59	-0.05	0.96	3.0	40	-0.47	0.64	3.1	19	0.67	0.50
216	38.3	424	0.55	0.58	38.3	262	0.56	0.58	38.2	162	0.05	0.96
218	21.9	331	0.37	0.71	21.3	214	-0.29	0.77	22.8	117	1.30	0.19
220	16.8	258	0.10	0.92	16.8	170	0.13	0.90	16.8	88	-0.03	0.98
222	9.7	192	0.04	0.97	10.4	132	0.72	0.47	8.4	60	-0.91	0.36
224	6.2	139	-1.18	0.24	6.0	91	-0.85	0.40	6.4	48	-0.63	0.53
226	0.6	15	0.34	0.73	0.6	10	-0.57	0.57	-	<10	-	-
3′												
181	0.9	21	-1.25	0.21	0.9	12	-1.09	0.27	-	<10	-	-
183	1.1	24	-0.04	0.97	1.1	18	0.06	0.95	-	<10	-	-
185	46.9	428	-0.72	0.47	46.1	280	-0.86	0.39	48.5	148	0.00	1.00
187	28.8	372	2.02	0.04*	29.2	236	1.90	0.06	27.9	136	0.80	0.43
189	18.4	301	-0.15	0.88	18.4	191	0.02	0.98	18.3	110	-0.33	0.74
191	2.9	64	-0.08	0.94	3.2	47	0.18	0.86	2.4	17	-0.44	0.66
193	1.0	20	-1.59	0.11	1.1	13	-1.44	0.15	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

Table 4. Association between alleles of *IGF1* polymorphisms and height in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′												
11	0.5	11	0.53	0.60	-	<10	-	-	-	<10	-	-
17	1.6	32	-1.16	0.24	1.5	21	-0.52	0.60	1.8	11	-1.52	0.13
18	6.3	126	-1.15	0.25	6.5	88	-1.38	0.17	5.8	38	0.39	0.69
19	61.2	417	1.21	0.22	60.6	271	2.18	0.03*	62.1	146	-1.51	0.13
20	21.2	315	-0.65	0.52	21.9	205	-0.96	0.34	19.9	110	0.19	0.85
21	6.7	130	-0.30	0.77	6.6	85	-1.20	0.23	7.0	45	1.77	0.08
22	2.2	45	0.85	0.39	2	26	0.68	0.50	2.5	19	0.73	0.47
Intron 2												
204	2.2	50	0.07	0.95	1.9	30	-0.50	0.62	2.9	20	1.46	0.15
212	0.8	15	-0.07	0.94	1.1	14	-0.19	0.85	-	<10	-	-
214	3.0	59	0.12	0.90	3.0	40	-0.09	0.93	3.1	19	0.84	0.40
216	38.3	426	0.68	0.50	38.3	262	0.52	0.60	38.2	164	0.42	0.68
218	21.9	336	0.91	0.36	21.3	217	0.89	0.37	22.8	119	0.30	0.77
220	16.8	256	-1.52	0.13	16.8	167	-1.60	0.11	16.8	89	-0.25	0.80
222	9.7	193	0.39	0.69	10.4	132	0.87	0.38	8.4	56	-0.71	0.48
224	6.2	141	-0.54	0.59	6	92	-0.27	0.78	6.4	49	-0.60	0.55
226	0.6	15	-1.52	0.13	0.6	10	-0.90	0.37	-	<10	-	-
3′												
181	0.9	21	0.88	0.38	0.9	12	1.13	0.26	-	<10	-	-
183	1.1	25	0.10	0.92	1.1	18	0.35	0.73	-	<10	-	-
185	46.9	426	-0.95	0.34	46.1	277	-1.68	0.09	48.5	149	0.92	0.36
187	28.8	373	1.50	0.13	29.2	234	1.22	0.22	27.9	139	0.88	0.38
189	18.4	306	-0.52	0.60	18.4	195	0.85	0.39	18.3	111	-2.15	0.03*
191	2.9	63	-1.05	0.29	3.2	47	-1.49	0.14	2.4	16	0.96	0.34
193	1.0	20	1.60	0.11	1.1	13	1.31	0.19	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

IGF1 haplotypes and anthropometric measures

Significant linkage disequilibrium was observed between adjacent polymorphisms and those at the 5′ and 3′ end (P=0.0002 for 5′ and 3′ polymorphisms in the OFBCR and P<0.00001 for all other pairs adjacent polymorphisms in the ABCFR and OFBCR). However, only adjacent polymorphisms had allele pairs with R^2 values greater than 0.1 (see Table 5 chapter 5). Therefore, haplotype analyses were restricted to adjacent polymorphisms. Haplotype analysis with BMI as the outcome produced a nominally significant result for one haplotype 22-214 constructed from the 5′ and intron 2 polymorphisms (P=0.02) (Table 5). This haplotype was the only one for which there was a nominally significant association with weight (P=0.05) (Table 6). One haplotype constructed from the intron 2 and 3′ polymorphism, 220-185, showed a nominally significant association with height (P=0.01) (Table 7). None of these results were significant after adjusting for multiple comparisons using the Benjamini-Hochberg correction.

Comparison of the associations of the 22-214 haplotype with BMI and weight in the ABCFR and OFBCR samples could not be made, since for this rare allele the minimum sample size of 10 informative families could only be achieved by combining samples. Inspection of Table 7 shows that a nominally significant association of the 220-185 allele with height was present in the ABCFR, in addition to the combined sample (P=0.004). However, this association is not supported by results from the OFBCR (P=0.69). The 5′-intron 2 20-220 allele showed a nominally significant association with height in the ABCFR (P=0.01), but again this result was not supported by the results from the OFBCR where the P-value was 0.68 and the effect was in the opposite direction to that observed in the ABCFR.

None of the other statistical tests of the association of haplotypes with any of the three outcomes produced nominally significant results.

Table 5. Association between *IGF1* haplotypes and BMI in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′ – intron 2												
19-216	24.2	355.8	-0.75	0.45	23.9	223.6	-0.81	0.42	24.8	130.5	-0.45	0.65
19-218	15.3	256.5	0.34	0.73	14.8	166.1	-0.01	1.00	16.2	88.6	0.98	0.33
20-216	9.0	158.1	0.85	0.40	9.1	94.7	0.80	0.43	8.5	60.0	0.25	0.80
19-222	8.3	160.1	0.18	0.85	8.6	109.0	0.45	0.65	7.7	47.5	-0.45	0.65
19-220	6.7	124.6	1.74	0.08	6.5	78.9	1.41	0.16	7.1	46.0	1.08	0.28
19-224	5.3	126.5	-1.41	0.16	5.4	82.5	-1.52	0.13	5.0	41.1	-0.21	0.83
20-220	5.3	103.5	0.83	0.41	5.1	68.8	1.35	0.18	5.4	35.3	-0.76	0.45
20-218	3.8	87.7	-0.68	0.50	3.9	56.6	-0.96	0.34	3.7	31.5	-0.60	0.55
18-216	3.7	87.9	0.83	0.41	4.1	62.7	0.61	0.54	3.0	26.0	0.36	0.72
21-220	2.8	63.0	-1.45	0.15	2.5	41.6	-1.27	0.20	3.4	20.1	-0.66	0.51
18-218	1.5	36.4	0.43	0.66	1.7	25.0	-0.11	0.91	1.2	10.1	0.66	0.51
19-214	1.3	29.8	1.13	0.26	1.2	19.1	0.73	0.46	-	<10	-	-
20-222	1.1	28.3	-0.83	0.40	1.3	23.6	0.02	0.99	-	<10	-	-
21-216	1.1	22.2	-1.00	0.32	1.0	14.0	-0.91	0.36	-	<10	-	-
21-218	1.1	14.8	-0.25	0.80	0.9	10.5	-0.70	0.48	-	<10	-	-
22-214	0.8	12.0	-2.30	0.02*	-	<10	-	-	-	<10	-	-
19-204	0.7	18.2	-0.99	0.32	0.7	13.1	-1.10	0.27	-	<10	-	-
22-216	0.7	14.0	0.41	0.68	0.9	11.0	0.80	0.43	-	<10	-	-
17-220	0.6	12.9	-1.28	0.20	-	<10	-	-	-	<10	-	-
19-212	0.6	13.3	-1.12	0.26	0.8	12.4	-1.02	0.31	-	<10	-	-
21-204	0.5	14.6	-0.20	0.85	-	<10	-	-	-	<10	-	-
17-218	0.5	12.9	0.93	0.35	-	<10	-	-	-	<10	-	-
21-222	0.4	10.7	1.44	0.15	-	<10	-	-	-	<10	-	-

Continued on the following page.

Table 5. Association between *IGF1* haplotypes and BMI in the ABCFR and OFBCR (continued from previous page).

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
intron 2-3′												
216-187	18.2	301.6	0.40	0.69	18.2	186.4	0.43	0.67	18.0	114.7	0.06	0.95
216-185	17.1	260.5	0.52	0.61	17.6	168.0	0.52	0.61	16.5	92.5	-0.02	0.98
218-185	16.8	277.2	0.04	0.97	16.4	179.0	-0.68	0.50	17.9	96.4	1.16	0.25
220-185	9.2	169.2	0.85	0.39	8.7	110.7	1.24	0.22	10.0	57.4	-0.72	0.47
222-189	7.1	132.7	0.00	1.00	7.5	96.5	0.09	0.93	6.4	39.7	-0.08	0.93
224-189	4.1	92.8	-0.59	0.56	3.9	55.7	-1.13	0.26	4.4	38.0	0.30	0.76
220-187	4.1	83.9	0.60	0.55	4.2	55.2	0.85	0.40	4.2	29.1	0.09	0.93
216-189	2.9	66.3	-0.72	0.47	3.0	41.4	0.12	0.91	2.5	23.1	-1.33	0.18
218-187	2.7	59.7	0.21	0.84	2.9	41.0	-0.02	0.99	2.3	17.6	0.33	0.74
218-189	2.0	45.9	0.37	0.71	2.1	27.2	0.04	0.97	1.6	15.4	0.25	0.80
204-185	1.8	45.8	-0.33	0.74	1.6	27.0	0.40	0.69	2.1	18.7	-1.12	0.26
220-189	1.7	30.0	0.67	0.50	1.2	14.4	-0.11	0.91	2.3	13.9	1.50	0.13
222-185	1.7	34.6	-0.76	0.45	1.7	23.0	0.43	0.67	1.6	10.4	-1.47	0.14
214-187	1.6	36.1	1.00	0.32	1.7	26.0	0.61	0.54	1.5	10.1	0.97	0.33
224-191	1.2	24.8	-0.29	0.78	1.4	20.0	-0.60	0.55	-	<10	-	-
214-185	1.0	22.0	-1.35	0.18	0.8	13.0	-1.45	0.15	-	<10	-	-
222-191	0.7	18.5	0.70	0.48	0.7	10.9	1.05	0.30	-	<10	-	-
216-183	0.4	11.0	0.75	0.46	-	<10	-	-	-	<10	-	-
218-181	0.3	10.8	-0.57	0.57	-	<10	-	-	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

Table 6. Association between *IGF1* haplotypes and weight in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′ – intron 2												
19-216	24.2	351.1	0.04	0.97	23.9	217.9	-0.21	0.83	24.8	130.5	0.13	0.90
19-218	15.3	254.0	0.41	0.68	14.8	162.6	0.35	0.73	16.2	89.6	0.43	0.67
20-216	9.0	155.1	0.73	0.47	9.1	92.7	0.91	0.36	8.5	59.0	-0.10	0.92
19-222	8.3	160.1	0.15	0.88	8.6	109.0	0.65	0.51	7.7	49.5	-0.71	0.48
19-220	6.7	124.7	1.53	0.13	6.5	79.0	1.58	0.11	7.1	46.0	0.38	0.71
19-224	5.3	124.5	-1.47	0.14	5.4	80.5	-1.33	0.18	5.0	41.1	-0.61	0.54
20-220	5.3	106.5	0.19	0.85	5.1	69.8	0.35	0.73	5.4	37.3	-0.26	0.80
20-218	3.8	87.8	-0.31	0.76	3.9	56.6	-0.73	0.46	3.7	31.5	-0.18	0.86
18-216	3.7	88.7	0.26	0.80	4.1	62.5	0.00	1.00	3.0	26.0	0.49	0.63
21-220	2.8	65.0	-1.84	0.07	2.5	41.6	-1.87	0.06	3.4	20.1	-0.46	0.64
18-218	1.5	35.7	0.36	0.72	1.7	24.2	-0.37	0.71	1.2	10.1	0.70	0.48
19-214	1.3	29.8	1.60	0.11	1.2	19.1	1.11	0.27	-	<10	-	-
20-222	1.1	27.3	-1.15	0.25	1.3	22.6	-0.11	0.91	-	<10	-	-
21-216	1.1	21.8	-0.46	0.64	1.0	13.8	-0.57	0.57	-	<10	-	-
21-218	1.1	14.8	-0.47	0.64	-	<10	-	-	-	<10	-	-
22-214	0.8	12.0	-1.92	0.05*	-	<10	-	-	-	<10	-	-
19-204	0.7	18.2	-0.63	0.53	0.7	13.1	-0.95	0.34	-	<10	-	-
22-216	0.7	14.0	0.38	0.70	0.9	11.0	0.87	0.38	-	<10	-	-
17-220	0.6	12.9	-1.16	0.25	-	<10	-	-	-	<10	-	-
19-212	0.6	13.3	-0.70	0.49	0.8	12.4	-0.75	0.45	-	<10	-	-
21-204	0.5	14.6	-0.82	0.41	-	<10	-	-	-	<10	-	-
17-218	0.5	12.9	0.41	0.68	-	<10	-	-	-	<10	-	-
21-222	0.4	10.7	1.13	0.26	-	<10	-	-	-	<10	-	-

Continued on the following page.

Table 6. Association between *IGF1* haplotypes and weight in the ABCFR and OFBCR (continued from previous page).

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
intron 2-3′												
216-187	18.2	302.6	0.80	0.42	18.2	186.4	0.70	0.49	18.0	113.7	0.33	0.74
216-185	17.1	257.5	0.39	0.70	17.6	165.0	0.20	0.85	16.5	92.5	0.31	0.76
218-185	16.8	277.7	-0.03	0.98	16.4	178.4	-0.40	0.69	17.9	98.4	0.74	0.46
220-185	9.2	172.2	-0.20	0.84	8.7	111.7	0.08	0.94	10.0	57.4	-0.73	0.47
222-189	7.1	132.7	-0.19	0.85	7.5	94.5	-0.06	0.95	6.4	38.7	-0.17	0.87
224-189	4.1	92.8	-0.85	0.40	3.9	54.7	-1.06	0.29	4.4	38.0	-0.22	0.82
220-187	4.1	84.9	1.27	0.21	4.2	55.2	1.51	0.13	4.2	30.1	0.39	0.70
216-189	2.9	65.3	-0.34	0.74	3.0	40.4	0.26	0.80	2.5	23.1	-1.07	0.28
218-187	2.7	59.5	0.22	0.83	2.9	39.8	0.13	0.90	2.3	18.6	-0.19	0.85
218-189	2.0	46.6	1.24	0.21	2.1	28.0	0.83	0.41	1.6	16.4	0.73	0.47
204-185	1.8	45.8	-0.30	0.76	1.6	27.0	0.19	0.85	2.1	18.7	-0.62	0.54
220-189	1.7	29.0	0.36	0.72	1.2	13.5	-0.01	0.99	2.3	13.9	1.36	0.17
222-185	1.7	33.6	-0.74	0.46	1.7	22.0	0.60	0.55	1.6	10.4	-1.53	0.13
214-187	1.6	36.1	0.95	0.34	1.7	26.0	0.56	0.58	1.5	10.1	0.94	0.35
224-191	1.2	24.8	-0.16	0.88	1.4	20.0	-0.30	0.77	-	<10	-	-
214-185	1.0	22.0	-0.90	0.37	0.8	13.0	-1.40	0.16	-	<10	-	-
222-191	0.7	19.5	0.78	0.43	0.7	10.9	1.08	0.28	-	<10	-	-
216-183	0.4	11.0	1.10	0.27	-	<10	-	-	-	<10	-	-
218-181	0.3	10.8	-0.54	0.59	-	<10	-	-	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

Table 7. Association between *IGF1* haplotypes and height in the ABCFR and OFBCR.

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
5′ – intron 2												
19-216	24.2	351.7	1.12	0.26	23.9	218.6	1.21	0.23	24.8	130.5	0.28	0.78
19-218	15.3	254.9	0.63	0.53	14.8	166.1	0.95	0.34	16.2	86.7	-0.32	0.75
20-216	9.0	156.4	0.11	0.91	9.1	92.4	0.36	0.72	8.5	61.0	-0.40	0.69
19-222	8.3	160.1	0.16	0.87	8.6	107.0	0.53	0.60	7.7	50.4	-0.52	0.60
19-220	6.7	126.1	-0.39	0.70	6.5	78.6	0.50	0.62	7.1	47.9	-1.60	0.11
19-224	5.3	126.5	-0.06	0.95	5.4	81.6	0.38	0.71	5.0	42.1	-0.65	0.51
20-220	5.3	103.1	-1.88	0.06	5.1	66.8	-2.44	0.01*	5.4	36.2	0.42	0.68
20-218	3.8	89.1	0.92	0.36	3.9	56.6	0.42	0.67	3.7	32.6	1.39	0.17
18-216	3.7	88.1	-1.35	0.18	4.1	62.7	-1.53	0.13	3.0	25.0	0.46	0.64
21-220	2.8	62.0	-0.97	0.33	2.5	40.6	-1.36	0.18	3.4	19.0	0.90	0.37
18-218	1.5	36.3	-0.18	0.86	1.7	25.0	-0.43	0.67	1.2	10.0	0.26	0.79
19-214	1.3	29.8	1.34	0.18	1.2	19.1	0.76	0.45	-	<10	-	-
20-222	1.1	28.3	-0.52	0.60	1.3	23.6	0.10	0.92	-	<10	-	-
21-216	1.1	21.8	1.11	0.27	1.0	13.7	0.80	0.42	-	<10	-	-
21-218	1.1	14.8	-0.17	0.87	0.9	10.5	-1.69	0.09	-	<10	-	-
22-214	0.8	13.0	0.08	0.94	-	<10	-	-	-	<10	-	-
19-204	0.7	19.2	0.89	0.37	0.7	13.1	0.49	0.62	-	<10	-	-
22-216	0.7	13.0	-0.20	0.84	0.9	10.0	0.03	0.98	-	<10	-	-
17-220	0.6	11.9	0.73	0.47	-	<10	-	-	-	<10	-	-
19-212	0.6	12.3	0.41	0.69	0.8	11.4	0.07	0.95	-	<10	-	-
21-204	0.5	14.6	-1.54	0.12	-	<10	-	-	-	<10	-	-
17-218	0.5	12.9	-1.11	0.27	-	<10	-	-	-	<10	-	-

Continued on the following page.

Table 7. Association between *IGF1* haplotypes and height in the ABCFR and OFBCR (continued from previous page).

Loci/ allele	ABCFR and OFBCR				ABCFR				OFBCR			
	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value	Allele frequency %	NIF	Z score	P-value
intron 2-3′												
216-187	18.2	301.6	1.11	0.27	18.2	183.4	0.95	0.34	18.0	115.7	0.49	0.62
216-185	17.1	256.5	-0.40	0.69	17.6	166.1	-0.84	0.40	16.5	91.6	0.69	0.49
218-185	16.8	281.2	0.41	0.68	16.4	180.0	0.74	0.46	17.9	99.4	-0.19	0.85
220-185	9.2	168.2	-2.78	0.01*	8.7	108.6	-2.89	0.004*	10.0	55.9	-0.40	0.69
222-189	7.1	132.7	-0.35	0.73	7.5	95.5	-0.05	0.96	6.4	40.7	-0.51	0.61
224-189	4.1	94.8	-0.36	0.72	3.9	55.7	0.10	0.92	4.4	40.0	-0.87	0.38
220-187	4.1	85.9	1.75	0.08	4.2	54.1	1.64	0.10	4.2	31.6	0.77	0.44
216-189	2.9	66.3	-0.41	0.68	3.0	41.4	-0.08	0.94	2.5	23.1	-1.48	0.14
218-187	2.7	60.7	0.17	0.87	2.9	41.0	0.31	0.76	2.3	18.6	-0.73	0.46
218-189	2.0	45.9	1.56	0.12	2.1	28.2	1.56	0.12	1.6	15.4	0.58	0.56
204-185	1.8	46.8	0.02	0.98	1.6	27.0	-0.53	0.59	2.1	19.2	1.18	0.24
220-189	1.7	29.0	-0.78	0.43	1.2	14.5	0.04	0.97	2.3	12.9	-0.66	0.51
222-185	1.7	33.6	0.16	0.87	1.7	22.0	0.48	0.63	1.6	10.3	-0.40	0.69
214-187	1.6	36.1	-0.19	0.85	1.7	26.0	-0.21	0.83	1.5	10.1	0.32	0.75
224-191	1.2	24.8	0.36	0.72	1.4	20.0	0.58	0.56	-	<10	-	-
214-185	1.0	22.0	0.95	0.34	0.8	13.0	0.32	0.75	-	<10	-	-
222-191	0.7	18.5	0.80	0.42	0.7	10.9	-0.05	0.96	-	<10	-	-
216-183	0.4	12.0	0.21	0.84	-	<10	-	-	-	<10	-	-
218-181	0.3	10.8	0.03	0.98	-	<10	-	-	-	<10	-	-

NIF: number of informative families.
* Nominally significant ($P \leq 0.05$).
- Results not reported when number of informative families is less than 10.

DISCUSSION

The results do not support a relationship between genetic variation at the *IGF1* gene and BMI, weight and height. In allele specific analyses there were a few nominally significant associations. These can, however, be explained by chance as there was a lack of consistency when significant associations were compared across samples, and none of the nominally significant associations met the criteria of significance after adjustment for multiple comparisons. Similarly, chance findings can also explain the few nominally significant associations observed in the haplotype analyses.

A major strength of this study is the family based design, which eliminates the possibility of confounding resulting from population stratification. In addition, this is a large sample of premenopausal women, a group for which sample size in other studies is often fairly small. A potential limitation of this study is that all measures were conducted on families where at least one member (the proband) had breast cancer. Therefore, the sample cannot be considered as representative of the populations they were sampled from.

The inclusion of women with breast cancer (probands and some sisters) in the sample, could conceivably lead to differential recall of for example, weight between women with and without breast cancer. However, differential recall between affected and unaffected family members should not be related to genotype, unless genotype was related to breast cancer risk. Since results from the previous chapter indicated there was limited evidence for an association between genotype and risk, differential recall between affected and unaffected family members is unlikely to bias the results. In addition analyses were also conducted excluding women with breast cancer (data not shown) and this did not alter the interpretation of the results.

The results presented here do not support the association between the 185 allele of the 3′ polymorphism with BMI and height reported in chapter 4. A nominally significant association of the 3′ 187 allele with weight was observed here, but this association was not found in analyses performed on the sample of healthy unrelated individuals described in chapter 4. In addition, results from the OFBCR did not support an association of the 3′ 187 allele with weight. Therefore, overall there is little evidence to support an association of the 3′ 187 allele with weight.

Several studies have reported on the association of the 5′ polymorphism with BMI or weight. In a study conducted in the Netherlands (Amsterdam), homozygote carriers of either the 19 or 20 alleles or carriers of both these alleles were reported to have significantly greater BMI relative to women with other genotypes, but only in the younger of two cohorts examined (*248*). Consistent with the results from the current study, no association of the 19 allele with BMI was reported in Hispanic or non-Hispanic White women in the U.S. (*249*), in chapter 4 of this thesis, or in a Chinese population (*189*), although in the latter study the 17 allele was associated with lower BMI (*189*).

In a large population based sample from the Netherlands (Rotterdam), women homozygous for either the 19 and 20 alleles or carriers of both these alleles were heavier than women who carried longer alleles (*186*). Although this analysis was not carried out in the families studied here, analyses where the 19 and 20 alleles were grouped and compared to all others under a recessive model, did not provide support for a potential positive association of these alleles with weight (data not shown).

A few studies have reported on the association of the 5′ polymorphism with height. No association between the 19 allele and height was observed in the U.S. Nurses Health Study

(*179*), or with the 19 and or 20 alleles of women from a population based sample in the Netherlands (*186*). In addition, there was no association between the 5′ 19 allele and height observed in chapter 4. In the current study results were inconsistent, an association of the 5′ 19 allele with greater height was observed in the ABCFR sample, but a non-significant association in the opposite direction was observed in OFBCR families. Overall, the evidence does not support an association between variant alleles of this polymorphism and height.

Some studies have examined other measures of body composition in relation to genetic variation at *IGF1*. There is some evidence for an association of the 19 allele of the 5′ polymorphism and waist hip ratio (*249*). In addition, an association between an *IGF1* CT repeat polymorphism with percentage fat, fat mass, fat free mass, and change in fat free mass, was observed in a family based exercise intervention study (*236*) and in lean body mass and leg mass in a sample of unrelated individuals (*251*). These results are not consistent with those observed here, although this could be attributed to the differences in body composition measures used, or the examination of a different *IGF1* polymorphism.

The results presented here provide little evidence to support an association between genetic variation at *IGF1* with either BMI, weight or height. There were, however, only three polymorphisms tested and linkage disequilibrium between them was not strong enough to fully capture genetic variation across *IGF1*. Other studies have examined only a single *IGF1* polymorphism in relation to anthropometric features. Future studies should include additional polymorphisms to better capture genetic variation at this gene. Inclusion of other measures of body composition would also be useful.

Chapter 7. Discussion

7.1. Summary of purpose

In chapter 1, a model was presented where in premenopausal women, genetic variation at the *IGF1* gene was hypothesized to modify circulating IGF-I concentration, with greater circulating IGF-I levels increasing the proliferative activity and quantity of breast stromal and epithelial cells, resulting in greater mammographic density and increased risk of developing breast cancer. Evidence to support aspects of this model comes from studies that although not entirely consistent, indicate that a greater circulating concentration of IGF-I is associated with greater mammographic density (*5-8*) and increased breast cancer risk in premenopausal women (*1-3*). Additional evidence comes from twin studies that indicate that circulating IGF-I levels, mammographic density, and the development of breast cancer are partly determined by genetic factors (*9-12*). This suggests that genes that influence the IGF-I pathway are good candidates to examine for association with mammographic density and breast cancer risk.

The association of allelic variants of *IGF1* with circulating IGF-I concentration has been examined in several studies, with most examining the 5′ polymorphism that was also investigated here. These studies provide little evidence to support an association between alleles at the 5′ polymorphism and IGF-I levels. There is, however, some support for an association of several *IGF1* tagging SNPs with IGF-I levels from a large study conducted in Great Britain (*187*).

Only two studies have examined the association between genetic variation at *IGF1* and mammographic density. One found no association between the 5′ 19 allele and mammographic density in either pre- or postmenopausal women (*99*). A second study found

a strong association between *IGF1* haplotypes and tagging SNPs and mammographic density in a sample comprised mainly of postmenopausal women (*100*). Little attention has been given to the association of *IGF1* with body composition (*186,189,248,249*).

Several studies examined the association of the 5′ polymorphism and breast cancer risk with inconsistent results (*172,179,189-193*). Results are also inconsistent among studies that investigated this polymorphism in relationship to breast cancer risk in large samples of premenopausal (or young) women, with either null results or different alleles showing association with disease (*189,190,193*).

Two studies (comprised mainly of postmenopausal women) found nominally significant associations with tagging SNPs at the 5′ end of *IGF1*, although results were not significant after adjustment for multiple comparisons (*194,195*). A third study (comprised mainly of premenopausal women) reported significant associations with five tagging SNPs across *IGF1* (located in introns 2 and 3 and the untranslated region of exon 4), (*187*) under a recessive model, although the interpretation of this study is complicated by the inclusion of prevalent cases. None of the studies reported associations stratified by menopausal status. Genome wide association studies have not reported significant associations with *IGF1*, but initial screens in these studies have not included large samples of premenopausal women (*38,39,197*).

A potential source of error highlighted as particularly important to many of the above studies was the lack of strong linkage disequilibrium across *IGF1*. If an ungenotyped causal variant is present, studies that examined only one polymorphism (e.g. the 5′ polymorphism) may have insufficient power to detect an association because of weak linkage disequilibrium between the two loci. Confounding due to population stratification is another factor that may

have led to inconsistencies in results. The studies presented here were designed to address these issues as much as possible. In addition, they address the need for more research into the association of mammographic density and anthropometric variables with allelic variants of *IGF1*, and larger sample size for studies examining *IGF1* and breast cancer risk in premenopausal women, where the association of IGF-I levels with risk has been reported.

7.2. Description of study

The studies described in chapters 4 through 6 examined the association between allelic variants of *IGF1* with circulating IGF-I levels, mammographic density, breast cancer risk and several related anthropometric measures (BMI, weight and height) in premenopausal women. Three CA repeat polymorphisms were examined to better capture genetic variation at *IGF1*. The issue of confounding due to population stratification was addressed in the breast cancer study by using families from the ABCFR and OFBCR. In the cross-sectional study examining *IGF1* and mammographic density, the possibility of confounding due to population stratification was reduced by restricting analyses to Caucasian women. Several anthropometric measures (BMI, weight and height) were measured in both the cross-sectional and family based study, in an attempt to replicate the results.

7.3. Summary of results

Main findings are summarized with reference to the original objectives in Table 1. Not included are results pertaining to potential effect modification by family history and oral contraceptive use as they are not part of the main objectives. However, effect modification by family history is discussed in section 7.4 (sources of error).

Table 1. Summary of main findings.

Original Objective	Main findings
Examine the association of each of the polymorphisms with circulating IGF-I levels.	Greater number of copies of the 5′ 19 allele were associated with lower IGF-I levels.
Examine the association of each of the polymorphisms with percentage breast density.	Greater number of copies of the 3′ 185 allele were associated with greater percentage breast density and a smaller amount of non-dense (fat) tissue. Analyses suggested the association of the 3′ 185 allele with percentage breast density is mediated through BMI (i.e. body fat).
Examine the association between genetic variation at *IGF1* and BMI, height and weight.	A significant association between the 3′ 185 allele and BMI and height was found in the cross-sectional study. An association of genetic variation at *IGF1* and either BMI, height or weight was not supported in the family based study.
Examine the association of each of the polymorphisms with breast cancer risk	*Primary analyses:* Some nominally significant associations were observed (5′ 21 allele, intron 2 212 allele, intron 216 allele) but these were not significant after adjustment for multiple comparisons and there was a lack of consistency across the ABCFR and OFBCR. *Exploratory analyses:* Associations were observed for the intron 2 216 allele under a recessive model and 5′ allele groupings resulted in a positive association for alleles of length 18 to 20, and a protective effect for alleles > than 20 repeats in length. These associations were not consistent across the ABCFR and OFBCR. Overall, there was limited evidence to support an association of allelic variants of the tested polymorphisms with breast cancer risk.
Determine linkage disequilibrium between markers. Construct haplotypes and examine their association with circulating IGF-I levels, percentage breast density and breast cancer risk.	Linkage disequilibrium between specific allele pairs of these polymorphisms was weak (No R^2 greater than 0.21 and generally less than 0.1). Haplotype analyses did not provide support for an association between genetic variation at *IGF1* and IGF-I levels, mammographic density or breast cancer risk.

7.3.1. *IGF1*, IGF-I concentration, mammographic density, and breast cancer risk

7.3.1.1. Hypothesized model

In the hypothesized model, variant alleles of *IGF1* were proposed to modify circulating IGF-I levels, with greater IGF-I levels promoting the proliferative activity and quantity of stromal and epithelial tissue in the breast, resulting in greater mammographic density and an increased risk of developing breast cancer. Although not discussed previously, greater percentage density has been shown to be associated with greater occurrence of benign proliferative disease, and atypical hyperplasia and ductal carcinoma *in situ*. These conditions are known to increase breast cancer risk and have been shown to be more likely to occur in women with greater percentage breast density (*252*). This suggests a more complex model, where IGF-I influence on breast tissue leading to benign proliferative disease, atypical hyperplasia and ductal carcinoma *in situ* in women with greater mammographic density results in increased risk of breast cancer. Results from the only study to examine the association of mammographic density and benign breast disease (carcinoma *in situ* was not included) with breast cancer risk, indicated that benign breast disease did not explain the association of percentage breast density with risk (*253*). Moreover, results indicated that atypical hyperplasia was most strongly associated with breast cancer risk in women with lower percentage breast density, while women with atypical hyperplasia and greater mammographic density were at reduced risk for developing breast cancer. However, the sample size for these sub-groups were small. As well, relatively few premenopausal women were included in the study. Nonetheless, current evidence, although limited, does not support a role for benign breast disease in the model. These results do not exclude an influence of

IGF-I levels on benign breast disease and ultimately breast cancer risk that is independent of mammographic density, although this discussion is beyond the scope of this thesis.

7.3.1.2. Interpretation of results in context of the model

Considering then the model as originally proposed, the results do not indicate that one specific allelic variant of *IGF1* is associated with all three outcomes of interest: IGF-I levels, mammographic density and breast cancer risk (Table 1). Furthermore, as elaborated on below, an association of *IGF1* with mammographic density is not strongly supported, and there is limited evidence for an association with breast cancer risk.

In chapter 4, the association of the 3′ 185 allele with mammographic density was not mediated through IGF-I levels, which were instead associated with the 5′ 19 allele. Although not entirely consistent with the originally proposed model, an influence of the 3′ 185 genotype could still be mediated through changes directly to the IGF-I gene product or through changes to breast tissue specific expression of IGF-I (i.e. autocrine/paracrine effects). If, however, mammographic density was influenced through either of these pathways results did not strongly indicate an effect on dense tissue was involved (which is comprised of stromal and epithelial cells), as would be predicted by the model. A possible influence of the 3′ 185 allele on amount of dense tissue could not be ruled out as the association with dense tissue was close to significant under the assumption of a dominant effect of this allele. Still, the evidence indicated that the observed association of the 3′ 185 allele with percentage density was mediated through a possible effect on BMI and in particular body fat. The amount of non-dense (fat) tissue and BMI were both inversely associated with the number of copies of this allele, and including BMI in regression models

resulted in loss of significance and substantial reduction in strength of effect of genotype on percentage density (Table 1).

The possible association of the 3′ 185 allele on percentage breast density through an effect on body fat stores was not, however, supported by the results in chapter 6. BMI was not observed to be associated with 3′ 185 genotype in the family based study in either the combined sample or in the ABCFR or OFBCR (Table 1). Sampling strategies were different between these two studies (stratified sampling was used in the cross-sectional study to maximize the range in percentage density, while in the family based study families were sampled based on a proband with breast cancer). However, this did not produce meaningful differences in ranges for BMI and is unlikely to explain the differences in results between the two studies. This indicates that the association of 3′ 185 allele with BMI observed in the cross-sectional study may in fact be spurious, and further suggests that the observed association of this allele with breast density is a false positive result.

With reference to the latter part of the model, there was no indication that a common *IGF1* variant was associated with both mammographic density and breast cancer. In fact, as indicated in Table 1, the evidence providing support for an association of variant alleles of *IGF1* with breast cancer risk was limited. Exploratory analyses did suggest an association of medium and long alleles of the 5′ polymorphism with risk, and a possible association with risk for the intron 2 216 allele under a recessive model. These associations were not, however, strongly supported in both the ABCFR and OFBCR. Furthermore, the positive association of medium length alleles with breast cancer risk was not consistent with the observed association of the 5′ 19 allele, the most common medium length allele, with lower IGF-I levels. Interestingly, results from chapter 4 provided qualitative support for the

association of the intron 2 216 allele with IGF-I levels under a recessive model (with P=0.07, results of analysis not shown) consistent with the association observed in the combined sample and the OFBCR for breast cancer. The data do not, however, provide support for an association of the intron 2 216 allele with mammographic density under a recessive model (analysis not shown).

Further discussion of these results in context with the published literature is returned to later in this chapter, following a discussion of the potential limitations of the studies presented in chapters 4 through 6.

7.4. Potential sources of error

7.4.1. Sampling

7.4.1.1. Limitations of cross-sectional studies

A limitation of cross-sectional studies is that exposure and outcome is measured concurrently and it is not always clear whether exposure precedes outcome. The causal direction of the study presented in chapter 4, however, is not at issue since genotype is a fixed exposure (i.e. not influenced by phenotype). Another limitation, length-biased sampling, where diseased cases with long duration are over sampled while diseased cases with short duration are under sampled (*254*), is not an issue since disease was not measured in this study. However, several outcomes related to disease (breast cancer) were measured, including mammographic density which is strongly related with risk. Women with more extensive mammographic density could potentially be under sampled, since a greater proportion of these women are expected to be diagnosed with breast cancer resulting in their being ineligible for this study. However, breast cancer is a rare disease in women in this age group (33-58 years) and diagnoses among potential study subjects should have little

influence on the final distribution of mammographic density in the study sample. In addition, the stratified sampling approach used would tend to capture additional numbers of women in the most extensive categories of density.

Interaction between genotype and mammographic density in relation to breast cancer risk could introduce bias into the study as fewer individuals with a combination of genotype and mammographic density related to breast cancer risk would appear in the sample. Again, the fact that breast cancer is a rare disease in this age group makes it unlikely that such an interaction would have an important effect. Furthermore, there was no indication of important distortion of allele or genotype frequencies (see allele frequencies and Hardy-Weinberg tests reported in results sections of chapters 4 and 5).

As with a case-control or cohort design, a cross-sectional study is prone to bias due to confounding resulting from population stratification, as other genetic or environmental factors associated with both genotype and phenotype through population substructure (e.g. ethnicity) could distort the relationship between genotype and outcome. Therefore the sample was restricted to include Caucasian women only, which should substantially reduce the potential for this form of confounding. Allele frequencies for the 5′ polymorphism in this sample were very similar to those in other Caucasian populations in the U.S., Europe and Israel (*175,177,191,234*). This suggests that within Caucasian populations, genetic variation at this locus is small, and if true for other *IGF1* loci, the chances of false positive or negative findings due to population stratification should be minimal. Still, the distribution of this and other alleles in Caucasian populations across Europe is unknown, and since this sample consists of individuals of European descent, variation in allele frequencies that is non-

causally correlated with phenotype could result in confounding due to population stratification (*201*).

7.4.1.2. Recruitment in cross-sectional study

The stratified sampling scheme used in the cross-sectional study was designed to improve efficiency by assembling a group of women with a wide range of mammographic density, since an association with outcome is less likely to be observed if its range is restricted or truncated (*254*). However, the distribution of outcome variables should not bias effect estimates unless the relationship between genotype and outcome changes across categories of density, and even if this were true, a spurious association would not result in the absence of population stratification, although the result would no longer be generalizable.

7.4.1.3. Limitations of family based studies

The advantage of appropriately analyzed family based designs examining genotype phenotype associations is that they are not susceptible to confounding bias due to population stratification. An important disadvantage of this design is reduced efficiency relative to case-control studies, particularly when one or more parents are missing and parental genotypes cannot be precisely reconstructed. As discussed in chapter 3, this was compensated for here by utilizing and combining samples from the ABCFR and OFBCR.

As with case-control studies, if genotypes of the polymorphism under investigation are related to mortality, then certain genotypes may be over- or underrepresented among affected individuals in a family based study. However, the potential influence of this bias on results was reduced by ensuring probands were incident primary breast cancer cases. There was no such restriction with respect to time since diagnosis for sisters, however, and if lower

survival rates were related to a specific genotype, then sisters with genotype and breast cancer would be less likely to be available. This could result in some sampling bias as other relatives in families with deceased sisters would not always be available to provide a blood sample for DNA analysis and these families would not be included in the sample. The influence of this bias should, however, be negligible as breast cancer is a rare disease in premenopausal women, survival rates are relatively high, and in many cases another family member would in fact be available to provide a DNA sample. Therefore, recruitment efforts would only miss a few families if breast cancer mortality was related to *IGF1* genotype.

Other factors might have an influence on recruitment of probands or family members. Weight and BMI, which were outcome variables in chapter 6, might influence recruitment into the study. For example, individuals with greater BMI might not agree to enter the study because of health related reasons. If BMI was positively associated with a specific *IGF1* genotype, then the power to detect an association between genotype and BMI could be decreased, because fewer individuals with the genotype related to greater BMI might enroll. It is unlikely, however, that there is a substantial proportion of people declining to enroll in the study because of extreme BMI or weight related to a common *IGF1* variant. Therefore, any resulting selection bias should not have a major influence on study power in chapter 6, or for that matter in the breast cancer study reported on in chapter 5. For the same reason, it is also unlikely that the association of *IGF1* variants with other health related factors would have an important influence on study power.

Another potential disadvantage in family based studies is bias in detecting associations in the presence of genotyping error. This is discussed in detail in the section on genotyping error (see below).

7.4.1.4. Recruitment in family based study

The OFBCR and the ABCFR were both population based samples, with study subjects being largely Caucasian (Table 2). Probands in the ABCFR were younger than those in the OFBCR as prior to 1996 all probands recruited were under the age of 40. However, age did not seem to have an important influence on study results (see age *IGF1* genotype interaction analyses in chapter 5).

In the OFBCR, probands with a family history of breast cancer were over sampled. Using a definition of family history where the proband had one mother and/or sister with a reported diagnosis of breast cancer, 27% of OFBCR probands and 11% of ABCFR probands were classified as having a family history of breast cancer (Table 2). The latter figure is similar to percentage of cases reported with a family history of breast cancer in mainly population based studies conducted in developed countries for this age group (10%, 11% and 13% in women age 20-29, 30-39 and 40-49) (*30*), and to previous population based case-control studies performed using the ABCFR (*255*). It is unlikely that this difference had an influence on overall results in the breast cancer study, as there was little evidence for effect modification by family history. Furthermore, where inconsistent associations were observed between the ABCFR and OFBCR, effect modification by family history could not explain the discrepancies.

As would be expected by the sampling method, probands in the OFBCR were more likely to have a *BRCA1* or *BRCA2* mutation, but the overall percentage of mutations (about 3% and 6% in the ABCFR and OFBCR) is fairly small. The number of *BRCA1* or *BRCA2* mutations among families classified as having a family history in chapter 5 was investigated further, since these families could be enriched for *BRCA1* and *BRCA2* mutations which could

potentially influence results of analyses of the association of *IGF1* with breast cancer risk stratified by family history. However, only 6 of 58 families in the ABCFR and 10 of 87 families in the OFBCR included a member with a *BRCA1* or *BRCA2* mutation carrier (all but one were found in affected individuals (probands) with breast cancer). Of particular interest in analyses stratified by family history was the association of the 5′ 19 allele with risk among probands with a family history of disease. Among the 65 informative families included in the analysis (see Table 8 chapter 5), only 8 included a family member with a *BRCA1* or *BRCA2* mutation (4 each in the ABCFR and OFBCR). Therefore, presence of *BRCA1* and *BRCA2* mutations among affected individuals in the ABCFR and OFBCR is unlikely to have influenced variation in results observed between these registries (a nominally significant inverse association of the 5′ 19 with risk in the ABCFR and no association in the OFBCR, see chapter 5).

Several other characteristics of probands related to breast cancer risk are compared across registries in Table 2. Probands were more likely to be Ashkenazi Jewish in the OFBCR but the percentage of Ashkenazi Jewish probands was small (3% in the OFBCR verses 1% in the ABCFR). Age at menarche, parity and age at birth of first child were similar between the ABCFR and OFBCR. Women in the ABCFR were more likely to have breast fed their children and attained higher education levels, but these differences were not large. Probands from the ABCFR also reported greater alcohol consumption but most (76% in the ABCFR and 90% in the OFBCR) reported having one drink a day or less. Few of these risk factors are thought to be related to IGF-I levels. A notable exception is oral contraceptive use, but oral contraceptive use was similar across the two populations and there was little evidence to indicate that oral contraceptive use modified risk (see chapter 5).

Other than the proportion of probands with a family history of disease, there were no striking differences in risk factor distributions between the ABCFR and OFBCR. It does not appear then that recruitment differences between the ABCFR and OFBCR (or for that matter differences in risk factors present across populations) could have an important influence on variation in results in the family based breast cancer study or the family based investigation of anthropometric measures (see chapter 6 for discussions of analyses including covariates in the family based study of *IGF1* and anthropometric measures). However, it of course cannot be ruled out that effect modification by unknown or unmeasured risk factors could have created discrepant results between the two registries.

Table 2. Descriptive characteristics for ABCFR and OFBCR families and probands.

	ABCFR	OFBCR
Number of Families	517	323
Age (S.D.) of diagnosis probands	37.3 (6.3)	42.2 (6.7)
Caucasian families	482 (93%)	303 (94%)
Affected sisters	5	5
Unaffected sisters	688	251
Family history[a]	58 (11%)	87 (27%)
BRCA1, Probands[b]	8 (1.5%)	11 (3.5%)
BRCA2, Probands[c]	8 (1.5%)	11 (3.5%)
Ashkenazi Jewish, Probands[d]	3 (1%)	11 (3.5%)
Mean age at menarche and (S.D.), Probands[e]	12.9 (1.6)	12.5 (1.4)
Mean age at first birth and (S.D.), Probands[f]	25.5 (4.6)	25.5(4.8)
Number of children, Probands		
0	129 (25%)	70(22%)
1-2	241 (47%)	176(54%)
3 or more	147 (28%)	77(24%)
Breast fed ever, Probands[g]	329 (66%)	187(58%)
Ever use of Oral contraceptives, Probands	489 (95%)	279(86%)
Alcoholic drinks per week, Probands[h]		
0	136(31%)	165(53%)
1-7	201(45%)	117(37%)
>7	106(24%)	31(10%)
Education, Probands[i]		
Up to High School	111(21%)	94(29%)
Vocational/Technical/Some college or University	280(54%)	124(39%)
University degree	126(24%)	101(32%)

All means and percentages calculated using all probands except as indicated below for d through i.
[a] Proband with mother or sister with breast cancer
[b,c] n=231 probands tested (ABCFR), n=209 probands tested
[d] n=349, unknown for pre-1996 probands (ABCFR)
[e] n=514 (ABCFR), n=322 (OFBCR)
[f] n=388 probands (ABCFR), n=252 (OFBCR)
[g] n=495 (ABCFR), n=322 (OFBCR)
[h] n= 443 (ABCFR), n=313 (OFBCR)
[i] n=319 probands (OFBCR)

7.4.1.5. Breast cancer rates in source populations

Sampling from both the ABCFR and OFBCR occurred from defined population based registries in New South Wales and Victoria (Australia) and Ontario (Canada). Breast cancer incidence rates from around the time of the study period were very similar, with rates ranging from 79.1 case per 100,000 in Ontario to 81.4 cases per 100,000 in Victoria (year range 1993 to 1997, standardized to the World Standard Population) across the three registries (*256*). Age specific rates show some variation between the Ontario and Australian populations. In the Australian provinces 5 year age specific breast cancer rates are 5% to 15% higher among women under the age of 55 (i.e. mainly premenopausal women). In older women rates are higher in Ontario, explaining the nearly identical age standardized rates between these three areas. These small differences in rates suggest differences in exposures to breast cancer related risk factors. However, considering the large variation in breast cancer rates observed internationally, these differences are small and do not give a strong indication that differences between populations influenced the interpretation of results in analyses that examined *IGF1* association with breast cancer risk or anthropometric measures (Chapter 5 and 6).

7.4.1.6. Discrepancies in results between cross-sectional and family based study

The 3′ 185 allele was observed to be associated with BMI and height in the cross-sectional study, but not the family based study. It seems unlikely that the differences in sampling schemes for the cross-sectional and family based studies should have contributed to the inconsistent results. Although a stratified sampling scheme based on mammographic density was employed in the cross-sectional study, the distribution of anthropometric

variables was similar to that observed in the ABCFR and OFBCR (see Table 2 in chapter 4 and Table 1 in chapter 6).

It is also unlikely that sampling through a diseased individual in the family based study contributed to the inconsistencies observed between the two study designs. Analyses that excluded women with breast cancer in the family based study did not have an appreciable influence on results when BMI or weight was set as outcome. Furthermore, it does not seem plausible that the major discrepancies observed between the cross-sectional and family based studies with respect to the association of the 3′ 185 allele with height can be explained by including women with breast cancer in the analysis of the family based sample (e.g. the association in the ABCFR although not significant (P=0.10) was in the opposite direction to that of the cross-sectional study).

Allele frequencies of the 3′ 185 allele was somewhat lower in the cross-sectional study compared to the ABCFR and OFBCR (43%, 46% and 49% in the cross-sectional, ABCFR and OFBCR samples respectively). These differences are, however, small. In fact, a comparison of 3′ 185 allele frequencies between the three samples, restricting ABCFR and OFBCR subjects to one randomly selected unaffected subject from each Caucasian family, did not indicate these differences were statistically significant (allele frequencies in the ABCFR and OFBCR for unaffected Caucasians were 45% and 50% respectively). As stated earlier, the cross-sectional study was susceptible to confounding due to population stratification, but based on the similarity of these allele frequencies among these populations, this may not have been an issue.

7.4.2. Measurement error

Measurement error for the main outcomes was discussed in chapter 3. Further discussion here emphasizes their importance in the interpretation of the results.

7.4.2.1. Genotyping

Genotyping error should not be an important source of error in the cross-sectional study since concordance rates on samples used for reliability was 100%. Although there could still be undetected errors, this would result in a minimal bias towards the null.

Genotyping error can introduce bias in family based association tests, as recently demonstrated in a study evaluating the use of the TDT test with multi-allelic markers. Specifically, genotyping errors can result in common alleles being incorrectly associated with increased risk and rare alleles with decreased risk (*243*). The reason for this is that error is most likely to occur in the most common allele and in parents as opposed to the offspring (parents outnumber offspring two to one). As a result, common alleles will appear less frequently than they should in parents relative to offspring. Therefore, these alleles may appear to be over transmitted relative to less common alleles. Similarly, rare alleles will appear to be over represented among parents and underrepresented in affected individuals due to genotyping error in common alleles (*243*).

The extent of this bias is greatly influenced by the extent of genotyping errors, which is reflected in the concordance rates of the family based study. Concordance rates for the 5′, intron 2 and 3′ polymorphisms were 97%, 94% and 97% respectively in the ABCFR and 99%, 98%, and 100% respectively in the OFBCR. The nominally significant association of the common intron 2 216 allele with breast cancer risk is the most likely candidate for a false positive result, since concordance was lowest for the intron 2 polymorphism. An error rate of

about 10% to 20% would be expected for the association of this allele with breast cancer risk in the ABCFR, if families were in fact all trios and a TDT test was performed (*243*). However, under the same assumptions, the false positive rate in the OFBCR, where in addition to the combined sample this allele showed a nominally significant association with risk, would only be about 5% to 7%.

In statistical tests using FBAT and PBAT, some family configurations will not be susceptible to this bias, meaning there is less potential for bias than indicated above. For example, genotyping error is equally probable in a pair of siblings, so families with an affected and unaffected sib will not show distorted transmission rates due to genotyping error. Since the potential for bias appears limited, particularly in the OFBCR where the significant association with risk was found, it is unlikely that genotyping errors would have influenced the interpretation of this study.

Unlike disease status, there is no reason to expect that genotyping error could produce an apparent over or under transmission of alleles to individuals with a specific BMI, weight or height unless there was some association of these with disease outcome. However, since no meaningful associations were identified in chapter 6, this is not an issue here.

Genotyping error is even more important in analyses using haplotypes, because there are more sources of error (two in the studies conducted here, since haplotypes in these studies were constructed using two loci). An interesting discrepancy in the results in haplotype analyses was the strong association of the 5′ 19 and intron 2 216 haplotype in the OFBCR with breast cancer risk under a recessive model, which was not supported by the results from the ABCFR. However, bias due to genotyping error is unlikely to have resulted in such incongruent findings, as the observed discrepancy in the results was too large. Furthermore,

association testing using the ABCFR sample, where the null result was observed, would have been more prone to producing a false positive finding because genotyping error was greater in this sample.

7.4.2.2. IGF-I and IGFBP-3

The RIA method which was used here is considered by some to be the gold standard for IGFBP-3 measurement (*156*). The RIA method is also a standard method for measuring IGF-I, and shows good correlation with other methods of IGF-I measurement (*257*).

Measurement error of IGF-I or IGFBP-3 could account for undetected genotype and phenotype associations in this study. Of particular interest here, is the lack of a relationship of the 3′ 185 allele with IGF-I levels, as this allele showed an association with mammographic density but inconsistent with the proposed model, not IGF-I levels. It appears unlikely, however, that error in measuring IGF-I levels would have resulted in such a discrepancy. Measurement of IGF-I levels was accurate enough to show an association with mammographic density in a previous study that included this sample and some additional subjects (*7*). IGF-I was also significantly associated with percentage density, after adjustment for age (results of analysis not shown) or age and IGFBP-3 levels (see chapter 4), using only the study subjects from chapter 4. Furthermore, other studies using comparable methods of measuring IGF-I have observed this same association (*5,6,258*).

7.4.2.3. Mammographic density

The computer assisted method used to assess mammographic density has been shown to produce similar gradients in risk estimates for breast cancer as those produced by categorization of mammographic density by radiologists. Reliability of measurement using

this method was high (0.90 or greater) in the study reported on in chapter 4, and similar estimates have been reported in other studies (*46,209*). In addition, this method of measuring mammographic density has produced high reproducibility in risk estimates for breast cancer and breast density heritability estimates (*10,46,58*). This suggests that at most, strength of association with mammographic density with variant *IGF1* alleles should only be affected slightly by error in measurement. However, bias should be towards the null and not produce false positive results.

7.4.2.4. Anthropometric measures

The cross-sectional study described in chapter 4 employed trained staff to measure height and weight. However, these measures were self-reported in the family based study described in chapter 6. The correlation between self-reported height and weight and actual measures of these same variables is high ($r > 0.9$) (*211*). Many studies, however, report that women generally overestimate height and underestimate weight (*212-215*), and this results in an underestimate of BMI (*214*). Furthermore, in women, increased weight is associated with increased error in reporting height and weight, therefore measurement error of height, weight and BMI will often be greatest in heavier women (*214,215*).

In general, the effect of this measurement error would be to create a small bias towards the null, and overall, the impact on study results should be minimal. Of specific interest is the failure to replicate the association of the 3′ 185 allele with BMI and height from the cross-sectional study, in analyses using the family based sample. It does not seem plausible, however, that this discrepancy in results can be attributed to error resulting from using self reported height and weight data. In the family based study, results relating the 3′ 185 allele genotype to BMI and weight were far from achieving nominal significance and the observed

direction of effect of the 3′ 185 allele with height in the ABCFR, although not statistically significant (P=0.09), was in the opposite direction to the statistically significant association reported in the cross-sectional study.

7.4.2.5. Breast cancer

Errors in measurement the accuracy of breast cancer diagnosis should not be an issue here, as all were verified by pathology report.

7.4.2.6. Density of polymorphisms

The 5′ CA repeat polymorphism is located in the untranslated region of the *IGF1* promoter. There is however, no experimental evidence to indicate that this repeat is functional (*189*). There is also no evidence supporting a functional role for the intron 2 or 3′ CA polymorphism, the latter of which lies in the 3′ untranslated region and to the authors knowledge has never been investigated. The selection of these polymorphisms was not, however, based on a hypothesis that they were functional, but on the possibility of their being in linkage disequilibrium with other ungenotyped functional variants in the gene.

Gabriel et al., recently reported a block like pattern of linkage disequilibrium across the human genome, with areas of low linkage disequilibrium interspersed with areas of high linkage disequilibrium (*171*). As described earlier, analysis of SNP data from the International HapMap Project (*196*) shows a block like pattern of linkage disequilibrium across *IGF1* (see chapter 2). In order to effectively capture genetic variation across *IGF1*, a large number of polymorphisms would have to be genotyped. This study investigated three CA repeat polymorphisms, which were in relatively weak linkage disequilibrium as no R^2 greater than 0.27 was observed in either sample (see Table 8 in chapter 4 and Table 5 in

chapter 5). An important limitation of this study then is that the few markers used captured only some of the genetic variation across *IGF1*. This implies that analyses testing individual polymorphisms or haplotypes may fail to capture an ungenotyped causal variant if present.

7.5. Power

In the family based studies, associations with breast cancer and anthropometric measures were not significant after adjustment for multiple comparisons. Therefore, a discussion of the power to detect an association in these two studies is of particular interest. Unfortunately, the PBAT software cannot produce power estimates for the family design employed in chapter 6. However, a primary objective of the family based study was to attempt to replicate the association of the 3′ 185 allele with BMI and height observed in the cross-sectional study. Given that there were over 800 families used in this analysis, there should have been ample power to replicate this association.

Power can be estimated for the association of variant *IGF1* alleles with breast cancer. The nominally significant association of the intron 2 216 allele (P=0.04) with risk was highlighted previously as being of particular interest. In order to illustrate the impact of the adjustment for multiple comparisons, power calculations for the intron 2 216 allele were performed using a relative risk of 2 for homozygotes and 1.5 for heterozygotes, with a sample size of 808 families, an allele frequency of 0.38 (both sample size and allele frequencies were obtained from the results in chapter 5) and assuming an additive model with complete linkage disequilibrium between the intron 2 216 allele and the putative causal variant. The power to detect a significant association was 0.89 with the P-value set to 0.05 and 0.68 with the P-value set to 0.007 (the latter being the adjusted P-value after implementing the Benjamini-Hochberg correction). There is then, as would be expected, a

substantial drop in power after applying a correction for multiple comparisons. Despite this loss in power, a formal approach to dealing with multiple comparisons was considered necessary in the family based study because of the many comparisons made. However, throughout this thesis, consistency of results was also stressed as a criterion for evaluating the importance of possible associations. Since the intron 2 216 allele was not strongly associated with risk in the ABCFR (under either an additive or recessive model) this association was not considered to be convincing. It is acknowledged, however, that comparing results across samples also has limitations, since there is reduced power to detect associations because of the smaller sample sizes.

This breast cancer study had the power to detect relative risks that approached 2.0 and 1.5 for homozygotes and heterozygotes respectively. As discussed in chapter 3, power to detect an association was relatively high compared to most other studies reporting on premenopausal women. However, complete linkage disequilibrium between the marker allele and the putative disease locus was assumed in power calculations. Power decreases with a reduction in the strength of linkage disequilibrium between markers and as mentioned above, using a limited number of markers could have an important impact on study power.

Recent genome wide scans have so far only detected susceptibility alleles that increase breast cancer risk by 50% or less (comparing homozygotes for low and high risk alleles) (*38,39,197*). As well, an association of similar magnitude is suggested by the results of a previous study investigating *IGF1* and breast cancer risk (*187*). Limitations of these studies with respect to *IGF1* and breast cancer risk were discussed previously and are discussed again further below. However, in addition to the discussion above, these results also suggest

that power in the current investigation could be limiting, despite the approximately 800 families used, even if more markers had been genotyped.

7.6. Final results interpreted in context with other studies

Table 3 shows the summary of results previously described in Table 1, next to major findings from the literature that were reviewed in previous chapters. Excluded from Table 3 are results presented earlier for haplotype analyses. These were previously discussed and will not be reviewed further here. The comparison of findings related to *IGF1* and anthropometric variables shown in Table 3 is also not discussed further (the reader is instead referred to chapter 6).

The following discussion compares and contrasts the most relevant results from this study to those in the published literature. General support for both the proposed model and specific aspects of the model (e.g. association of *IGF1* and mammographic density) is discussed, as are the most important limitations in studies that examined these associations.

Table 3. Summary of main findings compared to those in the published literature.

Original Objective	Main findings from chapter 4-6	Overview of findings from published studies
Examine the association of each of the polymorphisms with circulating IGF-I levels.	Greater number of copies of the 5′ 19 allele were associated with lower IGF-I levels.	Among women in general, results are inconsistent and contradictory (*172,175,177,179,181,182*) for the 5′ 19 allele. No significant associations reported in premenopausal women (*99,174*). An association of several *IGF1* tagging SNPs was observed in a British study (*187*).
Examine the association of each of the polymorphisms with percentage breast density.	Greater number of copies of the 3′ 185 allele were associated with greater percentage breast density and a smaller amount of non-dense (fat) tissue. Analyses suggested the association of the 3′ 185 allele with percentage breast density is mediated through BMI (i.e. body fat).	The 5′ 19 allele was not associated with percentage density in premenopausal women from Toronto (*99*). A strong association for *IGF1* haplotypes and several tagging SNPs with percentage breast density was found in the Nurses Health Study cohort (*100*).
Examine the association of each of the polymorphisms with breast cancer risk	*Primary analyses:* Some nominally significant associations were observed (5′ 21 allele, intron 2 212 allele, intron 216 allele) but these were not significant after adjustment for multiple comparisons and there was a lack of consistency between the ABCFR and OFBCR. *Exploratory analyses:* Associations were observed for the intron 2 216 allele under a recessive model and 5′ allele groupings resulted in a positive association for alleles of length 18 to 20, and a protective effect for alleles > than 20 repeats in length. These associations were not consistent across the ABCFR and OFBCR. Overall, there was limited evidence to support an association of allelic variants of the tested polymorphisms with breast cancer risk.	No association for the 5′ 19 allele in Caucasian populations (*172,179,190,191,193*) (including studies where young (under 50) (*190*) or premenopausal women (*193*) were examined). Positive association of the 5′ 19 allele with risk in a sample of mainly Black subjects (*192*) and in a Chinese population (*189*). Significant association reported for tagging SNPs at the 5′ end of the gene in two U.S. studies of primarily postmenopausal women (*194,195*)*4*). An association of several *IGF1* tagging SNPs with risk under a recessive model was reported in a sample of primarily premenopausal women from Great Britain (*187*).
Examine the association between genetic variation at *IGF1* and BMI, height and weight.	A significant association between the 3′ 185 allele and BMI and height was found in the cross-sectional study. An association of genetic variation at *IGF1* and either BMI, height or weight was not supported in the family based study.	An association with BMI was reported for 5′ 19 or 20 allele homozygotes or carriers in one of two cohorts investigated (*248*). No association of the 5′ polymorphism and BMI was reported in two other studies (*189,249*). Association of 5′ 19 or 20 allele carriers with weight reported in two of three Dutch cohorts (*186,248*). No association for 5′ 19 or 20 alleles with height (*186*).

7.6.1. *IGF1*, IGF-I concentration, mammographic density, and breast cancer risk

The results presented in Table 3 indicate that there is little evidence to support an association of the 5′ 19 allele with either IGF-I levels, mammographic density or breast cancer risk. In this thesis, evidence for an association of the 5′ 19 allele with IGF-I levels was presented, but this was not supported by other results in the published literature (Table 3).

Several studies have examined the association of *IGF1* tagging SNPs with one or more of these outcomes. A comparison of findings from these studies with those presented from chapters 4 through 6 provides some support for consistency of associations within outcomes. The observed association of the 3′ 185 allele with percentage breast density is consistent with results from the Nurses Health Study where percentage density was strongly associated with several *IGF1* haplotypes and tagging SNPs (Table 3). In addition, the association of the intron 2 216 allele with risk under a recessive model is consistent with results from a British study that examined the association of tagging SNPs with risk also under a recessive model. Similarly, the association of 5′ allele length with breast cancer risk can be viewed as consistent with two studies indicating that SNPs at the 5′ end of *IGF1* are associated with risk (Table 3).

These observations, however, require further qualification. The association of the 3′ 185 was found to be mediated through BMI, which was not consistent with results from the Nurses Health Study. As well, the associations of both medium length and long alleles of the 5′ polymorphism and the intron 2 216 allele with risk reported in chapter 5, were not strongly supported in both the ABCFR and OFBCR. Knowledge of linkage disequilibrium patterns between the polymorphisms investigated here and those in the published literature

would permit a more meaningful comparison of results. However, these data are not available.

In interpreting the results from the literature, it is important to note that there are inconsistencies both within and across outcomes among the studies that have used tagging SNPs. The SNP most strongly associated with mammographic density in the Nurses Health Study was also found be associated with IGF-I levels and breast cancer risk in a British study, but with opposing effects to what would be anticipated if the same allele was responsible for increased IGF-I levels, mammographic density and breast cancer risk (*187*). In addition, a study that examined the same genetic variants as those in the Nurses Health Study did not find an association with breast cancer risk (*195*). Inconsistent results also have been reported across studies where the same tagging SNP was investigated in association with breast cancer risk (*187,194*).

Not shown in Table 3 are the results of recent genome wide association studies which would have included *IGF1* tagging SNPs among the many SNPS they tested in association with breast cancer risk. These have not reported significant associations between *IGF1* and risk. However, true associations could have been missed particularly in premenopausal women, a group in which large samples have not been included during initial screens in these studies. As well, stringent methods employed in adjusting for multiple comparisons may have resulted in some associations being missed (*38,39,197*).

Currently, no definitive association between *IGF1* and either IGF-I levels, mammographic density or breast cancer risk can be said to have been identified. There are some intriguing results (e.g. association of *IGF1* with mammographic density) but additional studies are needed to replicate and further explore these associations, particularly in light of

the inconsistent and contradictory results observed across outcomes. Furthermore, inconsistent with the proposed model no single variant *IGF1* allele from any polymorphism has been observed to be associated with all outcomes (i.e., IGF-I levels, mammographic density and breast cancer risk). It is, however, premature to conclude that *IGF1* does not play a role in the model as outlined, as limitations of the studies discussed above may have contributed to inconsistencies in results and failures to detect associations.

Several limitations to studies that have examined allelic variation at *IGF1* with these outcomes were previously highlighted as being particularly important. One shared by many studies is the failure to capture genetic variation across *IGF1*. This is particularly relevant for studies that examined only the 5′ polymorphism. An effort was made in the current investigation to address this issue by including three polymorphisms, although as discussed earlier, failure to thoroughly capture genetic variation across *IGF1* was also a limitation of this study. The use of tagging SNPs is an approach that addresses this problem but only a few studies have assessed genetic variation at *IGF1* in this manner and notably only two of these have genotyped the approximately 25 SNPs needed to provide thorough coverage of the gene (see chapter 2) *(100,195).* Confounding due to population stratification is another important potential source of error. Only in the current investigation has a sample of women been investigated (specifically for breast cancer) using a family based design. As well, most of these studies reported on samples of pre- and postmenopausal women. Possible dependency of associations on menopausal status could be complicating comparisons across studies. In addition, results from recent genome wide association studies and research specifically focusing on *IGF1* suggest that risk from breast cancer susceptibility alleles are modest (increase in risk of less than 50% comparing homozygotes for high and low risk alleles).

Therefore, at least for studies where breast cancer was the outcome, sample size may often have been a limiting factor.

7.7. Conclusions and future research

As described above, a model of breast cancer risk was examined where variant alleles of *IGF1* are proposed to modify circulating IGF-I levels, with greater IGF-I levels promoting the proliferative activity and quantity of stromal and epithelial tissue in the breast, resulting in greater mammographic density and an increased risk of developing breast cancer. The results did not indicate that a single variant *IGF1* allele was associated with all three of these phenotypes as outlined in the proposed model. There was some support for an association of the 3′ 185 allele with mammographic density, but this was not mediated through IGF-I levels, which were instead associated with the 5′ 19 allele. As well, inconsistent with the proposed model, the positive association of the 3′ 185 allele with percentage density did not appear to result from an influence on dense tissue, but instead was mediated through a possible effect on body fat stores. The failure to replicate the association of the 3′ 185 allele with BMI in the family based study indicates that the association with BMI observed in the cross-sectional study is spurious, and therefore suggests that the association with mammographic density is a false positive result.

Examining the results in the context of the latter part of the proposed model, there was also no indication that a common *IGF1* variant was associated with both mammographic density and breast cancer risk. In fact, overall, evidence supporting an association of variant *IGF1* alleles with breast cancer risk was limited.

An important limitation of this study and many others is that few polymorphisms along *IGF1* were tested, and therefore genetic variation across the gene was not captured, reducing

power to detect an association. As well, this study did not have the power to detect small increases in risk that have been reported for genes in, for example, some recent genome wide scans. Studies that have examined multiple SNPs using large samples have provided some support for the association of variant *IGF1* alleles and IGF-I levels (*187*), mammographic density (*100*) and breast cancer risk (*187,194,195*). However, there are inconsistencies across studies, both when considering associations with specific outcome such as breast cancer risk (*187,194,195*) and when considering associations across outcomes, such as the opposing effects of alleles of the same SNP on IGF-I levels, breast cancer risk and mammographic density (*187*).

Further research is needed to determine whether genetic variation at *IGF1* is associated with breast cancer risk, and if so, whether risk might be influenced through a pathway that modifies circulating IGF-I levels and breast density. The role that *IGF1* may play in influencing body composition and the possible relationship with this proposed pathway is also of interest. These relationships should be thoroughly investigated in sufficiently large samples of premenopausal women, since it is in this group that the association of IGF-I levels with mammographic density and breast cancer risk has been observed. As well, future efforts should employ a tagging SNP approach to better capture genetic variation across *IGF1* and ensure efforts are made to control for potential confounding due to population stratification (e.g., genomic control). There are also a number of other factors, including genetic factors, that might influence IGF-I levels and possibly mammographic density and breast cancer risk. These include genes that regulate growth hormone release such as growth hormone, growth hormone releasing hormone, grehlin and somatostatin genes. Genes that code for IGF-I binding proteins such as IGFBP-1 through IGFBP-6 are also of interest. Of these, only

IGFBP-3 has received much attention. The influence of these genes on IGF-I levels and mammographic density and breast cancer risk both directly and through gene-gene interaction should be investigated. As well, extensive genome wide association approaches only have been employed in studies where breast cancer risk was the outcome. Application of this approach to include intermediate variables such as circulating IGF-I levels and mammographic density may provide important insights into breast cancer etiology.

References

(*1*) Shi R, Yu H, McLarty J, Glass J. IGF-I and breast cancer: a meta-analysis. Int J Cancer 2004;111:418-23.

(*2*) Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet 2004;363:1346-53.

(*3*) Renehan AG, Harvie M, Howell A. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and breast cancer risk: eight years on. Endocr Relat Cancer 2006;13:273-8.

(*4*) Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6:798-808.

(*5*) Diorio C, Pollak M, Byrne C, Masse B, Hebert-Croteau N, Yaffe M, et al. Insulin-like growth factor-I, IGF-binding protein-3, and mammographic breast density. Cancer Epidemiol Biomarkers Prev 2005;14:1065-73.

(*6*) Maskarinec G, Williams AE, Kaaks R. A cross-sectional investigation of breast density and insulin-like growth factor I. Int J Cancer 2003;107:991-6.

(*7*) Boyd NF, Stone J, Martin LJ, Jong R, Fishell E, Yaffe M, et al. The association of breast mitogens with mammographic densities. Br J Cancer 2002;87:876-82.

(*8*) Byrne C, Colditz GA, Willett WC, Speizer FE, Pollak M, Hankinson SE. Plasma insulin-like growth factor (IGF) I, IGF-binding protein 3, and mammographic density. Cancer Res 2000;60:3744-8.

(*9*) Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.

(*10*) Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med 2002;347:886-94.

(*11*) Harrela M, Koistinen H, Kaprio J, Lehtovirta M, Tuomilehto J, Eriksson J, et al. Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. J Clin Invest 1996;98:2612-5.

(*12*) Hong Y, Pedersen N, Brismar K, Hall K, de Faire U. Quantitative genetic analyses of insulin-like growth factor I (IGF-I), IGF-binding protein-1, and insulin levels in middle-aged and elderly twins. J Clin Endocrinol Metab 1996;81:1791-7.

(13) Cancer Surveillance On-Line. Public Health Agency of Canada; 2007. Available at: http://dsol-smed.phac-aspc.gc.ca/dsol-smed/cancer/index.html. [updated 2002-01-21] [accessed 2007-02-25].

(*14*) Veronesi U, Boyle P, Goldhirsch A, Orecchia R, Viale G. Breast cancer. Lancet 2005;365:1727-41.

(*15*) Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15:17-35.

(*16*) Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Annu Rev Public Health 1996;17:47-67.

(*17*) Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiol Rev 1993;15:36-47.

(*18*) Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet 2002;360:187-95.

(*19*) Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. J Mammary Gland Biol Neoplasia 2002;7:3-15.

(*20*) Key TJ. Serum oestradiol and breast cancer risk. Endocr Relat Cancer 1999;6:175-80.

(21) Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53297 women with breast cancer and 100239 women without breast cancer from 54 epidemiological studies. Lancet 1996;347:1713-27.

(*22*) Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. Jama 2002;288:321-33.

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

(*24*) Ross RK, Paganini-Hill A, Wan PC, Pike MC. Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. J Natl Cancer Inst 2000;92:328-32.

(*25*) Schairer C, Lubin J, Troisi R, Sturgeon S, Brinton L, Hoover R. Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk. Jama 2000;283:485-91.

(*26*) Food, nutrition and the Prevention of Cancer: a global perspective. World Cancer Research Fund / American Institute for Cancer Research; 1997.

(*27*) Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. Diet, nutrition and the prevention of cancer. Public Health Nutr 2004;7:187-200.

(*28*) Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., 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;87:1234-45.

(*29*) Oza AM, Boyd NF. Mammographic parenchymal patterns: a marker of breast cancer risk. Epidemiol Rev 1993;15:196-208.

*(*30*)* Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358:1389-99.

(*31*) Thompson D, Easton D. The genetic epidemiology of breast cancer genes. J Mammary Gland Biol Neoplasia 2004;9:221-36.

(*32*) Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72:1117-30.

(*33*) Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst 2005;97:813-22.

(34) CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. Am J Hum Genet 2004;74:1175-82.

(*35*) de Jong MM, Nolte IM, te Meerman GJ, van der Graaf WT, Oosterwijk JC, Kleibeuker JH, et al. Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. J Med Genet 2002;39:225-42.

(*36*) Varley J. TP53, hChk2, and the Li-Fraumeni syndrome. Methods Mol Biol 2003;222:117-29.

(*37*) Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007;39:352-8.

(*38*) Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870-4.

(*39*) Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087-93.

(*40*) Wolfe JN. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol 1976;126:1130-7.

(*41*) Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer 1976;37:2486-92.

(*42*) Boyd NF, Lockwood GA, Martin LJ, Knight JA, Byng JW, Yaffe MJ, et al. Mammographic densities and breast cancer risk. Breast Dis 1998;10:113-26.

(*43*) Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87:1622-9.

(*44*) Ursin G, Ma H, Wu AH, Bernstein L, Salane M, Parisky YR, et al. Mammographic density and breast cancer in three ethnic groups. Cancer Epidemiol Biomarkers Prev 2003;12:332-8.

(*45*) Torres-Mejia G, De Stavola B, Allen DS, Perez-Gavilan JJ, Ferreira JM, Fentiman IS, et al. Mammographic features and subsequent risk of breast cancer: a comparison of qualitative and quantitative evaluations in the Guernsey prospective studies. Cancer Epidemiol Biomarkers Prev 2005;14:1052-9.

(*46*) Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 1995;87:670-5.

(*47*) Boyd NF, O'Sullivan B, Campbell JE, Fishell E, Simor I, Cooke G, et al. Mammographic signs as risk factors for breast cancer. Br J Cancer 1982;45:185-93.

(*48*) Brisson J, Merletti F, Sadowsky NL, Twaddle JA, Morrison AS, Cole P. Mammographic features of the breast and breast cancer risk. Am J Epidemiol 1982;115:428-37.

(*49*) Brisson J, Morrison AS, Kopans DB, Sadowsky NL, Kalisher L, Twaddle JA, et al. Height and weight, mammographic features of breast tissue, and breast cancer risk. Am J Epidemiol 1984;119:371-81.

(*50*) Brisson J, Verreault R, Morrison AS, Tennina S, Meyer F. Diet, mammographic features of breast tissue, and breast cancer risk. Am J Epidemiol 1989;130:14-24.

(*51*) Maskarinec G, Meng L. A case-control study of mammographic densities in Hawaii. Breast Cancer Res Treat 2000;63:153-61.

(*52*) Thomas DB, Carter RA, Bush WH, Jr., Ray RM, Stanford JL, Lehman CD, et al. Risk of subsequent breast cancer in relation to characteristics of screening mammograms from women less than 50 years of age. Cancer Epidemiol Biomarkers Prev 2002;11:565-71.

(*53*) van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? Eur J Epidemiol 1998;14:315-20.

(*54*) van Gils CH, Hendriks JH, Otten JD, Holland R, Verbeek AL. Parity and mammographic breast density in relation to breast cancer risk: indication of interaction. Eur J Cancer Prev 2000;9:105-11.

(*55*) Wolfe JN, Saftlas AF, Salane M. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. AJR Am J Roentgenol 1987;148:1087-92.

(*56*) Kato I, Beinart C, Bleich A, Su S, Kim M, Toniolo PG. A nested case-control study of mammographic patterns, breast volume, and breast cancer (New York City, NY, United States). Cancer Causes Control 1995;6:431-8.

(*57*) Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, et al. Mammographic densities and risk of breast cancer. Cancer 1991;67:2833-8.

(*58*) Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356:227-36.

(*59*) Masala G, Ambrogetti D, Assedi M, Giorgi D, Del Turco MR, Palli D. Dietary and lifestyle determinants of mammographic breast density. A longitudinal study in a Mediterranean population. Int J Cancer 2006;118:1782-9.

(*60*) Maskarinec G, Pagano I, Lurie G, Kolonel LN. A longitudinal investigation of mammographic density: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 2006;15:732-9.

(*61*) Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Res 2005;7:R854-61.

(*62*) Heng D, Gao F, Jong R, Fishell E, Yaffe M, Martin L, et al. Risk factors for breast cancer associated with mammographic features in Singaporean chinese women. Cancer Epidemiol Biomarkers Prev 2004;13:1751-8.

(*63*) Jeffreys M, Warren R, Gunnell D, McCarron P, Smith GD. Life course breast cancer risk factors and adult breast density (United Kingdom). Cancer Causes Control 2004;15:947-55.

(*64*) Gapstur SM, Lopez P, Colangelo LA, Wolfman J, Van Horn L, Hendrick RE. Associations of breast cancer risk factors with breast density in Hispanic women. Cancer Epidemiol Biomarkers Prev 2003;12:1074-80.

(*65*) El-Bastawissi AY, White E, Mandelson MT, Taplin SH. Reproductive and hormonal factors associated with mammographic breast density by age (United States). Cancer Causes Control 2000;11:955-63.

(*66*) Stomper PC, D'Souza DJ, DiNitto PA, Arredondo MA. Analysis of parenchymal density on mammograms in 1353 women 25-79 years old. AJR Am J Roentgenol 1996;167:1261-5.

(*67*) Bartow SA, Pathak DR, Mettler FA, Key CR, Pike MC. Breast mammographic pattern: a concatenation of confounding and breast cancer risk factors. Am J Epidemiol 1995;142:813-9.

(*68*) Boyd NF, Connelly P, Byng J, Yaffe M, Draper H, Little L, et al. Plasma lipids, lipoproteins, and mammographic densities. Cancer Epidemiol Biomarkers Prev 1995;4:727-33.

(*69*) Grove JS, Goodman MJ, Gilbert FI, Jr., Mi MP. Factors associated with mammographic pattern. Br J Radiol 1985;58:21-5.

(*70*) Brisson J, Sadowsky NL, Twaddle JA, Morrison AS, Cole P, Merletti F. The relation of mammographic features of the breast to breast cancer risk factors. Am J Epidemiol 1982;115:438-43.

(*71*) Marchesoni D, Driul L, Ianni A, Fabiani G, Della Martina M, Zuiani C, et al. Postmenopausal hormone therapy and mammographic breast density. Maturitas 2006;53:59-64.

(*72*) Warren R. Hormones and mammographic breast density. Maturitas 2004;49:67-78.

(*73*) Cuzick J, Warwick J, Pinney E, Warren RM, Duffy SW. Tamoxifen and breast density in women at increased risk of breast cancer. J Natl Cancer Inst 2004;96:621-8.

(*74*) Brisson J, Brisson B, Cote G, Maunsell E, Berube S, Robert J. Tamoxifen and mammographic breast densities. Cancer Epidemiol Biomarkers Prev 2000;9:911-5.

(*75*) Atkinson C, Warren R, Bingham SA, Day NE. Mammographic patterns as a predictive biomarker of breast cancer risk: effect of tamoxifen. Cancer Epidemiol Biomarkers Prev 1999;8:863-6.

(*76*) Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. Cancer 2000;89:369-75.

(*77*) Vachon CM, Kuni CC, Anderson K, Anderson VE, Sellers TA. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). Cancer Causes Control 2000;11:653-62.

(*78*) Sala E, Warren R, McCann J, Duffy S, Luben R, Day N. High-risk mammographic parenchymal patterns and anthropometric measures: a case-control study. Br J Cancer 1999;81:1257-61.

(*79*) Boyd NF, Lockwood GA, Byng JW, Little LE, Yaffe MJ, Tritchler DL. The relationship of anthropometric measures to radiological features of the breast in premenopausal women. Br J Cancer 1998;78:1233-8.

(*80*) de Stavola BL, Gravelle IH, Wang DY, Allen DS, Bulbrook RD, Fentiman IS, et al. Relationship of mammographic parenchymal patterns with breast cancer risk factors and risk of breast cancer in a prospective study. Int J Epidemiol 1990;19:247-54.

(*81*) Gravelle IH, Bulstrode JC, Bulbrook RD, Hayward JL, Wang DY. The relation between radiological patterns of the breast and body weight and height. Br J Radiol 1982;55:23-5.

(*82*) Boyd N, Martin L, Stone J, Little L, Minkin S, Yaffe M. A longitudinal study of the effects of menopause on mammographic features. Cancer Epidemiol Biomarkers Prev 2002;11:1048-53.

(*83*) Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G, et al. Body size, mammographic density, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2006;15:2086-92.

(*84*) Maskarinec G, Takata Y, Pagano I, Lurie G, Wilkens LR, Kolonel LN. Alcohol consumption and mammographic density in a multiethnic population. Int J Cancer 2006;118:2579-83.

(*85*) Vachon CM, Kushi LH, Cerhan JR, Kuni CC, Sellers TA. Association of diet and mammographic breast density in the Minnesota breast cancer family cohort. Cancer Epidemiol Biomarkers Prev 2000;9:151-60.

(*86*) Herrinton LJ, Saftlas AF, Stanford JL, Brinton LA, Wolfe JN. Do alcohol intake and mammographic densities interact in regard to the risk of breast cancer? Cancer 1993;71:3029-35.

(*87*) Boyd NF, McGuire V, Fishell E, Kuriov V, Lockwood G, Tritchler D. Plasma lipids in premenopausal women with mammographic dysplasia. Br J Cancer 1989;59:766-71.

(*88*) Nagata C, Matsubara T, Fujita H, Nagao Y, Shibuya C, Kashiki Y, et al. Associations of mammographic density with dietary factors in Japanese women. Cancer Epidemiol Biomarkers Prev 2005;14:2877-80.

(*89*) Sala E, Warren R, Duffy S, Welch A, Luben R, Day N. High risk mammographic parenchymal patterns and diet: a case-control study. Br J Cancer 2000;83:121-6.

(*90*) Nordevang E, Azavedo E, Svane G, Nilsson B, Holm LE. Dietary habits and mammographic patterns in patients with breast cancer. Breast Cancer Res Treat 1993;26:207-15.

(*91*) Boyd NF, Greenberg C, Lockwood G, Little L, Martin L, Byng J, et al. Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. Canadian Diet and Breast Cancer Prevention Study Group. J Natl Cancer Inst 1997;89:488-96.

(*92*) Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. Behav Genet 1997;27:325-51.

(*93*) van Duijnhoven FJ, Bezemer ID, Peeters PH, Roest M, Uitterlinden AG, Grobbee DE, et al. Polymorphisms in the estrogen receptor {alpha} gene and mammographic density. Cancer Epidemiol Biomarkers Prev 2005;14:2655-60.

(*94*) Haiman CA, Hankinson SE, De Vivo I, Guillemette C, Ishibe N, Hunter DJ, et al. Polymorphisms in steroid hormone pathway genes and mammographic density. Breast Cancer Res Treat 2003;77:27-36.

(*95*) Haiman CA, Bernstein L, Berg D, Ingles SA, Salane M, Ursin G. Genetic determinants of mammographic density. Breast Cancer Res 2002;4:R5.

(*96*) Hong CC, Thompson HJ, Jiang C, Hammond GL, Tritchler D, Yaffe M, et al. Val158Met Polymorphism in catechol-O-methyltransferase gene associated with risk factors for breast cancer. Cancer Epidemiol Biomarkers Prev 2003;12:838-47.

(*97*) Maskarinec G, Lurie G, Williams AE, Le Marchand L. An investigation of mammographic density and gene variants in healthy women. Int J Cancer 2004;112:683-8.

(*98*) Mulhall C, Hegele RA, Cao H, Tritchler D, Yaffe M, Boyd NF. Pituitary growth hormone and growth hormone-releasing hormone receptor genes and associations with mammographic measures and serum growth hormone. Cancer Epidemiol Biomarkers Prev 2005;14:2648-54.

(*99*) Lai JH, Vesprini D, Zhang W, Yaffe MJ, Pollak M, Narod SA. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev 2004;13:573-82.

(*100*) Tamimi RM, Cox DG, Kraft P, Pollak MN, Haiman CA, Cheng I, et al. Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study. Breast Cancer Res 2007;9:R18.

(*101*) Pollak MN, Schernhammer ES, Hankinson SE. Insulin-like growth factors and neoplasia. Nat Rev Cancer 2004;4:505-18.

(*102*) Cohen P, Clemmons DR, Rosenfeld RG. Does the GH-IGF axis play a role in cancer pathogenesis? Growth Horm IGF Res 2000;10:297-305.

(*103*) Yang XF, Beamer WG, Huynh H, Pollak M. Reduced growth of human breast cancer xenografts in hosts homozygous for the lit mutation. Cancer Res 1996;56:1509-11.

(*104*) Pollak MN, Huynh HT, Lefebvre SP. Tamoxifen reduces serum insulin-like growth factor I (IGF-I). Breast Cancer Res Treat 1992;22:91-100.

(*105*) Kleinberg DL, Feldman M, Ruan W. IGF-I: an essential factor in terminal end bud formation and ductal morphogenesis. J Mammary Gland Biol Neoplasia 2000;5:7-17.

(*106*) Yu H, Rohan T. Role of the insulin-like growth factor family in cancer development and progression. J Natl Cancer Inst 2000;92:1472-89.

(*107*) Holmes MD, Pollak MN, Hankinson SE. Lifestyle correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev 2002;11:862-7.

(*108*) Biglia N, Ambroggio S, Ponzone R, Sgro L, Ujcic E, Dato FA, et al. Modification of serum IGF-I, IGFBPs and SHBG levels by different HRT regimens. Maturitas 2003;45:283-91.

(*109*) Campagnoli C, Biglia N, Altare F, Lanza MG, Lesca L, Cantamessa C, et al. Differential effects of oral conjugated estrogens and transdermal estradiol on insulin-like growth factor 1, growth hormone and sex hormone binding globulin serum levels. Gynecol Endocrinol 1993;7:251-8.

(*110*) Weissberger AJ, Ho KK, Lazarus L. Contrasting effects of oral and transdermal routes of estrogen replacement therapy on 24-hour growth hormone (GH) secretion, insulin-like growth factor I, and GH-binding protein in postmenopausal women. J Clin Endocrinol Metab 1991;72:374-81.

(*111*) Kelly JJ, Rajkovic IA, O'Sullivan AJ, Sernia C, Ho KK. Effects of different oral oestrogen formulations on insulin-like growth factor-I, growth hormone and growth hormone binding protein in post-menopausal women. Clin Endocrinol (Oxf) 1993;39:561-7.

(*112*) Cardim HJ, Lopes CM, Giannella-Neto D, da Fonseca AM, Pinotti JA. The insulin-like growth factor-I system and hormone replacement therapy. Fertil Steril 2001;75:282-7.

(*113*) Cano A, Castelo-Branco C, Tarin JJ. Effect of menopause and different combined estradiol-progestin regimens on basal and growth hormone-releasing hormone-

stimulated serum growth hormone, insulin-like growth factor-1, insulin-like growth factor binding protein (IGFBP)-1, and IGFBP-3 levels. Fertil Steril 1999;71:261-7.

(*114*) Heald A, Selby PL, White A, Gibson JM. Progestins abrogate estrogen-induced changes in the insulin-like growth factor axis. Am J Obstet Gynecol 2000;183:593-600.

(*115*) Jernstrom H, Deal C, Wilkin F, Chu W, Tao Y, Majeed N, et al. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev 2001;10:377-84.

(*116*) Balogh A, Kauf E, Vollanth R, Graser G, Klinger G, Oettel M. Effects of two oral contraceptives on plasma levels of insulin-like growth factor I (IGF-I) and growth hormone (hGH). Contraception 2000;62:259-69.

(*117*) Jernstrom H, Olsson H. Suppression of plasma insulin-like growth factor-1 levels in healthy, nulliparous, young women using low dose oral contraceptives. Gynecol Obstet Invest 1994;38:261-5.

(*118*) Morimoto LM, Newcomb PA, White E, Bigler J, Potter JD. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: personal and lifestyle factors (United States). Cancer Causes Control 2005;16:917-27.

(*119*) Thissen JP, Ketelslegers JM, Underwood LE. Nutritional regulation of the insulin-like growth factors. Endocr Rev 1994;15:80-101.

(*120*) Voskuil DW, Vrieling A, van't Veer LJ, Kampman E, Rookus MA. The insulin-like growth factor system in cancer prevention: potential of dietary intervention strategies. Cancer Epidemiol Biomarkers Prev 2005;14:195-203.

(*121*) Norat T, Dossus L, Rinaldi S, Overvad K, Gronbaek H, Tjonneland A, et al. Diet, serum insulin-like growth factor-I and IGF-binding protein-3 in European women. Eur J Clin Nutr 2007;61:91-8.

(*122*) Holmes MD, Pollak MN, Willett WC, Hankinson SE. Dietary correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev 2002;11:852-61.

(*123*) Baibas N, Bamia C, Vassilopoulou E, Sdrolias J, Trichopoulou A, Trichopoulos D. Dietary and lifestyle factors in relation to plasma insulin-like growth factor I in a general population sample. Eur J Cancer Prev 2003;12:229-34.

(*124*) Keinan-Boker L, Bueno De Mesquita HB, Kaaks R, Van Gils CH, Van Noord PA, Rinaldi S, et al. Circulating levels of insulin-like growth factor I, its binding proteins -1,-2, -3, C-peptide and risk of postmenopausal breast cancer. Int J Cancer 2003;106:90-5.

(*125*) Cruickshank JK, Heald AH, Anderson S, Cade JE, Sampayo J, Riste LK, et al. Epidemiology of the insulin-like growth factor system in three ethnic groups. Am J Epidemiol 2001;154:504-13.

(*126*) Janssen JA, Burger H, Stolk RP, Grobbee DE, de Jong FH, Lamberts SW, et al. Gender-specific relationship between serum free and total IGF-I and bone mineral density in elderly men and women. Eur J Endocrinol 1998;138:627-32.

(*127*) Goodman-Gruen D, Barrett-Connor E. Epidemiology of insulin-like growth factor-I in elderly men and women. The Rancho Bernardo Study. Am J Epidemiol 1997;145:970-6.

(*128*) Landin-Wilhelmsen K, Wilhelmsen L, Lappas G, Rosen T, Lindstedt G, Lundberg PA, et al. Serum insulin-like growth factor I in a random population sample of men and women: relation to age, sex, smoking habits, coffee consumption and physical activity, blood pressure and concentrations of plasma lipids, fibrinogen, parathyroid hormone and osteocalcin. Clin Endocrinol (Oxf) 1994;41:351-7.

(*129*) Laughlin GA, Barrett-Connor E, Criqui MH, Kritz-Silverstein D. The prospective association of serum insulin-like growth factor I (IGF-I) and IGF-binding protein-1 levels with all cause and cardiovascular disease mortality in older adults: the Rancho Bernardo Study. J Clin Endocrinol Metab 2004;89:114-20.

(*130*) Gram IT, Norat T, Rinaldi S, Dossus L, Lukanova A, Tehard B, et al. Body mass index, waist circumference and waist-hip ratio and serum levels of IGF-I and IGFBP-3 in European women. Int J Obes (Lond) 2006;30:1623-31.

(*131*) Lukanova A, Lundin E, Zeleniuch-Jacquotte A, Muti P, Mure A, Rinaldi S, 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;150:161-71.

(*132*) Sandhu MS, Gibson JM, Heald AH, Dunger DB, Wareham NJ. Association between insulin-like growth factor-I: insulin-like growth factor-binding protein-1 ratio and metabolic and anthropometric factors in men and women. Cancer Epidemiol Biomarkers Prev 2004;13:166-70.

(*133*) Allen NE, Appleby PN, Kaaks R, Rinaldi S, Davey GK, Key TJ. Lifestyle determinants of serum insulin-like growth-factor-I (IGF-I), C-peptide and hormone binding protein levels in British women. Cancer Causes Control 2003;14:65-74.

(*134*) Lukanova A, Soderberg S, Stattin P, Palmqvist R, Lundin E, Biessy C, et al. Nonlinear relationship of insulin-like growth factor (IGF)-I and IGF-I/IGF-binding protein-3 ratio with indices of adiposity and plasma insulin concentrations (Sweden). Cancer Causes Control 2002;13:509-16.

(*135*) Yamamoto H, Kato Y. Relationship between plasma insulin-like growth factor I (IGF-I) levels and body mass index (BMI) in adults. Endocr J 1993;40:41-5.

(*136*) Kunitomi M, Wada J, Takahashi K, Tsuchiyama Y, Mimura Y, Hida K, et al. Relationship between reduced serum IGF-I levels and accumulation of visceral fat in Japanese men. Int J Obes Relat Metab Disord 2002;26:361-9.

(*137*) De Pergola G, Zamboni M, Pannacciulli N, Turcato E, Giorgino F, Armellini F, et al. Divergent effects of short-term, very-low-calorie diet on insulin-like growth factor-I and insulin-like growth factor binding protein-3 serum concentrations in premenopausal women with obesity. Obes Res 1998;6:408-15.

(*138*) Rasmussen MH, Frystyk J, Andersen T, Breum L, Christiansen JS, Hilsted J. The impact of obesity, fat distribution, and energy restriction on insulin-like growth factor-1 (IGF-1), IGF-binding protein-3, insulin, and growth hormone. Metabolism 1994;43:315-9.

(*139*) Marin P, Kvist H, Lindstedt G, Sjostrom L, Bjorntorp P. Low concentrations of insulin-like growth factor-I in abdominal obesity. Int J Obes Relat Metab Disord 1993;17:83-9.

(*140*) Soderberg S, Ahren B, Eliasson M, Dinesen B, Brismar K, Olsson T. Circulating IGF binding protein-1 is inversely associated with leptin in non-obese men and obese postmenopausal women. Eur J Endocrinol 2001;144:283-90.

(*141*) Wolk K, Larsson SC, Vessby B, Wolk A, Brismar K. Metabolic, anthropometric, and nutritional factors as predictors of circulating insulin-like growth factor binding protein-1 levels in middle-aged and elderly men. J Clin Endocrinol Metab 2004;89:1879-84.

(*142*) Heald AH, Cruickshank JK, Riste LK, Cade JE, Anderson S, Greenhalgh A, et al. Close relation of fasting insulin-like growth factor binding protein-1 (IGFBP-1) with glucose tolerance and cardiovascular risk in two populations. Diabetologia 2001;44:333-9.

(*143*) Lukanova A, Toniolo P, Akhmedkhanov A, Hunt K, Rinaldi S, Zeleniuch-Jacquotte A, et al. A cross-sectional study of IGF-I determinants in women. Eur J Cancer Prev 2001;10:443-52.

(*144*) Gapstur SM, Kopp P, Chiu BC-H, Gann PH, Colangelo LA, Liu K. Longitudinal associations of age, anthropometric and lifestyle factors with serum total insulin-like growth factor-I and IGF binding protein-3 levels in Black and White men: the CARDIA Male Hormone Study. Cancer Epidemiol Biomarkers Prev 2004;13:2208-16.

(*145*) Chang S, Wu X, Yu H, Spitz MR. Plasma concentrations of insulin-like growth factors among healthy adult men and postmenopausal women: associations with body composition, lifestyle, and reproductive factors. Cancer Epidemiol Biomarkers Prev 2002;11:758-66.

(*146*) Helle SI, Ekse D, Holly JM, Lonning PE. The IGF-system in healthy pre- and postmenopausal women: relations to demographic variables and sex-steroids. J Steroid Biochem Mol Biol 2002;81:95-102.

(*147*) Bray I, Gunnell D, Holly JM, Middleton N, Davey Smith G, Martin RM. Associations of childhood and adulthood height and the components of height with insulin-like growth factor levels in adulthood: a 65-year follow-up of the Boyd Orr cohort. J Clin Endocrinol Metab 2006;91:1382-9.

(*148*) Rogers I, Metcalfe C, Gunnell D, Emmett P, Dunger D, Holly J. Insulin-like growth factor-I and growth in height, leg length, and trunk length between ages 5 and 10 years. J Clin Endocrinol Metab 2006;91:2514-9.

(*149*) Juul A, Bang P, Hertel NT, Main K, Dalgaard P, Jorgensen K, et al. Serum insulin-like growth factor-I in 1030 healthy children, adolescents, and adults: relation to age, sex, stage of puberty, testicular size, and body mass index. J Clin Endocrinol Metab 1994;78:744-52.

(*150*) Kaaks R, Lundin E, Rinaldi S, Manjer J, Biessy C, Soderberg S, et al. Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden. Cancer Causes Control 2002;13:307-16.

(*151*) Schernhammer ES, Holly JM, Pollak MN, Hankinson SE. Circulating levels of insulin-like growth factors, their binding proteins, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2005;14:699-704.

(*152*) Allen NE, Roddam AW, Allen DS, Fentiman IS, Dos Santos Silva I, Peto J, et al. A prospective study of serum insulin-like growth factor-I (IGF-I), IGF-II, IGF-binding protein-3 and breast cancer risk. Br J Cancer 2005;92:1283-7.

(*153*) Gronbaek H, Flyvbjerg A, Mellemkjaer L, Tjonneland A, Christensen J, Sorensen HT, et al. Serum insulin-like growth factors, insulin-like growth factor binding proteins, and breast cancer risk in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2004;13:1759-64.

(*154*) Baglietto L, English DR, Hopper JL, Morris HA, Tilley WD, Giles GG. Circulating insulin-like growth factor-I and binding protein-3 and the risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2007;16:763-8.

(*155*) Rinaldi S, Peeters PH, Berrino F, Dossus L, Biessy C, Olsen A, et al. IGF-I, IGFBP-3 and breast cancer risk in women: The European Prospective Investigation into Cancer and Nutrition (EPIC). Endocr Relat Cancer 2006;13:593-605.

(*156*) Schernhammer ES, Holly JM, Hunter DJ, Pollak MN, Hankinson SE. Insulin-like growth factor-I, its binding proteins (IGFBP-1 and IGFBP-3), and growth hormone and breast cancer risk in The Nurses Health Study II. Endocr Relat Cancer 2006;13:583-92.

(*157*) Toniolo P, Bruning PF, Akhmedkhanov A, Bonfrer JM, Koenig KL, Lukanova A, et al. Serum insulin-like growth factor-I and breast cancer. Int J Cancer 2000;88:828-32.

(*158*) Rinaldi S, Kaaks R, Zeleniuch-Jacquotte A, Arslan AA, Shore RE, Koenig KL, et al. Insulin-like growth factor-I, IGF binding protein-3, and breast cancer in young women: a comparison of risk estimates using different peptide assays. Cancer Epidemiol Biomarkers Prev 2005;14:48-52.

(*159*) Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet 1998;351:1393-6.

(*160*) Muti P, Quattrin T, Grant BJ, Krogh V, Micheli A, Schunemann HJ, et al. Fasting glucose is a risk factor for breast cancer: a prospective study. Cancer Epidemiol Biomarkers Prev 2002;11:1361-8.

(*161*) Krajcik RA, Borofsky ND, Massardo S, Orentreich N. Insulin-like growth factor I (IGF-I), IGF-binding proteins, and breast cancer. Cancer Epidemiol Biomarkers Prev 2002;11:1566-73.

(*162*) dos Santos Silva I, Johnson N, De Stavola B, Torres-Mejia G, Fletcher O, Allen DS, et al. The insulin-like growth factor system and mammographic features in premenopausal and postmenopausal women. Cancer Epidemiol Biomarkers Prev 2006;15:449-55.

(*163*) Bremnes Y, Ursin G, Bjurstam N, Rinaldi S, Kaaks R, Gram IT. Insulin-like growth factor and mammographic density in postmenopausal Norwegian women. Cancer Epidemiol Biomarkers Prev 2007;16:57-62.

(*164*) Maskarinec G, Takata Y, Chen Z, Gram IT, Nagata C, Pagano I, et al. IGF-I and mammographic density in four geographic locations: A pooled analysis. Int J Cancer 2007;121:1786-92.

(*165*) Aiello EJ, Tworoger SS, Yasui Y, Stanczyk FZ, Potter J, Ulrich CM, et al. Associations among circulating sex hormones, insulin-like growth factor, lipids, and mammographic density in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2005;14:1411-7.

(*166*) Hoppener JW, de Pagter-Holthuizen P, Geurts van Kessel AH, Jansen M, Kittur SD, Antonarakis SE, et al. The human gene encoding insulin-like growth factor I is located on chromosome 12. Hum Genet 1985;69:157-60.

(*167*) NCBI Annotation Project. Direct Submission Submitted (05-DEC-2001). National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA 2001.

(*168*) Delany AM, Pash JM, Canalis E. Cellular and clinical perspectives on skeletal insulin-like growth factor I. J Cell Biochem 1994;55:328-33.

(*169*) The International HapMap Consortium. The International HapMap Project. Nature 2003;426:789-96.

(*170*) Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-61.

(*171*) Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

(*172*) DeLellis K, Ingles S, Kolonel L, McKean-Cowdin R, Henderson B, Stanczyk F, et al. IGF1 genotype, mean plasma level and breast cancer risk in the Hawaii/Los Angeles multiethnic cohort. Br J Cancer 2003;88:277-82.

(*173*) Allen NE, Davey GK, Key TJ, Zhang S, Narod SA. Serum insulin-like growth factor I (IGF-I) concentration in men is not associated with the cytosine-adenosine repeat polymorphism of the IGF-I gene. Cancer Epidemiol Biomarkers Prev 2002;11:319-20.

(*174*) Jernstrom H, Chu W, Vesprini D, Tao Y, Majeed N, Deal C, et al. Genetic factors related to racial variation in plasma levels of insulin-like growth factor-1: implications for premenopausal breast cancer risk. Mol Genet Metab 2001;72:144-54.

(*175*) Rosen CJ, Kurland ES, Vereault D, Adler RA, Rackoff PJ, Craig WY, et al. Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density. J Clin Endocrinol Metab 1998;83:2286-90.

(*176*) Schut AF, Janssen JA, Deinum J, Vergeer JM, Hofman A, Lamberts SW, et al. Polymorphism in the promoter region of the insulin-like growth factor I gene is related to carotid intima-media thickness and aortic pulse wave velocity in subjects with hypertension. Stroke 2003;34:1623-7.

(*177*) Vaessen N, Heutink P, Janssen JA, Witteman JC, Testers L, Hofman A, et al. A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. Diabetes 2001;50:637-42.

(*178*) Rietveld I, Janssen JA, Hofman A, Pols HA, van Duijn CM, Lamberts SW. A polymorphism in the IGF-I gene influences the age-related decline in circulating total IGF-I levels. Eur J Endocrinol 2003;148:171-5.

(*179*) Missmer SA, Haiman CA, Hunter DJ, Willett WC, Colditz GA, Speizer FE, et al. A sequence repeat in the insulin-like growth factor-1 gene and risk of breast cancer. Int J Cancer 2002;100:332-6.

(*180*) Giovannucci E, Haiman CA, Platz EA, Hankinson SE, Pollak MN, Hunter DJ. Dinucleotide repeat in the insulin-like growth factor-I gene is not related to risk of colorectal adenoma. Cancer Epidemiol Biomarkers Prev 2002;11:1509-10.

(*181*) Frayling TM, Hattersley AT, McCarthy A, Holly J, Mitchell SM, Gloyn AL, et al. A putative functional polymorphism in the IGF-I gene: association studies with type 2 diabetes, adult height, glucose tolerance, and fetal growth in U.K. populations. Diabetes 2002;51:2313-6.

(*182*) Slattery ML, Baumgartner KB, Byers T, Guiliano A, Sweeney C, Herrick J, et al. Genetic, anthropometric, and lifestyle factors associated with IGF-1 and IGFBP-3 levels in Hispanic and non-Hispanic White women. Cancer Causes Control 2005;16:1147-57.

(*183*) Wong HL, Delellis K, Probst-Hensch N, Koh WP, Van Den Berg D, Lee HP, et al. A new single nucleotide polymorphism in the insulin-like growth factor I regulatory region associates with colorectal cancer risk in singapore chinese. Cancer Epidemiol Biomarkers Prev 2005;14:144-51.

(*184*) Kato I, Eastham J, Li B, Smith M, Yu H. Genotype-phenotype analysis for the polymorphic CA repeat in the insulin-like growth factor-I (IGF-I) gene. Eur J Epidemiol 2003;18:203-9.

(*185*) Kim JG, Roh KR, Lee JY. The relationship among serum insulin-like growth factor-I, insulin-like growth factor-I gene polymorphism, and bone mineral density in postmenopausal women in Korea. Am J Obstet Gynecol 2002;186:345-50.

(*186*) Rietveld I, Janssen JA, van Rossum EF, Houwing-Duistermaat JJ, Rivadeneira F, Hofman A, et al. A polymorphic CA repeat in the IGF-I gene is associated with gender-specific differences in body height, but has no effect on the secular trend in body height. Clin Endocrinol (Oxf) 2004;61:195-203.

(*187*) Al-Zahrani A, Sandhu MS, Luben RN, Thompson D, Baynes C, Pooley KA, et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet 2006;15:1-10.

(*188*) Morimoto LM, Newcomb PA, White E, Bigler J, Potter JD. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: genetic factors. Cancer Epidemiol Biomarkers Prev 2005;14:1394-401.

(*189*) Wen W, Gao YT, Shu XO, Yu H, Cai Q, Smith JR, et al. Insulin-like growth factor-I gene polymorphism and breast cancer risk in Chinese women. Int J Cancer 2005;113:307-11.

(*190*) Wagner K, Hemminki K, Israelsson E, Grzybowska E, Soderberg M, Pamula J, et al. Polymorphisms in the IGF-1 and IGFBP 3 promoter and the risk of breast cancer. Breast Cancer Res Treat 2005;92:133-40.

(*191*) Figer A, Karasik YP, Baruch RG, Chetrit A, Papa MZ, Sade RB, et al. Insulin-like growth factor I polymorphism and breast cancer risk in Jewish women. Isr Med Assoc J 2002;4:759-62.

(*192*) Yu H, Li BD, Smith M, Shi R, Berkel HJ, Kato I. Polymorphic CA repeats in the IGF-I gene and breast cancer. Breast Cancer Res Treat 2001;70:117-22.

(*193*) Cleveland RJ, Gammon MD, Edmiston SN, Teitelbaum SL, Britton JA, Terry MB, et al. IGF1 CA repeat polymorphisms, lifestyle factors and breast cancer risk in the Long Island Breast Cancer Study Project. Carcinogenesis 2006;27:758-65.

(*194*) Canzian F, McKay JD, Cleveland RJ, Dossus L, Biessy C, Rinaldi S, et al. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer 2006;94:299-307.

(*195*) Setiawan VW, Cheng I, Stram DO, Penney KL, Le Marchand L, Altshuler D, et al. Igf-I genetic variation and breast cancer: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 2006;15:172-4.

(*196*) The International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437:1299-320.

(*197*) Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865-9.

(*198*) Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, et al. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 2000;67:383-94.

(*199*) Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 2002;11:513-20.

(*200*) Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151-8.

(*201*) Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. Nat Genet 2005;37:868-72.

(*202*) Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004;36:388-93.

(*203*) Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003;73:1402-22.

(*204*) Thomas DC. Statistical Methods in Genetic Epidemiology. New York: Oxford University Press; 2004.

(*205*) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. J Natl Cancer Inst Monogr 1999:31-7.

(*206*) Lange C, Laird NM. Power calculations for a general class of family-based association tests: dichotomous traits. Am J Hum Genet 2002;71:575-84.

(*207*) Van Steen K, Lange C. PBAT: a comprehensive software package for genome-wide association analysis of complex family-based studies. Hum Genomics 2005;2:67-9.

(*208*) Hintze J. NCSS, PASS, and GESS. Kaysville, Utah: NCSS; 2006.

(*209*) Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol 1994;39:1629-38.

(*210*) Missmer SA, Spiegelman D, Bertone-Johnson ER, Barbieri RL, Pollak MN, Hankinson SE. Reproducibility of plasma steroid hormones, prolactin, and insulin-like growth factor levels among premenopausal women over a 2- to 3-year period. Cancer Epidemiol Biomarkers Prev 2006;15:972-8.

(*211*) Willett WC. Nutritional Epidemiology. New York: Oxford University Press; 1998.

(*212*) John U, Hanke M, Grothues J, Thyrian JR. Validity of overweight and obesity in a nation based on self-report versus measurement device data. Eur J Clin Nutr 2006;60:372-7.

(*213*) Gillum RF, Sempos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey. Nutr J 2005;4:27.

(*214*) Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. J Midwifery Womens Health 2003;48:338-45.

(*215*) Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. Public Health Nutr 2002;5:561-5.

(216) The Centre for Applied Genomics. 2003. Hospital for Sick Children. Toronto, Canada: http://www.tcag.ca/.

(*217*) Sobel E, Papp JC, Lange K. Detection and integration of genotyping errors in statistical genetics. Am J Hum Genet 2002;70:496-508.

(*218*) Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, et al. Identification and analysis of error types in high-throughput genotyping. Am J Hum Genet 2000;67:727-36.

(*219*) Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol 2000;19 Suppl 1:S36-42.

(*220*) Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506-16.

(*221*) Horvath S, Wei E, Xu X, Palmer LJ, Baur M. Family-based association test method: age of onset traits and covariates. Genet Epidemiol 2001;21 Suppl 1:S403-8.

(*222*) Lunetta KL, Faraone SV, Biederman J, Laird NM. Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. Am J Hum Genet 2000;66:605-14.

(*223*) Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 2001;9:301-6.

(*224*) Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society (Series B) 1995;57:289-300.

(*225*) Pounds S, Cheng C. Improving false discovery rate estimation. Bioinformatics 2004;20:1737-45.

(*226*) Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440-5.

(*227*) Wang L, Hirayasu K, Ishizawa M, Kobayashi Y. Purification of genomic DNA from human whole blood by isopropanol-fractionation with concentrated Nal and SDS. Nucleic Acids Res 1994;22:1774-5.

(*228*) Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989;44:388-96.

(*229*) Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 1999;27:573-80.

(*230*) Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361-72.

(*231*) Liu K, Muse S. PowerMarker: new genetic data analysis software. Version 3.0. Free program distributed by the author over the internet. http://www.powermarker.net

(*232*) SAS release 8.02 (1999-2001) by SAS Institute Inc. 1999-2001. SAS Institute Inc., Cary, NC, US.

(*233*) Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79-91.

(*234*) Takacs I, Koller DL, Peacock M, Christian JC, Hui SL, Conneally PM, et al. Sibling pair linkage and association studies between bone mineral density and the insulin-like growth factor I gene locus. J Clin Endocrinol Metab 1999;84:4467-71.

(*235*) Sachdev D, Yee D. The IGF system and breast cancer. Endocr Relat Cancer 2001;8:197-209.

(*236*) Sun G, Gagnon J, Chagnon YC, Perusse L, Despres JP, Leon AS, et al. Association and linkage between an insulin-like growth factor-1 gene polymorphism and fat free mass in the HERITAGE Family Study. Int J Obes Relat Metab Disord 1999;23:929-35.

(*237*) Hadsell DL. The insulin-like growth factor system in normal mammary gland function. Breast Dis 2003;17:3-14.

(*238*) Kleinberg DL. Role of IGF-I in normal mammary development. Breast Cancer Res Treat 1998;47:201-8.

(*239*) John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res 2004;6:R375-89.

(*240*) Sutherland HJ, Lacroix J, Knight J, Andrulis IL, Boyd NF. The Cooperative Familial Registry for Breast Cancer Studies: design and first year recruitment rates in Ontario. J Clin Epidemiol 2001;54:93-8.

(*241*) Couillault C. Pedagree: A program for detecting autosomal marker Mendelian incompatibilities in pedigree data. Version 1.00; 2002. http://pedagree.free.fr

(*242*) Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 2004;26:61-9.

(*243*) Mitchell AA, Cutler DJ, Chakravarti A. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. Am J Hum Genet 2003;72:598-610.

(*244*) van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol 2000;152:514-27.

(*245*) Silventoinen K, Krueger RF, Bouchard TJ, Jr., Kaprio J, McGue M. Heritability of body height and educational attainment in an international context: comparison of adult twins in Minnesota and Finland. Am J Hum Biol 2004;16:544-55.

(*246*) Silventoinen K. Determinants of variation in adult body height. J Biosoc Sci 2003;35:263-85.

(*247*) Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res 2003;6:399-408.

(*248*) Voorhoeve PG, van Rossum EF, Te Velde SJ, Koper JW, Kemper HC, Lamberts SW, et al. Association between an IGF-I gene polymorphism and body fatness: differences between generations. Eur J Endocrinol 2006;154:379-88.

(*249*) Sweeney C, Murtaugh MA, Baumgartner KB, Byers T, Giuliano AR, Herrick JS, et al. Insulin-like growth factor pathway polymorphisms associated with body size in Hispanic and non-Hispanic white women. Cancer Epidemiol Biomarkers Prev 2005;14:1802-9.

(*250*) Lange C, Silverman EK, Xu X, Weiss ST, Laird NM. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics 2003;4:195-206.

(*251*) Lopez-Alarcon M, Hunter GR, Gower BA, Fernandez JR. IGF-I polymorphism is associated with lean mass, exercise economy, and exercise performance among premenopausal women. Arch Med Res 2007;38:56-63.

(*252*) Boyd NF, Jensen HM, Cooke G, Han HL, Lockwood GA, Miller AB. Mammographic densities and the prevalence and incidence of histological types of benign breast disease. Reference Pathologists of the Canadian National Breast Screening Study. Eur J Cancer Prev 2000;9:15-24.

(*253*) Byrne C, Schairer C, Brinton LA, Wolfe J, Parekh N, Salane M, et al. Effects of mammographic density and benign breast disease on breast cancer risk (United States). Cancer Causes Control 2001;12:103-10.

(*254*) Rothman KJ, Greenland S. Modern Epidemiology. Philadelphia: Lippincott-Raven Publishers; 1998.

(*255*) Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95:448-57.

(*256*) International Agency for Research on Cancer. Cancer Incidence in Five Continents, Vol. VIII. IARC Scientific Publication No. 155. Available at: http://www-dep.iarc.fr/ [accessed 2007-07-24]

(*257*) Khosravi J, Diamandi A, Bodani U, Khaja N, Krishna RG. Pitfalls of immunoassay and sample for IGF-I: comparison of different assay methodologies using various fresh and stored serum samples. Clin Biochem 2005;38:659-66.

(*258*) Byrne C, Hankinson SE, Pollak M, Willett WC, Colditz GA, Speizer FE. Insulin-like growth factors and mammographic density. Growth Horm IGF Res 2000;10 Suppl A:S24-5.

Appendix 1. Contribution of the author to the design and implementation of the project

The author was not involved in the original collection of data that this thesis was based on. Recruitment into the cross-sectional study and the family based study were already complete for the time periods examined in this thesis when the author's study began. The author was key in the conceptualization and implementation of the studies presented here, and wrote the proposal to the Breast Cancer Registry Advisory Committee, which formed the basis for the studies presented in chapters 5 and 6 (no formal proposal was necessary for the cross-sectional study). The author used software and NCBI data to locate two of the three polymorphisms examined in this study (the third was obtained through the literature). In Dr. Ozcelik's laboratory at Mt. Sinai Hospital, the author designed and tested primers for DNA amplification, performed the majority of all PCR reactions and supervised the rest. The author also liased with The Centre for Applied Genomics (TCAG), inspected genotype assessments from the TCAG (with some assistance from an individual experienced in interpreting the output provided by the TCAG) and designed and implemented all genotyping quality control measures. The author evaluated the utility of various software tools required for statistical genetic analyses (e.g. FBAT 1.5, PBAT 2.5). All data analyses were performed by the author which were then presented with interpretations to the committee, prior to drafting the chapters of the thesis.