University of Calgary

Endometrial Cancer Risk Associated with a *CYP19* (Aromatase) Polymorphism and Anthropometric Measures

by

Amber Graystone

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

SEPTEMBER, 2006

University of Calgary

Faculty of Graduate Studies

The undersigned certify that they have read and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Endometrial Cancer Risk Associated with a CYP19 (Aromatase) Polymorphism and Anthropometric Measures" submitted by Amber Graystone in partial fulfillment of the requirements for the degree Master of Science

Supervisor, Dr. Linda Cook
Department of Community Health Sciences

Dr. Christine Friedenreich
Department of Community Health Sciences

Dr. Anthony Magliocco
Department of Pathology & Laboratory Medicine, Department of Oncology

Dr. Reg Sauve
Department of Community Health Sciences

Dr. Peter Bridge
Department of Medical Genetics

Sept. 19, 2006
Date

Abstract

A (TTTA)n repeat polymorphism in *CYP19* and endometrial cancer were investigated, as this gene encodes a rate limiting enzyme (aromatase) in estrogen biosynthesis from adipose tissue. Since estrogen production in adipose tissue places obese women at an increased risk of the disease compared to non-obese women, the independent and joint effects of *CYP19* and obesity were investigated. Study participants included postmenopausal endometrial cancer cases (n=127) and controls (n=271) who were not using hormones. Those with a current weight of ≥160 lbs had the largest risk of endometrial cancer of all anthropometric measures (OR=3.13, 95%CI: 1.89-5.17). This risk increased to 4.14 (95%CI: 1.94-8.85) for those with both alleles containing ≥10 (TTTA) repeats. Waist circumference was the next strongest predictor of risk, (OR=2.82, 95%CI: 1.73-4.60) indicating that upper body fat distribution may influence endometrial cancer risk. Since genotype does not influence risk in non-obese women, public health interventions for endometrial cancer should target obesity for disease prevention.

Acknowledgements

I am so grateful to have been given the opportunity to be supervised by Dr. Linda Cook. Being such a great teacher and mentor, she was fully dedicated to guiding me through all stages of my graduate program. I need to thank her for all her kindness and patience as I pushed deadlines to the very end. I have enjoyed every moment of my experience at the University of Calgary and will bring great memories back to Ontario.

I would also like to thank Dr. Tony Magliocco for offering his laboratory facilities and for providing invaluable guidance as I developed my protocol. I am also appreciative of the direction given by Dr. Christine Friedenreich, who made my project possible by providing resources from the parent Endometrial Cancer Study. Thanks, also, to Dr. Reg Sauve for taking the time to participate in committee meetings and guiding me through this project. I would also like to thank Dr. Peter Bridge for bringing his expertise to the defense. This project was made possible by funding provided by the Department of Obstetrics and Gynecology.

There are many people in the Division of Population Health and Information at the Tom Baker Cancer Center that I would also like to thank. I have enjoyed working with Rita Biel, Aleata Ryhorchuk, Jacquie Gregory, Heather Neilson and Hilton Chim, who were always willing to help. I also need to thank Sony Brar for helping me with statistics and Jennifer Yelland who has been supportive in so many ways. Ernest Amankwah took the time to guide me through my first couple days of lab work and continued to provide support throughout this project. I also would like to thank Shelly and Stephanie at the University Core DNA Services who went over and above providing AFLP data for this project.

Dedication

This thesis is dedicated to my parents; I attribute all my accomplishments to their continuous love and support.

"Teach me, and I will forget. Show me, and I will remember.
Involve me, and I will understand."
-Chinese Proverb

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Molecular epidemiology has recently become popular as the increasing availability of high throughput molecular techniques has provided more opportunity for research[1]. This emerging field differs from traditional epidemiology in that it involves the use of biomarkers, ranging from serum proteins to genetic sequences, to better understand disease etiology. If the biomarkers are chosen appropriately, molecular epidemiology studies have the potential to contribute to the understanding of the pathophysiology of disease, provide a screening target and even improve upon diagnosis of specific disease subtypes. Cancer, being a complex disease of multi-factorial causation, has been the target of many studies in molecular epidemiology. Biomarkers for cancer research have included markers of exposure, early biologic effect, altered molecular structure/function, susceptibility and prognosis[2]. With the opportunities that new technologies have created for utilizing these markers in cancer research, many exciting developments have been made. Advanced DNA technologies have improved upon the identification of the Human Papillomavirus (HPV), which has been determined as a necessary cause for cervical cancer[3]. Public health implications lie in the opportunity this discovery created for vaccine development to prevent HPV triggered cancer[4]. In fact, the vaccine has just recently been approved for use in the United States and Canada[5, 6].

To understand and identify different types of breast cancer tumors various biomarkers have been studied to including estrogen/progesterone receptors[7], human epidermal growth factor receptor (HER2) and genetic markers of susceptibility, such as BRCA1 and BRCA2. These research developments have led to customized treatment options, ideal for specific types of breast cancer[8-10]. Promising opportunities have been created in the field of molecular cancer epidemiology and for research aimed at effective prevention and treatment of this disease.

The use of genetic sequences as biomarkers in molecular epidemiology studies is possible due to the technological advancements in high-throughput sequencing which led to determination of the entire human genome in 2003[11]. Now, the task of understanding the significance of this code remains, and the implications on health and disease provides further challenges. Of interest in cancer related genetic research is the search for genetic variants that explain the between-person variation that influences cancer risk. Cancer is generally considered to develop from one or more genetic events that influence the regulation of cellular replication. Disruptions occurring in somatic cells can trigger cancer development, which may or may not involve an inherited risk derived from germ line cells[12]. Familial clustering of specific cancers has long been acknowledged to suggest inherited risk and now that the technology is available, molecular techniques can

identify the genes responsible for this phenomenon[2]. The limitation of the discovery of these high penetrance genes is that they are usually very rare and make small contributions to the population attributable risk (PAR)[13]. Rare mutations in the BRCA1 and BRCA2 genes carry a significant inherited risk, however only 5% of breast cancer overall is attributed to these high-penetrant mutations[14, 15]. Germline mutations in mismatch repair genes MSH2, MLH1 and MSH6 can be used to identify Hereditary Non-polyposis Colon Cancer (HNPCC), a syndrome that is associated with an increased risk of developing cancer of the colon and endometrium. Although the lifetime risk of obtaining endometrial cancer in those with HNPCC is 40-60%, only a small proportion of cases are attributed to this condition[16].

A new approach in genetic association studies of cancer involves identification of candidate genes of known metabolic pathways[17]. These genes are usually expressed in the target tissue and have a key function in the metabolism of the carcinogen[13]. They usually involve genes of low penetrance, which tend to be more common and are, therefore, more of a concern from a public health prespective[2]. The candidate genes chosen for genetic epidemiology studies are based on known molecular mechanisms of disease development, and therefore, limited to those with well established etiology. Less well known may be the functional significance of the sequences of interest. Many studies of genetic epidemiology have focused on variant sequences that are hypothesized

to influence gene splicing, regulation and post-translational modification, however, the exact biologic role of these regulatory elements may not be completely understood[17]. In spite of this limitation, the nucleotide code of these sequences can be obtained and used as reliable exposure markers.

The candidate gene chosen for this study encodes a key enzyme in biosynthesis of estrogen, a hormone that is a key risk factor for endometrial cancer. Although a low penetrant gene, the high risk alleles are fairly common and, therefore, have the potential to make a significant contribution to population attributable risk. The purpose of this study is to determine the risk of endometrial cancer associated with this particular genetic polymorphism in an Albertan population and to investigate how this risk may be modified with measures of obesity.

Chapter 2

Literature Review

2.1. Evolution of Gene-Environment Studies

Recent advancement of high-throughput technologies used to detect various disease-associated biomarkers has provided more opportunity for studies in molecular epidemiology. Now that the genome has been sequenced[11], genetic elements have been used as biomarkers of inherited susceptibility and risk factor exposure. Studies of genetic epidemiology have led to an interesting approach that combines traditional and genetic epidemiology to study gene-environment interactions[18]. These studies regularly utilize a low-penetrant polymorphism and a common environmental or lifestyle factor to determine how these factors co-participate through a shared mechanism to influence disease development. Since factors studied are typically common in the population, results of these studies tend to have an important impact from a public health perspective[2]. This approach has the potential to enhance the understanding of how common exogenous carcinogens (i.e., smoking and hormone use) or lifestyle factors, such as obesity or diet, and certain genotypes interact to influence cancer development[18]. This approach can improve understanding of various preventable and non-preventable factors found frequently in a given population.

2.2. Interaction of Genetic and Environmental Factors

Interaction, from a biological perspective, can be described as an interdependent operation between two or more causes to produce disease[19]. As more risk factors for diseases become known and we further discover that some operate along the same pathway, studies of multiple risk factors with joint effects provides a challenge to epidemiologists. Studies of gene-environment interactions encounter issues of organization and presentation of data that assess individual and joint effects, along with the synergistic or antagonistic relationship between the two in disease development. These studies differ from traditional epidemiology of case-control studies, which utilizes a two dimensional contingency table (i.e., a 2x2 table) to organize disease and exposure groups. Interaction studies expand to include risk estimates for a disease involving multiple exposure variables, where an effective display of these results requires a reformatted version of the traditional 2x2 table. A stratified 2x4 table allows each of the risk factors to be assessed separately among cases and controls[20] (see Table 1.). However, a more comprehensive table that further clarifies the role of each factor provides an effective summary of how two factors interact to effect disease risk (see Table 2.). This table presents a condensed version of multiple 2x2 tables, where the size of each exposure group is clearly outlined. These tables clarify the common limitations of studies of gene-environment interactions, mainly inadequacies of sample size in each of

the strata to detect statistically significant interactions[18, 20, 21]. Those well-planned studies that carefully consider the frequency of each risk factor in the target population can deliver relevant information to explain disease pathogenesis. It is well known that cancer can arise from both genetic and environmental factors, and now these investigations can help clarify the complex nature that cancer manifests within the general population.

Table 1. Traditional Two-Dimensional Contingency Table. The 2x2 contingency table used in most epidemiological studies is further stratified by the environmental factor.

	Environmental Factor present		Environmental Factor absent	
	Case	Control	Case	Control
Genetic Factor Present	a	b	e	f
Genetic Factor absent	c	d	g	h
Odds Ratio	OR=ad/cb		OR=eh/gf	
Crude Odds Ratio	OR=a+e*d+h/c+e*g+f			

Table 2. Gene-Environment Interaction Table. Comprehensive table highlighting interaction between genetic and environmental factors to influence disease risk.

G	E	Case	Control	Odds Ratio	Main Information
+	+	a	b	ah/bg	Joint genotype & environmental factor vs. none
+	-	c	d	ch/dg	Genotype alone vs. none
-	+	e	f	eh/fg	Environmental factor alone vs. none
-	-	g	h	1	Common referent

G: genetic factor, E: environmental factor

2.3. Candidate Genes

Studies of gene-environment interactions require identification of a candidate gene that is suspected to have a role in disease development. This approach allows for isolation of a biomarker of risk that can be further analyzed with an environmental factor to describe the contributions genetic predisposition and preventable risk factors make in

disease development. The benefit of this strategy lies in the opportunity to determine how common factors manifest in a given population and co-participate to influence disease risk.

As techniques used for genetic analysis continue to make gene identification cheaper, easier and faster, studies of genetic epidemiology have expanded to include haplotype analysis using a cluster of polymorphisms. Limitations to this approach exist when the functional significance of each polymorphism is undetermined, making interpretation of the results a complicated task. It is also difficult to determine whether the polymorphisms operate independently of each other or synergize to create a high risk phenotype. Utilizing numerous genetic markers can also make analysis of interaction of an environmental factor challenging. In contrast, when one locus is stratified by a relevant environmental factor, results not only increase understanding of disease pathology, but also allow for inferences that are relevant from a public health perspective.

Many genetic epidemiological studies have focused on genes from a family of enzymes, cytochrome-dependent monooxygenases (CYPs), as biomarkers for disease susceptibility. The genes encoding CYPs have been chosen as candidate genes since these enzymes are responsible for metabolizing a variety of carcinogenic compounds[22]. Of interest in this study is a CYP responsible for estrogen production in adipose tissue. Estrogen is considered carcinogenic in estrogen-responsive tissues such

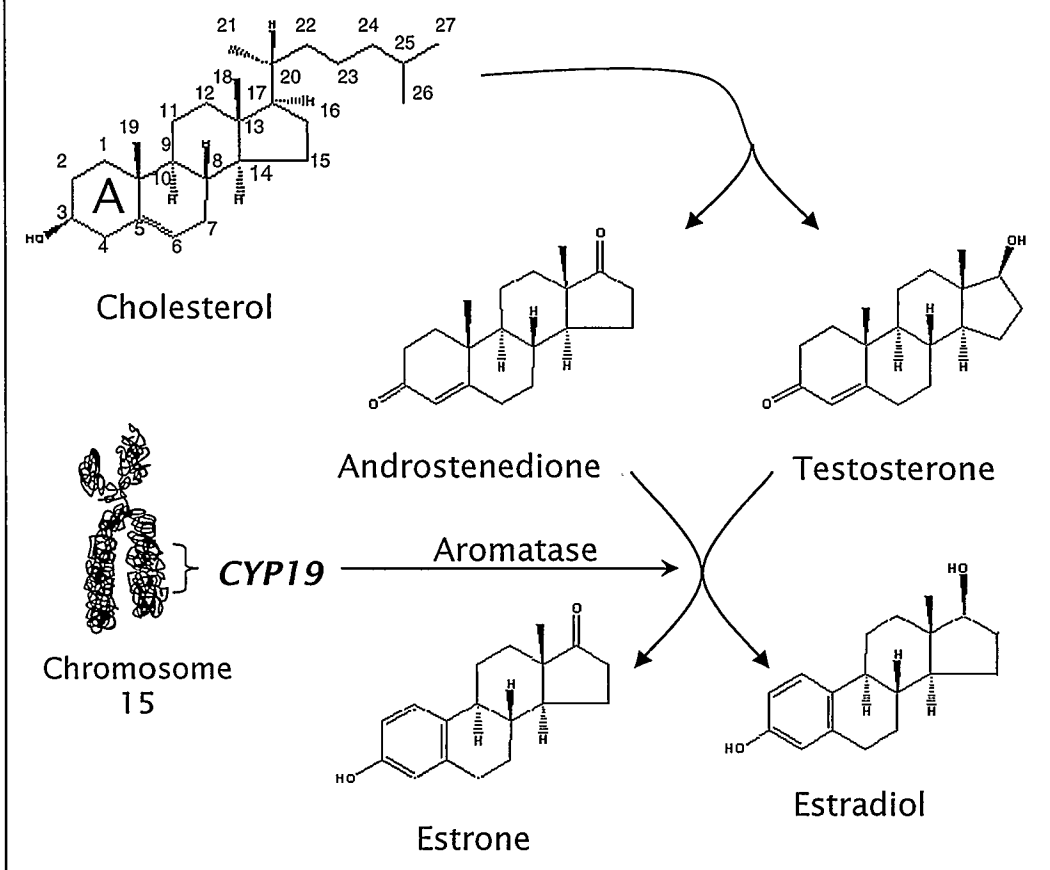
as the endometrium, and when unopposed by progesterone, has the ability to promote cellular proliferation. When bound to its receptor, the estrogen-estrogen receptor complex binds to specific sequences of DNA called estrogen-response elements to direct transcription of estrogen responsive genes[23]. Continual exposure to unopposed estrogen increases the risk of random genetic events that may result in loss of replicative control and tumor development[24].

2.4. Aromatase Cytochrome P450

In post-menopausal women, the primary source of estrogen is derived from cholesterol in adipose tissue[24]; a biosynthetic pathway that involves multiple enzymatic reactions catalyzed by a series of CYPs. Of most interest is the CYP involved in the rate-limiting step of estrogen production, where the androgens (testosterone and androstenedione) are converted to estrogens (estrone and estradiol). This last step is catalyzed by the enzyme aromatase, which is encoded by the *CYP19* gene[24, 25]. Aromatase is responsible for binding the C19 androgen substrates to produce C18 estrogens via a three-step process of aromatization. For every mole of C19 steroid metabolized, aromatase utilizes 3 mol of oxygen and 3 mol of NADPH[26]. It is believed that all three oxygen molecules are used in oxidizing the C19 angular methyl group, giving the gene encoding aromatase, *CYP19*, its designation. Subsequent aromatization of the A ring creates estrogen's phenolic moiety (see Figure 1.). The clinical and biological significance of aromatase is

supported by the occurrence of complications resulting from aromatase deficiency[27-30]. As genetic variations of the *CYP19* gene continue to demonstrate an influence on aromatase activity, further research is needed to determine how mutations ultimately affect hormone levels and cancer risk.

Figure 1. Rate Limiting Step in Estrogen Biosynthesis. Androgens (androstenedione and testosterone) are created from cholesterol in adipose tissue through a series of enzymatic reactions. Estrogen, in the form of estrone and estradiol is created through aromatization of androgens via aromatase, a protein encoded by the *CYP19* gene.



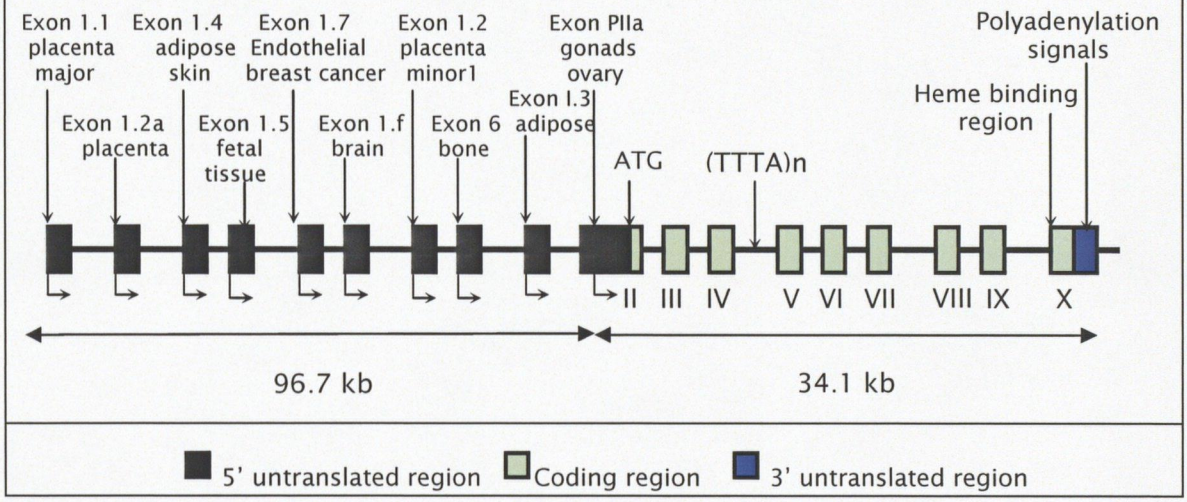
2.5. *CYP19* (TTTA)n Repeat Polymorphism

2.5.1. Description of *CYP19* Gene

The *CYP19* gene is found on chromosome 15 and spans 123 kilobases (kb) with a coding region of approximately 30 kb[31]. An alternatively spliced 5' untranslated region creates tissue specific promoters that mediate expression in bone, breast, ovary, brain, skin, placenta and adipose tissue. However, splicing at exon II, just upstream of these promoter regions, results in a coding region and protein product that is identical in all tissues[25] (see Figure 2.). Specific polymorphisms can affect enzymatic activity of aromatase, hormone levels and consequently lead to disease[32-35]. Many of these include single nucleotide polymorphisms, however, the most widely investigated polymorphism within *CYP19* is a variable tetranucleotide (TTTA)n repeat[36]. This repeat lies within intron 4 and has been found to naturally vary from 7-13 copies, where longer alleles have been shown to be associated with hormone related cancers[37-47]. This marker has been studied in relation to risk of breast and prostate cancer, however, the association between this polymorphism and endometrial cancer risk is of particular interest in this study. Aromatase has a key role in the pathway that provides the main source of estrogen in postmenopausal women, therefore, it would be appropriate to study this gene in relation to endometrial cancer. Furthermore, the (TTTA)n repeat polymorphism has been shown to be associated with increased estrogen levels, providing rationale for

investigation of this particular polymorphism and endometrial cancer[27, 43]. Previous research has found a significant association between an increased risk for endometrial cancer and a higher number of (TTTA)n repeats[48, 49]. These findings are summarized in Table 3.

Figure 2. Structural Organization of the *CYP19* Gene. Various promoters in the 5' untranslated region allow for tissue specific control, however, due to positioning of the ATG translational start codon, all protein products are identical. The tetranucleotide repeat polymporphism falls within the noncoding region of intron four, as indicated.



2.5.2. Possible Role of Polymorphism

Although the functional role of this particular *CYP19* tetranucleotide repeat has yet to be determined, research involving repeat polymorphisms provides some possible mechanisms to explain the relationship between this repeat sequences and endometrial cancer. The *CYP19* tetranucleotide repeat is found in an intronic region, one that is not within the protein coding information of aromatase. Therefore, it is likely that the repeat sequence may be responsible for a regulatory mechanism that may lead to disease development.

Fragile X syndrome and myotonic dystrophy are examples of diseases caused by complimentary triplet repeat polymorphisms. These repeat regions have the ability to form a folded hairpin structure when single stranded due to the complementary nucleotides in the repeating unit. This secondary structure may affect lagging strand DNA replication, methylation patterns and subsequent gene expression levels [50]. These repeating units have also been proposed to affect the affinity of RNA binding proteins, post-transcriptional splicing, RNA secondary structure stability and even regulation of splicing another gene product[51]. Ultimately, variation of *CYP19* RNA stability can lead to absolute gain/loss of aromatase function or altered protein conformations that may influence the reactivity of this enzyme[52, 53].

Most comparable to the *CYP19* (TTTA)n repeat is the (TCAT) intronic polymorphism found in the tyrosine hydroxlyase gene, a candidate gene for neuropsychiatric diseases. Alleles of this gene contain 5-10 (TCAT) repeats, where the 10 (TCAT) repeat allele was found to be associated with significantly higher levels of transcriptional activity. It was concluded that this polymorphism is a transcriptional regulatory element and acts as a binding site for nuclear proteins. This evidence suggests that perhaps a higher number of repeats efficiently stabilize proteins involved in DNA replication and transcription, improving the efficiency of protein expression.

Linkage disequilibrium may also explain the association between the intronic repeats and endometrial cancer. This effect exists when the biomarker of interest is associated with a nearby genetic sequence that may have a functional role in disease development. When these two elements are in linkage disequilibrium, they are commonly inherited together, which may explain a consistent association between the biomarker and disease[54]. This possibility is particularly important to consider since the functional role of the tetranucleotide repeat marker is unclear.

Table 3. Previous CYP19 Research. Summary of studies investigating *CYP19* (TTTA)n repeat according to hormone levels and endometrial cancer risk.

	Author	Year	Study Design /Population	Exposure	Findings
Hormone concentrations	Tworoger, SS[27]	2004	Case-control nested in an RCT / 173 postmenopausal women	Associations in estrogen levels	• Shortest repeats associated with ↓estrogen levels
	Haiman, CA[43]	2000	Case-control nested within NHS cohort / 190 postmenopausal women with no history of hormone therapy	Estrogen levels associated with 7 and 8 repeat alleles	• 7 repeats associated with ↓estrogen levels, 8 with ↑estrogen levels
Endometrial cancer Risk	Paynter, RA[49]	2005	Case-control nested within NHS cohort / 161 cases and 398 controls, all postmenopausal	Risk of having at least one >7 repeat allele	• OR=2.0 (1.3-3.3)
	Berstein, LM[48]	2001	Case-control / 85 cases (23 premenopausal, 62 postmenopausal) 110 controls	Risk of having at least one 11 repeat allele	• OR=1.9 (1.0-3.6)

NHS: Nurses' Health Study

2.5.3. Implications of Previous *CYP19* Research in Present Study

Previous *CYP19* studies have based risk estimates on the three most common alleles, containing 7, 8 and 11 (TTTA) repeats. Lacking, however, is the significance of these particular repeats at the molecular level. Results summarized in Table 3. do suggest that longer repeats are associated with higher estrogen levels and an increased risk of endometrial cancer. Investigations of breast cancer and the *CYP19* (TTTA)n polymorphism further support the hypothesis that longer repeats are associated with a higher cancer risk. These studies are addressed in section 5.2. of the discussion. In addition to these cancer studies, two additional studies summarized in Table 3. also suggest that longer repeats are associated with elevated estrogen levels, compared to those with shorter repeats.

Studies of endometrial and breast cancer and hormone levels with the *CYP19* (TTTA)n polymorphism suggests that the longer repeat alleles may affect aromatase activity, stability or rates of expression to consequently alter hormonal profile. Indeed, there is mounting evidence for the association between longer repeating genetic regions and disease[55-58]. Given the evidence that longer repeating genetic regions affect hormone profile levels and disease risk, we choose the midpoint of the allele distribution to categorize our alleles as low risk (shorter than the midpoint) and high risk (longer than the midpoint). Therefore, based on the evidence of the *CYP19* (TTTA)n repeat, our *a priori* hypothesis

states that an increase in risk would be present in those with alleles containing ≥10 (TTTA) repeats. We suggest that the length of the allele (i.e., the number of repeats) is predictive of risk, as opposed to one genotype in particular. With this consideration, and given the evidence in Table 3., the middle most allele (10 being the center from the natural variation of 7-13) seems to be an appropriate measure of risk.

2.6. Endometrial Cancer

2.6.1. Epidemiology of the Disease

The median age for endometrial cancer is 61 years, where the largest percentage of patients fall within the 50-59 age category[59]. In Canada, an estimated 3,500 new cases were diagnosed in 2001[60]. Previously, the incidence of endometrial cancer in Canada (and Alberta) increased dramatically from 1963-1978[61], a trend partially attributed to unopposed estrogen use[62]. This steady increase through the mid-1970s was also observed throughout North America, where the effect subsided since the late 1980s. In general, these variations in incidence were not accompanied by an increase in the mortality rate for endometrial cancer[60, 61]. Within Canada, incidence rates (based on data from 1994-1998) vary slightly across provinces, from 15.1/100,000 in New Brunswick to 21.5/100,000 in Manitoba. Alberta's incidence rate was second highest in Canada at ~20.5/100,000[60]. In Alberta's most recent Cancer Report, the age standardized incidence of endometrial cancer was 21.8/100,000[63].

The most common type of endometrial cancer by pathologic classification, Type I, accounts for ~80% of cases and is considered to be 'estrogen-dependent'[64]. It is believed to develop from endometrial hyperplasia (tissue growth), express estrogen and progesterone receptors[65] and be associated with elevated levels of serum estradiol[66]. This type of endometrial cancer is typically of endometrioid morphology and is associated with genetic level changes during tumor progression. Hypermethylation in stable microsatellite regions is associated with atypical endometrial hyperplasia, suggesting the role of microsatellite instability in this phenotype. Further genetic mutations in numerous susceptibility genes, K-ras, PTEN and p53, occur as the tumor progresses[67].

Type II accounts for approximately 20% of all endometrial cancers and does not develop from unopposed estrogen exposure. Type II is associated with endometrial atrophy (tissue deterioration), is generally serous or clear-cell carcinomas and has a higher grade and a poorer prognosis[64, 68]. Microsatellite instability and hypermethyation are not associated with this phenotype, however, mutations in p53 are frequent in Type II serous tumors. HER2/neu mutations are typically found in this type of endometrial cancer, however, the time of onset of these mutations are unknown[67].

2.6.2. Unopposed Estrogen Hypothesis

The influence of hormones on endometrial cancer first gained attention after a steady rise in the incidence of endometrial cancer in the United States and Canada during the 1970's, as mentioned above. This trend coincided with a sharp rise in the use of hormone-replacement therapy (HRT) in postmenopausal women [61, 69]. As well, sequential oral contraceptives (OCs) doubled the risk of endometrial cancer in pre-menopausal women, until they were removed from the market in 1976[70]. Introduction of combination OCs, which contained a higher dose of progesterone, accordingly attenuated this risk[70]. These events suggested that a balance of estrogen and progesterone hormones is crucial for influencing endometrial cancer development. Indeed, these two hormones induce specific and fairly opposite biological responses and provide a basis for understanding most well known risk factors of endometrial cancer. A model for endometrial cancer development, named the "Unopposed Estrogen Hypothesis" explains how estrogen is a risk factor when present without the counterbalance of progesterone[24]. Where estrogen drives cell proliferation, progesterone neutralizes this effect by promoting differentiation, attenuating cell proliferation of endometrial tissue and leads to removal of the endometrial tissue during menstruation in premenopausal women[69]. The biological mechanism behind most risk factors is explained through this model, as they significantly affect risk through influencing circulating hormone levels.

This hypothesis is further supported by consistent findings of significantly higher circulating levels of estrogen and androgen among endometrial cancer cases, compared to controls[70-73].

2.6.3. Risk Factors

Factors that potentially influence the fluctuation of estrogen and progesterone have been investigated as potential contributors to the risk of endometrial cancer. The number of reproductive events is a well known example of risk attributed to changes in hormone exposure. Nulliparity is the state where a woman has not experienced a full-term pregnancy and is subsequently at an increased risk of endometrial cancer. This, in part, is due to the lack of a reduction in mitotic activity of the endometrium that normally occurs during pregnancy due to the persistently high progesterone levels[24]. OC use has consistently been shown to reduce risk of endometrial cancer, as these exogenous hormones contain a balance of both estrogen and progesterone[74]. Another factor that has been found to be protective is smoking. Smokers have about a 50% reduction in risk, which may be attributed to antiestrogenic properties[75, 76]. Menstrual events also influence risk. Early menarche and late menopause have been found to be associated with an increasing risk of endometrial cancer[71, 77]. HPNCC is an inherited syndrome that increases risk of mainly endometrial and colorectal cancer[78]. This syndrome is attributed to mutations in mismatch repair

genes which have a penetrance of approximately 60%, particularly for endometrial cancer[79].

Obesity is believed to be attributed to ≥ 40% of the incidence of endometrial cancer[80]. It is hypothesized that obese, post menopausal women have an increased conversion of androgens to estrogens in adipose tissue, which subsequently increases serum levels of estrogen and increases endometrial cancer risk among obese women[81].

2.6.4. Obesity and Endometrial Cancer Risk

Adiposity is strongly associated with increased estrogen levels; therefore, weight, body mass index (BMI, kg/m^2), hip and waist circumference, waist-to-hip ratio and weight gain since adulthood are all possible risk factors for endometrial cancer risk. Choosing an appropriate measure is complex since hormonal profiles differ between women with upper versus lower body obesity. Women with upper body obesity have significantly higher serum testosterone and estradiol levels[82]. However, women with lower body obesity have significantly higher aromatization of androstenedione to estrone[82] and it has also been shown that aromatase expression is greatest in buttock and thigh regions, compared to abdomen[83]. Further evidence is summarized in Table 4., which includes studies of obesity measures, hormone levels and endometrial cancer risk.

A multi-centre study done in the United States found that waist circumference was a strong predictor of endometrial cancer risk,

suggesting fat distribution is an important consideration. However, weight gain since early adulthood was the most significant predictor of risk[84]. A similar study done in Baltimore, USA found that women with upper body fat had a higher risk of endometrial cancer than women with lower body fat and that both quantity and location of fat appeared to be important risk factors[85]. A study in Florida found a large attributable risk associated with waist-to-hip ratio, where upper body adiposity was a significant risk factor. Although waist-to-hip ratio and waist circumference were significantly different among cases and controls, hip circumference was not[86].

One study included measures of estrogen and androgen levels to explain risk attributed to obesity. In addition to finding a larger relative risk of endometrial cancer with measures of BMI than with waist-to-hip ratio, BMI also had a stronger association with serum estrogen (but not with androgen levels)[87]. A large study done in Wisconsin, USA investigated various changes in body weight and endometrial cancer risk. They found significant risk associated with high body weight, BMI and weight gain since adulthood[88]. Three additional studies not involving endometrial cancer were found to contribute to this body of evidence, as they investigated various measurements of adiposity. One study compared anthropometric measurements with dual-energy x-ray absorptiometry readings. Although BMI was most highly correlated with total body fat, hip girth and age were the strongest predictors of total

and abdominal body fat[89]. Another study incorporated serum hormone levels and found that BMI, but not waist-to-hip ratio was associated with estrogen levels. Abdominal obesity, measured by waist-to-hip ratio was associated with androgenic hormone levels[90]. This is consistent with the findings of Kirschner *et al.*[82] who found that upper body fat was associated with increased androgen production, where lower body obesity was associated with increased aromatization of androstenedione to estrone. Other studies that have suggested that the relationship between obesity and increased endometrial cancer risk may be the result of increased production and level of estrogens rather than the increased availability of androgens[87, 91]. Fluctuations in serum hormone binding globulin (SHBG) detected in overweight and obese women are also important to consider. SHBG is a protein that binds and inactivates testosterone and estrogen and has been shown to have lower serum concentrations in those with upper body obesity[87, 90, 92], and be negatively associated with waist circumference but not BMI[93].

This literature review included recent articles that investigated various measures of obesity with endometrial cancer and hormone levels. Only studies that predominantly involved postmenopausal women were included, as the main source of estrogen in these women is derived from adipose tissue and is not confused with hormone production from the ovaries. In order to make informative comparisons with the present study, only articles that investigated similar anthropometric measures

(and evaluated more than one measure) were selected. The inconsistent evidence in Table 4. highlights the complicated relationship between obesity, body fat distribution and endometrial cancer. In addition to this evidence, one review article explains three main consequences of obesity: a decrease in SHBG, an increase in levels of estrogen and an increase in unbound testosterone[80]. Also noted, is a stronger association of increased weight and testosterone levels in women with upper body obesity, where levels of androstenedione are unaffected by excess weight. This review addresses the need for more epidemiological studies that investigate associations between obesity, hormone levels and endometrial cancer. Specifically, the need to identify common genetic predisposition factors that interact with obesity to influence endometrial cancer risk is also mentioned[80]. The present study addresses this gap in the literature by further including a genetic factor an investigation of obesity and endometrial cancer risk. A polymorphism within *CYP19* gene, encoding aromatase, is studied here, as it is a key enzyme involved in estrogen biosynthesis from adipose tissue. The combined effect of the *CYP19* (TTTA)n polymorphism with various anthropometric measurements will be investigated among cases of endometrial cancer and controls. Since endometrial cancer risk has been shown to be associated with weight, BMI, hip and waist circumference, waist-to-hip ratios and weight gain since adulthood, these measures will all be investigated within an Albertan population.

Table 4. Previous Research on Anthropometric Measures. Summary of various studies of endometrial cancer, hormone levels and anthropometric measurements.

	Author/Year	Study	Main Findings
Endometrial cancer studies	Swanson, CA 1993[81]	403 EC cases (74% menopausal), 247 controls (70% menopausal)	• BMI>30vs.<23: RR=2.0(1.2-3.3)[a] • Weight>78.3kgvs.<58.6: RR=2.3(1.4-3.7)[a] • WC>104cmvs.<81.7: RR=3.9(1.1-4.5)[b] • Weight gain since adulthood: ≥80lbsvs.0-19: RR=6.7(3.4-13)[a]
	Elliot, EA 1990[91]	46 EC cases, 140 controls (mean ages 62, 54 yrs of age)	• BMI≥27.3vs.<27.3: OR=2.3(1.1-4.9) • W-H>0.84vs<0.78: OR=3.2(1.2-8.9)[c] • W-H>84/BMI≥27.3vsW/H<0.78/BMI<27.3: OR=5.8(1.7-19.9)[c]
	Schapira, DV 1991[83]	40 EC cases and 40 controls (age,BMI matched); ~40% >55 yrs of age)	• W-H≥1.14vs<1.14: RR=15.0(1.98-58.0) • W-H (p<0.01) and WC (0.05) are significantly higher in cases, HC and weight not significant
	Austin, H 1991[84]	168 EC cases, 224 controls (most postmenopausal)	• BMI(kg/m$^{1.5}$)>36.4vs.≤28.4: RR=2.3(1.3-3.9)[d] • W-H>0.86vs≤28.4: RR=1.1(0.6-2.1)[d]
	Trentham-Dietz, A 2006[85]	740 cases, 2342 controls (83.4 and 90.6% postmenopausal)	• BMI>29.0vs.<22.7: OR=3.2(2.4-4.2) • Weight>77.1kgvs.<58.9: OR=3.4(2.6-4.5) • Change in weight since 20 yrs of age≥21kgvs.0-7kg: OR=2.5(1.7-3.8)
Anthropometric studies	Raja, C 2004[86]	75 postmenopausal women, compared DXA with anthropometric measurements	• Correlations with total body fat: BMI=0.93, WC=0.68, HC=0.86 • Hip girth and age strong predictors of total and abdominal body fat
	Austin, H 1991[84]	334 women (304 were postmenopausal)	• BMI[e] associated with estrone (p=<0.0001), estradiol (p=0.001) and androstenedione (p=0.08) but not with W/H[f] (p=0.83, 0.11, 0.085)
	Kirschner, MS 1990[79]	15 women w/ upper body obesity (W-H>0.85), 14 lower, age:20-40 yrs	• Upper body obesity associated with higher androgen levels (p<0.05) • Lower obesity associated increased aromatization of adrostenedione to estrone
	Kirchenglast, S 1994[90]	171 postmenopausal women	• Waist and hip girth (but not BMI) positively correlated with estradiol, (p<0.05), negatively associated with SHBG (p<0.05)
	Kaye, SA 1991[87]	88 postmenopausal women	• BMI more highly correlated with free estradione (0.45), where W-H was not (0.14) • Abdominal fat associated with androgenic but not estrogenic hormones

EC=Endometrial cancer BMI=kg/m² DXA=Dual-Energy X-Ray Absorptiometry
HC=hip circumference WC=waist W-H=waist-to-hip ratio

[a] adjusted for body fat distribution, age, education, number of births, estrogen use and smoking

[b] adjusted for age, education, number of births, menopausal estrogen use and smoking status

[c] adjusted for age, W-H, parity, BMI and smoking

[d] adjusted for age, race, years of schooling and BMI (or indicator variables for BMI quartiles)

[e] adjusted for age, race and years of schooling

[f] adjusted for age, race, years of schooling and BMI

2.7. Research Questions

The present study evaluates the risk of endometrial cancer attributed to the tetranucleotide repeat polymorphism in *CYP19* specifically among postmenopausal women. This association is further stratified by various measures of obesity including weight, BMI, waist and hip circumference, waist-to-hip ratio and weight gain since adulthood. The following questions will be addressed:

1. Is the risk of endometrial cancer higher in women with both alleles having $\geq$10 (TTTA) repeats compared to those with at least one allele having <10 (TTTA) repeats?
2. Which anthropometric measure (weight, BMI, waist and hip circumference, waist-to-hip ratio and weight gain since adulthood) is most associated with endometrial cancer risk?
3. Is the expected association between longer alleles of CYP19 and endometrial cancer strongest among heavier women?

In assessing the extent to which obesity modifies the risk attributed to the *CYP19* tetranucleotide repeat in aromatase, this study may further contribute to understanding the relationship between estrogen production, obesity, fat distribution and endometrial cancer.

Chapter 3

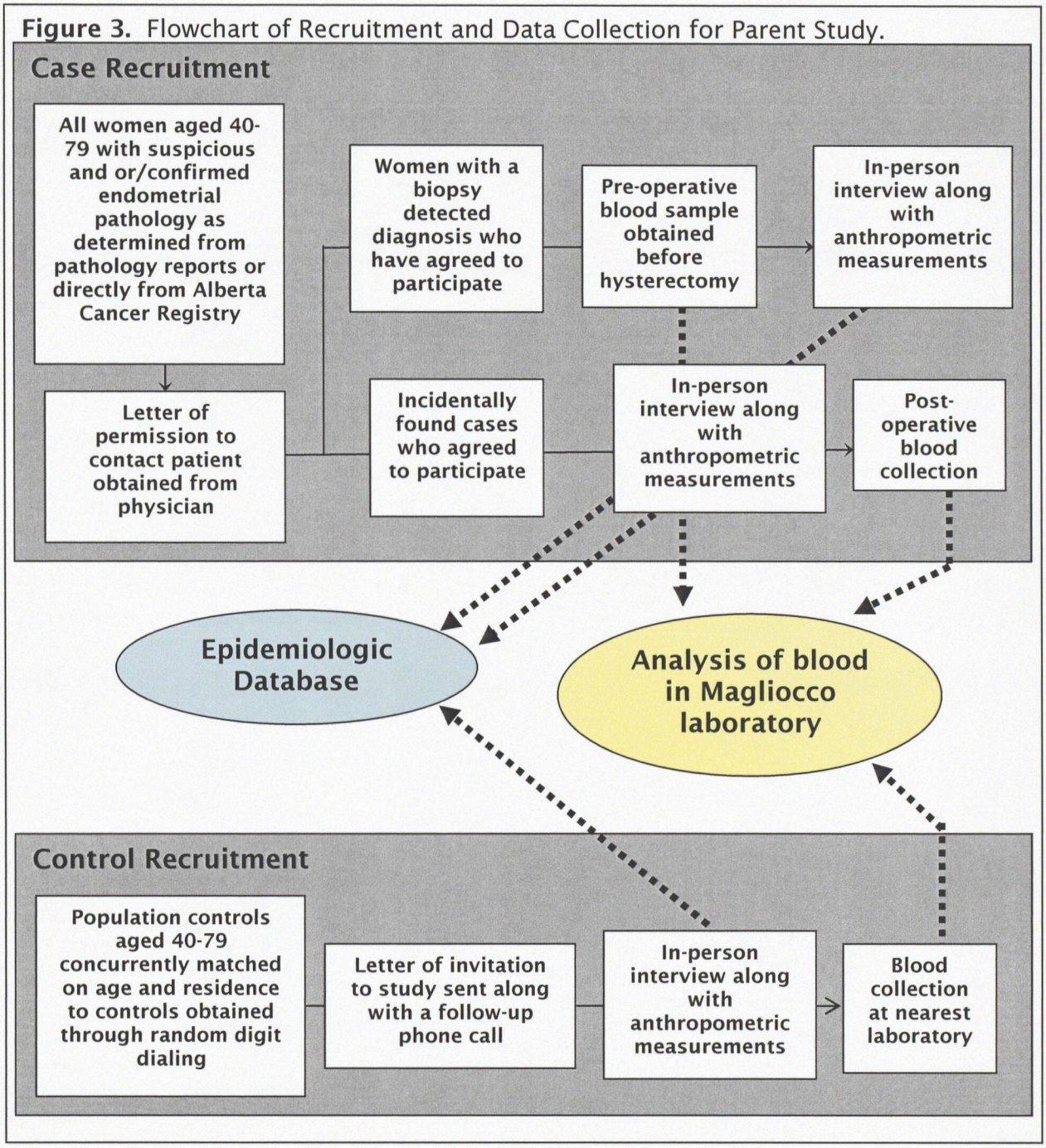
Study Design and Methods

3.1. Parent Study

3.1.1. Description of Study Design

This study is a sub-sample of a larger study of endometrial cancer conducted in Alberta from 2001-2006 led by Drs. Friedenreich, Cook and Magliocco. An interview-administered questionnaire and blood data were available for this study. For the current analysis, epidemiologic and biologic data were used to investigate how genetic predisposition and various measures of obesity and adiposity influence endometrial cancer risk.

Figure 3. Flowchart of Recruitment and Data Collection for Parent Study.



3.1.2. Population: Case and Control Definition

Cases recruited for the parent study included Albertan women with incident, histologically confirmed, invasive primary endometrial cancer. Cases needed to be able to speak English to complete the interview, were between the ages of 40 to 80 years and had not been excluded by their physicians because of their illness. Pathology reports were sent electronically (Edmonton) or by mail (Calgary) to the Alberta Cancer Registry. With cooperation of the directors of pathology laboratories in Alberta, pathology reports were expedited throughout the course of the study. The Study Coordinator reviewed the pathology reports in consultation with the study Pathologist. Once a diagnosis was confirmed, the Study Coordinator contacted the referring physician for permission to contact the cases.

Controls were Albertan women who had not been diagnosed with endometrial cancer and had an intact uterus. They were identified using random digit dialing (RDD) and were frequency matched to cases on age (±5 years) and place of residence (rural versus urban). The RDD method involved an initial screening and interview process to determine if any women in the household were eligible according to age, hysterectomy status and history of cancer. Eligible women were then offered a study package.

Cases with urban residence (Calgary or Edmonton) were matched with controls of the same city. For other major cities such as Red Deer

and Medicine Hat, controls were sampled within the same city, whenever possible. Controls were sampled concurrently with each case ascertainment to avoid possible biases introduced by secular changes over time that may influence any exposures.

3.1.3. Data Collection

i. Blood Collection

Once permission was obtained from the referring physician, cases were sent an introductory letter and consent form, along with a laboratory requisition to donate a preoperative blood sample. Since a pre-operative blood sample is required before each surgery, it was requested that extra blood be drawn for the study. For cases in the parent study, an average of 109 days passed from time of blood collection to interview. Post-operative blood samples were obtained from cases who were incidentally detected through a hysterectomy performed for another reason besides that of endometrial cancer or if a pre-operative sample could not be taken for whatever reason. The blood requisition form was given to each post-operative case at the time of the interview. Blood samples were collected, on average, 39 days after the interview for these participants. Blood kits and requisition forms were given to each control at the time of the interview. Collection of blood samples for controls was done at their nearest convenient laboratory. The lag time between interview and blood collection for controls averaged 53 days.

Once collected, the blood samples were processed, aliquoted and frozen at -80°C. At regular intervals, the participating laboratories in Calgary, Edmonton, Medicine Hat and Lethbridge sent the frozen samples on dry ice to the Tom Baker Cancer Centre where they were stored in ultralow freezers until they were retrieved for the laboratory analysis.

ii. Interview Process

Letters inviting participants for an interview were sent out immediately after identification to schedule an interview after their surgery for cases and after recruitment for controls and incidentally found cases. The interview process is intended to extract a detailed history of past exposure events such as reproductive history and exogenous hormone use. Cognitive interviewing techniques developed by the Office of Research and Methodology, National Center for Health Statistics[95] were used to aid in participant recall of past events. Blaise® interview software assists the interviewer with collecting all necessary information for the study[96].

iii. Anthropometric measurements

After the interview, all study participants were weighed three times. The average value of these measurements was used to determine current weight. Height was measured, along with waist and hip circumference. A history of weight gain, loss and cycling was obtained by participant recall for each decade from 20 to up to 60 years of age.

3.2. Research Methods

3.2.1. Population

This case-control study utilized participants of the Parent study described in Section 2.7. A flow chart found in Figure 3. summarizes patient recruitment and data. Cases for the present study are post-menopausal Albertan women, 79 years of age or younger, with incident, invasive primary endometrial cancer diagnosed between September 1, 2002 and June 30, 2006. Controls are Albertan women selected through RDD and frequency-matched to cases on age and urban versus rural residence. All participants were without a previous cancer diagnosis (except non-melanoma skin cancer).

Restriction criteria based on menopausal status and HRT use were established to isolate for the effect of *CYP19* genotype and obesity. Pre-menopausal women and women with significant exposure to HRT may have estrogen levels that can overwhelm the influence attributed to the *CYP19* polymorphism. Therefore, the present study only includes post-menopausal women who were never-users or very short-term users (<6 months) of HRT.

3.2.2. Laboratory Methods for Genetic Polymorphism

All laboratory work was funded by a research grant from the Department of Obstetrics and Gynecology through a Centre of Advancement in Health funding competition. Pre-existing aliquots of buffy coat (stored at -80^{c}) derived from the parent study were used for

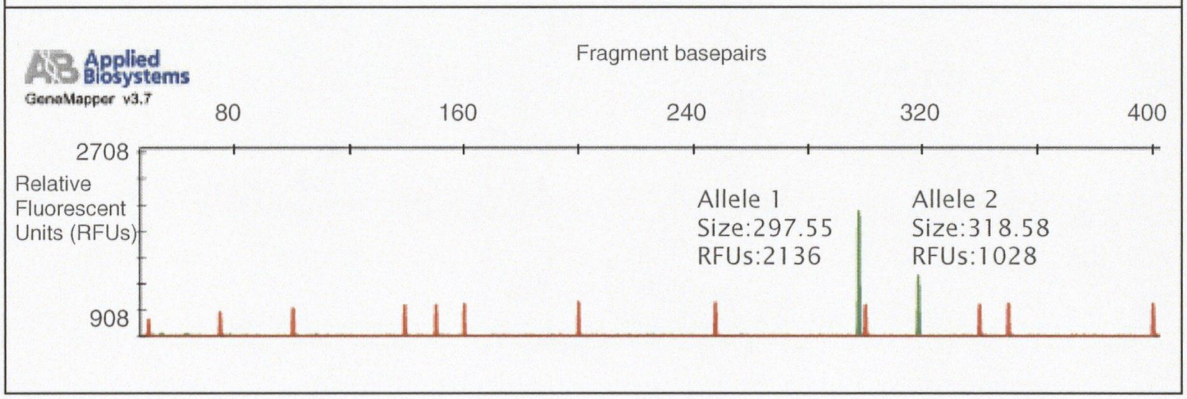
genetic analysis of the *CYP19* tetranucleotide polymorphism. DNA was extracted from white blood cells using a QIAamp® DNA Maxi Blood Extraction Kit (Qiagen Inc., Mississagauga, ON). Afterwards, DNA concentrations were then measured by UV absorbance spectrometry. The isolated DNA samples were amplified using the polymerase chain reaction (PCR) (forward primer: 5'-GTC TAT GAA TGT GCC TTT TT-3', and reverse primer: 5'-GTT TGA CTC CGT GTG TTT GA -3') (primer sequence courtesy Dr. Chu Chen, Fred Hutchinson Cancer Research Center, Seattle, Washington), in a Bio-rad iCycler DNA thermocycler (Applied Biosystems, Foster City, CA). PCR products were verified using agarose gel electrophoresis with ethidium bromide stain and illuminated under UV light. DNA extractions and amplification were all done in the laboratory of Dr. Anthony Magliocco by the student.

Polymorphism analysis was done using a technique called Amplified Fragment Length Polymorphism (AFLP). This method of DNA fingerprinting offers the required sensitivity to detect four nucleotide differences, which is necessary to identify the variable number of tetranucleotide repeats. AFLP analysis was done by the University of Calgary Core DNA and Protein Services, in the Faculty of Medicine. Applied Biosystems had adapted this AFLP technique for use with its ABI PRISM™ fluorescent dye-labeling and detection technology. The 96 well plate, in-capillary detection system utilizes dual-side illumination to detect the hexachlorofluorescein (HEX) fluorescently labeled forward

primer. During electrophoresis, the instrument monitors the passage of the fragments through polymer capillary system by detecting fluctuations in emitted light when the fragments migrate past a fixed argon-ion, multi-line laser beam[97].

After the sample data were collected, GeneMapper® Analysis software, version 3.7 (Applied Biosystems, Foster City, CA) was used to analyze and display tabular data and electropherograms. The electropherograms provides visual data, describing the size of the fragment detected and the quantity of the product (see Figure 4.). This graph plots the relative fluorescent units obtained from the radiolabelled primers and the number of nucleotides detected in the fragment.

Figure 4. Sample Electropherogram of AFLP Results from Individual X. Using Applied Biosystems 3730 DNA Analyzer and GeneMapper software, results indicate that Individual X is heterozygous with 7 and 11 (TTTA) repeat alleles. Red peaks indicate the set of standards labeled with ROX, while the green peaks represent the HEX labeled sample.



Each sample was labeled with HEX, which is displayed in green, and was run with its own set of standards labeled with rhodamine X (ROX), shown in red. The standards include a collection of DNA fragments of a

variety of lengths, which provide a benchmark to determine fragment size of each sample. The primer pair above was found in the *CYP19* sequence (Homo sapiens chromosome 15 clone RP11-522G20 map 15q21.2) derived from the National Center for Biotechnology Information's online nucleotide alignment program[98]. From this sequence information, the expected fragment size was found to range from 298-323 nucleotides and the repeat number that corresponds to each fragment was determined.

3.2.3. Data Analysis

All data was initially organized in Microsoft Excel® version 11.2.3 and then exported into Intercooled STATA® version 8.2. Tests for differences in *CYP19* allele and genotype frequencies, anthropometric measurements and selected potential confounding factors between cases and controls were done using unpaired two sample mean and proportion t-tests. Categorical variables were assessed using ANOVA analysis. Difference in weight gain since adulthood was calculated by subtracting current weight from reported weight at 20 years of age. Also considered was risk attributed to HNPCC. Identification of those with HNPCC was done using family history given by recall from each participant and the Amsterdam Criteria II[16]. Any cancer diagnosis was recorded for each family member affected, along with the age at diagnosis and type of cancer. The entire family history of each participant was examined to determined HNPCC status according to the following criteria:

1. Having at least three relatives with an HNPCC-associated cancer (colon, endometrium, ureter or renal pelvis).

2. One affected person was a first-degree relative of the other two

3. Two successive generations were affected and at least one relative was diagnosed before 50 year of age.

This tool has been found to have a sensitivity and specificity of up to 95 and 65%[16]. Use of HNPCC exposure allows for consideration of family history of endometrial cancer that may be independent of the *CYP19* (TTTA)n polymorphism, as no gene association studies have linked *CYP19* and the mismatch repair genes of HNPCC.

Multivariate logistic regression was used to compute odds ratios and 95% confidence intervals (95%CI) for the main effects of genotype and anthropometric measurements. Regression analysis was adjusted for age and residence, as cases and controls were frequency matched for these factors in the parent study. Further adjustment included other factors that potentially influence endometrial cancer risk including smoking status, parity, OC use, age at onset of menarche and menopause, education, ethnicity and HNPCC status. Regression analysis for interaction of genotype and anthropometric measurements were done by creating dummy variables to categorize each risk group. An example of the regression model assessing interaction of genotype with the anthropometric measurement of current weight is found below. The

dependent variable log(p/1-p) is the logarithm of the odds of endometrial cancer, where p is the probability of endometrial cancer.

$$\log(p/1-p)= \beta_0 + \beta_1(\text{age}) + \beta_2(\text{residence}) + \beta_3(\text{births}) + \beta_4(\text{OC use}) + \beta_5(\text{menarche age}) + \beta_6(\text{menopause age}) + \beta_7(\text{smoking status}) + \beta_8(\text{genotype/weight risk groups})$$

where age is coded as

0	40-44
1	45-49
2	50-54
3	55-59
4	60-64
5	65-69
6	70-74
7	75-79

residence is coded as

0	rural
1	urban

smoking status is coded as

0	never smoked
1	(ex) occasional
2	ex-smoker
3	current smoker

genotype/weight risk group dummy variables are coded as

d1	d2	genotype/weight risk groups
0	0	≥10<10 or <10/<10/<160lbs
1	0	≥10<10 or <10/<10/<160lbs
0	1	≥10/≥10/≥160lbs
1	1	≥10/≥10/≥160lbs

Number of births, months of OC use, age at menarche and age at menopause were coded as continuous variables. Since the addition of education, ethnicity and HNPCC status did not notably change the

coefficient corresponding to the main effects, these factors were not included in the final analysis. This regression analysis was further used to assess interaction of genotype and body mass index (BMI), waist and hip circumference, waist-to-hip ratio and weight gain since adulthood on endometrial cancer risk.

3.2.4 Sample Size and Power Considerations

Based on previous studies, *a priori* estimates included a prediction that 53% of controls would have at least one ≥10(TTTA) repeat allele and that this factor would be associated with an elevated endometrial cancer risk (RR=1.9)[48]. With 132 cases and 264 controls, 80% power would be achieved to detect an odds ratio of 1.9 with α= 0.05 (two-tailed) in comparing women with at least one ≥10(TTTA) repeat allele compared to those with <10(TTTA) repeat alleles.

3.2.5. Ethical Consideration

This approval for the study was based on the consent the women provided for the parent study. Please note the following excerpt below under the *Confidentiality* section, which informs the study subjects that additional analysis may be done with the information and materials collected in the parent study.

"All material and data obtained from this study will be stored and may be used for future analysis without obtaining further consent from you. Future studies will be based on important research questions and the up-to-date methods that are available to address these questions.

Even though the exact tests that will be done in the future are not known, it is possible that they would include genetic (inherited) factors, nutrients such as vitamins and minerals, hormones, infectious agents such as viruses and bacteria, and indicators of environmental exposures."

Approval for the present study was updated by the Conjoint Health Research Ethics Board at the University of Calgary.

CHAPTER 4

Results

4.1 Characteristics of Study Population

Cases identified through Alberta Cancer Registry had a response percentage of 62%. Potential controls were identified using RDD. The screening response rate was 61%, which represents the percent residential phone lines for which we were able to determine household eligibility. The interview response rate of 53%, indicates the percent of those screened eligible that completed the in-person inteview. Therefore, an overall response rate of 32% was achieved during recruitment of all controls in the parent study.

For the present study, 379 cases and 1014 controls were identified from the parent study between September 1st, 2002 and April 30th, 2005. Of those excluded because of their recent history of HRT use, 173 were cases and 458 were controls. Since only postmenopausal women were included, 61 cases and 236 controls were further eliminated because they were pre or perimenopausal. Blood samples were not collected from 6 cases and 24 controls and 10 cases and 27 controls were excluded because of incomplete data. After these considerations, participants included 398 postmenopausal women who had no recent use (within the last 6 months) of HRT. The final analysis included 127 endometrial cancer cases and 271 controls.

Characteristics of cases and controls are described in Table 5. On average, cases were younger than controls (60.35 versus 63.37 years of age, $p<0.01$) and the proportion of cases and controls in urban versus rural areas did not differ significantly (0.66 versus 0.60, $p=0.22$). No significant difference between cases and controls was found for education ($p=0.94$), ethnicity ($p=0.98$), smoking status ($p=0.08$) and HNPCC status ($p=0.94$). Based on averages, cases had given birth fewer times than controls (2.36 versus 2.97, $p<0.01$) and had a shorter history of OC use (26.01 versus 41.85 months, $p<0.01$). The average age at menarche and menopause did not differ significantly between cases and controls (p-value=0.27, 0.26). In Table 5. the distribution suggests that cases are younger at menarche and older entering menopause, as shown in other studies[77, 99].

Table 5. Characteristics of Study Population. Distribution and frequency (%) of demographic, reproductive and menstrual characteristics among endometrial cancer cases (N=127) and controls (N=271).			
Characteristic		**Cases** N (%)	**Controls** N (%)
Age	40-44	2 (2)	1 (<1)
	45-49	3 (2)	2 (<1)
	50-54	19 (15)	39 (14)
	55-59	43 (34)	52 (19)
	60-64	26 (20)	48 (18)
	65-69	19 (15)	60 (22)
	70-74	8 (6)	45 (17)
	75-79	7 (6)	24 (9)
Residence	Urban	84 (66)	162 (60)
	Rural	43 (34)	109 (40)
Education	High school or less	45 (36)	101 (37)
	College/technical	60 (47)	125 (46)
	University	22 (17)	45 (17)
Ethnicity	Caucasian/European	123 (97)	261 (96)
	other/mixed	4 (3)	10 (4)
Number of births	0	20 (16)	22 (8)
	1-3	78 (61)	165 (61)
	4+	29 (23)	84 (31)
Months of OC use	≤6	62 (49)	109 (40)
	7-59	41 (15)	89 (33)
	≥60	24 (19)	73 (27)
Age at menarche	<11	11 (9)	13 (5)
	11-12	61 (48)	119 (44)
	>13	55 (43)	139 (51)
Age at menopause	<45	12 (10)	24 (9)
	45-55	92 (72)	220 (81)
	>55	23 (18)	27 (10)
Smoking status	never smoked	76 (60)	146 (54)
	(ex) occasional	7 (6)	9 (3)
	ex-smoker	31 (24)	89 (33)
	current smoker	13 (10)	27 (10)
HNPCC status	Yes	2 (2)	4 (2)
	No	125 (98)	267 (98)

Table 6. displays the distribution of selected anthropometric measurements among cases of endometrial cancer and controls. On average, cases had consistently larger measurements of current weight (192.34 versus 165.16 lbs, $p<0.01$), BMI (33.36 versus 28.23 kg/m^2,

p<0.01), waist circumference (99.14 versus 89.99 cm, p<0.01), hip circumference (116.98 versus 108.43 cm, p<0.01), waist-to-hip ratio (0.85 versus 0.83, p=0.04) and weight gain since adulthood (63.73 versus 43.44 lbs, p<0.001). Using dichotomous variables for anthropometric categories seemed appropriate for this analysis due to the nature of the data and the study objectives. Creating dichotomous categories allows for the comparison of endometrial cancer risk between anthropometric measures. We would not expect to see any peculiar pattern if each anthropometric category was further subdivided (ie: risk is expected to increase as each measure increases). Also, using dichotomous variables allows for a feasible interaction analysis with the *CYP19* (TTTA)n polymorphism[20].

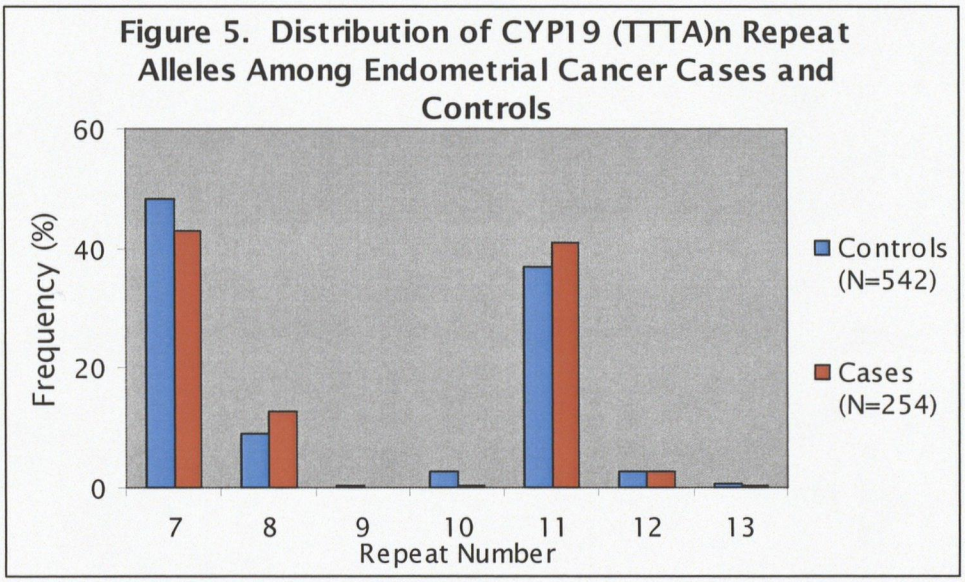
Exposure groups for current weight, hip circumference, waist-to-hip ratio and weight gain since adulthood were chosen based on the 50th percentile of the control group. Dichotomous categorization of BMI and waist circumference were derived from criteria established from the National Institutes of Health (NIH)[100], although, the 50th percentile of waist circumference of this dataset equaled the median circumference for this control population. Current height was also assessed (data not shown), however it did not differ significantly between cases and controls.

Table 6. Anthropometric Measures. Distribution of selected anthropometric measures among cases of endometrial cancer (N=127) and controls (N=271).			
Anthropometric Measure		**Cases** N (%)	**Controls** N (%)
Current weight (lbs)	<160	32 (25)	138 (51)
	≥160	95 (75)	133 (49)
BMI (kg/m²)	<30	51 (40)	174 (64)
	≥30	76 (60)	97 (36)
Waist circumference* (cm)	<88	33 (26)	132 (49)
	≥88	92 (74)	138 (51)
Hip circumference* (cm)	<106	39 (31)	136 (50)
	≥106	86 (69)	134 (50)
Waist-to-hip ratio*	<0.80	38 (30)	109 (40)
	≥0.80	87 (70)	161 (60)
Weight gain since adulthood (lbs)**	<40	42 (33)	137 (51)
	≥40	84 (66)	134 (49)
* missing data: two case and one control ** missing data: one case			

4.2. Allele and Genotype Distribution of *CYP19* (TTTA)n Polymorphism

Figure 5. highlights the distribution of (TTTA)n repeat alleles in *CYP19* among cases of endometrial cancer and controls (data supporting this graph can be found in Table 5a. of the appendix). Note that each individual contributes a paternal and maternal allele to this measure of frequency. Allele length varied from 7-13 (TTTA) repeats, where the 7 and 11 (TTTA) repeat alleles were the most common in this population of Albertan women. While the 8 and 11 (TTTA) repeats were slightly more common in cases than controls, and the 7 (TTTA) repeat was slightly more common in controls than cases, none of these differences were statistically significant. Other alleles were so rare that no conclusions can

be made regarding their distributions; with an increased sample size, the distribution of these rare alleles is subject to change. How these alleles combine into genotypes of case and controls is shown in Table 7. A significant difference between the proportion of cases and controls homozygous for 11 (TTTA) repeats was detected (p=0.04). More controls than cases exhibited the 7/11 genotype, however the difference between cases and controls was not significant.



4.3. Main Effects of (TTTA)n Repeat Polymorphism

The main effect of genotype was assessed individually with endometrial cancer risk. Odds ratios were adjusted for age and residence, as cases and controls were crudely frequency matched on these factors, according to the study design. Odds ratios were further adjusted for weight and then for number of births, OC use, age at menarche, age at menopause and smoking status. The risk associated

with having two alleles with ≥10 (TTTA) repeats, compared to those with at least one allele with <10 (TTTA) repeats, was associated with an odds ratio of 1.41 (95% CI: 0.83-2.40), as shown in Table 8. With limited sample size, a non-significant risk estimate was anticipated due to the high frequency of higher risk alleles within the control population. After adjustment for weight, this risk estimate is attenuated (OR=1.05, 95%CI: 0.58-1.90).

Table 7. Genotype Frequencies.
Distribution of (TTTA)n *CYP19* genotypes among case and controls.

Genotype	Frequency in Cases (%)	Frequency in Controls (%)
7/7	27 (21)	59 (22)
7/8	15 (12)	24 (9)
7/9	0	1 (<1)
7/10	1 (<1)	6 (2)
7/11	36 (30)	102 (38)
7/12	2 (2)	9 (3)
7/13	1 (<1)	2 (1)
8/8	2 (<1)	3 (1)
8/10	0	1 (<1)
8/11	11 (9)	16 (6)
8/12	0	1 (<1)
8/13	0	1 (<1)
10/10	0	1 (<1)
10/11	0	5 (2)
11/11	27 (21)	36 (13)*
11/12	3 (2)	4 (2)

*difference is significant (p<0.05)

Although the *a priori* hypothesis predicted that alleles containing ≥10 (TTTA) repeats would be associated with increased risk of endometrial cancer, allele length from each possible repeating unit was also assessed. Table 8a. of the appendix summarizes these findings. An elevated risk of 1.65 (95%CI: 1.03-2.63) was found for having with both

alleles with ≥8 (TTTA) repeats, compared to those with at least one allele with 7 (TTTA) repeats. Similarly, this risk estimate became non-significant after adjustment for weight (OR=1.35, 95%CI: 0.60-1.88).

Table 8. Main effects of *CYP19* (TTTA)n Polymorphism. Risk associated with having both alleles with ≥10 (TTTA) repeats compared to having at least one allele with <10 (TTTA) repeats.

Genotype	Case (N=127)	Control (N=271)	OR (95%CI)[a]	OR (95%CI)[b]	OR (95%CI)[c]
≥10/≥10 vs. ≥10/<10 or <10/<10	30	46	1.41 (0.83-2.40)	1.05 (0.60-1.88)	1.05 (0.58-1.90)
	97	225	1	1	1

[a] adjusted for age and residence [b] adjusted for age, residence and weight
[c] adjusted for age, residence, weight, number of births, OC use, age at menarche, age at menopause and smoking status

4.4. Main Effects of Anthropometric Measures

Significant elevations in endometrial cancer risk were found with all anthropometric measurements, as described in Table 9. Current weight was associated with the highest risk, where those individuals that weighed 160 lbs or more had an odds ratio of 3.13 (95% CI:1.89-5.17), when compared to those who weighed less that 160 lbs. Waist circumference of 88 cm or more versus measures less than 88 cm was associated with an odds ratio of 2.82 (95% CI:1.73-4.60). This risk estimate is larger than that of hip circumference, which was 2.27 (95% CI:1.41-3.66) in those with a measure of 106 cm or more, compared to measures of less than 106 cm. BMI of 30 or more is associated with an odds ratio of 2.68 (95% CI:1.69-4.27), when compared to those with a BMI of less than 30. Those with a waist-to-hip ratio of 0.80 or greater is associated with a 1.92 (95% CI:1.18-3.13) increase in risk, when

compared to those with a waist-to-hip ratio of <0.80. Finally, those who had a weight gain since adulthood of 40 lbs or greater had an elevation of risk of 2.31 (95% CI:1.43-3.74), compared to those with a weight gain of < 40 lbs.

Table 9. Main Effects of Anthropometric Measurements. Risk associated with endometrial cancer according to independent effects of *CYP19* genotype and various measures of obesity.

Anthropometric measure	Risk Groups	Case N=127	Control N=271	OR(95%CI)[a]	OR(95%CI)[b]
Current weight (lbs)	≥160	95	133	2.99(1.86-4.82)	3.13(1.89-5.17)
	<160	32	138	1	1
BMI (kg/m²)	≥ 30	76	97	2.63(1.69-4.08)	2.68(1.69-4.27)
	<30	51	174	1	1
Waist circumference (cm)*	≥ 88	92	138	2.68(1.68-4.27)	2.82(1.73-4.60)
	<88	33	132	1	1
Hip circumference (cm)*	≥ 106	86	134	2.22(1.41-3.50)	2.27(1.41-3.66)
	<106	39	136	1	1
Waist-to-hip ratio*	≥ 0.80	87	161	1.80(1.12-2.87)	1.92(1.18-3.13)
	<0.80	38	109	1	1
Weight gain since adulthood (lbs)**	≥ 40	84	134	1.93(1.23-3.03)	2.31(1.43-3.74)
	<40	42	137	1	1

[a] adjusted for age and residence

[b] adjusted for age, residence, number of births, OC use, age at menarche, age at menopause and smoking status

* missing data: two cases and one control ** missing data: one case

4.5. Interaction between Polymorphism and Anthropometric Measurements

Each anthropometric measurement was assessed for interaction with the genotype in the endometrial cancer risk relationship. Each reference group was based on the low risk genotype, which includes those with at least one allele with <10 (TTTA) repeats and the low risk anthropometric measure. A consistent pattern is shown in each evaluation, where the highest risk group includes those with both alleles

with ≥10 (TTTA) repeats and with a high risk anthropometric measure. It is clear from this table that genotype and anthropometric measures synergize to influence endometrial cancer risk. Risk estimates were further stratified to include a heterozygous risk group and a group homozygous for the <10 (TTTA) repeat allele. These results are found in Table 10a. of the appendix. Similar risk estimates were found for heterozygous individuals (having one ≥10 (TTTA) repeat allele and one <10 (TTTA) repeat allele) and homozygous individuals (having both alleles with <10 (TTTA) repeats). Therefore, these subgroups were collapsed to produce the reference group of individuals with at least one allele containing <10 (TTTA) repeats.

Those who reported a current weight of ≥160 lbs and had both alleles with ≥10 (TTTA) repeats had the greatest endometrial cancer risk (OR=4.14, 95%CI:1.94-8.85). Age and residence adjusted odds ratio were similar to the more fully adjusted odds ratios, indicating that the other risk factors of endometrial cancer are only weakly related to the exposure of *CYP19* genotype, as anticipated. The likelihood ratio test was used as a test of the goodness-of-fit between models with and without an interaction term. However, based on a chi-square distribution, the difference in the likelihood scores did not differ significantly ($p>0.10$).

Table 10. Gene-environment Interactions. The influence of the *CYP19* (TTTA)n polymorphism on endometrial cancer risk is modified by measurements of current weight (lbs), BMI (kg/m^2), waist and hip circumference (cm), waist-to-hip ratio and weight gain since adulthood (lbs).

Anthropo-metric Measure	G	E	Case N=127	Control N=271	OR(95%CI)[a]	OR(95%CI)[b]
Current weight (lbs)	≥10/≥10	≥ 160	24	25	3.98(1.94-8.19)	4.14(1.94-8.85)
		< 160	6	21	1.25(0.46-3.44)	1.08(0.39-3.04)
	<10/<10 or <10<10	≥ 160	71	108	2.91(1.71-4.94)	2.95(1.69-5.16)
		< 160	26	117	1	1
BMI (kg/m^2)	≥10/≥10	≥ 30	22	21	3.26(1.62-6.56)	3.24(1.56-6.75)
		< 30	8	25	1.11(0.46-26.7)	1.01(0.41-2.47)
	≥10/<10/ or <10/<10	≥ 30	54	76	2.50(1.52-4.09)	2.52(1.49-4.24)
		< 30	43	149	1	1
Waist circumfe-rence (cm)*	≥10/≥10	≥ 88	23	24	3.57(1.72-7.37)	3.68(1.71-7.92)
		< 88	6	21	1.25(0.48-3.26)	1.16(0.44-3.08)
	≥10/<10/ or <10/<10	≥ 88	69	114	2.60(1.54-4.38)	2.72(1.58-4.69)
		< 88	27	111	1	1
Hip circumfe-rence (cm)*	≥10/≥10	≥ 106	22	25	2.92(1.44-5.94)	2.84(1.35-5.94)
		< 106	7	20	1.23(0.47-3.21)	1.17(0.44-3.12)
	≥10/<10/ or <10/<10	≥ 106	64	109	2.13(1.29-3.57)	2.21(1.29-3.76)
		< 106	32	116	1	1
Waist-to-hip Ratio*	≥10/≥10	≥ 0.80	21	27	2.54(1.23-5.25)	2.64(1.23-5.63)
		< 0.80	8	18	1.19(0.47-3.07)	1.06(0.40-2.81)
	≥10/<10/ or <10/<10	≥ 0.80	66	134	1.71(1.01-2.89)	1.80(1.05-3.09)
		< 0.80	30	91	1	1
Weight gain since adulthood (lbs)**	≥10/≥10	≥ 40	25	25	2.80(1.41-5.54)	3.11(1.52-6.40)
		< 40	5	21	0.73(0.25-2.10)	0.69(.24-2.03)
	≥10/<10/ or <10/<10	≥ 40	59	109	1.63(0.99-2.67)	1.96(1.15-3.32)
		< 40	37	116	1	1

G: genetic factor E: environmental factor

[a] adjusted for age and residence

[b] adjusted for age, residence, number of births, OC use, age at menarche, age at menopause and smoking status

* missing data: two cases and one control ** missing data: one case

CHAPTER 5

Discussion

5.1. Main Effect of *CYP19* (TTTA)n Polymorphism

There was a suggestion of an elevated risk of endometrial cancer for individuals with alleles containing ≥10 (TTTA) repeats. Where this estimate did not reach statistical significance, risk associated with having both alleles with ≥8 (TTTA) repeats was significant. These findings suggest that an elevated risk may exist in those with longer CYP19 (TTTA)n repeat alleles. However, these elevated risks attenuated in the present study when weight was included in the analysis. A measure of obesity is an important consideration in studying endometrial cancer, as it is such a large contributor to risk of the disease. Further controlling for other risk factors of endometrial cancer did not greatly change these findings. This was anticipated since factors such as nulliparity and smoking status were not expected to be associated with CYP19 genotype.

These genotype results emphasize the importance of considering obesity when estimating risk of endometrial cancer associated with the *CYP19* (TTTA)n polymorphism that was not done in previous studies[48, 49].

5.2. Review of *CYP19* Evidence

The distribution of *CYP19* (TTTA)n allele frequencies found in this study was analogous to the distribution in studies that investigated this polymorphism in other populations[38, 39, 43, 48, 101-105]. These studies also found the most common allele to contain 7 (TTTA) repeats, followed by

the 11 (TTTA) repeat allele. Since *CYP19* encodes aromatase, a key enzyme in estrogen synthesis, associations between polymorphisms in this gene and breast and endometrial cancer have been examined. Kirstensen *et al.* found that individuals carrying the 12 (TTTA) allele in a Scandinavian population have an increased risk of breast cancer (OR=2.42, 95%CI: 1.03-5.80)[47]. Haimen *et al.* reported an increased risk in those carrying one 10 (TTTA) allele (OR=2.87, 95%CI: 1.20-6.87) among women in the Nurses' Health Study[43]. A study done in Philidelphia, Pennsylvania found that the 12 (TTTA) allele was over-represented in breast cancer cases, and that a shorter allele of 171 base pairs (presumably the 8 (TTTA) repeat allele) was associated with an odds ratio of 1.47 (95%CI: 0.99-2.17)[38]. Han *et al.* found an odds ratio of 1.83 (95%CI: 1.14-2.93) for those heterozygous with the 10 (TTTA) repeat allele in a Chinese population[103] and Miyoshi *et al.* found an odds ratio of 1.80 (95%CI: 0.97-3.36) for those with ≥10 (TTTA) repeat alleles in a Japanese population[41]. A slightly increased breast cancer risk was found for those with the 10 (TTTA) repeat alleles in two other studies, however these estimates were not significant[44, 102].

Fewer studies have investigated this polymorphism with risk of endometrial cancer. One study contained a subset of women from the Nurses' Health study. An increased risk was detected for post-menopausal women with at least one allele >7 (TTTA) repeats (OR=1.97, 95%CI: 1.25-3.12) and for those with both alleles >7 (TTTA) repeats

(OR=1.92, 95%CI:1.17-3.14)[49]. Haplotype analysis of the *CYP19* gene was also performed. One particular haplotype was associated with a small but significantly higher risk of endometrial cancer (OR=1.30, 95%CI: 1.03-1.65) when compared to a pooled group of all other haplotypes (no anthropometric measure was considered). Alleles containing 8, 10 and 11 (TTTA) repeats cosegregate with this haplotype.

Another study conducted in a Russian population detected similar distributions of the repeat alleles in postmenopausal women with and without endometrial cancer. Results from this study were not translated in terms of repeating units, however, cases had a higher frequency of the A6 and A7, as was also found in our population (these alleles are identified as the 11 and 12 (TTTA) repeat alleles in the present study). They also found that controls were more likely to have the A2/A2 genotype (homozygous for the 7 (TTTA) repeat allele, p=0.03)[48]. The only inconsistency when comparing the Russian study's allelic distribution with the present study data involved the frequency of the 8 (TTTA) repeat. They found this allele over represented in controls, where it was found in a higher frequency of cases in our Albertan population. Although no estimate for risk of endometrial cancer was presented in this study, estradiol and testosterone levels were found to be non-significantly higher in cases with A6 and A7 genotypes (p>0.20).

One factor overlooked in these two studies of endometrial cancer risk and the *CYP19* (TTTA)n polymorphism is the contribution to risk made by obesity. The present study addresses this limitation.

5.3. Interaction of Anthropometric Measurements

Interpreting the extent to which obesity and aromatase activity interact to influence endometrial cancer risk is a main objective in this study. An analysis of the *CYP19* (TTTA)n polymorphism and endometrial cancer risk stratified by various anthropometric measurements allows for an examination of the co-participation between these two factors in influencing risk of endometrial cancer. The lack of statistical significance of this interaction does not take away from the biological impact and rationale for public health intervention. For each evaluation, the risk of endometrial cancer attributed to the high risk genotype is further increased by high risk anthropometric measures. Therefore obese individuals are at an even higher risk of endometrial cancer with the high risk genotype, where non-obese individuals with the high risk genotype do not have a significant risk of endometrial cancer. Nevertheless, justification of a population health strategy targeting obesity does not require detection of a significant statistical interaction.

A common problem of studies of gene-environment interactions is lack of power to detect significant interactions. With the rapidly emerging studies of gene-environment interactions, many methodological and analytical strategies are still being developed[22]. It is argued that

using one statistical model for interaction cannot logistically be generalized to a variety of mechanisms of disease pathogenesis. Detecting significant interactions using this study design is made further challenging since polymorphisms that have large effects of risk factors are usually rare[106]. In the present study, one particular group of individuals, those with the low risk genotype and the high risk anthropometric measurement, was fairly small. This created difficulty in detecting significant interactions. With increasing sample size, it can be anticipated that this group would contain more individuals and, therefore, improve statistical power to detect an interaction.

Understanding how aromatase and obesity influence hormonal profiles can further explain the biological significance our results. Many significant associations have been reported between increased hormone levels and various measures of obesity[87, 89, 90, 92, 93], however, the exact mechanism for this association, and the effect of body fat distribution, remains unclear. Upper body obesity has been associated with a male phenotype and an androgenic hormone profile, where lower body obesity is more common in females and is associated with an estrogenic hormone profile[92]. Subsequently, upper body obesity has been shown to be associated with increased testosterone levels, while lower body obesity has been shown to be associated with increased aromatization of androgens to estrogens[82, 90, 92]. Since the hormonal profiles of individuals with upper and lower body obesity differ, the effect of fat distribution on

endometrial cancer risk and hormone levels has been investigated. General measures of obesity, such as current weight and BMI, are consistently associated with significantly elevated risk of endometrial cancer [84, 88, 94, 107], however waist circumference and waist-to-hip ratio have also been shown to be strong predictors of risk[84], even after adjusted for BMI[94]. Yet, this is not a consistent finding across all studies[87]. When considering hormone levels according to measures of obesity, BMI has a stronger relationship with estrogen levels, than waist-to-hip ratio[87, 90]. However, one study reported that waist and hip girth were positively correlated with estrogen, where as BMI was not[93]. These results are further complicated by the consideration of LH (leutinizing hormone), FSH (follicle stimulating hormone) and SHBG, all of which also influence levels of circulating estrogen levels and are correlated with BMI and waist-to-hip ratio[90, 93].

Our results indicate that all measures of obesity are fairly strong predictors of endometrial cancer risk. When considering the interaction of these measures with *CYP19*, this pattern remains. The largest risk exists for those who weighed ≥160 lbs, and had both alleles with ≥10 (TTTA) repeat. The effect is second largest among those with a waist circumference ≥88 cm, where the lowest risk was determined by waist-to-hip ratio. Therefore, it can be concluded that a synergistic effect exists between overall weight and upper body obesity with the (TTTA)n polymorphism to influence endometrial cancer risk (more than any other

measure in this population). Also, using a waist-to-hip ratio to measure fat distribution may not be a strong predictor of risk, compared to general measures of obesity.

Since lower body obesity may be associated with increased aromatization, we would expect that the effect of the *CYP19* polymorphism would be largest in those with a waist-to-hip ratio of <0.80, but this was not observed. However, when considering the high risk genotype in all low risk anthropometric measures, the highest risk exists with those with a waist circumference <88 cm. This may suggest that upper versus lower body obesity is important to consider when studying the interaction of aromatase and obesity.

In summary, our data suggests that the genetic factor exacerbates the effect of the environmental factor but does not have an effect on disease risk in the absence of the environmental factor. This pattern corresponds to a model for gene-environment interaction established by Yang and Khourey. This Type 2 interaction exists when the effect of the genotype is null (OR~1.0), while the risk according to the environmental factor is further elevated according to the genetic polymorphism[108].

5.4. Impact of Results

The purpose of studying gene-environment interaction is to enhance our understanding of disease development and pathophysiology as opposed to identifying a high risk susceptibility gene or deciphering inheritance patterns that may influence disease development[108]. When

common exposures are used, studies of gene-environment interactions have the potential to also have an impact from a public health perspective. This can be assessed by considering the frequency of these main factors among controls, along with the population attributable risk (PAR). The frequency of the anthropometric measures are much more common among the controls than the *CYP19* (TTTA)n polymorphism, as indicated in Table 11. PAR further estimates the proportion of cases that could be prevented if the risk factor is removed from the population. Relative risk estimates used to calculate each PAR can be found in Table 8 and 9. Almost 50% of controls had a high risk weight (≥160 lbs), where the PAR associated this measure is ~50%, indicating that almost half of cases of endometrial cancer could be prevented if this Albertan population were of a healthier weight. This data could further support the need for public health strategies to target obesity as a means of promoting health and preventing cancer.

Table 11. Population Level Implications. Distribution of high risk groups in each main factor among controls is presented along with population attributable risk.

Factor	Frequency in Controls (%)	Population Attributable Risk (PAR%)
***CYP19*: both alleles ≥10 (TTTA) repeats**	16.97	7.97
Current weight (≥160 lbs)	49.08	50.52
BMI (≥30 kg/m²)	35.79	37.45
Waist circumference (≥88 cm)*	49.45	42.73
Hip circumference (≥106 cm)*	49.45	37.97
Waist-to-hip ratio (≥0.80)*	59.41	24.62
Weight gain since adulthood (≥40 lbs)**	49.45	34.05

* missing data: two cases and one control ** missing data: one case

Since our results are limited to the specific (TTTA)n polymorphism, which is a non-modifiable risk factor for endometrial cancer, our *CYP19* results may not have any *direct* implications for population health. However, investigating the effect that genetic variants *CYP19* have on endometrial cancer risk can better our understanding of how aromatase influences this disease. Aromatase is a key target for treatment of other hormone related conditions such as breast cancer and endometriosis through the use of aromatase inhibitors[109]. However, the effect aromatase inhibitors have on endometrial cancer is complicated, as evidence exists that it can both treat and increase the risk of endometrial cancer[110,111]. In further understanding how aromatase activity influences endometrial cancer risk, this knowledge may possibly have implications in improving prevention and treatment options for those with different genetic variants of *CYP19.*

5.5. The Effect of Linkage Disequilibrium

Results of the present study may be due to a true association between the (TTTA)n polymorphism, anthropometric measures and endometrial cancer, or may be the result of an association of this factor with another factor that is the true biomarker of endometrial cancer risk. This possibility results from the event of recombination, or chromosomal exchange, which creates particular allelic combinations that are subsequently inherited together. Linkage disequilibrium occurs and frequencies of particular genes become non-randomly associated with

each other[112]. Therefore, the results found according to the *CYP19* (TTTA)n polymorphism may be an artifact of the co-inheritance of the *CYP19* loci with other functional polymorphisms.

Single nucleotide polymorphisms found in coding regions of *CYP19* have been shown to change amino acid composition and subsequent enzymatic activity[101]. The (TTTA)n polymorphism has been shown to be linked to these polymorphisms that create missence mutations in aromatase protein sequence[49, 113]. Paynter *et al.* developed haplotypes that describe which 7 and 11 (TTTA) repeat alleles co-segregate with single nucleotide polymorphisms that functionally change protein composition. The longer repeats have been identified in haplotypes shown to have significantly higher risk of endometrial cancer[49]. A breast cancer study investigated the association of the *CYP19* (TTTA)n polymorphism with a C-T substitution polymorphism exon 10. The longer *CYP19* repeats have been linked to the TT genotype which is in turn associated with breast cancer risk and more aggressive tumors. These polymorphisms have also indicated a change in aromatase expression levels in ovaries and adipose tissue[113]. In an osteoporosis study, the 7 (TTTA) repeat alleles have been shown to be linked to two other polymorphisms found in intronic regions and untranslated regions that are related to bone mineral density in postmenopausal women[46]. This evidence further supports the association between *CYP19* (TTTA)n repeat polymorphism and cancer risk, however, since risk estimates

found according to genotype are small and the functional significance of the polymorphism investigated in this study is unknown, the effect of linkage disequilibrium is important to consider.

5.6. Study Strengths

Investigations of endometrial cancer etiology can contribute to the understanding of this disease, as well as hormone related cancers and other diseases influenced by fluctuations in exogenous and endogenous hormones[23, 24, 114, 115]. This detailed analysis of endometrial cancer in an Albertan population is made possible by high quality resources provided by the parent study including blood samples, anthropometric measurements and extensive interview data. Examination of the population characteristics, main of effects of *CYP19* (TTTA)n polymorphism and obesity and the interaction of the two produced an assessment of how gene/environment co-participate to influence the risk of endometrial cancer. Other studies have investigated the effect of the (TTTA)n polymorphism on cancer and other diseases. The influence of obesity and body fat distribution on endometrial cancer risk has also been previously researched, however, there has been no known investigation of these two factors together. The advantage of a gene/environment interaction study is that it allows for an examination of disease pathogenesis according to factors common in the population. This requires identification of a potential candidate gene that is suspected to influence disease risk, which can be limiting due to the

complex nature of our genome. Now that the technology is available, other studies of genetic epidemiology have begun to utilize gene chip technology and microarrays to study multiple loci from more than one gene. These are very powerful tools that will exponentially increase our understanding of genes and polymorphisms involved in cancer development. However, utilizing multiple polymorphisms without understanding the biological significance of each may result in spurious associations without functional significance. Also, the methods to interpret results of microarray technology have not been fully developed. We had *a priori* evidence that the number of *CYP19* (TTTA)n repeats were associated with physiological changes in hormone levels and endometrial cancer risk[27, 43, 48, 49]. Thus, we were able to develop our study to address the independent and combined effects of the *CYP19* repeat polymorphism along with measures of obesity.

5.7. Study Limitations

Certain methodological issues can introduce biases that may or may not affect results. In the present study, selection bias in using RDD to identify controls may create a control population that is not necessarily representative of the target population. Since the response rate in control recruitment was 32%, there may be systematic differences between those who agreed to be in the study, compared to those who refused. For example, women with larger body sizes may have been more likely to refuse participation. If women with more healthy body

sizes were more likely to participate than those with larger body sizes, this biased control population may create exaggerated risk estimates, especially in the main effects of anthropometric measurements .

According to the natural history of endometrial cancer, control group participants may also introduce bias if they contained early manifestations of endometrial cancer. It is possible that unaffected individuals may develop endometrial cancer in the future. A control group containing any of these individuals would attenuate risk estimates, since these individuals are more likely to be similar to cases in main effect variables and potential confounding variables.

This study provides an understanding of how genetic variation at the *CYP19* (TTTA)n polymorphism differs between cases of endometrial cancer and controls. However, limitations to the study include epigenetic considerations that will affect gene expression and subsequent phenotypic outcome. More simply, this study does not address how the (TTTA)n polymorphism regulates *CYP19* gene expression and post-transcriptional consideration such as mRNA stability, splicing and aromatase glycosylation.

5.8. Future Research

Additional research into the expression profiling of *CYP19* and sequencing of the mRNA and protein products according to the (TTTA)n polymorphism will give insight into the impact this genetic factor has at the molecular level. Identification of the functional role of this

polymorphism would further confirm that risk attributed to this polymorphism is not an artifact of linkage disequilibrium with the true genetic factor. Also, an epidemiological study that additionally included hormonal profiles of study participants would be an appropriate follow-up this study. This could include assessment of androgen and estrogen concentrations as well as those factors that influence hormone levels such as LH and SHBG. These additional considerations would help link the polymorphism to an increase in serum estrogen levels and a subsequent increase in endometrial cancer risk. A large population based cohort study could be done to assess the exposure according to various measures of obesity and hormone levels with follow-up to identify cases of endometrial cancer. Examination of the (TTTA)n *CYP19* polymorphism, various anthropometric measurements and the hormonal profile of participants would allow for further understanding how obesity and *CYP19* polymorphisms impact endometrial cancer risk through their effect on peripheral estrogen biosynthesis.

References

1. Boffetta, P. Molecular epidemiology. J Intern Med 248, 447-54 (2000).
2. Chen, Y. C. & Hunter, D. J. Molecular epidemiology of cancer. CA Cancer J Clin 55, 45-54; quiz 57 (2005).
3. Walboomers, J. M. et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 189, 12-9 (1999).
4. Lowy, D. R. & Schiller, J. T. Prophylactic human papillomavirus vaccines. J Clin Invest 116, 1167-73 (2006).
5. CTV.ca News Staff. Cervical cancer vaccine wins approval in U.S. Available at http://www.ctv.ca/servlet/ArticleNews/story/CTVNews/20060608/cervicalcancer_vaccine_060608/20060608?hub=CTVNewsAt11. Accessed on June 13th 2006.
6. Bridges, A. Cervical cancer vaccine approved in the U.S., Canadian approval expected. Available at http://bodyandhealth.canada.com/channel_health_news_details.asp?news_id=10103&news_channel_id=145&channel_id=145&rot=11. Accessed on June 13, 2006.
7. Grann, V. R. et al. Hormone receptor status and survival in a population-based cohort of patients with breast carcinoma. Cancer 103, 2241-51 (2005).
8. Honrado, E., Benitez, J. & Palacios, J. The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications. Mod Pathol 18, 1305-20 (2005).
9. Ayers, M. et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 22, 2284-93 (2004).
10. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351, 2817-26 (2004).
11. Schmutz, J. et al. Quality assessment of the human genome sequence. Nature 429, 365-8 (2004).
12. Thomas, D. C. Genetic Epidemiology with a Capitol "E". Genetic Epidemiology 19, 289-3000 (2000).
13. Taioli, E. Biomarkers of genetic susceptibility to cancer: applications to epidemiological studies. Future Oncol 1, 51-56 (2005).
14. Bennett, I. C., Gattas, M. & Teh, B. T. The management of familial breast cancer. Breast 9, 247-63 (2000).

15. Newman, B. et al. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. Jama 279, 915-21 (1998).
16. Lipton, L. R. et al. Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic. J Clin Oncol 22, 4934-43 (2004).
17. Peltonen, L. & McKusick, V. A. Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291, 1224-9 (2001).
18. Brennan, P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? Carcinogenesis 23, 381-7 (2002).
19. Rothman, K. J., Greenland, S. & Walker, A. M. Concepts of interaction. Am J Epidemiol 112, 467-70 (1980).
20. Botto, L. D. & Khoury, M. J. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. Am J Epidemiol 153, 1016-20 (2001).
21. Ahmed, F. E. Gene-gene, gene-environment & multiple interactions in colorectal cancer. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 24, 1-101 (2006).
22. Mucci, L. A., Wedren, S., Tamimi, R. M., Trichopoulos, D. & Adami, H. O. The role of gene-environment interaction in the aetiology of human cancer: examples from cancers of the large bowel, lung and breast. J Intern Med 249, 477-93 (2001).
23. Gruber, C. J., Tschugguel, W., Schneeberger, C. & Huber, J. C. Production and actions of estrogens. N Engl J Med 346, 340-52 (2002).
24. Henderson, B. E. & Feigelson, H. S. Hormonal carcinogenesis. Carcinogenesis 21, 427-33 (2000).
25. Meinhardt, U. & Mullis, P. E. The aromatase cytochrome P-450 and its clinical impact. Horm Res 57, 145-52 (2002).
26. Simpson, E. R. et al. Tissue-specific promoters regulate aromatase cytochrome P450 expression. J Steroid Biochem Mol Biol 44, 321-30 (1993).
27. Tworoger, S. S. et al. Association of CYP17, CYP19, CYP1B1, and COMT polymorphisms with serum and urinary sex hormone concentrations in postmenopausal women. Cancer Epidemiol Biomarkers Prev 13, 94-101 (2004).
28. Mullis, P. E. et al. Aromatase deficiency in a female who is compound heterozygote for two new point mutations in the P450arom gene: impact of estrogens on hypergonadotropic hypogonadism, multicystic ovaries, and bone densitometry in childhood. J Clin Endocrinol Metab 82, 1739-45 (1997).

29. Ito, Y., Fisher, C. R., Conte, F. A., Grumbach, M. M. & Simpson, E. R. Molecular basis of aromatase deficiency in an adult female with sexual infantilism and polycystic ovaries. Proc Natl Acad Sci U S A 90, 11673-7 (1993).
30. Fukami, M. et al. Cytochrome P450 oxidoreductase deficiency in three patients initially regarded as having 21-hydroxylase deficiency and/or aromatase deficiency: diagnostic value of urine steroid hormone analysis. Pediatr Res 59, 276-80 (2006).
31. Sebastian, S. & Bulun, S. E. A highly complex organization of the regulatory region of the human CYP19 (aromatase) gene revealed by the Human Genome Project. J Clin Endocrinol Metab 86, 4600-2 (2001).
32. Mitrunen, K. & Hirvonen, A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. Mutat Res 544, 9-41 (2003).
33. Kadohama, N., Zhou, D., Chen, S. & Osawa, Y. Catalytic efficiency of expressed aromatase following site-directed mutagenesis. Biochim Biophys Acta 1163, 195-200 (1993).
34. Grumbach, M. M. & Auchus, R. J. Estrogen: consequences and implications of human mutations in synthesis and action. J Clin Endocrinol Metab 84, 4677-94 (1999).
35. Petry, C. J. et al. Associations between common variation in the aromatase gene promoter region and testosterone concentrations in two young female populations. J Steroid Biochem Mol Biol 98, 199-206 (2006).
36. Cook, L. S., Doherty, J. A., Weiss, N. S. & Chen, C. Endometrial Cancer: Epidemiology and Molecular Endocrinology. In: Henderson, B. E., Ponder, B. & Ross, R. K. (ed.) Endometrial Cancer: Epidemiology and Molecular Endocrinology. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 2003.
37. Tsuchiya, N. et al. Impact of IGF-I and CYP19 gene polymorphisms on the survival of patients with metastatic prostate cancer. J Clin Oncol 24, 1982-9 (2006).
38. Siegelmann-Danieli, N. & Buetow, K. H. Constitutional genetic variation at the human aromatase gene (Cyp19) and breast cancer risk. Br J Cancer 79, 456-63 (1999).
39. Ribeiro, F. S. et al. CYP19 (TTTA)(n) polymorphism and breast cancer risk in Brazilian women. Toxicol Lett 164, 90-5 (2006).
40. Okobia, M. N. et al. Simple tandem repeat (TTTA)n polymorphism in CYP19 (aromatase) gene and breast cancer risk in Nigerian women. J Carcinog 5, 12 (2006).
41. Miyoshi, Y., Iwao, K., Ikeda, N., Egawa, C. & Noguchi, S. Breast cancer risk associated with polymorphism in CYP19 in Japanese women. Int J Cancer 89, 325-8 (2000).

42. Li, L., Cicek, M. S., Casey, G. & Witte, J. S. No association between a tetranucleotide repeat polymorphism of CYP19 and prostate cancer. Cancer Epidemiol Biomarkers Prev 13, 2280-1 (2004).
43. Haiman, C. A. et al. A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk. Int J Cancer 87, 204-10 (2000).
44. Dialyna, I., Tzanakakis, G., Dolapsakis, G. & Tsatsakis, A. A tetranucleotide repeat polymorphism in the CYP19 gene and breast cancer susceptibility in a Greek population exposed and not exposed to pesticides. Toxicol Lett 151, 267-71 (2004).
45. Ahsan, H. et al. Variants in estrogen-biosynthesis genes CYP17 and CYP19 and breast cancer risk: a family-based genetic association study. Breast Cancer Res 7, R71-81 (2005).
46. Riancho, J. A. et al. Aromatase gene and osteoporosis: relationship of ten polymorphic loci with bone mineral density. Bone 36, 917-25 (2005).
47. Kristensen, V. N. et al. A rare CYP19 (aromatase) variant may increase the risk of breast cancer. Pharmacogenetics 8, 43-8 (1998).
48. Berstein, L. M. et al. CYP19 gene polymorphism in endometrial cancer patients. J Cancer Res Clin Oncol 127, 135-8 (2001).
49. Paynter, R. A. et al. CYP19 (aromatase) haplotypes and endometrial cancer risk. Int J Cancer 116, 267-74 (2005).
50. Mitas, M. Trinucleotide repeats associated with human disease. Nucleic Acids Res 25, 2245-54 (1997).
51. Ranum, L. P. & Day, J. W. Pathogenic RNA repeats: an expanding role in genetic disease. Trends Genet 20, 506-12 (2004).
52. Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet 6, 743-55 (2005).
53. Lian, Y. & Garner, H. R. Evidence for the regulation of alternative splicing via complementary DNA sequence repeats. Bioinformatics 21, 1358-64 (2005).
54. Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease (eds. Khoury, M. J., Little, J. & Burke, W.) (Oxford University Press New York, 2004).
55. Li, Y. C., Korol, A. B., Fahima, T. & Nevo, E. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21, 991-1007 (2004).
56. Pearson, C. E., Nichol Edamura, K. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6, 729-42 (2005).
57. Petruska, J., Arnheim, N. & Goodman, M. F. Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet

repeats associated with neurological diseases. Nucleic Acids Res 24, 1992-8 (1996).
58. De Temmerman, N. et al. Intergenerational instability of the expanded CTG repeat in the DMPK gene: studies in human gametes and preimplantation embryos. Am J Hum Genet 75, 325-9 (2004).
59. DiSaia, P. J. & Creasman, W. T. Clinical Gynecological Oncology (The C.V. Mosby Company St. Louis, Missouri, 1989).
60. Duarte-Franco, E. & Franco, E. L. Women's Health Surveillance Report. Public Health Agency of Canada. Canadian Institute for Health Information (2003).
61. King, H. S., Wigle, D. T., Hill, G. B. & Silins, J. Uterine cancers in Alberta: trends of incidence and mortality. Can Med Assoc J 127, 591-4 (1982).
62. Wigle, D. T., Grace, M. & Smith, E. S. Estrogen use and cancer of the uterine corpus in Alberta. Can Med Assoc J 118, 1276-8 (1978).
63. Cree, M., Kamruzzaman, A. & Karosas, A. Cancer in Alberta: A Regional Picture. Division of Population Health and Information. Alberta Cancer Board (2006).
64. Acharya, S., Hensley, M. L., Montag, A. C. & Fleming, G. F. Rare uterine cancers. Lancet Oncol 6, 961-71 (2005).
65. Lax, S. F., Pizer, E. S., Ronnett, B. M. & Kurman, R. J. Clear cell carcinoma of the endometrium is characterized by a distinctive profile of p53, Ki-67, estrogen, and progesterone receptor expression. Hum Pathol 29, 551-8 (1998).
66. Sherman, M. E. et al. Risk factors and hormone levels in patients with serous and endometrioid uterine carcinomas. Mod Pathol 10, 963-8 (1997).
67. Ryan, A. J., Susil, B., Jobling, T. W. & Oehler, M. K. Endometrial cancer. Cell Tissue Res 322, 53-61 (2005).
68. Amant, F. et al. Endometrial cancer. Lancet 366, 491-505 (2005).
69. Key, T. & Pike, M. The dose-effect relationship between "unopposed" oestrogens and endometrial mitotic rate: its central role in explaining and predicting endometrial cancer risk. Br J Cancer 57, 205-12 (1998).
70. Potischman, N. & al, e. Case-control study of endogenous steroid hormones and endometrial cancer. J natl Cancer Inst 88, 1127-35 (1996).
71. Xu, W. H. et al. Menstrual and reproductive factors and endometrial cancer risk: Results from a population-based case-control study in urban Shanghai. Int J Cancer 108, 613-9 (2004).
72. Nyholm, H. C. et al. Plasma oestrogens in postmenopausal women with endometrial cancer. Br J Obstet Gynaecol 100, 1115-9 (1993).
73. Lukanova, A. et al. Circulating levels of sex steroid hormones and risk of endometrial cancer in postmenopausal women. Int J Cancer 108, 425-32 (2004).

74. Purdie, D. M. & Green, A. C. Epidemiology of endometrial cancer. Best Pract Res Clin Obstet Gynaecol 15, 341-54 (2001).
75. Viswanathan, A. N. et al. Smoking and the risk of endometrial cancer: results from the Nurses' Health Study. Int J Cancer 114, 996-1001 (2005).
76. Newcomer, L. M., Newcomb, P. A., Trentham-Dietz, A. & Storer, B. E. Hormonal risk factors for endometrial cancer: modification by cigarette smoking (United States). Cancer Causes Control 12, 829-35 (2001).
77. Ewertz, M., Schou, G. & Boice, J. D., Jr. The joint effect of risk factors on endometrial cancer. Eur J Cancer Clin Oncol 24, 189-94 (1988).
78. Lu, H. K. & Broaddus, R. R. Gynecologic Cancers in Lynch Syndrome/HNPCC. Fam Cancer 4, 249-54 (2005).
79. de la Chapelle, A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 4, 769-80 (2004).
80. Kaaks, R., Lukanova, A. & Krurzer, M. Obesity, Endogenous Hormones, and Endometrial Cancer Risk, A Synthetic Review. Cancer Epidemiol Biomarkers Prev 11, 15531-1543 (2002).
81. MacDonald, P., Edmona, C., Hemsell, D., Porter, J. & Siiteri, P. Effect of obesity on conversion of plasma androstenedione to estrone in postmenopausal women with and without endometrial cancer. Am J Obstet Gynecol 130, 448-55 (1978).
82. Kirschner, M. A. et al. Androgen-estrogen metabolism in women with upper body versus lower body obesity. J Clin Endocrinol Metab 70, 473-9 (1990).
83. Bulun, S. E. & Simpson, E. R. Competitive reverse transcription-polymerase chain reaction analysis indicates that levels of aromatase cytochrome P450 transcripts in adipose tissue of buttocks, thighs, and abdomen of women increase with advancing age. J Clin Endocrinol Metab 78, 428-32 (1994).
84. Swanson, C. et al. Relation of Endometrial Cancer Risk to Past and Contemporary Body Size and Body Fat Distribution. Cancer Epidemiology, Biomarker & Prevention 2, 321-327 (1993).
85. Elliot, E., Matanoski, G., Rosenshein, N., Grumbine, F. & Diamond, E. Body Fat Patterning in Women with Endometrial Cancer. Gynecologic Oncology 39, 253-258 (1990).
86. Schapira, D. et al. Upper-Body Fat Distribution and Endometrial Cancer Risk. JAMA 266, 1808-1811 (1991).
87. Austin, H., Austin, J., Partridge, E., Hatch, K. & Shingleton, H. Endometrial Cancer, Obesity, and Body Fat Distribution. Cancer Research 51, 568-572 (1991).
88. Trentham-Dietz, A., Nichols, H. B., Hampton, J. M. & Newcomb, P. A. Weight change and risk of endometrial cancer. Int J Epidemiol 35, 151-8 (2006).

89. Raja, C., Hansen, R., Baber, R. & Allen, B. Hip girth as a predictor of abdominal adiposity in postmenopausal women. Nutrition 20, 772-7 (2004).
90. Kaye, S. A., Folsom, A. R., Soler, J. T., Prineas, R. J. & Potter, J. D. Associations of body mass and fat distribution with sex hormone concentrations in postmenopausal women. Int J Epidemiol 20, 151-6 (1991).
91. Judd, H. L. et al. Serum androgens and estrogens in postmenopausal women with and without endometrial cancer. Am J Obstet Gynecol 136, 859-71 (1980).
92. Kirschner, M. A. & Samojlik, E. Sex hormone metabolism in upper and lower body obesity. Int J Obes 15 Suppl 2, 101-8 (1991).
93. Kirchengast, S. Interaction between sex hormone levels and body dimensions in postmenopausal women. Hum Biol 66, 481-94 (1994).
94. Elliott, E. A., Matanoski, G. M., Rosenshein, N. B., Grumbine, F. C. & Diamond, E. L. Body fat patterning in women with endometrial cancer. Gynecol Oncol 39, 253-8 (1990).
95. Willis, G. Cognitive Interviewing and Questionnaire Design: A Training Manual. (National Center for Health Statistics, 1994).
96. Blaise for Windows Version 4.2. Department of Statistical Informatics: Statistics Netherlands. Voorburg/ Heerlen.1999.
97. Applied Biosystems. AFLP® System Analysis Getting Started Guide. 2005
98. National Center for Biotechnology Information. Basic Local Alignment Search Tool Available at http://130.14.29.110/BLAST/. Accessed on Nov. 14, 2005.
99. Kelsey, J. L. et al. A case-control study of cancer of the endometrium. Am J Epidemiol 116, 333-42 (1982).
100. Janssen, I., Katzmarzyk, P. T. & Ross, R. Body mass index, waist circumference, and health risk: evidence in support of current National Institutes of Health guidelines. Arch Intern Med 162, 2074-9 (2002).
101. Ma, C. X. et al. Human aromatase: gene resequencing and functional genomics. Cancer Res 65, 11071-82 (2005).
102. Healey, C. S. et al. Polymorphisms in the human aromatase cytochrome P450 gene (CYP19) and breast cancer risk. Carcinogenesis 21, 189-93 (2000).
103. Han, D. F. et al. Polymorphisms of estrogen-metabolizing genes and breast cancer risk: a multigenic study. Chin Med J (Engl) 118, 1507-16 (2005).
104. Dick, I. M., Devine, A. & Prince, R. L. Association of an aromatase TTTA repeat polymorphism with circulating estrogen, bone structure, and biochemistry in older women. Am J Physiol Endocrinol Metab 288, E989-95 (2005).

105. Almeida, S. et al. Estrogen-metabolizing gene polymorphisms and lipid levels in women with different hormonal status. Pharmacogenomics J 5, 346-51 (2005).
106. Clayton, D. & McKeigue, P. M. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 358, 1356-60 (2001).
107. Bershtein, L. M. et al. [A polymorphism study of the CYP19 gene in endometrial cancer patients]. Vopr Onkol 46, 302-5 (2000).
108. Yang, Q. & Khoury, M. J. Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. Epidemiol Rev 19, 33-43 (1997).
109. Bulun, S. E. et al. Regulation of aromatase expression in estrogen-responsive breast and uterine disease: from bench to treatment. Pharmacol Rev 57, 359-83 (2005).
110. Lai, C. H. & Huang, H. J. The role of hormones for the treatment of endometrial hyperplasia and endometrial cancer. Curr Opin Obstet Gynecol 18, 29-34 (2006).
111. Ursic Vrscaj, M. et al. Endometrial and other primary cancers after tamoxifen treatment of breast cancer -- results of retrospective cohort study. Eur J Obstet Gynecol Reprod Biol 95, 105-10 (2001).
112. Hartl, D. L. & Clark, A. C. Principles of Population Genetics (Sinauer Associates, Inc. Sunderland, MA, 1997).
113. Kristensen, V. N. et al. Genetic variants of CYP19 (aromatase) and breast cancer risk. Oncogene 19, 1329-33 (2000).
114. Simpson, E. R. et al. Aromatase expression in health and disease. Recent Prog Horm Res 52, 185-213; discussion 213-4 (1997).
115. Simpson, E. R., Clyne, C., Speed, C., Rubin, G. & Bulun, S. Tissue-specific estrogen biosynthesis and metabolism. Ann N Y Acad Sci 949, 58-67 (2001).

Appendix

Table 5a. Distribution of (TTTA) CYP19 repeat alleles among case and controls.

Repeat number	Frequency in Cases (%)	Frequency in Controls (%)
7	109 (43)	262 (48)
8	32 (13)	49 (9)
9	0	1 (<1)
10	1 (<1)	14 (3)
11	104 (41)	199 (37)
12	7 (3)	14 (3)
13	1 (<1)	3 (<1)

Table 8a. Additional Analysis of the Main Effect of Genotype. Risk of endometrial cancer according to the CYP19 allele at the (TTTA)n polymorphism is assessed from each repeating unit.

Genotype	Case N=127	Control N=271	OR (95%CI)[a]	OR (95%CI)[b]	OR (95%CI)[c]
≥8/≥8 vs. ≥8/<8 or <8/<8	45 82	68 203	1.65 (1.03-2.63) 1	1.35 (0.82-2.21) 1	1.46 (0.88-2.45)
≥8/≥8 or ≥8/<8 vs. <8/<8	100 27	212 59	1.06 (0.63-1.79) 1	0.97 (0.56-1.67) 1	1.06 (0.60-1.88) 1
≥9/≥9 vs. ≥9/<9 or <9/<9	30 97	46 225	1.41 (0.83-2.40) 1	1.05 (0.60-1.87) 1	1.04 (0.58-1.90)
≥9/≥9 or ≥9/<9 vs. <9/<9	83 44	185 86	0.87 (0.55-1.37) 1	0.78 (0.48-1.26) 1	0.82 (0.50-1.34) 1
≥10/≥10 vs. ≥10/<10 or <10/<10	30 97	46 225	1.41 (0.83-2.40) 1	1.05 (0.60-1.88) 1	1.05 (0.58-1.90) 1
≥10/≥10 or ≥10/<10 vs. <10/<10	83 44	184 87	0.88 (0.58-1.39) 1	0.80 (0.50-1.28) 1	0.82 (0.50-1.34) 1
≥11/≥11 vs. ≥11/<11 or <11/<11	30 97	40 231	1.70 (0.99-2.92) 1	1.22 (0.68-2.21)	1.26 (0.68-2.32)
≥11/≥11 or ≥11/<11 vs. <11/<11	82 45	176 95	0.98 (0.62-1.53) 1	0.89 (0.55-1.43) 1	0.91 (0.56-1.49) 1
≥12/≥12 vs. ≥12/<12 or <12/<12	0 127	0 271	-	-	
≥12/≥12 or ≥12/<12 vs. ≥12/≥12	8 119	17 254	1.07 (0.44-2.61) 1	1.22 (0.48-3.14) 1	1.38 (0.52-3.61) 1
13/13 vs. 13/<13 or <13/<13	0 127	0 271	-	-	
13/<13 or <13/<13 vs. 13/13	1 126	3 268	0.79 (0.08-8.0) 1	1.44 (0.14-15.11) 1	2.82 (0.26-30.96) 1

[a] adjusted for age and residence
[b] adjusted for age, residence and weight
[c] adjusted for age, residence, weight, number of births, OC use, age at menarche, age at menopause and smoking status

Table 10a. Gene-environment interaction. The influence of the CYP19 (TTTA) polymorphism on endometrial cancer risk is modified by anthropometric measurements. Genotype further stratified to distinguish between heterozygous individuals (having one ≥10 (TTTA) repeat allele and one <10 (TTTA) repeat allele) and homozygous individuals (both alleles with <10 (TTTA) repeats).

	G	E	Case N=127	Control N=271	OR(95%CI)[a]	OR(95%CI)[b]
Current weight (lbs)	≥10/≥10	≥160	25	25	2.56(1.12-5.88)	2.71(1.13-6.52)
	≥10/≥10	<160	5	21	0.80(0.27-2.40)	0.71(0.23-2.17)
	≥10/<10	≥160	33	63	1.90(0.92-3.90)	2.00(0.93-4.28)
	≥10/<10	<160	20	75	0.43(0.18-1.04)	0.45(0.18-1.10)
	<10/<10	≥160	26	46	1.83(0.86-3.93)	1.83(0.82-4.09)
	<10<10	<160	17	41	1	1
BMI (kg/m^2)	≥10/≥10	≥30	23	24	2.32(1.05-5.12)	2.39(1.04-5.46)
	≥10/≥10	<30	6	21	0.79(0.31-2.05)	0.75(0.28-1.97)
	≥10/<10	≥30	41	65	1.88(0.94-3.74)	1.97(0.96-4.07)
	≥10/<10	<30	12	73	0.55(0.27-1.10)	0.58(0.28-1.19)
	<10/<10	≥30	28	49	1.65(0.78-3.48)	1.71(0.78-3.72)
	<10<10	<30	15	38	1	1
Waist circumf-erence (cm)*	≥10/≥10	≥88	22	25	2.41(1.05-5.58)	2.58(1.07-6.24)
	≥10/≥10	<88	7	20	0.85(0.30-2.41)	0.82(0.28-2.38)
	≥10/<10	≥88	38	62	1.78(0.87-3.66)	1.94(0.92-4.12)
	≥10/<10	<88	15	76	0.49(0.21-1.12)	0.52(0.22-1.24)
	<10/<10	≥88	25	42	1.73(0.80-3.72)	1.85(0.83-4.13)
	<10<10	<88	18	45	1	1
Hip circumf-erence (cm)*	≥10/≥10	≥106	24	25	1.97(0.88-4.43)	1.99(0.85-4.63)
	≥10/≥10	<106	6	21	0.83(0.30-2.33)	0.82(0.28-2.37)
	≥10/<10	≥106	42	64	1.48(.74-2.95)	1.59(0.77-3.29)
	≥10/<10	<106	11	74	0.47(0.21-1.06)	0.51(0.22-1.17)
	<10/<10	≥106	29	44	1.40(0.66-2.96)	1.48(0.68-3.23)
	<10<10	<106	15	43	1	1
Waist-to-hip ratio *	≥10/≥10	≥0.80	21	22	2.06(0.87-4.91)	2.20(0.89-5.45)
	≥10/≥10	<0.80	8	25	0.97(0.34-2.79)	0.89(0.30-2.64)
	≥10/<10	≥0.80	32	43	1.30(0.61-2.76)	1.42(0.65-3.11)
	≥10/<10	<0.80	21	95	0.70(0.30-1.64)	0.73(0.30-1.76)
	<10/<10	≥0.80	22	33	1.54(0.8-3.5)	1.62(0.71-3.67)
	<10<10	<0.80	22	54	1	1
Weight gain since adulthood (lbs)**	≥10/≥10	≥40	21	27	2.17(0.97-4.87)	2.55(1.09-5.98)
	≥10/≥10	<40	8	18	0.56(0.18-1.77)	0.57(0.18-1.82)
	≥10/<10	≥40	37	81	1.23(0.60-2.54)	1.58(0.74-3.40)
	≥10/<10	<40	16	57	0.65(0.30-1.40)	0.71(0.32-1.57)
	<10/<10	≥40	29	53	1.29(0.61-2.75)	1.62(0.73-3.61)
	<10<10	<40	14	34	1	1

G: genetic factor E: environmental factor

[a] adjusted for age and residence

[b] adjusted for age, residence, number of births, OC use, age at menarche, age at menopause and smoking status

* missing data: two cases and one control ** missing data: one case