Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Kira C. Taylor (date)

Genetic and Environmental Contributions to Reproductive Characteristics

By

Kira C. Taylor
Doctor of Philosophy

Epidemiology

_____
Michele Marcus, Ph.D.
Advisor

_____
Celia E. Dominguez
Committee Member

_____
Michael P. Epstein
Committee Member

_____
W. Dana Flanders
Committee Member

_____
Stephanie L. Sherman
Committee Member

_____
Chanley M. Small
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Genetic and Environmental Contributions to Reproductive Characteristics

By

Kira C. Taylor
B.A., Rice University, 1997
M.S., Stanford University, 1999
M.A.Ed., Wake Forest University, 2000

Advisor: Michele Marcus, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2009

Abstract

Genetic and Environmental Contributions to Reproductive Characteristics
By Kira C. Taylor

Reproductive traits such as age at menarche, menstrual function, and perhaps fertility are heritable, but few specific genes have been identified that influence these traits. This dissertation consists of five studies that evaluate the effects of candidate genetic and environmental influences and their interactions on reproductive traits in women. The first study examined the relative contributions of genes and environment on age at menarche and menstrual cycle length. These traits were heritable, and heritability varied across levels of an environmental exposure (polybrominated biphenyls, PBBs). PBB levels were also inversely associated with estradiol and follicle stimulating hormone levels in a subset of this population. Three candidate gene studies were conducted in a prospective cohort study of women office workers. Variation in the progesterone receptor gene was associated with age at menarche and menstrual cycle length. Inhibin gene polymorphisms were not associated with menstrual function. Heavy alcohol use (>1 drink/day) and smoking were significant predictors of longer time to pregnancy in this population, but only among carriers of the "slow" acetylator haplotype of the enzyme N-acetyltransferase-2. This dissertation emphasizes the need to incorporate genetic information about metabolic enzymes when studying the effects of xenobiotics on human health. The results support the utility of candidate gene studies when there is prior evidence. This dissertation also adds to a growing body of literature characterizing the endocrine-disrupting properties of brominated flame retardants.

Genetic and Environmental Contributions to Reproductive Characteristics

By

Kira C. Taylor
B.A., Rice University, 1997
M.S., Stanford University, 1999
M.A.Ed., Wake Forest University, 2000

Advisor: Michele Marcus, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2009

Acknowledgments

There are many people who have helped me, in small and large ways, both indirectly and directly, to pursue this degree and complete this dissertation. First, I would like to thank Michele Marcus, who has stood behind me 100% from the first time I talked to her over five years ago. Her passion for science, combined with her unwavering support and unselfish guidance, made her a truly remarkable advisor. Chanley Small gave up countless hours of her time to help me through classes, grant proposals, examinations, research, and writing. Without her daily encouragement, this dissertation would not have been possible. Stephanie Sherman, Mike Epstein, Chely Dominguez and Dana Flanders provided invaluable expertise and thoughtful evaluations of my work at every step. I was truly fortunate to have such a talented and supportive committee.

I would also like to thank my lifelong friends; my newfound friends at Emory; those I met through UUCA; and those who joined me in my other passion, singing alongside me in the Atlanta Sacred Chorale. They have constantly reminded me that there is life beyond exams, and have helped me keep my sense of perspective and my sense of humor. I would finally like to thank both of my truly amazing parents, who taught me from a very early age that I could do whatever I dreamed, and have been a constant source of inspiration, support, and love.

TABLE OF CONTENTS

TABLES

FIGURES

Chapter 1

Introduction and Background

1.1 Introduction

This dissertation examines several genetic and environmental influences on reproductive outcomes. Reproductive outcomes examined in this dissertation are age at menarche, menstrual function, and fecundability. This chapter presents background information for each of these outcomes, including the basic biology involved as well as epidemiologic risk factors and measurement issues.

1.2 Age at Menarche

1.2.1 Biology and Epidemiology

Puberty is a complex process whose initiation and regulation is controlled by both the nervous and endocrine systems and is influenced by genetic, endogenous, and exogenous factors (**1**). The initiation of puberty is controlled by the hypothalamic-pituitary-gonadal (HPG) axis. The HPG axis is active during the fetal period, and becomes inactive during childhood during a period called the juvenile pause. The HPG axis becomes active again at the initiation of puberty, when the hypothalamus begins releasing regular pulses of gonadotropin-releasing hormone (GnRH). The role of GnRH is to stimulate the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the anterior pituitary gland. These hormones, in turn, regulate the maturation of the ovary and its production of sex steroid hormones such as estrogen. The gonadal steroid hormones then provide feedback to regulate the pulses of GnRH as well as the

secretion of FSH and LH. Once estrogen and progesterone pulses become regular and high enough, menstrual bleeding is caused by estrogen and progesterone withdrawal (**2**). Although there is a significant correlation between initiation of puberty and age at menarche, the correlation coefficient is only 0.37-0.38, suggesting that there are some factors that affect puberty and menarche independently (**3**).

A recent analysis of NHANES data in the U.S. revealed a secular trend in decreasing age at menarche within race/ethnicity subgroups between women born from the 1920's to the mid-1980's, although they could not adjust for possible changes in body mass index (BMI) or adiposity (**4**). Anderson and Must (**5**) claim that the unadjusted secular decline in observed in the U.S. is probably not due to environmental etiology, but instead may be due to changing distributions of risk factors**.** Changing BMI and prevalence of obesity, increasing presence of harmful environmental contaminants, changing social environments (such as being raised in a household without a father), and genetic polymorphisms have been hypothesized as the culprits (**6**). It should be noted that within race/ethnicity subgroups, age at menarche only slightly decreased when comparing data from NHANES III to NHANES 99-02, approximately 10 years later (**5**).

Age at menarche has broad social and biological implications. Earlier age at menarche is associated with earlier sexual activity and delinquent behavior (**7, 8**), and alcohol and substance abuse (**9, 10**).

Earlier age at menarche is also associated with earlier exposure to higher levels of estrogens, and may indicate a higher basal estrogen level throughout life (**11**). Studies suggest that cumulative estrogen exposure increases the risk of breast cancer (**12, 13**).

Correspondingly, later age at menarche is significantly associated with reduced risk of breast cancer (for a review, see Kelsey, Gammon et al. (**14**)), with a delay of 2 years resulting in approximately a 10% decrease in risk (**15**). Early age at menarche may also be associated with endometrial cancer (**16, 17**). Earlier age at menarche is associated with higher parity; therefore, the increased parity may be the cause of the increased risk of endometrial cancer.

Age at menarche has also been associated with syndromes such as hyperthyroidism, encephalitis, uremia, congenital heart disease, cystic fibrosis, and diabetes mellitus (**18, 19**). Later menarche is associated with increased risk of diseases such as Alzheimer's disease (**20**) and osteoporosis (**21, 22**). Later menarche is also associated with decreased risk of coronary heart disease and ischemic heart disease (**20, 23**).

Age at menarche has been associated with time to pregnancy (**24**). Notably, later age at menarche was associated with higher parity after adjusting for total number of fertile years (**25**), although this was not found in two other studies (**26, 27**). Early and late menarche were associated with risk of ectopic pregnancy. Increasing age at menarche has been found to be associated with risk of spontaneous abortions, but this trend was not found when first pregnancies alone were considered (**28, 29**).

The causal pathways in all of these associations are unclear. Age at menarche could be associated with these reproductive cancers because of greater lifetime estrogen exposure or associations with behaviors that are also associated with higher cancer risk. In these causal pathways, age at menarche is considered to be a causal factor in cancer

risk. In cases of childhood diseases, the situation may be reversed; the disease may cause an advanced or delayed age at menarche. Another explanation is that the associations between age at menarche and disease risk are confounded by other factors that influence both. For example, perhaps a specific genetic variant leads to increasing age at menarche and independently leads to reduced breast cancer risk. In this case, age at menarche would not be the cause itself; rather it is just a marker for the true underlying cause.

The primary endogenous influence on age at menarche is body composition, which itself is influenced by multiple factors. Menarche has been found to be associated with nutritional intake; these associations act primarily through accumulation of a critical percentage of body fat (**30, 31**). A recent study found an association between central adiposity and estradiol levels at age 8 and earlier age at menarche (**32**). Body fat may influence age at menarche both directly and indirectly. For example, adipose tissue can convert androgens to estrogens though aromatization. Indirectly, body fat increases circulating leptin levels, and leptin can then act as a messenger to communicate with the brain, acting through receptors in the hypothalamus and indicating the body's readiness for menarche; correspondingly, higher leptin levels are associated with lower age at menarche (**33**). Skeletal maturity and height are also associated with age at menarche (**34**); however, a recent study using NHANES data found that body fat distribution was a better predictor of earlier menarche, with gluteofemoral fat being the most important factor for menarche (**35**). Exercise and athletic training could affect menarche through body fat levels (**36**), or directly by altering hormonal secretion and metabolism (**37**).

Environmental estrogens such as polybrominated biphenyls (**38**), polychlorinated hydrocarbons, (**39**), polyhalogenated aromatic hydrocarbons (**40**) or lead and other contaminants (**41**) have been associated with pubertal development and may influence age at menarche. However, these chemicals are generally mixtures of estrogenic and anti-estrogenic congeners, and may influence age at menarche through *in utero* or breastfeeding exposure (e.g., **38**). The window of susceptibility may also vary by compound.

Social factors are also known to influence age at menarche. Race and ethnicity may be proxies for both socioeconomic and genetic factors. Although it is a common observation in the U.S. that black girls presently achieve menarche earlier than white girls, whereas Asians achieve it later than whites, there may not be a difference in age at menarche among different races and ethnicities living in similar socioeconomic circumstances (**42**). The difference in age at menarche between U.S. blacks and whites seems to be a recent phenomenon. Examining age at menarche by decade and race, blacks have only had a significantly earlier age at menarche since the 1970's; in fact, from the 1920's to the 1940's, they had an older age at menarche than whites (**4**). Having a stepfather in the home advances menarche; the number of younger brothers also influences age at menarche (**43**).

1.2.2 Genetic Influences on Age at Menarche

Studies investigating familial correlations in age at menarche date back to the 1920's when Popenoe (**44**) calculated the correlation between sisters' age at menarche. Heritability estimates using same-generation siblings or twins range from 0.45 to 0.95,

but for many of these studies, the heritability is likely biased upward because shared environmental influences were not accounted for in the study design or analysis (**45**).

Treloar and Martin (**46**) found that the Pearson correlation between DZ twins was 0.19±0.04, and between MZ twins was 0.64±0.03; if the genetic component was only additive, and shared environmental effects are ignored, then the correlation between MZ twins should be exactly twice that of DZ twins. When the authors dropped the dominance variance component from the model, there was a significantly worse fit. They thus inferred that the genetic contribution to age at menarche is largely composed of dominance or epistasis effects; because they included no parent-offspring pairs in their study, they cannot distinguish between dominance effects between alleles or interaction between loci (epistasis).

Towne et al. (**45**) estimated the heritability of age at menarche in a population drawn from the Fels Longitudinal Study (**47**), from a suburban area near Dayton, Ohio. Using 371 women from 112 nuclear and extended families, they estimated the heritability of age at menarche to be 0.49±.13 (95% CI, 0.24 to 0.73), indicating that approximately half of the variation in age at menarche in this population could be attributed to additive genetic factors. They tested for dominant genetic and household environmental factors by comparing sister-sister correlations with mother-daughter correlations and found that the correlations were not significantly different from each other, indicating a lack of dominance or household effects. This study is unique in that it collected all age at menarche data prospectively during biyearly interviews and included a wide variety of

familial relationships, and therefore could distinguish between shared genetic and environmental effects.

Polymorphisms in the estrogen receptor alpha gene are inconsistently associated with age at menarche (**48-50**). Polymorphisms in the steroid hormone binding globulin gene (**51**), and cytochrome p450 (**52-54**) have also been associated with age at menarche. While the same is true for the androgen receptor, this is hypothesized to be through the father's behavior (**55**). A polymorphism in the vitamin D receptor is also associated with age at menarche (**56**). One SNP and a corresponding haplotype in IGF-1 were also associated with age at menarche in 354 Caucasian families, using an empirical permutation-based significance level (**57**). A study of estrogen metabolism pathway genes in a small (N=152) population of white women found an association of age at menarche with CYP1B1*2 and CYP19 7r, though associations were modest and no longer reached significance after adjustment for multiple comparisons. No associations were found with other genes in the estrogen metabolism pathways, including the estrogen receptor (**58**).

Guo et al. (**59**) conducted a genome-wide scan for quantitative trait loci associated with age at menarche among 2461 females from 402 pedigrees. Using the software SOLAR, they conducted a variance component linkage analysis. The heritability of age at menarche in this population was 0.59±0.05. They identified three genomic regions (22q13, 22q11, and 11q23) with significant LOD scores that explained 20%, 16%, and 16% of the total variance, respectively. Candidate genes in the regions Guo et al. identified include sterol regulatory element binding transcription factor-2, which

stimulates the transcription of sterol-regulated genes and controls cholesterol synthesis; the progesterone receptor; and catechol-O-methyltransferase, which is involved in estrogen catabolism and elimination. They also detected one significant epistatic interaction between two loci on chromosomes 22 and 3 (p=.005). They did not examine environmental effects or gene-environment interaction. Subsequently, another genome-wide linkage scan for age at menarche in 92 pairs of twin girls whose adolescent BMI was known also found LOD scores >1 at regions 22q11 and 22q13 (**60**). The linkage to the region on chromosome 11 was not replicated in this population.

1.2.3 The Progesterone Receptor: A Candidate Gene for Age at Menarche

The gene encoding the progesterone receptor (*PGR*) is located on chromosome 11q22-23 and is expressed mainly in the female genital tract, breast and brain (**61**). It spans 92kb and contains 8 exons that consist of a DNA binding domain, a ligand (hormone)-binding domain, and an activation domain (for review, see (**62**). There are 16 nonsynonymous SNPs identified in dbSNP. Of these, only one has a reported minor allele frequency >5% in both dbSNP and HapMap. There are three isoforms, PR-A, PR-B, and PR-C, which result from different translational start sites. Its concentration is down-regulated by progestins and up-regulated by estrogens. Decreasing progesterone levels (progesterone withdrawal) is necessary for the onset of bleeding (shedding) during the menstrual cycle (**63**). The increase in the ratio of PR-A relative to PR-B may be the primary cause of functional progesterone withdrawal (**64**). Chronic low doses of the antiprogestogen mifepristone (RU486) inhibit ovulation and cause amenorrhea, probably by impairing the gene regulatory activity of the progesterone receptor (**65, 66**). Thus, in

addition to its role in the HPG axis during puberty, the progesterone receptor also plays a critical role in the menstrual cycle, and could thereby influence age at menarche.

Polymorphisms in *PGR* have been examined with respect to reproductive cancers because estrogen is a potent mitogen and, when unopposed by progesterone, may predispose women to cancer. One polymorphism in the promoter region of *PGR* has been identified as a risk factor for endometrial cancer (**67**). Another variant coined PROGINS represents a combination of three linked polymorphisms: a 320 bp Alu insertion in an intron, and two single nucleotide polymorphisms that result in Val660→Leu amino acid change in exon 4 and His770→His (a synonymous mutation) in exon 5 (**68, 69**) (Figure 1.1). The three mutations are in 100% linkage disequilibrium according to De Vivo et al. (**67**), Pijnenborg et al. (**70**), Schweikert et al. (**71**), and Wang-Gohrke et al. (**72**). These polymorphisms are all present in the hormone-binding domain, which comprises exons 4 through 8 (**61**). All three isoforms of the progesterone receptor contain exons 4 through 8.

The Val660Leu polymorphism is in the "hinge" region of the receptor and is involved in receptor dimerization and ligand binding (**73**), transcriptional activation (**74**); nuclear localization; and interactions with co-repressors (**75**). A recent *in vitro* study shows that this variant may have decreased response to progestin, and is not as efficient at opposing estrogen's proliferative effects, due to decreased transcription and protein activity (**76**). It is not clear whether the PROGINS allele decreases or increases stability of the transcript. Work by Kieback et al. indicates the PROGINS variant has increases stability and transcriptional activity (**68, 77, 78**). However, according to work by

Romano et al. (**76**), the Val660Leu allele seems to be transcribed at a lower rate than the wild-type. On the other hand, transcription factors such as *ER-alpha* and *Sp1* bind to the Alu insertion and in their presence, the amount of *PGR* transcript increased. They also found that the ALU insertion decreased stability of the transcript and that the Val660Leu polymorphism conferred reduced activity. Clearly more studies need to be done to identify whether these effects are specific to particular tissues or cell lines, or depend on interaction with other factors.

The PROGINS variant may be a risk factor for ovarian cancer, possibly only among women who have not taken oral contraceptives (**79-81**). It may also be a risk factor for endometrial cancer (**82**). In addition, the Val660→Leu SNP was associated with increased risk of breast cancer in a large study (**83**). It was also shown to be protective of uterine fibroids in Brazilian non-white women (**84**).

PGR polymorphisms have also been examined with respect to other reproductive characteristics. Carriers of the PROGINS variant were more likely to be nulliparous, infertile, and experience irregular menstrual cycles; and were less likely to have premenstrual weight gain or breast pain, among controls in a case-control study of ovarian cancer (**85**). The PROGINS variant was also significantly associated with spontaneous abortion in another case-control study (**71**). No studies were identified that examined the association between progesterone polymorphisms and age at menarche. This variant may affect age at menarche by altering progesterone levels and/or permitting higher baseline estrogen activity, thus altering hormonal feedback mechanisms.

1.3 Menstrual Function

1.3.1 Biology and Epidemiology

Historically, women of reproductive age spent much of their lives in pregnancy or lactation-induced amenorrhea and likely only experienced 30-40 periods in their lifetimes. However, the average woman today menstruates over 400 times in her lifetime (**86**). The menstrual cycle provides epidemiologists with a sensitive and non-invasive measure of reproductive health. Although daily urinary hormone levels are the "gold standard" for measuring menstrual function in reproductive epidemiology, additional information on menstrual function may be measured noninvasively by calculating menstrual cycle length and menstrual bleed length. Other possible measurements include the volume of blood loss or perceived pain during menstruation. Using symptoms reported by patients, menstrual dysfunction may be classified as menorrhagia (heavy menstrual bleeding, or more than 80 mLs of blood lost) (**87**); dysmenorrhea (painful menstrual bleeding); or irregular, frequent, prolonged, sporadic, or absent periods (**88**). Alternatively, it could be classified in a more specific scientific manner, such as luteal phase deficiency (LPD), which is the "recurrent postovulatory deficiency in the production and/or effect of progesterone from the corpus luteum" (**89**). However, even LPD has heterogeneous causes.

The monthly pattern of menstruation directly reflects ovarian and endocrine function. Abnormal patterns may indicate subclinical physiological states that contribute to future disease development. For example, menstrual cycle patterns and the frequency of ovulation have been associated with a woman's risk of breast and ovarian cancer (**90-**

95). Particular cycle characteristics are also associated with age at menopause (**96**), osteoporosis, fracture risk (**97**), heart disease (**98, 99**), ovarian cancer (**90**), uterine fibroids (**100**), and diabetes and cardiovascular disease (**101**). There is accumulating evidence (some of which was contributed by some committee members on this dissertation) that menstrual cycle characteristics are related to reproductive success as measured by fertility or risk of spontaneous abortion (**102-104**).

Costs of medical treatment of menstrual dysfunction are suspected to be large. Approximately one-third of hysterectomies performed in England are reported to be a result of menstrual dysfunction (**86**). In a survey of women in the Oxford area, 50% reported at least one menstrual problem, and 20% had consulted their general practitioner about menstrual problems in the past 12 months (**105**). Luteal phase deficiency may be may be responsible for as much as 10% of infertility and 25% of habitual abortion (**89**). Estimates of the current prevalence of dysmenorrhea range from 43 to 90%; the lack of standardized diagnostic criteria makes it difficult to estimate the prevalence accurately (**2**). In a recent study of Norwegian athletes and controls, 15.2% of the control group reported current menstrual dysfunction (**106**). The prevalence of hysterectomies is also alarmingly large, with estimated lifetime prevalence approaching 50% for American populations, a large proportion of which is indicated by menstrual dysfunction (**86, 107**).

Factors associated with menstrual function include race, age, cigarette smoking, alcohol consumption, and physical activity (**108**). Smoking was associated with several types of abnormal menstrual characteristics in a study of 2115 premenopausal women (**109**), with current smoking having more adverse effects than ever smoking. Smoking

has also been associated with short cycles (**104**). Body mass index, parity, and age at menarche (**110**), in addition to caffeine consumption (**111**), may also be associated with menstrual function. Increasing alcohol intake was associated with higher estradiol levels throughout the menstrual cycle (**112**). A BMI above 35 was associated with a 5 fold increase in the odds of a long cycle (36 days or more) in one study (**104**). Occupational exposures, such as glycol ethers, may be associated with longer menstrual cycle length and subfertility (**113**).

1.3.2 Genetic Influences on Menstrual Function

Little is known about genetic influences on menstrual function. Genetic factors known to be associated with menstrual function are polymorphisms in some of the CYP enzymes, which are involved in the estrogen metabolism pathway (**112, 114, 115**). A polymorphism in the FSH receptor, which appears to increase the ovarian threshold to FSH, is associated with longer menstrual cycles (**116**). Epidemiologic studies of monozygotic and dizygotic twins suggest that menorrhagia may have a genetic cause (**117**). Common genetic polymorphisms that alter menstrual function may not only affect reproductive function, but also may affect long term risk of chronic diseases.

1.4 Fecundability

1.4.1 Biology and Epidemiology

Longer time to achieve pregnancy is an indication of subfertility or lower fecundability. The proportion of women undergoing fertility treatment has dramatically increased due to increased age when trying to conceive and increased availability of

assisted reproductive technologies (ART). An international study of infertility prevalence and treatment found a median of 9% infertility (12 months of unprotected intercourse without conception); the prevalence was similar in developed and less developed countries, and approximately 50% of infertile couples sought medical care, while 22% received it (**118**). In 1995, 15% of women reported ever seeing a doctor for infertility; 2% reported seeing a doctor within the past year; and 7% of couples reported trying for 12 months or more without conception (thereby meeting the clinical definition of infertility) (**119**). Among those seeking infertility services in 1995, approximately 26% reported problems with ovulation or menstruation, 26% with ovarian cysts, 22% had sperm or semen problems, 18% had endometriosis and 17% reported blocked tubes (**120**). Overall, of all infertility cases, 20% are idiopathic, meaning there is no known cause. Of those infertility cases that have an identified cause, 40% are thought to be due to female factors, 40% due to male factors, and 20% due to both genders or an unknown cause (**121**).

Costs associated with infertility are direct and indirect. There is direct financial burden of undergoing ART itself, and indirect, in terms of lost time at work. The public health burden of ART is an effect of both its costs up-front, and also because over 50% of ART result in twins and higher order multiples, according to the CDC report of ART (**122**). Although average IVF cycle charges are between $10,000 and $15,000, analysts have reported costs of about $35,000 per delivery in younger women, while medical costs per delivery are over $132,000 for women over 40 due to increased complications and multiple births (**123, 124**). There are also emotional consequences of infertility such as anxiety, depression, anger and guilt (**125, 126**).

Epidemiologic studies using time to pregnancy as an outcome have traditionally been complicated by the inconsistent use of birth control, inconsistent timing of intercourse across the menstrual cycle and dependence on recall. Factors associated with infertility or reduced fertility include increased maternal age (**127**), black race (**128**), timing and frequency of sexual intercourse (**129**), IUD contraceptive use (**130**), obesity (**131**), nulliparity (**132**), a history of STD's (**127**), tubal impairments (**121**), abdominal surgery and endometriosis (**133**), cigarette smoking or exposure to maternal smoking *in utero* (**127, 134**), exposure to DES *in utero* (**135**), rigorous exercise (**136**), poor nutrition (**137**), stress (**138**), environmental chemical exposures such as glycol ethers (**139**), and possibly caffeine (**140**) and alcohol (**141**).

1.4.2 Genetic Influences on Fecundability

As in the case with menstrual function, polymorphisms affecting fecundability have not been well-studied. Known genetic factors influencing fertility are summarized in Layman (**142**). Most of these genes cause sterility and also have pronounced effects on pubertal development or timing or gonadal development. Polymorphisms with subtle effects causing subfertility or normal variation in fertility have not been presented in the literature. However, evidence shows that fertility in animals does have a genetic component; this can be measured by heritability of litter size, for example (**143, 144**).

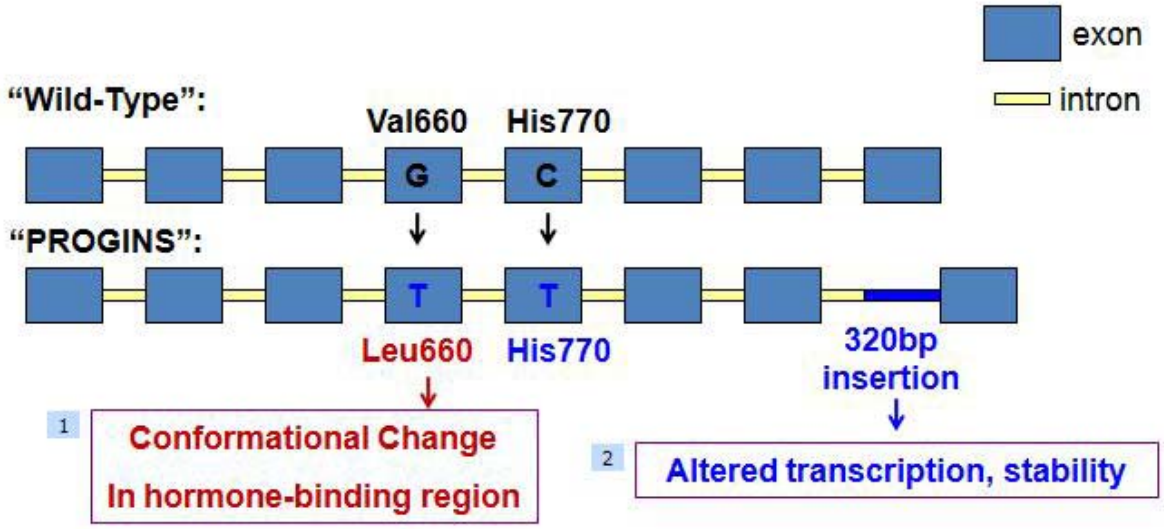
1.5 Summary

Age at menarche, menstrual function, and fecundability are reproductive characteristics that share underlying endocrine influences. These traits are influenced by

both genetic and environmental factors. This dissertation examines the relative contributions of genes and environmental factors to these reproductive traits, examines three specific candidate genetic and environmental influences, and evaluates candidate gene-environment interactions.

1.6 Figure

Figure 1.1 Structure of the Progesterone Receptor Gene, Highlighting Differences between the Wild-Type and "PROGINS" Variant



1.7 References

1. Nathan BM, Palmert MR. Regulation and disorders of pubertal timing. *Endocrinology and Metabolism Clinics of North America* 2005;**34**(3): 617-41, ix.

2. Jabbour HN, Kelly RW, Fraser HM, Critchley HOD. Endocrine regulation of menstruation. *Endocrine Reviews* 2006;**27**(1): 17-46.

3. Biro FM, Huang B, Crawford PB, et al. Pubertal correlates in black and white girls. *The Journal of Pediatrics* 2006;**148**(2): 234-40.

4. McDowell MA, Brody DJ, Hughes JP. Has age at menarche changed? Results from the National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Journal of Adolescent Health* 2007;**40**(3): 227-31.

5. Anderson SE, Must A. Interpreting the continued decline in the average age at menarche: results from two nationally representative surveys of U.S. girls studied 10 years apart.. *Journal of Pediatrics* 2005;**147**(6): 753-60.

6. Herman-Giddens ME. The decline in the age of menarche in the United States: should we be concerned? *Journal of Adolescent Health* 2007;**40**(3): 201-3.

7. Johansson T, Ritzen EM. Very long-term follow-up of girls with early and late menarche. *Endocrine Development* 2005;**8**: 126-36.

8. Burt SA, McGue M, DeMarte JA, Krueger RF, Iacono WG. Timing of menarche and the origins of conduct disorder. *Archives of General Psychiatry* 2006;**63**(8): 890-6.

9. Stice E, Presnell K, Bearman SK. Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. *Developmental Psychology* 2001;**37**(5): 608-19.

10. Wilson DM, Killen JD, Hayward C, et al. Timing and rate of sexual maturation and the onset of cigarette and alcohol use among teenage girls. *Archives of Pediatric and Adolescent Medicine* 1994;**148**(8): 789-95.

11. MacMahon B, Trichopoulos D, Brown J, et al. Age at menarche, urine estrogens and breast cancer risk. *International Journal of Cancer* 1982;**30**(4): 427-31.

12. Zeleniuch-Jacquotte A, Shore RE, Koenig KL, et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. *British Journal of Cancer* 2004;**90**(1): 153-9.

13. Bernstein L, Ross RK. Endogenous hormones and breast cancer risk. *Epidemiology Reviews* 1993;**15**(1): 48-65.

14. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiology Reviews* 1993;**15**(1): 36-47.

15. Hsieh C, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. *International Journal of Cancer* 1990;**46**: 796-800.

16. Purdie DM, Green AC. Epidemiology of endometrial cancer. *Best Practice & Research in Clinical Obstetrics & Gynaecology* 2001;**15**(3): 341-54.

17. Kaaks R, Lukanova A, Kurzer MS. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiology, Biomarkers & Prevention* 2002;**11**(12): 1531-43.

18. Golub S. Menarche: the beginning of menstrual life. *Women's Health* 1983;**8**(2-3): 17-36.

19. Graham C, Maude GH, Serjeant GR. Delayed menarche in homozygous sickle cell disease. *West Indian Medical Journal* 1986; **35**(1): 18-22.

20. Rees M. The age of menarche. *Orgyn* 1995(4): 2-4.

21. Osei-Hyiaman D, Satoshi T, Ueji M, Hideto T, Kano K. Timing of menopause, reproductive years, and bone mineral density: a cross-sectional study of postmenopausal Japanese women. *American Journal of Epidemiology* 1998; **148**(11): 1055-61.

22. Ito M, Yamada M, Hayashi K, Ohki M, Uetani M, Nakamura T. Relation of early menarche to high bone mineral density. *Calcified Tissue International* 1995;**57**(1): 11-4.

23. Cooper GS, Ephross SA, Weinberg CR, Baird DD, Whelan EA, Sandler DP. Menstrual and Reproductive Risk Factors for Ischemic Heart Disease. *Epidemiology* 1999;**10**(3): 255-9.

24. Axmon A, Rylander L, Albin M, Hagmar L. Factors affecting time to pregnancy. *Human Reproduction* 2006; **21**(5): 1279-84.

25. McKibben SL, Poston DL, Jr. The influence of age at menarche on the fertility of Chinese women. *Social Biology* 2003;**50**(3-4): 222-37.

26. Otor SC, Pandey A. Puberty and the family formation process in Sudan: age-at-menarche differential fecundity hypothesis revisited. *Social Biology* 1998;**45**(3-4): 246-59.

27. Kirk KM, Blomberg SP, Duffy DL, Heath AC, Owens IP, Martin NG. Natural selection and quantitative genetics of life-history traits in Western women: a twin study. *Evolution* 2001;**55**(2): 423-35.

28. Sandler DP, Wilcox AJ, Horney LF. Age at menarche and subsequent reproductive events. *American Journal of Epidemiology* 1984;**119**(5): 765-74.

29. Bracken MB, Bryce-Buchanan C, Stilten R, Holford T. Menarcheal age and habitual miscarriage: evidence for an association. *Annals of Human Biology* 1985;**12**(6): 525-31.

30. Frisch RE, McArthur JW. Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance or onset. *Science* 1974 13;**185**(4155): 949-51.

31. Maclure M, Travis LB, Willett W, MacMahon B. A prospective cohort study of nutrient intake and age at menarche. *American Journal of Clinical Nutrition* 1991;**54**(4): 649-56.

32. Tam CS, de Zegher F, Garnett SP, Baur LA, Cowell CT. Opposing influences of prenatal and postnatal growth on the timing of menarche. *Journal of Clinical Endocrinology and Metabolism* 2006;**91**(11): 4369-73.

33. Matkovic V, Ilich JZ, Skugor M, et al. Leptin is inversely related to age at menarche in human females. *Journal of Clinical Endocrinology and Metabolism* 1997;**82**(10): 3239-45.

34. Ellison PT. Skeletal growth, fatness and menarcheal age: a comparison of two hypotheses. *Human Biology* 1982;**54**(2): 269-81.

35. William D. Lassek SJCG. Menarche is related to fat distribution. *American Journal of Physical Anthropology* 2007;**133**(4): 1147-51.

36. Frisch RE. Body fat, menarche, fitness and fertility. *Human Reproduction* 1987;**2**(6): 521-33.

37. Rebar RW, Cumming DC. Reproductive function in women athletes. *Journal of the American Medical Association* 1981;**246**(14): 1590.

38. Blanck HM, Marcus M, Tolbert PE, et al. Age at menarche and tanner stage in girls exposed in utero and postnatally to polybrominated biphenyls. *Epidemiology* 2000;**11**(6): 641-7.

39. Den Hond E, Roels HA, Hoppenbrouwers K, et al. Sexual maturation in relation to polychlorinated aromatic hydrocarbons: Sharpe and Skakkebaek's hypothesis revisited. *Environmental Health Perspectives* 2002;**110**(8): 771-6.

40. Brouwer A, Ahlborg UG, Van den Berg M, et al. Functional aspects of developmental toxicity of polyhalogenated aromatic hydrocarbons in experimental animals and human infants. *European Journal of Pharmacology* 1995;**293**(1): 1-40.

41. Denham M, Schell LM, Deane G, Gallo MV, Ravenscroft J, DeCaprio AP. Relationship of lead, mercury, mirex, dichlorodiphenyldichloroethylene, hexachlorobenzene, and polychlorinated biphenyls to timing of menarche among Akwesasne Mohawk girls. *Pediatrics* 2005;**115**(2): e127-34.

42. Goodman MJ. Secular changes in recalled age at menarche. *Annals of Human Biology* 1983;**10**(6): 585.

43. Jones B, Leeton J, McLeod I, Wood C. Factors influencing the age of menarche in a lower socio-economic group in Melbourne. *Medical Journal of Australia* 1972;**2**(10): 533-5.

44. Popenoe P. Inheritance of age of onset of menstruation. *Eugenics News* 1928;**13**(101).

45. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at menarche in girls from the Fels Longitudinal Study. *American Journal of Physical Anthropology* 2005;**128**(1): 210-9.

46. Treloar SA, Martin NG. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. *American Journal of Human Genetics* 1990;**47**(1): 137-48.

47. Roche A. *Growth, maturation and body composition: the Fels Longitudinal Study 1929-1991*. Cambridge: Cambridge University Press; 1992.

48. Long JR, Xu H, Zhao LJ, et al. The oestrogen receptor alpha gene is linked and/or associated with age of menarche in different ethnic groups. *Journal of Medical Genetics* 2005;**42**(10): 796-800.

49. Stavrou I, Zois C, Ioannidis JP, Tsatsoulis A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. *Human Reproduction* 2002;**17**(4): 1101-5.

50. Stavrou I, Zois C, Chatzikyriakidou A, Georgiou I, Tsatsoulis A. Combined estrogen receptor alpha and estrogen receptor beta genotypes influence the age of menarche. *Human Reproduction* 2006;**21**(2): 554-7.

51. Xita N, Tsatsoulis A, Stavrou I, Georgiou I. Association of SHBG gene polymorphism with menarche. *Molecular Human Reproduction* 2005;**11**(6): 459-62.

52. Feigelson HS, Coetzee GA, Kolonel LN, Ross RK, Henderson BE. A polymorphism in the CYP17 gene increases the risk of breast cancer. *Cancer Research* 1997;**57**(6): 1063-5.

53. Gorai I, Tanaka K, Inada M, et al. Estrogen-metabolizing gene polymorphisms, but not estrogen receptor-alpha gene polymorphisms, are associated with the onset of menarche in healthy postmenopausal Japanese women. *Journal of Clinical Endocrinology and Metabolism* 2003;**88**(2): 799-803.

54. Lai J, Vesprini D, Chu W, Jernstrom H, Narod SA. CYP Gene Polymorphisms and Early Menarche. *Molecular Genetics and Metabolism* 2001;**74**(4): 449-57.

55. Jorm AF, Christensen H, Rodgers B, Jacomb PA, Easteal S. Association of adverse childhood experiences, age of menarche, and adult reproductive behavior: does the androgen receptor gene play a role? *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 2004;**125**(1): 105-11.

56. Kitagawa I, Kitagawa Y, Kawase Y, Nagaya T, Tokudome S. Advanced onset of menarche and higher bone mineral density depending on vitamin D receptor gene polymorphism. *European Journal of Endocrinology* 1998;**139**(5): 522-7.

57. Zhao J, Xiong D-H, Guo Y, Yang T-L, Recker RR, Deng H-W. Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in caucasian females. *Human Reproduction* 2007;**22**(6); 1789-94.

58. Mitchell ES, Farin FM, Stapleton PL, et al. Association of estrogen-related polymorphisms with age at menarche, age at final menstrual period, and stages of the menopausal transition. *Menopause* 2008;**15**(1): 105-11.

59. Guo Y, Shen H, Xiao P, et al. Genomewide linkage scan for quantitative trait loci underlying variation in age at menarche. *Journal of Clinical Endocrinology and Metabolism* 2006;**91**(3): 1009-14.

60. Rothenbuhler A, Fradin D, Heath S, et al. Weight-adjusted genome scan analysis for mapping quantitative trait Loci for menarchal age. *Journal of Clinical Endocrinology & Metabolism* 2006;**91**(9): 3534-7.

61. Misrahi M, Venencie P-Y, Saugier-Veber P, Sar S, Dessen P, Milgrom E. Structure of the human progesterone receptor gene. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1993;**1216**(2): 289-92.

62. Lee K, Jeong J, Tsai MJ, Tsai S, Lydon JP, DeMayo FJ. Molecular mechanisms involved in progesterone receptor regulation of uterine function. *The Journal of Steroid Biochemistry and Molecular Biology* 2006;**102**(1-5): 41-50.

63. Brenner R, Slayden O. Cyclic changes in the primate oviduct and endometrium. In: Knobil E, Neill J, editors. *The Physiology of Reproduction* 1994; p. 541–69.

64. Mesiano S, Chan E-C, Fitter JT, Kwek K, Yeo G, Smith R. Progesterone Withdrawal and Estrogen Activation in Human Parturition Are Coordinated by Progesterone Receptor A Expression in the Myometrium. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(6): 2924-30.

65. Baulieu EE. Contragestion and other clinical applications of RU 486, an antiprogesterone at the receptor. *Science* 1989;**245**(4924): 1351-7.

66. Brown A, Cheng L, Lin S, Baird DT. Daily low-dose mifepristone has contraceptive potential by suppressing ovulation and menstruation: a double-blind randomized control trial of 2 and 5 mg per day for 120 days. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(1): 63-70.

67. De Vivo I, Huggins GS, Hankinson SE, et al. A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. *Proceedings of the National Academy of Sciences* 2002;**99**(19): 12263-8.

68. Kieback D, Tong X-W, Weigel N, Agoulnik I. A genetic mutation in the progesterone receptor (PROGINS) leads to an increased risk of non-familial breast and ovarian cancer causing inadequate control of estrogen receptor driven proliferation. *Journal of the Society for Gynecologic Investigation* 1998;**5**.

69. Rowe SM, Coughlan SJ, McKenna NJ, et al. Ovarian carcinoma-associated TaqI restriction fragment length polymorphism in intron G of the progesterone receptor gene is due to an Alu sequence insertion. *Cancer Research* 1995;**55**(13): 2743-5.

70. Pijnenborg JM, Romano A, Dam-de Veen GC, et al. Aberrations in the progesterone receptor gene and the risk of recurrent endometrial carcinoma. *Journal of Pathology* 2005 **205**(5): 597-605.

71. Schweikert A, Rau T, Berkholz A, Allera A, Daufeldt S, Wildt L. Association of progesterone receptor polymorphism with recurrent abortions. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 2004;**113**(1): 67-72.

72. Wang-Gohrke S, Chang-Claude J, Becher H, Kieback DG, Runnebaum IB. Progesterone Receptor Gene Polymorphism Is Associated with Decreased Risk for Breast Cancer by Age 50. *Cancer Research* 2000;**60**(9): 2348-50.

73. Tetel MJ, Jung S, Carbajo P, Ladtkow T, Skafar DF, Edwards DP. Hinge and amino-terminal sequences contribute to solution dimerization of human progesterone receptor. *Molecular Endocrinology* 1997;**11**(8): 1114-28.

74. Tsai MJ, O'Malley BW. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annual Reviews of Biochemistry* 1994;**63**: 451-86.

75. Jackson TA, Richer JK, Bain DL, Takimoto GS, Tung L, Horwitz KB. The partial agonist activity of antagonist-occupied steroid receptors is controlled by a novel hinge domain-binding coactivator L7/SPA and the corepressors N-CoR or SMRT. *Molecular Endocrinology* 1997;**11**(6): 693-705.

76. Romano A, Delvoux B, Fischer D-C, Groothuis P. The PROGINS polymorphism of the human progesterone receptor diminishes the response to progesterone. *J Molecular Endocrinology* 2007;**38**(2): 331-50.

77. Agoulnik IU, Tong XW, Fischer DC, et al. A germline variation in the progesterone receptor gene increases transcriptional activity and may modify ovarian cancer risk. *Journal of Clinical Endocrinology and Metabolism* 2004; **89**(12): 6340-7.

78. McKenna NJ, Kieback DG, Carney DN, Fanning M, McLinden J, Headon DR. A germline TaqI restriction fragment length polymorphism in the progesterone receptor gene in ovarian carcinoma. *British Journal of Cancer* 1995;**71**(3): 451-5.

79. Runnebaum IB, Wang-Gohrke S, Vesprini D, et al. Progesterone receptor variant increases ovarian cancer risk in BRCA1 and BRCA2 mutation carriers who were never exposed to oral contraceptives. *Pharmacogenetics* 2001;**11**(7): 635-8.

80. Romano A, Lindsey PJ, Fischer D-C, et al. Two functionally relevant polymorphisms in the human progesterone receptor gene (+331 G/A and progins) and the predisposition for breast and/or ovarian cancer. *Gynecologic Oncology* 2006;**101**(2): 287-95.

81. Pearce CL, Hirschhorn JN, Wu AH, et al. Clarifying the PROGINS Allele Association in Ovarian and Breast Cancer Risk: A Haplotype-Based Analysis. *Journal of the National Cancer Institute* 2005;**97**(1): 51-9.

82. Junqueira MG, da Silva IDCG, Nogueira-de-Souza NC, et al. Progesterone receptor (PROGINS) polymorphism and the risk of endometrial cancer development. *International Journal of Gynecological Cancer* 2007;**17**(1): 229-32.

83. Pooley KA, Healey CS, Smith PL, et al. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. *Cancer Epidemiology, Biomarkers & Prevention* 2006 **15**(4): 675-82.

84. Gomes MTV, Castro RdA, Villanova FE, et al. The progesterone receptor gene polymorphism, PROGINS, may be a factor related to the development of uterine fibroids. *Fertility and Sterility* 2007;**87**(5): 1116-21.

85. Terry KL, De Vivo I, Titus-Ernstoff L, Sluss PM, Cramer DW. Genetic Variation in the Progesterone Receptor Gene and Ovarian Cancer Risk. *American Journal of Epidemiology* 2005;**161**(5): 442-51.

86. Coulter A, McPherson K, Vessey M. Do British women undergo too many or too few hysterectomies? *Social Science and Medicine* 1988;**27**(9): 987-94.

87. Hallberg L, Hogdahl AM, Nilsson L, Rybo G. Menstrual blood loss--a population study. Variation at different ages and attempts to define normality. *Acta Obstetricia et Gynecologica Scandinavicad* 1966;**45**(3): 320-51.

88. Livingstone M, Fraser IS. Mechanisms of abnormal uterine bleeding. *Human Reproduction Update* 2002;**8**(1): 60-7.

89. Soules MR. Luteal Phase Deficiency: The Most Common Abnormality of the Menstrual Cycle? In: Pirke KM, Wuttke W, Schweiger U, editors. *The Menstrual Cycle and Its Disorders*. New York: Springer-Verlag; 1989. p. 97-109.

90. Parazzini F, La Vecchia C, Negri E, Gentile A. Menstrual factors and the risk of epithelial ovarian cancer. *Journal of Clinical Epidemiology* 1989;**42**(5): 443-8.

91. Garland M, Hunter DJ, Colditz GA, et al. Menstrual cycle characteristics and history of ovulatory infertility in relation to breast cancer risk in a large cohort of US women. *American Journal of Epidemiology* 1998;**147**(7): 636-43.

92. Rockhill B, Moorman PG, Newman B. Age at menarche, time to regular cycling, and breast cancer (North Carolina, United States). *Cancer Causes Control* 1998; **9**(4): 447-53.

93. Bell DA, Taylor JA, Butler MA, et al. Genotype/phenotype discordance for human arylamine N-acetyltransferase (NAT2) reveals a new slow-acetylator allele common in African-Americans. *Carcinogenesis* 1993;**14**(8): 1689-92.

94. Bernstein JL, Thompson WD, Risch N, Holford TR. Risk factors predicting the incidence of second primary breast cancer among women diagnosed with a first primary breast cancer. *American Journal of Epidemiology* 1992;**136**(8): 925-36.

95. Bosetti C, Tavani A, Negri E, Trichopoulos D, La Vecchia C. Reliability of data on medical conditions, menstrual and reproductive history provided by hospital controls. *Acta Obstetricia et Gynecologica Scandinavica* 2001;**54**(9): 902-6.

96. Harlow BL, Signorello LB. Factors associated with early menopause. *Maturitas* 2000;**35**(1): 3-9.

97. Cooper GS, Sandler DP. Long-term effects of reproductive-age menstrual cycle patterns on peri- and postmenopausal fracture risk. *American Journal of Epidemiology* 1997;**145**(9): 804-9.

98. Nduka EU, Agbedana EO. Total cholesterol, high density lipoprotein cholesterol and steroid hormone changes in normal weight women during the menstrual cycle. *International Journal of Gynecology and Obstetrics* 1993;**41**(3): 265-8.

99. de Kleijn MJJ, van der Schouw YT, van der Graaf Y. Reproductive history and cardiovascular disease risk in postmenopausal women: A review of the literature. *Maturitas* 1999;**33**(1): 7-36.

100. Chen CR, Buck GM, Courey NG, Perez KM, Wactawski-Wende J. Risk factors for uterine fibroids among women undergoing tubal sterilization. *American Journal of Epidemiology* 2001;**153**(1): 20-6.

101. Solomon CG, Hu FB, Dunaif A, et al. Menstrual cycle irregularity and risk for future cardiovascular disease. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(5): 2013-7.

102. Small CM, Manatunga AK, Klein M, et al. Menstrual cycle characteristics: associations with fertility and spontaneous abortion. *Epidemiology* 2006;**17**(1): 52-60.

103. Hjollund NH, Jensen TK, Bonde JP, et al. Distress and reduced fertility: a follow-up study of first-pregnancy planners. *Fertility & Sterility* 1999;**72**(1): 47-53.

104. Rowland AS, Baird DD, Long S, et al. Influence of medical conditions and lifestyle factors on the menstrual cycle. *Epidemiology* 2002;**13**(6): 668-74.

105. Brown S. *The epidemiology of menstrual patterns.* Oxford: Oxford; 1987.

106. Torstveit MK, Sundgot-Borgen J. Participation in leanness sports but not training volume is associated with menstrual dysfunction: a national survey of 1276 elite athletes and controls. *British Journal of Sports Medicine* 2005;**39**(3): 141-7.

107. Merrill RM. Prevalence Corrected Hysterectomy Rates and Probabilities in Utah. *Annals of Epidemiology* 2001;**11**(2): 127-35.

108. Liu Y, Gold EB, Lasley BL, Johnson WO. Factors affecting menstrual cycle characteristics. *American Journal of Epidemiology* 2004;**160**(2): 131-40.

109. Brown S, Vessey M, Stratton I. The influence of method of contraception and cigarette smoking on menstrual patterns. *British Journal of Obstetrics & Gynaecology* 1988;**95**(9): 905-10.

110. Kato I, Toniolo P, Koenig KL, et al. Epidemiologic correlates with menstrual cycle length in middle aged women. *European Journal of Epidemiology* 1999;**15**(9): 809-14.

111. Fenster L, Quale C, Waller K, et al. Caffeine Consumption and Menstrual Function. *American Journal of Epidemiology* 1999;**149**(6): 550-7.

112. García-Closas M, Herbstman J, Schiffman M, Glass A, Dorgan JF. Relationship between serum hormone concentrations, reproductive history, alcohol consumption and genetic polymorphisms in pre-menopausal women. *International Journal of Cancer* 2002;**102**(2): 172-8.

113. Hsieh GY, Wang JD, Cheng TJ, Chen PC. Prolonged menstrual cycles in female workers exposed to ethylene glycol ethers in the semiconductor manufacturing industry. *Occupational and Environmental Medicine* 2005;**62**(8): 510-6.

114. Lurie G, Maskarinec G, Kaaks R, Stanczyk FZ, Le Marchand L. Association of Genetic Polymorphisms with Serum Estrogens Measured Multiple Times During a 2-Year Period in Premenopausal Women. *Cancer Epidemiology Biomarkers and Prevention* 2005;**14**(6): 1521-7.

115. Small CM, Marcus M, Sherman SL, Sullivan AK, Manatunga AK, Feigelson HS. CYP17 genotype predicts serum hormone levels among pre-menopausal women. *Human Reproduction* 2005;**20**(8): 2162-7.

116. Greb RR, Grieshaber K, Gromoll J, et al. A common single nucleotide polymorphism in exon 10 of the human follicle stimulating hormone receptor is a major determinant of length and hormonal dynamics of the menstrual cycle. *Journal of Clinical Endocrinology & Metabolism* 2005;**90**(8): 4866-72.

117. Rybo G, Hallberg L. Influence of heredity and environment on normal menstrual blood loss. A study of twins. *Acta Obstetricia et Gynecologica Scandinavicad* 1966;**45**(4): 389-410.

118. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Human Reproduction* 2007;**22**(6): 1506-12.

119. Abma JC, Chandra A, Mosher WD, Peterson LS, Piccinino LJ. Fertility, family planning, and women's health: new data from the 1995 National Survey of Family Growth. *Vital Health Statistics* 1997; **23**(19): 1-114.

120. Stewart GK. Infertility. In: Hatcher RA, Trussell J, Stewart F, et al., editors. *Contraceptive Technology*. 17 ed. New York: Ardent Media; 1998.

121. Nelson AL, Marshall MD. Impaired Fertility. In: Hatcher RA, Trussell J, Stewart F, et al., editors. *Contraceptive Technology*. New York: Ardent Media; 2004. p. 651-71.

122. *Assisted Reproductive Technology Surveillance --- United States, 2000*. 2000 [cited 2007 4/4]; Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5209a1.htm.

123. Van Voorhis BJ, Stovall DW, Allen BD, Syrop CH. Cost-effective treatment of the infertile couple. *Fertility and Sterility* 1998;**70**(6): 995-1005.

124. Perloe M. *Duration, Cost of IVF Treatment*. 2007 [Accessed April 4. 2007]; Available from: http://health.ivillage.com/infertility/infertivf/0,,677v,00.html.

125. Mahlstedt PP. The psychological component of infertility. *Fertility and Sterility* 1985;**43**(3): 335-46.

126. Golombok S. Psychological functioning in infertility patients. *Human Reproduction* 1992;**7**(2): 208-12.

127. Westhoff CL. The epidmemiology of infertility. In: Kiely M, editor. *Reproductive and Perinatal Epidemiology*. CRC Press; 1991.

128. Mosher WD, Aral SO. Factors related to infertility in the United States, 1965-1976. *Sexually Transmitted Diseases* 1985;**12**(3): 117-23.

129. Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. *New England Journal of Medicine* 1995;**333**(23): 1517-21.

130. Cramer DW, Schiff I, Schoenbaum SC, et al. Tubal infertility and the intrauterine device. *New England Journal of Medicine* 1985;**312**(15): 941-7.

131. Gesink Law DC, Maclehose RF, Longnecker MP. Obesity and time to pregnancy. *Human Reproduction* 2007;**22**(2): 414-20.

132. Chandra A. Infertility. In: Wilcox LS, Marks J, editors. *From Data to Action: CDC's Public Health surveillance for women, infants, and children* U.S. Department of Health and Human Services; 1994.

133. Healy DL, Trounson AO, Andersen AN. Female infertility: causes and treatment. *Lancet* 1994;**343**(8912): 1539-44.

134. Weinberg CR, Wilcox AJ, Baird DD. Reduced fecundability in women with prenatal exposure to cigarette smoking. *American Journal of Epidemiology* 1989;**129**(5): 1072-8.

135. Palmer JR, Hatch EE, Rao RS, et al. Infertility among women exposed prenatally to diethylstilbestrol. *American Journal of Epidemiology* 2001;**154**(4): 316-21.

136. Green BB, Daling JR, Weiss NS, Liff JM, Koepsell T. Exercise as a risk factor for infertility with ovulatory dysfunction. *American Journal of Public Health* 1986;**76**(12): 1432-6.

137. Wade GN, Jones JE. Neuroendocrinology of nutritional infertility. *American Journal Regulatory Integrative, and Comparative Physiol* 2004;**287**(6): R1277-96.

138. Wilson JF, Kopitzke EJ. Stress and infertility. *Current Womens Health Reports* 2002;**2**(3): 194-9.

139. Correa A, Gray RH, Cohen R, et al. Ethylene Glycol Ethers and Risks of Spontaneous Abortion and Subfertility. *American Journal of Epidemiology* 1996;**143**(7): 707-17.

140. Wilcox A, Weinberg C, Baird D. Caffeinated beverages and decreased fertility. *Lancet* 1988;**2**(8626-8627): 1453-6.

141. Eggert J, Theobald H, Engfeldt P. Effects of alcohol consumption on female fertility during an 18-year period. *Fertility and Sterility* 2004;**81**(2): 379-83.

142. Layman LC. Human gene mutations causing infertility. *Journal of Medical Genetics* 2002;**39**(3): 153-61.

143. D'Alessandro A, Martemucci G, Iaffaldano N. Active immunization with a synthetic fragment of pig inhibin alpha-subunit increases ovulation rate and embryo production in superovulated ewes but season affects its efficiency. *Journal of Reproduction and Fertility* 1999;**115**(1): 185-91.

144. Leyhe B, Hiendleder S, Jaeger C, Wassmuth R. Pronounced differences in the frequency of TaqI beta A-inhibin alleles between sheep breeds with different reproductive performance. *Animal Genetics* 1994;**25**(1): 41-3.

Chapter 2

Heritability of Age at Menarche and Menstrual Cycle Length in a Population Exposed to a Brominated Flame Retardant

2.1 Abstract

Objectives. We determined the heritability of age at menarche and menstrual cycle length in a population exposed to a brominated flame retardant, polybrominated biphenyls (PBBs). We further investigated whether exposure to this potential endocrine disruptor affected the heritability of these traits.

Methods. Our study population consisted of 373 families with variable PBB exposure. We estimated the heritability of self-reported age at menarche and menstrual cycle length using variance components methodology and assessed whether such heritability estimates varied by PBB exposure.

Results. Age at menarche and menstrual cycle length were both heritable (heritability of age at menarche, 0.53±0.05; cycle length, 0.42±0.10). Heritability of age at menarche was higher in those with PBB levels at or below the limit of detection (0.60±0.07) than in the highly exposed group (0.38±0.12), although the difference did not reach statistical significance (χ^2=2.50, p=0.11). Menstrual cycle length exhibited a non-significant trend similar to the menarche analysis when stratified by PBB exposure.

Conclusions. This study provides support that age at menarche and menstrual cycle length are heritable. These data suggest further investigation into potential gene-

environment interactions, in which exposure to PBBs potentially reduces or modifies the influence of genetic factors on these reproductive traits.

2.2 Introduction

Heritability is the proportion of variance in a trait that can be attributed to genetic factors. Studies of the heritability of age at menarche date to the 1920's, and it is well-established that age at menarche is highly heritable, with estimates ranging from 0.49 to 0.70 in recent studies (**1-4**). To our knowledge, the heritability of menstrual cycle length has not been studied.

Environmental exposures may affect age at menarche and menstrual function. Endocrine disruptors such as brominated flame retardants (**5**), other persistent organic pollutants (**6, 7**), and lead (**8**) have been associated with the timing of pubertal development and menarche (reviewed in **9**). There is also substantial evidence that environmental exposures such as diet and exercise affect age at menarche and menstrual function (**10-12**).

In 1973, an incident occurred in which cattle in Michigan consumed feed contaminated with polybrominated biphenyls (PBBs). Between mid-1973 and late 1974, an estimated 6.5 million Michigan residents consumed PBB-tainted beef, poultry, and dairy products (**13**). Because PBBs are lipophilic, they are concentrated and excreted in milk, and a substantial proportion of the meat and dairy products in Michigan during and shortly after the contamination period contained PBBs. A cross-sectional survey conducted in 1978 estimated that a majority of Michigan residents had detectable concentrations of PBBs (**14**).

In a study of females aged 5-24 in this population, it was demonstrated that high *in utero* exposure to PBBs (≥7 parts per billion [ppb]) was associated with earlier age at menarche among girls who were breastfed (mean age=11.6 among breastfed girls exposed to high concentrations of PBBs *in utero*; mean age=12.2-12.6 years among breastfed girls exposed to lower concentrations of PBBs *in utero*; mean age=12.7 among girls not breastfed) (**5**). Previous research from our group suggested a shorter menstrual cycle (3.5-5.5 days) and longer menstrual bleed (0.87 days) among women with the highest exposure to PBBs (≥12 ppb) who had also experienced weight loss (**15**).

Heritability of traits varies across and within populations. This can be conceptualized as heritability-environment interaction, previously described in studies of psychosocial characteristics. For example, heritability of IQ in children was modified by socioeconomic status (**16**) and parental education (**17**). We extend this concept to include the modification of heritability of reproductive characteristics by environmental exposures. In this study we estimate the heritability of self-reported age at menarche and menstrual cycle length, and assess whether such heritability estimates vary by serum concentrations of PBBs. We hypothesized that these traits would be less heritable in families with high exposure to PBBs because the influence of PBBs may reduce the proportion of total phenotypic variance attributable to genetic factors.

2.3 Methods

2.3.1 Population

In 1973-74, Michigan residents were exposed to substantial PBB concentrations through contaminated animal and dairy products when NutriMaster®, a nutritional cattle feed supplement, was inadvertently replaced with FireMaster®, a brominated flame-retardant. In 1976-77, the Michigan Department of Community Health (MDCH) enrolled ~4,000 individuals with a range of exposure levels into a registry for long term health monitoring (**13**). The study population, drawn from the Michigan Long-Term PBB Study, includes female cohort members 18 years and older who participated in at least one of two computer-assisted telephone interviews (CATIs) conducted between 1997 and 2006.

Of 1425 eligible women in 1997 who could be located, 1185 (83%) participated in the CATI. Of 581 eligible women in 2003 who could be located, 479 (82%) participated in the second CATI. The 2003 CATI included 157 women who had not previously participated in the 1997 CATI, bringing the total number of participants to 1342. Only women with a female relative who also participated were eligible for the current study (N=1056). Inclusion of women who were not relatives did not change fixed effect estimates.

To be included in the age at menarche analysis, a woman had to report an age at menarche in the range of 9-16 (N=1041) and have at least one relative who also reported age at menarche in this range (N=1033). For age at menarche, values outside the range

(mean ± two standard deviations, SD) were excluded. Extreme values were excluded because they may have a distinct etiology. For menstrual cycle length, we limited the analysis to premenopausal women who were not using hormonal contraceptives (N=544), reported a menstrual cycle length in the range of 20-40 days (N=519), and had a female relative in the study who also reported a menstrual cycle length in this range (N=367). The chosen range of 20-40 days was taken from the mean ±2 SD from a prospective study of menstrual function (**18**). The menstrual cycle sample was nearly a subset of the menarche sample; there were just four women who reported cycle length but not age at menarche, bringing the total sample size to 1037.

For the women who participated in both interviews and had a discrepancy between the two self-reported ages at menarche (N= 58 of 322 women, median difference=1 year), the age at menarche reported in the earlier interview was used because it was closer to the event. In adult women, mean cycle length decreases gradually with age (**12, 19**). For the 137 women who reported cycle length in both interviews, cycle length and covariate information from the most recent interview was used.

2.3.2 PBB Measurements

PBB serum measurements were taken at the time of enrollment into the cohort (1976-1977), or later (1978-1993) for women born after 1977 and for a few women whose enrollment PBB was not available. Quantification of PBBs was based on the main congener, 2,2',4,4',5,5' hexa-bromobiphenyl, also known as PBB-153. The methods of

PBB detection used have coefficients of variation of 7.1-14.0% and recovery ranges of 80-90% (**20**).

2.3.3 Determination of Relatedness

Relatedness among women in the cohort was determined at enrollment into the cohort in 1976-1977 or by reproductive histories taken in the 1997 or 2003 interviews. When ambiguous, relationships were verified by records kept by the Michigan Department of Community Health. If the relationship of a pair from the cohort could not be verified, then we conservatively assumed the pair was unrelated. The degree of relatedness was determined and taken into account in the heritability analyses.

2.3.4 Determination of Shared Household

Shared household was determined by comparing addresses of women in the same family during the most recent telephone interview. If the address was not available, then a household was considered shared if the same home telephone number was reported. This information was used only to assign shared households in the menstrual cycle length analysis. It was assumed that siblings shared a household when they experienced menarche, and that a parent and child did not share the same household when each experienced menarche. Therefore, a household variance component could not be estimated for age at menarche because it was completely unidentifiable from dominance genetic effects, which are also shared by siblings but not by parent and child (**21**).

2.3.5 Covariates

We used linear regression to model the following variables as potentially significant predictors of menstrual cycle length: age at interview, body mass index (BMI, kg/m^2), current smoking status (yes/no), and physical activity (Never/1-2 times per week/3-4 times per week/5 or more times per week). Age at interview, BMI, and smoking were significant predictors of cycle length and were included as covariates in the heritability analyses of cycle length. We did not have access to information on potential predictors of age at menarche, such as childhood BMI, and therefore no covariates were included in the heritability analysis of age at menarche.

2.3.6 Heritability Analyses

Heritability analyses were conducted using the genetic analysis software program MENDEL (**22**), which provides estimates of heritability adjusting for covariate effects. MENDEL allows for estimation of variance components including both additive and dominance genetic components, as well as shared household effects. The covariance between two individuals (Y_i, Y_j) is the sum of their shared variance components, including additive and dominant genetic components, and a component for shared environment (household). The degree of relatedness determines the covariance due to additive and dominance genetic effects. The coefficient for the covariance due to additive genetic effects is the kinship coefficient, which is the expected proportion of alleles shared identical by descent. The coefficient for dominance genetic effects is the probability that two individuals share *both* alleles at a locus identical by descent. For example, for full siblings, the coefficients for the covariance for additive and dominance

genetic effects are ½ and ¼, respectively. For a grandparent-grandchild relationship, the coefficients are ¼ and 0. Thus all degrees of relationships are taken into account in the heritability analyses.

There was evidence of digit bias among the self-reported menstrual cycle length, with many women reporting a cycle length of either 28 or 30 days. Because heritability estimates are sensitive to departures from normality, we transformed the cycle length variable by assigning women into 10 normally distributed "bins" of cycle length, creating a normally distributed trait.

Neither age at menarche nor menstrual cycle length had a significant dominance genetic variance component, as assessed with a likelihood ratio test comparing models with both dominance and additive components to models that included only an additive component. Therefore the heritability of these traits was estimated assuming an additive genetic component only. Statistical significance of heritability was determined using a likelihood ratio test comparing models with and without an additive genetic variance component. We also determined whether the shared household variance component for the menstrual cycle length analysis was significant using a likelihood ratio test. Under the null hypothesis, each of these tests follow a 50:50 mixture of a chi-square distribution with 1 degree of freedom and a point mass of 0. This distribution follows from testing the variance parameter on its boundary value of 0 under the null hypothesis (**23**).

2.3.7 Determination of Heritability by Concentrations of PBB Exposure

The analyses stratified by PBB concentrations were limited to women whose own serum PBB value or whose mother's serum PBB value was available (N=1034). In our primary analysis, we assigned anyone missing a PBB value (N=202) to their mother's PBB exposure category. Of the 506 mother-daughter pairs where PBB was measured in both individuals, the correlation coefficient was 0.40 ($p<0.0001$). The similarity in mothers' and daughters' PBB concentrations may result from similar diets around the time of the contamination incident, *in utero* exposure, or exposure through breastfeeding (**24**).

For our primary analysis we divided the population into those with PBB concentrations at or below the limit of detection (≤1 part per billion, ppb), and greater than the limit of detection (>1 ppb). For the age at menarche analysis, the sample size was large enough that approximate tertiles were also created, so that a dose-response relationship could be investigated. We divided the menarche sample at the limit of detection (1 ppb) and at the median of detectable values (3 ppb). If menarche occurred before PBB entered the food chain in Michigan, the individual was assigned to the lowest exposure group for the menarche analysis. In some cases, families had individual women that fell into different PBB exposure strata. When at least two women from a family were assigned to the same stratum, their data would contribute to the heritability analysis within that stratum; otherwise their data would contribute to fixed effects, but not heritability.

We estimated heritability of these traits within these strata and tested whether heritability was significantly different among the strata using a Wald test, which follows a chi-square distribution with one degree of freedom.

2.4 Results

Table 2.1 shows demographic information for the population samples used in the two analyses. The population was 99% white. Figure 2.1 shows that self-reported age at menarche was normally distributed with ages 12 and 13 years being the most common (median =13 years). Figure 2.2 illustrates the digit bias (28 and 30 days) in self-reported menstrual cycle length (median=28 days).

Kindreds were composed of women in up to three generations, and Table 2.2 shows that families contained between 2 and 8 women. Age at menarche was heritable (h^2=0.53±0.05; p<0.001). Menstrual cycle length, adjusted for age at interview, BMI, and current smoking, was also heritable (h^2=0.27±0.11, p=0.018 using the original cycle length data; h^2=0.42±0.10, p<0.001 using the transformed distribution). For menstrual cycle length we also estimated a shared household variance component. The household variance component was not significant; however, only eighteen pairs of women were confirmed as living at the same residence during the latest cohort update, and therefore we had little power to detect covariance due to shared household.

Heritability of age at menarche and menstrual cycle length both varied by levels of PBB exposure, though the variation did not reach statistical significance (Tables 2.3 and 2.4). Heritability of age at menarche was higher in the group with PBB

concentrations ≤1 ppb (0.60±0.07) than in the exposed group (0.38±0.12) although the difference was not statistically significant (χ^2=2.50, p=0.11). When divided into tertiles, heritability was highest in the unexposed (h^2=0.61±0.08), lower in the moderately exposed (0.45±0.16) and lowest in the highly exposed (0.40±0.16), although pairwise differences were not statistically significant (Wald test comparing tertile 1 to tertile 2, χ^2=0.80, p=0.37; comparing tertile 1 to tertile 3: χ^2=1.37, p=0.24). The same nonsignificant trend was observed in the cycle length analysis: the heritability was higher in the PBB≤1 group than in the exposed group (0.58±0.18 and 0.46±0.11, respectively; χ^2=0.32, p=0.57). The total variance of age at menarche and menstrual cycle length did not vary by PBB exposure.

In Tables 2.3 and 2.4, women with missing PBB values were assigned to the same tertile or category as their mother. As a comparison, we conducted sensitivity analyses excluding those who had missing PBB values. The heritability values changed slightly but the overall trend was maintained and our conclusions were not altered. Alternatively, we assigned all second-generation women (women who were born after the incident and therefore exposed to PBB *in utero* or through breastfeeding only) to their mother's PBB category, regardless of whether the daughter also had her own serum PBB value measured. The results of this analysis were almost identical to the original analysis.

2.5 Discussion

This study provides further evidence that age at menarche is heritable and that the heritability is a result of the effects of genetic factors that act in additive fashion on this trait. We have demonstrated that self-reported menstrual cycle length is heritable; to our

knowledge, this is the first time this has been reported in the literature. We have also shown that heritability may vary by levels of a chemical exposure. Age at menarche and cycle length were both less heritable among the highly exposed, although the difference did not reach statistical significance. The same trend was observed in both the menarche and the cycle length analysis, further supporting the possible influence of brominated flame retardants on the heritability of reproductive traits.

We did not find evidence of dominance genetic effects on either trait. Treloar and Martin (**2**) found that the Pearson correlation for age at menarche between dizygotic (DZ) twins was 0.19±0.04, and between monozygotic (MZ) twins was 0.64±.0.03. If the genetic components consisted only of additive effects, and there were no shared environmental effects, then the correlation between MZ twins should be exactly twice that of DZ twins. They concluded that the genetic contribution to age at menarche is largely composed of dominance or epistasis effects. More consistent with our study design and our results, a recent study of kindreds of various sizes estimated the heritability of age at menarche to be 0.49±.13, indicating that approximately half of the variation in age at menarche in this population could be attributed to additive genetic factors (**1**). Dominance and household effects were not significant in their study.

Heritability is defined as the proportion of total phenotypic variance due to genetic factors (σ^2_g/σ^2). Therefore, decreased heritability in the higher categories of PBB exposure could be a result of larger σ^2 in those categories, or a smaller σ^2_g. However, the total variance of age at menarche and menstrual cycle length did not change substantially across categories of PBB exposure, suggesting that the observed altered heritability is not

a simple result of increased total variance in the trait, and is instead a result of a smaller σ^2_g in the exposed categories. This analysis suggests that exposure to PBB may reduce the influence of genetic factors on both age at menarche and menstrual cycle length.

PBBs have demonstrated both estrogenic and anti-estrogenic activity in vitro (**25-27**), and may exert these effects through the aryl hydrocarbon receptor, or estrogen receptors alpha or beta (**28, 29**). PBBs have been shown to alter timing of vaginal opening in rats (**30**). In rodents, oral exposure to PBBs has been found to accelerate steroid sex hormone metabolism (**27**) and demonstrated anti-estrogenic effects by attenuating the normal physiologic responses to estrogen (**31, 32**) and lengthening estrous cycles (**33**). Rhesus monkeys fed PBB-contaminated food for 6 months showed signs of ovarian hormone dysfunction including lengthened menstrual cycles and decreased serum progesterone concentrations (**34, 35**). We hypothesize that reduced heritability may result from altered activity or expression of genes that are involved with both menarche and/or menstrual function and are potentially affected by exposure to PBBs (e.g. genes downstream of the ER-β receptor).

A limitation of this study is the potential misclassification of age at menarche and menstrual cycle length. This retrospective data was self-reported in a telephone interview. While misclassification of age at menarche increases with age, studies have shown that self-reported age at menarche is generally reliable (**36, 37**). Self-reported menstrual cycle length is less reliable (**38, 39**). Consistent with the literature, older age at interview was associated with shorter cycle length in our study, providing some confidence in this measure (**12**). The presence of digit bias in the menstrual cycle length

distribution may increase heritability estimates if closely related people tend to mis-report their cycle length in the same way. Alternatively, digit preference may decrease heritability if related and unrelated women misreport their cycle length to the same degree, and therefore would introduce random error. A prospective study of cycle length that includes related women would be ideal to confirm or refine the results of the present study.

This unique cohort, with large family sizes and wide variation in PBB exposure, provides an excellent opportunity to estimate heritability and to determine if an environmental exposure could potentially alter heritability. Because both traits were highly heritable, we hypothesize that genetic factors may influence the underlying hormonal milieu and feedback mechanisms that affect pubertal development, menstrual function, and possibly other reproductive characteristics such as fertility. Heritability analyses such as these can determine whether it is worthwhile to pursue studies to elucidate the genetic mechanisms of these complex traits. They can also provide guidance into assessing whether complex traits are potentially influenced by gene-environment interaction.

2.6 Acknowledgments

This work has previously been published in the International Journal of Child and Adolescent Health. Permission was obtained from the journal to include it in this dissertation. Citation:

Taylor, Kira C., Chanley M. Small, Michael P. Epstein, Metrecia L. Terrell, and Michele Marcus. (2009). Heritability of age at menarche and menstrual cycle length in a population exposed to a brominated flame retardant. *International Journal of Child and Adolescent Health* 2(3).

We thank the staffs of the Michigan Department of Community Health and Michigan Public Health Institute for their assistance in data collection.

This manuscript was developed under a STAR Research Assistance Agreement No. 916891 awarded by the U.S. Environmental Protection Agency. It has **not** been formally reviewed by the EPA. The views expressed in this manuscript are solely those of Kira Creswell Taylor and the EPA **does not endorse** any products or commercial services mentioned in this manuscript.

Funding for this research was also provided by the U.S. Environmental Protection Agency (R 825300), National Institute of Environmental Health Sciences (RO1 ES08341 and RO1 ES012014), and by Centers for Disease Control and Prevention Cooperative Agreement U37/CCU500392.

The Institutional Review Boards at Emory University (691-2002) and the Michigan Department of Community Health (58-PHAEPI(R)) have approved this study.

2.7 Figures and Tables

Table 2.1 Characteristics of a Population Exposed to Polybrominated Biphenyls

Characteristic	Age at Menarche Sample N=1033 N (%)	Menstrual Cycle Length Sample N=367 N (%)
Age at Interview (yrs)		
18-30	354 (34)[a]	106 (29)[a]
31-40	300 (29)	160 (44)
41-60	269 (26)	101 (28)
60-85	110 (11)	--
BMI (kg/m2)		
<18.5	17 (2)	9 (2)
18.5-25	453 (44)	163 (44)
25-30	287 (28)	99 (27)
>30	272 (26)	95 (26)
Missing	*4 (<1)*	*1 (<1)*
Serum PBB value (ppb)		
<1	189 (18)	50 (14)
1-3	378 (37)	117 (32)
>3	264 (26)	128 (35)
Missing	*202 (20)*	*72 (20)*
Smoker	325 (31)	118 (32)
Education		
High school graduate or less	411 (40)	122 (33)

Some college	328 (32)	123 (34)
College graduate	292 (28)	122 (33)
Refused	*2 (<1)*	--

[a] There were two interviews. If a woman conducted both interviews, her age is taken from the interview that provided the data used in the analysis (the earlier interview for menarche, and the later interview for cycle length).

Table 2.2 Kindred Sizes in a Population Exposed to Polybrominated Biphenyls

	Number of Kindreds	
Kindred size	Age at Menarche study	Cycle length study
2	178	114
3	115	29
4	48	9
5	16	2
6	5	1
7	2	--
8	2	--
Total Families	366	155
Total Individuals	1033	367

Table 2.3 Heritability of Age at Menarche and Menstrual Cycle Length in a Population Exposed to Polybrominated Biphenyls

	Total Population	Low or no exposure (PBB≤1 ppb)	Exposed (PBB>1 ppb)
Age at Menarche	N=1033[a]	N=681[b]	N=349
Total variance in trait	1.86	1.97	1.80
Additive genetic variance	0.99	1.19	0.69
Heritability ± SE	0.53±0.05	0.60±0.07	0.38±0.12
Menstrual Cycle Length[c]	N=367	N=124	N=241
Total variance in trait	3.69	3.69	3.67
Additive genetic variance	1.56	2.10	1.70
Heritability ± SE	0.42±0.10	0.57±0.18	0.46±0.11

[a]The N of the total population is greater than the sum of the other two columns because for three women, neither they nor their mothers had PBB measurements taken.
[b]The N is large in the unexposed group for the menarche analysis because many women experienced menarche before the PBB contamination incident and were assigned to the lowest exposure category.
[c]Adjusted for age at interview, BMI, and current smoking. The transformed distribution was used in this analysis (transformed values are normally distributed and range from 1-10).

Table 2.4 Heritability of Age at Menarche According to Tertiles of PBB Exposure

	Unexposed (PBB<1 ppb) N=608[a]	Tertile 2 (1≤PBB≤3 ppb) N=242	Tertile 3 (PBB>3 ppb) N=180
Total variance in trait	1.94	1.66	2.05
Additive genetic variance	1.18	0.74	0.82
Heritability ± SE	0.61±0.08	0.45±0.16	0.40±0.16

[a]N is large in the unexposed group because many women experienced menarche before the PBB contamination incident and were assigned to the lowest exposure category.

Figure 2.1 Self-Reported Age at Menarche in a Population Exposed to Polybrominated Biphenyls

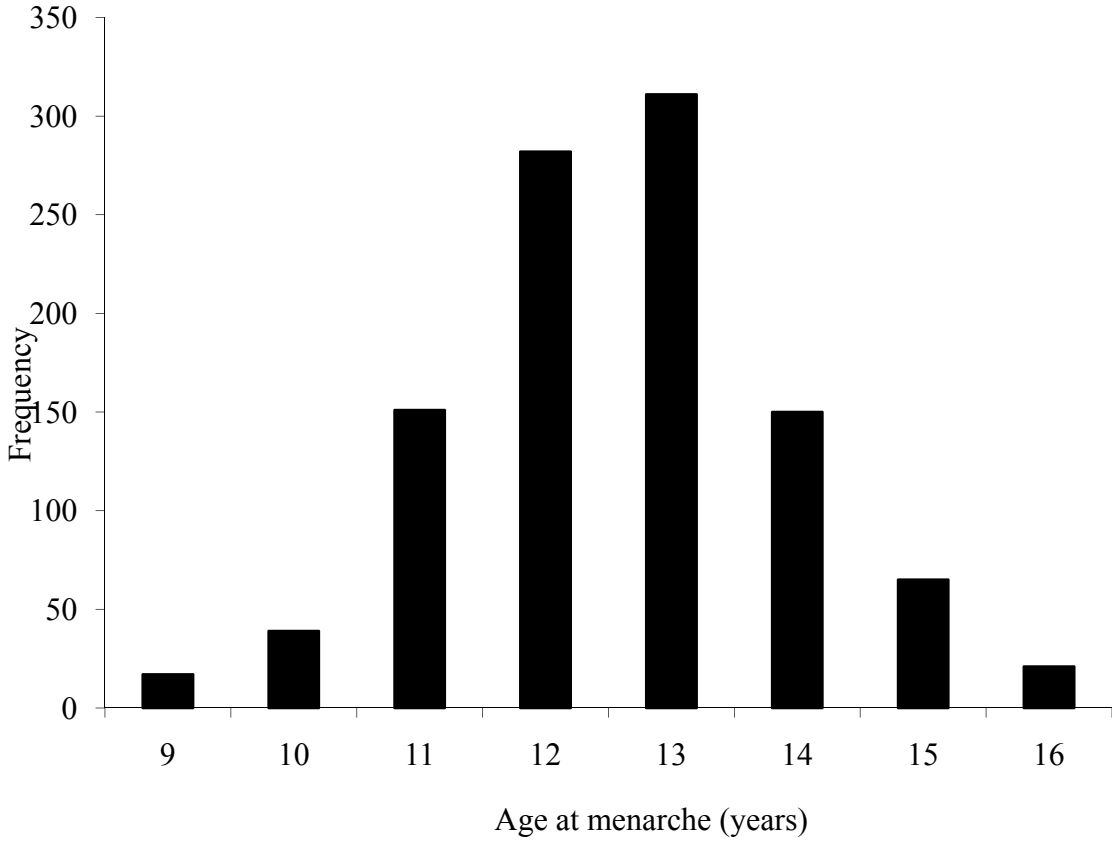
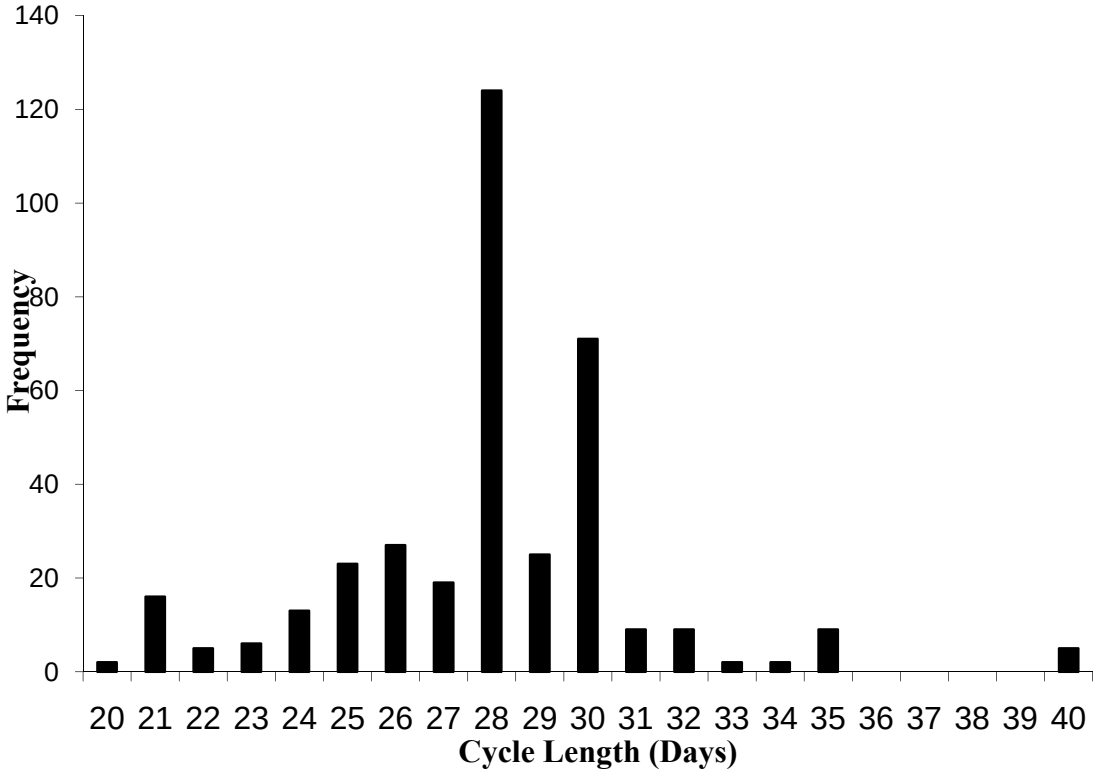


Figure 2.2 Self-Reported Menstrual Cycle Length in a Population Exposed to Polybrominated Biphenyls



2.8 References

1. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at menarche in girls from the Fels Longitudinal Study. *American Journal of Physical Anthropology* 2005;**128**(1): 210-9.

2. Treloar SA, Martin NG. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. *American Journal of Human Genetics* 1990;**47**(1): 137-48.

3. Guo Y, Shen H, Xiao P, et al. Genomewide linkage scan for quantitative trait loci underlying variation in age at menarche. *Journal of Clinical Endocrinology & Metabolism* 2006;**91**(3): 1009-14.

4. van den Berg SM, Boomsma DI. The familial clustering of age at menarche in extended twin families. *Behavior Genetics* 2007;**37**(5): 661-7.

5. Blanck HM, Marcus M, Tolbert PE, et al. Age at Menarche and Tanner Stage in Girls Exposed In Utero and Postnatally to Polybrominated Biphenyl. *Epidemiology* 2000;**11**(6): 641-7.

6. Denham M, Schell LM, Deane G, Gallo MV, Ravenscroft J, DeCaprio AP. Relationship of lead, mercury, mirex, dichlorodiphenyldichloroethylene, hexachlorobenzene, and polychlorinated biphenyls to timing of menarche among Akwesasne Mohawk girls. *Pediatrics* 2005;**115**(2): e127-34.

7. Vasiliu O, Muttineni J, Karmaus W. In utero exposure to organochlorines and age at menarche. *Human Reproduction* 2004;**19**(7): 1506-12.

8. Selevan SG, Rice DC, Hogan KA, Euling SY, Pfahles-Hutchens A, Bethel J. Blood lead concentration and delayed puberty in girls. *New England Journal of Medicine* 2003;**348**(16): 1527-36.

9. Buck Louis GM, Lynch CD, Cooney MA. Environmental influences on female fecundity and fertility. *Seminars in Reproductive Medicine* 2006;**24**(3): 147-55.

10. Maclure M, Travis LB, Willett W, MacMahon B. A prospective cohort study of nutrient intake and age at menarche. *American Journal of Clinical Nutrition* 1991;**54**(4): 649-56.

11. Frisch RE. Body fat, menarche, fitness and fertility. *Human Reproduction* 1987;**2**(6): 521-33.

12. Liu Y, Gold EB, Lasley BL, Johnson WO. Factors affecting menstrual cycle characteristics. *American Journal of Epidemiology* 2004;**160**(2): 131-40.

13. Fries GF. The PBB episode in Michigan: an overall appraisal. *Critical Reviews in Toxicology* 1985;**16**(2): 105-56.

14. Wolff MS, Anderson HA, Selikoff IJ. Human tissue burdens of halogenated aromatic chemicals in Michigan. *Journal of the American Medical Association* 1982;**247**(15): 2112-6.

15. Davis SI, Blanck HM, Hertzberg VS, et al. Menstrual function among women exposed to polybrominated biphenyls: a follow-up prevalence study. *Environmental Health: A Global Access Science Source* 2005;**4**: 15.

16. Turkheimer E, Haley A, Waldron M, D'Onofrio B, Gottesman, II. Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* 2003;**14**(6): 623-8.

17. Rowe DC, Jacobson KC, Van den Oord EJ. Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development* 1999;**70**(5): 1151-62.

18. Harlow SD, Zeger SL. An application of longitudinal methods to the analysis of menstrual diary data. *Journal of Clinical Epidemiology* 1991;**44**(10): 1015-25.

19. Brodin T, Bergh T, Berglund L, Hadziosmanovic N, Holte J. Menstrual cycle length is an age-independent marker of female fertility: results from 6271 treatment cycles of in vitro fertilization. *Fertility and Sterility* 2008;**90**(5): 1656-61.

20. Needham LL, Burse VW, Price HA. Temperature-programmed gas chromatographic determination of polychlorinated and polybrominated biphenyls in serum. *Journal of the Association of Official Analytic Chemists* 1981;**64**(5): 1131-7.

21. Khoury MJ, Beaty TH, Cohen BH. Genetic Approaches to Familial Aggregation (Chapter 7). *Fundamentals of Genetic Epidemiology*. New York: Oxford University Press, Inc.; 1993. p. 200-32.

22. Lange K, Cantor R, Horvath S, et al. Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *American Journal of Human Genetics* 2001;69 (Supplement): A1886.

23. Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 1987;**82**(398): 605-10.

24. Jacobson JL, Fein GG, Jacobson SW, Schwartz PM, Dowler JK. The transfer of polychlorinated biphenyls (PCBs) and polybrominated biphenyls (PBBs) across the human placenta and into maternal milk. *American Journal of Public Health* 1984;**74**(4): 378-9.

25. Bonhaus DW, McCormack KM, Braselton WE, Jr., Hook JB. Effect of polybrominated biphenyls on hepatic microsomal metabolism of estrogens and uterotropic action of administered estrogen in rats. *Journal of Toxicology and Environmental Health* 1981;**8**(1-2): 141-50.

26. Nakari T, Pessala P. In vitro estrogenicity of polybrominated flame retardants. *Aquatic Toxicology* 2005;**74**(3): 272-9.

27. McCormack KM, Arneric SP, Hook JB. Action of exogenously administered steroid hormones following perinatal exposure to polybrominated biphenyls. *Journal of Toxicology & Environmental Health* 1979;**5**(6): 1085-94.

28. Shepherd EC, Phillips TD, Irvin TR, Safe SH, Robertson LW. Aflatoxin B1 metabolism in the rat: polyhalogenated biphenyl enhanced conversion to aflatoxin M1. *Xenobiotica* 1984;**14**(9): 741-50.

29. ATSDR. *Toxicological Profile for Polybrominated Biphenyls and Polybrominated Diphenyl Ethers (PBBs and PBDEs)*. Agency for Toxic Substances and Disease Registry 2004 [Accessed February, 2008]; Available from: http://www.atsdr.cdc.gov/.

30. Harris SJ, Cecil HC, Bitman J. Embryotoxic effects of polybrominated biphenyls (PBB) in rats. *Environmental Health Perspectives* 1978;**23**: 295-300.

31. McCormack KM, Lepper LF, Wilson DM, Hook JB. Biochemical and physiological sequelae to perinatal exposure to polybrominated biphenyls: a multigeneration study in rats. *Toxicology and Applied Pharmacology* 1981;**59**(2): 300-13.

32. Bonhaus DW, McCormack KM, Braselton WE, Jr., Hook JB. Effect of polybrominated biphenyls on hepatic microsomal metabolism of estrogens and uterotropic action of administered estrogen in rats. *Journal of Toxicology & Environmental Health* 1981;**8**(1-2): 141-50.

33. Johnston CA, Demarest KT, McCormack KM, Hook JB, Moore KE. Endocrinological, neurochemical, and anabolic effects of polybrominated biphenyls in male and female rats. *Toxicology and Applied Pharmacology* 1980;**56**(2): 240-7.

34. Allen JR, Barsotti DA, Lambrecht LK, Van Miller JP. Reproductive effects of halogenated aromatic hydrocarbons on nonhuman primates. *Annals of the New York Academy of Sciences* 1979;**320**: 419-25.

35. Lambrecht LK, Barsotti DA, Allen JR. Responses of nonhuman primates to a polybrominated biphenyl mixture. *Environmental Health Perspectives* 1978;**23**: 139-45.

36. Bosetti C, Tavani A, Negri E, Trichopoulos D, La Vecchia C. Reliability of data on medical conditions, menstrual and reproductive history provided by hospital controls. *Acta Obstetricia et Gynecologica Scandinavica* 2001;**54**(9): 902-6.

37. Must A, Phillips SM, Naumova EN, et al. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *American Journal of Epidemiology* 2002;**155**(7): 672-9.

38. Small CM, Manatunga AK, Marcus M. Validity of Self-Reported Menstrual Cycle Length. *Annals of Epidemiology* 2007;**17**(3):163-70.

39. Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *American Journal of Epidemiology* 1979;**109**(2): 181-5.

Chapter 3

Polybrominated Biphenyl Exposure and Reproductive Hormone Levels

3.1 Abstract

Background. Polybrominated biphenyls (PBBs) are persistent halogenated organic chemicals used as fire retardants in plastics, and are potential endocrine-disruptors. Contamination of cattle feed in the 1970's in Michigan resulted in widespread exposure of farming populations to PBBs. The purpose of this study is to determine whether there is an association between serum PBB concentrations and baseline levels of urinary estrogen metabolites and follicle-stimulating hormone (FSH), and whether the route of exposure (dietary vs. *in utero* or breastfeeding) modifies these effects.

Methods. Menstruating women between the ages of 20 and 47 who were not using hormonal contraception were eligible. Urinary levels of E13G (an estrogen metabolite) and FSH were measured on day 3 of each menstrual cycle for 96 women with dietary exposure and 29 women exposed *in utero*. Multivariate linear mixed models examining associations between PBB concentrations and hormone levels were adjusted for age and urinary creatinine levels.

Results. Mean adjusted urinary E13G levels in exposed women (with PBB levels greater than the limit of detection) were approximately 2 ng/mL lower than in the unexposed ($p=0.03$). Mean adjusted urinary FSH levels in exposed women were approximately 2 mIU/mL lower than in the unexposed among women with dietary exposure ($p<0.01$), but not among women exposed *in utero* (p for interaction = 0.008).

Conclusions. These results suggest an effect of exposure to PBBs on both FSH and estrogen metabolite levels. In addition, they suggest that the effects of endocrine disruptors may depend on the route of exposure. Exposure during gestation or infancy may have developmental effects that alter reproductive characteristics throughout the lifespan.

3.2 Introduction

Polybrominated biphenyls (PBBs) are brominated hydrocarbon compounds that were first produced commercially in 1970, primarily for use as a fire retardant in plastics. PBBs, along with other halogenated hydrocarbons such as polychlorinated biphenyls (PCBs), and polybrominated diphenyl ethers (PBDEs) are potential "endocrine disrupters," and the question of whether these compounds affect the reproductive health of humans has gained a great deal of attention in both the popular press and scientific journals (**1-4**). There is accumulating evidence that this family of structurally similar halogenated organics causes endocrine disruption in humans and wildlife (**5, 6**). These chemicals are widespread in occurrence, resistant to degradation, and accumulate in fatty tissue.

It is estimated that 13.3 million pounds of PBBs were produced in the U.S. until all production ceased in 1979 due to concerns about its toxicity (**7**). In contrast to other halogenated organics such as PBDE's, exposure of the general population to PBBs has been minimal, with serum levels typically less than 1 part per billion. However, in animal experiments and *in vitro* studies, PBBs have demonstrated toxic effects similar to those of other more prevalent exposures such as PCBs, PCDDs and PCDFs; therefore it can be argued that PBBs can be used as a model to study this family of compounds. Like the other halogenated organics, PBBs bind to the aryl hydrocarbon (Ah) receptor, a cytosolic molecular receptor believed to mediate many of the toxic effects of this class of chemicals (**7**). Many halogenated organics (e.g., DDT, DDE, some PCB congeners) have tested positive in *in vitro* estrogenicity screening, and others (e.g., other PCB congeners, a commercial PBB mixture, and TCDD) have been found to be anti-estrogenic (**8-13**).

For congeners of PCBs, these effects appear to be produced by inhibiting the effects of natural ligands on their receptors (**14**).

Differential route and timing of exposure to PBBs may modify any effects of these chemicals on health outcomes. While the effect of an adult dietary exposure may be reversible as the exposure is removed from the body, *in utero* and lactational exposures to endocrine disruptors may have developmental effects, especially on hormonally sensitive systems such as the urogenital tract and the nervous system (**15**). In rats, oral exposure to PBBs has been associated with acceleration of steroid sex hormone metabolism (**16**), reduction of adrenal cortex hormone serum levels (**17**), alterations in plasma prolactin levels (**18**), and lengthening of estrous cycles (**19**). Animal research on lactational or *in utero* exposures to endocrine disruptors has demonstrated adverse reproductive effects. Female pup rats whose mothers were exposed to 250 ng/kg bodyweight of PCB126 (resulting in lipid PCB levels in the pups of about 23 times that of the control group) had significantly delayed vaginal openings, reduction of serum levels of estrogen and progesterone metabolites, fewer antral follicles, and more atretic follicles when compared with the control group (**20**).

In humans, significant dietary exposure to PBBs resulted from widespread contamination of cattle-feed in Michigan in the early 1970's, resulting in high exposure to farming families and nearby residents. A cohort was formed in the mid-1970's, and multiple generations have been followed and investigated for effects of PBB exposure on various health outcomes. Supporting the hypothesis that PBBs have endocrine-disrupting properties, daughters exposed to PBBs *in utero* and through lactation had earlier

menarche than unexposed daughters (11.6 years for those who were breastfed and were exposed to high levels *in utero*; 12.2-12.7 years for those exposed to lower levels *in utero* or not breastfed) (**21**). Therefore we were motivated to further investigate the effects of PBBs on reproductive characteristics of both first generation women (who were exposed through diet) and second generation women (who were exposed *in utero* and/or through breastfeeding).

There are very few studies examining the association between halogenated organic chemicals and hormone levels in humans. The baseline level of follicle-stimulating hormone, measured on day 3 of the menstrual cycle, is commonly accepted as an indicator of ovarian aging and ovarian reserve (**22**), and day 3 estradiol levels have also been used as an indicator of IVF success and overall ovarian function (**23, 24**). In this study, we investigate the association between serum PBB levels with day 3 levels of FSH and an estrogen metabolite, E13G. Furthermore, we determine whether these effects are modified by exposure route (dietary vs. *in utero*/breastfeeding) by examining effects in both first-generation and second-generation women. By determining whether there is an association between serum PBBs and these intermediate hormonal endpoints, we examine a possible causal pathway by which PBB may alter reproductive outcomes.

3.3 Methods

3.3.1 Population

In 1976-77, the Michigan Department of Community Health (MDCH) organized nearly 4000 individuals potentially exposed to PBB into a registry and obtained baseline

health information as well as serum samples. Since the initial enrollment period, the MDCH has been contacting cohort members periodically to update their files and to enroll children born to exposed mothers. In 2003-06, 479 women completed a phone interview including questions on reproductive health. During this interview, women were asked to participate in a prospective study of menstrual function. Women who were between the ages of 18 and 46, not using oral contraceptives, and not pregnant or lactating were eligible for this study. During the menstrual function study, women collected first morning daily urines and completed daily diaries with information on menstrual bleeding and covariates for up to 7 menstrual cycles.

3.3.2 Exposure Assessment

Serum samples collected at enrollment into the registry were assayed for PBBs. Quantification of PBBs was based on the main congener, 2,2', 4,4', 5,5' hexa-bromobiphenyl. PBB was measured as Aroclor 1254. The methods of PBB detection used have coefficients of variation of 7.1-14.0%, recovery ranges of 80-90% (**25**) and a level of detection of 1 part per billion (ppb). The mean serum PBB level in the registry participants at enrollment was 21 ppb (range, 1 to 2260 ppb; median, 3 ppb), compared to a mean of less than 1 ppb for the general U.S. population. Exposures less than or equal to 1 ppb (the limit of detection, or LOD) were categorized as "low"; levels between 1 and 4 ppb as "medium"; and levels greater than or equal to 4 ppb as "high".

Our research group previously developed and validated a decay model for serum PBB levels (**26, 27**). The decay model was used to estimate serum PBB levels in women at the time of pregnancy, thereby creating a measure of *in utero* exposure for the

daughters. Because the timing and route of exposure (dietary vs. *in utero*/lactational) may be crucial modifiers of PBB's effects on the reproductive system, and because the methods of estimating exposure differ, first-generation women and daughters were considered separately in this study and were only combined if it was deemed that the effects were independent of exposure route (i.e. there was no interaction between PBB exposure and generation in multivariate models).

3.3.3 Covariate Assessment

Information on covariates such as age, reproductive history, body mass index (BMI, in kg/m^2) and age at menarche was obtained from the 2003-06 telephone interview. Information on time-dependent covariates such as smoking, alcohol intake, intercourse, menstrual cycle information, and stress was obtained from the daily diaries. For time-dependent exposures such as smoking and alcohol, exposure information from the current cycle and the previous cycle were both assessed as potential confounders.

3.3.4 Assessment of Urinary Hormone Levels

First morning urine samples (5-10 mL) were collected daily for up to eight consecutive menstrual cycles. Urine samples were stored in the participants' home freezers in vials containing 7% glycerol to prevent freeze-induced activity loss of FSH (**28**). Samples were shipped in dry ice by express courier to the laboratory and were stored frozen at -80°C until assayed for urinary ovarian hormone metabolites. Urinary FSH was assayed in duplicate using non-competitive, two-site time-resolved immunofluorometric assays (**29, 30**). E13G was assayed in triplicate using competitive, double-antibody time-resolved fluoroimmunoassays (**31**). Urinary creatinine

concentrations were measured using a Vitros 250 Chemistry System (Ortho-Clinical Diagnostics). The Vitros system employs a slide comprised of a dry, multilayered analytical element coated on a polyester support.

Urine samples from day 3 of each menstrual cycle (where Day 1 is the start of menses) were used in this study. Adjacent days were utilized if day 3 was not available. Menses was defined as at least two consecutive days of bleeding, where at least one day was not defined as "spotting." The first day of bleeding that was not spotting was defined as Day 1 of the menstrual cycle. Pregnant cycles were identified by testing levels of hCG on days -1 and -2 the next menstrual cycle, and were excluded if hCG levels exceeded 3 ng/mL. The Immulite® 2000 hCG assay from Siemens (Los Angeles, CA) was used to measure hCG levels in duplicate.

3.3.5 Analyses

Distributions of urinary hormone levels were examined for normality and transformed if necessary. Multivariate mixed linear regression models were used to investigate the associations between PBB exposure (in 3 categories, as defined above) with day 3 E13G and FSH. A random intercept for each woman was included to account for correlation of hormone levels within a woman across cycles and to account for variability across women. Separate models were initially run for first- and second-generation women. Potentially confounding variables were those known to be associated with hormone levels or reproductive outcomes such as parity, age at menarche, age, body mass index, smoking, alcohol, caffeine consumption, and stress. Because some of these covariates may in fact be in the causal pathway between PBB exposure and menstrual

function, models were run both with and without their inclusion (except age, which was always included). Models were also run that combined both generations and included an indicator variable for generation. A product term between generation and PBB category was included in multi-generational models to test the interaction between these variables, to determine whether any effect of PBBs on hormone levels is modified by exposure route.

3.4 Results

Of the 479 women who completed the phone interview, 314 (66%) were eligible for the menstrual function study (Figure 3.1). Major reasons for ineligibility included oral contraceptive use (n=94) and pregnancy or breastfeeding (n=25). Of the 314 eligible women, 219 (70%) initially agreed to participate. However, only 134 (61%) of these women returned at least one diary card and one urine sample. In addition, only 125 women returned eligible urine samples (collected on days 1-4 of the menstrual cycle, and not directly after a cycle in which pregnancy occurred). These 125 women were included in the analysis, and consisted of 96 first-generation women and 29 second-generation women. These women were followed up to 7 menstrual cycles (median = 5). Two menstrual cycles were excluded because of high hCG levels, indicating an early pregnancy loss. After exclusions, 533 eligible menstrual cycles were available for analysis.

All women reported that they were white. Two first-generation women and one second-generation woman also indicated that they were Native American. Other demographic and follow-up characteristics are shown in Table 3.1. First generation

women were, on average, about 12 years older than second-generation women (median ages: 39 and 27, respectively). First-generation women had higher median PBB levels than second-generation women. Only eleven of the second-generation women had their own serum PBB level measured, and seven of these women had levels below the limit of detection (≤1 ppb). However, all 29 of these women's mothers had their PBB levels measured; these values could be used to estimate the daughters' exposure *in utero* or through breastfeeding.

Women in both generations were mostly of normal BMI or overweight, and about three-fourths were nonsmokers. Urinary levels of E13G and FSH were measured on day 3 of the menstrual cycle for 521 menstrual cycles; day 1 for two cycles; day 2 for nine cycles, and day 2 for two cycles, where day 3 samples were not available. First generation women had higher median FSH levels than second generation women, which was expected because first generation women were older. First generation women also had higher median E13G levels; this also may be expected because FSH stimulates the ovaries to produce estrogen.

Day 3 urinary FSH and E13G levels were not normally distributed. After transforming each of these variables by taking the natural logarithm, the distributions appeared normal and could be modeled as continuous outcomes in mixed linear regression models. Initial multivariate models included BMI, parity, smoking (cigarettes/day), alcohol use (drinks/day), caffeine (drinks/day), and stress (on a scale from 1-4) as covariates. However, it was determined that none of these biological characteristics or lifestyle factors confounded the association between PBB exposure and hormone levels, and they were removed from the multivariate models. This was true

whether time-varying covariate information was taken from the current cycle or the previous cycle. Final models included the categorical PBB predictor variables, age, and urinary creatinine levels as a covariate.

There was an inverse association between serum PBB levels and urinary E13G levels in this population (Table 3.2 and Figure 3.2). For ease of interpretation, mean values of urinary E13G levels are shown by PBB category in Table 3.2. Mean values are shown for women of the median age (39 for 1st generation women and 27 for 2nd generation women) and median creatinine level (122 for 1st generation women and 136 for 2nd generation women). Unexposed women had mean E13G levels approximately 2 ng/mL higher than women in either of the exposed categories. This was true for both generations; when these generations were combined into a single model, the association reached statistical significance (p-value for the Score test for the association between PBB category and E13G levels = 0.03). There was no significant interaction between generation and PBB category in this combined model (p for interaction= 0.39).

There was a significant inverse association between serum PBB levels and urinary FSH levels in the first generation women, but not the second generation women (Table 3.2 and Figure 3.3). Again, mean urinary FSH levels are shown for women of the median age and creatinine level for each generation. When the two generations were combined into a single model that included an interaction term for generation and PBB category and another interaction term between generation and age to account for the increasing effect of age on FSH levels as age increases, the interaction between generation and PBB category was highly significant (p=0.008).

3.5 Discussion

In this small study, we observed a significant effect of exposure to PBBs, a potential endocrine disruptor, on urinary estrogen metabolite levels. This effect was observed independently in two separate samples of women (first- and second-generation women). The direction of the effect was anti-estrogenic. There was also a significant inverse effect of PBBs on urinary FSH levels, but only among first-generation women, who were exposed to PBBs through diet.

There are few studies of the effects of organic chemicals on reproductive hormone levels in humans, but those that exist have shown both estrogenic and anti-estrogenic effects. A recent prospective study examining the association between DDT and its isomers and metabolites with measures of urinary prenanediol-3-glucuronide (PdG) and estrone conjugate (E_1C) levels found that increased serum levels of DDT were associated with decreased levels of PdG and E_1C (**32**). Among 17 daughters of PCB and PCDF-exposed women from the Yu-Cheng incident in Taiwan, day 3 serum levels of estradiol ($p=0.02$) and FSH ($p=0.06$) were higher in exposed girls compared to 16 controls (**33**). It is not surprising that different halogenated organic compounds have seemingly conflicting effects on reproductive characteristics; PCBs and PBBs are both a mixture of estrogenic and anti-estrogenic congeners that may be present in varying amounts and may have tissue-specific effects.

The E13G levels for exposed women in this study were approximately 2 ng/mL lower, on average, than those of unexposed women (PBB$\leq$1 ppb). In addition, for first-generation women, the FSH levels of exposed women were approximately 2 mIU/mL

lower than unexposed women. These differences amount to approximately a quartile, and was also approximately the difference that was observed when comparing the two generations. Although these associations achieved statistical significance, the question remains as to whether the differences are clinically meaningful and would result in an impairment of reproductive function or quality of life. At the very least, these results should be considered a warning that halogenated organic compounds can potentially alter reproductive hormone levels; effects of more common compounds (e.g., PBDE's) may be more or less pronounced.

PBB exposure was associated with FSH levels in first-generation but not second-generation women. One possible explanation for this is that PBB exposure *in utero* or through breastfeeding did not have long-term consequences on the production or metabolism of FSH, whereas dietary exposure affected FSH levels more directly. The serum PBB levels in the second-generation women were much lower than in first-generation women. Another explanation is that FSH naturally fluctuates more in older women, and therefore there may be a greater opportunity for endocrine-disrupting compounds to disturb the hormone levels.

A major limitation of this study was the small number of women present in some categories of PBB exposure. For example, there were only 9 unexposed women in the second generation. Individual characteristics of these women may have influenced the associations observed. Selection bias could have also affected the results. Women in the PBB cohort are aware of their own PBB serum levels. Bias could occur if women highly exposed to PBBs were only motivated to participate if they suspected a disruption in their

reproductive function, for example. Confounding was unlikely to cause the observed associations, since PBB exposure was likely independent of confounding factors in the first-generation women. If, however, women with very high exposure to PBBs also share a particular diet that affects reproductive function, the diet (or other lifestyle factors) could have been at least partially responsible for the observed association. However, PBB exposure would have preceded any other factors in the second-generation women, and so it could be argued that any reproductive or lifestyle factors associated with PBB exposure in these women are in the causal pathway between PBB exposure and reproductive hormone levels, and therefore should not confound the association.

In summary, this study supports the hypothesis that PBBs act in an endocrine-disrupting manner. The results suggest an urgent need to investigate effects of other halogenated organic chemicals on hormone levels and other aspects of reproductive function, particularly as more of these chemicals accumulate in the environment. It also demonstrates that the effects of these chemicals may depend on the route of exposure, and that exposure to these chemicals *in utero* may have lifelong consequences. The results of this study could have implications for policy and environmental regulations to improve public health, by providing crucial evidence needed to demonstrate the mechanism by which endocrine disruptors affect reproductive health outcomes that threaten both the survival and well-being of animals and humans.

3.6 Acknowledgments

We thank the staffs of the Michigan Department of Community Health and Michigan Public Health Institute for their assistance in data collection.

This manuscript was developed under a STAR Research Assistance Agreement No. 916891 awarded by the U.S. Environmental Protection Agency. It has **not** been formally reviewed by the EPA. The views expressed in this manuscript are solely those of Kira Creswell Taylor and the EPA **does not endorse** any products or commercial services mentioned in this manuscript.

Funding for this research was also provided by the U.S. Environmental Protection Agency (R 825300), National Institute of Environmental Health Sciences (RO1 ES08341 and RO1 ES012014), and by Centers for Disease Control and Prevention Cooperative Agreement U37/CCU500392.

The Institutional Review Boards at Emory University (691-2002) and the Michigan Department of Community Health (58-PHAEPI(R)) have approved this study.

3.7 Figures and Tables

Table 3.1 Demographic and Follow-Up characteristics of Women in the PBB Long-Term Cohort Menstrual Function Study

	1st generation N=96	2nd generation N=29
Total menstrual cycles of observation	425	111
Age at interview		
Mean±SD	38.7±3.6	26.4±3.8
20-29	0 (0)[a]	21 (72)
30-39	59 (61)	8 (28)
40-47	37 (39)	0 (0)
Serum PBB levels (ppb)		
Median, IQR	2.8 (2.0, 5.6)	0.5 (0.5, 1.9)
≤1 (the LOD)	20 (21)	7 (64)
1.1-3.9	38 (40)	2 (18)
≥4	38 (40)	2 (18)
Not measured	--	18 (62)
Mother's estimated PBB levels at conception (ppb)		
Median, IQR	--	2.1 (1.0, 4.1)
≤1 (the LOD)	--	9 (31)
1.1-3.9	--	11 (38)
≥4	--	9 (31)
Route of PBB exposure		
Diet	96 (100)	--
In utero only	--	9 (31)
In utero and breastfed	--	18 (62)
Breastfeeding status unknown	--	2 (7)
BMI (kg/m^2)		
<18.5	1 (1)	2 (7)
18.5-25	43 (45)	14 (48)
25-30	33 (34)	9 (31)
>30	20 (21)	4 (14)

Smoking		
Never	71 (74)	22 (76)
Ex-smoker	12 (13)	2 (7)
Current	13 (14)	5 (17)
Day 3 Urinary E13G (ng/mg creatinine) (median, IQR)	8.3 (6.0, 10.7)	6.0 (4.2, 7.6)
Day 3 Urinary FSH (mIU/mg creatinine) (median, IQR)	4.9 (4.0, 7.2)	3.8 (2.6, 5.2)

[a]Values are N (%) unless otherwise indicated.

Table 3.2 Mixed Linear Model Results: Mean Day 3 E13G and FSH Levels According to PBB Exposure Categories

Generation	PBB exposure[a] (ppb)	N women	N cycles	Mean day 3 E13G (ng/mL) (95% CI)[b]	Mean day 3 FSH (mIU/mL) (95% CI)
1	≤1	20	95	9.9 (7.9, 12.4)	7.2 (6.1, 8.6)
	1.1-3.9	38	155	7.3 (6.2, 8.6)	5.2 (4.6, 5.9)
	≥4	38	175	8.4 (7.2, 9.9)	5.1 (4.5, 5.7)
	Score test p-value[c]			0.11	<0.01
2	≤1	9	36	8.1 (6.4, 10.2)	3.2 (2.3, 4.4)
	1.1-3.9	11	39	5.8 (4.6, 7.2)	4.5 (3.3, 6.1)
	≥4	9	36	6.4 (5.1, 8.0)	3.9 (2.8, 5.5)
	Score test p-value			0.13	0.39

[a]For the second-generation women, PBB level refers to the mother's estimated serum PBB level at conception.
[b]Means and 95% CI are adjusted for correlation across cycles within women, age, and urinary creatinine, and are presented for the median age and creatinine level for each generation.
[c]For the effect of the categorical PBB variable on the outcome.

Figure 3.1 Population flowchart for Participation in the Menstrual Function Study

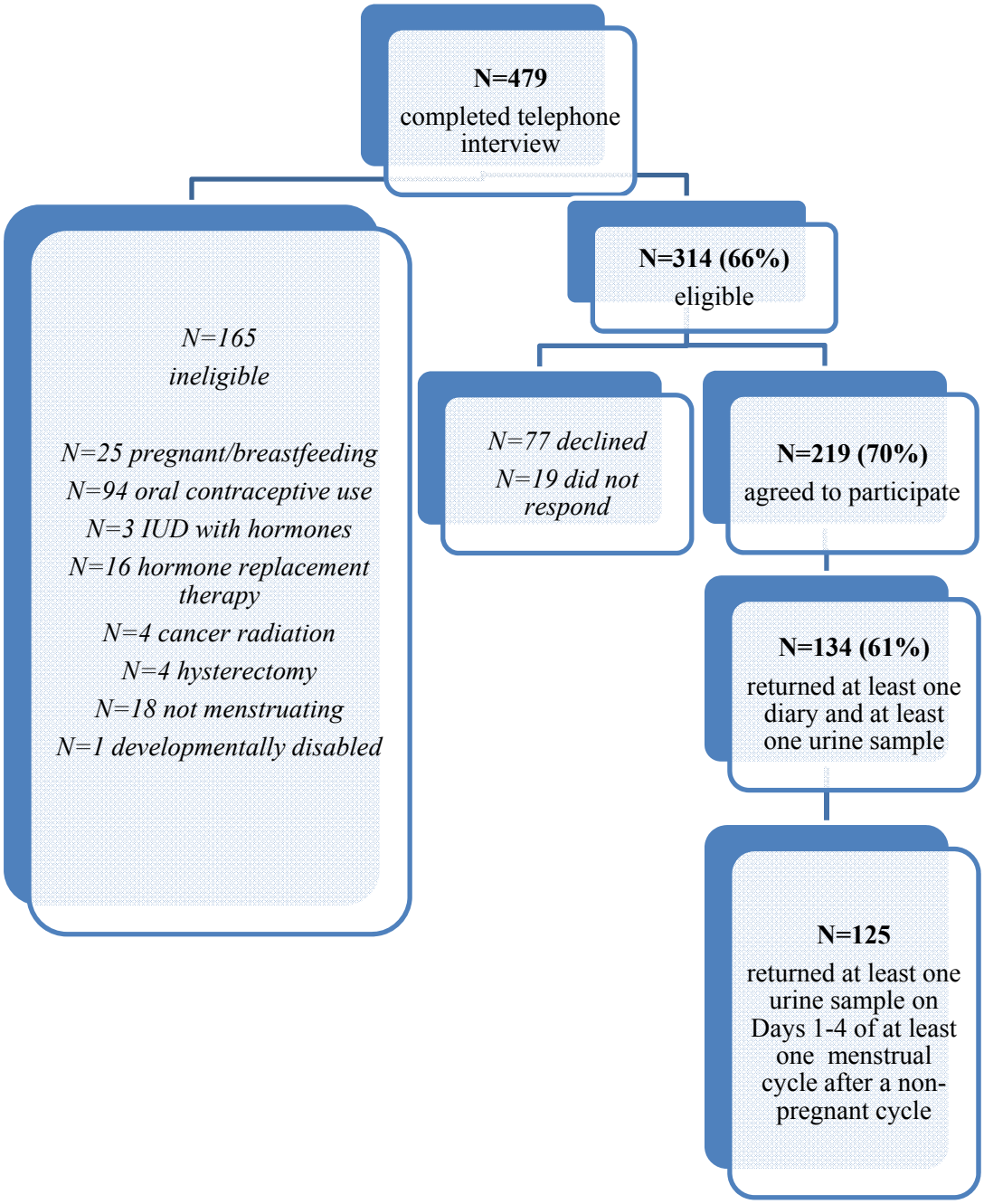


Figure 3.2 Mean Day 3 Estrogen Metabolite Levels and 95% Confidence Intervals for First and Second-Generation Women, Presented for the Median Age and Creatinine Levels for Each Generation

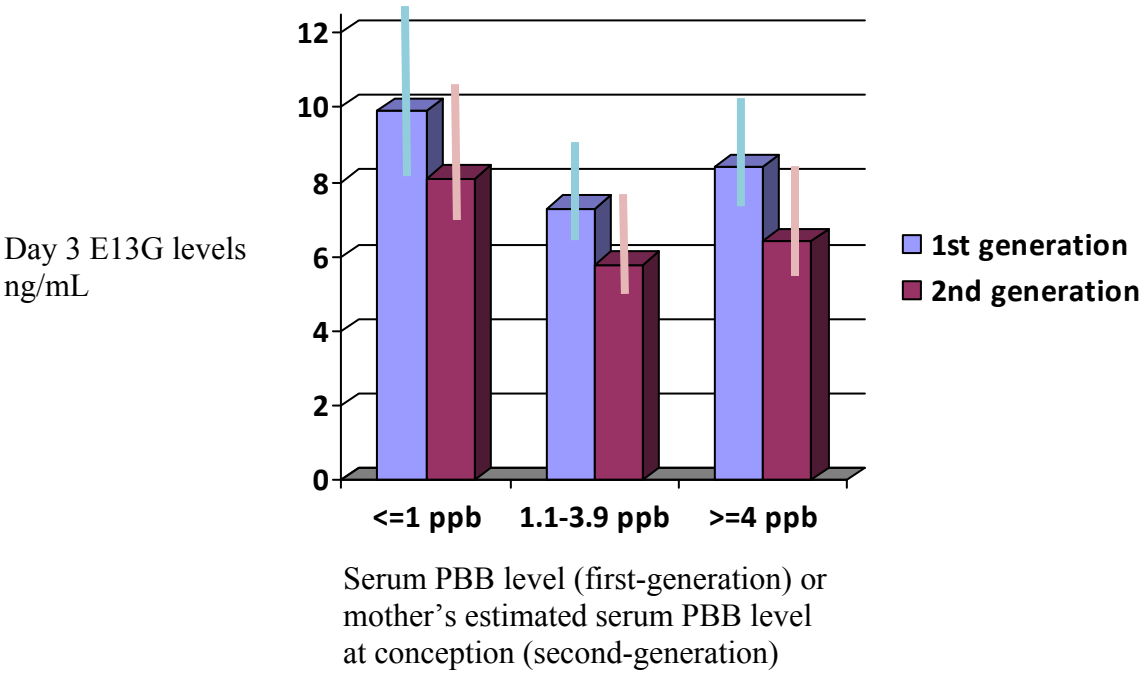
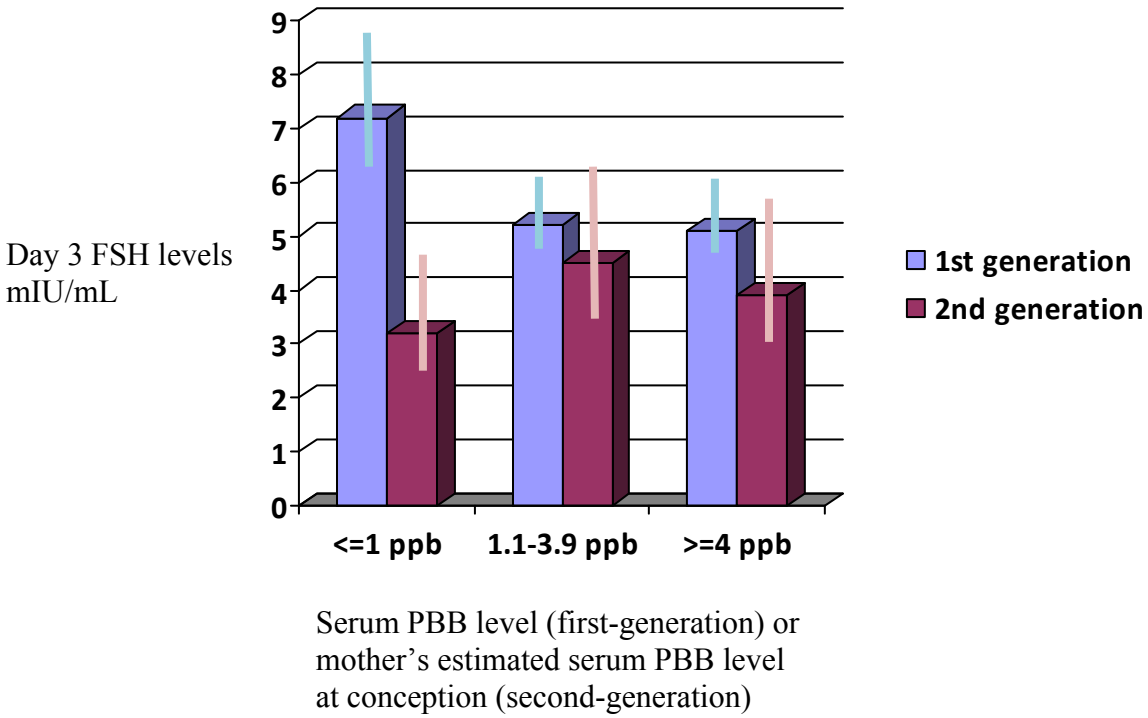


Figure 3.3 Mean Day 3 FSH Levels and 95% Confidence Intervals for First and Second-Generation Women, Presented for the Median Age and Creatinine Levels for Each Generation



3.8 References

1. Sharpe RM, Skakkeback NE. Are estrogens involved in falling sperm counts and disorders of the male reproductive tract? *Lancet* 1993;**341**: 1392-5.

2. Sharpe RM, Turner KJ, Sumpter JP. Endocrine disruptors and testis development. *Environmental Health Perspectives* 1998;**106**(5): A220-1.

3. Hauser R, Chen Z, Pothier L, Ryan L, Altshul L. The relationship between human semen parameters and environmental exposure to polychlorinated biphenyls and p,p'-DDE. *Environmental Health Perspectives* 2003;**111**(12): 1505-11.

4. Buck Louis GM, Lynch CD, Cooney MA. Environmental influences on female fecundity and fertility. *Seminars in Reproductive Medicine* 2006;**24**(3): 147-55.

5. Colborn T, vom Saal FS, Soto AM. Developmental effects of endocrine-disrupting chemicals in wildlife and humans. *Environmental Health Perspectives* 1993;**101**(5): 378-84.

6. Rasier G, Toppari J, Parent AS, Bourguignon JP. Female sexual maturation and reproduction after prepubertal exposure to estrogens and endocrine disrupting chemicals: A review of rodent and human data. *Molecular and Cellular Endocrinology* 2006;**254-255**: 187-201.

7. ATSDR. Toxicological Profile for Polybrominated Biphenyls and Polybrominated Diphenyl Ethers (PBBs and PBDEs). Agency for Toxic Substances and Disease Registry 2004 [Accessed February, 2008]; Available from: http://www.atsdr.cdc.gov/.

8. Bonhaus DW, McCormack KM, Braselton WEJ, Hook JB. Effect of polybrominated biphenyls on hepatic microsomal metabolism of estrogens and uterotropic action of administered estrogen in rats. . *Journal of Toxicology and Environmental Health* 1981;**8**: 141-50.

9. Bulger WH, Kupfer D. Estrogenic Action of DDT analogs. *American Journal of Industrial Medicine* 1983;**4**(1-2): 163-73.

10. Cooke PS, Sato T, Buchanan DL. Disruption of steroid hormone signalling by PCBs. In: Robertson LWH, L.G., editor. *PCBs: Recent Advances in Environmental Toxicology and Health Effects*. Lexington: The University Press of Kentucky; 2001. p. 257-63.

11. Korach KS, Sarver P, Chae K, McLachlan JA, McKinney JD. Estrogen receptor-binding activity of polychlorinated hydroxybiphenyls: conformationally restricted structural probes. *Molecular Pharmacology* 1988;**33**(1): 120-6.

12. McCormack KM, Melrose P, Rickert DE, Dent JG, Gibson JE, Hook JB. Concomitant dietary exposure to polychlorinated biphenyls and polybrominated biphenyls: Tissue distribution and arylhydrocarbon hydroxylase activity in lactating rats. *Toxicology and Applied Pharmacology* 1979;**47**(1): 95-104.

13. Wolff MS, Toniolo PG. Environmental Organochlorine Exposure as a Potential Etiologic Factor in Breast Cancer. *Environmental Health Perspectives* 1995;**103**(Supp. 7): 141-5.

14. Bonefeld-Jorgensen EC, Andersen HR, Rasmussen TH, Vinggaard AM. Effect of highly bioaccumulated polychlorinated biphenyl congeners on estrogen and androgen receptor activity. *Toxicology* 2001;**158**(3): 141-53.

15. Singleton DW, Khan SA. Xenoestrogen exposure and mechanisms of endocrine disruption. *Frontiers in Bioscience* 2003;**8**(s): 110-8.

16. McCormack KM, Arneric SP, Hook JB. Action of exogenously administered steroid hormones following perinatal exposure to polybrominated biphenyls. *Journal of Toxicology & Environmental Health* 1979;**5**(6): 1085-94.

17. Dent JG. Characteristics of cytochrome P-450 and mixed function oxidase enzymes following treatment with PBBs. *Environmental Health Perspectives* 1978;**23**: 301-7.

18. McCormack KM, Melrose P, Rickert DE, Dent JG, Gibson JE, Hook JB. Concomitant dietary exposure to polychlorinated biphenyls and polybrominated biphenyls: tissue distribution and arylhydrocarbon hydroxylase activity in lactating rats. *Toxicology & Applied Pharmacology* 1979;**47**(1): 95-104.

19. Johnston CA, Demarest KT, McCormack KM, Hook JB, Moore KE. Endocrinological, neurochemical, and anabolic effects of polybrominated biphenyls in male and female rats. *Toxicology & Applied Pharmacology* 1980;**56**(2): 240-7.

20. Muto T, Imano N, Nakaaki K, et al. Estrous cyclicity and ovarian follicles in female rats after prenatal exposure to 3,3',4,4',5-pentachlorobiphenyl. *Toxicology Letters* 2003;**143**(3): 271-7.

21. Blanck HM, Marcus M, Tolbert PE, et al. Age at menarche and tanner stage in girls exposed in utero and postnatally to polybrominated biphenyl. *Epidemiology* 2000;**11**(6): 641-7.

22. Scott R, Toner J, Muasher S. Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. *Fertility and Sterility* 1989;**51**: 651-54.

23. Licciardi FL, Liu HC, Rosenwaks Z. Day 3 estradiol serum concentrations as prognosticators of ovarian stimulation response and pregnancy outcome in patients undergoing in vitro fertilization. *Fertility and Sterility* 1996;**66**(4): 991-4.

24. Smotrich DB, Widra EA, Gindoff PR, Levy MJ, Hall JL, Stillman RJ. Prognostic value of day 3 estradiol on in vitro fertilization outcome. *Fertility and Sterility* 1995;**64**(6): 1136-40.

25. Needham LL, Burse VW, Price HA. Temperature-programmed gas chromatographic determination of polychlorinated and polybrominated biphenyls in serum. *Journal of the Association of Official Analytic Chemists* 1981;**64**(5): 1131-7.

26. Blanck HM, Marcus M, Hertzberg V, et al. Determinants of polybrominated biphenyl serum decay among women in the Michigan PBB cohort. *Environmental Health Perspectives* 2000;**108**(2): 147-52.

27. Terrell ML, Manatunga AK, Small CM. A Decay Model for Assessing Polybrominated Biphenyl Exposure Among Women in the Michigan Long-Term PBB Study. *Journal of Exposure Science and Environmental Epidemiology* 2008; 18(4): 410-420.

28. Kesner JS, Knecht EA, Krieg EF, Jr. Stability of urinary biomarkers of female reproductive hormones stored under various conditions. *Reproductive Toxicology* 1995;**9**: 239-244.

29. Kesner JS, Knecht EA, Krieg EF, Jr. Time-resolved immunofluorometric assays for urinary lutenizing hormone and follicle stimulating hormone. *Analytica Chimica Acta* 1994;**285**: 13-22.

30. Kesner JS, Knecht EA, Krieg EF, Jr., Wilcox AJ, O'Connor JF. Detecting pre-ovulatory luteinizing hormone surges in urine. *Human Reproduction* 1998;**13**(1): 15-21.

31. Kesner JS, Knetch EA, Krieg EF, Jr., et al. Validation of time-resolved fluoroimmunoassays for urinary estrone 3-glucuronide and pregnanediol 3-glucuronide. *Steroids* 1994;**58**: 205-11.

32. Perry MJ, Ouyang F, Korrick SA, et al. A Prospective Study of Serum DDT and Progesterone and Estrogen Levels across the Menstrual Cycle in Nulliparous

Women of Reproductive Age. *American Journal of Epidemiology* 2006;**164**(11): 1056-64.

33. Yang CY, Yu ML, Guo HR, et al. The endocrine and reproductive function of the female Yucheng adolescents prenatally exposed to PCBs/PCDFs. *Chemosphere* 2005;**61**(3): 355-60.

Chapter 4

Inhibin Polymorphisms and Menstrual Cycle Characteristics

4.1 Abstract

Background. Menstrual cycle length and menstrual cycle variability are interrelated outcomes, connected by their dependence on ovulation and the corresponding hormonal milieu surrounding ovulation. Menstrual dysfunction is associated with adverse outcomes such as infertility. Inhibins α, β_A and β_B (*INHA*, *INHBA* and *INHBB)* are three related genes that are integral to the process of ovulation. The purpose of this study is to saturate the known genetic variability in these three genes to determine whether there is an association between these genes and menstrual function.

Methods. 470 women office workers participated in a prospective study of fertility and menstrual function and recorded menstrual bleeding and covariates such as smoking and intercourse in daily diaries for 1-19 cycles (median=7 cycles). First-morning urine was used as the DNA source. Women were genotyped for linkage disequilibrium-tagging single nucleotide polymorphisms (SNPs) across the extended gene regions of *INHA, INHBA*, and *INHBB*. Linear regression was used to model the effect of all SNPs on cycle length, whereas logistic regression was used to determine any associations with the probability of having highly variable cycles. Least-squares kernel machines (LSKM) were also used to examine the effects of the global variation in each gene on these outcomes. SNP-SNP interaction was examined among SNPs that were individually significantly associated with cycle characteristics.

Results. One SNP in *INHA*, four in *INHBA*, and one in *INHBB* were associated with cycle length, and one SNP in *INHA* was associated with cycle variability, but none of these associations reached statistical significance after accounting for multiple comparisons. LSKM analyses did not reveal any further associations. There was no evidence of interaction among these SNPs.

Conclusion. The results of this study do not generally support the hypothesis that variation in the inhibin genes is associated with menstrual function. However, all SNPs in this study were either intronic or in regulatory regions. Resequencing these genes could reveal functional polymorphisms or other rare variants that have more influence on menstrual function.

4.2 Introduction

The inhibins and activins are proteins produced by ovarian granulosa cells that are essential to ovarian function. These glycoprotein dimers are encoded by three genes, inhibin alpha (*INHA*), inhibin beta A (*INHBA*), and inhibin beta B *(INHBB*). Each beta subunit can combine with the alpha subunit to form heterodimer proteins known as inhibin A ($\alpha\beta_A$) and B ($\alpha\beta_B$). The beta subunits themselves also form heterodimers and homodimers, known as activins AB ($\beta_A\beta_B$), A ($\beta_A\beta_A$), and B ($\beta_B\beta_B$) (**1**). Based on their sequences, the beta subunits belong to the TGF-β superfamily of growth and differentiation factors (**2**).

Inhibins and activins play essential roles in follicular development and the regulation of ovarian steroidogenesis. Tight regulation of follicular development is important to ensure adequate oocyte quality and limit multiple ovulations. In this controlled system, inhibins exert negative feedback on the gene expression and pituitary secretion of follicle stimulating hormone (FSH) while activins stimulate FSH release. Human and animal evidence of this feedback are reviewed in Messinis (**3**).

Inhibin and activin protein levels are used as markers of ovarian function. Decreases in inhibins and corresponding increases in FSH mark normal reproductive aging (**4-6**), including shortened follicular phases (**7**), diminished ovarian reserve, and changes in corpus luteum function (**8, 9**). As menopause approaches, inhibin B levels are closely associated with changes in menstrual cycle length (**10**).

Beyond their endocrine role, the inhibins and activins may directly influence follicular development. Human *in vitro* and animal evidence suggest that activins promote granulosa cell proliferation and differentiation, prevent premature luteinization, and may cause preantral follicles to remain dormant (**11-17**). In addition, activins may promote FSH stimulated production of estrogen by increasing FSH receptor expression on granulosa cells (**18, 19**) and may increase androgen production by stimulating thecal cell proliferation (**20**). Animal models suggest that inhibins promote the growth and differentiation of follicles (**15, 21, 22**) and may influence the final stages of follicular development (**23**). Should genetic variability influence inhibin or activin levels or activity, the effects on folliculogenesis may be manifest in menstrual cycle characteristics.

Few studies have examined the relationship between genetic variability in the inhibin genes and reproductive function. Several previous studies have found associations between polymorphisms in *INHA* (769G→A and -16C→T) and premature ovarian failure (POF) (**24-28**). Additionally, a polymorphism in the inhibin-binding domain of betaglycan/*TGF-βRIII* was significantly associated with POF in a population of women from New Zealand (**29**). In sheep, genetic variation at the *INHBA* locus was associated with litter size (**30**). This effect of *INHBA* on litter size may be a result of increased ovulation, because inhibin-immunized ewes displayed a 90% increase in ovulation rate (**31**).

No studies have yet investigated the relationship between these three genes and normal variation in reproductive function in humans. In this study, we investigate whether

there are associations between polymorphisms in *INHA, INHBA* and *INHBB* and menstrual cycle length and variability in a population of 470 women office workers. Because of the interdependence of these three subunits, it is plausible that the effect of a polymorphism in one gene may depend on polymorphisms present in the others. The protein levels also interact in the sense that the opposing effects of activins and inhibins are not additive--the highest concentrations of activins completely abolish the effects of inhibins (**32**). Therefore, SNP-SNP interactions are also investigated.

4.3 Methods

4.3.1 Population

This study takes advantage of existing data and stored urine samples from the Mount Sinai Study of Women Office Workers (MSSWOW), a prospective study of menstrual function and fertility (R01 HD24618). Women from 14 companies and government agencies in New York, New Jersey and Massachusetts were enrolled from 1990 through 1994. 4,640 women completed self-administered questionnaires including questions on work practice, musculoskeletal symptoms and psychosocial stress, as well as questions on reproductive health and current birth control practices (**33**). Women were eligible for the prospective study of fertility and spontaneous abortion if they had been sexually active in the month prior to completing the questionnaire while using inconsistent or no birth control, and were age 18-40 (N=855). The study required the completion of daily diaries and urine collection at the onset of each cycle for at least two days. Daily diaries included menstrual cycle information as well as covariates such as intercourse, smoking, caffeine, and alcohol. The urine samples were the source of DNA

for this study. Women were asked to participate for one year or until the end of a clinical pregnancy. 563 women agreed to participate. Of these, 14 became newly ineligible at the time of the intake interview. Of the remaining 549 women, 79 did not collect any urine. This left 470 women for the prospective study of menstrual function and fertility (Figure 4.1).

4.3.2 Selection of Single Nucleotide Polymorphisms

The three inhibin/activin genes are fairly small. The smallest is *INHA*, which maps to 2q33-36, contains 2 exons and spans 3.5 kilobases. The largest and most variable is *INHBA,* which maps to 7p15-p13, contains 3 exons and spans 18.0 kb. *INHBB* maps to 2cen-q13, contains 2 exons and spans 5.8 kb (genetic information from *Ensembl*, http://www.ensembl.org). SNP data was downloaded from the CEPH population of the International HapMap Project, Phase II (**34**). The CEPH population consists of Utah residents with ancestry from northern and western Europe. We downloaded SNP data for each gene's extended gene region, which includes the region starting 20kb upstream to 10kb downstream. We then used HaploView (**35**) to identify SNPs with a minor allele frequency greater than 5%. Because some SNPs are correlated with each other, the program Tagger within Haploview was used to identify linkage disequilibrium-tagging SNPs, requiring that tagging SNPs be highly correlated ($R^2 \geq 0.95$) with SNPs not genotyped (**36**). This reduced the number to 8 LD-tagging SNPs in *INHA*, 15 in *INHBA*, and 14 in *INHBB*.

4.3.3 DNA Extraction and Genotyping

DNA from the frozen, stored urine samples was extracted, amplified and genotyped in the Emory Biomarker Service Center. Urine samples were extracted in duplicate on 20% of the women. Five mL aliquots of urine were centrifuged for 5 minutes at 3000 rpm to pellet cells and debris. DNA was extracted using the Qiagen MagAttract DNA Mini M48 kit in combination with the BioRobot M48 workstation.

The Beckman-Coulter GenomeLab SNPstream system was used to genotype the women for the SNPs using primers designed by Autoprimer.com (**37**). Up to 6 ng of DNA was used for genotyping, depending on the sample concentration. SNPs passed or failed genotyping based on default parameters in the GenomeLab SNPstream Genotyping System Software Suite v2.3 and manual quality control (signal intensity and clustering pattern). The 20% of samples extracted in duplicate were also genotyped in duplicate. Discordant duplicate genotypes were excluded from the analysis.

4.3.4 Menstrual Cycle Length Analysis

Cycle length was calculated from the daily dairies by taking the number of days from the first day of menstrual bleeding of one cycle until the first day of menstrual bleeding of the next cycle. The effect of each SNP (using genotype-based tests with 2 degrees of freedom) on menstrual cycle length was estimated using mixed linear models. A random effect for woman was included to account for correlation of cycle lengths within a woman. To reduce the effect of outliers and to remove cycles that may have a pathological basis, only cycles within two standard deviations of the mean

cycle length (14-46 days) were included in this analysis. Cycles in which a pregnancy occurred (identified using a hCG test at the start of the next cycle) were excluded because they may result in artificially longer cycles. Multiple comparisons were taken into account using the Nyholt procedure, which accounts for linkage disequilibrium between SNPs (**38**).

We then conducted a global test to determine the association between all SNPs in each inhibin gene and menstrual cycle length. We fit a semiparametric regression model using the method of least-squares kernel machines (LSKM), which yields a single global score statistic that measures the association between all SNPs (in each gene) and menstrual cycle length (**39**). This model is identical to analysis using a linear-mixed model that models the SNP data as a vector of random effects. This method has been shown to be more powerful than testing tagSNPs individually because it incorporates linkage-disequilibrium information from multiple SNPs simultaneously in analysis. However, it does not permit missing data, and thus the analysis could only accommodate those women who had 100% genotyping success, thereby leading to a reduced sample size. This LSKM analyses were adjusted for race, ethnicity, and age.

4.3.5 Menstrual Cycle Variability Analysis

Using logistic regression, we modeled the effect of each inhibin SNP on the probability of having highly variable cycles. The standard deviation of cycle length (a woman-level variable) was used as the measurement of cycle variability. The standard deviation of cycle length could only be calculated for women with two or more cycle

length measurements (N=402). The LSKM analysis could not be done with cycle variability because it requires a normally distributed outcome.

Age, body mass index (BMI, in kg/m^2), smoking, alcohol, and caffeine intake were evaluated as potential confounders. Only age and smoking met *a priori* criteria for confounding; they are predictors of menstrual cycle characteristics in the literature, and they were also associated with some inhibin SNP genotypes in this population. All analyses were either adjusted for self-reported race (White, Black, Asian, or Other) and ethnicity (Hispanic or non-Hispanic), or restricted to non-Hispanic whites, to reduce potential confounding by population stratification.

Interactions between SNPs were assessed by testing the significance of product terms between SNPs that were individually associated with either menstrual cycle length or variability.

The Emory Institutional Review Board approved the study protocol after complete de-identification of all samples, surveys and interviews.

4.4 Results

Of the 37 original SNPs, 25 were successfully genotyped, and 12 failed genotyping. Of the 470 women we attempted to genotype, 78 women failed genotyping, likely due to the age or quality of the frozen urine samples. The remaining 392 women were genotyped for at least one SNP. Demographic and reproductive characteristics of these women were not different from the 470 original women (Table 4.1).

Of the 25 SNPs genotyped, the minor allele frequencies were very similar to those reported in the CEPH population in HapMap (Table 4.2). Only one SNP violated Hardy-Weinberg equilibrium at the α=0.05 level (rs2059693, p=0.04), which could be expected by chance given the number of SNPs genotyped. Genotyping accuracy, as measured by duplicate concordance, ranged from 88% to 100% for each SNP. The mean overall genotyping accuracy was 97%. Discordant genotypes were excluded from the analysis.

The distribution of menstrual cycle lengths (N=3120) among the 392 women genotyped for at least one SNP is shown in Figure 4.2. Cycle lengths greater than 99 days were not shown in this histogram and they were excluded from all analyses because they are indicative of anovulation. The median cycle length was 28 days and the mean was 30.3 days.

Associations between each SNP and menstrual cycle length and variability for non-Hispanic whites (N=306 of 392 genotyped women), adjusted for age (continuous) and current smoking (yes/no), are shown in Tables 3-5. Results for combined races and ethnicities were similar. Effect estimates of SNPs varied within genes and generally had confidence intervals that spanned the null. One SNP in *INHA* (rs7588807) was associated with cycle length at the α=0.05 level (p=0.03). Carrying either one or two alleles of this intronic SNP was associated with a shorter cycle length by approximately 1.5 days (Table 4.3).

In *INHBA*, three SNPs were associated with menstrual cycle length and one was associated with cycle variability at the alpha=0.05 level (Table 4.4). The three SNPs associated with menstrual cycle length (rs3501158, rs2237432, and rs2237435) are in linkage disequilibrium (D' ranges from 0.90 to 1.00 and R^2 ranges from 0.04 to 0.52). For these

three SNPs, having one copy of the minor allele was associated with a longer cycle length by approximately one day. However, the effects of two copies of the minor alleles were inconsistent, likely due to small numbers. One SNP (rs7782324) was associated with cycle variability in an apparent dose-response manner (OR = 1.2 for carrying one copy of the minor allele and 3.1 for carrying two copies), but confidence intervals were wide.

In *INHBB*, carrying one copy of the minor allele of rs10201826 was associated with a longer cycle length of 1.2 days (95% CI, 0.1 to 2.2 days; p=0.05; Table 4.5). However, carrying two copies was not associated with cycle length.

Using the Nyholt procedure to correct for multiple comparisons, the p-value corresponding to α=0.05 is 0.0028 for this study. No associations achieved statistical significance according to this criterion. In addition, none of the three genes were significantly associated with menstrual cycle length using the LSKM method, indicating that the overall global variation in each gene was not associated with cycle length (Table 4,6).

Two-way SNP-SNP interactions were examined for all SNPs that were individually significantly associated with either menstrual cycle length (rs7588807, rs3801158, rs2237432, rs2237435, and rs10201826). Ten (five choose two) Wald tests were conducted in ten separate multivariate models. The strongest interaction was between rs2237435 and rs10201826, with a p-value of 0.11. No significant interactions were detected (data not shown).

4.5 Discussion

In this study, we examined whether polymorphisms in the extended gene regions of three inhibin genes (*INHA, INHBA*, and *INHBB*) were associated with menstrual cycle length or variability. A few SNPs were significantly associated at the α=0.05 level, including two that were associated with menstrual function in a dose-response manner. However, these associations did not persist after correcting for multiple comparisons. Using a global test to examine genetic variation across each gene in a single test statistic, no further associations were revealed. No statistically significant SNP-SNP interactions were present.

Inhibin and activin protein levels are unquestionably important to ovarian function. There are several possible explanations (other than chance) for why we did not find any associations between inhibin polymorphisms and menstrual function. First, inhibin and activin protein levels may not be influenced by polymorphisms in these genes, either because no such regulatory polymorphisms exist or because post-translational regulation is far more important to final protein levels. Second, polymorphisms in these genes may be important to transcription and to protein levels, but these polymorphisms were not selected for genotyping. Third, we may have selected influential polymorphisms, but low-quality DNA, low genotyping rates, and imperfect genotyping accuracy may have clouded true associations. Finally, it is possible that the selected polymorphisms do affect protein levels and ovarian functioning but that these effects were not reflected in menstrual cycle length or variability in this population. Each of these possibilities is evaluated below.

A pertinent issue for this study is whether inhibin protein levels are controlled by transcriptional, translational, or post-translational processes. If regulation is at the transcriptional level, polymorphisms in regulatory regions surrounding the coding sequence may be important. Follicle-stimulating hormone uses G-protein coupled receptors and can regulate expression of other genes through a cAMP-dependent mechanism (**40**). Cyclic AMP response elements are located in the promoter regions of human *INHBA* (**41**) and the rat homologue of *INHA* (**42**) but not human *INHBB* (**43**). The α and βA subunit mRNAs, but not βB mRNA, have been isolated from corpus lutea, suggesting at least some degree of transcriptional regulation. Further support for transcriptional (as opposed to translational or post-translational) regulation is that the secretion of inhibins A and B throughout the menstrual cycle mirrors their mRNA levels (**44, 45**). Therefore it appears that at least some important regulation does occur at the transcriptional level.

If regulation does occur at the transcriptional level, we may expect some polymorphisms or mutations to be important in determining inhibin mRNA levels. We did not genotype rare mutations (MAF<5%) in this study, reasoning that lack of power would prevent us from discovering any associations. In addition, all polymorphisms were selected from the HapMap database, which only provides information regarding a selection of SNPs along a strand of DNA. To discover all SNPs in the extended gene regions of these three genes, we would need to sequence every individual in this study in those regions. This could uncover mutations and rare variants important to transcriptional regulation, protein levels, protein sequence and activity, and ultimately, menstrual function. Other types of genetic variation (e.g., copy number variants or

epigenetic changes) could also serve regulatory functions. It is possible that some the SNPs in this study were in linkage disequilibrium with presumptive latent causal SNPs or variants. This could explain some of the marginally significant associations observed.

The inhibin genes encode both inhibin and activin proteins, which have opposing effects. Therefore another possible explanation of our null findings is that the polymorphisms in these genes are affecting both inhibin and activin protein levels and are therefore not detectable since changes inhibin function may be balanced by changes in activin function.

The use of twenty-year-old frozen urine samples as the DNA source undoubtedly impacted our results. It decreased the sample size, since high-quality DNA could not be extracted from some women; and of the women who were genotyped, many were only genotyped for a subset of the SNPs. In addition, the genotyping accuracy ranged from 88-100%, indicating some degree of misclassification. Discordant genotypes were excluded, but the mere existence of discordant genotypes indicates that some of the concordant genotypes may be incorrect as well, causing misclassification. Misclassification may cause bias if, for example, certain alleles are more likely to be detected than others and those alleles are associated with menstrual function. This misclassification could be obscuring any true associations (causing type II errors) or causing there to be an apparent association when in fact there is none (type I error), although type I errors did not occur in this study since no associations were found to be statistically significant.

Under the circumstance that polymorphisms affecting protein levels were in fact selected and accurately genotyped, and that these protein levels are important to ovarian functioning, it is still possible that any effects on ovarian function were not detectable as measurable changes in menstrual cycle length or variability. Inhibins and activins regulate follicle-stimulating hormone levels. Higher FSH levels are associated with shorter menstrual cycle length, likely in a causal manner, so if these polymorphisms do affect inhibin and activin protein levels, it should be reflected in cycle length (**46**). However, the effect could have been too small to be detected in this study since FSH levels are also regulated by other factors (e.g., follistatin) that may add to regulation by inhibins and activins; it is a complex milieu of factors that ultimately determines FSH levels (**47**). Likewise, many factors may affect cycle variability (smoking, stress, diet, exercise, etc.) and although we adjusted for smoking, the combination of factors (known and unknown) may have overcome any effect of the polymorphisms that were genotyped. In addition, the interaction between the proteins that form the inhibin and activin dimers may diminish the observed effects of individual polymorphisms.

In conclusion, the inhibins and activins are hormones that play key roles in follicular development and reproductive function. There is accumulating evidence that a woman's risk of chronic diseases, including breast and ovarian cancer, uterine fibroids, diabetes, and cardiovascular disease may be influenced by lifelong menstrual patterns and lifetime exposure to estrogen. Identifying common polymorphisms associated with menstrual function and fertility will not only contribute to our understanding of the genetic landscape of reproductive function, but also may help identify women who are at risk for future disease. Our study was the first to examine one component of a woman's

hormonal profile with respect to menstrual cycle characteristics. Although we did not find statistically significant associations, we have identified specific gaps that will help guide future research in this area.

4.6 Acknowledgments

This work was funded by an Emory University Research Committee grant, NIH RO1 grant #HD24618, and an NIH RO3 grant #HD055176. This manuscript was developed under a STAR Research Assistance Agreement No. 916891 awarded by the U.S. Environmental Protection Agency. It has **not** been formally reviewed by the EPA. The views expressed in this manuscript are solely those of Kira Creswell Taylor and the EPA **does not endorse** any products or commercial services mentioned in this manuscript.

4.7 Figures and Tables

Table 4.1 Population Characteristics of 470 Women Office Workers and 440 Women Who Were Genotyped for at Least One SNP in *INHA*, *INHBA*, and *INHBB*.

	All women N=470 N (%)	Genotyped women N=378 N (%)
Age (years)		
19-24	41 (9)	32 (8)
25-29	150 (32)	125 (32)
30-34	164 (35)	134 (34)
35-41	115 (24)	101 (26)
Race		
White	376 (80)	314 (80)
African-American or Black	58 (12)	50 (13)
Asian	13 (3)	8 (2)
Other	23 (5)	20 (5)
Ethnicity		
Hispanic	22 (5)	20 (5)
Non-Hispanic	445 (95)	369 (94)
Missing	3 (1)	3 (1)
Marital status		
Married	309 (66)	260 (66)
Single (never married)	127 (27)	103 (26)
Divorced/separated/widowed	34 (7)	29 (8)
Ever Pregnant		
Yes	285 (61)	239 (61)
No	185 (39)	153 (39)
Highest education		
High school or technical school	110 (23)	94 (24)
Some college	156 (33)	136 (35)
College graduate	203 (43)	162 (41)
Missing	1 (0)	0 (0)

BMI (kg/m^2)		
<20	80 (17)	61 (16)
20-25	240 (51)	199 (51)
26-30	91 (19)	78 (20)
>30	58 (12)	53 (14)
Missing	1 (0)	1 (0)
Current smoker		
Yes	120 (26)	108 (27)
No	350 (74)	284 (73)
Mean cycle length during follow-up		
<25 days	43 (9)	37 (9)
25-30 days	238 (51)	198 (51)
31-35 days	122 (26)	102 (26)
>35 days	67 (14)	55 (14)

Table 4.2 SNPs in the Extended Gene Region of *INHA*, *INHBA*, and *INHBB*

Gene	SNP rs number	Location of SNP	N[a]	MAF[b]	MAF in HapMap[c]	Hardy-Weinberg p-value	Genotyping Accuracy[d]
INHA	rs1039900	5’	269	0.47	0.48	0.29	100
	rs907141	5’	256	0.34	0.34	0.52	100
	rs1039898	5’	273	0.10	0.07	0.24	100
	rs7588807	Intron	245	0.50	0.53	0.08	93
	rs2059693	3’	220	0.29	0.30	0.04	95
INHBA	rs17776182	5’	219	0.14	0.16	0.07	95
	rs7782324	5’	258	0.46	0.39	0.49	98
	rs1003291	5’	202	0.20	0.19	0.57	94
	rs998190	5’	253	0.22	0.14	0.20	89
	rs1122291	5’	215	0.28	0.21	0.47	100
	rs2877098	5’	224	0.34	0.31	0.10	98
	rs3801158	Intron	255	0.20	0.15	0.39	100
	rs2237432	Intron	261	0.27	0.23	0.24	98
	rs11770488	Intron	248	0.17	0.20	0.14	98
	rs2237435	Intron	237	0.27	0.24	0.16	93
	rs2237436	Intron	209	0.40	0.37	0.70	100
	rs12701929	3’	219	0.28	0.22	0.10	94
	rs17705333	3’	228	0.27	0.25	0.81	95
INHBB	rs7593535	5’	264	0.21	0.17	0.90	98
	rs17625845	5’	232	0.20	0.17	0.86	98
	rs7589138	5’	200	0.44	0.41	0.96	94
	rs10201826	5’	196	0.44	0.40	0.83	88
	rs745723	3’	246	0.12	0.09	0.57	98
	rs7579169	3’	259	0.40	0.43	0.92	98
	rs1548039	3’	259	0.14	0.14	0.42	100

[a]Non-Hispanic whites genotyped.
[b]Minor allele frequency among non-Hispanic whites in this study.
[c]Minor allele frequency reported among the CEPH population in HapMap.
[d]As measured by concordance of duplicates.

Table 4.3 *INHA* and Cycle Characteristics among Non-Hispanic Whites (N=306)

INHA SNP	Contrast[a] # of minor alleles	Cycle length β (95% CI)	Cycle variability OR (95% CI)
rs1039900	2 vs. 0	0.9 (-0.2, 2.1)	1.0 (0.4, 2.6)
	1 vs. 0	0.7 (-0.3, 1.7)	1.0 (0.5, 2.2)
rs907141	2 vs. 0	-0.6 (-2.2, 1.0)	1.0 (0.3, 3.5)
	1 vs. 0	-0.2 (-1.1, 0.7)	0.8 (0.4, 1.7)
rs1039898	2 vs. 0	-0.9 (-7.2, 5.5)	--
	1 vs. 0	0.5 (-0.6, 1.6)	0.5 (0.2, 1.5)
rs7588807	2 vs. 0	-1.4 (-2.6, -0.1)*	0.4 (0.1, 1.1)
	1 vs. 0	-1.5 (-2.6, -0.4)*	0.6 (0.3, 1.2)
rs2059693	2 vs. 0	-1.9 (-3.8, 0.1)	--
	1 vs. 0	-0.4 (-1.3, 0.6)	0.9 (0.4, 1.8)

[a]2 vs. 0: 2 copies of the minor allele vs. 0 copies
1 vs. 0: 1 copy of the minor allele vs. 0 copies
*$p < 0.05$

Table 4.4 INHBA and cycle characteristics among non-Hispanic whites (N=306)

INHBA SNP	# of minor alleles	Cycle length β (95% CI)	Cycle variability OR (95% CI)
rs17776182	2 vs. 0	-2.0 (-8.7, 4.7)	--
	1 vs. 0	-0.03 (-1.1, 1.0)	1.3 (0.5, 3.0)
rs7782324	2 vs. 0	1.0 (-0.3, 2.2)	3.1 (1.2, 8.2)*
	1 vs. 0	0.3 (-0.7, 1.3)	1.2 (0.5, 2.9)
rs1003291	2 vs. 0	-1.3 (-3.6, 1.1)	1.4 (0.2, 12.8)
	1 vs. 0	0.8 (-0.1, 1.8)	1.9 (0.8, 4.5)
rs998190	2 vs. 0	-0.2 (-2.7, 2.3)	1.9 (0.4, 10.8)
	1 vs. 0	-0.2 (-1.2, 0.7)	1.7 (0.8, 3.5)
rs1122291	2 vs. 0	-0.2 (-2.0, 1.5)	1.9 (0.5, 8.2)
	1 vs. 0	-0.1 (-1.0, 0.8)	1.8 (0.8, 4.1)
rs2877098	2 vs. 0	1.3 (-0.3, 2.9)	1.0 (0.3, 3.6)
	1 vs. 0	0.04 (-0.9, 1.0)	0.8 (0.3, 1.7)
rs3801158	2 vs. 0	-0.2 (-2.1, 1.7)	0.6 (0.1, 4.8)
	1 vs. 0	1.3 (0.4, 2.3)*	1.8 (0.9, 3.7)
rs2237432	2 vs. 0	0.1 (-1.4, 1.6)	1.2 (0.4, 3.8)
	1 vs. 0	1.1 (0.2, 2.0)*	1.5 (0.7, 3.0)
rs11770488	2 vs. 0	-3.1 (-6.6, 0.4)	1.6 (0.2, 16.0)
	1 vs. 0	-0.8 (-1.8, 0.2)	0.4 (0.2, 1.1)
rs2237435	2 vs. 0	0.8 (-0.7, 2.3)	2.3 (0.8, 6.9)
	1 vs. 0	1.0 (0.04, 1.9)*	1.9 (0.4, 2.7)
rs2237436	2 vs. 0	1.0 (-0.3, 2.3)	2.1 (0.7, 6.4)
	1 vs. 0	0.8 (-0.2, 1.8)	1.2 (0.5, 3.0)
rs12701929	2 vs. 0	-0.1 (-2.1, 1.8)	0.5 (0.1, 3.8)
	1 vs. 0	0.2 (-0.7, 1.2)	1.1 (0.5, 2.5)
rs17705333	2 vs. 0	0.5 (-1.4, 2.3)	2.3 (0.7, 7.9)
	1 vs. 0	-0.1 (-1.0, 0.8)	0.7 (0.3, 1.6)

*$p<0.05$

Table 4.5 *INHBB* and Cycle Characteristics among Non-Hispanic Whites (N=306)

INHBB SNP	Contrast # of minor alleles	Cycle length β (95% CI)	Cycle variability OR (95% CI)
rs7593535	2 vs. 0	-0.8 (-2.8, 1.2)	0.4 (0.04, 2.9)
	1 vs. 0	-0.7 (-1.6, 0.2)	0.6 (0.3, 1.3)
rs17625845	2 vs. 0	-0.04 (-2.3, 2.3)	0.5 (0.1, 4.1)
	1 vs. 0	-0.6 (-1.5, 0.4)	0.5 (0.2, 1.1)
rs7589138	2 vs. 0	0.2 (-1.1, 1.5)	1.2 (0.4, 4.5)
	1 vs. 0	-0.8 (-0.3, 1.8)	2.1 (0.8, 5.8)
rs10201826	2 vs. 0	0.5 (-0.9, 1.8)	1.5 (0.4, 5.8)
	1 vs. 0	1.2 (0.1, 2.2)*	2.6 (0.9, 7.5)
rs745723	2 vs. 0	-1.5 (-5.9, 2.9)	1.7 (0.1, 19.4)
	1 vs. 0	0.5 (-0.6, 1.5)	1.3 (0.6, 2.9)
rs7579169	2 vs. 0	1.3 (-0.03, 2.5)	1.0 (0.4, 2.7)
	1 vs. 0	-0.5 (1.5, 0.4)	0.8 (0.4, 1.8)
rs1548039	2 vs. 0	-0.5 (-4.2, 3.2)	4.2 (0.6, 31.4)
	1 vs. 0	0.5 (-0.5, 1.5)	1.2 (0.6, 2.5)

*$p<0.05$

Table 4.6 Global Tests (Least-Squares Kernel Machines) Examining Associations between Variation in INHA, INHBA, and INHBB, and Menstrual Cycle Length

Gene	N[a]	Score statistic (df)[b]	p-value
INHA	286	3.9 (5.0)	0.57
INHBA	236	10.1 (6.3)	0.14
INHBB	249	5.8 (5.7)	0.42

[a]Sample sizes vary because each gene had a different subset of women who were genotyped for all SNPs in that gene.
[b]Adjusted for age, race (white, black, or other) and ethnicity. Further adjustment for smoking did not alter the results.

Figure 4.1 Population Flowchart for the Prospective Pregnancy Study

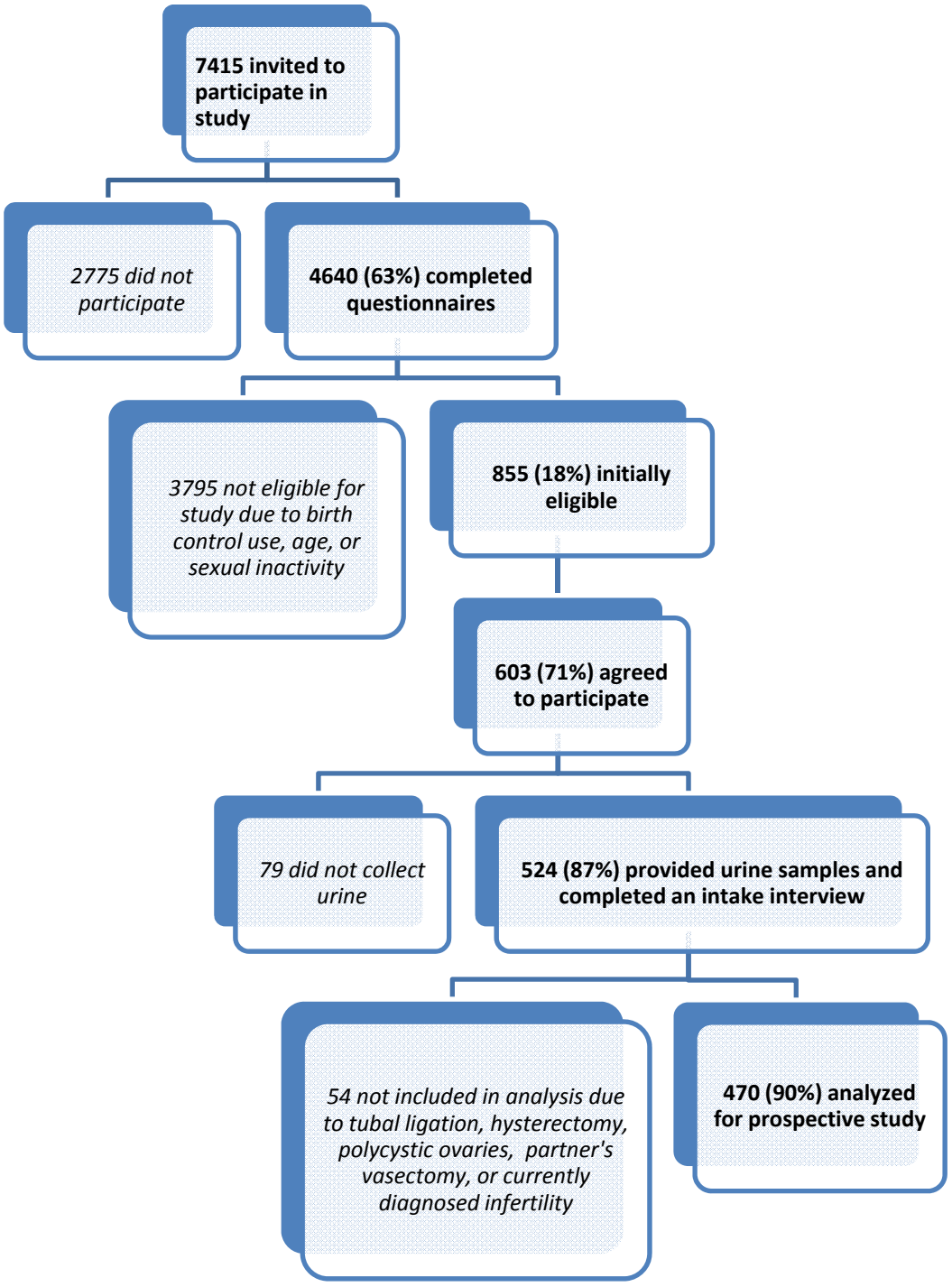
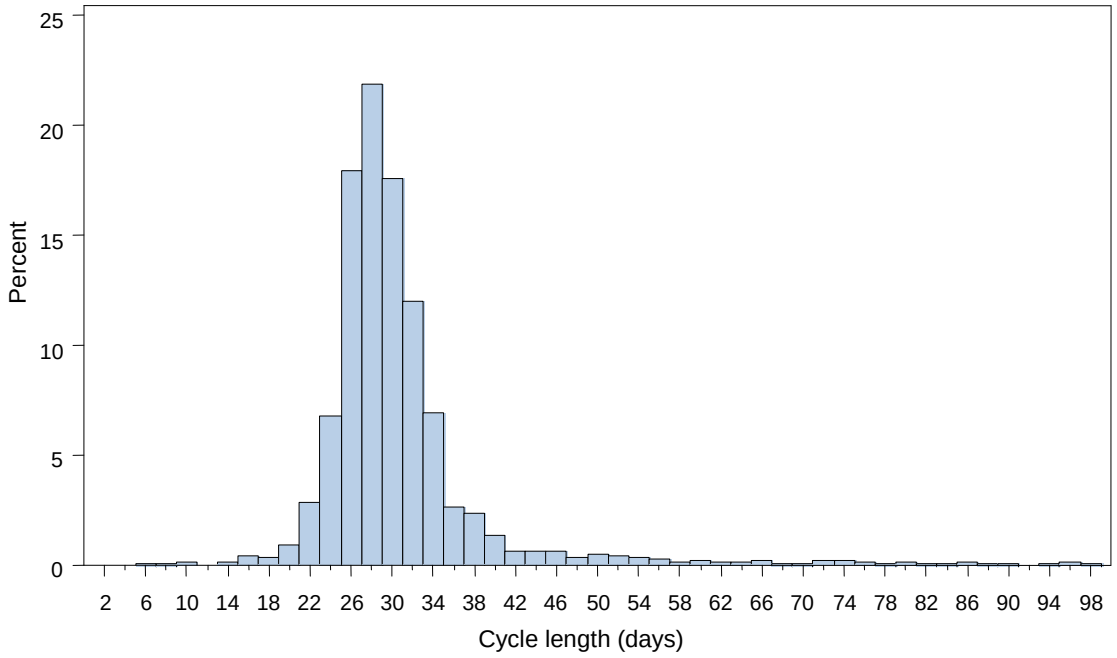


Figure 4.2 Distribution of Menstrual Cycle Length (N=3120) among 392 Women Office Workers.



References

1. Burger HG, Groome NP, Robertson DM. Both inhibin A and B respond to exogenous follicle-stimulating hormone in the follicular phase of the human menstrual cycle. *Journal of Clinical Endocrinology and Metabolism* 1998;**83**(11): 4167-9.

2. Mason AJ, Hayflick JS, Ling N, et al. Complementary DNA sequences of ovarian follicular fluid inhibin show precursor structure and homology with transforming growth factor-beta. *Nature* 1985;**318**(6047): 659-63.

3. Messinis IE. From menarche to regular menstruation: endocrinological background. *Annals of the New York Academy of Sciences* 2006;**1092**: 49-56.

4. Welt CK, McNicholl DJ, Taylor AE, Hall JE. Female reproductive aging is marked by decreased secretion of dimeric inhibin. *Journal of Clinical Endocrinology and Metabolism* 1999;**84**(1): 105-11.

5. Klein NA, Illingworth PJ, Groome NP, McNeilly AS, Battaglia DE, Soules MR. Decreased inhibin B secretion is associated with the monotropic FSH rise in older, ovulatory women: a study of serum and follicular fluid levels of dimeric inhibin A and B in spontaneous menstrual cycles. *Journal of Clinical Endocrinology and Metabolism* 1996;**81**(7): 2742-5.

6. Santoro N, Adel T, Skurnick JH. Decreased inhibin tone and increased activin A secretion characterize reproductive aging in women. *Fertility and Sterility* 1999;**71**(4): 658-62.

7. Klein NA, Harper AJ, Houmard BS, Sluss PM, Soules MR. Is the short follicular phase in older women secondary to advanced or accelerated dominant follicle development? *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(12): 5746-50.

8. Hall JE, Weltand CK, Cramer DW. Inhibin A and inhibin B reflect ovarian function in assisted reproduction but are less useful at predicting outcome. *Human Reproduction* 1999;**14**(2): 409-15.

9. Danforth DR, Arbogast LK, Mroueh J, et al. Dimeric inhibin: a direct marker of ovarian aging. *Fertility and Sterility* 1998;**70**(1): 119-23.

10. Gracia CR, Sammel MD, Freeman EW, et al. Defining menopause status: creation of a new definition to identify the early changes of the menopausal transition. *Menopause* 2005;**12**(2): 128-35.

11. Findlay JK, Russell DL, Doughton B, Tsonis CG, Borchers C, Forage RG. Effect of active immunization against the amino-terminal peptide (alpha N) of the alpha 43 kDa subunit of inhibin (alpha 43) on fertility of ewes. *Reproduction, Fertility and Development* 1994;**6**(2): 265-7.

12. Li R, Phillips DM, Mather JP. Activin promotes ovarian follicle development in vitro. *Endocrinology* 1995;**136**(3): 849-56.

13. Miro F, Smyth CD, Hillier SG. Development-related effects of recombinant activin on steroid synthesis in rat granulosa cells. *Endocrinology* 1991;**129**(6): 3388-94.

14. El-Hefnawy T, Zeleznik AJ. Synergism between FSH and activin in the regulation of proliferating cell nuclear antigen (PCNA) and cyclin D2 expression in rat granulosa cells. *Endocrinology* 2001;**142**(10): 4357-62.

15. Garrett WM, Mack SO, Rohan RM, Guthrie HD. In situ analysis of the changes in expression of ovarian inhibin subunit mRNAs during follicle recruitment after ovulation in pigs. *Journal of Reproduction and Fertility* 2000;**118**(2): 235-42.

16. Rabinovici J, Spencer SJ, Jaffe RB. Recombinant human activin-A promotes proliferation of human luteinized preovulatory granulosa cells in vitro. *Journal of Clinical Endocrinology and Metabolism* 1990;**71**(5): 1396-8.

17. Takagi T, Mizunuma H, Andoh K, et al. Changes in serum immunoreactive inhibin during ovulation induction in women with amenorrhea. *Endocrine Journal* 1994;**41**(6): 703-8.

18. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *European Journal of Human Genetics* 2001;**9**(4): 291-300.

19. Xiao S, Robertson DM, Findlay JK. Effects of activin and follicle-stimulating hormone (FSH)-suppressing protein/follistatin on FSH receptors and differentiation of cultured rat granulosa cells. *Endocrinology* 1992;**131**(3): 1009-16.

20. Duleba AJ, Pehlivan T, Carbone R, Spaczynski RZ. Activin stimulates proliferation of rat ovarian thecal-interstitial cells. *Biology of Reproduction* 2001;**65**(3): 704-9.

21. Woodruff TK, Lyon RJ, Hansen SE, Rice GC, Mather JP. Inhibin and activin locally regulate rat ovarian folliculogenesis. *Endocrinology* 1990;**127**(6): 3196-205.

22. Safi M, Buys N, Onagbesan OM, Vleugels B, Decuypere E. Quantification of inhibin/activin alpha and betaA subunit messenger ribonucleic acid by competitive reverse transcription-polymerase chain reaction in chicken granulosa cells during follicular development. *Biology of Reproduction* 1998;**59**(5): 1047-54.

23. Brown RL, Pollock AN, Couchman KG, et al. A novel ryanodine receptor mutation and genotype-phenotype correlation in a large malignant hyperthermia New Zealand Maori pedigree. *Human Molecular Genetics* 2000;**9**(10): 1515-24.

24. Shelling AN, Burton KA, Chand AL, et al. Inhibin: a candidate gene for premature ovarian failure. *Human Reproduction* 2000;**15**(12): 2644-9.

25. Dixit H, Deendayal M, Singh L. Mutational analysis of the mature peptide region of inhibin genes in Indian women with ovarian failure. *Human Reproduction* 2004;**19**(8): 1760-4.

26. Woad KJ, Pearson SM, Harris SE, Gersak K, Shelling AN. Investigating the association between inhibin alpha gene promoter polymorphisms and premature ovarian failure. *Fertility and Sterility* 2009;**91**(1): 62-6.

27. Marozzi A, Porta C, Vegetti W, et al. Mutation analysis of the inhibin alpha gene in a cohort of Italian women affected by ovarian failure. *Human Reproduction* 2002;**17**(7): 1741-5.

28. Harris SE, Chand AL, Winship IM, et al. INHA promoter polymorphisms are associated with premature ovarian failure. *Molecular Human Reproduction* 2005;**11**(11): 779-84.

29. Chand AL, Robertson DM, Shelling AN, Harrison CA. Mutational analysis of betaglycan/TGF-[beta]RIII in premature ovarian failure. *Fertility and Sterility* 2007;**87**(1): 210-2.

30. Leyhe B, Hiendleder S, Jaeger C, Wassmuth R. Pronounced differences in the frequency of TaqI beta A-inhibin alleles between sheep breeds with different reproductive performance. *Animal Genetics* 1994;**25**(1): 41-3.

31. Wrathall JH, McLeod BJ, Glencross RG, Knight PG. Effects of active immunization against a synthetic peptide sequence of the inhibin alpha-subunit on plasma gonadotrophin concentrations, ovulation rate and lambing rate in ewes. *Journal of Reproduction & Fertility* 1992;**95**(1): 175-82.

32. Yen SSC. The human menstrual cycle: neuroendocrine regulation. In: Yen SSC, Jaffe RB, Barbieri RL, editors. *Reproductive Endocrinology: Physiology,*

Pathophysiology, and Clinical Management. 4 ed. Philadelphia: W.B. Saunders; 1999. p. 191-217.

33. Marcus M, Gerr F. Upper extremity musculoskeletal symptoms among female office workers: associations with video display terminal use and occupational psychosocial stressors. *American Journal of Industrial Medicine* 1996;**29**(2): 161-70.

34. The International HapMap Consortium: The International HapMap Project. *Nature* 2003;**426**(6968): 789-96.

35. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**(2): 263-5.

36. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nature Genetics* 2005;**37**(11): 1217-23.

37. Beckman Coulter I. GenomeLab SNPstream Genotyping System 48-plex Training Guide. In: Beckman Coulter I, editor. Fullerton, CA: Beckman Coulter, Inc.; 2005.

38. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *American Journal of Human Genetics* 2004;**74**(4): 765-9.

39. Kwee LC, Liu D, Lin X, Ghosh D, Epstein MP. A powerful and flexible multilocus association test for quantitative traits. *American Journal of Human Genetics* 2008;**82**(2): 386-97.

40. Richards JS, Hedin L. Molecular Aspects of Hormone Action in Ovarian Follicular Development, Ovulation, and Luteinization. *Annual Review of Physiology* 1988;**50**(1): 441-63.

41. Tanimoto K, Yoshida E, Mita S, Nibu Y, Murakami K, Fukamizu A. Human Activin beta A Gene. Identification of novel 5prime exon, functional promoter, and enhancers. *Journal of Biological Chemistry* 1996;**271**(51): 32760-9.

42. Pei L, Dodson R, Schoderbek WE, Maurer RA, Mayo KE. Regulation of the alpha inhibin gene by cyclic adenosine 3',5'- monophosphate after transfection into rat granulosa cells. *Molecular Endocrinology* 1991;**5**(4): 521-34.

43. Feng ZM, Wu AZ, Chen CL. Characterization and regulation of two testicular inhibin/activin beta B-subunit messenger ribonucleic acids that are transcribed from alternate initiation sites. *Endocrinology* 1995;**136**(3): 947-55.

44. Groome NP, Illingworth PJ, O'Brien M, et al. Measurement of dimeric inhibin B throughout the human menstrual cycle. *Journal of Clinical Endocrinology and Metabolism* 1996;**81**(4): 1401-5.

45. Roseff SJ, Bangah ML, Kettel LM, et al. Dynamic changes in circulating inhibin levels during the luteal-follicular transition of the human menstrual cycle. *Journal of Clinical Endocrinology and Metabolism* 1989;**69**(5): 1033-9.

46. Miro F, Parker SW, Aspinall LJ, Coley J, Perry PW, Ellis JE. Relationship between follicle-stimulating hormone levels at the beginning of the human menstrual cycle, length of the follicular phase and excreted estrogens: the FREEDOM study. *Journal of Clinical Endocrinology and Metabolism* 2004;**89**(7): 3270-5.

47. Phillips DJ, de Kretser DM. Follistatin: a multifunctional regulatory protein. *Frontiers in Neuroendocrinology* 1998;**19**(4): 287-322.

Chapter 5

Associations of Progesterone Receptor Polymorphisms with Age at Menarche and Menstrual Cycle Length

5.1 Abstract

Background. Age at menarche and menstrual cycle characteristics are indicators of endocrine function and may be risk factors for diseases such as reproductive cancers. The progesterone receptor gene (*PGR*) has been identified as a candidate gene for age at menarche, menstrual cycle variability and fertility. This study examines associations of genetic variation in the progesterone receptor with age at menarche and menstrual cycle characteristics.

Methods. 515 women office workers self-reported age at menarche in an interview and 470 participated in a prospective study of fertility in which they recorded menstrual bleeding and covariates in daily diaries for 1-19 cycles. First-morning urine was used as the DNA source. Women were genotyped for a functional variant in *PGR*, rs1042838 (Val660Leu), and 29 other SNPs across the extended gene region**.**

Results. Genetic variation across all 30 *PGR* SNPs was associated with age at menarche using a global score statistic (p=0.03). Women carrying two copies of the Val660Leu variant experienced menarche at a significantly older age than women carrying one or no copies of the variant (13.6±0.6 vs. 12.6±0.1; p=0.04 among non-Hispanic whites). The Val660Leu variant was also associated with decreased odds of

short menstrual cycles (17-24 days) (odds ratio, 95% confidence interval: 0.54 (0.36, 0.80); p=0.002).

Conclusion. Genetic variation in *PGR* is associated with age at menarche and menstrual cycle length in this population. Further investigation of these associations in a replication dataset is warranted.

5.2 Introduction

Age at menarche is highly heritable, with approximately 50-60% of the variance attributable to genetic factors (**1, 2**). Menstrual cycle characteristics, such as menstrual cycle length, may also have a genetic component (**3**). A recent genome-wide linkage scan for loci affecting age at menarche identified three genomic regions with significant LOD scores (**1**). One of these regions contains the progesterone receptor gene (*PGR*). Genetic variants in the promoter and coding regions of this gene have been associated with breast, endometrial, and ovarian cancer, although there is some inconsistency among the studies (**4-9**).

A genetic variant of *PGR* coined "PROGINS" consists of a 320 base pair (bp) Alu insertion in an intron, which is in 100% linkage disequilibrium (LD) with a functional (non-synonymous) variant Val660Leu and a synonymous variant His770His. An *in vitro* study showed that the PROGINS variant may have decreased response to progestins, and is not as efficient at opposing estrogen's proliferative effects, due to decreased mRNA stability and protein activity (**10**). Carriers of the Val660Leu variant were more likely to be nulliparous, infertile, and experience irregular menstrual cycles, and were less likely to have premenstrual weight gain or breast pain among controls in a case-control study of ovarian cancer (**11**). The Val660Leu variant was also significantly associated with spontaneous abortion in another case-control study (**12**).

No studies were identified that examined the association between progesterone receptor polymorphisms and age at menarche. Because puberty and menstruation are complex processes that are dependent on feedback mechanisms involving the action of

progesterone (**13**), we hypothesized that variation in the progesterone receptor may influence age at menarche and menstrual cycle characteristics. To investigate this hypothesis, we genotyped a population of female office workers for a set of single-nucleotide polymorphisms (SNPs) that tag variation within the extended gene region of *PGR* and subsequently assessed whether such SNPs were associated with age at menarche. We then examined whether there was any association between the Val660Leu variant and menstrual cycle characteristics or age at menarche.

5.3 Methods

5.3.1 Population

Women office workers in New York, New Jersey and Massachusetts were enrolled in a study of reproductive health from 1990 through 1994 (**14**). A total of 4,640 women completed self-administered questionnaires including questions on reproductive health and current birth control practices. Women were eligible for a prospective study of fertility if they were between the ages of 18 and 40 and had been sexually active in the month prior to completing the questionnaire while using inconsistent or no birth control (n=855, Figure 5.1). The study required first-morning urine collection at least two days each cycle at the onset of menstrual bleeding. 603 eligible women agreed to participate. Of these, 524 women collected at least one urine sample, which is the source of DNA for this study. The women reported age at menarche, as well as covariates such as race, ethnicity, and year of birth, during an interview at the onset of the prospective study. Three women who did not report age at menarche and six women who did not report either race or ethnicity were excluded, resulting in a starting sample size of 515 for the

study of age at menarche. Of the 524 women, 470 women were eligible for the menstrual cycle analysis based upon the following criteria: completed follow-up for at least one menstrual cycle; no history of hysterectomy, polycystic ovaries, or tubal ligations; not currently infertile; and partner has not had a vasectomy. These women completed daily dairies that included information on menstrual bleeding as well as covariates such as intercourse; birth control use; smoking; and alcohol and caffeine consumption (see Daily Diary, Appendix 5.A). Three women who did not report race or ethnicity were excluded, resulting in a sample size of 467 for the menstrual cycle analyses.

5.3.2 Selection of Single Nucleotide Polymorphisms

The Progesterone Receptor is Entrez gene NM_000926, Chromosome 11 q22-23, position 100414313-100506465. SNP data were downloaded from the CEPH population of the International HapMap Project, Phase II (**15**). The CEPH population consists of Utah residents with ancestry from northern and western Europe. We identified 293 SNPs in the extended gene region of the progesterone receptor, which includes the region starting 20 kb upstream of the gene to 10kb downstream. We used HaploView (**16**) to identify SNPs with a minor allele frequency greater than 5% (145 SNPs). Because some SNPs are correlated with each other, the program Tagger within Haploview was used to identify linkage disequilibrium (LD) -tagging SNPs, requiring that tagging SNPs be highly correlated ($R^2 \geq 0.95$) with SNPs not genotyped (**17**). This reduced the number to 37 LD-tagging SNPs, including the Val660Leu variant, rs1042838 (Appendix 5.B).

5.3.3 DNA Extraction and Genotyping

DNA from the frozen, stored urine samples was extracted, amplified and genotyped in the Emory Biomarker Service Center. Urine samples were extracted in duplicate on 20% of the women. Five mL aliquots of urine were centrifuged for 5 minutes at 3000 rpm to pellet cells and debris. DNA was extracted using the Qiagen MagAttract DNA Mini M48 kit in combination with the BioRobot M48 workstation. For women with unsatisfactory results (<3 ng/μL DNA) using the first method, we extracted DNA from additional urine samples using the PureGene protocol "DNA Purification from Body Fluid" using the Gentra PureGene Blood Kit.

The Beckman-Coulter GenomeLab SNPstream system was used to genotype the women for the 37 SNPs using primers designed by Autoprimer.com (**18**) (Appendix 5.C). Up to 6 ng of DNA was used for genotyping, depending on the sample concentration. The 20% of samples extracted in duplicate were also genotyped in duplicate. SNPs passed or failed genotyping based on default parameters in the GenomeLab SNPstream Genotyping System Software Suite v2.3 and manual quality control (signal intensity and clustering pattern).

We wished to increase our sample size for the rs1042838 SNP, given the large body of literature concerning this SNP. We re-genotyped this SNP from additional samples for every woman on a separate SNPstream chip to increase the genotyping success rate for this SNP of particular interest.

To assess genotyping accuracy, we examined the concordance among duplicate genotypes. We investigated whether the SNPs were in Hardy-Weinberg equilibrium using the calculator in HaploView, which uses an exact test (**16, 19**).

5.3.4 Age at Menarche Analyses

We examined whether the mean age at menarche varied by genotype using an ANOVA for a 3-way comparison across genotypes among non-Hispanic whites. Linear regression was also conducted to estimate the effect of each SNP on age at menarche in the total population, adjusting for the effects of race (white, black or African-American, Asian, or other) and ethnicity (Hispanic or non-Hispanic) to partially control for potential population stratification. Results were adjusted for multiple comparisons using the Nyholt method, which takes linkage disequilibrium between SNPs into account (Nyholt 2004).

We identified eight haplotype blocks in this region using HaploView, where a haplotype block was defined according to the 95% confidence bounds on D', according to Gabriel et al.(**20**). We used the haplo.score function of the R package haplo.stats to infer the phase of haplotypes and determine the association between each haplotype block and age at menarche (**21**). Haplo.stats uses the expectation-maximization algorithm to phase haplotypes from unphased genotype data.

We then conducted a global test to determine the association between all SNPs in the gene and age at menarche. We fit a semiparametric regression model using the method of least-squares kernel machines (LSKM), which yields a single global score

statistic that measures the association between all 30 SNPs and age at menarche (**22**). This model is identical to analysis using a linear-mixed model that models the SNP data as a vector of random effects. This method has been shown to be more powerful than testing tagSNPs individually because it incorporates linkage-disequilibrium information from multiple SNPs simultaneously in analysis. However, it does not permit missing data, and thus the analysis could only accommodate those women who had 100% genotyping success, thereby leading to a reduced sample size. Therefore, as a sensitivity analysis among the non-Hispanic whites, we repeated the analysis on all women who had more than 50% genotyping success, imputing missing genotype data using PHASE v 2.1.1, which uses a Bayesian method of haplotype reconstruction given genotype data (**23, 24**). PHASE may output several possible haplotypes for an individual, each with an assigned probability. We first used the haplotypes with the highest probabilities for each individual to rerun the LSKM program, and repeated the analysis using the haplotypes with the lowest probabilities.

An important epidemiologic predictor of age at menarche is nutrition during childhood, which could be measured as childhood body mass index (BMI) or central adiposity (**25-27**). Because the study included adult women, we were unable to control for childhood BMI. As a crude surrogate, we explored models including adult BMI, but recognize that the temporal sequence of this relationship may be erroneous if age at menarche influences adult BMI.

5.3.5 Menstrual Function Analysis

For women who were followed for at least one full menstrual cycle, cycle length was calculated from the daily dairies by taking the number of days from the first day of menstrual bleeding of one cycle until the first day of menstrual bleeding of the next cycle. Cycles less than 17 days in length were excluded on the basis that spotting may have been misinterpreted as a menstrual bleed. Cycles longer than 99 days were excluded because they are indicative of an anovulatory condition. Cycles were categorized as short (17-24 days), standard (25-35 days), or long (36-99 days), based on the top and bottom deciles of cycle length, and are consistent with other definitions in the literature. The standard deviation of cycle length was used as the measure of cycle variability. Therefore a woman had to have completed at least two complete menstrual cycles to be included in the cycle variability analysis.

Generalized estimating equations were used to model the effect of the rs1042838 genotype and covariates on the probability of experiencing long or short cycles, adjusting for multiple cycles within a single woman. Dose-response models and recessive genetic models were examined.

All menstrual function analyses were either adjusted for self-reported race (White, Black, or other) and ethnicity (Hispanic or non-Hispanic), or restricted to non-Hispanic whites, to reduce potential confounding by population stratification. Other predictors of menstrual function include age, BMI, smoking, alcohol, and caffeine use. These were explored as covariates; however, it should be noted that if including these covariates in regression models alters the parameter estimates of SNPs on any menstrual function

outcomes, it may be because these variables are in fact in a causal pathway, and therefore may not be true confounders.

The Emory Institutional Review Board approved the study protocol after complete de-identification of all samples, surveys and interviews.

5.4 Results

Thirty-one SNPs were successfully genotyped of the 37 attempted. This is within the expected range for the Beckman SNPstream system (W. Tang, personal communication). Of the 515 women genotyped for these 31 SNPs, 118 women failed genotyping completely in the first round of genotyping, likely due to the age of the urine samples and the low DNA concentration in some urine. Genotypes were obtained on at least one SNP for 397 women (Figure 5.1). The mean DNA concentration of the women who were successfully genotyped for at least one SNP was 8.6 ng/μL, compared with 1.1 ng/μL for the failed samples. Women with DNA concentrations >3 ng/μL experienced >99% genotyping success (Appendix 5.D). Genotyping accuracy was greater than 99%, as calculated by determining the percent of duplicate genotypes that were concordant (N=2190 of 2194). The few discordant genotypes (n=4) were set to missing.

To be included in the gene-wide analysis of *PGR* and age at menarche, we required that a woman be genotyped for at least 50% of the SNPs. This reduced the sample size to 350 women (250 non-Hispanic whites). We then excluded one SNP with less than 90% genotyping success. Thus our sample for the gene-wide analysis included data on 350 women and 30 SNPs. The mean success rate of SNPs in this sample was

98% (median 99%). The linkage disequilibrium and haplotype block structure for the 30 SNPs among the non-Hispanic whites in this population are shown in Figure 5.2.

Genotypes were obtained on all 30 SNPs for 264 women; the remaining women (n=86) had a subset of the SNPs available for analysis. None of the 30 SNPs violated Hardy-Weinberg equilibrium (Table 5.1). Among the non-Hispanic whites, the minor allele frequencies (MAF) for all SNPs were very similar to those reported in the CEU population in HapMap.

During the second round of genotyping (in which the PROGINS SNP, rs1042838, was re-genotyped using DNA extracted from additional urine samples from each woman, and again in duplicate for 20% of the women), additional genotypes were obtained for this SNP, resulting in a total of 444 women available for the PROGINS analyses.

The study population was mostly white and non-Hispanic (Table 5.2). Of the original 515 women we attempted to genotype, the mean age at interview was 31±0.21 (median=31), and the mean age at menarche was 12.7±0.07 (median=13). The characteristics of the 264 women who were successfully genotyped for all 30 SNPs differed markedly in the DNA concentration of their final extracted DNA samples when compared to the other 251 women (11.5±0.7 ng/μL vs. 3.9±0.5 ng/μL, $p<0.001$). The two groups did not differ by age at interview, decade of birth, or age at menarche. Being white or non-Hispanic (both $p=0.04$) and having higher education ($p=0.01$) were associated with lower genotyping success. However, in a multiple logistic regression model that included all variables in Table 5.2 and in a reduced model that included DNA concentration, race, ethnicity, and education, DNA concentration was the only significant

predictor of genotyping success ($p<0.001$) and no other variables approached significance (data not shown).

None of the covariates listed in Table 5.2 were significantly associated with age at menarche in a multivariate model or in models adjusted for each covariate individually, indicating that there was not an association of race or ethnicity with age at menarche (Appendix 5.E). There was not a secular trend of age at menarche during this time period in this population (Appendix 5.F).

Missing data are not permitted for the LSKM method that tests the gene-wide association described in Kwee et al. (2008), and therefore only women with data for all 30 SNPs could be included in the LSKM analysis (N=264). There was a weak association between global variation and *PGR* and age at menarche, adjusted for race and ethnicity ($p=0.088$, Table 5.3). When we restricted the population to the non-Hispanic whites with complete genotype data (N=181), the association was strengthened ($p=0.03$).

As a sensitivity analysis, we imputed missing genotypes for non-Hispanic white women with >50% genotyping success (N=250) and repeated the LSKM analysis. Results were nearly identical to the original analysis: $p=0.03$ using the most likely haplotypes, and $p=0.04$ using the least likely haplotypes (Table 5.3).

Of the eight haplotype blocks identified in HaploView (Figure 5.2), two were marginally associated with age at menarche (Table 5.4): block 1, including upstream SNPs rs474320 and rs521488 (2 df, $p=0.035$), and block 5, including the intronic SNP rs660541 and the functional PROGINS SNP rs1042838 (2 df, $p = 0.041$). These

associations were no longer significant after a Bonferroni correction for multiple comparisons (threshold p-value = 0.006). A Bonferroni correction could be used here because the haplotype blocks are not in strong LD with each other. Regression coefficients for all haplotypes in the eight haplotype blocks are also shown in Table 5.4. As an example, women who carried the TT haplotype in haplotype block 1 experienced menarche 0.35±0.16 years earlier, on average, than women with the reference haplotype (TC).

Mean ages at menarche for the 3 genotypes for each of the 30 SNPs are shown in Table 5.5 for non-Hispanic whites. Results for all races combined were similar. To maintain a type I error rate of 0.05, the threshold p-value was 0.0018, according to the Nyholt method for adjustment for multiple comparisons in the presence of linkage disequilibrium. Only one SNP was associated with age at menarche at this level of significance, using a genotype-based model (rs948516; p=0.0006).

Carriers of the TT genotype (N=13) of rs1042838 experienced menarche a year later than those with the GT or GG genotypes (Table 5.6). This association was significant in unadjusted analyses (p=0.03, TT vs others); in analyses restricted to non-Hispanic whites (p=0.04, TT vs. others); and in a linear regression model adjusted for race, ethnicity, and adult BMI (p=0.05, TT vs. others; Appendix 5.G).

The mean menstrual cycle lengths were not significantly different across the three rs1042838 genotypes (Table 5.7; p=0.30 in a mixed linear model adjusted for race and ethnicity). However, each additional T allele was associated with significantly decreased odds of having short cycles (17-25 days) in a clear dose-response manner. These

findings were remarkably robust to adjustment for potential confounders including age at interview, smoking, caffeine, alcohol, and BMI. There was no association of rs1042838 genotype with either the odds of long cycles or cycle variability. The distributions of cycle lengths in this population are shown according to rs1042838 genotype in Appendix 5.H. Individual women contributed multiple cycles to the analysis (n=1 to 19, median=7), thus cycle lengths shown in Appendix 5.H are not independent.

5.5 Discussion

We observed an association of variation in the progesterone receptor gene with age at menarche using a global analysis that examined multiple SNPs in a single test. This finding is consistent with a whole genome linkage scan for age at menarche, in which the progesterone receptor was identified as a candidate gene (**1**). Two possible haplotypes and one SNP located in the promoter region may be partially responsible for the global association. We also observed a significant association between the Val660Leu variant with older age at menarche and with decreased probability of short menstrual cycles.

The precise sequence of physiological events that result in menarche is unknown, and the mechanism is likely complex. Progesterone is thought to be an important factor. Progesterone and estrogen levels increase at the inception of puberty, and progesterone production and subsequent withdrawal is necessary for a menstrual bleed to occur (**13**). In addition, high adrenal progesterone levels can prevent menarche (**28**). Therefore, it is plausible that genetic variation in progesterone or the progesterone receptor may affect pubertal development and age at menarche.

The stability and transcriptional activity of the PROGINS variant of *PGR* differs from the wild-type receptor (**10, 29**); thus the response to progesterone (and the threshold levels required for menstruation and menarche) could vary according to *PGR* genotype. Variants in both coding and non-coding regions, including those analyzed in this study and those in linkage disequilibrium with SNPs in this study, could be responsible for variation in transcriptional, translational, or functional activity of *PGR*. One study examined the levels of PgR protein across women of varying PROGINS genotypes and found no difference; however, the maximum level of PgR protein was lower for PROGINS carriers than non-carriers (~150 vs. 550 fmol/mg protein) (**30**). Hypothetically, even if protein levels are similar across PGR genotypes, the different forms of the protein could have different transactivational activities and therefore different downstream effects.

While the sample size is small, the older age at menarche among those with the TT genotype for rs1042838 is consistent with what is known about the biology. The PROGINS variant (which includes Val660Leu or the T allele of rs1042838) has been shown to have decreased response to progestins *in vitro,* and may be less efficient at suppressing estrogen-induced proliferation (**10**). If the PROGINS variant has a reduced response to progestins *in vivo*, this could result in a delay of menarche. It may seem conflicting that high adrenal progesterone levels can prevent menarche (**28**); however, it is plausible that adrenal progesterone levels may be inversely related to progesterone levels in the reproductive tract; in addition, presence of the PROGINS variant could result in higher progesterone levels through feedback mechanisms.

Three studies have found that the PROGINS variant is associated with a modestly increased risk of breast cancer, possibly due to its reduced opposition to estrogen's mitogenic effects (**5, 9, 31**). In a meta-analysis of ovarian cancer cases and controls, the rs1042838 SNP was associated with a significantly increased risk of endometrioid ovarian cancer (**32**). Either this SNP may have pleiotropic effects on multiple reproductive outcomes, or its effects on cancer risk may be mediated through age at menarche. However, this study does not support the latter hypothesis, since the variant was associated with older age at menarche, but appears to increase the risk of reproductive cancers.

A simple mechanism for the observed association of the rs1042838 with a decreased risk of short menstrual cycles is not clear. In general, low progesterone levels may result in luteal phase deficiency and shorter menstrual cycles. To be consistent with this mechanism, the PROGINS variant should be associated with higher progesterone levels if it causes decreased risk of short cycles. If the PROGINS variant of *PGR* is less responsive to progesterone, as hypothesized by Romano et al. (2007), it is plausible that this could result in increased progesterone levels through homeostasis and feedback mechanisms. If progesterone levels are higher, this could have far-reaching complex effects independent of any effects through the progesterone receptor. For example, higher progesterone levels may affect estrogen levels, which could possibly result in a lower risk of short menstrual cycles. While the mechanism may be elusive, the observed dose-response relationship bolsters the plausibility of a causal association, and future studies should further characterize association between progesterone receptor variants and menstrual cycle characteristics. *In vitro* studies or animal studies investigating

feedback mechanisms between progesterone receptor activity and progesterone and estrogen levels could also help elucidate the biological mechanism underlying this association.

A unique aspect of this study was the demonstration of urine as a valid DNA source for epidemiological studies. Studies have shown that urine can be as valid as blood or other sources for genotyping, depending on the age and storage conditions of the samples (**33, 34**). Our DNA yield was lower than these studies. However, our genotyping accuracy was over 99%, as measured by the concordance of genotypes among duplicate samples, and all of our SNPs were in Hardy-Weinberg equilibrium. The minor allele frequency for all SNPs in our dataset were almost identical to those found in HapMap, providing evidence that there was not bias in genotyping success. In addition, when we conducted a sensitivity analysis examining the worst case scenario (assigning all missing genotypes to their least likely value), the association remained. Furthermore, genotyping success was not related to age at menarche. Therefore, while missing genotype data reduced the power of this study, we have substantial evidence that it was unlikely to bias results.

We did not observe an association between race or ethnicity and age at menarche in this study. The women in this study experienced menarche mainly in the 1960's and 1970's. Consistent with our study, the National Health and Nutrition Survey shows no significant difference of age at menarche among races in the 1960's (**35**). In addition, no association between year of birth and age at menarche was observed. Although a secular

trend of decreasing age at menarche was observed over the course of the 20th century, (**35**) it may not be apparent in this study because of the limited time frame.

In conclusion, variation in the progesterone receptor was associated with age at menarche and menstrual cycle length in this population. Understanding the genetic contributions to menstruation and menarche can help elucidate the biological pathways and causal mechanisms involved, including clarifying the roles of genes, environment, and gene-environment interactions. A more complete picture of the factors affecting age at menarche may eventually help identify those at risk for disorders and chronic diseases associated with menarche. Likewise, understanding the influences of genes on menstruation will add to the body of knowledge concerning menstrual dysfunction and associated morbidities such as infertility. Replication of this study, with particular attention to the Val660Leu variant and the novel SNP identified in the promoter region, is needed to clarify the relationship between progesterone receptor genotypes, age at menarche, menstrual function, and risk of reproductive cancers.

5.6 Acknowledgments

This research was supported by funding from the National Institutes of Health (RO1HD24618, RO3HD55176, R01HG03618), and the University Research Committee of Emory University. This manuscript was developed under a STAR Research Assistance Agreement No. 916891 awarded by the U.S. Environmental Protection Agency. It has **not** been formally reviewed by the EPA. The views expressed in this manuscript are solely those of Kira Creswell Taylor and the EPA **does not endorse** any products or commercial services mentioned in this manuscript.

5.7 Figures and Tables

Table 5.1 SNPs Genotyped in the Extended Gene Region of *PGR*

SNP #	dbSNP reference	Minor/Major Allele	N[a]	N[b]	HW p-value	MAF[c]	MAF in HapMap[d]
Upstream							
1	rs474320	A/T	317	223	1.00	0.19	0.18
2	rs521488	C/T	346	247	0.97	0.49	0.48
3	rs948516	G/A	325	233	0.73	0.29	0.33
In gene region[e]							
4	rs550778	C/A	338	240	0.40	0.37	0.35
5	rs537681	G/A	349	249	0.20	0.17	0.17
6	rs471715	C/T	348	248	0.65	0.19	0.24
7	rs590688	C/G	339	243	0.21	0.50	0.49
8	rs694070	G/A	349	250	0.60	0.29	0.26
9	rs601040	A/G	341	244	0.07	0.08	0.09
10	rs7116336	A/T	345	246	1.00	0.08	0.08
11	rs572943	T/C	334	238	0.65	0.11	0.13
12	rs653752	C/G	349	249	0.29	0.35	0.35
13	rs665617	C/G	349	249	0.81	0.11	0.13
14	rs10895057	G/A	347	247	0.68	0.12	0.13
15	rs495997	G/A	347	249	0.25	0.41	0.42

16	rs660541	T/C	324	232	0.79	0.45	0.44
17	rs1042838[f]	T/G	346	246	0.77	0.20	0.21
18	rs553272	C/T	349	249	1.00	0.09	0.09
19	rs1144133	T/C	347	249	1.00	0.08	0.07
20	rs492457	G/A	347	247	0.05	0.25	0.27
21	rs11224575	A/G	347	247	0.78	0.13	0.15
22	rs569857	A/T	349	249	1.00	0.09	0.08
23	rs1042839[g]	T/C	348	248	0.79	0.19	0.19
24	rs572402	C/T	340	244	0.55	0.28	0.27
Downstream							
25	rs523535	C/T	336	241	0.94	0.33	0.29
26	rs471767	G/A	327	236	0.30	0.31	0.33
27	rs11224561	A/G	344	245	0.65	0.10	0.13
28	rs3740751	A/G	348	248	1.00	0.08	0.13
29	rs1046982	G/A	348	248	0.45	0.10	0.13
30	rs17728653	T/C	326	232	0.54	0.33	0.28

[a] Number successfully genotyped (all races) out of 350 women with >50% genotyping success.
[b] Number successfully genotyped of 250 non-Hispanic whites.
[c] Minor allele frequency in the HapMap CEU population.
[d] Minor allele frequency in non-Hispanic whites in this study.
[e] All are intronic unless otherwise noted.
[f] Functional SNP (nonsynonymous)
[g] SNP in coding region (synonymous)

Table 5.2 Overall Study Population Characteristics Compared to Women with 100% Genotyping Success[a]

	Frequency(%) N=515	Frequency(%) N=264[a]
Age at Interview		
19-25	42 (8)	23 (9)
26-30	158 (31)	89 (34)
31-35	201 (39)	90 (34)
36-41	114 (22)	62 (23)
Decade of birth		
1950's	222 (43)	112 (42)
1960's	284 (55)	147 (56)
1970's	9 (2)	5 (2)
Age at Menarche		
8-10	39 (8)	22 (8)
11	73 (14)	46 (17)
12	128 (25)	65 (25)
13	144 (28)	72 (27)
14	76 (15)	30 (11)
15-19	55 (11)	29 (11)
Race		
White	393 (76)	188 (71)
Black or African-American	71 (14)	43 (16)
Asian	14 (3)	7 (3)
Other	37 (7)	26 (10)
Ethnicity		
Non-Hispanic	478 (93)	239 (91)
Hispanic	37 (7)	25 (9)

DNA concentration[b]		
<1 ng/μL	106 (21)	29 (11)
1-3 ng/μL	139 (27)	47 (18)
3-10 ng/μL	144 (28)	80 (30)
>10 ng/μL	126 (24)	108 (41)
Education		
High school or less	95 (19)	57 (22)
Some college	205 (40)	115 (44)
College graduate	209 (41)	91 (35)
Missing[c]	6	1

[a]264 women were successfully genotyped for all 30 SNPs and were included in the LSKM analysis.
[b]More than one urine sample was extracted and genotyped for 20% of the women. This table includes only the sample with the highest concentration.
[c]Not included in percentages.

Table 5.3 LSKM Analysis for Variation in *PGR* and Age at Menarche

	N	Score statistic (df)	P-value
Adjusted for race and ethnicity	264	15.2 (9.1)	0.088
Non-Hispanic whites (NHW) only	181	17.5 (7.9)	0.025
Sensitivity Analysis (NHW only)			
-using most likely genotypes	250	17.7 (8.5)	0.029
-using least likely genotypes	250	16.8 (8.4)	0.038

Table 5.4 Associations of 8 Haplotype Blocks in *PGR* with Age at Menarche

Block	SNPs included	Haplotype (freq)[a]	Linear regression beta (SE)	Mean age at menarche[b]	Global score test p-value[c]
1	1, 2	AT 0.19	0.16 (0.19)	12.9	0.035
		TT 0.32	-0.35 (0.16)	12.4	
		TC 0.49	Referent	12.8±0.2	
2	3, 4, 5, 6, 7, 8	ACATCA 0.06	0.33 (0.32)	12.4	0.188
		AAGTCA 0.15	0.14 (0.22)	12.2	
		AAACGA 0.18	0.37 (0.21)	12.5	
		AAATGG 0.27	0.48 (0.19)	12.6	
		GCATCA 0.28	Referent	12.1±0.3	
3	11, 12, 13	TGC 0.11	0.27 (0.23)	12.7	0.644
		CCG 0.34	0.12 (0.16)	12.6	
		CGG 0.54	Referent	12.4±0.2	
4	14, 15	AG 0.29	0.19 (0.17)	12.7	0.184
		GG 0.12	-0.26 (0.23)	12.3	
		AA 0.60	Referent	12.5±0.2	
5	16, 17	CG 0.36	0.40 (0.16)	12.6	0.041
		CT 0.19	0.30 (0.19)	12.5	
		TG 0.45	Referent	12.2±0.2	
6	18, 19	CT 0.08	0.23 (0.25)	12.8	0.580
		TC 0.91	Referent	12.5±0.1	
7	20, 21, 22, 23, 24, 25	AGACCT 0.09	0.33 (0.26)	12.8	0.155
		AGTTCT 0.19	0.15 (0.20)	12.6	
		AATCTT 0.13	-0.38 (0.24)	12.1	
		GGTCTT 0.26	0.21 (0.19)	12.7	
		AGTCTC 0.34	Referent	12.5±0.2	

8	27, 28	AG	0.02	0.23 (0.50)	12.9	0.142
		AA	0.08	-0.47 (0.27)	12.2	
		GG	0.89	Referent	12.7±0.1	

[a] Only haplotypes with frequency ≥0.02 were included in this table and in regression models. Numbers may not add to 1 due to rounding or rare haplotypes.
[b] P value for heterogeneity across haplotypes. Using a conservative Bonferroni correction for multiple comparisons, the threshold p-value for alpha=0.05 is 0.05/8 = 0.0063. None of the associations were significant at this level.

Table 5.5 Genotype-Specific Mean Ages at Menarche for 30 SNPs Across the Extended Gene Region of *PGR*, Among 250 Non-Hispanic Whites

SNP #	dbSNP reference	N	Minor Homozygote Mean (SE)	Heterozygote Mean (SE)	Major Homozygote Mean (SE)	ANOVA p-value (2 df)
1	rs474320	223	14.0 (0.76)	12.5 (0.18)	12.5 (0.13)	0.034
2	rs521488	247	13.0 (0.22)	12.3 (0.13)	12.7 (0.21)	0.018
3	rs948516	233	12.6 (0.37)	12.1 (0.15)	12.9 (0.15)	<0.001
4	rs550778	240	12.5 (0.32)	12.4 (0.13)	12.9 (0.17)	0.112
5	rs537681	249	12.4 (0.70)	12.5 (0.17)	12.6 (0.12)	0.673
6	rs471715	248	14.1 (0.86)	12.5 (0.16)	12.6 (0.12)	0.027
7	rs590688	243	13.0 (0.20)	12.5 (0.13)	12.4 (0.21)	0.027
8	rs694070	250	13.3 (0.38)	12.6 (0.14)	12.5 (0.15)	0.081
9	rs601040	244	12.5 (1.85)	12.4 (0.27)	12.6 (0.11)	0.812
10	rs7116336	246	12.0 --	12.4 (0.20)	12.6 (0.11)	0.675
11	rs572943	238	13.0 (1.00)	12.7 (0.23)	12.5 (0.11)	0.615
12	rs653752	249	12.7 (0.35)	12.6 (0.13)	12.5 (0.17)	0.875
13	rs665617	249	13.0 (1.00)	12.8 (0.22)	12.5 (0.11)	0.566
14	rs10895057	247	12.0 (0.00)	12.3 (0.19)	12.7 (0.12)	0.338
15	rs495997	249	12.9 (0.30)	12.5 (0.12)	12.7 (0.18)	0.266
16	rs660541	232	12.3 (0.21)	12.4 (0.14)	13.0 (0.21)	0.013
17	rs1042838	246	13.8 (0.84)	12.5 (0.16)	12.6 (0.12)	0.086
18	rs553272	249	10.5 (1.5)	13.0 (0.22)	12.5 (0.11)	0.026
19	rs1144133	249	10.5 (1.5)	13.0 (0.24)	12.5 (0.11)	0.032
20	rs492457	247	13.7 (0.82)	12.6 (0.14)	12.5 (0.13)	0.100

21	rs11224575	247	12.0 (0.00)	12.2 (0.19)	12.8 (0.11)	0.033
22	rs569857	249	10.5 (1.50)	13.1 (0.24)	12.5 (0.11)	0.015
23	rs1042839	248	13.8 (0.84)	12.5 (0.16)	12.6 (0.12)	0.089
24	rs572402	244	13.2 (0.50)	12.5 (0.15)	12.5 (0.14)	0.193
25	rs523535	241	12.8 (0.29)	12.3 (0.14)	12.7 (0.17)	0.108
26	rs471767	236	12.8 (0.44)	12.6 (0.14)	12.5 (0.15)	0.682
27	rs11224561	245	12.0 --	12.3 (0.22)	12.7 (0.11)	0.377
28	rs3740751	248	12.0 --	12.1 (0.25)	12.6 (0.11)	0.128
29	rs1046982	248	12.5 (0.29)	12.2 (0.24)	12.7 (0.11)	0.167
30	rs17728653	232	12.9 (0.27)	12.4 (0.15)	12.7 (0.16)	0.149

Table 5.6 Mean Ages at Menarche According to rs1042838 Genotype

				Linear regression models Beta±SE (TT vs. other) Wald p-value	
Genotype	GG	TG	TT	Adjusted for race and ethnicity	Adjusted for race, ethnicity, and BMI (kg/m^2)
Sample size	N=275	N=107	N=13		
Mean age at menarche±SE	12.6±0.1	12.6±0.1	13.6±0.5	0.96±0.47 p=0.04	0.87±0.44 p=0.05

Table 5.7 Associations of rs1042838 Genotype with Menstrual Cycle Characteristics

Genotype	GG	TG	TT	Logistic regression models[a]	
N=395	N=275	N=107	N=13	Dose-response (per additional T allele)	
				Odds ratio (95% CI)	Score test p-value
Mean cycle length (days)[b]	28.9±0.2	29.1±0.3	30.2±0.9	--	--
SD of cycle length (mean±SE of SD)	2.0±0.1	1.9±0.1	1.6±0.2	0.81 (0.49, 1.34)	0.42
n (short cycles)/ N(total cycles) (%)	237/1775 13.4%	54/715 7.6%	1/63 1.6%	0.54 (0.36, 0.80)	0.002
n (long cycles)/ N (total cycles) (%)	182/1775 10.3%	61/715 8.5%	5/63 7.9%	0.91 (0.63, 1.33)	0.64

[a]Adjusted for race and ethnicity. Generalized estimating equations were used for modeling the probability of long and short cycles.
[b]Means are adjusted for within-women correlation across cycles.

FIGURES

Figure 5.1 Population and Genotyping Flowchart for *PGR* Study

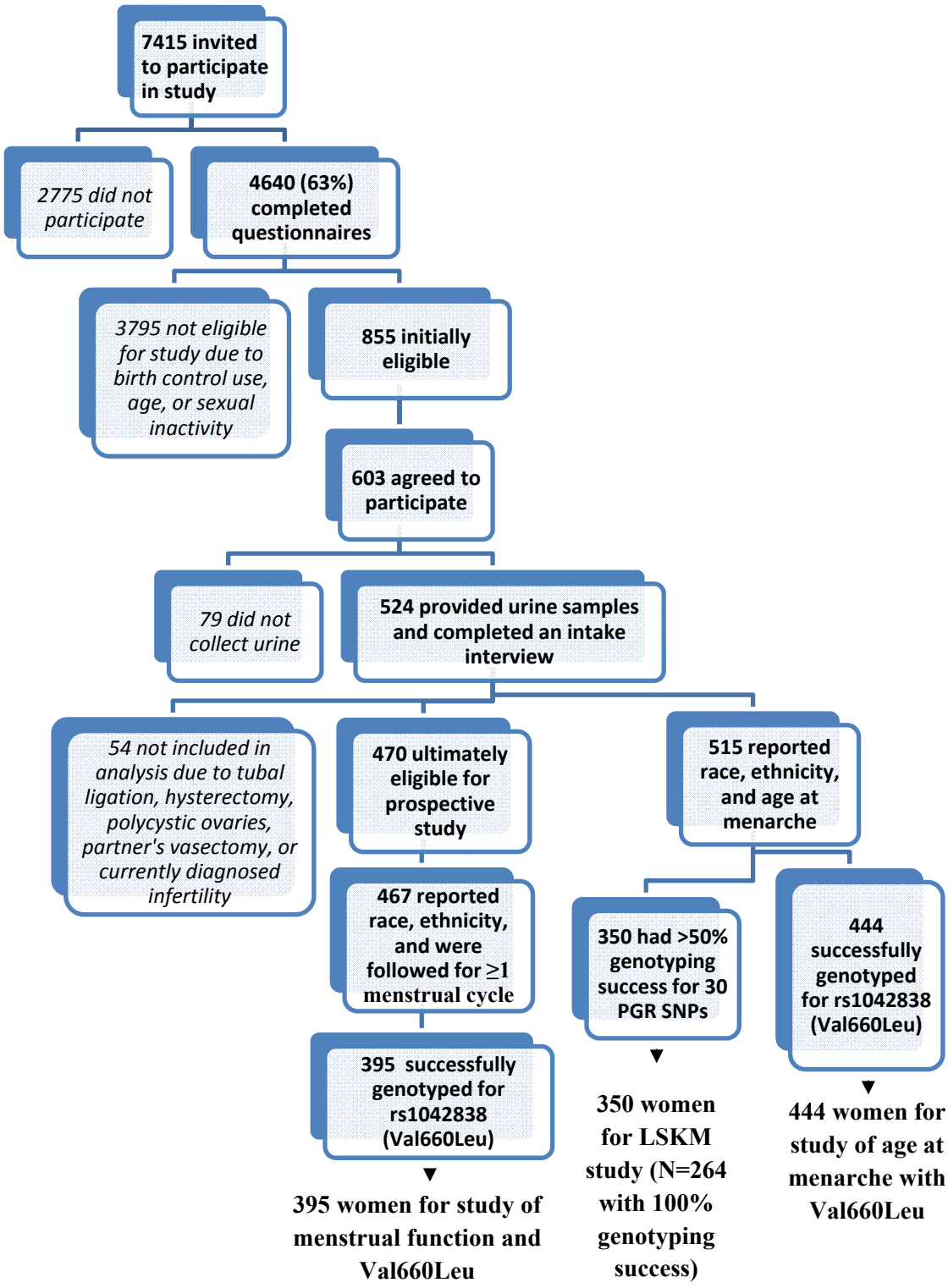
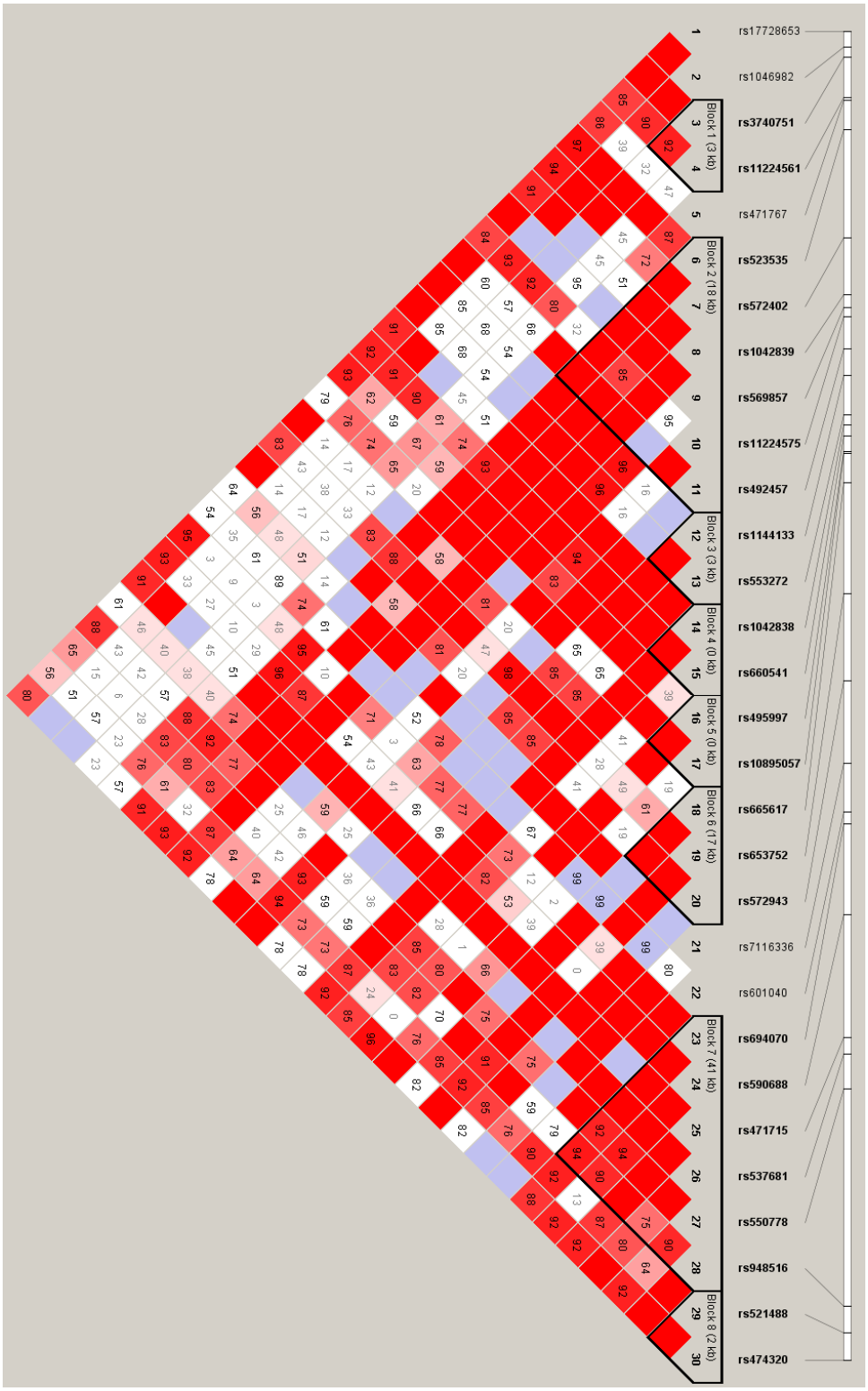


Figure 5.2 Linkage Disequilibrium Pattern and Haplotype Blocks for 30 SNPs Genotyped in the Extended Gene Region of *PGR*, Among 250 Non-Hispanic Whites



5.8 References

1. Guo Y, Shen H, Xiao P, et al. Genomewide linkage scan for quantitative trait loci underlying variation in age at menarche. *Journal of Clinical Endocrinology and Metabolism* 2006;**91**(3): 1009-14.

2. Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM. Heritability of age at menarche in girls from the Fels Longitudinal Study. *American Journal of Physical Anthropology* 2005;**128**(1): 210-9.

3. Taylor KC, Small CM, Epstein MP, Terrell ML, Marcus M. Heritability of age at menarche and menstrual cycle length in a population exposed to a brominated flame retardant. . *International Journal of Child and Adolescent Health* 2009;**2**(3).

4. Leite DB, Junqueira MG, de Carvalho CV, et al. Progesterone receptor (PROGINS) polymorphism and the risk of ovarian cancer. *Steroids* 2008;**73**(6): 676-80.

5. Johnatty SE, Spurdle AB, Beesley J, et al. Progesterone receptor polymorphisms and risk of breast cancer: results from two Australian breast cancer studies. *Breast Cancer Research and Treatment* 2007;**109**(1).

6. Wang-Gohrke S, Chang-Claude J, Becher H, Kieback DG, Runnebaum IB. Progesterone Receptor Gene Polymorphism Is Associated with Decreased Risk for Breast Cancer by Age 50. *Cancer Research* 2000;**60**(9): 2348-50.

7. De Vivo I, Huggins GS, Hankinson SE, et al. A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. *Proceedings of the National Academy of Sciences* 2002;**99**(19): 12263-8.

8. Romano A, Lindsey PJ, Fischer D-C, et al. Two functionally relevant polymorphisms in the human progesterone receptor gene (+331 G/A and progins) and the predisposition for breast and/or ovarian cancer. *Gynecologic Oncology* 2006;**101**(2): 287-95.

9. Pooley KA, Healey CS, Smith PL, et al. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. *Cancer Epidemiology, Biomarkers & Prevention* 2006;**15**(4): 675-82.

10. Romano A, Delvoux B, Fischer D-C, Groothuis P. The PROGINS polymorphism of the human progesterone receptor diminishes the response to progesterone. *Journal of Molecular Endocrinology* 2007;**38**(2): 331-50.

11. Terry KL, De Vivo I, Titus-Ernstoff L, Sluss PM, Cramer DW. Genetic Variation in the Progesterone Receptor Gene and Ovarian Cancer Risk. *American Journal of Epidemiology* 2005;**161**(5): 442-51.

12. Schweikert A, Rau T, Berkholz A, Allera A, Daufeldt S, Wildt L. Association of progesterone receptor polymorphism with recurrent abortions. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 2004;**113**(1): 67-72.

13. Jabbour HN, Kelly RW, Fraser HM, Critchley HOD. Endocrine Regulation of Menstruation. *Endocrine Reviews* 2006;**27**(1): 17-46.

14. Marcus M, Gerr F. Upper extremity musculoskeletal symptoms among female office workers: associations with video display terminal use and occupational psychosocial stressors. *American Journal of Industrial Medicine* 1996;**29**(2): 161-70.

15. The International HapMap Consortium: The International HapMap Project. *Nature* 2003;**426**(6968): 789-96.

16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**(2): 263-5.

17. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nature Genetics* 2005;**37**(11): 1217-23.

18. Beckman Coulter I. GenomeLab SNPstream Genotyping System 48-plex Training Guide. In: Beckman Coulter I, editor. Fullerton, CA: Beckman Coulter, Inc.; 2005.

19. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics* 2005;**76**(5): 887-93.

20. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;**296**(5576): 2225-9.

21. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score Tests for Association between Traits and Haplotypes when Linkage Phase Is Ambiguous. *The American Journal of Human Genetics* 2002;**70**(2): 425-34.

22. Kwee LC, Liu D, Lin X, Ghosh D, Epstein MP. A powerful and flexible multilocus association test for quantitative traits. *American Journal of Human Genetics* 2008;**82**(2): 386-97.

23. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 2001;**68**(4): 978-89.

24. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 2003;**73**(5): 1162-9.

25. Tam CS, de Zegher F, Garnett SP, Baur LA, Cowell CT. Opposing influences of prenatal and postnatal growth on the timing of menarche. *Journal of Clinical Endocrinology and Metabolism* 2006;**91**(11): 4369-73.

26. Frisch RE, McArthur JW. Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance or onset. *Science* 1974;**185**(4155): 949-51.

27. Maclure M, Travis LB, Willett W, MacMahon B. A prospective cohort study of nutrient intake and age at menarche. *American Journal of Clinical Nutrition* 1991;**54**(4): 649-56.

28. Stikkelbroeck NMML, Sweep CGJ, Braat DDM, Hermus ARMM, Otten BJ. Monitoring of menstrual cycles, ovulation, and adrenal suppression by saliva sampling in female patients with 21-hydroxylase deficiency. *Fertility and Sterility* 2003;**80**(4): 1030-6.

29. Agoulnik IU, Tong XW, Fischer DC, et al. A germline variation in the progesterone receptor gene increases transcriptional activity and may modify ovarian cancer risk. *Journal of Clinical Endocrinology and Metabolism* 2004;**89**(12): 6340-7.

30. Dan Tong, Gerhild Fabjani, Georg Heinze, Andreas Obermair, Sepp Leodolter, Robert Zeillinger. Analysis of the human progesterone receptor gene polymorphism progins in Austrian ovarian carcinoma patients. *International Journal of Cancer* 2001;**95**(6): 394-7.

31. Breast Cancer Association C. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. *Journal of the National Cancer Institute* 2006;**98**(19): 1382-96.

32. Pearce CL, Wu AH, Gayther SA, et al. Progesterone receptor variation and risk of ovarian cancer is limited to the invasive endometrioid subtype: results from the ovarian cancer association consortium pooled analysis. *British Journal of Cancer* 2008;**98**(2): 282-8.

33. Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL, Smith MT. Accuracy of Multiplexed Illumina Platform-Based Single-Nucleotide Polymorphism Genotyping Compared between Genomic and Whole Genome Amplified DNA Collected from Multiple Sources. *Cancer Epidemiology Biomarkers and Prevention* 2006;**15**(12): 2533-6.

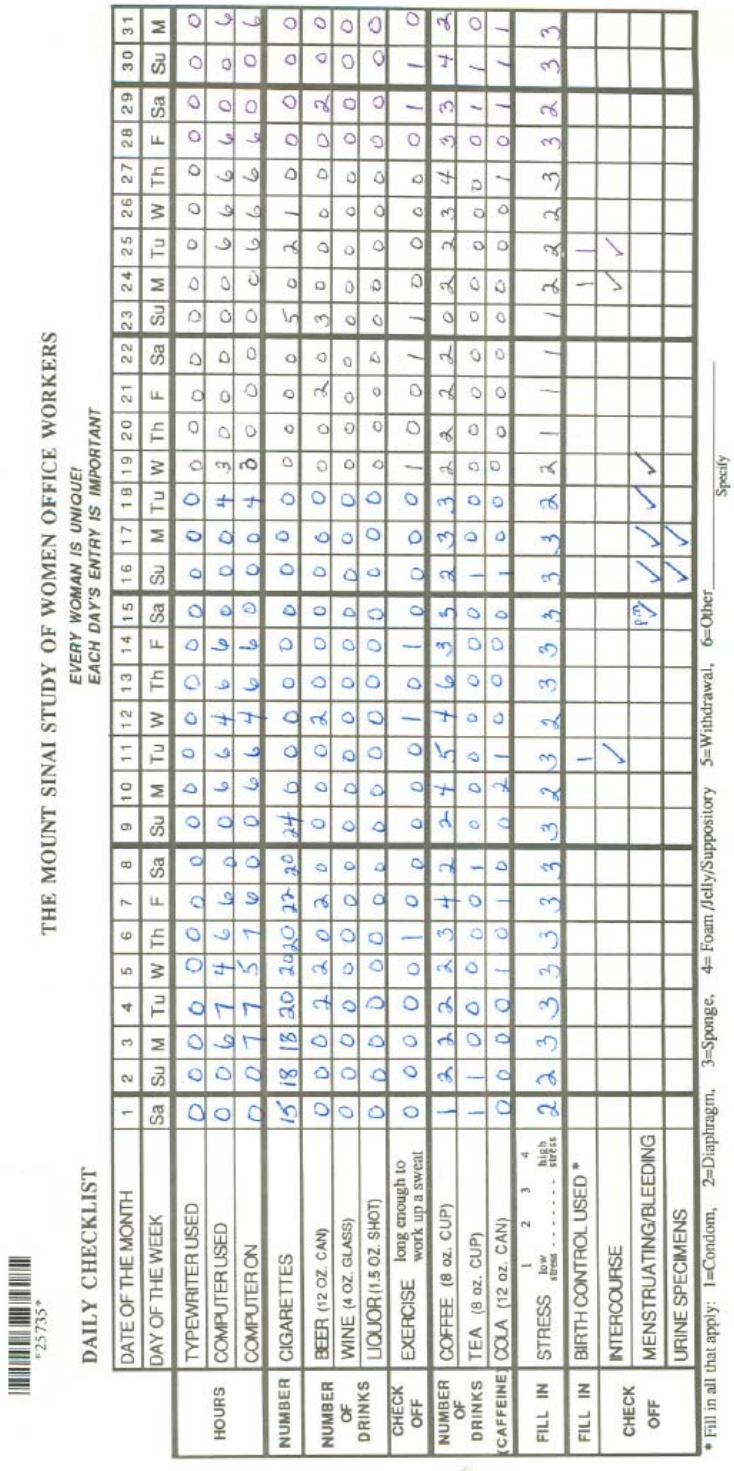
34. van der Hel OL, van der Luijt RB, Bas Bueno de Mesquita H, et al. Quality and Quantity of DNA Isolated from Frozen Urine in Population-Based Research. *Analytical Biochemistry* 2002;**304**(2): 206-11.

35. McDowell MA, Brody DJ, Hughes JP. Has age at menarche changed? Results from the National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Journal of Adolescent Health* 2007;**40**(3): 227-31.

5.9 Appendices

Appendix 5.A

The Daily Diary Used in the Mount Sinai Study of Women Office Workers



25735

THE MOUNT SINAI STUDY OF WOMEN OFFICE WORKERS

EVERY WOMAN IS UNIQUE!
EACH DAY'S ENTRY IS IMPORTANT

DAILY CHECKLIST

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	DATE OF THE MONTH																															
	DAY OF THE WEEK	Sa	Su	M	Tu	W	Th	F	Sa	Su	M	Tu	W	Th	F	Sa	Su	M	Tu	W	Th	F	Sa	Su	M	Tu	W	Th	F	Sa	Su	M
HOURS	TYPEWRITER USED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	COMPUTER USED	0	0	6	7	4	6	6	0	0	6	6	4	6	6	6	0	0	4	3	0	0	0	0	0	6	6	6	6	0	0	6
	COMPUTER ON	0	0	7	7	5	7	4	0	0	6	6	7	6	6	6	0	0	4	8	0	0	0	0	0	6	6	6	6	0	0	6
NUMBER	CIGARETTES	15	18	18	20	20	20	22	20	24	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	2	1	0	0	0	0	0
NUMBER OF DRINKS	BEER (12 OZ. CAN)	0	0	0	2	2	0	2	0	0	0	0	3	0	0	0	0	0	6	0	0	0	3	0	3	0	0	0	0	2	0	0
	WINE (4 OZ. GLASS)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	LIQUOR (1.5 OZ. SHOT)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHECK OFF	EXERCISE long enough to work up a sweat	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	1	0
NUMBER OF DRINKS	COFFEE (8 oz. CUP)	1	2	2	2	2	3	4	2	2	4	5	5	4	6	3	3	3	3	2	2	2	2	0	2	2	3	4	3	3	4	2
	TEA (8 oz. CUP)	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
CAFFEINE	COLA (12 oz. CAN)	0	0	0	0	1	0	1	0	0	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1
FILL IN	STRESS 1 2 3 4 low stress ..... high stress	2	2	3	3	3	3	3	3	3	2	3	2	3	3	3	3	3	2	1	1	1	1	1	2	2	2	3	3	2	3	3
FILL IN	BIRTH CONTROL USED *											1													1	1						
CHECK OFF	INTERCOURSE											✓													✓	✓						
	MENSTRUATING/BLEEDING																✓	✓	✓	✓												
	URINE SPECIMENS																✓	✓														

* Fill in all that apply: 1=Condom, 2=Diaphragm, 3=Sponge, 4= Foam /Jelly/Suppository 5=Withdrawal, 6=Other_____ Specify

Appendix 5.B

Linkage Disequilibrium-Tagging SNPs in the Extended Gene Region of *PGR*

dbSNP reference	Position on Chromosome 11 (reference assembly)	MAF in HapMap CEU	Tags SNPs	Passed/ failed genotyping
rs17728653	100404528	0.33	rs1870019, rs17728653	P
rs481883	100405345	0.09	rs481883	F
rs1046982	100405926	0.10	rs1046982, rs11224556	P
rs3740751	100406809	0.08	rs3740751	P
rs561610	100408204	0.29	rs561610	F
rs11224561	100410266	0.10	rs11224561	P
rs471767	100410507	0.31	rs471767	P
rs523535	100413084	0.33	rs523535	P
rs11224563	100414396	0.20	rs10895054, rs11224563	F
rs11571271	100414774	0.05	rs11571271	F
rs572402	100422465	0.28	rs547378, rs606789, rs478850, rs504372, rs500760, rs542491, rs1217841, rs572698, rs572402, rs492827, rs563656, rs511298, rs1145460, rs550382, rs1217839, rs601046, rs588913, rs592080, rs572580, rs511484, rs558959, rs546763, rs504402, rs1217840, rs518382	P
rs1042839	100427412	0.19	rs1042839	P
rs569857	100428510	0.09	rs569857	P
rs11224575	100429243	0.13	rs1379131, rs11224575	P

rs492457	100432070	0.25	rs502465, rs578029, rs519792, rs543936, rs503362, rs577615, rs1218633, rs492457, rs666207, rs552916, rs545845, rs660149, rs679275	P
rs1144133	100434397	0.08	rs936269, rs547565, rs1144133	P
rs553272	100437813	0.09	rs553272	P
rs1042838	100438622	0.20	rs11224567, rs10895055, rs1145465, rs568801, rs502471, rs559700, rs585447, rs548711, rs507141, rs1545611, rs1042838, rs12364291, rs11571259, rs9282823, rs505819, rs541463, rs11571201, rs673943, rs12365216, rs562894, rs482765	P
rs660541	100439577	0.45	rs540622, rs660541, rs545835, rs635984	P
rs495997	100440990	0.41	rs495997	P
rs10895057	100441139	0.12	rs1824128, rs10895057	P
rs665617	100443654	0.11	rs665617	P
rs666553	100443878	0.13	rs1456765, rs538915, rs503602, rs508653, rs666553, rs7106686, rs516693	F
rs653752	100453320	0.35	rs653752, rs508533, rs486992, rs529359	P
rs572943	100460828	0.11	rs572943	P
rs7116336	100468083	0.08	rs11224579, rs11224580, rs10895058, rs7116336	P
rs601040	100472262	0.08	rs518162, rs601040, rs582691	P

rs578938	100472265	0.30	rs1824126, rs11224589, rs555572, rs11571171, rs10895063, rs481855, rs572483, rs11224592, rs523323, rs11224598, rs578938, rs10895065	F
rs694070	100473333	0.29	rs694070	P
rs590688	100481184	0.50	rs555653, rs590688	P
rs47171	100491807	0.19	rs520017, rs565186, rs471715	P
rs537681	100493244	0.17	rs618032, rs501732, rs543215, rs542384, rs537681	P
rs550778	100496255	0.37	rs550778, rs485283, rs480851, rs493957, rs1145463, rs619487, rs560291, rs493220, rs506487, rs596223	P
rs948516	100515143	0.29	rs948516, rs4754732	P
rs521488	100517417	0.49	rs521488	P
rs499590	100519136	0.30	rs499590	F
rs474320	100519759	0.19	rs474320	P

Appendix 5.C

SNPs and Primers Chosen for *PGR* Genotyping for the Beckman SNPstream System

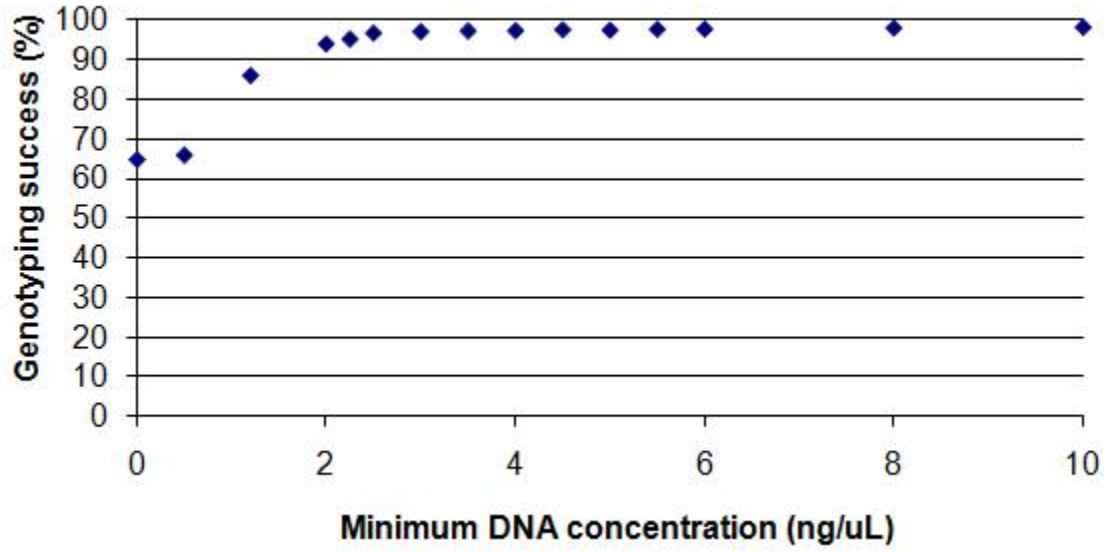
dbSNP reference	Upper strand primer	Lower strand primer	Extension primer
rs17728653	CAATAAGAAACAAGTATTTTGACATTTC	ACCATGTTAGGTCTTGGAGACA	ATACCTACCACGCTACAGCCTCAACTCAAACTTACAGCAAGAATC
rs481883	AAACACTGTGTAGTTGGTTTCAA	ATTATAGAGAAAATATCCTTGACTGGG	AGCCGAACTACCACTGAGTATTGATTGCCTGAGAATCACTCTTTG
rs1046982	AATCTAAAGTCATACACCTTGCTCC	CCAGCCCCAGGCATACAC	ACGTAAGACCACTCAAGACCGACAATTGTTCTGAAGGTTTTTGCC
rs3740751	ATTCTAAAACCTGACAGAAGCCT	ATTTGTCCTCTTTGCCCTACA	CACGACAAGACAACAGATACAGCACATAGCAAGAGGAAGTGAGAG
rs561610	AATTTATATGGTGTATTTCATCTCCTTT	AGAACATTTTTGAGGAAGTTGC	AGACTTCTACGCAAGCACTGTTGCACTATTTTGGTGAAAATGATG
rs11224561	AGCAGTCCTGCAACAGTCTT	AGAGCTCAGGTCACAGGC	CAAGACCGCAACTAGATACAACAGGGGCTGCGCCCAAGCTTGTCC
rs471767	TTTAACTTACTACCAACACCCCC	TAATAGAACCTACACTTCTAAAGTTCGG	CACTACATACGACCGCAGAAGGTGGACAAATTATTGAAGAAAACT
rs523535	TTCTTGGGATGACATTTGTG	ATAACACCTTTATGTGTCAGTAATTCAT	TCCAGAATAGACAACAGACGAGTAATTCATTTGCATTAAATCTTA
rs11224563	TGGAAACACAAATCTCTTAGGAA	TCCCCAAAAAAGGTTATTTTAAA	AACATCCACGCAACTCATACTTAATAAGTAGATTCATATCATTAT
rs11571271	TGAAGAATTTACAAGATTGAAAAAGTAC	GATGATTTGTTAGATGCAAAAGTTAA	GATCCATCAACAGACATCACTTAAAAATCCTCACCTACATGGTAT
rs572402	GCAAAGTAAAACTTTGATAATTTGGT	TAAAAATGTTTGTTCAACCTACTGTC	CGCAGAAGCAACTCACTTCTTGTCTATTCCTGGAGATTTATATTT
rs1042839	TATTCTTGGATGAGCTTAATGGT	ATTTAGTATTAGATCAGGTGCAAAATACA	CCACTCAACTCCACGAATACGTCTAGGATGGAGATCCTACAAACA

rs569857	TTTAGAGATGTAGGCTGGGTCA	TAAGAATGATCCATCTACAATGCA	GCAGACAACGAACAACTACCATCATCCAAGGTCTTGCAAGTCTAT
rs11224575	AAAGTTATTTCCTGACAATGTTAATC	TATTCACATGAAGATGGGACG	ACAATCAACATACGAACAGCTGTCAAGGTAGGTAGCAAAAAAGAA
rs492457	ATTCATTACTTTATGCTGAAACAAAG	ATTTTGGTCATCTCAAATGTACTAAGT	ACAACTACCGACGACAAGACGAACACAGAGATTAAATTCCCA
rs1144133	AAAAAATCCTCTGTCAAGATGC	ATACTTTATTCTTTGTGTCATTTCACC	ATCTAACGCACCTACGACCTCACTGACTCAAAAGTTGTTTTGGCA
rs553272	AAAAATAGCTGATAAGTAAAGTGACTGA	CTAATTTATACCCAAAGAGAGTATATCG	CAACAAGACATAACAACGCATCTTCTGTGATGATATAATCCAAGT
rs1042838	AATAAAGTCAGAGTTGTGAGAGCAC	AAAAAGTGAATCTCTGGCTTAGG	CGATCACCTCACTAGAACAAGGATGCTGTTGCTCTCCCACAGCCA
rs660541	TTGAGGAGAATTAGAGAAAGTTGTG	ATTGCAACCTCTGCTTAGATAATT	ACCGCACTAAGCAATGTATCAGCTACTTCATACTGATAAGTAGCC
rs495997	GTGTGTAGTAGGCTATACCATCTAAGTC	AAGCTATGTTCTCTGAAGGCAA	CAGCCATCCATTCACTATCTGCGTAAATGCATTCTATGATGTTCA
rs10895057	CAGGCTTGCTGCCCTGTA	AGTCCACTGTGTAGTTAGCAGTTTATAA	CACCGCTATCAACAGACTTGGATAGGGGCTAGGTCCATTTCACTC
rs665617	ATGTAAAATAATGTTCACAATTGAACTG	TACACTGTATGATTTCATTGTATGAGTATC	CCAGATCCTCACCATGTAAGTCCCATTTATTTATTCTCATGTATT
rs666553	ATTTCTATGGAGTAGAAAAAAACATCTC	TCTATATCTTTAATTTGGAGTTGGTCA	AGTAGCCTAACAGCACTCGAGTTGGTCATTTTCTTATATTATTCT
rs653752	TTGCCTTCCATTTTTCAGG	TCTAGCATCTTTGTTTTCTTTGAAG	GCAAGCCATCAGCTAATACACACCTCCAATTTCCAGATCAAAACA
rs572943	AGTGTTGAGGGCGTATGTATTTA	TTGCTATTCTTAGATTAGATAAAACAGACTT	CTCAGACTACGAATCCACGTTAAATCTTCTTATTGAACTCTTTAT
rs7116336	ATAGCTGAAGAATCAATCCTAGGG	TACAACTAATGCTTATCAGAAACGTG	AGACCGACAAGCAATCTACAATCTTTTGCAAGAGATTTCTGTTTA

rs601040	TTTATATTCTGGTTTTGAGAAACACA	TTTAATATAGTGACGTGTTTTACACTTGT	CAAGCAACGACCTACTACAAAAATTGCGTATTATTTCTTCTCCTC
rs578938	TTAAAGAAATCTGGAAGCTCAATT	TTTAAGCCAAAACATTGCTTG	CAGAATAGCCACGCCTAGATACTCTAACTTTTCATCATCTTTAAC
rs694070	TATAGACACAAAGTATACAATAGGAATACATT	CTCATCTGGTCTTACCTGACTG	GCAACATAAGACCGCTCAACAATGTGCAGTTGGTACACAGAAGAA
rs590688	CACTGTAAGAAATTTGATTTTTCTCTAA	TTACCTTCCTAGATTCCAGAGGT	TACCTATGACCAGCAAGCACCCACTGGAGACACTGGAGCAGACGG
rs471715	GGCTGCTACATTGCAAAAA	TTCAGCTAAGAGATTTTGGACAT	AACATACAGACGCACTCCTCCCAGCAGGCACCACTCTCAACATAG
rs537681	TAATCAGTAATGCTTATTAGTGAGTTGAG	AACAGGAAGAAAAATGAACAAAAA	TACAAGCACGCACTAGACATAAAAAATATTAAGACAGTTCACAAA
rs550778	GAAGTATTTTGATACAGTTCTAAATGAACTAG	ATACATTTTCATCTTCAGTCCCA	AGCAAGACCACCTAGACCAGTCACTAATTTATTCACTTCAGATAC
rs948516	TTGGGAATTCAACAAATAGTGTC	AAAATAAGTGGGCTATTAAATTCATTC	CCGCCAGTAAGACCTAGACGTAAGTGATACTTAATCTAAACTGAG
rs521488	TTTATTTGTATAAAGTGCAGCAAGA	AGCTCAGCTTTAAGAAAAGGTCTT	ACAACTCACGCAAGTACCATGCTTGTAAAATTGGCTTTGATGGAA
rs499590	AAATGGTCTGGTGATAGGG	TATTACTATTGAGAAGGACATTGGATT	CACTAGTCATAACGCAGCCTTTTTTTTTTTTTGCACTTAATC
rs474320	ATTTCTGTGATTACAGGAGTTTTAATTC	AAAGAATGAGAGAAACAGTTTATTGG	CCATAACAACTTACCAGCCAAACTTATTTCAAAGCCTATACCATG

Appendix 5.D

Genotyping Success According to DNA Concentration Cutoffs



Appendix 5.E

Mean Age at Menarche According to Race and Ethnicity

N=515	Non-Hispanic white N=382	Non-Hispanic black N=69	Non-Hispanic Asian N=13	Hispanic N=37	Non-Hispanic Other N=14
Mean age at menarche ±SE	12.7±0.1	12.5±0.2	12.8±0.4	12.8±0.3	12.9±0.6

Appendix 5.F

Mean Age at Menarche During the Study Period According to Year of Birth

Birth year N=515	1950-1954 n=64	1955-1959 n=158	1960-1964 n=183	1965-1972 n=110
Mean age at menarche ±SE	12.7±0.2	12.5±0.1	12.8±0.1	12.7±0.1

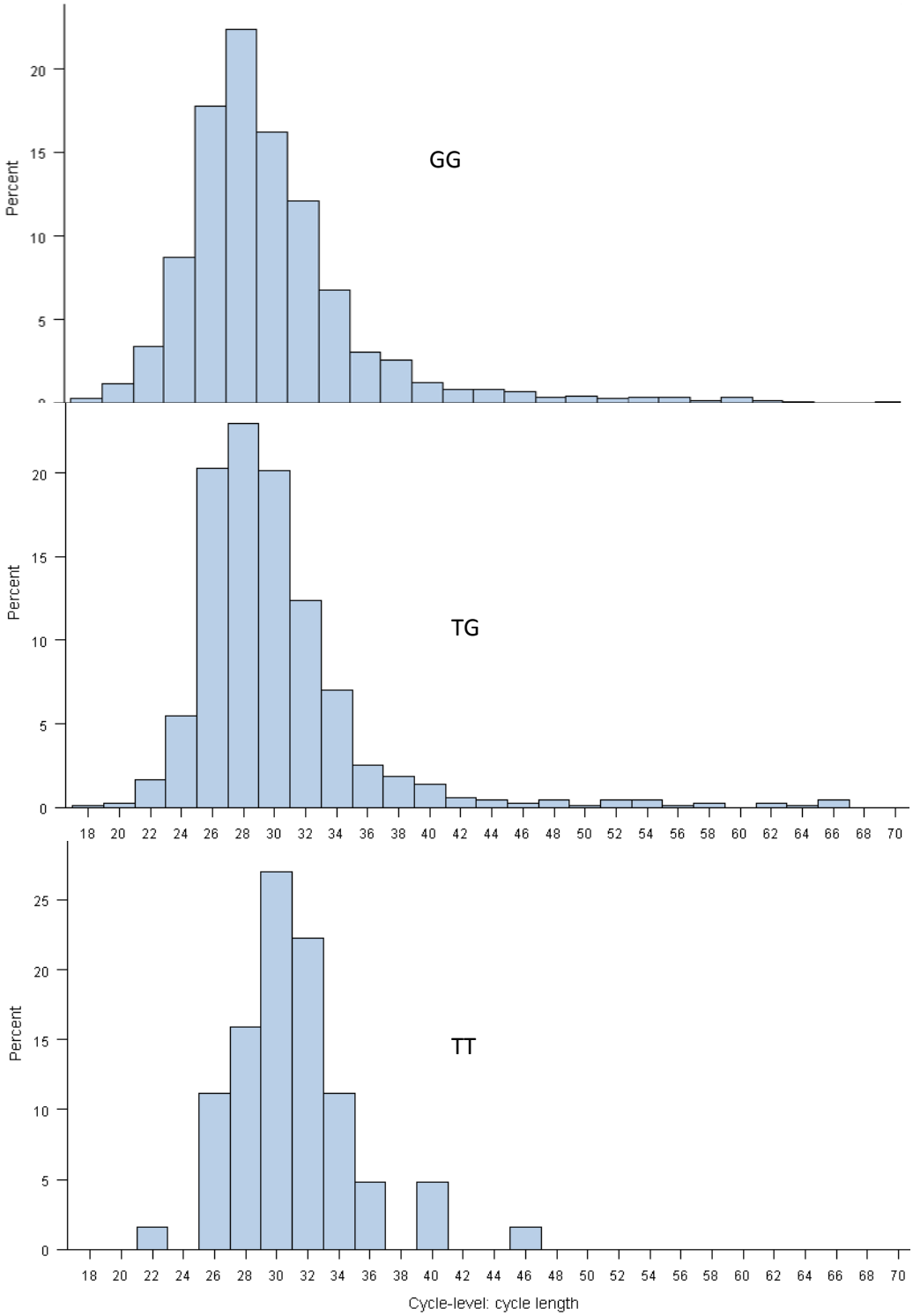
Appendix 5.G

Multiple Linear Regression Results for the Effect of the rs1042838 Variant (Homozygous TT vs. Other) on Age at Menarche

Variable	Beta±SE	p-value
rs1042838		
GT or GG	Referent	
TT	0.94±0.46	0.04
BMI		
<25	Referent	
25-<30	-0.52±0.20	<0.01
≥30	-0.51±0.24	0.04
Race		
White	Referent	
Black or African-American	-0.07±0.23	0.78
Other	0.01±0.52	0.99
Hispanic ethnicity		
No	Referent	
Yes	-0.06±33	0.87

Appendix 5.H

Distribution of Menstrual Cycle Lengths According to rs1042838 Genotype



Chapter 6

Associations of Alcohol, Smoking, and Caffeine with Fecundability: Effect Modifications by *NAT2*

6.1 Abstract

Background. The enzyme N-acetyltransferase 2 (*NAT2*) is responsible for metabolizing and detoxifying xenobiotics such as caffeine, tobacco smoke, pesticides, and prescription drugs. Common polymorphisms in the *NAT2* gene determine haplotypes that have slow or fast acetylator phenotypes and follow distinct metabolic pathways. Caffeine, alcohol and smoking are candidate risk factors for infertility. The purpose of this study is to investigate the effects of these exposures on time to pregnancy, and determine whether the effects are modified by *NAT2*.

Methods. The population consisted of 470 women office workers ages 20-41 who were at risk for pregnancy. Fertility was measured by counting the menstrual cycles until a pregnancy occurred (time to pregnancy). Exposure and covariate information were collected in an interview and in daily diaries. Urine samples were collected and served as the source of DNA. Three *NAT2* polymorphisms (rs1799929, rs1799930, and rs1208) were genotyped in 319 women using the Beckman SNPstream system. Discrete survival analysis was used to determine whether *NAT2* haplotypes modified any effects of alcohol, smoking, or caffeine on time to pregnancy.

Results. The 319 women were each followed for an average of 8 menstrual cycles, resulting in 124 pregnancies. 161 women carried two copies of the slow haplotype and

thus were identified as slow acetylators. There was no effect of caffeine on time to pregnancy in this population, regardless of haplotype. Heavy alcohol use (>1 drink/day) and smoking were significant predictors of longer time to pregnancy, but only among slow acetylators. These associations and interactions remained after adjustment for potential confounders.

Conclusion. This study demonstrates the importance of including genetic information about relevant metabolic enzymes when studying the effects of xenobiotics on human health. This is of particular importance when estimating the effects of these exposures among those with varying levels of genetic susceptibility.

6.2 Introduction

Longer time to pregnancy is an indication of subfertility. Subfertility and infertility are growing public health issues because women are waiting until older ages to attempt pregnancy. Smoking, alcohol, and caffeine are modifiable lifestyle risk factors that have long been suspected as being related to reduced fertility, but results for these exposures have been inconsistent in previous studies. Caffeine was found to reduce fecundability in a prospective study by Wilcox et al. (**1**), where consuming more than 3150 mg/month (about 1 cup of coffee/day) was associated with a 50% reduction in fecundability. Other prospective studies have found either no clear effect (**2**) or even increased fecundability for tea consumption or for moderate levels of caffeine intake (**3, 4**). A smoking-caffeine interaction may be important; a detrimental effect of caffeine was observed among non-smokers but not among smokers in two studies (**5, 6**), possibly because smoking induces CYP1A2, a liver enzyme responsible for metabolizing caffeine (**7**).

Results for alcohol and smoking have been more consistent; typically, both are associated with increased time to pregnancy, although some studies have found no effect. Among prospective pregnancy studies, Buck Louis et al. (**8**) found a fecundability odds ratio (FOR) of 0.96 (95% CI, 0.93 to 0.99) for each alcoholic beverage consumed in a 28-day period, and Jensen et al. (**9**) observed a dose response between alcohol intake and increased time to pregnancy from as little as 1-5 drinks per week. Wilcox (**1**) did not find an association between alcohol and time to pregnancy. Jensen et al. (**10**) also found reduced fecundability associated with a woman's smoking (FOR: 0.67 (0.42, 1.06),

especially if she was also exposed to her mother's smoking *in utero* (FOR: 0.53 (0.31,0.91)). However, Buck Louis et al. (2008) reported no effect of smoking in their study.

Reasons for inconsistencies among studies may be methodological (e.g., retrospective vs. prospectively collected data); a result of using differing categorization schemes for exposures or ways of defining subfertility; the inclusion or exclusion of subclinical spontaneous abortions; or adjustment for a varying number of potential confounders. Studies of one of these exposures (e.g., smoking) have not always controlled for the others (e.g., alcohol and caffeine), resulting in a mixing of effects. Finally, there may be biological effect modification resulting from genetic heterogeneity in the ability to metabolize and detoxify these compounds.

N-acetyltransferase-2 (NAT2) is an enzyme that catalyzes the N-acetylation of aromatic and heterocyclic amines (**11**). NAT2 is highly expressed in the liver and gut and is responsible for metabolizing and detoxifying xenobiotics such as caffeine, tobacco smoke, cooked meat, pesticides, and prescription drugs (**12-15**). Common polymorphisms in the *NAT2* gene determine haplotypes that correspond with slow or rapid acetylator phenotypes (**16-18**). The slow acetylator phenotype is at least partly due to reduced protein levels (**19**). The ratio of caffeine metabolites depends on NAT2 acetylator status (**20**).

It has been shown that *NAT2* haplotypes may modify associations between toxins and various outcomes. For example, slow acetylators are especially susceptible to hepatotoxicity during prescription drug treatment (**21**). Acetylator status also modified

the association between smoking and risk of colorectal adenoma, with slow acetylators being more susceptible (**22**). Slow acetylator genotype is also a risk factor for bladder cancer, likely by decreasing the ability of the liver to detoxify cigarette smoke or occupational exposure to arylamines (**23**); in addition, slow status may increase the effect of smoking on bladder cancer (**24**). No studies to date have investigated *NAT2* polymorphisms as risk factors or as effect modifiers of fecundability.

This study attempts to address the inconsistencies and limitations of previous studies by using prospectively collected daily data on caffeine, alcohol, and smoking; prospectively measuring time to pregnancy, including subclinical pregnancies; and investigating genetic heterogeneity by looking at whether *NAT2* acetylator status (rapid vs. slow) modifies any of these effects.

6.3 Methods

6.3.1 Population

The Study of Women Office Workers was a prospective study conducted from 1990-1994 to investigate the effects of computer terminal usage on musculoskeletal characteristics and fertility (**25**). A total of 4640 women completed questionnaires to determine eligibility in the study (see flowchart, Appendix 6.A). Women between the ages of 18 and 40 who were sexually active in the preceding month were eligible for the study. Exclusion criteria included the use of hormonal contraception, an intrauterine device, or current infertility (>1 year of unprotected intercourse without pregnancy). Women with a history of hysterectomy, infertility, polycystic ovaries, or those with

partners who had vasectomies were also excluded. 855 were initially eligible for the prospective study of fertility and menstrual function. Of these, 563 (64%) agreed to participate. Fourteen of these women became newly ineligible before the start of the study. We excluded seventy-nine women who did not collect any urine samples. The remaining 470 women comprise the sample for the current study. These women were requested to complete daily diaries for 12 months or until pregnancy. They were asked to collect urine samples on the first two days of every menstrual cycle (where the first day of bleeding is considered day 1). Urine samples were stored in the woman's freezer until a courier picked them up.

6.3.2 Exposure Information

The women recorded daily information on caffeine, alcohol, and smoking in the diaries, which were mailed on a monthly basis (Appendix 5.A). Caffeine consumption was recorded as cups of caffeinated tea, coffee, and cola. Alcohol was recorded as cans of beer, glasses of wine and shots of liquor. Smoking was recorded as the number of cigarettes. Menstrual bleeding and urine collection were indicated (yes/no), as well as intercourse and whether any birth control (e.g., barrier method, withdrawal, etc.) was used. Other covariates and demographic variables such as year of birth, race, ethnicity, body mass index (kg/m^2), age at menarche, previous months of unprotected intercourse (less than 1 year), desire to become pregnant, reproductive and medical histories were recorded during an interview at the onset of the study.

Caffeine was converted from cups into milligrams of caffeine using the following conversion factors: 1 cup caffeinated coffee= 150 mg; 1 cup caffeinated tea=55 mg; 1

cup caffeinated cola=45 mg, which are within the ranges of reported values. Caffeine was averaged over each cycle and categorized into low (mean caffeine intake <150 mg/day), medium (150-300 mg/day) and high (>300 mg/day) groups, on the basis that these are meaningful categories in terms of equivalent cups of coffee, and also allow a substantial number of cycles to fall into each category. Alcohol was averaged over each cycle and categorized into nondrinkers, light drinkers (mean alcohol intake <1 drink/day), and heavy drinkers (1+ drink/day). Categories were created to be practically meaningful while maintaining stable sample sizes; in addition, drinking more than 1 drink/day was recently associated with increased risk of many cancers (**26**). Smoking was either dichotomized or split into three categories: nonsmokers, <10 cigarettes/day, or 10+ cigarettes/day (half a pack). These variables are all cycle-specific; in other words, a woman may be categorized as a smoker for one cycle and a non-smoker for the next. Because knowledge of pregnancy may alter behavior, for cycles in which a pregnancy occurred, the mean alcohol, caffeine, and smoking values were averaged over the woman's mean cycle length and subsequent days were excluded.

6.3.3 NAT2 Genotyping

DNA was extracted from the frozen urine samples using the Qiagen MagAttract DNA Mini M48 kit in combination with the BioRobot M48 workstation. DNA was extracted in duplicate from 20% of the women. We then genotyped three single nucleotide polymorphisms in *NAT2:* rs1799929 (C481T); rs1799930 (G590A), and rs1208 (A803G). SNPs were chosen based on minor allele frequencies in HapMap (>5%) and their ability to differentiate slow from rapid metabolizers (**27**). The *4

haplotype (wild-type) is the rapid acetylator haplotype. Only one copy of this haplotype is necessary to confer a rapid acetylator phenotype. Good concordance between genotype and acetylator phenotype has been documented (**27-29**). Slow acetylators were assigned by having either a C→ T polymorphism at rs1799929, a G→A polymorphism at rs1799930, or an A→G polymorphism at rs1208. Ambiguous haplotypes were assigned using PHASE (v 2.1.1) using the expectation-maximization algorithm (**30, 31**).

The three *NAT2* SNPs were genotyped using the Beckman-Coulter GenomeLab SNPstream system using primers designed by Autoprimer.com (**32**). The primer sequences are as follows: for rs1799929: upper strand, GTGCCTTGCATTTTCTGC; lower strand, AAATTCTTTGTTTGTAATATACTGCTCTC; extension primer, CAACAAGTAATCCGCAGACTTACTGCTCTCTCCTGATTTGGTCCA; for rs1799930: upper strand, AAAGAATTTCTTAATTCTCATCTCCTG; lower strand, AAAATGATGTGGTTATAAATGAAGATG; extension primer, GCAGACAACGAACAACTACCATATACTTATTTACGCTTGAACCTC; for rs1208; upper strand, ATAAAGACAATACAGATCTGGTCGA; lower strand, TTTGGGCACGAGATTTCTC; extension primer, AGACTTCTACGCAAGCACTGGAGGAAGAGGTTGAAGAAGTGCTGA. Up to 6 ng of DNA was used for genotyping, depending on the sample concentration. To check genotyping accuracy, 20% of the samples were genotyped in duplicate from separate urine samples. The GenomeLab SNPstream Genotyping System Software Suite v2.3 was used to check the quality of the genotyping results, along with manual quality control (signal intensity and clustering pattern) to make individual calls where questionable.

6.3.4 Analysis

Time to pregnancy was measured as the number of menstrual cycles up until and including the cycle when a pregnancy was achieved. Some women became pregnant more than once during the study period. Only first pregnancies (N=179) were included in the present analysis; subsequent cycles and pregnancies were excluded. Pregnancies had previously been ascertained from hCG analysis of urine samples (**33**). Subclinical pregnancies were detected by measuring hCG levels in urine samples collected on days 1 and 2 of the subsequent menstrual cycle. The hCG was analyzed and pregnancies were diagnosed as described in Small et al. (**33**). Briefly, two samples of hCG greater than >0.25 ng/uL was used to determine pregnancy. We used questions from the entry interview to determine an approximate number of cycles a woman had been at risk for pregnancy at study entry, which was incorporated into the discrete survival analysis as either a covariate or added to prospective cycles at risk. Pregnancy outcomes were live births (n=126), subclinical spontaneous abortions (n=38), clinical spontaneous abortions (n=25), ectopic (n=2), molar (n=1), induced abortions (n=9), or unknown (n=6).

The effect of caffeine, alcohol, smoking, and covariates on time to pregnancy was assessed using discrete survival analysis. This approach is statistically more powerful than dichotomizing reproductive success as fertile/infertile (**34**). The discrete time hazard is defined as the conditional probability that a woman became pregnant in a given menstrual cycle conditional on a pregnancy not occurring in prior cycles. The likelihood for a discrete time hazard rate is equivalent to that for binary regression models (**35**). We modeled a per-cycle probability of conception (fecundability) using logistic regression and generated fecundability odds ratios (FORs), representing the odds of conception in

one group compared with the odds of conception in the referent group, with 95% confidence intervals (CIs).

Interactions between caffeine, alcohol, and smoking, both with each other and with NAT2 were also assessed by conducting stratified analyses or including product terms in the multivariate models.

The Emory Institutional Review Board approved the study protocol after complete de-identification of all samples, surveys and interviews.

6.4 Results

The women in this study were mostly white, non-Hispanic, married, and had some college education (Table 6.1). There were a substantial number of women who smoked (39%) and drank alcohol (89%) during follow-up. Although women had to have been at risk for pregnancy to be included in the study, only 23% reported that they were trying to become pregnant. Of the women we attempted to genotype, high quality DNA (resulting in successful genotyping) could only be obtained from 319 of the women. The 319 women who were genotyped for *NAT2* were not different from the other 151 women for any of the demographic or follow-up characteristics shown in Table 6.1.

The minor allele frequencies of the three *NAT2* SNPs among non-Hispanic whites were virtually identical to those in the CEU population of HapMap (Appendix 6.B). The concordance of genotype duplicates was 100% for all three SNPs, indicating high genotyping accuracy. The frequency of the wild-type, rapid acetylator haplotype in the population was 0.28±0.003. Of 319 women genotyped for *NAT2*, 161 (50%) carried two

copies of a slow haplotype and were assigned "slow" acetylator status (Table 6.2). Those who carried either one or two copies of the rapid haplotype were assigned "intermediate" and "rapid" acetylator status. Of the non-Hispanic whites, 54% were slow acetylators; this is consistent with other estimates (59%, 62%) in other Caucasian populations (**28, 36**). African-Americans had a higher proportion of rapid acetylators than whites (65% vs. 46%). Only one Asian was a slow acetylator. Because so few Asians were present in the population, and they may have distinct haplotype frequencies and/or reproductive characteristics, they were not included in multivariate models. Slow, intermediate, and rapid acetylators did not differ by other characteristics such as age, age at menarche, or pregnancy rate during the study (Table 6.3). However, slow acetylators consumed significantly more caffeine than intermediate or rapid acetylators. Rapid acetylators smoked more cigarettes and consumed more alcohol than slow or intermediate acetylators.

In a multivariate model without any interaction terms, age over 35, 6+ prior cycles at risk, obesity (BMI>30), and heavy alcohol use (1+ drink/day) were all significantly associated with longer time to pregnancy (FOR < 1) (Table 6.4). Alcohol, prior cycles at risk, and alcohol use were also significantly associated with longer time to pregnancy when modeled continuously instead of categorically; categorical results are presented to facilitate interpretation. Notably, a dose-response effect was apparent for alcohol intake, with light drinkers experiencing a 27% decrease in the FOR and heavy drinkers experiencing a 55% reduction in the FOR when compared to nondrinkers. Using a more extreme cutoff point for alcohol (>2 drinks/day) resulted in an even smaller FOR of 0.39 (95% CI, 0.11 to 1.32), but the confidence interval was wider, likely because of

the small number of cycles that fell into this category (N=148 cycles, or 4% of all cycles). Including prior cycles at risk as prospectively measured cycles (as opposed to a covariate) in the survival analysis did not alter the results. A higher frequency of unprotected intercourse and the intention to become pregnant were associated with a shorter time to pregnancy (FOR >1). Smoking and caffeine were not associated with time to pregnancy in this no-interaction model whether modeled continuously or categorically. *NAT2* haplotype, age at menarche, marital status, ethnicity, black race, partner's age, and ever pregnant were not associated with time to pregnancy when included as covariates (data not shown).

When *NAT2* and a smoking**NAT2* interaction term were included in the model, the FOR for smoking among slow acetylators was 0.39 (95% CI, 0.16 to 0.87) and among rapid acetylators was 1.68 (95% CI, 0.91 to 3.12; p for interaction = 0.005). When *NAT2* and an alcohol**NAT2* interaction term were included in the model, the FOR for alcohol (dichotomized, >1 drink vs. <1 drink/day) was 0.16 (95%, 0.04 to 0.72) among slow acetylators and 1.08 (95% CI, 0.53 to 2.21) among rapid acetylators (p for interaction = 0.02).

To further evaluate interactions between smoking, alcohol use, and *NAT2* acetylator status, the population was first divided into four groups: low-exposure women who did not smoke and who drank less than 1 drink/day (N=258); women who smoked but drank <1 drink/day (N=142); women who drank heavily (1+ drink/day) but were nonsmokers (N=29); and the highly exposed women, who both smoked and were heavy drinkers (N=41). Restricted to those with *NAT2* genotypes, the sample size in each group

reduces to 170, 94, 24, and 31, respectively. Table 6.5 shows the crude pregnancy rates among 8 exposure strata of various combinations of smoking, alcohol consumption, and *NAT2*. Low crude pregnancy rates and rate ratios demonstrate the lower fecundability among slow acetylators who smoke, drink heavily, or do both. Because none of the women in the highest exposure group (who smoke and drink heavily) became pregnant, a rate ratio could not be estimated. However, an exact 95% confidence interval for the rate ratio was calculated and it excluded 1.0 (95% CI: 0.0, 0.69), indicating that these women have significantly lower fecundability than the referent group (light/nondrinkers and nonsmokers).

Unadjusted Kaplan-Meier curves also show the significant interaction between smoking, alcohol, and *NAT2* haplotype on time to pregnancy (Figures 6.1 and 6.2). Among rapid acetylators (Figure 6.1), unadjusted survival curves comparing the four groups were not significantly different from one another (log-rank test, p=0.78) suggesting that smoking and drinking alcohol do not affect fecundability among rapid acetylators. However, among slow acetylators, a pattern emerged (Figure 6.2). Women who smoked and drank heavily did not become pregnant. Women who either smoked or drank heavily appeared to have reduced fecundability when compared to women who did neither. Thus a significant gradation of risk was apparent among slow acetylators (log-rank test, p=0.003).

Interactions between caffeine and alcohol, caffeine and smoking, and caffeine and *NAT2* were also investigated, but none were found, either in unadjusted analyses or in

multivariate models including product terms of each of these variables with caffeine (data not shown).

In order to generate an adjusted estimate of the reduced fecundability observed in Figures 6.1 and 6.2, a multivariate model accounting for smoking/alcohol/*NAT2* interaction was implemented. Women in the moderate and highly exposed strata were grouped together because of the low numbers of pregnancies in these strata. Thus the women were dichotomized into low exposure (women who do not smoke and drink <1 alcoholic drink/day) vs. highly exposed (smokers and/or women who consume >1 drink/day), and then further dichotomized into acetylator status (slow vs. rapid), resulting in four groups. Because this model included genetic information (*NAT2)*, the analysis was restricted to non-Hispanic whites to reduce the potential of confounding by population stratification or effect modification by race. The multivariate model included age, prior cycles at risk, intention to become pregnant, frequency of unprotected intercourse, BMI, and the 4-level interaction variable. Numbers are small because of the restriction to non-Hispanic whites who were genotyped. The highly exposed slow acetylators had markedly reduced fecundability when compared to the other three groups, which all had similar fecundability (Table 6.6).

6.5 Discussion

This study aimed to estimate the effects of caffeine, alcohol, and smoking on time to pregnancy, and determine whether these effects were modified by *NAT2* acetylator activity. No effect of caffeine on time to pregnancy was found, regardless of *NAT2* acetylator status. Smoking and alcohol, both separately and jointly, significantly

increased time to pregnancy, but only among women who were slow acetylators. Women who both smoked and drank alcohol and were slow acetylators had the lowest fecundability. These associations and interactions were observed both in unadjusted measures and in measures adjusted for age, prior cycles at risk, frequency of unprotected intercourse, intention to become pregnant, and BMI, and also persisted when the population was restricted to non-Hispanic whites.

Although most studies have observed a detrimental effect of smoking on fecundability (**10, 37-40**), others have not (**3, 8, 41**). Without stratifying on *NAT2*, no effect of smoking was observed in this study. Thus, it is possible that genetic heterogeneity clouded the effect of smoking on fecundability in the studies that did not find an effect. This emphasizes the importance of including *NAT2* haplotype information in studies where smoking is an exposure of interest.

There are several mechanisms by which smoking may affect fecundability. Current smoking has been related to increased FSH levels, indicating accelerated ovarian aging (**42**). There could also be effects on estrogen synthesis and metabolism (**43**). Smoking has also been associated with abnormal menstrual cycle characteristics and shorter menstrual cycles (**44, 45**). Passive and active smokers had a thicker zona pellucida than nonsmokers, which may affect fertilization (**46**). Smoking has been shown to affect the meiotic spindle in oocytes, resulting in chromosomal errors (**47**). These effects could theoretically be mediated by a combination of nicotine, cotinine, or other toxins present in cigarette smoke; however, the interaction between smoking and *NAT2* on time to pregnancy suggests that any effects of smoking on time to pregnancy are

mediated by the aromatic and heterocyclic amine carcinogens found in cigarette smoke that are metabolized by *NAT2* (**15, 48**).

A significant effect of alcohol on time to pregnancy was observed in this population, even without stratifying on *NAT2*. However, further analysis revealed that the detrimental effect of alcohol on fecundability was, in fact, limited to the slow acetylators. The combined effect was the result of a very low fecundability among heavy drinkers combined with the normal fecundability of light/nondrinkers.

There is presently no evidence that *NAT2* could directly metabolize ethanol, and therefore a mechanism for effect modification of alcohol by *NAT2* remains to be elucidated. However, other studies have found an interaction between *NAT2* and alcohol use with respect to various outcomes. Chen et al. (**49**) found that increasing alcohol intake was associated with increased risk of oral squamous cell cancer, but only among rapid or intermediate acetylators. In another case control study of bladder cancer, there was a marked interaction observed between alcohol and *NAT2* genotype, with alcohol having a much stronger effect among slow acetylators (**50**). There is also the possibility that alcohol intake is correlated with intake of other substances (e.g., cooked meat) which contain toxins that are metabolized by NAT2 (i.e., confounding). In that case, the effect modification observed with alcohol could be a reflection of (1) the effect modification of the substances correlated with alcohol intake, or (2) the alcohol adding to total body burden of toxins, surpassing some threshold so that NAT2 and other enzymes cannot metabolize all toxins as efficiently.

There was no effect of caffeine on time to pregnancy in this population. This is the first study to report findings from daily, prospective information on caffeine intake and prospectively ascertained pregnancies, including subclinical pregnancies, while controlling for frequency of intercourse, smoking and alcohol. Other studies of caffeine and fecundability shared one or more important differences. For example, Wilcox et al. used a 3-month recall of average caffeine exposure and had a small sample size (n=104 women who had not become pregnant in the first 3 months of the study) (**1**). Because these women shared a reduced baseline fecundability, they may have been more sensitive to any effects of caffeine.

No interaction was observed between smoking and caffeine on time to pregnancy. Smoking increases the metabolism of caffeine by inducing CYP1A2 activity, which catalyzes the initial step in caffeine metabolism (**7, 51**). In our study, smokers drank significantly more caffeine than nonsmokers (1817 mg/week vs. 1313 mg/week: $p<0.0001$); however, even among nonsmokers, there was no effect of caffeine on time to pregnancy.

Misclassification of caffeine intake could have also cause biased any effect estimates towards the null. In a recent validity study, self-reported caffeine use in a prospective daily diary was correlated with salivary caffeine concentrations ($r=0.68$), confirming the validity of using self-reported caffeine measures (**52**). However, even with a correlation of 0.68, there could still be some misclassification of caffeine use. In addition, this study assumed that each cup of coffee contained 150 mg caffeine. The

amount of caffeine in cups of coffee varies; this would cause further misclassification of caffeine intake.

A limitation of this study was the small numbers in some of the risk groups, resulting in wide confidence intervals for some measures. Only a subset of the women could be genotyped. However, being genotyped was not associated with any of the characteristics in this study, and *NAT2* SNP frequencies were virtually identical to those found in HapMap, so this subsetting was unlikely to cause any bias.

Another limitation could be the misclassification of acetylator status. Genotypes for six *NAT2* SNPs are required to assign acetylator phenotype with greater than 99% sensitivity and specificity (**53**). We downloaded linkage disequilibrium data from the CEU population in HapMap to determine whether the three SNPs we genotyped were correlated with the ones not genotyped. We genotyped rs1799929, which had a D' of 1.0 with all three ungenotyped SNPs (rs1041983, rs1801280 and rs1208), an R^2 of 89% with rs1801280, and an R^2 of 91% with rs1041983. The third SNP not genotyped (rs1799931) had a very low minor allele frequency (MAF <0.01 among the CEU population) and therefore should not greatly impact our classification scheme. In summary, the SNPs not genotyped were either in high linkage disequilibrium with our SNPs or had a very low minor allele frequency, and therefore likely resulted in only a small degree of misclassification.

This population consisted of a mixture of women who reported that they were trying to get pregnant and some who reported that they were not. We controlled for this in the analysis, in addition to controlling for frequency of unprotected intercourse.

Women who were trying were not different from women who were not trying by age, smoking, alcohol intake, or caffeine intake (Appendix 6.C). Women who reported that they were trying to become pregnant did have significantly more unprotected intercourse per cycle ($p<0.001$) and were more likely to achieve pregnancy during the study ($p<0.001$) than women who were not trying.

The apparent effects of alcohol and smoking on time to pregnancy could be a result of very early pregnancy losses (before the expected start of the next menstrual cycle), as opposed to a delay in conception. If women who drink alcohol and smoke actually have the same fecundability as other women, but have very early pregnancy losses, this could lead to an apparent increase in time to pregnancy. This is an unlikely explanation since in our study we tested hCG at the start of every menstrual cycle and included all subclinical pregnancies in the analysis. Thus, the "very early" losses would have to be extremely close to implantation to have been missed in our analysis.

In summary, we observed a strong effect of alcohol and smoking on time to pregnancy among women who were slow *NAT2* acetylators. This research emphasizes the need to incorporate genetic information about metabolic enzymes to better interpret and understand the effects of xenobiotics on human health.

6.6 Acknowledgments

This research was supported by funding from the National Institutes of Health (RO1HD24618, RO3HD55176, R01HG03618), and the University Research Committee of Emory University. This manuscript was developed under a STAR Research

Assistance Agreement No. 916891 awarded by the U.S. Environmental Protection Agency. It has **not** been formally reviewed by the EPA. The views expressed in this manuscript are solely those of Kira Creswell Taylor and the EPA **does not endorse** any products or commercial services mentioned in this manuscript.

6.7 Figures and Tables

Table 6.1 Population Characteristics of 470 Women Office Workers and 319 Women Who Were Genotyped for *NAT2*

	All women N=470 N (%)[a]	Genotyped women N=319 N (%)	χ^2 p-value[b]
Age (years)			0.74
19-24	41 (9)	30 (9)	
25-29	150 (32)	102 (32)	
30-34	164 (35)	113 (35)	
35-41	115 (24)	74 (23)	
Race			0.21
White	376 (80)	248 (78)	
African-American or Black	58 (12)	45 (14)	
Asian	13 (3)	8 (3)	
Other	23 (5)	18 (6)	
Ethnicity			0.36
Hispanic	22 (5)	18 (6)	
Non-Hispanic	445 (95)	299 (94)	
Missing	3 (1)	2 (1)	
Marital status			0.91
Married	309 (66)	208 (65)	
Single (never married)	127 (27)	87 (27)	
Divorced/separated/widowed	34 (7)	24 (8)	
Ever Pregnant			0.91
Yes	285 (61)	194 (61)	
No	185 (39)	125 (39)	
Highest education			0.23
High school or technical school	110 (23)	76 (24)	
Some college	156 (33)	112 (35)	
College graduate	203 (43)	131 (41)	
Missing	1 (0)	0 (0)	

BMI (kg/m^2)			0.53
<20	80 (17)	49 (15)	
20-25	240 (51)	164 (51)	
26-30	91 (19)	62 (19)	
>30	58 (12)	43 (13)	
Missing	1 (0)	1 (0)	
Trying to get pregnant			0.39
Yes	108 (23)	73 (23)	
No	332 (71)	229 (72)	
Missing	30 (6)	17 (5)	
Smoking during follow-up			0.87
Any	183 (39)	125 (39)	
None	287 (61)	194 (61)	
Mean alcohol intake during follow-up			0.10
None	54 (11)	34 (11)	
<1 drink/day	346 (74)	230 (72)	
≥1 drink/day	70 (15)	55 (17)	
Mean caffeine intake during follow-up			0.99
≤150 mg/day	256 (54)	174 (55)	
151-300 mg/day	137 (29)	93 (29)	
>300 mg/day	77 (16)	52 (16)	

[a]Percents may not add to 100 due to rounding.
[b]Chi-square test comparing category frequencies of N=319 women who were genotyped with N=151 women who were not.

Table 6.2 Association of Haplotype Carrier Status with Race and Ethnicity Among 319 Women with Assigned Haplotypes

	Slow N=161 N (row %)	Intermediate N=132 N (row%)	Rapid N=26 N (row%)	Chi-square p-value
Race				0.04
White	134 (54)	98 (40)	16 (6)	
Black	16 (36)	22 (49)	7 (16)	
Asian	1 (13)	5 (63)	2 (25)	
Other	10 (56)	7 (39)	1 (6)	
Ethnicity				0.40
Hispanic	9 (50)	9 (50)	0 (0)	
Non-Hispanic	151 (51)	123 (41)	25 (8)	

Table 6.3 Reproductive and Follow-Up Characteristics According to *NAT2* Haplotype

	Slow (0 copies of CGA haplotype) N=161	Intermediate (1 copy of CGA) N=132	Rapid (2 copies of CGA) N=26	P_{trend}
Mean age±SE (years)	30.5±0.4	30.8±0.4	32.0±0.8	0.19
Pregnancy rate (pregnancies/100 cycles at risk)	4.5±0.6	5.3±0.7	5.6±1.7	0.30
Caffeine during follow-up				
Mean mg per week	1308±84	1021±93	988±213	0.05
≤150 mg/day (N, col %)	65 (41)	69 (52)	12 (46)	0.06
150-300 mg/day	42 (26)	35 (26)	8 (31)	
>300 mg/day	52 (33)	28 (21)	6 (23)	
Smoking during follow-up				
Mean # cigarettes/day	3.1±0.6	3.1±0.6	6.0±1.4	0.16
Any (N, col %)	61 (38)	48 (36)	16 (62)	0.16
None	100 (62)	84 (64)	10 (38)	
Alcohol during follow-up				
Mean # drinks/week	3.2±0.4	3.9±0.4	6.4±1.0	0.01
None (N, col %)	14 (9)	17 (13)	3 (12)	0.19
<1 drink/day	126 (78)	90 (68)	14 (54)	
≥1 drink/day	21 (13)	25 (19)	9 (35)	

Table 6.4 Predictors of Time to Pregnancy in a Multivariate, No-Interaction Model

Variable	Fecundability odds ratio (95% CI)	Score test p-value
Age (years)		
19-34	Ref	0.02
35-41	0.57 (0.35, 0.91)	
Unprotected intercourse (times per week)	1.52 (1.32, 1.76)	<0.0001
Intention to become pregnant		
No	Ref	0.002
Yes	1.86 (1.25, 2.76)	
Prior cycles at risk		
0-5	Ref	0.003
6+	0.51 (0.32, 0.80)	
Body mass index (kg/m^2)		
<25	Ref	0.04
25-30	0.93 (0.59, 1.47)	
30+ (obese)	0.43 (0.22, 0.83)	
Average caffeine intake		
<150 mg/day	Ref	0.80
150-300 mg/day	1.15 (0.75, 1.76)	
300+ mg/day	1.01 (0.64, 1.59)	
Alcohol intake		
None	Ref	0.04
<1 drink/day	0.73 (0.48, 1.10)	
≥1 drink/day	0.45 (0.24, 0.84)	
Smoking		
None during cycle	Ref	0.76
Any during cycle	1.06 (0.71, 1.58)	

Table 6.5 Crude Pregnancy Rates Examining Interaction of Speed, Smoking (Smoker vs. Non), and Alcohol (≥1 Drink/Day vs. Less)

≥1 Alcoholic drink/day	Smoker	Slow speed	N women	Total cycles at risk	Total preg-nancies	Crude rate (pregnancies/ 100 cycles at risk)	Crude rate ratio 95% CI[a]
			79	570	34	6.0	Ref
		X	91	693	44	6.3	1.06 (0.68, 1.67)
	X		45	390	17	4.4	0.73 (0.41, 1.31)
	X	X	49	443	12	2.7	0.45 (0.24, 0.88)
X			15	107	6	5.6	0.93 (0.39, 2.24)
X		X	9	71	2	2.8	0.47 (0.05, 1.84)[b]
X	X		19	161	9	5.6	0.93 (0.45, 1.95)
X	X	X	12	95	0	0.0	-- (0.0, 0.69)[b]

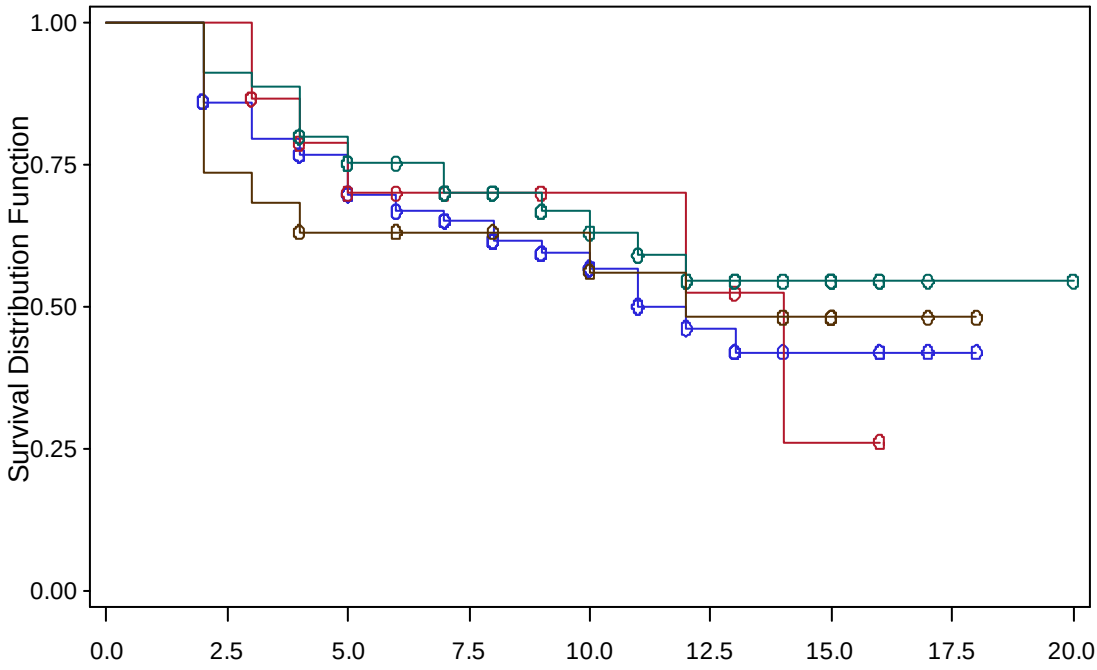
[a]Large sample 95% confidence limits are given unless otherwise noted.
[b]Fisher exact 95% confidence limits.

Table 6.6 Adjusted Fecundability Odds Ratios Demonstrating the Interaction Between Smoking, Heavy Drinking, and *NAT2* Haplotype Among non-Hispanic Whites

	Fecundability odds ratio[a] (95% CI)
Nonsmokers, light/nondrinkers; rapid acetylators	1.00 (Ref)
Nonsmokers, light/nondrinkers; slow acetylators	0.94 (0.52, 1.72)
Smokers and/or heavy drinkers; rapid acetylators	1.11 (0.60, 2.06)
Smokers and/or heavy drinkers; slow acetylators	0.32 (0.14, 0.74)**

[a]Adjusted for age, frequency of unprotected intercourse, prior cycles at risk, intention to become pregnant, and BMI, in the categories shown in Table 6.4.
**p<0.01

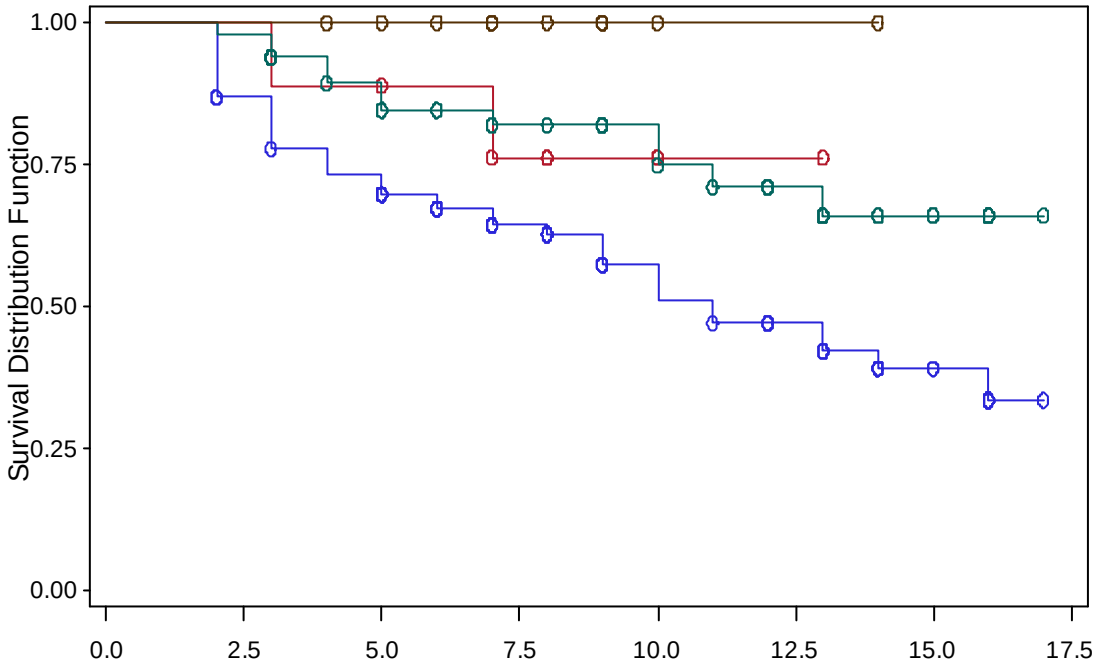
Figure 6.1 Joint Effects of Alcohol and Smoking on Time to Pregnancy Among Fast Acetylators



Legend: Blue =light/nondrinkers, nonsmokers
Red = heavy drinkers, nonsmokers
Green = light/nondrinkers, smokers
Black = heavy drinkers, smokers

Circles indicate censored observations

Figure 6.2 Joint Effects of Alcohol and Smoking on Time to Pregnancy Among Slow Acetylators



Legend: Blue =light/nondrinkers, nonsmokers
Red = heavy drinkers, nonsmokers
Green = light/nondrinkers, smokers
Black = heavy drinkers, smokers

Circles indicate censored observations

6.8 References

1. Wilcox A, Weinberg C, Baird D. Caffeinated beverages and decreased fertility. *Lancet* 1988;**2**(8626-8627): 1453-6.

2. Hakim RB, Gray RH, Zacur H. Alcohol and caffeine consumption and decreased fertility. *Fertility and Sterility* 1998;**70**(4): 632-7.

3. Florack EI, Zielhuis GA, Rolland R. Cigarette smoking, alcohol consumption, and caffeine intake and fecundability. *Preventive Medicine* 1994;**23**(2): 175-80.

4. Caan B, Quesenberry CP, Jr., Coates AO. Differences in fertility associated with caffeinated beverage consumption. *American Journal of Public Health* 1998;**88**(2): 270-4.

5. Jensen TK, Henriksen TB, Hjollund NH, et al. Caffeine intake and fecundability: a follow-up study among 430 Danish couples planning their first pregnancy. *Reproductive Toxicology* 1998;**12**(3): 289-95.

6. Stanton CK, Gray RH. Effects of caffeine consumption on delayed conception. *American Journal of Epidemiology* 1995;**142**(12): 1322-9.

7. Zevin S, Benowitz NL. Drug interactions with tobacco smoking. An update. *Clinical Pharmacokinetics* 1999;**36**(6): 425-38.

8. Buck Louis GM, Dmochowski J, Lynch C, Kostyniak P, McGuinness BM, Vena JE. Polychlorinated biphenyl serum concentrations, lifestyle and time-to-pregnancy. *Human Reproduction* 2008:24(2):451-458.

9. Jensen TK, Hjollund NH, Henriksen TB, et al. Does moderate alcohol consumption affect fertility? Follow up study among couples planning first pregnancy. *British Medical Journal* 1998;**317**(7157): 505-10.

10. Jensen TK, Henriksen TB, Hjollund NH, et al. Adult and prenatal exposures to tobacco smoke as risk indicators of fertility among 430 Danish couples. *American Journal of Epidemiology* 1998;**148**(10): 992-7.

11. Blum M, Grant DM, McBride W, Heim M, Meyer UA. Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA Cell Biology* 1990;**9**(3): 193-203.

12. Windmill KF, Gaedigk A, Hall PM, Samaratunga H, Grant DM, McManus ME. Localization of N-acetyltransferases NAT1 and NAT2 in human tissues. *Toxicological Sciences* 2000;**54**(1): 19-29.

13. Weber WW, Hein DW. N-acetylation pharmacogenetics. *Pharmacology Reviews* 1985;**37**(1): 25-79.

14. Cross AJ, Sinha R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environmental and Molecular Mutagenesis* 2004;**44**(1): 44-55.

15. Kataoka H, Kijima K, Maruo G. Determination of mutagenic heterocyclic amines in combustion smoke samples. *Bulletin of Environmental Contamination and Toxicology* 1998;**60**(1): 60-7.

16. Sim E, Lack N, Wang CJ, et al. Arylamine N-acetyltransferases: structural and functional implications of polymorphisms. *Toxicology* 2008;**254**(3): 170-83.

17. Kalow W. Variability of caffeine metabolism in humans. *Arzneimittelforschung* 1985;**35**(1A): 319-24.

18. Hein DW, Doll MA, Fretland AJ, et al. Molecular Genetics and Epidemiology of the NAT1 and NAT2 Acetylation Polymorphisms. *Cancer Epidemiology Biomarkers and Prevention* 2000;**9**(1): 29-42.

19. Grant DM, Morike K, Eichelbaum M, Meyer UA. Acetylation pharmacogenetics. The slow acetylator phenotype is caused by decreased or absent arylamine N-acetyltransferase in human liver. *Journal of Clinical Investigation* 1990;**85**(3): 968-72.

20. Welfare MR, Bassendine MF, Daly AK. The effect of NAT2 genotype and gender on the metabolism of caffeine in nonsmoking subjects. *British Journal of Clinical Pharmacology* 2000;**49**(3): 240-3.

21. Makarova SI. Human N-acetyltransferases and drug-induced hepatotoxicity. *Current Drug Metabolism* 2008;**9**(6): 538-45.

22. Moslehi R, Chatterjee N, Church TR, et al. Cigarette smoking, N-acetyltransferase genes and the risk of advanced colorectal adenoma. *Pharmacogenomics* 2006;**7**(6): 819-29.

23. Risch A, Wallace DM, Bathers S, Sim E. Slow N-acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer. *Human Molecular Genetics* 1995;**4**(2): 231-6.

24. Ouerhani S, Rouissi K, Marrakchi R, et al. Combined effect of NAT2, MTR and MTHFR genotypes and tobacco on bladder cancer susceptibility in Tunisian population. *Cancer Detection and Prevention* 2009;**32**(5-6): 395-402.

25. Marcus M, Gerr F. Upper extremity musculoskeletal symptoms among female office workers: associations with video display terminal use and occupational psychosocial stressors. *American Journal of Industrial Medicine* 1996;**29**(2): 161-70.

26. Allen NE, Beral V, Casabonne D, et al. Moderate alcohol intake and cancer incidence in women. *Journal of the National Cancer Institute* 2009;**101**(5): 296-305.

27. Cascorbi I, Drakoulis N, Brockmoller J, Maurer A, Sperling K, Roots I. Arylamine N-acetyltransferase (NAT2) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. *American Journal of Human Genetics* 1995;**57**(3): 581-92.

28. Gross M, Kruisselbrink T, Anderson K, et al. Distribution and Concordance of N-Acetyltransferase Genotype and Phenotype in an American Population. *Cancer Epidemiology Biomarkers and Prevention* 1999;**8**(8): 683-92.

29. Jorge-Nebert LF, Eichelbaum M, Griese EU, Inaba T, Arias TD. Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. *Pharmacogenetics* 2002;**12**(1): 39-48.

30. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 2001;**68**(4): 978-89.

31. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 2003;**73**(5): 1162-9.

32. Beckman Coulter I. GenomeLab SNPstream Genotyping System 48-plex Training Guide. In: Beckman Coulter I, editor. Fullerton, CA: Beckman Coulter, Inc.; 2005.

33. Small CM, Manatunga AK, Klein M, et al. Menstrual cycle characteristics: associations with fertility and spontaneous abortion. *Epidemiology* 2006;**17**(1): 52-60.

34. Baird DD, Wilcox AJ, Weinberg CR. Use of time to pregnancy to study environmental exposures. *American Journal of Epidemiology* 1986;**124**(3): 470-80.

35. Scheike TH, Keiding N. Design and analysis of time-to-pregnancy. *Statistical Methods in Medical Research* 2006;**15**(2): 127-40.

36. Bech BH, Autrup H, Nohr EA, Henriksen TB, Olsen J. Stillbirth and slow metabolizers of caffeine: comparison by genotypes. *International Journal of Epidemiology* 2006;**35**(4): 948-53.

37. Olsen J, Rachootin P, Schiodt AV, Damsbo N. Tobacco use, alcohol consumption and infertility. *International Journal of Epidemiology* 1983;**12**(2): 179-84.

38. Suonio S, Saarikoski S, Kauhanen O, Metsapelto A, Terho J, Vohlonen I. Smoking does affect fecundity. *European Journal of Obstetrics, Gynecology and Reproductive Biology* 1990 ;**34**(1-2): 89-95.

39. Curtis KM, Savitz DA, Arbuckle TE. Effects of cigarette smoking, caffeine consumption, and alcohol intake on fecundability. *American Journal of Epidemiology* 1997;**146**(1): 32-41.

40. Hassan MA, Killick SR. Negative lifestyle is associated with a significant reduction in fecundity. *Fertility and Sterility* 2004;**81**(2): 384-92.

41. de Mouzon J, Spira A, Schwartz D. A prospective study of the relation between smoking and fertility. *International Journal of Epidemiology* 1988;**17**(2): 378-84.

42. Kinney A, Kline J, Kelly A, Reuss ML, Levin B. Smoking, alcohol and caffeine in relation to ovarian age during the reproductive years. *Human Reproduction* 2007;**22**(4): 1175-85.

43. Shiverick KT, Salafia C. Cigarette smoking and pregnancy I: ovarian, uterine and placental effects. *Placenta* 1999;**20**(4): 265-72.

44. Rowland AS, Baird DD, Long S, et al. Influence of medical conditions and lifestyle factors on the menstrual cycle. *Epidemiology* 2002;**13**(6): 668-74.

45. Brown S, Vessey M, Stratton I. The influence of method of contraception and cigarette smoking on menstrual patterns. *British Journal of Obstetrics & Gynaecology* 1988;**95**(9): 905-10.

46. Shiloh H, Lahav-Baratz S, Koifman M, et al. The impact of cigarette smoking on zona pellucida thickness of oocytes and embryos prior to transfer into the uterine cavity. *Human Reproduction* 2004;**19**(1): 157-9.

47. Zenzes MT. Smoking and reproduction: gene damage to human gametes and embryos. *Human Reproduction Update* 2000;**6**(2): 122-31.

48. Regina Stabbert, Karl-Heinz Schäfer, Christine Biefel, Klaus Rustemeier. Analysis of aromatic amines in cigarette smoke. *Rapid Communications in Mass Spectrometry* 2003;**17**(18): 2125-32.

49. Chen C, Ricks S, Doody DR, Fitzgibbons ED, Porter PL, Schwartz SM. N-Acetyltransferase 2 polymorphisms, cigarette smoking and alcohol consumption, and oral squamous cell cancer risk. *Carcinogenesis* 2001;**22**(12): 1993-9.

50. Lu CM, Chung MC, Huang CH, Ko YC. Interaction Effect in Bladder Cancer between N-Acetyltransferase 2 Genotype and Alcohol Drinking. *Urologia Internationalis* 2005;**75**(4): 360-4.

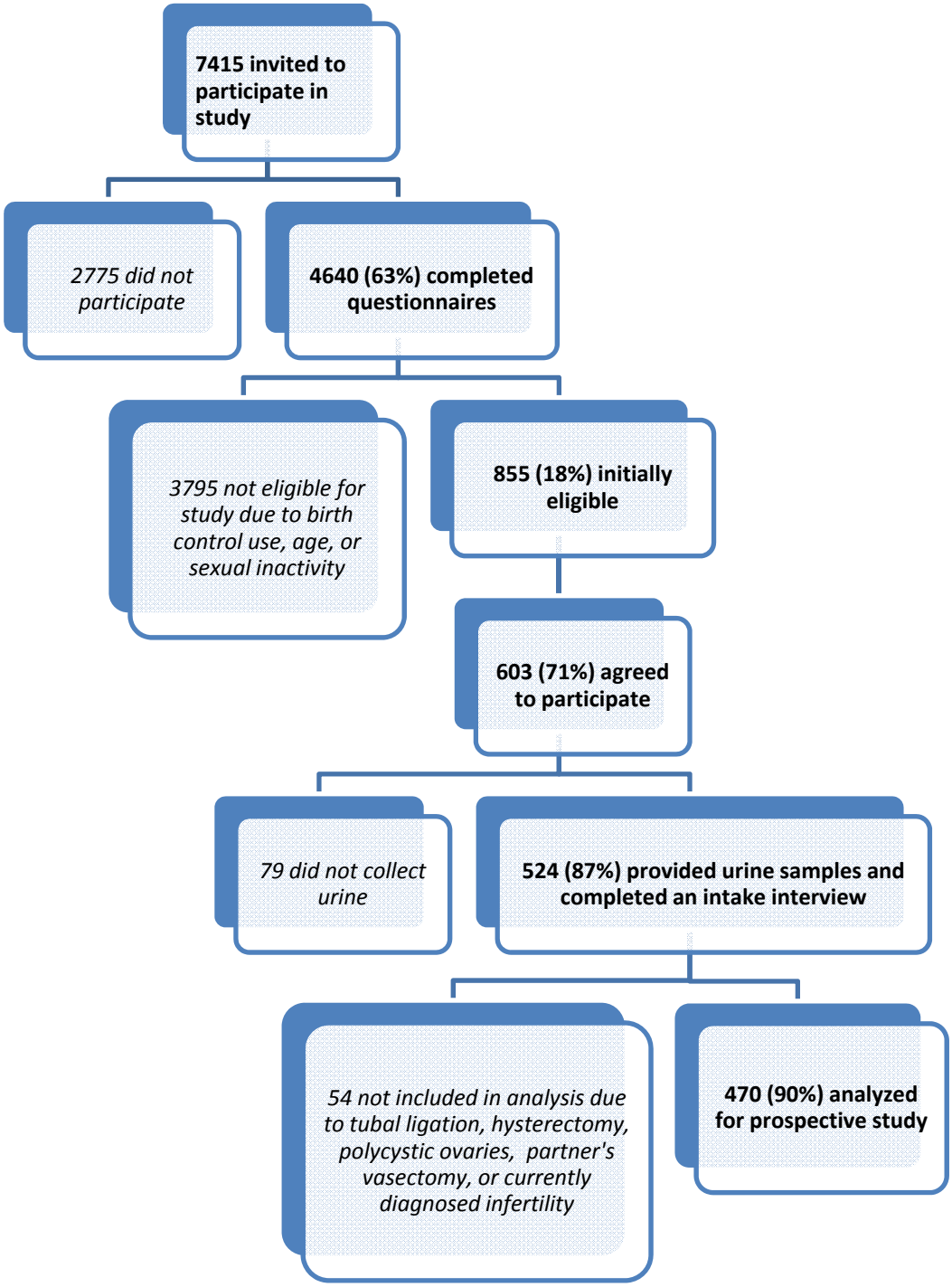
51. Benowitz NL, Peng M, Jacob P. Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism. *Clinical Pharmacology and Therapeutics* 2003;**74**(5): 468-74.

52. Addicott M, Yang L, Peiffer A, Laurienti P. Methodological considerations for the quantification of self-reported caffeine use. *Psychopharmacology* 2009;**203**(3): 571-8.

53. Kuznetsov IB, McDuffie M, Moslehi R. A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. *Bioinformatics* 2009;**25**(9): 1185-6.

6.9 Appendices

Appendix 6.A Participation in MSSWOW and the Prospective Pregnancy Study



Appendix 6.B Three SNPs genotyped in *NAT2*

dbSNP rs number	Position along chr8p22	Alleles (Major/ Minor)	MAF[a]	MAF[b]	Amino acids
rs1799929	18302274	C481T	0.42	0.43	Leu/Leu
rs1799930	18302383	G590A	0.29	0.30	Arg/Gln
rs1208	18302596	A803G	0.43	0.42	Lys/Arg

[a]HapMap CEU population
[b]Non-Hispanic whites in this study

Appendix 6.C Follow-Up Characteristics of Women Attempting Pregnancy and Women Not Attempting Pregnancy

Follow-up variable	Reported attempting pregnancy N=108 N(%)	Not attempting pregnancy N=332 N(%)	p-value
Caffeine			
Mean mg per week±SE	1098±96	1246±58	0.20
<150 mg/day	63 (58)	169 (51)	0.13
150-300 mg/day	31 (29)	102 (31)	
>300 mg/day	14 (13)	61 (18)	
Smoking			
Mean cigs/day±SE	3.3±0.7	3.5±0.4	0.80
Any	41 (38)	134 (40)	0.66
None	67 (62)	198 (60)	
Alcohol			
Mean drinks per week±SE	3.2±0.4	3.7±0.3	0.33
None	11 (10)	30 (9)	0.39
<1 drink/day	83 (77)	247 (74)	
≥1 drink/day	14 (13)	55 (17)	
Unprotected intercourse			<0.0001
Never	1 (1)	87 (26)	
<5 times/month	65 (61)	205 (62)	
≥5 times/month	40 (38)	37 (11)	
Achieved pregnancy during study			
Yes	62 (57)	96 (29)	<0.0001
No	46 (43)	236 (71)	
Pregnancy outcomes			
Live births	48 (77)	52 (54)	0.08
Induced abortion	2 (3)	7 (7)	
Clinical spontaneous abortion	7 (11)	17 (18)	
Subclinical spontaneous abortion	3 (5)	9 (9)	
Blighted ovum/Ectopic pregnancy/Molar pregnancy	2 (3)	6 (6)	
Unknown	0	5 (5)	

Chapter 7

Discussion

7.1 Discussion

The overarching aims of this dissertation were to: (1) investigate the relative contributions of genes and environmental factors to reproductive traits; (2) determine whether specific genetic and environmental factors affect reproductive traits; and (3) assess gene-gene and gene-environment interactions using both agnostic and *a priori* approaches. Although these aims were broad and ambitious, a valuable picture emerged when the results of all of these studies were considered in concert. This section highlights the novel contributions that each of these studies made to the literature in the context of discussing major concepts and developments in the fields of genetic and environmental epidemiology.

Genes do not function in a vacuum. Genes function within a particular micro-environment, which is part of a cellular environment, which is part of a tissue in which cells communicate with one another. Tissues comprise the organism, and the organism is exposed to the external environment. Expression and function of some genes may appear to be independent of environmental factors, but in fact they may depend greatly on environmental factors that do not vary appreciably. The first study presented in this dissertation ("Heritability of age at menarche in a population exposed to polybrominated biphenyls") indicated that the *degree* to which genes influence a trait may depend on the influence of environmental factors. These environmental factors vary across populations, within populations, and even across cells and tissues within individual organisms. To our knowledge, this is the first study to compare the heritability of a trait across families with

varying levels of an environmental biological or chemical exposure. Future heritability studies should consider this "shifting landscape" of the influence of genes and the varying, complex environments in which they must act.

Similarly, when studying the effects of environmental factors on health outcomes, it must be considered that the environmental factors act on very complex and adaptable biological organisms. Genes, along with environments in which they act, are ultimately responsible for biological characteristics that may confer a greater or lesser susceptibility to particular environmental exposures. The study of alcohol, smoking and time to pregnancy in this dissertation provided a very clear example of how genetic susceptibility, in the form of the "slow" *NAT2* haplotype, permitted (or even caused) an environmental exposure to be more detrimental in susceptible individuals. This was the first study to examine *NAT2* acetylator status as an effect modifier for risk factors of fecundability or fertility. Even in this relatively small study, a clinically important and statistically significant interaction was observed. Such interactions are undoubtedly both present and measurable far more often than they are considered or evaluated in the literature.

Once we accept that gene-environment interaction is always present in this "shifting landscape", understanding the relevant biology and chemistry is imperative for properly designing and analyzing an epidemiologic study in the area of reproductive health. Endocrine function is latent and each measurable trait is the culmination of complex biological pathways. Animal studies and *in vitro* studies have shown that polybrominated biphenyls and other related compounds may act in an estrogenic or anti-estrogenic fashion. Without these studies to observe the direct and indirect effects of

these chemicals on model cells and organisms, we would be less able to conceive of plausible hypotheses or explanations for seemingly conflicting results in human epidemiologic studies. For example, consider the study in this dissertation that examined PBBs and day 3 levels of FSH and E13G. We could not have hypothesized that PBBs would affect reproductive hormone levels, let alone understand how they could result in lower levels of estrogen or FSH, without the previous work of biologists and chemists to understand the basic science of these compounds and their effects in controlled environments. This study was the first to evaluate the effects of polybrominated biphenyls on reproductive hormone levels in humans. Findings from this study will help investigators understand the effects of PBBs on more indirect reproductive outcomes such as age at menarche, fertility, or reproductive cancers.

Although the current trend in genetic epidemiology is to invest heavily in genome-wide association scans (GWAS), this dissertation emphasizes the continued value of candidate gene and gene-environment interaction studies. For example, *PGR* was not identified in recent whole-genome association scans for age at menarche. The significance criterion for whole-genome association scans is usually a p-value of 10^{-7} or even less. Therefore if *PGR* were associated with menarche but the p-value in the GWAS was only 10^{-6}, it would not be identified as an important player. Clearly, with such stringent significance criteria, GWAS studies always run the risk of many type II errors. In addition, specific genes and SNPs may be missed by using the GWAS approach. Information obtained from these "agnostic" studies must be combined with candidate gene studies conceived by biology or prior knowledge in order for the field to move forward.

For candidate gene studies, replication is usually now required for publication. Although we did not have a dataset in which to replicate this finding, variation in *PGR* was associated with both age at menarche and menstrual cycle length in this population. This bolsters the evidence that variation in *PGR* can alter reproductive function in general because it is less plausible that both individual associations were simply due to type I errors. This pleiotropy (the effect of a single gene on different independent outcomes) may be considered a "replication" of sorts, if the outcomes are independent. This study was the first epidemiological study that specifically examined *PGR* genotypes and age at menarche, and it was the first study to investigate (and find) an association between *PGR* and menstrual function.

In this dissertation, no association was observed between inhibin polymorphisms and menstrual function despite considerable biological plausibility. Even a "simply" measured outcome such as menstrual cycle length is a result of many factors, such as: (1) interactions between hormones and their receptors causing follicular maturation and egg release; (2) interactions between hormones and their receptors causing maturation of the corpus luteum and vascularization of the endometrium in preparation for pregnancy; (3) the influence of genetic polymorphisms, environmental influences (e.g., smoking), biological characteristics (e.g., obesity), and their interactions. These complexities and interactions may partially account for the lack of association observed in this dissertation; therefore, future studies of inhibin polymorphisms must consider these factors in both the study design and the analysis.

7.2 Future Directions

Genetic studies must consider the environment and vice versa. Although gene-environment interaction is not often reported in epidemiologic literature, possibly because it is not often investigated, I found evidence of gene-environment interaction in both studies where I inquired. In particular, it was notable that the smoking effect on time to pregnancy was not detectable unless *NAT2* genotype was taken into account. This was a situation with much *a priori* evidence to suggest the interaction. In studies where there is no previous literature to suggest an interaction, perhaps an agnostic approach should be taken to investigate all possible gene-environment interactions within the realm of the study. On the other hand, one may also argue that careful consideration be given to each possible interaction tested. Researchers disagree on the best approach; the former may allow for more new discoveries, but also runs the risk of increased type I errors. A combination of approaches may be the best way for scientists to proceed.

Consideration of gene-environment interaction has become far more important as effects of both individual genes and environmental factors have become more and more subtle in recent years. Historically, the field of human genetics has focused on major single-gene disorders that have drastic phenotypic consequences and have high "penetrance" (appear to be less susceptible to environmental influences). Similarly, epidemiologic studies from decades ago focused on very strong associations, such as that between smoking and lung cancer. We now must look for more subtle associations, which may only be observable when interaction is properly taken into account. At the same time, analysis techniques are becoming more and more sophisticated in this era of biomarkers, gene "chips" that are able to analyze a million SNPs at a time, and analysis

methods that take even the most powerful computer several hours or days to complete. We are able to assess millions of associations simultaneously, resulting in a vast number of false positive and false negative associations. Analyses that focus on interactions, and not main effects, will be a useful tool to detect etiologically and clinically meaningful associations.

Few mechanisms of gene-environment interaction are well-understood. A recent development in this area is the conceptualization of "epigenetics," or the discovery of heritable gene modification other than traditional mutations in DNA sequence. One example of epigenetic change is the methylation of DNA that has been shown to alter gene expression. Methylated nucleotides are now being considered as the 5^{th} and 6^{th} nucleotides (5-methylcytosine and 5-hydroxymethylcytosine) (**1**). Some methylation patterns are conserved as cells divide during mitosis or meiosis. It is now known that some environmental factors are able to cause epigenetic changes. As an example, benzene and traffic particle pollutants have recently been shown to alter genome-wide as well as gene-specific methylation patterns (**2, 3**). Future studies in genetic epidemiology must find a way to measure genetic variability including these new nucleotides, which are heavily influenced by the environment.

7.3 Conclusion

This dissertation presents evidence of several genetic and environmental influences and their interactions on reproductive characteristics. In this era of rapid scientific and technological progress, consideration of gene-environment interaction is critical to both discovering and evaluating associations in diverse epidemiologic fields of study.

7.4 References

1. Kriaucionis S, Heintz N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* 2009;**324**(5929): 929-30.

2. Bollati V, Baccarelli A, Hou L, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Research* 2007;**67**(3): 876-80.

3. Baccarelli A, Wright RO, Bollati V, et al. Rapid DNA methylation changes after exposure to traffic particles. *American Journal of Respiratory and Critical Care Medicine* 2009;**179**(7): 572-8.