Women's Health: Genetic Variation In Complex Traits

By

Jennifer Renée Malinowski

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

December, 2014

Nashville, Tennessee

Approved:

Dana C. Crawford, Ph.D.

Ellen W. Clayton, M.D., J.D.

Melinda Aldrich, Ph.D., M.P.H.

William S. Bush, Ph.D., M.S.

Digna Velez-Edwards, Ph.D.

Copyright © 2014 Jennifer Renée Malinowski

All Rights Reserved

To Adam and Harry, may this serve as a reminder that you can accomplish your dreams,

and

To my husband, Jason, for always supporting my adventures with unyielding love.

ACKNOWLEDGEMENTS

The work in this dissertation was supported, in part, by the NIH Human Genetics Training Grant (T32GM080178). The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center). The Electronic Medical Records and Genomics (eMERGE) Network is funded by NHGRI, with additional funding from NIGMS through the following grants: U01HG04599 and U01HG006379 to Mayo Clinic; U01HG004610 and U01HG006375 to Group Health Cooperative; U01HG004608 To Marshfield Clinic; U01HG006389 to Essentia Institute of Rural Health; U01HG004609 and U01HG006388 to Northwestern University; U01HG004603 and U01HG006378 to Vanderbilt University; U01HG006385 to the Coordinating Center; U01HG006382 to Geisinger Clinic; U01HG006380 to Mount Sinai School of Medicine; U01HG006830 to The Children's Hospital of Philadelphia; and U01HG006828 to Cincinnati Children's Hospital and Boston Children's Hospital. The data and materials included in the age at menarche and age at menopause study result from collaboration between the following studies: ARIC and WHI. The authors thank the staff and participants of the ARIC study for their important contributions. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. The dataset(s) used for the algorithm development analyses described herein were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1TR00045 from NCATS/NIH. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core also provided computational and/or analytic support for those analyses.

This work would not have been possible without the guidance and expertise of my dissertation committee: Dana Crawford, Ellen Clayton, Melinda Aldrich, Digna Velez-Edwards, and William Bush. I greatly appreciate the constructive criticism and helpful advice from each committee member. As Isaac Newton remarked in a letter to Robert Hooke, "If I have seen further it is by standing on the sholders [sic] of Giants." I believe that statement represents that this body of work would not have

been possible without my committee's help and hope that it adequately conveys my appreciation for each member's dedication to my success.

Graduate training can often be a lonely and difficult task; the friendships and scientific discourse from my fellow Center for Human Genetics Research graduate students have helped to make my journey much more pleasant. Present and former members of the Crawford Lab (Sarah Pendergrass, Logan Dumitrescu, Janina Jeff, Sabrina Mitchell, Matt Oetjens, and Nicole Restrepo), Laura Wiley, and Jake Hall were immeasurably helpful throughout this process. Their collegiality and friendship are greatly appreciated.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF APPENDICES

OVERVIEW

The sequencing of the first draft of the human genome in 2003 was met with great enthusiasm from scientists and the general public alike. Heralding a new age of biomedical discovery, the Human Genome Project has brought forth significant advances in genomic technologies, scientific breakthroughs, and has the potential to improve health outcomes and reduce healthcare disparities in myriad ways. Though technological advancements like large-scale genotyping arrays, next-generation sequencing, and whole exome and whole genome sequencing have produced considerable data, from its conception, the Human Genome Project was designed to also consider the ethical, legal, and social implications of this new era in genetics. It is at the intersection of modern genomics and ELSI where this thesis occurs.

Personalized, or precision, medicine can be thought of as the ability to identify the correct medical or lifestyle intervention at the optimal time for each individual. Chapter I provides an overview of personalized medicine: how it is currently implemented and ways in which it differs from current clinical practice. Cancer diagnosis and treatment, and pharmacogenomics are used as examples of how personalized medicine can lead to improved health outcomes while highlighting the challenges that are faced in extending this approach to more common, complex diseases. Women and individuals from diverse populations, historically underrepresented in clinical research, are at risk of widening health disparities unless additional emphasis is placed upon these individuals for future research—putting the "personalization" in personalized medicine. I conclude Chapter I with an overview of the types of statistical models and study designs currently used in genetic studies.

I present the first of three case studies in Chapter II. Women's health, long overlooked by the clinical research enterprise, is now fully recognized as an important facet to understanding the nuances of disease. The timing of the reproductive lifespan, in particular, dictates fertility and influences disease risk. After an overview of the female reproductive lifespan, I consider what genetic variants are associated with age at menarche and age at natural menopause in African American women from two population-based epidemiologic studies: the Atherosclerosis Risk in Communities (ARIC) and the Women's Health Initiative (WHI). Most large scale or genome-wide association studies have been performed in European-descent populations; this association study was the first to examine these traits

in an African American cohort and provided us an opportunity to compare results to prior publications.

In Chapter III, I document the creation of an algorithm to extract ages of reproductive milestones from electronic medical records (EMRs). These data have significant research utility independently and as covariates in genome-wide association studies for a variety of diseases and have the potential to inform clinical care. After the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, secondary uses of EMRs for research have become more common and include selecting cohorts for clinical and research studies and reporting health statistics on aggregated data for disease monitoring. However, these data are not consistently documented and are often missing due to a lack of standardized data fields, requiring data-mining techniques for extraction. I present the development process for a data-mining technique to extract the ages at reproductive milestones from the Vanderbilt University Medical Center (VUMC) Synthetic Derivative, a de-identified version of the VUMC EMR for research purposes, and the performance of this algorithm. I also consider potential uses of the algorithm for personalized medicine and in future genetic studies.

I present the second case study in Chapter IV. Endometrial cancer, the most common gynecologic cancer, affects more than 50,000 women in the United States yearly and is responsible for 8,590 estimated annual deaths. Despite the prevalence of endometrial cancer, few genetic association studies have been performed and the etiology of this complex disease is not fully understood. I begin Chapter IV with an overview of the known molecular mechanisms, environmental risk factors, and genetic associations attributed to endometrial cancer. I hypothesized variants previously associated with other cancers may also play a role in the development of endometrial cancer in a small cohort from VUMC. Using a candidate-gene association study approach, I present the results from our small sample and that of the larger meta-analysis in the Population Architecture using Genomics and Environment (PAGE) Study to which our results were contributed for meta-analysis.

The last case study is presented in Chapter V, where I used a genome-wide association study (GWAS) to identify genetic variants associated with serum thyroid stimulating hormone (TSH) levels in both African American and European descent individuals from the Electronic Medical Records and Genomics (eMERGE) Network. The eMERGE Network is a collaboration of nine medical centers with EMRs linked to biobanks and a coordinating center, which allows investigation of genotype-phenotype associations in larger sample sizes than each individual site can provide. With different EMRs utilized

at each eMERGE member site, phenotype harmonization and data extraction are important considerations for genetic studies in this consortium. TSH levels are measured to diagnose thyroid diseases, such as hypo- and hyperthyroidism and women are disproportionately affected by these disorders. To identify genetic variants associated with the distribution of TSH levels, I performed a GWAS in euthyroid (non-thyroid disease) subjects from the eMERGE Network. I present the results of this analysis for both European descent and African Americans and compare our results to previously published GWAS for this and other related traits. Given that environmental factors, such as BMI, influence TSH factors, I describe an interaction analysis between BMI and single nucleotide polymorphisms (SNPs). Lastly, I examine what role population differentiation plays as a possible reason for the disease burden faced by European descent and African American individuals.

In Chapter VI, I consider the ethical, legal, and social implications of personalized medicine and the analytic evidence supporting its use in the clinical setting for common, complex diseases. The methods currently used to obtain the necessary analytic evidence to recommend genetic testing in this context are both time- and resource-intensive. I propose a rapid structured review model using a hypothetical genetic test for hypothyroidism risk to address this issue and demonstrate how this method can be used to identify gaps in evidence at academic medical centers with limited resources. Finally, I examine in greater detail the ethical, legal and social impacts on both the health care system and the general public that personalized medicine for common, complex diseases will have.

This thesis encompasses several phenotypes that play a role in women's health and highlight the challenges faced in generalizing research findings to diverse populations. Personalized medicine has the potential to reduce health disparities and improve health outcomes, but faces significant barriers. In Chapter VII, I look ahead and discuss how this might occur, from genetic study design, to the role of research findings in clinical care.

CHAPTER I

INTRODUCTION

Personalized medicine

Personalized, or precision, medicine has generally come to mean the use of genetic data to inform clinical care for individual patients, including decision-making for prevention, diagnosis, and treatment (National Human Genome Research Institute2014b). Personalized medicine (PM) gained traction and increased societal awareness after sequencing of the first draft of the human genome was completed: researchers and physicians were motivated to utilize the new wealth of genetic information in the clinical setting to improve health outcomes. New businesses based on bringing genetic testing directly to the masses fought for consumers and against regulation. In the clinical setting, PM is used to determine which chemotherapy a tumor is susceptible to (Kim et al.2013), guide medication choice and drug dosage (Scott et al.2013), and diagnose unknown genetic disorders (Yang et al.2013). Despite direct-to-consumer marketing suggesting your genome can accurately predict risk for developing hundreds of complex diseases and traits at the present time, expansion to the clinical setting of PM for common, complex disease risk prediction and management of care is essentially nonexistent.

Cancer is the poster child for successful integration of PM in the clinical setting. *BRCA1/2* testing is routine in breast cancer diagnosis and is used to determine lifetime risk of developing breast and/or ovarian cancer in affected individuals and their family members (Peshkin et al.2002). Characterizing breast tumors as estrogen receptor positive/negative or overexpressing human epidermal growth-factor receptor 2 (HER2) results in targeted chemotherapies and improved clinical outcomes (Slamon et al.2001). Similarly, non-small-cell lung cancer patients now benefit from the identification of gene expression signatures that are sensitive to synthetic indolotriazine (Kim et al.2013) and from dosing recommendations for fluoropyrimidines based on *DYPD* genotypes (Caudle et al.2013). Numerous cancer and medical centers now advertise the use of a patient's tumor genome

to guide chemotherapy decisions for some types of cancer (The University of Arizona Cancer Center2014;Vanderbilt Ingram Cancer Center2014;University of California2014).

Personalized medicine is also being used to guide medication choice and drug dosage at several institutions (Rossolatos and Aitchison2014). Pharmacogenomics, the study of how the genome affects drug response, seeks to predict which patients will respond favorably, unfavorably, or not at all to a particular medication (National Library of Medicine2014). Though adverse drug events are rare, they are a significant cause of morbidity and mortality in the United States and over 200 drug labels carry warnings suggesting pharmacogenetic testing to guide therapy (Dodson2011). The Clinical Pharmacogenetics Implementation Consortium (CPIC) is a collaboration of investigators who are studying the role of genetics in pharmacology. CPIC publishes recommendations using a standardized format and grading system to evaluate the strength of the genotype-phenotype association (Caudle et al.2014). Nine studies, to date, on the use of genotypes to guide medication choice and dosage have been published by CPIC (Caudle et al.2014), including recommendations for codeine therapy based on *CYP2D6* genotype (Crews et al.2014) and the anti-clotting agent clopidogrel based on *CYP2C19* genotypes (Scott et al.2013). These guidelines are being implemented in the clinical setting at institutions such as Vanderbilt University Medical Center (VUMC) and at St. Jude Children's Research Hospital(Hoffman et al.2014). The VUMC Pharmacogenetic Resource for Enhanced Decisions in Care and Treatment (PREDICT) Program identifies patients at-risk for adverse events and prospectively genotyped them (Pulley et al.2012). Obtaining the genetic data prior to the clinical need allows for physicians to integrate the data into clinical care through decision support mechanisms implemented in the electronic health record.

Current practice

The questions of when to screen a patient for a disorder, how to accurately identify those who are risk of disease, and when to provide prophylactic interventions (e.g., thyroid replacement hormone for subclinical hypothyroidism or statin therapy for hypercholesterolemia) to prevent disease or improve health outcomes are key points of PM research. The decision of whether to screen for a particular disease has primarily relied upon guidelines developed in 1968 by Wilson and Jungner (Table 1) (Wilson and Jungner 1968). These recommendations laid out ten criteria that a screening test should possess before a screening program is initiated, including that the condition should be an important health problem, that there should be a recognizable latent or early symptomatic stage, that a suitable test and treatment be available and are generally accepted by the public (Wilson and Jungner

1968). The Wilson and Jungner guidelines also require that the natural history of the condition, including development of the disease from latency to diagnosis be adequately understood and that the screening program be generally cost-effective compared to medical costs incurred by not screening (Wilson and Jungner 1968). These criteria have been used to justify some public screening programs, (e.g., newborn screening for phenylketonuria (PKU)(Petros 2012)), and not others (screening for adult celiac disease(Evans, Hadjivassiliou, and Sanders 2011)). Advances in genomics have led some to consider if the Wilson-Jungner criteria should be updated in order to be more flexibly applied to genetic testing (Table 1) (Petros 2012;Andermann et al. 2008). These updated criteria may provide an initial guideline when determining which diseases a PM approach is appropriate.

Table 1. Screening criteria for disease.

Wilson & Jungner 1968(Wilson and Jungner1968)	Andermann et al. 2008(Andermann et al.2008)	Petros 2012(Petros2012)
The condition should be an important health problem.	The screening should respond to a recognized need.	The test may be multiplexed or overlaid onto an existing structure or system.
There should be an accepted treatment for patients with recognized disease.	The objectives of screening should be defined at the start of the program.	The "diagnostic odyssey" for the patient/family may be reduced or eliminated.
Facilities for diagnosis and treatment should be available.	There should be a defined target population.	Adverse outcome(s) are rare with a false-positive test.
There should be a recognizable latent or early symptomatic stage.	There should be scientific evidence of screening program effectiveness.	Treatment costs may be covered by third parties (either private or public).
There should be a suitable test or examination.	The program should integrate education, testing, clinical services and program management.	Testing may be declined by parents/guardians.
The test should be acceptable to the population.	There should be quality assurance, with mechanisms to minimize potential risks from screening.	Adequate pretesting information or counseling is available to parents/guardians.
The natural history of the condition, including development from latency to diagnosis, should be adequately understood.	The program should ensure informed choice, confidentiality, and respect for autonomy.	Screening in the newborn period is critical for prompt diagnosis and treatment.
There should be an agreed policy on whom to treat as patients.	The program should promote equity and access to screening for the entire target population.	Public health infrastructure is in place to support all phases of the testing, diagnosis, and interventions.
The cost of case-finding (including diagnosis and treatment of diagnosed patients) should be economically balanced in relation to possible expenditure on medical care as a whole.	The program evaluation should be planned from the start.	If carriers are identified, genetic counseling is provided.
Case-finding should be a continuing process and not a "once and for all" project.	The overall benefits of screening should outweigh the harm.	Treatment risks and the impact of a false-positive test are explained to parents/guardians.
		The limitations of screening and risks of a false-negative test are explained to parents/guardians.

Aspects of health "personalization"

Though the nomenclature may be recent, personalized clinical care has existed for hundreds of years (Murray 2012). Physicians routinely personalize clinical care based on sex, age, race/ethnicity, family history, and environmental exposures. For example, screening for colorectal cancer is age and family history dependent: the United States Preventative Services Task Force (USPSTF) recommends screening beginning at age 50 unless a family history or other risk factors cause a patient to be at higher risk of developing the disease (U.S.Preventative Services Task Force 2008). The pneumococcal vaccination is recommended for all children under age 5 years and adults with certain medical conditions, such as immunodeficiency, sickle cell disease, or chronic lung diseases (National Center for Immunization and Respiratory Diseases 2012).

For some diseases, family and health histories provide sufficient basis to accurately assign risk to an individual. Huntington's disease, an autosomal dominant genetic disorder caused by a triplet repeat expansion in the *HTT* gene, provides an example of this. An individual with an affected parent has a 50% risk of inheriting the mutation and developing the disease. But for more common diseases, such as cancer or type 2 diabetes (T2D), that have both environmental and complex genetic factors and interactions, determining risk for an individual patient can be more challenging.

Sex and race/ethnicity have roles in disease risk, providing an opportunity for PM based, in part, upon those traits (Burchard et al. 2003). Women are disproportionately affected by numerous complex diseases, including autoimmune and reproductive disorders. There are notable differences in the incidences and severity of diseases between men and women, from Alzheimer's disease (Irvine et al. 2012) to inflammatory arthritis (Barnabe et al. 2012), which may stem, in part, from hormone differences between men and women (Carter et al. 2012). Similarly, population-specific genetic differences have already been identified for ECG traits (Ramirez et al. 2011), and age-related macular degeneration (Klein et al. 2011). Given the underlying biological mechanisms for many complex diseases are not fully understood, these sex- and population-specific differences emphasize the benefit in a personalized medicine approach.

Current implementation

EHR utilization

Personalized medicine often relies substantially on software systems that can identify at-risk patients across large patient populations and guide clinical decision-making. Electronic health/medical records (EHR/EMR) systems are a key factor in successful PM initiatives in medical centers. Through the HITECH Act, medical institutions benefit financially from meaningful uses of their EHR systems, such as monitoring communicable disease incidence and immunization statistics for national surveillance programs and recording patient data (e.g., demographic, medication allergy, smoking status), (Stevens et al. 2013; Blumenthal 2010; Blumenthal and Tavenner 2010; Blumenthal 2011; Jha et al. 2009; Kukafka et al. 2007). With bioinformatic and computational biology approaches, EHR systems can scan patient populations to find cohorts for clinical trials, identify patients delinquent in immunization schedules (Stevens et al. 2013), and target interventions to specific populations.

EHR utilization goes beyond the clinical space, however. Researchers also benefit from the ability to access EHRs to perform epidemiologic and genetic studies on patient populations to better understand how genetic variations contribute to health and disease. The Electronic Medical Records and Genomics (eMERGE) Network is an example of a group of medical centers where researchers access EHR data linked to biobanks in order to perform genetic studies (McCarty et al. 2011;Crawford et al. 2014). eMERGE has published numerous studies on a broad range of phenotypes, including hypothyroidism (Denny et al. 2011), low density lipoprotein (LDL) levels (Rasmussen-Torvik et al. 2012), and cardiac conduction (Ritchie et al. 2013). However, this secondary use of clinical data by non-clinical researchers leads to ethical, legal, and social issues (Clayton et al. 2010; Fullerton et al. 2012) that remain to be fully addressed.

Decision support mechanisms

Identification of patients for an intervention is only one half of the PM implementation strategy. Moving from identification to intervention relies on decision support mechanisms. Decision support mechanisms may include computerized alerts, reminders to clinicians, generation of patient data reports, and automatic order set suggestions--approaches that integrate multiple pieces of healthcare data and may involve strategies to engage the patient in the healthcare decision making process (Downing et al. 2009). These may be built into the EHR system or exist as a secondary system

depending on the use. For example, at VUMC, prescriptions are ordered electronically, allowing for a decision support mechanism to alert a prescribing clinician when a medication that is contraindicated for a patient due to genetics or drug allergy has been ordered (Pulley et al. 2012). This alert can provide the relevant data for the clinician about the contraindication and suggest alternate medications or dosing strategies (Pulley et al. 2012). Monitoring patients with chronic health conditions, such as asthma or T2D, is a key attribute of the Kaiser Permanente health system (McCarthy, Mueller, and Wrenn 2009). This allows Kaiser to target information to their patients through multiple methods and interactions with healthcare providers; current smokers may receive information about smoking cessation programs and asthmatics may receive information about reducing the frequency of attacks (McCarthy, Mueller, and Wrenn 2009).

Despite the potential to improve health outcomes using decision support mechanisms, there are numerous challenges to implementing a decision support process. EHR systems may not have been designed for this type of use and require modification or replacement and a lack of interoperability between specialized EHR systems may impede clinical decision support (Blavin et al. 2013). Even when a decision support system is implemented at an institution, the decision to change orders based on an alert may not occur. Alert fatigue, a phenomenon where the physician becomes desensitized to decision support alerts and ignores the information, is a key problem (Ancker et al. 2014; McCoy et al. 2014). Additionally, ethical and legal concerns about clinical decision support systems have been noted by several (Berner 2002; Goodman 2007; Castillo and Kelemen 2013). The reliability and accuracy of the clinical decision support system needs to be verified and its limitations communicated to the end users (Castillo and Kelemen 2013). The challenges noted for clinical decision support systems and the use of EHRs should be addressed for successful implementation of personalized medicine.

Current examples

Despite the abovementioned challenges, there are numerous examples of successful implementation of PM for pharmacogenetics and cancer treatment.

Pharmacogenetics

One use of pharmacogenetics is to match drug therapy to the patient in which it will be effective. Cystic fibrosis, an autosomal recessive genetic disorder caused by mutations in *CFTR,* is associated with a reduced lifespan and pulmonary events such as mucus buildup, infection,

inflammation, bronchiecstasis and respiratory failure (Rowe, Miller, and Sorscher 2005). Approximately 1,900 variants in *CFTR* have been associated with the disease (Online Mendelian Inheritance in Man 2014) and the variants can be grouped into classes based on the mechanism of CFTR defect (Clancy et al. 2014). Current treatment of cystic fibrosis relies primarily on targeting the symptoms resulting from the *CFTR* mutations; however, there is interest in approaches that restore function to the mutant CFTR protein (Clancy and Jain 2012). Ivacaftor, is the first FDA-approved drug to target a specific *CFTR* defect: gating of CFTR at the plasma membrane (Van et al. 2009). The effectiveness of ivacaftor relies on CFTR expression on the cell surface and the ability of CFTR activation through normal intracellular signaling mechanisms (Eckford et al. 2012); therefore, only patients who are heterozygous or homozygous for the *G551D-CFTR* variant are recommended for ivacaftor therapy (Davies et al. 2013; Ramsey et al. 2011; Accurso et al. 2010).

Pharmacogenetics has also been used to identify which patients are at greater risk of adverse drug events. Codeine, an opioid analgesic, is activated by cytochrome p450 2D6 (*CYP2D6*). Variants in *CYP2D6* have been associated with variations in drug efficacy and toxicity (Crews et al. 2014). Individuals with decreased codeine metabolism have demonstrated poor analgesic effects from codeine (Lotsch et al. 2009) and severe or life-threatening toxicity following normal doses have been observed in fast metabolizers (Gasche et al. 2004). As a standard starting dose can result in toxicity for ultrafast metabolizers, identifying these at-risk patients by their *CYP2D6* genotypes may decrease potential adverse events through administration of alternate analgesics (Crews et al. 2014).

Cancer treatment

Cancer medicine is the area where personalized medicine has arguably been the most successful. Substantially increased lifetime risk of breast and ovarian cancers has been associated with mutations in *BRCA1* and *BRCA2* (National Cancer Institute at the National Institutes of Health 2014a). Genetic testing for *BRCA1/2* variants may enable carriers of deleterious mutations to seek chemoprevention, intensive cancer screening (Peshkin et al. 2002), or prophylactic surgery to reduce their risk of developing breast or ovarian cancers (U.S.Preventative Services Task Force 2013). Gene-expression data has also contributed to the success of PM in cancer care. Increased expression of the human epidermal growth-factor receptor 2 (HER2) is present in approximately one quarter of breast cancers and is indicative of an aggressive subtype of disease with poor prognosis (Slamon et al. 1987).Testing breast cancer tumors for increased HER2 expression allows oncologists to use

trastuzumab antibody therapy in these patients, which has significantly improved health outcomes for this subset of breast cancer patients (Slamon et al. 2001).

Other cancers have also benefitted from personalized medicine approaches, including guiding treatment decisions and determining disease prognosis. Non-small cell lung carcinoma is a type of lung cancer where the tumor develops in the small alveoli of the lungs (National Cancer Institute at the National Institutes of Health 2014c). Prognosis and treatment depend upon whether the tumor has mutations in particular genes, such as the epidermal growth factor receptor (*EGFR*) or the anaplastic lymphoma kinase (*ALK*) gene. Specific *ALK* mutations are used to guide therapy with crizotinib, a tyrosine kinase inhibitors in lung cancer. Additionally, *ALK* overexpression has been associated with poor prognosis in colorectal cancer (Bavi et al. 2013).

Common, complex diseases

Personalized medicine approaches, (i.e., patient-finding, treatment guiding, and intervention timing) have not been as successful for other non-cancer common, complex diseases. Though the public health burden from common diseases like cardiovascular disease, T2D, asthma, or autoimmune disorders is substantial, PM approaches have not generally been implemented on a population-level scale. Common disorders result from both genetic and environmental factors and complex interactions of those factors. The biologic mechanisms for these diseases are not fully understood, so identifying the optimal time and type of intervention for PM is difficult. Randomized clinical trials, considered to be the "gold standard" in evaluating an intervention, have been performed for pharmacogenetics and cancer treatments, but are rarely utilized for other common, complex diseases. The downstream consequence of this is a lack of evidence resulting in challenges to establishing policy for PM.

Despite these challenges, PM approaches have been used successfully for age-related macular degeneration. Age-related macular degeneration (AMD) is a phenotypically heterogeneous ocular disease characterized by central vision loss, from damage to the macula that presents with one of two subtypes: wet or dry (National Eye Institute at the National Institutes of Health 2013). Genetic (e.g., *ARMS2/HTRA1, CFH*) and environmental (e.g., cigarette smoking, elevated body mass index) factors contribute to disease development (Chakravarthy et al. 2013; Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005; National Eye Institute at the National Institutes of Health 2013). A recent study has used a PM approach to classify neovascular AMD (nAMD) patients into subtypes for potential therapeutic interventions. Feehan et al. classified affected AMD patients into four discrete clusters

based on phenotypic and genotypic heterogeneity (Feehan et al. 2011). History of hypertension or of hypercholesterolemia were significant risk factors, as was the *ARMS2/HTRA1* rs1049331 TT genotype (Feehan et al. 2011). Currently, injectable anti-VEGF therapies are used to treat nAMD; however, it has been proposed that these therapies are contraindicated in patients with cardiovascular risk factors (Enseleit, Michels, and Ruschitzka 2010). The ability to classify patients into these clusters, based on both environmental and genetic risk factors, may provide clinicians an opportunity to target pharmacologic treatments to specific patients and not others.

Response to AMD therapy based on genotype was investigated by Wang et al. (Wang et al. 2012). Wang et al. identified an association between response to ranibizuman/bevacizumab therapy in nAMD patients and *PLA2G12A* rs2285714, though this association was not significant after Bonferroni correction for multiple testing (Wang et al. 2012). Interactions between genetic risk (*CFH* rs1061170 or *ARMS2* rs10490924) and environment (dietary intake of antioxidants, zinc, and omega-3 fatty acids) were evaluated in a recent study (Wang et al. 2014a). Participants were grouped by number of risk alleles in *CFH* or *ARMS2* and a regression analysis was performed to identify associations between genetic risk and the environmental variables. The authors found interactions between intake of antioxidants and fish consumption with decreased AMD risk, but only in participants with two or more risk alleles in *CFH* or *ARMS2* (Wang et al. 2014a). This association highlights the potential opportunity for clinicians to identify at-risk patients who are most likely to benefit from specific interventions and demonstrates the use of PM for a common, complex disease.

Personalization: Understanding Women's Health

Tailoring clinical care to an individual can take many forms and may be based on several factors, including genetics, ethnicity, and sex. Known differences in disease prevalence and severity between men and women, and female-specific issues of reproduction and the reproductive lifespan, emphasize the need to perform disease research in women as well as men.

Historic research inequality

Women comprise more than 50% of the US population (Bureau of the Census 2011; U.S.Department of Commerce 2014) and there are notable differences in the incidences and severity of diseases between men and women, from Alzheimer's disease (Irvine et al. 2012) to inflammatory arthritis (Barnabe et al. 2012). Parity may affect disease manifestation or outcomes such as body mass

index (BMI) (Bobrow et al. 2013) and diabetes (Simons et al. 2012). The timing of the reproductive lifespan is associated with numerous disease traits; an earlier age at menarche and/or later age at menopause increases risk for several cancers, while an earlier age at menopause place women at increased risk for osteoporosis and cardiovascular disease, in addition to decreased fertility. It is unclear to what extent sex-specific genetic architecture and environmental factors (such as hormone fluctuations or behavior) contributes to observed disease prevalence or severity variation. Gene regulation differences and genotype-sex interactions have been suggested to play a role in sex-specific heritability for some diseases (Ober, Loisel, and Gilad 2008). Nevertheless, only in the last few decades has the importance of women's health and physiologic differences between males and females in the research setting come to the forefront of researchers and government agencies (Taylor 1994).

Endometrial Cancer research

One example of a female-specific disease is endometrial cancer (EC). Endometrial cancer is the most common invasive gynecological cancer and has the fourth highest cancer incidence rate in the United States, with an estimated 52,630 new cases and 8,590 deaths in 2014 (National Cancer Institute at the National Institutes of Health 2014b). Worldwide, an estimated 287,000 new cases arise yearly, with twice the age-standardized incidence rate in developed countries as in developing countries (Jemal et al. 2011). Diagnosis of EC currently occurs after the patient is symptomatic. There is no standard screening test, making the development of biomarkers (including genetic risk scores) very important. Common forms of EC can be roughly categorized into three distinct subtypes: endometrioid, serous, and clear cell. Several rarer histological subtypes (e.g., mixed Müllerian, squamous cell carcinoma) also exist. Endometrioid cancer is estrogen dependent and is the most common of the subtypes; prognosis for endometrioid EC is often the best of the common subtypes and many of the cancers are diagnosed at early stages. Serous and clear cell carcinoma are both estrogen-independent and are typically higher grade; outcomes for these cancers are worse than for endometrioid cancers, even when adjusted for stage (O'Hara and Bell 2012). EC risk increases with age, type 2 diabetes, obesity, hypertension, and gynecological disorders such as polycystic ovarian syndrome (PCOS), endometriosis, and uterine fibroids. Smoking, nulliparity, and hypertension are associated with increased risk of EC, while parity and oral contraceptive use have a protective effect against EC (Haidopoulos et al. 2010).

Genetic contributions to endometrial cancer

A family history of EC is a known risk factor for the disease and its heritability has been estimated at 52% (Schildkraut, Risch, and Thompson 1989). Several studies performed in women with European ancestry have reported odds ratios (OR) ≈ 2.0 for EC for women with a family history of EC and/or colorectal cancers, however, results have been largely inconsistent (Lucenteforte et al. 2009). It has been suggested that a family history of different cancers may predispose women to EC, an effect more pronounced in young women diagnosed at an early (<40 years) age (Lucenteforte et al. 2009). Lynch syndrome has been well documented to increase risk of several cancers, including EC, providing further evidence of a genetic contribution to EC through mutations in DNA mismatch repair genes (Wang et al. 2013; Gruber and Thompson 1996; Ma, Ledbetter, and Glenn 2013). HER-2/neu expression, also implicated in breast cancer, has been found to be associated with EC overall, cancer-related, and disease-free survival (Kalogiannidis et al. 2014).

Several candidate gene association studies (CGAS) have been published for EC (Lee et al. 2010; Ashton et al. 2009b; Chen et al. 2012b; Xu et al. 2009a; Xu et al. 2009c). Given that endometrioid ECs are estrogen-dependent, one CGAS found *ESR1* and *ESR2* variants to be associated with increased EC risk, an effect which was increased in subjects with both polymorphisms (Ashton et al. 2009a). Variants in *PGR*, the progesterone receptor, have been associated with increased EC risk (Lee et al. 2010; O'Mara et al. 2011a); though a recent meta-analysis of eight studies found no significant association between the *PGR +331G/A* polymorphism and endometrial cancer risk (Pabalan et al. 2014). Polymorphisms in caspases, more broadly associated with the progression of cancer through the apoptotic pathway, were found to be associated with increased risk of EC in a cohort of Chinese women (Xu et al. 2009a). Obesity is a known risk for EC and is associated with adiponectin and leptin levels; it was hypothesized that obesity-related genes may play a role in EC risk (Chen et al. 2012b). Chen *et al.* identified variants in both *LEP* and *ADIPOQ* associated with reduced EC risk in their Chinese cohort (Chen et al. 2012b). In general, though several studies have identified variants associated with EC risk, these have typically been unreplicated or of mixed results, and represent a small fraction of genes involved in cancer initiation and progression pathways.

Four genome-wide association studies (GWAS) have been published for EC (Spurdle et al. 2011; Long et al. 2012; De, I et al. 2014; Kuhn et al. 2012). These studies have identified a single significant locus, *HNF1B* rs4430796, associated with EC after multiple testing corrections ($p=7.1x10^{-10}$) (Spurdle et al. 2011; De, I et al. 2014). Three SNPs, all in LD in the first four exons of *HNF1B*, were associated with

endometrial cancer at $p<10^{-07}$ in a European-descent population; one of these SNPs was tested in a Chinese cohort from Shanghai (Shanghai Endometrial Cancer Genetics Study), but failed to generalize to that population in a supplemental analysis (Spurdle et al. 2011). The second GWAS published for endometrial cancer performed a multiple-stage analysis, beginning with the Shanghai cohort from the first GWAS, and identified a variant near *CAPN9* on chromosome 1 associated with EC, an effect that was more significant when the analysis was limited to endometrioid ECs (Long et al. 2012). A replication study was performed in two studies from the PAGE Consortium; two SNPs, rs4430796 and rs7501939, identified in the Spurdle et al. GWAS for EC were tested for association with EC in 1,357 incident cases of invasive EC in women of diverse ethnic ancestries from the Multiethnic Cohort Study (MEC) and the Women's Health Initiative (WHI) (Setiawan et al. 2012). These SNPs were found to be protective against EC (both type I and type II tumors) in their overall study population and in women of European-descent, with similar trends found in the African American, Asian/Pacific Islander and Latina women (Setiawan et al. 2012). Despite the association between EC and *HNF1B*, no other variants have been associated with EC at genome-wide significance, suggesting common variants may not explain a significant amount of heritability for EC (Chen et al. 2014b). Therefore, an exome-wide association study (EXWAS) design was used to identify rare variants associated with EC in multiethnic participants from the Epidemiology of Endometrial Cancer Consortium (E2C2) (Chen et al. 2014b). No variants in the EXWAS reached genome-wide significance after Bonferroni adjustment for multiple tests, nor did a gene-based analysis identify significant associations with EC; however this study was small and powered only to detect ORs>2.53 for low frequency (MAF<0.02) variants (Chen et al. 2014b). The limited variants consistently identified through a relatively small number of genetic studies highlight the opportunities for additional studies to meaningfully contribute to the field.

Timing of the Reproductive lifespan

Age at menarche (AM) and age at natural menopause (ANM) define the boundaries of the reproductive lifespan in women. Cross-sectional and longitudinal studies have shown a recent secular trend of earlier attainment of pubertal milestones (breast development, appearance of pubic hair, menarche) from the 1960s to present (Kaplowitz 2006; Herman-Giddens 2006). This tendency is accelerated in girls of African American and Hispanic ancestry, a bias that remains after adjusting for socioeconomic variables and body mass index (BMI) (Wu, Mendola, and Buck 2002). Known environmental modifiers of age at menarche include exposure to organochlorine chemicals and polybrominated biphenyls (Wolff and Landrigan 2002; Blanck et al. 2000). Increased BMI is associated

with earlier AM (Biro, Khoury, and Morrison 2006) and later ANM (Palmer et al. 2003), while active smoking is associated with earlier ANM (Gold 2011). These and other known environmental factors explain only a small amount of individual variation in the timing of these reproductive measures.

Menarche

The gonadotropin-releasing hormone (GnRH) can be considered the 'master switch' controlling the timing of puberty and release of hormones from the hypothalamus. Secretion of GnRH varies with age; in utero release of GnRH leads to pituitary release of follicle-stimulating hormone (FSH) and luteinizing hormone (LH). This hormonal cascade reaches a peak at about the same time the maximum number of oocytes is reached (DiVall and Radovick 2008). LH and FSH levels begin to decline as placental estrogen provides negative feedback on the hypothalamus-pituitary-gonadotropin (HPG) axis. After birth and the removal of placental estrogen, FSH and LH levels rise again, only to fall in the first two years of life to nearly undetectable levels. From this point, the HGP axis enters what is known as the 'juvenile pause,' a state of relative quiet in the HPG axis (Nathan and Palmert 2005).

The process of puberty begins some time prior to menarche/initiation of menstruation. Kisspeptins, encoded by the *KISS1* gene, stimulate pulsatile release of GnRH from the hypothalamus to the pituitary (Okamura et al. 2013); pituitary production of FSH and LH increases (Garcia-Galiano, Pinilla, and Tena-Sempere 2012; Matzuk and Lamb 2008). FSH leads to maturation of the follicles in the ovary and production of estradiol and anti-Mullerian hormone (AMH) (Burger et al. 2007; Hale et al. 2007). Estradiol is the strongest of three types of estrogens: estradiol, estrone, and estriol. Estradiol is the primary circulatory estrogen before the onset of menopause. In the pubertal girl, circulating estrogen leads to uterine and breast growth, and the lining of the uterus, the endometrium, becomes vascularized. AMH acts to inhibit overstimulation of follicles due to the increased levels of FSH. LH increases lead to ovulation and maintenance of the endometrium. In absence of pregnancy and hormone levels to sustain the vascularized tissue, the endometrium is sloughed off and menstruation occurs. It has been hypothesized that the first few anovulatory menses act to mature the HPG axis (Henriet, Gaide Chevronnay, and Marbaix 2012). Though ovulation may occur at the time of menarche, it is not common, and cyclic menstruation may occur in its absence (Zhang et al. 2008). As the HPG axis matures, menstrual cycles become regular and the cyclic rise and fall of the various hormones stabilizes into a regular pattern (Ruiz-Alonso, Blesa, and Simon 2012).

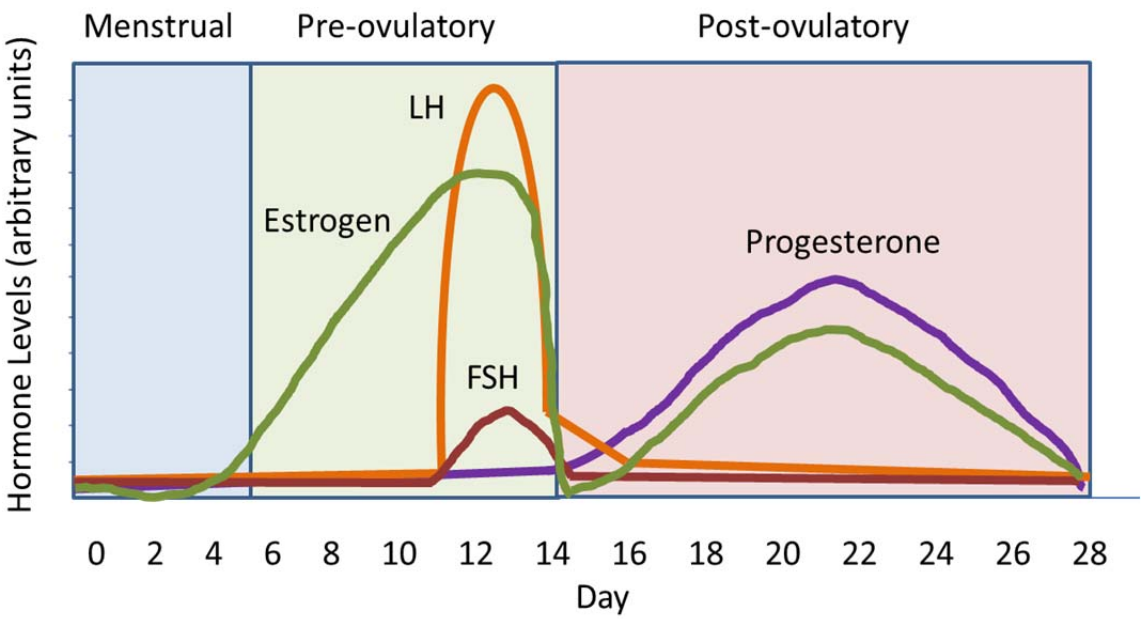


Figure 1. Hormonal fluctuations during the menstrual cycle.

Shown are the relative hormonal fluctuations that occur during a typical (non-disordered state) 28-day menstrual cycle. Colored boxes represent the three phases of the menstrual cycle as labeled above the figure. Abbreviations: luteinizing hormone (LH); follicle stimulating hormone (FSH). Adapted from Ruiz-Alonso et al. 2012.

In the clinical setting, female puberty can be subdivided into five stages based on breast (thelarche) and pubic hair (pubarche) development (Table 2) (Marshall and Tanner 1969; Tanner and Whitehouse 1976). Menarche, the initiation of the menstrual cycle, occurs typically at some time after Tanner stage 3 or 4, at which point breast and pubic hair development are continuing, but not mature (Tanner stage 5) (Marshall and Tanner 1969). In the Harpenden Growth Study, a longitudinal study of British girls from an institutional setting, the mean age at menarche was 13.47 years, with a standard deviation of 1.02 years (Marshall and Tanner 1969).

Table 2. Tanner stages of female puberty.

Tanner Stage	Breast Development	Pubic Hair Development
1	pre-adolescent; papilla elevation only	pre-adolescent; no pubic hair
2	breast bud stage; enlargement of breast and papilla as a small mound; enlargement of areola diameter	sparse growth of long, slightly curled, and slightly pigmented hair; appears mainly on the labia
3	continued enlargement of breast and areola	darker, coarser, and more curled hair spread sparsely
4	projection of papilla and areola to form a secondary mount above the level of the breast	hair is adult in type but coverage is significantly less than in adults
5	mature breast	adult in quantity and type, spread to medial surface of the thighs but not above the base of the inverse triangle

Concern of a secular trend toward earlier attainment of the stages of pubertal development served as the impetus for several studies (Sun et al. 2005; Sun et al.2002). Using data collected from the National Health and Nutrition Examination Survey (NHANES) III (1988-1994), differences in timing of menarche and breast/pubic hair development were found between non-Hispanic blacks, Mexican-Americans, and non-Hispanic whites (Chumlea et al. 2003; Sun et al. 2002). Comparing the sexual maturity data collected from the National Health Examination Survey (NHES) 1966, the Hispanic Health and Nutrition Examination Survey (HHANES) 1982-1984, and the NHANES III (1988-1994), no obvious secular trend for earlier attainment of breast or pubic hair development was found for non-Hispanic blacks or non-Hispanic whites between NHES (1966) and NHANES III (1988-94) (Sun et al. 2005). In Mexican-American girls, a greater proportion had attained Tanner stage 2 or higher in NHANES III than in HHANES; however, completion of pubertal development (Tanner stage 5) occurred later in NHANES III compared to HHANES (Sun et al. 2005). Age at menarche for all girls declined slightly in NHANES III (median: 12.43 years) compared to NHES (median: 12.77 years) (Sun et al. 2005). Non-Hispanic blacks reached menarche earliest of the three groups (median: 12.06 years), compared to non-Hispanic whites (median: 12.55 years) or Mexican-Americans (median: 12.25 years). The largest difference in age at menarche between NHES 1966 and NHANES III 1988-94 was observed in non-Hispanic blacks (difference: 0.46 years)(Sun et al. 2005). Despite the earlier age at menarche and breast development observed in NHANES III compared to earlier population surveys, the authors determined overall pubertal development had not substantially declined(Sun et al. 2005). Similarly,

data from NHANES (1999-2004) demonstrated a downward trend in age at menarche for all women and within each racial/ethnic category(McDowell, Brody, and Hughes 2007).

In the Copenhagen Puberty Study, using a cross-sectional population study design, investigators collected data for Tanner stages of puberty, weight and height, and hormone levels (estradiol, serum follicle-stimulating hormone (FSH), and luteinizing hormone (LH)) (Aksglaede et al. 2009). The findings were similar to those found in the US study (Sun et al. 2005); the mean age at menarche declined from 13.42 years to 13.13 years and Tanner stage 2 (breast) declined from 10.88 years to 9.86 years (Aksglaede et al. 2009). Increased gonadotropins were not associated with the observed earlier breast development; however, there was a small but significant decrease in estradiol levels among 8-10 year olds (Aksglaede et al. 2009). Notably, adjustment for BMI did not change the significance of the menarche or pubic hair development results (Aksglaede et al. 2009). These data, considered together, suggest that earlier breast development may occur as a result of estrogenic actions, rather than earlier activation of the HPG axis, and this is not associated with increasing incidence of childhood obesity (Aksglaede et al. 2009).

The Avon Longitudinal Study of Parents and Children (ALSPAC) evaluated the timing of puberty in a British cohort (Christensen et al. 2010b). This study used questionnaires mailed to the participants to obtain self-reported/parent-reported data on Tanner stages and age at menarche; drawings were provided to aid in the determination of Tanner stage (Christensen et al. 2010b). Several characteristics were assessed for association with Tanner stage of the child, including mother's pre-pregnancy BMI, mother's age at delivery, mother's age at menarche, mother's level of education, social class, child's birth weight, race, birth order, and BMI at time of questionnaire (Christensen et al. 2010b). The median age at menarche in ALSPAC was 12.9 years, though this was found to vary according to whether the participant began thelarche before pubarche, pubarche before thelarche, or entered thelarche and pubarche simultaneously (Christensen et al. 2010b). Unlike the Copenhagen cohort, increased BMI was associated with higher Tanner stages for breast and pubic hair development (Christensen et al. 2010b).

Though several studies have found a decrease in median age at menarche and earlier attainment of Tanner stages, the cause of this is not well understood. BMI is not consistently associated with earlier thelarche, pubarche, or menarche (Cousminer et al. 2014). Others have considered the influence of environmental factors such as exposure to chemicals that mimic estrogenic compounds (e.g., bisphenol A (BPA), phthalates) (Wolff and Landrigan 2002; Blanck et al. 2000). It is unclear if

interactions between genetic variants and environmental factors play a role in the timing of puberty. Given the impact on future disease risk associated with timing of pubertal milestones, additional studies that examine both the genetic and environmental components of puberty are warranted.

Menopause

The median age at which menopause occurs is 51 years, though unlike the relatively short timeframe for menarche, there is substantial variation (Cramer and Xu 1996; Gold 2011). At the start of the menopausal transition, the follicular count continues to decrease and AMH levels are low (Burger et al. 2007; Harlow et al. 2012). Cycle length does not change initially. However, changes in the length of the cycle and FSH levels in the early follicular phase of the cycle increase as a woman approaches menopause. As the menopausal transition progresses, variability is seen in the length of the cycle >7 days from normal and is persistent across multiple cycles as the levels of FSH increase but are more variable (Harlow et al. 2012; Burger et al. 2007). As FSH levels continue to rise, estrogen levels further decline and missed menstruation (amenorrhea) occurs, though some women may experience normal cycles and normal or increased levels of estradiol up to age 55 (Harlow et al. 2012; Weiss et al. 2004; Santoro and Randolph, Jr. 2011). The Stages of the Reproductive Aging Workshop (STRAW) group have delineated the reproductive lifespan and menopausal transition into three main phases with seven stages based on frequency and variation of the menstrual cycle and supporting evidence from antral follicle counts and FSH and AMH levels (Table 3) (Harlow et al. 2012). The final menstrual period marks the boundary of the menopausal transition into post-menopause. At the end of the menopausal transition, the antral follicle count is very low, FSH levels continue to rise before stabilizing, and both estradiol and AMH levels are very low (Harlow et al. 2012).

Table 3. Stages of the Reproductive Lifespan

	Reproductive				**Menopausal Transition**		**Postmenopause**			
Stage	-5	-4	-3b	-3a	-2	-1	+1a	+1b	+1c	+2
Menstrual Cycle	variable to regular	regular	regular	some changes to flow	variable length	amenorrhea ≥60 days				
Endocrine Levels										
FSH			low	variable	variable, elevated	elevated, >25 IU/L	variable, elevated		stabilizes	
AMH			low	low	low	low	low		very low	
Inhibin B				low	low	low	very low		very low	
Antral Follicle Count			low	low	low	low	very low		very low	

Shown are the stages of the reproductive lifespan, identified by the Stages of Reproductive Aging Workshop (STRAW). Menstrual Cycle, selected hormone levels and antral follicle counts are presented. Stage (-5) begins with menarche. The postmenopausal period is divided into two main stages; stage (+1a) begins after a 12 month period of amenorrhea and defines that the final menstrual cycle has occurred. Shaded stage boxes indicate perimenopause. Abbreviations: follicle stimulating hormone (FSH), anti-Mullerian hormone (AMH). Adapted from Harlow et al. 2012.

Genetic contributions to timing of the female reproductive lifespan

The genetic component for the timing of menarche and natural menopause has been investigated in multiple twin, family, and large population studies, with heritability estimates of nearly 50% for both AM and ANM (He and Murabito 2012). The associations between AM/ANM with disease underscore the importance of elucidating the mechanisms responsible for timing of these events and the genetic predisposition to timing which could affect future disease risk.

Genome-wide linkage analyses (GWLA) have been used to identify regions of the genome linked to AM and ANM—four for AM and two for ANM (Guo et al. 2006a; Anderson et al. 2008; Rothenbuhler et al. 2006; Murabito et al. 2005; Pan et al. 2008; van Asselt et al. 2004). All of the GWLA were performed in European-descent populations (He and Murabito 2012). There was little concordance in the results of the GWLA; only one region was identified in two separate GWLA for AM (22q13) (Pan et al. 2008; Guo et al. 2006a). Furthermore, these results have not been replicated in other studies, nor have the causal genes under the linkage peaks been identified. Multiple candidate gene association studies have also been performed for AM and ANM, many of these in the last few years (reviewed in (He and Murabito 2012)). While earlier candidate gene studies focused on estrogen biosynthesis pathways and yielded inconsistent findings, more recent studies have considered other biological pathways with some concordance. In a study evaluating biologic pathways for associations with AM and ANM, *FSHB* was associated with later onsets of menarche and menopause, and *ESR2* was associated with AM (He et al. 2010). In general, although linkage and candidate gene studies have identified several potential associations with AM and/or ANM, a lack of replication and inconsistency has resulted in few variants of accepted significance.

Genome-wide association studies (GWAS) have identified many novel loci associated with AM or ANM. Four GWAS published simultaneously for AM identified many of the same SNPs (He et al. 2009a; Ong et al. 2009; Sulem et al. 2009; Perry et al. 2009). Variants in or near *LIN28B* (6q21) and 9q31.2 reached genome-wide significance in these studies; however, they were responsible for less than 1.0% of the variation in AM (Perry et al. 2009; Ong et al. 2009). The International ReproGen Consortium published a meta-analysis of 87,802 women with a replication cohort of 14,731 women, all of European ancestry. This meta-analysis confirmed the associations between *LIN28B*/6q21 and 9q31.2 with AM and identified 30 novel associations and ten suggestive associations with AM (Elks et al. 2010). Despite this increase in sample size, these 42 variants account for <10% of the observed variation in AM (Elks et al. 2010).

The findings from GWAS on ANM are similar to those for AM, with little concordance between studies. Studies have identified the same locus (19q13.42/*BRSK1*) at genome-wide significance (He et al. 2009a; Stolk et al. 2009). A recent meta-analysis from the ReproGen Consortium in 38,968 women with replication in 14,435 women, all of European ancestry, confirmed associations with four menopause loci and identified thirteen novel ANM associations in DNA repair and immune pathways (Stolk et al. 2012). Additionally, the WHI-SHARe (Women's Health Initiative-SNP Health Association Resource) has generalized associations with eight menarche loci and two menopause loci to a cohort of Hispanic ancestry (Chen et al. 2012a). A pentanucleotide repeat polymorphism in *SHBG* was associated with an earlier age at menopause in a Greek study, though this effect was slight and the study size small (n=210) (Markatseli et al. 2014).

A lack of data for genetic variants associated with AM/ANM in non-European-descent populations has been highlighted as a known barrier to identifying loci associated with these traits in all race/ethnicities, and emphasizes the need to extend GWAS analyses to more diverse populations (Dvornyk and Waqar 2012). This disparity in research has recently been addressed in a GWAS meta-analysis in African American women (n=6,510) from eleven studies across the US (Chen CT et al. 2014). Though no new associations with age at menopause were identified, the authors generalized to their participants six variants previously associated with ANM in European cohorts (Chen et al. 2014a).

A recent study has assessed the interaction between cigarette smoking and genetic variants on the timing of natural menopause in European-Americans (Butts et al. 2014). Smoking is a risk factor for natural menopause, decreasing the ANM by 1-2 years (Cramer et al. 1995; Gold et al. 2001; Sievert et al. 2013). It has been hypothesized that smoking influences age at menopause through a mechanism that results in cytotoxicity to oocytes, leading to oocyte depletion to hastening time to menopause (Mattison, Nightingale, and Shiromizu 1983; Jurisicova et al. 2007; Matikainen et al. 2001) or though chemicals in cigarettes that lead to hypoestrogenism (Zhu and Conney 1998; Michnovicz et al. 1986). Candidate genes were selected for their association with timing of menopause or bioactivation of polyaromatic hydrocarbons that are in cigarettes (Butts et al. 2014). Significant associations of earlier age at menopause were found in carriers of *CYP1B1*3* or *CYP3A4*1B* who were smokers, resulting in a more than two-fold increased risk of menopause than nonsmokers (Butts et al. 2014). However, this was a small study (n=410) and the findings only applied to the European-American participants (n=205), not the African Americans in their study, and the strata for each genotype and smoking status

were small (Butts et al. 2014). Nevertheless, these findings suggest interactions between genetic variants and environmental factors may play a role in the timing of menopause.

Autoimmunity and Thyroid disease

Autoimmune disorders are a group of disorders characterized by a loss of immune tolerance to self antigens (Murphy, Travers, and Walport 2008). Autoimmune diseases can be generally organized into two groups: organ-specific and systemic; Graves' disease and Hashimoto's thyroiditis are examples of organ-specific autoimmune disorders (Murphy, Travers, and Walport 2008). For many autoimmune diseases, women are disproportionately affected (Lawrence et al. 1998; Weyand et al. 1998) and there may be differences in disease incidence across racial/ethnic groups (Murphy, Travers, and Walport 2008; Cooper and Stroehla 2003; Okayasu et al. 1994). Though the human leukocyte antigen (HLA) region of chromosome six has been implicated in many autoimmune disorders (reviewed in (Gough and Simmonds 2007)), these diseases are not fully understood and represent an important area in women's and minority health research.

Epidemiology of autoimmune thyroid disease

The thyroid is a small, butterfly-shaped gland on the front and sides of the neck. As a part of the hypothalamus-pituitary-thyroid (HPT) axis, the thyroid helps to regulate many physiological processes, including metabolic processes and mitochondrial function (Bassett and Williams 2008; Costa-e-Sousa RH and Hollenberg 2012; Vidali et al. 2014). Thyroid disease is a broad term that encompasses several related, though clinically distinct, disorders including thyroid nodules and goiter, thyroid cancer, congenital hypothyroidism, hyperthyroidism, and hypothyroidism. Both hyperthyroidism and hypothyroidism may develop from autoimmune causes or from environmental factors, such as iodine insufficiency; in developed countries, where iodine is replete, autoimmune causes are the most common (Vanderpump 2011). Both hyperthyroidism and hypothyroidism are diagnosed by measuring the serum thyroid stimulating hormone (TSH), free thyroxine (free T4), and free tri-iodothyronine (free T3) levels (Vanderpump 2011;National Endocrine and Metabolic Diseases Information Service (NEMDIS), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and National Institutes of Health 2014) (Table 4).

Table 4. Reference ranges and typical thyroid function test results.

	TSH (mIU/L)	free thyroxine (fT4)(ng/dL)	free triiodothyronine (fT3) (pg/mL)
Euthyroid (normal)	0.5-4.50	0.5-1.2	2.3-4.2
Hypothyroid	↑	↓	↓
Hyperthyroid	↓	↑	↑

The reference range for TSH reflects the position of a joint conference of the American Association of Clinical Endocrinologists, the American Thyroid Association, and The Endocrine Society (JAMA 2004; 291:228). Arrows indicate direction of test value from the reference level in a typical thyroid diseased patient. Adapted from: (National Endocrine and Metabolic Diseases Information Service (NEMDIS), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and National Institutes of Health 2014)

Hypothyroidism is most commonly caused by Hashimoto's disease and is four times more common in women than in men; it is primarily seen in middle-aged women, though it may occur at any age in either sex (Vanderpump 2011; Hollowell et al. 2002). Symptoms of hypothyroidism include fatigue and/or muscle weakness, sensitivity to cold, constipation, unexplained weight gain, excessive or prolonged menstruation, and depression (Dubbs and Spangler 2014). Complications from hypothyroidism include development of a goiter, increased risk of heart disease (Bai et al. 2014), depression and other mental health issues, and myxedema; untreated myxedema is critical and may result in myxedemic coma or death (Dubbs and Spangler 2014). Pregnant women with untreated Hashimoto's disease are at increased risk of adverse maternal and fetal outcomes including recurrent miscarriage and birth defects (Nathan and Sullivan 2014; Dosiou et al. 2012).

Genetic contributions to the development of Hashimoto's disease

The heritability of autoimmune hypothyroidism and thyroid hormones (TSH, T3, T4) has been assessed in several twin- and family-based studies with substantial variability (Brix et al. 2000; Panicker et al. 2008b; Samollow et al. 2004; Hansen et al. 2004; Meikle et al. 1988). Intra-individual variability has been shown to be roughly half that of inter-individual variability in these hormone levels for healthy individuals (Andersen et al. 2002). Nevertheless, these studies demonstrate a strong genetic component to thyroid hormone levels and corresponding disease state. Both candidate gene and genome-wide association studies have been used to identify genetic variants associated with hypothyroidism and corresponding thyroid hormone levels. Variants in *PDE8B* are associated with serum TSH levels (Arnaud-Lopez et al. 2008; Medici et al. 2011; Taylor et al. 2011); this gene encodes a cAMP-specific

protein expressed in thyroid tissue (Horvath et al. 2010). *FOXE1*, a thyroid transcription factor believed to be important in thyroid morphogenesis (Cuesta, Zaret, and Santisteban 2007), has been associated with hypothyroidism (Eriksson et al. 2012; Denny et al. 2011; Gudmundsson et al. 2009) and may be weakly associated with serum TSH levels (Medici et al. 2011). Despite the strong autoimmune component to hypothyroidism, few variants in known autoimmune loci (e.g., HLA region, *CTLA-4)* have been found to be associated with the clinical disease (Eriksson et al. 2012). In summary, the known genetic variants associated with TSH levels or hypothyroidism account for little of the expected heritability, emphasizing the need for future studies to identify additional genetic risk factors.

Personalization: Understanding Race/Ethnicity

Historic research inequality

As women have historically been overlooked in clinical research, so too, have been populations of non-European ancestry. The NIH Revitalization Act of 1993 was put into place to require biomedical researchers include more women and minorities in their clinical research studies unless compelling reasons against their inclusion existed (National Institutes of Health 1993). This and other initiatives to include women and minorities has resulted in some success: a recent report by the Federal Food and Drug Administration (FDA) determined that sex composition of clinical trials in 2011 was reflective of the disease prevalence differences between men and women (Food and Drug Administration and U.S.Department of Health and Human Services 2013). However, more than a decade later, minority participation in clinical research is still lacking (Food and Drug Administration and U.S.Department of Health and Human Services 2013; Ford et al. 2005). For example, the African American participation in T2D clinical studies was less than 5%, despite the higher prevalence of T2D in African Americans and their overall percent of the US population (13.2%) (Food and Drug Administration and U.S.Department of Health and Human Services 2013; U.S.Department of Commerce 2014; Office of Minority Health and U.S.Department of Health and Human Services 2014a). This FDA report concluded that, for 2011 clinical trials, non-white clinical trial study participants were underrepresented (Food and Drug Administration and U.S.Department of Health and Human Services 2013).

There are several barriers to participation in biomedical research by underrepresented groups from both the participant and the institutional perspectives. These include ineligibility per study design, cost, language differences, low literacy, and practical obstacles (e.g., number of visits required per study design) (Williams 2009). A lack of awareness and mistrust are challenges that result, in part,

from historic injustices to some groups (Armstrong et al. 1999; Gorelick et al. 1998). For example, the Tuskegee Syphilis Study, in which African American men with syphilis were untreated in order to study the natural progression of the disease, has had lingering effects on minority participation in biomedical research and clinical trials (Thomas and Quinn 1991; Freimuth et al. 2001; Gamble 1997; Tuskegee University Centers of Excellence Bioethics Center 2014). These challenges may be successfully overcome through use of social marketing, referrals from family or friends, and recruitment through healthcare providers (UyBico, Pavel, and Gross 2007). Use of EHRs to identify individuals who may qualify for participation in clinical research may improve minority participation.

Clinical differences across diverse populations

Population-specific differences have already been identified for many diseases. For example, Native Hawaiians and Pacific Islanders are 30% more likely to have asthma than non-Hispanic whites, and from 2003-2005, African American children were 7 times more likely to die from asthma than non-Hispanic white children (Asthma Disparities Working Group 2012). Type 2 diabetes disproportionately affects racial and ethnic minority groups; risk of diabetes compared to non-Hispanic whites is 18% higher in Asian Americans, 66% higher in Hispanics, and 77% higher in non-Hispanic blacks (Centers for Disease Control and Prevention and U.S.Department of Health and Human Services 2011). Complications from T2D, such as end stage renal disease and lower limb amputation are also more common in non-white individuals (Office of Minority Health and U.S.Department of Health and Human Services 2014b). Other population-specific differences have been identified for ECG traits (Ramirez et al. 2011), age-related macular degeneration (Klein et al. 2011), and heart disease (Office of Minority Health and U.S.Department of Health and Human Services 2014c). For reproductive traits, it is well-established that certain race/ethnicities differ in the median AM (Salsberry, Reagan, and Pajer 2009), though the cause of those differences is not completely understood and is likely to arise from both genetic and environmental factors. Similarly, endometrial cancer rates differ between race/ethnicities, and these differences cannot be attributed solely to socioeconomic issues that impair access to care (1987). The incidence rate for cervical cancer is more than five times higher for Vietnamese women in the US than for white women (Agency for Healthcare Research and Quality 2001). Given the public health burden of common, complex diseases, understanding the genetic, environmental, and socioeconomic factors that contribute to these disorders is imperative and necessitates studies in diverse populations.

Strategies for Building Statistical Models in Personalized Medicine

There are multiple approaches to designing a genetic study, each with a unique set of challenges and benefits. As the technology used to perform these studies changes, methods used to analyze the data must adapt to maximize the utility of the results. Many genetic studies follow the common disease/common variant hypothesis (CDCV), where common genetic variants confer the majority of disease susceptibility (Reich and Lander 2001). Because these risk alleles are common (typically defined as having a minor allele frequency >5%) and not subject to strong natural selections as are mutations, these risk alleles are likely ancient and shared across most populations. Each common variant is expected to confer a small effect towards disease risk; therefore, it is expected that common diseases will have multiple risk alleles and interactions with environmental factors. Genome-wide association studies, discussed below, were developed on the CDCV premise (Manolio et al. 2009). For most common disorders, current studies have failed to identify more than a small fraction of the genetic component using approaches based on CDCV, suggesting that rare variants with large effect sizes and/or complex interactions between genetic variants and environmental factors may play a substantial role in disease susceptibility (Manolio et al. 2009; Eichler et al. 2010; Cirulli and Goldstein 2010). For women's health, two main approaches have been used to identify the genetic variants associated with disease: candidate gene and genome-wide association studies.

Candidate gene approach

Based on prior knowledge or biological plausibility, the candidate gene association study interrogates specific variants, genes, or regions for association with the disease or quantitative trait. Candidate gene association studies (CGAS) were the first type of association study performed and are still widely used. Benefits to this approach include lower costs than other methods, the hypothesis-driven nature of the study, and limited number of tests performed. However, this design is not without drawbacks. If the correct variant/gene is not selected, no association with the phenotype will be found—a potential hazard due to genetic (both locus and allelic) heterogeneity. Candidate gene studies with small sample sizes have led to few of these studies replicating, though completely excluding a gene based on negative results is difficult. As in GWAS studies (see below), properly designed (appropriately powered) studies are essential when attempting to replicate a genotype-phenotype association.

Despite the challenges inherent with CGAS, they have been successful in identifying genes associated with numerous phenotypes important in women's health. Early CGAS evaluating the role of hormone biosynthesis pathways in the timing of the reproductive lifespan were largely unsuccessful; however, more recent studies have identified *FSHB* and *ESR2* associations with timing of AM (He et al. 2010). A CGAS was used to identify inflammatory pathway genes that were associated with endometrial cancer in the Shanghai Endometrial Cancer Genetics Study (Delahanty et al. 2013). CGAS may also be used to confirm results obtained from genome-wide association studies (GWAS) (see below); O'Mara et al. assessed five SNPs previously associated with endometrial cancer in GWAS, but failed to confirm these past associations(O'Mara et al. 2011b). In addition, CGAS may be used to prioritize studies hoping to generalize results from association studies in one population to another. With the majority of genetic studies performed in European-descent populations, extension of the findings to more diverse populations may suggest underlying biological disease mechanisms, while those that fail to associate may suggest population-specific disease risk or false positives. A recent CGAS assessed forty SNPs, previously identified in GWAS of European women, for association with breast cancer in Chinese women. rs9693444 was associated with overall breast cancer in this Chinese cohort ($p=6.44x10^{-04}$), while others were associated with various breast cancer subtypes (Zhang et al. 2014). Though there are significant challenges to the CGAS approach, it is a useful and relevant method to identify genotype-phenotype associations.

Genome-wide association studies

Unlike candidate gene studies where an *a priori* hypothesis about a relationship between the genetic variant and phenotype exists, genome-wide association studies (GWAS) require no previous knowledge about such a relationship, relying instead on the CDCV hypothesis. In a GWAS, interrogation occurs across the genome, generally capturing common variants in European-descent populations; the exact number of variants tested varies by platform and assay. It is common for hundreds of thousands or millions of SNPs to be tested. Linkage disequilibrium, the non-random association of alleles, allows a fraction of the genome to be genotyped while inferring information about untyped variants. Though the GWAS approach offers researchers the ability to discover new genotype-phenotype associations, it comes with a high statistical price: correcting for multiple statistical tests. This statistical burden is often corrected for using the Bonferroni method, where the alpha value for a single hypothesis test is divided by the total number of tests performed. For GWAS, the rule of thumb is that a result is significant if the p-value $<5x10^{-8}$ (0.05/1 million) (Dudbridge and

Gusnanto 2008). With newer GWAS chips capable of genotyping 5 million SNPs, this threshold may be inadequate; however, some have suggested the Bonferroni method is too stringent and other methods, such as false-discovery rates may be better (Zablocki et al. 2014; Pan 2013; Lin and Lee 2012; Wei 2012). This stringency arises from linkage disequilibrium; many of the SNPs tested in a GWAS are not independent, and the Bonferroni correction of these non-independent tests can result in a greater number of false negatives (missed true interactions) (De, Bush, and Moore 2014). In addition, GWAS chips primarily focus on common SNPs with allele frequencies greater than 0.05, limiting their ability to identify rare variants that are associated with a particular phenotype. The ability of genotyping chips to tag common variants also depends on the population under study (Hoffmann et al. 2011; Eberle et al. 2007). Furthermore, most SNPs associated with a disease phenotype have small effect sizes, explaining only a small amount of the phenotypic variance. These small effect sizes may fail to be clinically meaningful and frustrate replication attempts, as increasingly larger sample sizes are required to replicate the initial discovery. For example, a recent meta-analysis performed in the GIANT consortium with more than 250,000 cases and controls identified novel variants associated with overweight with an OR=1.04 (Berndt et al. 2013); successful replication of these results will require many more thousands of individuals derived from the same population, not already used in one of the contributing studies, demonstrating the practical challenges of replicating results with very small effect sizes.

Despite these limitations, GWAS have been successful in identifying genetic variants associated with many women's health traits and complex diseases and have led to additional hypotheses about the biological mechanisms responsible for disease. For example, a recent GWAS for systemic lupus erythematosus (SLE) identified novel HLA-region genes and replicated four genes previously associated with the autoimmune disorder (Armstrong et al. 2014). Numerous GWAS have been performed for breast cancer (briefly, (Low et al. 2013; Garcia-Closas et al. 2013; Michailidou et al. 2013)), endometriosis (Albertsen et al. 2013; Nyholt et al. 2012; Painter et al. 2011), and cervical cancer (Chen et al. 2013; Shi et al. 2013). Traits like gestational diabetes (Hayes et al. 2013) and fibroid tumors (Cha et al. 2011) have also been assessed with GWAS. These studies represent only a few women's health traits that have been investigated using GWAS. Though GWAS has not been successful in identifying causal variants with large effect sizes for most diseases/traits, the findings may point to underlying genetic architecture and biological mechanisms.

Interactions

Interactions, both gene-gene (GxG) and gene-environment (GxE), have been suggested as explanations for the "missing heritability" from GWAS studies (Zuk et al. 2012; Manolio et al. 2009). Interactions are challenging to identify in human genetic studies for a variety of reasons. Sample size requirements differ based on the study design (e.g., case-only vs. matched case-control), what type of interaction (GxG, GxE), and expected effect size of the interaction (Gauderman 2002a; Gauderman 2002b). Testing for statistical interactions among all the genetic variants is computationally intensive and leads to sparse/no data for some interactions; limiting GxG testing to variants with significant associations with the phenotype improves the computational challenges, including corrections for multiple tests, but compromises the ability to identify interactions between variants without main effects. Testing for interactions may be done using data reduction methods (e.g., combinatorial partitioning (Nelson et al. 2001), restricted partitioning (Culverhouse 2007), multifactor dimensionality reduction (Ritchie et al. 2001)), extensions to regression analysis (e.g., classification and regression trees (CART) (Breiman, Friedman, and Olshen), multivariate adaptive regression splines (MARS) (Lin et al. 2008)), and pattern recognition methods (e.g., neural networks (Turner, Dudek, and Ritchie 2010)). Prioritizing variants for GxG or GxE by biological plausibility reduces the number of statistical tests and computational burdens, yet restricts the potential to identify novel interactions that may be clinically meaningful.

Gene-gene interactions

Despite the issues addressed above, GxG interactions have been identified for a variety of phenotypes. Gene-level interactions between *SMAD3* and *NEDD9* affecting lipid levels was found in the Atherosclerosis Risk in Communities (ARIC) study and replicated in an independent sample from the Multi-Ethnic Study of Atherosclerosis (MESA) (Ma, Clark, and Keinan 2013). Samples with age-related macular degeneration (Klein et al. 2005) were used to find variants in several genes interacting with *CFH,* a well-characterized AMD gene (Zhang, Long, and Ott 2014). Notably, the AMD results are not only biologically plausible, but the interaction between *BBS9* and *CFH* replicates earlier studies performed using different methodology to detect the interactions (Chen et al. 2007; Wang et al. 2009).

Gene-environment interactions

Based on epidemiologic studies, environmental factors, which include such variables as body mass index (BMI), dietary intake, and carcinogen exposure, are known to play a role in the susceptibility of numerous disorders and complex traits (Cecchini et al. 2012; Turati et al. 2014; Steenland et al. 1996). How environmental exposures in conjunction with genetic variants contribute to the genetic architecture of complex diseases and traits is not fully understood. Examples of GxE interactions include exposure to farming with genetic variants on asthma risk (Ege et al. 2011; Ober and Vercelli 2011) and the effect of early childhood environment with genetic predisposition on mental health traits (Cicchetti and Rogosch 2012; Forsyth et al. 2013).

Despite some success, there is no systematic approach to identifying GxG or GxE and relatively few phenotypes have been adequately assessed for these potentially important interactions. The potential importance of GxG and GxE interactions should be considered in the context of personalized medicine. Individuals with significantly higher or lower disease risks based on GxG interactions may benefit from modified screening schedules; for example, in absence of a family history predisposing to colorectal cancer (CRC), someone with a GxG interaction that significantly increases their risk of developing CRC could benefit from more frequent colonoscopies. In addition, modifying exposures, such as alcohol intake, may reduce the risk of some diseases, like liver cancer. For individuals with increased genetic risk for a specific disease, understanding how environmental factors may interact with genetic factors may be a motivational tool to encourage healthy lifestyle choices.

Summary

Personalized medicine offers the potential to improve health outcomes by tailoring clinical care, including preventative medicine, to the individual patient. Though PM for cancer treatment and pharmacogenetics have been successfully integrated into clinical care at some institutions, expansion to other common, complex diseases has yet to be realized. Importantly, PM initiatives may reduce the historic and continued health and research disparities faced by women and some populations as researchers seek to understand the genetic architecture of complex traits and translate these findings to clinical care, or may lead to increased health disparities if access to and coverage of genetic testing is not universal or if decision support rules are invalid for non-European descent individuals. This body of work presents three genetic studies with an emphasis on complex traits important through the

course of a woman's life, from menarche through menopause and beyond, and then considers what work remains to effectively use this information to improve health outcomes in the future.

CHAPTER II

CASE STUDY: GENETICS OF THE FEMALE REPRODUCTIVE LIFESPAN[1]

Introduction

As noted in Chapter 1, age at menarche and age at natural menopause are heritable traits that influence a variety of phenotypes along the female life course. In this study, we used data from the Metabochip genotyping array to characterize previously identified variants associated with menarche and menopause in African Americans in a combined cohort of African-American women from the Women's Health Initiative (WHI) and Atherosclerosis Risk in Communities (ARIC) studies (Matise et al. 2011) as part of the Population Architecture using Genomics and Epidemiology (PAGE) Study (Buyske et al. 2012). The Metabochip array is based on the Illumina iSelect platform and contains approximately 200,000 single nucleotide polymorphisms (SNPs) consisting of GWAS index variants and fine-mapping common and less common variants for GWAS-identified regions relevant to metabolic and cardiovascular traits (Voight et al. 2012; Buyske et al. 2012). Using current GWAS and candidate gene literature as a guide, we attempted to generalize previously identified menarche and menopause SNPs and gene regions identified in European-descent populations to African Americans in the PAGE Study. We then sought to identify novel SNPs associated with AM and/or ANM.

Menarche

Study population

Women participants from two cohorts of the PAGE Study (Matise et al. 2011), Atherosclerosis Risk in Communities Study (ARIC) and the Women's Health Initiative (WHI), were included in these analyses. ARIC is a population-based prospective study of cardiovascular diseases and their causes in ~16,000 men and women aged 45-64 at baseline (1989). Participants were recruited in Forsyth County, N.C., Jackson, M.S., Minneapolis, M.N., and Washington County, M.D. From this group, 2,070 women,

[1] Adapted from: Spencer KL*, Malinowski J*, Carty CL, Franceschini N, *et al*. Genetic variation and reproductive timing: African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS One.* (2013). 8(2), e5528 PMID: 23424626 *these authors contributed equally to the work.

all of self-reported African American race/ethnicity and with information on reproductive timing, were selected for study. The WHI is a long term national health study investigating the leading causes of mortality and frailty in post-menopausal women in the United States, including heart disease, breast and colorectal cancer, and osteoporotic fractures (1998). A subset of 2,455 self-reported African American women selected based on consent to use DNA and availability of DNA, blood lipids, and glucose and insulin measurements were included in this study. The appropriate institutional review board at each participating study site approved all procedures, and written informed consent was obtained from all participants. Age at menarche was defined as the age when menstrual periods started in years, with extreme values pooled in groups of 9 years or less and 17 years or older.

Genotyping and statistical methods

Genotyping and quality control methods were the same for the age at menarche and the age at natural menopause analyses. Genotyping was performed on the Metabochip, a custom Illumina iSelect genotyping chip designed to genotype SNPs associated with metabolic traits and cardiovascular disease (Buyske et al. 2012; Voight et al. 2012). The array also includes 2,207 SNPs associated at genome-wide significance to any trait published in the NHGRI GWAS catalog as of August 1, 2009. For each of these GWAS-identified SNPs, an additional proxy SNP with $r^2>0.90$ in the CEU HapMap II dataset, plus up to four additional SNPs with $r^2>0.5$ in the YRI HapMapII dataset were also included on the array. Lastly, SNPs selected to fine-map regions of interest related to metabolic traits, copy number variant-tagging SNPs, Major Histocompatibility Complex (MHC) SNPs, SNPs on the X and Y chromosomes, mitochondrial DNA SNPs, and "wildcard" SNPs were also targeted, for a total of approximately 200,000 SNPs. Of these, 161,098 (81.9%) passed quality control filters for tests of Hardy-Weinberg Equilibrium ($>1x10^{-7}$) and genotyping efficiency (>95% call rate). There was no filter for minor allele frequency due to PAGE quality control protocol. The design and performance of this genotyping chip in this African American sample has been described in detail elsewhere (Buyske et al. 2012). All analyses were carried out in either METAL or the R software package, and data were plotted using LocusZoom (Pruim 2010; R Development Core Team 2012). Statistical power to detect an expected association was estimated in Quanto (Gauderman 2002a) assuming the observed sample size and coded allele frequency in this African American cohort and the genetic effect size previously reported in the literature.

All participants self-reported African American ancestry. To adjust for potential population stratification, we used the principal components method implemented in EIGENSTRAT (Price et al.

2006). We excluded any ancestry outliers further than eight standard deviations away from the mean for the first ten principal components determined by EIGENSOFT.

For menarche, linear regression was performed assuming an additive genetic model to test for associations between individual SNPs and the outcomes of age at menarche in years. We examined two models for menarche: 1) a minimally adjusted model that accounted only for study sites and principal components, and 2) a fully adjusted model that included study site, year of birth, principal components, and body mass index at ascertainment, with the understanding that BMI at ascertainment may be a poor proxy for BMI at age of menarche. Age at menarche was self-reported many years later at time of examination, which has been shown to be fairly accurate (Must et al. 2002).

We studied one model for natural menopause using Cox's proportional hazards for time-to-event (natural menopause) analysis, which adjusted for study site, principal components, and year of birth. Women with a missing age at menopause, an age at menopause <40 years, or hysterectomy, oophorectomy, or hormone replacement therapy after age 40 but prior to menopause, were excluded from the study. Women who had menopause >60 years had their ANM set as censored at age 60. A fixed effects meta-analysis was then performed using METAL to obtain effect size and standard error (SE) estimates (Willer, Li, and Abecasis 2010).

We looked to generalize to our population of African American women genes, gene regions (400 kb upstream and downstream of a gene of interest), and SNPs described in previous GWAS and candidate gene studies associated with AM. We tested all SNPs in the regions regardless of linkage disequilibrium (LD) with the index SNP, although we only considered a test of association generalized if the tested SNPs were identical to the index SNP or in strong LD with the index variant in HapMap CEU samples. For each candidate gene, we plotted results of single SNP tests of association using LocusZoom and examined regions 400kb upstream and downstream of the gene/gene region of interest. Tests of association were considered significant for generalization at a liberal threshold of $p<0.05$. For previously reported variants not genotyped in our study, we identified SNPs in LD with our directly genotyped SNPs (Johnson et al. 2008) and reported results from our minimally adjusted model (Model 1) for the proxy SNPs.

In addition to generalization, we sought to discover novel SNP-trait associations using the entire Metabochip. Significance in this discovery phase was defined as $p<3.1\times10^{-7}$, after Bonferroni correction (0.05/161,098). Because this threshold is highly conservative given the correlation among the

SNPs on the Metabochip, we also defined an arbitrary suggestive significance level as $p<1x10^{-4}$ in the discovery phase.

Results

A total of 4,159 African American female participants met the study definitions for AM and both PAGE studies were represented roughly equally (Table 5). In ARIC, the mean age at menarche was 12.9 years, which was slightly greater than the mean age at menarche in WHI (12.6 years) (Table 5). Groups' participants' heights, weights, and body mass indices were comparable. In both ARIC and WHI, the majority of participants' decade of birth was the 1930s (Table 5).

Table 5. Population characteristics of African American women from the PAGE Study for age at menarche (AM) analysis.

		Age at Menarche (AM) Study Population	
		ARIC	**WHI**
Participants (n)		2078	2081
Age at menarche, yrs		12.89 (1.76)	12.56 (1.64)
Age at enrollment, yrs		53.36 (5.73)	61.01 (6.87)
Body mass index, kg/m²		30.86 (6.63)	31.34 (6.83)
Weight, lbs.		181.05 (39.68)	182.87 (41.26)
Height, in.		64.24 (2.43)	64.00 (2.63)
Decade of birth, #(%)	**1910s**	-	26 (1.24)
	1920s	504 (24.07)	414 (19.82)
	1930s	1083 (51.72)	981 (46.96)
	1940s	507 (24.21)	668 (31.98)

Data presented as means (sd) unless otherwise noted. Abbreviations: Population Architecture using Genomics and Epidemiology (PAGE), Atherosclerosis Risk in Communities (ARIC), Women's Health Initiative (WHI), years (yrs), standard deviation (sd).

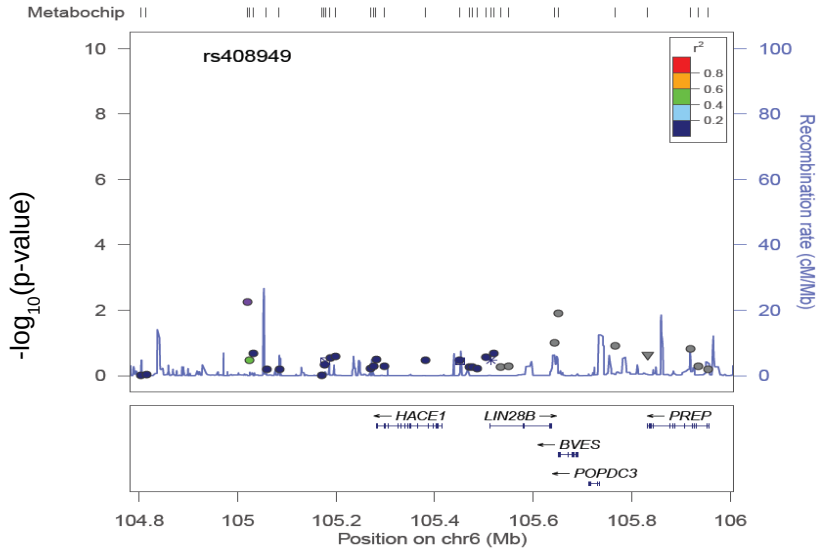


Figure 2. Regional association plot for AM in African American women from PAGE Studv.

Locus Zoom plot for *LIN28B* region in age at menarche (AM) analysis. Vertical axis is $-\log_{10}$ of the p-value, the horitzontal axis is the chromosomal position. Each dot represents a SNP tested for association with AM in 4,159 African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study. Approximate linkage disequilibrium between the most significant SNP, rs408949, listed at the top of the plot, and the other SNPs in the plot is shown by the r^2 legend.

Generalization to PAGE African Americans

To generalize previously-associated genetic variants in our African American population, we examined regions/genes previously associated with AM from either published candidate gene studies or GWAS: *CYP19A1, CYP17, CYP1B1*, *FTO, LIN28B*, 9q31.2 region, *IGF1, TNFSF11, TNFRSF11A*, and *LHCGR* (Guo et al. 2006b; Mitchell et al. 2008; Elks et al. 2010; Zhao et al. 2007; Perry et al. 2009; Ong et al. 2009; Lu et al. 2010; He et al. 2009a; He et al. 2010). We also evaluated forty-two SNPs associated with AM identified in a recent meta-analysis by Elks *et al.* of >87,000 European-descent women from forty-nine studies (Elks et al. 2010).

Overall, 11/21 (52%) SNPs previously identified for AM from earlier studies and 15/42 (36%) from the Elks *et al.* meta-analysis were directly genotyped or in strong (r^2>0.70) LD in the CEU panel of HapMap with those genotyped (Table 6 and Appendix A, respectively), and one generalized to this

African American cohort: rs9385399, in LD with previously reported rs1361108 (r^2=1.00, p=0.01) (Appendix A). A LocusZoom plot of the results of association tests and LD in this African American sample is given for *LIN28B* -- previously associated with AM (Figure 2) (He et al. 2009a; Ong et al. 2009; Sulem et al. 2009; Perry et al. 2009).

Three SNPs in *LIN28B* were included on the Metabochip (rs314277, rs4946651, and rs7759938), and while the direction of genetic effect was consistent with previous reports, all failed to reach statistical significance in this sample (p>0.30) (Table 6). Four additional SNPs in LD with these *LIN28B* SNPs were also not significant. At the 9q31 locus, rs7861820 and rs4452860, both located downstream of *TMEM38B,* had betas opposite to prior reports (Perry et al. 2009; He et al. 2009a). Neither SNP nor their proxy SNPs were significant at p<0.05. Similarly, SNPs in LD (rs1856142 and rs605765) with previously associated variants in and around *FSHB* were not significantly associated with AM in this African American sample, though rs605765 (β=-0.06) had the same direction of effect and comparable magnitude as rs1782507 (β=-0.07) (He et al. 2010).

Table 6. Comparison of GWAS-identified age at menarche (AM) variants in African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study.

Prior GWAS in European descent women							African American women from the PAGE Study						
SNP	CHR	Gene/Region	Coded Allele	Beta	P-value	Ref.	Best Proxy SNP from present study	r² in HapMap CEU/YRI	Coded Allele	Model 1 Beta (SE)	Model 1 P-value	Model 2 Beta (SE)	Model 2 P-value
rs314277	6	*LIN28B*	A	0.16	2.7E-13	(He et al.2009a)	rs314277	-	A	0.03(0.04)	0.34	0.03(0.04)	0.36
rs369065	6	*LIN28B*	C	0.11	2.4E-11	(He et al.2009a)	rs7759938	1.00/0.34	A	-0.02(0.04)	0.61	-0.02(0.04)	0.55
rs7759938	6	*LIN28B*	C	0.09	7.0E-09	(Perry et al.2009)	rs7759938	-	A	-0.02(0.04)	0.61	-0.02(0.04)	0.55
rs314276	6	*LIN28B*	C	-0.22	1.5E-08	(Ong et al.2009)	rs314274	1.00/0.73	A	0.05(0.04)	0.22	0.05(0.04)	0.24
rs314280	6	*LIN28B*	T	0.09	2.3E-08	(Sulem et al.2009;He et al.2009a)	rs7759938	0.64/0.28	A	-0.02(0.04)	0.61	-0.02(0.04)	0.55
rs4946651	6	*LIN28B*	A	0.09	3.1E-08	(He et al.2009a)	rs4946651	-	A	0.03(0.04)	0.55	0.03(0.04)	0.55
rs314262	6	*LIN28B*	C	0.08	9.7E-08	(He et al.2009a)	rs7759938	0.60/0.29	A	-0.02(0.04)	0.61	-0.02(0.04)	0.55
rs7861820	9	9q31	C	-0.09	3.4E-09	(He et al.2009a)	rs7861820	-	A	-0.10(0.06)	0.10	-0.09(0.06)	0.12
rs12684013	9	9q31	T	-0.10	3.6E-08	(He et al.2009a)	rs4452860	0.81/0.01	A	-0.03(0.04)	0.43	-0.03(0.04)	0.42
rs4452860	9	9q31	G	-0.09	7.9E-08	(He et al.2009a)	rs4452860	-	A	-0.03(0.04)	0.43	-0.03(0.04)	0.42
rs7028916	9	9q31	A	-0.09	9.7E-08	(He et al.2009a)	rs4452860	0.98/0.85	A	-0.03(0.04)	0.43	-0.03(0.04)	0.42
rs2090409	9	9q31	A	-0.10	1.7E-09	(Perry et al.2009)	rs4452860	0.83/0.82	A	-0.03(0.04)	0.43	-0.03(0.04)	0.42
rs555621	11	*FSHB*	C	0.06	0.001	(He et al.2010)	rs1856142	0.43/0.71	A	0.03(0.04)	0.44	0.03(0.04)	0.36
rs1782507	11	*FSHB*	T	-0.07	0.006	(He et al.2010)	rs605765	0.83/0.87	A	-0.06(0.04)	0.14	-0.06(0.04)	0.13
rs4953616	2	*LHCGR*	T	-0.07	0.006	(He et al.2010)	rs1589749	0.17/0.05	A	0.002(0.07)	0.97	-0.01(0.07)	0.87
rs7579411	2	*LHCGR*	T	0.06	0.01	(He et al.2010)	rs1589749	0.17/0.05	A	0.002(0.07)	0.97	-0.01(0.07)	0.87
rs4374421	2	*LHCGR*	C	0.06	0.02	(He et al.2010)	rs17326321	0.19/0.69	A	-0.01(0.06)	0.86	-0.01(0.06)	0.84
rs2470144	15	*CYP19A1*	G	-	5.9E-06	(Guo et al.2006b)	rs12148492	0.23/0.01	A	-0.01(0.07)	0.91	-0.02(0.07)	0.73
rs2445761	15	*CYP19A1*	G	-	1.2E-06	(Guo et al.2006b)	rs4774585	0.28/0.02	A	0.04(0.05)	0.47	0.03(0.05)	0.58
rs9525641	13	*TNFSF11/RANKL*	T	-	0.04	(Lu et al.2010)	rs931273	0.05/0.03	A	0.11(0.09)	0.24	0.11(0.09)	0.21
rs3826620	18	*TNFRSF11A/RANK*	A	-	0.02	(Lu et al.2010)	rs8092336	0.16/0.22	A	0.16(0.17)	0.33	0.17(0.17)	0.29
rs6214	12	*IGF1*	G	-	0.02	(Zhao et al.2007)	rs6214	-	A	-0.01(0.04)	0.71	-0.02(0.04)	0.61

Comparison of previously reported SNPs associated with AM in European descent women to 4,159 African American women from the PAGE Study in a minimally adjusted model for AM (Model 1) and a model adjusted for study site, year of birth, principal components, and body mass index (Model 2). Data are presented for the previously identified SNP. If the previously identified SNP was not directly genotyped in present study, data shown are for best proxy SNP based on linkage disequilibrium from the International HapMap Project CEU panel.

We also examined SNPs associated with AM that were reported in a recent meta-analysis performed by Elks *et al*. for the ReproGen Consortium (Elks et al. 2010) (Appendix A). Of the forty-two SNPs associated with AM in Elks et al., we detected an association with rs9385399 ($p=0.01$), located downstream of *CENPW,* which is a perfect proxy ($r^2=1.00$) for previously associated variant rs1361108, and the only SNP to generalize to our African American sample. We also identified an association with rs2947411 ($p=0.02$) with AM (Appendix A), though the directions of effect were opposite. One additional SNP, rs4929923 ($p=0.06$), nearly reached the significance threshold and had a similar magnitude and direction of effect compared with the previous report. Overall, AM SNPs from previously published studies of European-descent women, including the Elks et al. meta-analysis, did not generalize to our PAGE African American population.

Discovery

We tested all SNPs genotyped on the Metabochip for an association with AM adjusted for study site and principal components (Model 1) and adjusted for study site, year of birth, principal components, and body mass index (Model 2) (Appendix C). After accounting for multiple testing ($p<3.1\times10^{-7}$), no SNPs were significantly associated with AM in either model (Appendix C). The most significant SNP in both models was rs11604207 (Model 1: $p=1.59\times10^{-6}$; Model 2: $p=1.82\times10^{-6}$), which is located upstream of *RSF1*, a gene encoding a chromatin remodeling protein implicated in ovarian and breast cancers (Maeda et al. 2011; Choi et al. 2009; Brown et al. 2008) (Appendix C).

Menopause

Study population

Age at natural menopause was defined as the age at which cessation of regular menstrual periods due to the body's natural aging process occurred. In ARIC, women were asked, "Was your menopause natural or the result of surgery or radiation?" Only women who indicated natural menopause were included. Women in WHI who underwent hysterectomy, oophorectomy, or hormone replacement therapy before the onset of natural menopause were excluded. In both studies, women reporting age at natural menopause <40 years were excluded; women reporting age at natural menopause >60 years were censored at age 60. All women included in the present study were post-menopausal.

Genotyping and statistical methods

Genotyping, quality control, and statistical methods are detailed above for age at menarche.

Results

A total of 1,860 African American female participants met the study definitions for ANM and both PAGE studies were represented roughly equally (Table 7). In ARIC, the mean age at natural menopause was 48, which was slightly younger than the WHI group (Table 7). The body mass indices, heights, and weights of both groups were comparable. In ARIC and WHI, the majority of participants' decade of birth was the 1930s (Table 7).

Table 7. Population characteristics of African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study for age at natural menopause (ANM) analysis.

		Age at Natural Menopause (ANM)	
		Study Population	
		ARIC	**WHI**
Participants (n)		994	866
Age at menopause, yrs		47.97 (3.83)	50.84 (4.50)
Age at enrollment, yrs		53.07 (5.75)	61.30 (6.78)
Body mass index, kg/m²		31.29 (6.94)	30.95 (6.76)
Weight, lbs.		183.78 (40.80)	181.05 (40.63)
Height, in.		64.31 (2.38)	64.05 (2.75)
Decade of birth, #(%)	**1910s**	-	12 (1.39)
	1920s	221 (22.23)	183 (21.13)
	1930s	522 (52.52)	414 (47.81)
	1940s	251 (25.25)	257 (29.68)

Data presented as means (sd) unless otherwise noted. Abbreviations: Atherosclerosis Risk in Communities (ARIC), Women's Health Initiative (WHI), years (yrs), standard deviation (sd).

Generalization to PAGE African Americans

As with AM, to generalize results to our African American population, we examined previously identified 400kb regions around genes associated with ANM from published candidate gene studies and GWAS (Table 8) (He et al. 2010; He et al. 2009a; He et al. 2007; Murray et al. 2011; Stolk et al. 2009; Meng et al. 2011a; Long et al. 2006; Voorhuis et al. 2011; Lu et al. 2010). We also examined twenty SNPs associated with ANM that were identified in a recent study by Stolk et al. (Stolk et al. 2012) (Appendix B).

Overall, 14/23 (40%) SNPs previously identified for ANM via GWAS and 6/20 SNPs from the Stolk et al. meta-analysis were directly genotyped on the Metabochip or were in strong LD (r^2>0.70) in CEU panel of HapMap. One twelfth (8%) of the tested SNPs in these regions/genes generalized to this African American sample: rs8113016 (Table 8). rs8113016, located in an intron of *TMEM150B/TMEM224* and downstream of *BRSK1*, is in LD with previously reported rs897798 (r^2=0.72) and was associated with ANM in our sample (p=0.03). An intronic *APOE* variant, rs769450, was associated with ANM (p=0.03), though the nonsynonymous *APOE* rs7412 was not (p=0.55); these SNPs are not in LD with each other (r^2=0.04). In *BRSK1*, no previously reported SNPs were genotyped in our study; however, directly genotyped intronic *TMEM150B* rs4806660 was in very strong LD with intronic *BRSK1* rs1172822 (r^2=0.98). *BRSK1* rs1168309, in strong LD with rs2384687 (r^2=0.85) was not associated with ANM in this African American sample (p=0.59).

Three of the twenty SNPs recently identified by Stolk et al. as associated with ANM were directly genotyped on the Metabochip (Appendix B). Two of the three genotyped SNPs (rs2303369 and rs2153157) had the same directions of effect, though the magnitudes were smaller. Of the remaining 17 SNPs not directly targeted by the Metabochip, three were in strong LD (HapMap CEU r^2 ranging from 0.86 to 0.91) with the SNPs identified by Stolk et al: rs1176133, rs4668368, and rs12593363. For seven SNPs, no proxy SNP could be identified on the Metabochip (Appendix B). Of the twenty SNPs identified in the Stolk et al. meta-analysis and directly or indirectly represented on the Metabochip, none were associated with ANM in this African American sample (Appendix B).

Table 8. Comparison of GWAS-identified age at natural menopause (ANM) variants in African American women in the Population Architecture using Genomics and Environment (PAGE) Study.

		Prior GWAS in European descent women					African American women from the PAGE Study				
SNP	Chr	Gene/Region	Coded Allele	Beta	P-value	Ref.	Best Proxy SNP from present study	r^2 in HapMap CEU/YRI	Coded Allele	Beta (SE)	P-value
rs16991615	20	*MCM8*	A	1.07	1.21E-21	(He et al.2009a;Murray et al.2011)	rs16991615	-	A	-0.17(0.15)	0.25
rs236114	20	*MCM8*	A	0.50	9.71E-11	(Stolk et al.2009)	rs236114	-	A	0.02(0.06)	0.69
rs1172822	19	*BRSK1*	T	-0.49	1.8E-19	(Stolk et al.2009;He et al.2009a)	rs4806660	0.98/0.64	A	0.002(0.03)	0.97
rs2384687	19	*BRSK1*	C	-0.47	2.4E-18	(He et al.2009a)	rs11668309	0.85/0.43	A	0.02(0.04)	0.59
rs897798	19	*BRSK1*	G	-0.40	1.1E-14	(He et al.2009a)	rs8113016	0.72/0.02	A	0.12(0.05)	0.03
rs1065778	15	*CYP19A*	A	-	0.05	(He et al.2007)	rs10519297	0.90/0.32	A	-0.01(0.05)	0.84
rs2255192	15	*CYP19A*	A	-	0.04	(He et al.2007)	rs10459592	0.32/0.02	A	-0.02(0.04)	0.52
rs621686	11	*FSHB*	A	0.32	0.007	(He et al.2010)	rs1856142	0.27/0.32	A	0.04(0.03)	0.29
rs7951733	11	*FSHB*	A	-0.32	0.02	(He et al.2010)	rs7951733	-	A	0.11(0.13)	0.37
rs769450	19	*APOE*	A	-	0.007	(He et al.2009b)	rs769450	-	A	-0.07(0.03)	0.03
rs7412	19	*APOE*	-	-	0.001	(Meng et al.2011b)	rs7412	-	A	-0.03(0.05)	0.55
rs1019731	12	*IGF1*	C	-0.28	0.005	(He et al.2010)	rs1019731	-	A	-0.03(0.11)	0.82
rs9457827	17	*IGF2R*	T	0.37	0.04	(He et al.2010)	rs9457827	-	A	0.04(0.04)	0.28
rs4135280	3	*PPARG*	T	0.54	0.005	(He et al.2010)	rs4135280	-	A	-0.14(0.18)	0.42
rs1256044	14	*ESR2*	G	-	0.03	(He et al.2007)	rs1268656	0.08/0.004	A	-0.01(0.06)	0.88
rs1256059	14	*ESR2*	A	-	0.05	(He et al.2007)	rs1268656	0.08/0.004	A	-0.01(0.06)	0.88
rs1056836	2	*CYP1B1*	G	-	0.04	(Long et al.2006)	rs10495874	0.04/0.03	A	-0.03(0.05)	0.60
rs346578	13	*TNFSF11*	A	-	0.007	(Lu et al.2010)	rs6561072	0.07/0.07	A	0.04(0.04)	0.22
rs9525641	13	*TNFSF11*	T	-	0.01	(Lu et al.2010)	rs931273	0.05/0.03	A	-0.02(0.08)	0.81
rs8086340	18	*TNFRSF11A*	G	-	0.02	(Lu et al.2010)	rs8094440	0.10/0.01	A	0.03(0.03)	0.38
rs2002555	12	*AMHR2*	G	0.30	0.02	(Voorhuis et al.2011)	rs7131938	0.59/0.54	A	0.01(0.04)	0.84
rs2384687	19	*TMEM224*	C	0.38	1.39E-10	(Stolk et al.2009)	rs11668309	0.85/0.43	A	0.02(0.04)	0.59
rs897798	19	*TMEM224*	G	0.31	3.91E-08	(Stolk et al.2009)	rs8113016	0.72/0.02	A	0.12(0.05)	0.03

Comparison of previously reported SNPs associated with ANM in European and Chinese descent women to 1,860 African American women from the PAGE Study. Data presented are for the previously identified SNP. If the previously identified SNP was not directly genotyped in present study, data shown are for the best proxy SNP based on linkage disequilibrium calculated from the International HapMap Project CEU data.

Age at Natural Menopause: Discovery

We tested all SNPs on the Metabochip for associations with ANM adjusted for study site and principal components. Three SNPs were significant after Bonferroni correction ($p<3.1x10^{-7}$): *LDLR* (rs189596789, $p=4.98x10^{-8}$), *KCNQ1* (rs79972789, $p=1.90x10^{-7}$), and *COL4A3BP* (rs181686584, $p=2.85x10^{-7}$) (Table 9). The most significant association was with rs189596789, located approximately 10kb upstream of the low-density lipoprotein receptor (*LDLR*) gene, which has been associated with familial hypercholesterolemia (Diakou et al. 2011; De Castro-Oros et al. 2011).

Several of the most significant SNPs for ANM were located in/near genes previously associated with obesity, type 2 diabetes (T2D), coronary artery disease and lipid metabolism, e.g., *LDLR* (rs189596789), *NOS1AP* (rs76078015), *DGKB* (rs74486449), *LYPLAL1* (rs78696400), and *CDKAL1* (rs114158228) (Appendix D). We were unable to generalize the previously reported association between ANM and *PPARG* rs4135280 in this African American sample.

Table 9. Significant SNP associations for ANM in African Americans from the PAGE Study.

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA	SE	P VALUE
19	rs189596789	*LDLR*	upstream	A	0.006	1.09	0.20	$4.98x10^{-8}$
11	rs79972789	*KCNQ1*	intronic	C	0.997	-1.76	0.34	$1.90x10^{-7}$
5	rs181686584	*COL4A3BP*	intronic	A	0.002	2.35	0.46	$2.85x10^{-7}$

Tests of association at $p\leq3.1x10^{-7}$ (Bonferroni correction) from individual SNP linear regressions adjusted for study site and principal components in 1,860 African American women from the PAGE study are shown. For each significant test of association, the chromosome, rs number, nearest gene, location, coded allele, beta, standard error, and p-value are given. Genes listed are nearest genes to SNP as measured from the transcription start site for upstream SNPs or the transcription stop site for downstream SNPs. Abbreviations: chromosome (CHR), single nucleotide polymorphism (SNP), coded allele frequency (CAF), standard error (SE).

Two genes were suggestively associated with both ANM and AM at a nominal significance threshold. *PHACTR1* was suggestively associated with AM (rs73725617;Appendix C) and ANM (rs117124693;Appendix D). Though the direction of effects was similar for each SNP in *PHACTR1*, the SNPs are not in LD with each other. Likewise, SNPs in *ARHGAP42*, located at the 11q22.1 locus, were

suggestively associated with AM (rs11224447;Appendix C) and ANM (rs11224401; Appendix D), but are not in LD with each other, though the direction of effects was the same.

Summary

Here we demonstrated the use of the Metabochip genotyping array to identify SNPs associated with AM and ANM in a sample of African American women. Previous GWAS studies for AM and ANM have been performed in primarily European descent populations; generalization to diverse populations has largely been lacking (Dvornyk and Waqar 2012). Our study is the first, to our knowledge, to consider this trait in a large African American cohort. We were able to generalize only one previously identified variant for AM and two variants for ANM to our African American cohort [AM: rs1361108; ANM: rs897798 and rs9385399 (proxy for rs1361108)]. Overall, however, we were unable to generalize the majority of significant associations for previously identified SNPs associated with AM, including *LIN28B* or the 9q31 locus, or with ANM, including *MCM8* or *TMEM150b/TMEM224*, which have recently been identified in several GWAS of European-descent women. Our inability to replicate earlier findings in our African American sample may have, in part, resulted from scant Metabochip coverage of these regions. The emphasis of the Metabochip on genes involved in lipid metabolism and cardiovascular traits is evident comparing coverage in the *FTO* region (1053 SNPs) to the *LIN28B* region (28 SNPs).

In the discovery phase of our AM analysis, none of our results reached genome-wide significance. However, the ANM analysis yielded three associations that were significant after multiple testing corrections. Broadly, we demonstrate the ability to potentially uncover new variants associated with age at natural menopause in our African American cohort using the Metabochip.

Several studies have shown relationships between a woman's reproductive milestones (AM, ANM, parity) and menstrual characteristics and risk for breast cancer, endometrial cancer, and ovarian cancer (Milne et al. 2011; Opdahl et al. 2011; Narod 2011a; Narod 2011b; Jasen 2011) and chronic diseases such as diabetes, osteoporosis and cardiovascular disease (briefly (Dishi et al. 2011; Campbell Jenkins et al. 2011; Kallen and Pal 2011)). Interestingly, a recent study that assessed breast cancer susceptibility loci for associations with timing of menarche, menopause and the reproductive lifespan found only two SNPs associated with ANM: *CASP8* rs17468277 and 5p12 rs10941679; however, these results were not replicated in a validation cohort (Warren et al. 2014). The relationships between the timing of these reproductive events and their influences on complex diseases remain to be elucidated.

The most significant result in the ANM analysis was a SNP located upstream of *LDLR* (rs189596789) which encodes a low density lipoprotein receptor implicated in familial cholesterolemia. *KCNQ1* (rs79972789) also reached genome wide significance in our ANM analysis. Numerous variants in *KCNQ1* have also been implicated in type 2 diabetes in several populations, though none were in linkage disequilibrium with rs79972789 (Cui et al. 2011; Saif-Ali et al. 2011a; Saif-Ali et al. 2011b; Rees et al. 2011; Yasuda et al. 2008; Unoki et al. 2008). Recently, Buber *et al.* evaluated the role of menopausal hormonal changes with cardiac events in women with mutations in *KCNQ1* and congenital long-QT syndrome (LQTS) and determined the onset of menopause was associated with an increase in the risk of cardiac events in LQTS women (Buber et al. 2011). Though not significant, suggestive AM associations included *LPL* and *CYP4F22*, which are associated with type 2 diabetes and lipid metabolism (rs1372339, rs4922116, rs1273516), and *TMEM18* (rs2947411), associated with obesity and body mass index (Jurvansuu and Goldman 2011; Speliotes et al. 2010). These ANM associations and suggestive AM associations with genes involved in cardiovascular function, lipid metabolism, and type 2 diabetes concur with research showing later AM lowers obesity and diabetes risk while earlier ANM increases risk for cardiovascular disease, obesity and insulin resistance (Carr 2003; Salpeter et al. 2006).

Different pathways appear to be involved in the initiation and cessation of menses. Other GWAS and linkage studies performed in European descent or Asian populations for AM and ANM show little concordance with specific genes (reviewed in (Hartge 2009) (Perry et al. 2014). Our analysis is consistent with this observation. Only *PHACTR1* and *ARHGAP42* SNPs were suggestively significant in both our AM and ANM analyses. *PHACTR1* is a phosphatase and actin regulator which has been implicated in coronary artery disease (Schunkert et al. 2011; Ripatti et al. 2010). Its role in menarche and menopause is yet to be determined. *ARHGAP42*, a Rho GTPase activating protein, has not yet been evaluated for a role in menarche or menopause. A GWAS identified intronic *ARHGAP42* rs633185 is associated with blood pressure (Ehret et al. 2011) , but this variant is not in strong LD with *ARHGAP42* variants suggestively associated with either AM or ANM in this study. A recent study by Lu et al., found SNPs in both *TNFSF11* and *TNFRSF11A* significant for AM and ANM (Lu et al. 2010). SNPs genotyped on the Metabochip were in weak LD with the reported SNPs and failed to reach significance in this African American sample. Given the role that both *PHACTR1* and *ARHGAP42* play in atherosclerosis, osteoporosis and the development of lactation glands in pregnancy, further investigation on the influence of these genes in AM and ANM is warranted (Hofbauer and Schoppet 2004; Boyce and Xing 2008).

The Metabochip was designed to be a cost-effective method of genotyping approximately 200,000 metabolic and cardiovascular SNPs and SNPs in other useful regions of the genome, such as the HLA region and the X and Y chromosomes. Overall, median SNP density on the Metabochip is approximately one SNP per 370 bases (Buyske et al. 2012). This coverage appears sufficient to replicate some loci associated with both cardiovascular or metabolic traits and AM/ANM. However, we found instances of previously identified genes for AM/ANM with little/no Metabochip coverage (*CYP1B1, LIN28B, ESR2, and BRSK1*) which may have impacted our results. Additionally, prior studies that identified SNPs associated with AM and ANM were performed primarily in European-descent cohorts. Though our study included over 4,000 African American women, we had limited power to identify significant associations in most previously identified loci, which may explain why we failed to detect the same associations identified in European-descent GWAS. For specific tests of association, our power was impacted by sample size and by minor allele frequencies. For example, the allele frequency for rs7861820 in this African American cohort was 0.11 compared to a higher frequency observed in HapMap CEU (0.57; Table 10). Interestingly, we were adequately powered (>98%) (Appendix E) to generalize the intronic *LIN28B* SNP, rs314277, with AM in our sample, yet failed to find an association with this SNP or with SNPs in strong LD with it (Table 6, Appendix A).

Table 10. Minor allele frequency comparisons of African American women in the Population Architecture using Genomics and Epidemiology (PAGE) Study to HapMap CEU Panel.

SNP	Gene/Region	HapMap CEU AF (Allele)	PAGE Study AF (Allele)
rs314277	*LIN28B*	0.13 (A)	0.39 (A)
rs7759938	*LIN28B*	0.64 (T)	0.46 (A)
rs4946651	*LIN28B*	0.48 (A)	0.75 (A)
rs7861820	9q31	0.57 (T)	0.11 (A)
rs4452860	9q31	0.72 (A)	0.67 (A)
rs16991615	*MCM8*	0.09 (A)	0.01 (A)
rs236114	*MCM8*	0.21 (A)	0.09 (A)
rs7951733	*FSHB*	0.95 (A)	0.99 (A)
rs769450	*APOE*	0.38 (A)	0.38 (A)
rs7412	*APOE*	0.09 (T)	0.10 (A)
rs1019731	*IGF1*	0.14 (T)	0.02 (A)
rs9457827	*IGF2R*	0.05 (T)	0.28 (A)
rs4135280	*PPARG*	0.98 (T)	0.99 (A)

Comparison of allele frequencies between PAGE Study African American women and HapMap CEU Panel. SNPs compared were previously associated with age at menarche or age at natural menopause and directly genotyped on the Metabochip. Abbreviation: single nucleotide polymorphism (SNP), allele frequency (AF).

Metabochip performance in non-European populations was recently evaluated in a pilot study in African American PAGE participants (Buyske et al. 2012). In this pilot study, Buyske *et al*. demonstrated that the majority (89%) of SNPs targeted by the Metabochip passed rigorous quality control with high call rates (Buyske et al. 2012). Using lipid traits as an example, Buyske et al. demonstrated that Metabochip data can be used to replicate known GWAS-identified SNP-trait relationships (Buyske et al. 2012). Furthermore, the pilot study demonstrated that Metabochip data can be used to fine-map GWAS-identified regions to uncover potential novel index SNPs specific to African Americans in an established locus for that trait. Fine-mapping data for AM/ANM was not included in the Metabochip content. While we were able to use the Metabochip to identify potentially novel SNP-trait relationships for AM/ANM, additional fine-mapping efforts of other loci already implicated for these traits are needed. Furthermore, additional studies in general are warranted for diverse (non-European descent) populations using Metabochip or other arrays designed for fine-mapping. Admixture in the African American population and its associated decreased LD compared to European Americans challenge identification of trait-associated SNPs. Targeted fine mapping, such as use of the

Metabochip, may be more appropriate in some circumstances than GWAS to evaluate specific SNPs and regions associated with particular traits.

A GWAS meta-analysis for age at menarche in African Americans was published shortly after publication of the PAGE study (Demerath et al. 2013). In a much larger cohort (n=18,089), Demerath et al. found only suggestive associations ($p>5x10^{-8}$) between AM and *FLRT2* and *PIK3R1*; conditional analyses identified *RORA* rs339978 and rs980000 as additional variants, independently associated with AM (Demerath et al. 2013). Additionally, they generalized to their cohort 25/42 (60%) variants previously associated in the Elks meta-analysis (Elks et al. 2010). This is higher than the 36% that we were able to generalize to the PAGE African American cohort (Appendix A), though the substantial increase in sample size for the Demerath et al. meta-analysis likely played a role.

Although the Metabochip was designed for genotyping of cardiovascular and metabolic SNPs, this study demonstrates the feasibility of utilizing such a targeted chip to identify SNP associations with age at menarche and age at natural menopause. We identified potentially novel associations with AM/ANM at loci implicated in cardiovascular traits, obesity and cancer. This may result from pleiotropic loci or may suggest that the AM/ANM timing mechanisms influence underlying disease process. With numerous genes implicated in both metabolic and cardiovascular phenotypes and both AM and ANM, further studies will allow us to consider how specific genes may influence the reproductive lifespan in women.

CHAPTER III

ALGORITHMIC EXTRACTION OF FEMALE REPRODUCTIVE MILESTONES FROM ELECTRONIC MEDICAL RECORDS[2]

Introduction

As described in Chapter 1, the rich phenotypic data existing in EMR systems allows clinicians and researchers to identify potential cohorts, while EMRs that are linked to biobanks extend this framework to genotype-phenotype association studies. Traditional epidemiologic studies are costly and require significant amounts of time to complete; furthermore, these studies may not include sufficient numbers of individuals from diverse ancestries. The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) Study seeks to address these limitations by enabling high-throughput identification and generalization of genotype-phenotype associations in diverse research populations. Accessing data from EMRs for use in research may prove to be a cost effective alternative to traditional ascertainment and data collection. One challenge to research use of EMR-derived data is the lack of consistency in recording certain types of data in the EMR. Despite the obvious health implications, AM and AAM/ANM are not recorded consistently or in a standardized manner in the EMR. This presents a challenge for researchers and suggests algorithm development is a necessary first step in developing a resource for women's health studies in diverse populations.

BioVU

BioVU is the Vanderbilt University Medical Center (VUMC) biorepository linked to the EMR system. Beginning in 2007, discarded blood samples from routine clinical testing have the DNA extracted, stored, and linked to a de-identified version of the EMR termed the Synthetic Derivative

[2] Adapted from: Malinowski J, Farber-Eger E, Crawford DC. Development of a data-mining algorithm to identify ages at reproductive milestones in electronic medical records. *Pacific Symposium in Biocomputing*. (2014) 19:376-87 PMID: 24297563

(SD). As of mid-2012, more than 150,000 samples have been collected for BioVU, including more than 16,000 pediatric samples. Patients are given the opportunity to opt-out of BioVU at any time. Once a sample has been accepted into the system, a unique ID is generated through a one-way hash mechanism and linked to that patient's SD. The SD removes or de-identifies Health Insurance Portability and Accountability Act (HIPAA) information, such as names, geographical locations, and social security numbers, and replaces dates with dates that have been randomly shifted by up to six months. The date shifting is consistent within a single SD record. The SD enables researchers to examine genome-phenome associations and identify cohorts for research.

Methods

Study population

As part of the Population Architecture using Genomics and Epidemiology (PAGE) I Study, EAGLE genotyped all non-European descent patients in BioVU (EAGLE BioVU, n=15,863) on the Metabochip, a custom genotyping array with an emphasis on cardiovascular disease and metabolic traits (see Chapter II, Methods and (Voight et al. 2012) for full details on the design). Overall, 11,521 African Americans, 1,714 Hispanics, 1,122 Asians and others were genotyped on the Metabochip in EAGLE. For the AM study, all females age>6 in EAGLE BioVU as of January 31, 2013 were eligible for inclusion. For the AAM study, all females >18 years were eligible for inclusion; for the ANM study, only women ages≥41 were eligible for inclusion. All patients were of diverse race/ethnicity.

Algorithm development

We developed a flow chart to visualize the inclusion/exclusion processes for the algorithms (Figure 3A-C). AM and age at menopause or age at natural menopause (AAM/ANM) are not consistently recorded in the EMR system at VUMC; individuals may enter BioVU through numerous outpatient clinics with different data field requirements. The lack of a pre-specified field for AM and AAM/ANM in the EMR necessitated a combination of free text data mining using regular expressions/pattern matching, billing (ICD-9) codes, and procedure (CPT) codes to identify AM and AAM/ANM in the subsequently generated SD. All analysis for this study was performed using the SD.

Age at menarche (AM)

Primary exclusion criteria for AM phenotype consisted of four components: age<7 years, male sex, ICD-9 codes for delayed puberty/sexual development (259.0) and precocious puberty/sexual

development (259.1), and keywords (Figure 3A). Inclusion of any of the preceding criteria in the SD resulted in exclusion for the AM study. As part of the de-identification data scrubbing to convert a patient's EMR to the SD, ages and dates may be masked and listed as "birth-12" or "in teens." Dates and ages which are not masked were date shifted by up to six months forward or backward from the actual date.

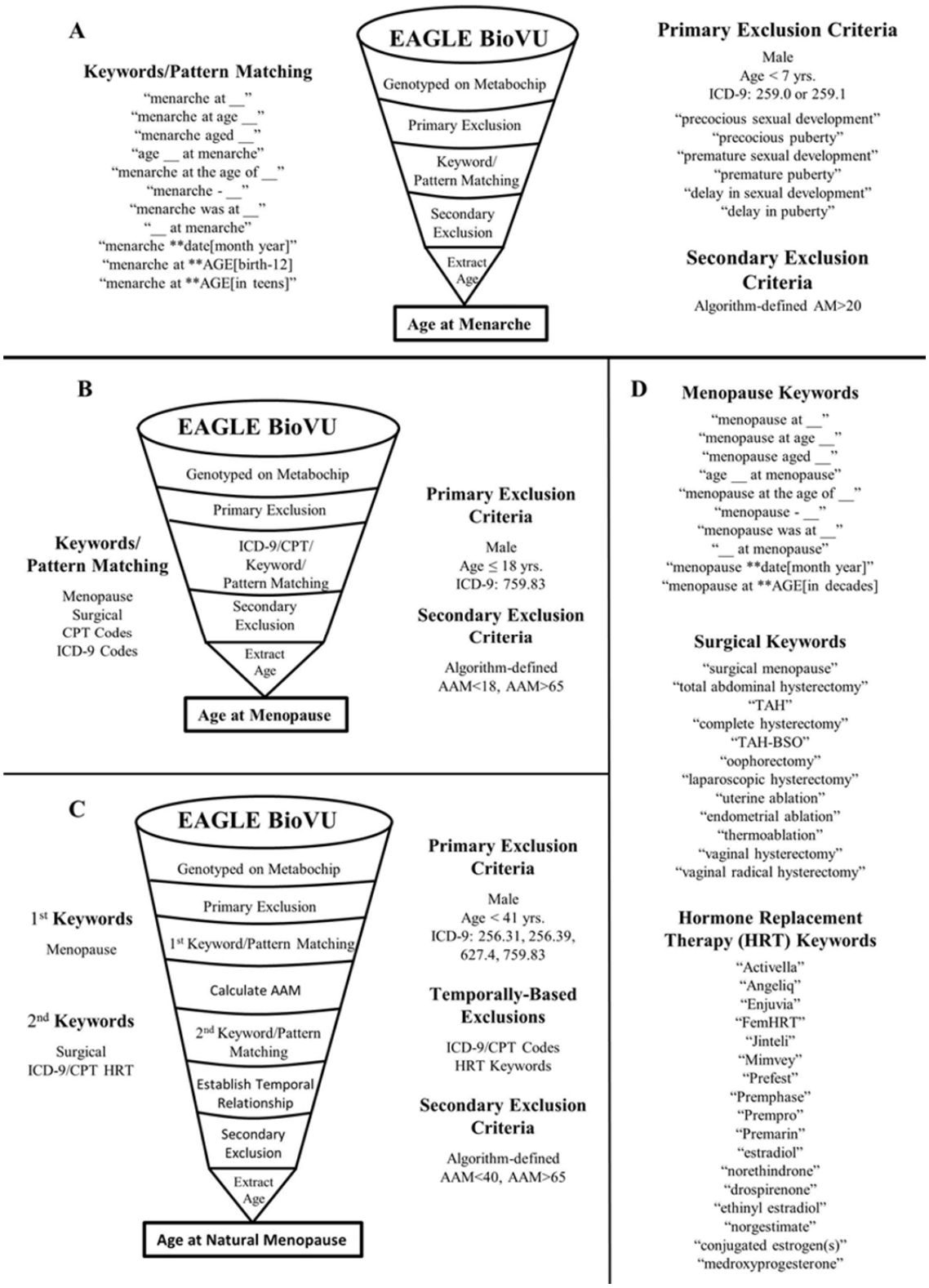


Figure 3. Flow chart for algorithm development for reproductive milestones.

To identify a listed AM for an individual, we utilized pattern matching to seek instances with menarche keyword phrases (Figure 3A). Numbers and dates were allowed to be included as numerals only. Instances where the AM was listed as a date used the subject's birthdate to calculate the age (in years) at menarche. In cases of ties, where more than one AM was identified and recorded an equal number of times in the SD, the AM was determined to be the one listed first in the SD. If the algorithm identified multiple versions of the AM (an exact age, an age calculated from a date, or a de-identified age), a hierarchy was used to determine the AM for the output, where an exact age or date was prioritized over de-identified age ranges. Instances where multiple different ages were listed in the SD as AM defaulted to the age listed most frequently. We considered situations where the algorithm identified an exact AAM and a de-identified AAM range containing the exact AAM to be the same for purpose of calculating sensitivity, specificity, and positive predictive value (PPV), but different for the purpose of calculating accuracy. The resulting output file contained the subject's unique research id (RUID), date of birth, and either an algorithm-generated AM or null value.

Age at menopause (AAM)

For an algorithm to identify all post-menopausal women and their age at menopause (AAM), we initially excluded all males, set a minimum age of 18 years, and excluded patients with a Fragile X diagnosis (ICD-9 759.83) (Figure 3B). Pattern matching was utilized to find keyword phrases similar to those used in the menarche algorithm, substituting "menopause" for "menarche" (Figure 3D). Furthermore, we included keywords pertaining to surgical procedures that induce cessation of menses/menopause (Figure 3D). We excluded instances where the word "possible" immediately preceded a keyword. For instances where the SD had scrubbed the exact age, decade-specific results (e.g. "in 30s", "in 50s") were captured by our algorithm. CPT (Table 11) and ICD-9 (Table 12) codes were used to identify women with surgical menopause or menses-ceasing procedures.

Table 11. CPT codes for age at menopause (AAM) and age at natural menopause (ANM) algorithms.

CPT code	Procedure
58150	Total abdominal hysterectomy (corpus and cervix), with or without removal of tube(s), with or without removal of ovary(s)
58152	Total abdominal hysterectomy (corpus and cervix), with or without removal of tube(s), with or without removal of ovary(s); with colpo-urethrocystopexy (eg, Marshall-Marchetti-Krantz, Burch)
58180	Supracervical abdominal hysterectomy (subtotal hysterectomy), with or without removal of tube(s), with or without removal of ovary(s)
58200	Total abdominal hysterectomy, including partial vaginectomy, with para-aortic and pelvic lymph node sampling, with or without removal of tube(s), with or without removal of ovary(s)
58260	Vaginal hysterectomy, for uterus 250g or less
58262	Vaginal hysterectomy, for uterus 250g or less, with removal of tube(s), and/or ovary(s)
58263	Vaginal hysterectomy, for uterus 250g or less, with removal of tube(s), and/or ovary(s), with repair of enterocele
58267	Vaginal hysterectomy, for uterus 250g or less, with colpo-urethrocystopexy (eg, Marshall-Marchetti-Krantz type, Pereyra type) with or without endoscopic control
58270	Vaginal hysterectomy, for uterus 250g or less, with repair of enterocele
58275	Vaginal hysterectomy, with total or partial vaginectomy
58280	Vaginal hysterectomy, with total or partial vaginectomy, with repair of enterocele
58285	Vaginal hysterectomy, radical (schauta type operation)
58290	Vaginal hysterectomy, for uterus greater than 250g
58291	Vaginal hysterectomy, for uterus greater than 250g, with removal of tube(s), and/or ovary(s)
58292	Vaginal hysterectomy, for uterus greater than 250g, with removal of tube(s), and/or ovary(s), with repair of enterocele
58293	Vaginal hysterectomy, for uterus greater than 250g, with colpo-urethrocystopexy (eg, Marshall-Marchetti-Krantz type, Pereyra type) with or without endoscopic control
58294	Vaginal hysterectomy, for uterus greater than 250g, with repair of enterocele
58353	Endometrial ablation, thermal, without hysteroscopic guidance
58541	Laparoscopy, surgical, supracervical hysterectomy, for uterus 250 g or less
58542	Laparoscopy, surgical, supracervical hysterectomy, for uterus 250 g or less, with removal of tube(s), and/or ovary(s)
58543	Laparoscopy, surgical, supracervical hysterectomy, for uterus greater than 250 g
58544	Laparoscopy, surgical, supracervical hysterectomy, for uterus greater than 250 g, with removal of tube(s), and/or ovary(s)
58548	Laparoscopy, surgical, with radical hysterectomy, with bilateral total pelvic lymphadenectomy and para-aortic lymph node sampling (biopsy), with removal of tube(s) and ovary(s), if performed
58550	Laparoscopy, surgical, with vaginal hysterectomy, for uterus 250 g or less
58552	Laparoscopy, surgical, with vaginal hysterectomy, for uterus 250 g or less, with removal of tube(s) and ovary(s)
58553	Laparoscopy, surgical, with vaginal hysterectomy, for uterus greater than 250 g
58554	Laparoscopy, surgical, with vaginal hysterectomy, for uterus greater than 250 g, with removal of tube(s) and ovary(s)
58563	Hysteroscopy, surgical, with endometrial ablation (eg endometrial resection, electrosurgical ablation, thermoablation)
58570	Laparoscopy, surgical, with total hysterectomy, for uterus 250 g or less
58571	Laparoscopy, surgical, with total hysterectomy, for uterus 250 g or less, with removal of tube(s) and ovary(s)
58572	Laparoscopy, surgical, with total hysterectomy, for uterus greater than 250 g
58573	Laparoscopy, surgical, with total hysterectomy, for uterus greater than 250 g, with removal of tube(s) and ovary(s)

CPT codes used to identify menopausal women for the age at menopause (AAM) algorithm and to identify women to exclude in a time-dependent manner for the age at natural menopause (ANM) algorithm in EAGLE BioVU. Abbreviations: Current procedural terminology (CPT).

Table 12. ICD-9 codes for age at menopause (AAM) and age at natural menopause (ANM) algorithms.

65.5	Bilateral oophorectomy
65.51	Other removal of both ovaries at same operative episode
65.52	Other removal of remaining ovary
65.53	Laparoscopic removal of both ovaries at same operative episode
65.54	Laparoscopic removal of remaining ovary
65.6	Bilateral Salpingo-oophorectomy
65.61	Other removal of both ovaries and tubes at same operative episode
65.62	Other removal of remaining ovary and tube
65.63	Laparoscopic removal of both ovaries and tubes at the same operative episode
65.64	Laparoscopic removal of remaining ovary and tube
68.23	Endometrial ablation
68.3	Subtotal abdominal hysterectomy
68.31	Laparoscopic supracervical hysterectomy
68.39	Other and unspecified subtotal abdominal hysterectomy
68.4	Total abdominal hysterectomy
68.41	Laparoscopic total abdominal hysterectomy
68.49	Other and unspecified total abdominal hysterectomy
68.5	Vaginal hysterectomy
68.51	Laparoscopically assisted vaginal hysterectomy (LAVH)
68.59	Other and unspecified vaginal hysterectomy
68.6	Radical abdominal hysterectomy
68.61	Laparoscopic radical abdominal hysterectomy
68.69	Other and unspecified radical abdominal hysterectomy
68.7	Radical vaginal hysterectomy
68.71	Laparoscopic radical vaginal hysterectomy (LRVH)
68.79	Other and unspecified radical vaginal hysterectomy
68.9	Other and unspecified hysterectomy

ICD-9 codes used to identify menopausal women for the age at menopause (AAM) algorithm and to identify women to exclude in a time-dependent manner for the age at natural menopause (ANM) algorithm in EAGLE BioVU.
Abbreviations: International Classification of Diseases, Ninth Revision (ICD-9).

After SD review of initial algorithms and subject matter knowledge, we implemented secondary exclusion criteria based on the algorithm-identified AAM and excluded subjects with a calculated AAM<18 or AAM>65 (Figure 3B). A hierarchy was used to determine the AAM for the output, with an exact age or date identified by keyword or pattern matching and ICD-9/CPT codes prioritized over de-identified age ranges. In rare instances where the algorithm identified more than one AAM for a subject, the age recorded most frequently was determined to be the AAM for that patient. In cases of ties, where more than one AAM was identified and recorded an equal number of times in the SD, the AAM was determined to be the one listed first in the SD. We considered situations where the algorithm identified an exact AAM and a de-identified AAM range containing the exact AAM to be the same for purpose of calculating sensitivity, specificity, and PPV, but different for the purpose of calculating accuracy. The resulting output file contained the subject's unique research id (RUID), date of birth,

race/ethnicity, either an algorithm-generated AAM or null value, the method by which the AAM was calculated (e.g., from ICD-9 code, keyword), and the date in the SD corresponding to the AAM identification.

Age at natural menopause (ANM)

To discriminate age at natural menopause (ANM) from all instances of menopause (AAM), we extended the AAM algorithm to exclude women aged <41 years, men, and subjects with ICD-9 codes signifying premature ovarian failure/premature menopause (256.31), artificially induced menopause (627.4), ovarian failure (256.39), and Fragile X syndrome (759.83) (Figure 3C). We used pattern matching with the menopause keywords to identify an age at menopause (Figure 3D). We did not use ICD-9 codes, CPT codes, or keywords associated with procedures that induce menopause to identify subjects for the ANM cohort.

Medication delivery and prescriptions are captured by the EMR at VUMC and are included in the SD. To ascertain the temporal relationship between AAM and menopause-inducing/menses-ceasing surgery or hormone replacement therapy (HRT) use, we first calculated the AAM with the alternate algorithm (Figure 3C). Surgery-inducing menopause, determined through CPT and/or ICD-9 codes or keywords, and HRT were not exclusion criteria unless the first instance of surgery or HRT occurred prior to the extended algorithm-identified AAM. Keyword pattern matching was performed using surgical keywords (Figure 3D). We used a combination of brand-name and generic names for HRT identification (Figure 3D). If AAM was identified and no keywords or CPT/ICD-9 codes were found to indicate artificially induced menopause, the subject was deemed to have undergone natural menopause. If surgery or HRT occurred after the algorithm-determined ANM, the subject was also considered to have undergone natural menopause. If the subject had either surgery or used HRT prior to menopause, they were excluded from the cohort and the resulting output was a null value.

We implemented secondary exclusion criteria (Figure 3C) based on the algorithm-identified age at menopause and excluded subjects with a calculated ANM<18 or ANM>65 based on subject matter knowledge and review of early versions of our algorithms. A hierarchy was used to determine the ANM for the output. If the algorithm determined more than one ANM for a subject, we used the same procedure as described above to determine the final ANM generated by our query. We again considered situations where the algorithm identified an exact ANM and a de-identified ANM range containing the exact ANM to be the same for purpose of calculating sensitivity, specificity, and PPV,

but different for the purpose of calculating accuracy. The resulting output file contained the subject's unique research id (RUID), date of birth, race/ethnicity, either an algorithm-generated ANM or null value, the method by which the ANM was calculated (e.g., from exact date, de-identified age), and the date in the SD corresponding to the ANM identification.

Manual review

To determine the sensitivity, specificity, PPV, and accuracy of the AM, AAM, and ANM algorithms, extensive manual chart review was performed by a single individual for consistency. Each algorithm output contained three types of values: exact ages, de-identified ages, and null values. For each algorithm, a random number generator was used to randomize RUIDs within each of the three types of output and the subjects were then sorted in ascending value by the random number. The first 50 subjects in the exact age and de-identified age categories and the first 100 subjects with a null value had their SD reviewed manually to determine the AM, AAM, or ANM. Sensitivity, specificity, PPV and accuracy were calculated by comparing the automated algorithm result to the manual review result for each subject.

Results

Population characteristics

A total of 10,051 females were genotyped on the Metabochip in BioVU in EAGLE for various studies. We identified an age for menarche (exact or de-identified) in 1,618 individuals. For the AAM algorithm, we identified an AAM (exact age or de-identified decade) for 1281 individuals. We identified 83 individuals with an ANM (exact or de-identified decade) (Table 13). The algorithm-extracted mean AM in our population was 12.7 (+/- 2.1) yrs. The mean AAM in our population was 44.6 (+/- 9.8) yrs. and the mean ANM was 49.7 (+/- 5.6) yrs. (Table 13). Approximately half of the algorithm extracted AM (54.7%) and ANM (47.0%) were exact ages, while the majority of AAM (92.5%) were exact ages (Table 13).

Table 13. Population characteristics for women with algorithm-identified age at menarche (AM), age at menopause (AAM), and age at natural menopause (ANM) from EAGLE BioVU.

	AM	AAM	ANM
N, total	1618	1281	83
exact age (n)	885	1185	39
de-identified age (n)	733	96	44
Age at event, mean +/- sd (yrs)	12.7 (2.1)	44.6 (9.8)	49.7 (5.6)
Age range at event (yrs)	8-20	18-65	40-65
Race/ethnicity (n)			
African American	1232	1112	62
Hispanic	120	45	4
Asian	115	66	11
Other	151	58	6

Abbreviations: standard deviation (sd), years (yrs).

AM algorithm performance

We manually reviewed 200 SD entries for the AM algorithm to determine sensitivity, specificity, PPV, and accuracy. Of the 100 subjects with an algorithm-specified AM, 94 were confirmed by manual review. For the 100 subjects without an AM captured by the algorithm, 99 were not found to have an identifiable AM upon manual review. The AM algorithm had a sensitivity and specificity of 99.0% and 94.3%, respectively, and a PPV of 94.0% (Table 14). We calculated the accuracy of the algorithm by comparing the results for the 94 subjects with both manually identified and algorithm identified AMs, requiring identical results for concordance. Of these 94 subjects, we found 87 where the AM matched in both manual and algorithm identification for an accuracy of 92.6% (Table 14). We observed instances where the algorithm calculated an exact AM (e.g., 8) and manual review found a de-identified AM (e.g., birth-12), or vice-versa. If we allow these to be concordant, accuracy increases to 94.7%.

Table 14. Performance of the age at menarche (AM), age at menopause (AAM), and age at natural menopause (ANM) algorithms in women from EAGLE BioVU.

	Sensitivity	Specificity	Accuracy	PPV
AM (n=200)	99.0%	94.3%	92.6%	94.0%
AAM (n=200)	94.4%	85.6%	52.4%	84.0%
ANM (n=183)	89.8%	75.8%	75.5%	63.9%

Abbreviations: positive predictive value (PPV).

AAM algorithm performance

For the AAM algorithm, we manually reviewed 200 SD entries to determine sensitivity, specificity, PPV, and accuracy. Of the 100 subjects with an algorithm-identified AAM, we identified 82 with AAM via manual review. Only five of the 100 subjects without an algorithm-identified AAM were found to have an identifiable AAM with manual review. Overall, our algorithm was found to have 94.4% sensitivity, 85.6% specificity, and a PPV of 84.0% (Table 14). We also calculated the accuracy of our AAM algorithm by comparing the algorithm-obtained AAM to the manual review-obtained AAM. We observed a 52.4% exact concordance within our 82 subjects with AAMs calculated from both manual review and the algorithm. If we allowed a de-identified age range encompassing an exact age to be considered concordant with the exact age obtained from the other method, our accuracy improved to 61.9%.

ANM algorithm performance

The ANM algorithm identified 83 individuals with an ANM; therefore, we manually reviewed 183 SD entries to determine the specificity, sensitivity, PPV, and accuracy of our ANM algorithm. Of the 100 individuals with no algorithm-identified ANM, manual review of the SD found 6 instances with an identifiable ANM (Table 14). Of the 83 individuals with an algorithm-specified ANM, manual review confirmed 53. Overall, the sensitivity and specificity of the ANM algorithm were 89.8% and 75.8%, respectively, and the PPV was 63.9%. Of the 53 subjects with both algorithm- and manually-identified ANM, 40 were an exact match, yielding an accuracy of 75.5%. We again observed instances where the algorithm yielded an exact age, but manual review of the SD obtained only a de-identified ANM range that encompassed the exact age, and vice-versa; if we considered these as concordant, our accuracy increased to 81.1%.

Summary

Potential use in personalized medicine

Menarche and menopause are the bookends of the reproductive lifespan in women. The timing of these events may increase risk for various complex disorders and cancers, such as osteoporosis and breast cancer (Hartge 2009). Precocious or delayed menarche may signal the occurrence of hormonal imbalance, inadequate nutrition or caloric intake, or pituitary diseases (Hartge 2009). The timing of menopause directly affects reproductive capabilities. In addition, premature menopause may result from hormonal imbalances, genetic disorders such as Fragile X Syndrome, metabolic disorders, or autoimmune diseases such as thyroid disease or rheumatoid arthritis (Okeke, Anyaehie, and Ezenyeaku 2013). Though the timing of menarche and menopause may increase risk for disease or indicate underlying pathologies, this information is not consistently included in electronic health records, leading to missed opportunities to inform clinical care and represents a challenge to clinicians and researchers alike. A clinical decision support mechanism to identify women who are at risk of developing certain diseases as a result of the timing of their reproductive milestones may be useful in clinical practice.

Potential use for genetic studies of reproductive timing

Data-mining EMRs has been used to identify cohorts for research studies (Newton et al. 2013; Stratton-Loeffler et al. 2012; Brownstein et al. 2010; Warren et al. 2012), determine smoking status (Wiley et al. 2013), and predict disease, such as sepsis (Mani et al. 2014). Our development of algorithms to extract these important data is notable for the emphasis on diverse populations and attention to women's health, both historically underrepresented in health outcomes research. The menarche (AM) and menopause (AAM) algorithms have PPV>80% and high specificity and sensitivity, though accuracy of the AAM algorithm was just over 50%. The age at natural menopause (ANM) algorithm had moderately high (>75%) sensitivity and specificity but the lowest PPV, at 63.9%. However, the accuracy of the ANM algorithm bested that of the AAM (75.5% vs. 52.4%, respectively). In addition, the algorithm-extracted ages at menarche, menopause, and natural menopause are consistent with published research, validating our methodology.

Several factors may have reduced the performance of our menopause algorithms. We observed many instances where the ages calculated by the algorithm and by manual review differed by one year. This may have been the result of the date-shifting done within each individual's SD for de-

identification purposes. If the method for calculating the age differed between the methods, it is possible this could result in the observed one-year difference. When we allowed a +/- 1 year difference in the algorithm and manual identified AAM and ANM, the accuracy of our algorithms improved to 70.2% and 90.6%, respectively. The timing of menopause is challenging to identify, as the menstrual cycle becomes more erratic as a woman moves through perimenopause into menopause. Months may lapse between cycles; hormone levels may change substantially. In addition, the normal menopausal age range is quite large, taking place between the ages of 40 and 60. These factors challenge the accurate dating of the onset of menopause.

Furthermore, an algorithm designed to identify the age at menopause may not accurately reconcile multiple mentions in an EMR of menopause. Discerning between natural menopause and medically/surgically induced menopause is an additional challenge. Our extensive list of time-dependent exclusions for HRT and surgical procedures was not exhaustive and may have led to the algorithm identifying an ANM where manual review identified HRT and/or a procedure artificially inducing menopause. Correctly identifying the temporal relationship between attainment of natural menopause and surgical procedures that result in menopause may perform inconsistently in the absence of these data in structured fields in an EMR. Addressing some of these issues by including structured fields for age at menarche, age at menopause, and type of menopause (natural/medical), and standardizing the reporting of these data could greatly improve the performance of our algorithms.

We have demonstrated the performance of algorithms designed to extract the age at menarche and age at menopause from the Synthetic Derivative, a de-identified version of the electronic medical record at Vanderbilt University Medical Center. Furthermore, we have developed an algorithm to discriminate naturally occurring menopause from artificially-induced menopause. Our method combining text-mining for regular expressions and pattern matching, and structured data derived from the EMR to obtain the age at menarche and the age at menopause is likely to be easily transferable to other institutions, given the simplicity of the approach. Overall, these algorithms provide an opportunity for researchers and clinicians to obtain these valuable, though inconsistently reported data.

CHAPTER IV

CASE STUDY: GENETIC VARIANTS ASSOCIATED WITH ENDOMETRIAL CANCER

Introduction

As discussed in Chapter I, endometrial cancer (EC) is the most common invasive gynecological cancer in the United States; an estimated 52,630 new cases arise yearly with approximately 8,590 deaths in 2014 (National Cancer Institute at the National Institutes of Health 2014b). EC typically occurs in post-menopausal women; the average age at diagnosis is 60 years (Jick, Walker, and Rothman 1980). Symptoms of EC include irregular or post-menopausal bleeding, pelvic pain, presence of a pelvic mass, and weight loss (National Cancer Institute at the National Institutes of Health 2014b). Known risk factors for Type I EC include hormone imbalances, greater than average number of menstrual cycles over the reproductive lifespan, early menarche, late age at menopause, nulliparity, tamoxifen use for breast cancer, estrogen-only hormone replacement therapy (HRT), obesity, and hypertension (Haidopoulos et al. 2010). Risk factors for Type II ECs include long term use of tamoxifen, history of pelvic radiation, and hereditary conditions like childhood retinoblastoma and hereditary leiomyomatosis and renal cell carcinoma syndrome. Prolonged exposure to estrogen-only therapy is associated with an increased risk of EC while progesterone plus estrogen has a protective effect (Jick, Walker, and Jick 1993a; Jick, Walker, and Jick 1993b). There are notable differences in EC incidence internationally, with European-Americans and US Hawaiian Japanese immigrants having the highest rates among twenty-two international cancer registries (Katanoda and Qiu 2006). Variation in EC incidence and survival across racial/ethnic groups has been observed in US cohorts (Setiawan et al. 2007; Madison et al. 2004). In the Multiethnic Cohort Study, Setiawan et al. identified lower relative risks for EC in African Americans and Latinas compared to European Americans, while the risks in Native Hawaiians and Japanese Americans were similar (Setiawan et al. 2007).

EC can be categorized into different subtypes by histology: endometrioid, serous, clear cell, and poorly differentiated (2007; Creasman et al. 2006; Creasman et al. 2004). ECs that arise in the inner layer of the uterus (endometrium) are known as endometrial carcinomas. These comprise most ECs. ECs that arise in the muscle layer (myometrium) or connective tissues are called sarcomas. EC carcinomas that begin in the cells that form the glands of the endometrium are known as adenocarcinomas; of these, the most common form is endometrioid adenocarcinomas that comprise 80% of ECs. The endometrioid type of EC tends to be less aggressive than the more uncommon forms of EC (e.g. clear cell or serous) and are the type of EC that presents clinically most often (O'Hara and Bell 2012).

EC tumors can be further divided into three histological grades (1-3) based on the percentage of tissue forming glands. Grade 1 tumors have more than 95% of the cancerous tissue forming glands; grade 2 tumors have between 50-94%. Grade 3 tumors are more aggressive, may have metastasized at the time of diagnosis, and tend to have a poor outlook compared to grades 1 and 2 (O'Hara and Bell 2012). EC tumors are also divided into two types based on estrogen dependence. Type I cancers are typically slow to spread and are not usually aggressive. They appear to be estrogen-dependent and immunostaining has demonstrated that they do not stain for p53. Grades 1 and 2 endometrioid ECs are often Type I. Type II EC cancers are estrogen-independent and appear not to be related to endometrial hyperplasia; grade 3 endometrioid ECs (serous, clear cell, and poorly differentiated) are all Type II (Kitchener and Trimble 2009). Type II cancers physiologically do not resemble normal endometrium; these cancers tend to be more aggressive than Type I and are more likely to be high-grade serous and clear cell ECs (O'Hara and Bell 2012). Type II cancers immunostain for p53, supporting the hypothesis that Types I and II ECs have a different etiology (Kitchener and Trimble 2009). A small fraction (~4%) of uterine cancers are uterine carcinosarcoma, also known as malignant mixed mesodermal or malignant mixed Mullerian tumors. These cancers contain features of both carcinomas and sarcoma, with similar natural history to Type II ECs. Along with type and grade, EC tumors are also classified by their stage; these three classifications comprise the International Federation of Gynecology and Obstetrics (FIGO) score for tumors (Table 15) (Compton 2012).

Table 15. FIGO scoring for endometrial carcinoma tumors.

Category	FIGO Stage for EC carcinomas	Description
Tumor	I, IA, IB	tumor is confined to corpus uteri, tumor may invade >1/2 or more of myometrium
	II	tumor invades stromal connective tissue of the cervix but doesn't invade beyond uterus
	IIIA, IIIB	tumor involves vagina
	IVA	tumor invades bladder and/or bowel mucosa
Lymph Node Involvement	IIIC1, IIIC2	regional lymph node metastasis pelvic (IIIC1) or para-aortic (IIIC2) lymph nodes
Metastasis	IVB	distant metastasis to lung, liver, bone or inguinal lymph nodes

Description and FIGO scoring criteria for endometrial carcinoma tumors. Abbreviations: Federation of Gynecology and Obstetrics (FIGO).

Molecular mechanisms

The molecular mechanisms responsible for development of EC are not well understood. Some forms of Type II EC have mutations in p53, a tumor suppressor gene, suggesting germline or somatic mutations play a role in development of Type II EC (Garg et al. 2010; Tashiro et al. 1997). Other risk factors include Lynch syndrome and a first-degree relative with EC, though these risk factors are not entirely understood and account for a small percent of ECs.

Changes to the HPG axis that results in excess estrogen can lead to unregulated proliferation and vascularization of the endometrium, yielding an environment amenable to the development of Type I EC. A similar result can be found in women with intact uteri who take tamoxifen for breast cancer treatment. Though tamoxifen acts in an anti-estrogenic way in breast tissue, tamoxifen behaves like estrogen in the uterus, causing growth of the endometrium (Lumachi et al. 2012). Ovarian tumors known as granulosa-theca tumors also lead to increased levels of estrogen and an increased risk for EC. Gynecological disorders, such as polycystic ovarian syndrome (PCOS) and endometriosis, with increased androgens and estrogens, also lead to an increased risk for EC, as the estrogens and androgens are accompanied by lower than normal levels of progesterone. Diminished levels of progesterone lead to unopposed levels of estrogen, disrupting the homeostatic balance of the HPG axis. This is also consistent with evidence that the risk of Type I EC is twice as high in overweight women as in normal weight women, and three times as high in obese women (Everett et al. 2003), suggesting a dose-dependent effect of estrogen on Type I EC. In general, the more menstrual cycles a woman experiences and fewer instances of pregnancy, the greater her exposure to estrogens. This leads to the repeated proliferation and vascularization of the endometrium, supporting the known increased risk of EC due to early menarche, late menopause, and nulliparity.

Genetic associations

Heritability of EC has been estimated at 0.52 (Schildkraut, Risch, and Thompson 1989), but the unknown molecular mechanisms responsible for EC have made identifying the genetic risk factors challenging. As a result, numerous candidate gene studies have been performed for EC, primarily for genes involved in sex hormone biosynthesis (Figure 4). The strong relationship between estrogen and EC risk has been explored through candidate gene studies (Ashton et al. 2009a). Two estrogen receptor SNPs in *ESR1* and four *ESR2* SNPs were assessed in a case-control study of European-descent women from Australia (Ashton et al. 2009a). The authors observed SNPs in each estrogen receptor gene that were associated with increased EC risk after adjusting for risk factors: ESR1 rs2234693 and ESR2 rs1255998 and rs944050 (Ashton et al. 2009a). Thirty-six variants involved in sex hormone metabolism, including *CYP17, ESR1,* and *ESR2* were considered in a pathway-based case-control study of Polish women (Yang et al. 2010). Only *AR*, involved in the synthesis of androgen response elements (Figure 4), was significantly associated in a gene-based analysis with EC risk in this population; however, SNP-based analyses uncovered associations with multiple SNPs in genes: *AKR1C2*, *AR*, *CYP11B1*, *HSD17B2*, *CYP19A1*, *CYP1A1/A2, SHBG,* and *SRD5A1,* though these results were not corrected for multiple tests (Yang et al. 2010). The role of *CYP17*, a gene that encodes an enzyme in the estrogen biosynthesis pathway (Figure 4), was assessed in a recent meta-analysis in a multi-ethnic study (Xu et al. 2013). In this meta-analysis, rs743572(C) was significantly associated with EC using a recessive model in European-descent and East Asian populations, but not others (Xu et al. 2013). Given the differences in EC incidence across diverse populations, additional studies to evaluate the population-specific frequencies of *CYP17* variants and their effect on EC risk may be warranted.

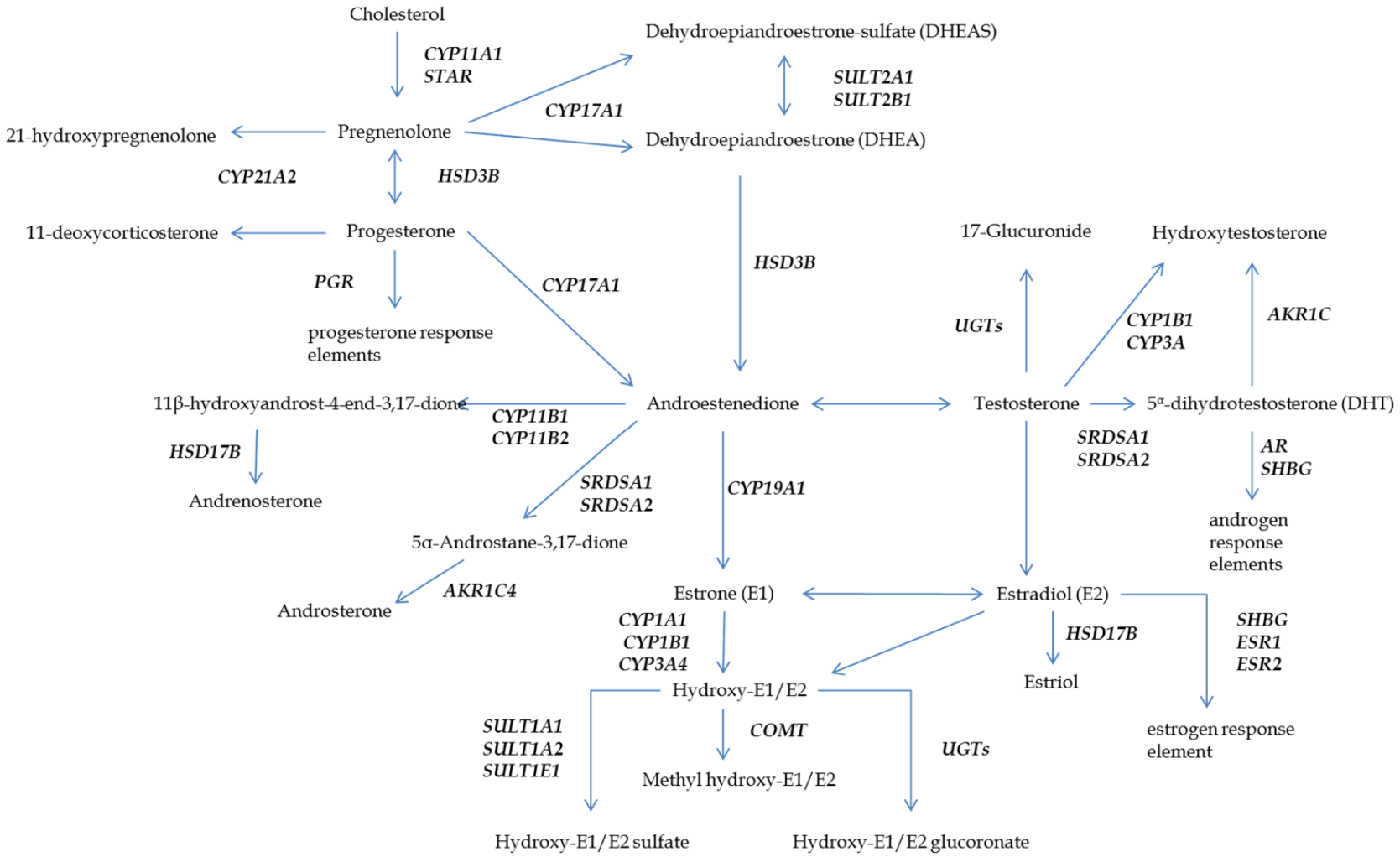


Figure 4. Sex hormone metabolic pathway.

Shown are the genes (bold) involved in the biosynthesis of sex hormones and their products. Adapted from Yang HP et al. 2010 and Hyland PL et al. 2013.

Other genes not involved in sex hormone metabolism have been considered for association with EC based on known risk factors and biologic mechanisms of disease. A proposal by Modugno et al. (Modugno et al. 2005) that inflammation may be a factor in the development of EC led Ashton *et al*. to investigate the role of *NOD* and *TLR* polymorphisms and EC in their Australian cohort (Ashton et al. 2010). Haplotype analysis demonstrated *TLR9* rs5743836(C) and rs187084(C) were protective for EC (Ashton et al. 2010), though these results have not yet been replicated in an independent cohort.

The strong positive association between obesity and endometrial cancer risk was further assessed by a candidate gene study for variants in adipokine and leptin genes (Chen et al. 2012b). Adipokines are hormones that play a role in energy regulation and insulin sensitivity (Rasmussen-Torvik et al. 2009) and have been shown to inhibit cell growth and angiogenesis and induce apoptosis (Korner et al. 2007; Dieudonne et al. 2006; Jarde et al. 2009); women with the highest levels of adiponectin were observed to have a 50% reduced risk of EC compared to women with the lowest levels, independent of BMI, in a nested case-control study in the European Prospective Investigation into Cancer and Nutrition (Cust et al. 2007). Leptins, conversely, promote cell proliferation and angiogenesis (Korner et al. 2007; Dieudonne et al. 2006; Jarde et al. 2009). In a case-control study of women from the Shanghai Endometrial Cancer Study and the Shanghai Breast Cancer Study, Chen et al. observed three SNPs in *ADIPOQ* (rs3774262, rs1063539, rs12629945) and one SNP in *LEP* (rs2071045) associated with EC risk (Chen et al. 2012b). In addition, women carrying two or more minor alleles for the *ADIPOQ* SNPs were found to have a 22% lower EC risk than women with no minor alleles (Chen et al. 2012b). Though promising, there are limitations to this study. Study subjects' adiponectin and leptin levels were not measured, preventing the authors from assessing the relationship between the significant variants and hormone levels. Additionally, this study has not been validated in an independent cohort and the authors did not correct for multiple tests (e.g., Bonferroni correction). Furthermore, the generalizability of the study results to other (non-Chinese) populations with differing obesity levels is unknown. Despite these limitations, the results are suggestive of a role for adipokines in EC risk.

As with many other diseases and traits, GWAS have been performed for EC. The first GWAS for EC was published in 2011 by Spurdle et al. in European-descent women from Australia and the UK (Spurdle et al. 2011). The authors identified *HNF1B* rs4430796(G) associated with endometrioid EC (OR=0.84, $p=7.1 \times 10^{-10}$), an association that held in their independent validation cohort of European-descent women (Spurdle et al. 2011). This locus has been previously associated with prostate cancer

(Elliott et al. 2010) and with increased risk for type 2 diabetes (Voight et al. 2010). Subsequent studies have attempted to replicate the association of EC with *HNF1B* variants, with mixed results (De, I et al. 2014; Setiawan et al. 2012; Long et al. 2012). A recent exome-wide association study failed to identify any novel variants associated with EC in a multiethnic population from the Epidemiology of Endometrial Cancer Consortium (E2C2) (Chen et al. 2014b). Given the limited number of variants associated with EC, additional studies performed in diverse populations may improve our understanding of the molecular mechanisms responsible for EC and the genetic factors influencing population-specific disease burdens.

Influence of the female reproductive lifespan

The length of the reproductive lifespan (age at menarche) and parity play a role in the risk for developing EC. An early age at menarche and/or later age at natural menopause are risks for EC, as the lifetime exposure to estrogen increases (Hartge 2009). Similarly, nulliparous women are at higher risk than parous women. EC risk in women with ten or more deliveries was found to be substantially lower than a Finnish reference population (Hognas et al. 2014). Whether this protective effect is dose-dependent (number of deliveries) or occurs at a given threshold is unknown. Use of estrogen-only hormone replacement therapy (HRT) in post-menopausal women is also a significant risk factor for EC and other poor outcomes, though absolute risk for an individual patient may differ (Ali 2014; Manson 2014; Harman 2014).

Influence of cancer-associated variants

There is evidence that some cancers may be associated with the same genetic variants, suggesting pleiotropic effects. For example, a meta-analysis of PAGE, Genetic Epidemiology of Colorectal Cancer (GECCO), and the Colon Cancer Family Registry (CCFR) identified associations with colorectal cancer in 8q24, a known cancer locus (Cheng et al. 2014). As noted above, *HNF1B* has been associated with both EC (Spurdle et al. 2011) and with prostate cancer (Elliott et al. 2010). Females with Lynch syndrome, a hereditary colorectal cancer syndrome caused by mutations in DNA mismatch repair genes, have a 40-60% lifetime risk of developing endometrial cancer (Sehgal et al. 2014). Whether additional variants, associated with other cancers, are also associated with EC remains unknown.

Methods

Selection of genetic variants

Variants were selected based upon their association with other cancers and identified by PAGE I Study investigators from the NHGRI GWAS catalog as of January 2010, fine mapping literature, and review of recent cancer GWAS. One hundred twenty-two (122) candidate SNPs plus 128 ancestry informative markers (AIMS) were genotyped using Sequenom and TaqMan assays (Appendix G).

Study population

The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study, as part of the Population Architecture using Genomics and Epidemiology (PAGE) I Study, accessed DNA samples from BioVU, the Vanderbilt University Medical Center biorepository linked to de-identified medical records. A description of BioVU was presented in Chapter III. Women 18 years of age and older were eligible for this study. Endometrial cancer cases were required to have a tumor registry entry for endometrial (uterine) cancer with a primary site of C540 (Isthmus uteri), C549 (Corpus uteri), or C559 (Uterus NOS); women with lymphoma or leukemia primary histology (9590-9989) were excluded. Controls for this study were females from the eligible BioVU population, excluding women with any ICD-9-CM code for EC, a tumor registry entry, or cancer-associated terminology in their synthetic derivative record.

Statistical methods

Quality control (QC) methods were performed based on the QC consensus pipeline from the PAGE Coordinating Center (Cheng et al. 2014). The principal components method implemented in EIGENSTRAT was used to adjust for potential population stratification (Price et al. 2006). Samples were excluded based on call rates (<90%), departure from Hardy-Weinberg equilibrium ($p<0.001$), and concordance of blinded replicates (≤98%). A total of 114 candidate SNPs passed the QC process and were assessed for association with EC. An additive genetic model was selected for logistic regression between each SNP and EC; the 'risk' allele was defined as the allele previously associated with increased cancer risk in prior publications. We considered a minimally adjusted model and a fully adjusted model that included age, principal components, and body mass index (BMI). We defined significance as $p<4.39 \times 10^{-4}$, which is Bonferroni corrected (0.05/114).

Unexpectedly, only 20/206 EC cases and 156/2227 controls (European descent) and 2/20 cases and 22/335 controls (African American) remained for analysis after QC. Approximately half of the

samples had BMI data, a critical covariate in the EC study. Therefore, we contributed our European-descent samples' data to a larger meta-analysis (Setiawan et al. 2014) from the PAGE Study (Matise et al. 2011) and the E2C2 (Olson et al. 2009). Log odds regression estimates were combined across study sites in a fixed-effect, inverse-variance weighted meta-analysis implemented in METAL (Willer, Li, and Abecasis 2010). Heterogeneity across studies was estimated based on Cochran's Q statistic. For the meta-analysis, significance was defined as a Bonferroni-corrected $p<2.35x10^{-4}$ (0.05/213) (Setiawan et al. 2014).

Results

Presented are the single-site (EAGLE BioVU) results for the EC study. The controls were slightly older than the cases; BMI for the cases and controls was comparable (Table 16). No SNPs met the Bonferroni-corrected significance threshold of $4.39x10^{-4}$ (Appendix H). The most significant SNPs at $p<0.05$ are listed in Table 17. Full logistic regression results for the EAGLE BioVU study are presented in Appendix H.

Table 16. Population characteristics for EAGLE BioVU endometrial cancer study.

	Cases	Controls
Number of subjects (n)	20	156
Mean age (yrs.)	88.1	89
Mean BMI (kg/m^2)	25.8	25.5

Table 17. Significant results from EAGLE BioVU endometrial cancer analysis.

CHR	SNP	GENE	A1	NMISS	OR	SE	P-value
8	rs10086908	intergenic	T	92	8.06	0.80	$9.20x10^{-3}$
2	rs1465618	*THADA*	T	92	7.13	0.82	0.02
12	rs10778826	*PPFIA2*	A	92	4.41	0.63	0.02
8	rs6983267	*CCAT2*	G	92	0.26	0.59	0.02
8	rs10505477	LOC101930033	A	92	0.26	0.59	0.02

Data shown are significant SNP associations at $p<0.05$ for EAGLE BioVU endometrial cancer analysis. For each SNP, p-values, risk allele, odds ratio, number of cases and controls are given for a fully adjusted model that included age, principal components, and body mass index (BMI). Gene shown is the closest gene to the SNP. Abbreviations: chromosome (CHR), single nucleotide polymorphism (SNP), risk allele (A1), number of cases and controls used in each logistic regression (NMISS), odds ratio (OR), standard error (SE).

The most significant associations from the EC logistic regression include SNPs previously implicated in prostate cancer and colorectal cancer. The most significant result from our analysis ($p<0.05$) was rs10086908, located in an intergenic region in chromosome 8q24. This locus has been identified in several studies to be associated with prostate cancer (Liu et al. 2011; Xu et al. 2009b; Liu et al. 2008; Robbins et al. 2007). *THADA* rs1465618 at 2p21 has been associated with prostate cancer in both Europeans (Eeles et al. 2009) and Chinese males (Zhao et al. 2014). rs6983267, in 8q24.21, a colorectal cancer susceptibility SNP has recently been associated with prostate cancer and tumor volume in European descent men (Reinhardt et al. 2013) and with prostate cancer in a Hispanic (Chilean) population (San Francisco et al. 2014). rs10505477, another colorectal cancer associated SNP is also located in the 8q24 region (Table 17).

Comparing our single-site results to those of the larger meta-analysis, we failed to find agreement between any of the most significant associations ($p<0.05$) observed in EAGLE BioVU (Table 17) and the associations reported by the PAGE I Study meta-analysis ($p<0.05$) (Setiawan et al. 2014). The lack of agreement is expected given the smaller sample size and limited power of EAGLE BioVU (20 cases, 156 controls) compared with the overall PAGE I Study meta-analysis (3758 cases, 5966 controls) (Setiawan et al. 2014). Despite this lack of similarity, in both the single-site and meta-analysis, the most significant results were also associated with prostate cancer, suggesting pleiotropic effects for these SNPs and perhaps a shared biological mechanism for the development of both cancers (Table 17, (Setiawan et al. 2014)).

Summary

Pleiotropy, the association of a genetic locus with more than one distinct phenotype (Stearns 2010; Solovieff et al. 2013), has been investigated for a few diseases, most recently with the PheWAS—the phenome-wide association study (Pendergrass et al. 2011; Denny et al. 2010). The PheWAS design is similar to that of the GWAS; where a GWAS interrogates multiple SNPs for association with a single phenotype, the PheWAS interrogates multiple phenotypes with a single variant (Pendergrass et al. 2011; Denny et al. 2010). This approach may uncover previously unknown genotype-phenotype associations and suggest underlying biological mechanisms that are shared across phenotypes. Though PheWAS is a new approach for assessing pleiotropy on a large scale, testing individual variants where prior studies have shown associations across multiple phenotypes remains a valid approach.

Studies have identified multiple genetic variants that are associated with several cancer phenotypes. In a recent analysis, SNPs in the 5p15.33 region, were assessed for their association with six distinct cancers (Wang et al. 2014b). This region has been previously linked to ten different cancers, including bladder (Rothman et al. 2010), glioma (Shete et al. 2009), lung (Wang et al. 2008), ovarian (Beesley et al. 2011), melanoma (Rafnar et al. 2009), and prostate cancer (Kote-Jarai et al. 2011) and contains *TERT*, a gene coding for a subunit of telomerase reverse transcriptase (Kim et al. 1994) and *CLPTM1L*, a gene implicated in lung and pancreatic cancer (James et al. 2012; Jia et al. 2014). Wang et al. identified five SNPs in *TERT* and one in *CLPTM1L* with significant pleiotropic effects (Wang et al. 2014b).

This study and the meta-analysis to which it contributed, considered how GWAS-identified variants from other cancer studies might be associated with EC. For example, *HNF1B* is associated with endometrial cancer (Spurdle et al. 2011) and prostate cancer (Elliott et al. 2010) and has been linked to increased risk for type 2 diabetes (Voight et al. 2010). Interestingly, we failed to find *HNF1B* variants associated with EC at a nominal ($p<0.05$) significance level in EAGLE BioVU (Appendix H). However, we did observe SNPs in the 8q24 region, previously associated with colorectal (Real et al. 2014) and prostate cancer (Robbins et al. 2007), to be nominally significant in our study (Table 17). Additionally, *THADA* rs1465618 was nominally significant in our study and has been previously associated with prostate cancer (Eeles et al. 2009). The mechanism by which the 8q24 intergenic region contributes to cancer susceptibility remains unknown. This locus appears to generalize to diverse populations--African American (Han et al. 2014), European (Eeles et al. 2009), and Chinese males (Hui et al. 2014); its association with multiple cancer phenotypes further suggests a role as a part of a general cancer mechanism.

A significant limitation of this single-site study is the sample size. Though we were adequately powered (power=0.80) to identify an odds ratio (OR) ≈ 2.25 with our pre-QC sample size of 206 cases and 2227 controls (European descent), once quality control procedures were completed, our sample size dropped to 20 cases. We would have needed to detect an OR>4.75 with an allele frequency ≥0.15 to have sufficient power with this reduced sample size (Appendix I). Given that we were underpowered to detect associations of modest size, we contributed our data to a larger effort that combined studies from two consortia: PAGE and E2C2. In addition, this study only assessed a relatively small number of SNPs that were previously associated with a variety of cancers, primarily through GWAS. It is likely that other SNPs, not considered in this analysis, may also demonstrate pleiotropic effects in multiple

cancer phenotypes. Future studies investigating this potential are needed. Furthermore, despite initial intentions of including African American samples in this analysis, the number of EAGLE BioVU cases was insufficient (n=2). The lack of EC minority cases makes generalizing results to a non-European population challenging; the larger meta-analysis also limited their results to European-descent women. Additional studies in more diverse populations should be performed to verify the associations identified here and in other studies. Despite these limitations, this single-site study identified nominally significant associations between prostate cancer SNPs and endometrial cancer, suggesting these variants may have pleiotropic effects across multiple cancer phenotypes.

CHAPTER V

CASE STUDY: A GENOME-WIDE ASSOCIATION STUDY FOR SERUM THYROID STIMULATING HORMONE LEVELS[3]

Introduction

Hyperthyroidism and hypothyroidism are important endocrine diseases caused by over- or under-production of thyroid hormone, which is regulated by thyroid stimulating hormone (TSH) produced in the anterior pituitary gland. Hypothyroidism, the most common thyroid disease, can be caused by iodine insufficiency, autoimmunity, pregnancy, pituitary disease (leading to increased TSH production), or other conditions. Thyroid diseases occur more often in women than in men (Vanderpump 2011) and the risk of developing hypothyroidism increases with age (Laurberg et al. 2005; Bagchi, Brown, and Parish 1990). Diagnosis of thyroid diseases involves measuring TSH levels and circulating thyroxine (T4) and triiodothyronine (T3) in the blood; elevated TSH levels and depressed T4 levels signify clinical hypothyroidism (Laurberg et al. 2005; Means 1940) while elevated TSH levels and normal T4 levels indicate mild (subclinical) hypothyroidism (Hollowell et al. 2002). TSH is produced by a normally functioning pituitary gland in response to decreased thyroid hormone levels; as thyroid hormone levels decrease, TSH signals to the thyroid to produce additional thyroid hormone. When the thyroid gland does not maintain sufficient production of thyroid hormone, serum TSH levels become elevated, and the individual develops hypothyroidism. Similarly, elevated thyroid hormone levels from primary hyperthyroidism result in decreased TSH levels.

Both genetic and environmental factors influence serum TSH levels. Physical and emotional stress, poor nutrition, increased body mass index (BMI), current smoking, and pregnancy are all risk factors for elevated serum TSH levels (Brix et al. 2000; Jorde and Sundsfjord 2006; Nyrnes, Jorde, and Sundsfjord 2006). Normal serum TSH levels range from 0.3 μIU/mL – 4.0 μIU/mL but are tightly

[3] Adapted from: Malinowski JR, Denny JC, Bielinski SJ, Basford MA, Bradford Y, *et al*. Genetic variants associated with serum thyroid stimulating hormone (TSH) levels in European Americans and African Americans: an eMERGE Network analysis. *PLoS One.* In press.

regulated within an individual, suggesting a genetic 'set point' for individual thyroid hormone levels (Hollowell et al. 2002; Chiamolera and Wondisford 2009; Arnaud-Lopez et al. 2008). A cross-sectional population study demonstrated differences in mean TSH levels between race/ethnicities, with higher mean TSH levels in non-Hispanic whites than in Mexican Americans or non-Hispanic blacks (Hollowell et al. 2002). The etiology behind the observed differences in mean TSH levels across ethnic groups has not been elucidated, and it is unclear if those differences lead to lower prevalence of hypothyroidism in populations of diverse ancestry. A recent study identified differences in prevalence of thyroid cancer across racial/ethnic groups living in England (Finlayson et al. 2014), and TSH antibodies were demonstrably lower in non-Hispanic blacks compared to non-Hispanic whites or Mexican-Americans in the National Health and Nutrition Examination Survey (NHANES) III (Spencer et al. 2007); however, studies evaluating hypothyroidism or hyperthyroidism burden among different racial/ethnic groups have not been performed. Twin and family-based studies have suggested heritability estimates of 32%-67% for TSH, T4, and T3 levels (Panicker et al. 2008b; Panicker et al. 2008a; Panicker 2011). Several genetic association studies have been performed, including two meta-analyses of GWAS (Porcu et al. 2013; Rawal et al. 2012). These studies have identified common variants associated with serum TSH levels: rs2046045 (*PDE8B*), rs10917477 (*CAPZB*), rs10028213 (*NR3C2*), and rs3813582 (16q23) (Panicker et al. 2010; Porcu et al. 2013). Altogether, the known loci explain <5% of the variance in TSH levels (Rawal et al. 2012). However, these GWAS and meta-analyses have been performed in populations of European ancestry, and it is unclear if these findings generalize to other race/ethnicities.

In this study, we sought to identify variants associated with normal variability of serum TSH levels in euthyroid (thyroid disease free) European Americans and African Americans from the Electronic Medical Records and Genomics (eMERGE) Network. We looked to replicate in our study known associations between SNPs and serum TSH levels. We hypothesized variants associated with serum TSH levels might also be associated with thyroid disorders, such as hyperthyroidism (Grave's disease), hypothyroidism (Hashimoto's disease), and thyroid cancer. Given that increased BMI is a risk factor for elevated serum TSH levels, we also tested for evidence that TSH-associated SNPS are modified by BMI in this study of euthyroid European and African Americans from the eMERGE Network.

Methods

The eMERGE Network is a collaboration of institutions with biobanks linked to EMRs. The data for these analyses included Phase I of the eMERGE Network whose members included Group Health Cooperative/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, Vanderbilt University and the eMERGE Administrative Coordinating Center (McCarty et al. 2011).

Study population and phenotype

This study was performed in the eMERGE Network which includes approximately 17,000 individuals who were genotyped for a variety of complex diseases (e.g. dementia, cataracts, peripheral arterial disease (PAD), type 2 diabetes) and medically relevant quantitative traits (e.g. cardiac conduction) (Denny et al. 2011). To qualify for euthyroid designation in this analysis, individuals were required to have at least one test of thyroid function (i.e., TSH and T3 or T4 if available) with no abnormal results, must not have any billing codes for hypothyroidism or history of myasthenia gravis in his/her EMR or evidence of thyroid replacement medication, and must have at least two past medical history sections (non-acute visits) and medication lists. For individuals with multiple TSH tests, the median TSH level was used in the analysis. Individuals were excluded if they had any cause of hypothyroidism or hyperthyroidism, any other thyroid diseases (e.g. Graves, thyroid cancer) as indicated by billing (ICD-9) codes, procedure (CPT) codes or text word diagnoses, or were on thyroid-altering medication (e.g., lithium) (Denny et al. 2011). From this group, 6,086 European Americans and 633 African Americans qualified as euthyroid, of which 4,501 European Americans and 351 African Americans had body mass index (BMI). The appropriate institutional review board at each participating study site approved all procedures.

Genotyping

Genotyping was performed using the Illumina Human660W-Quadv1_A and the Illumina1M BeadChips for European Americans and African Americans, respectively, as previously described (Denny et al. 2011). Of the SNPs on each array, 474,366 SNPs and 905,285 SNPs, respectively, passed quality control filters for tests of genotyping efficiency (>99% call rate), and minor allele frequency (>5%). Details of eMERGE quality control have been previously published (Turner et al. 2011; Zuvich et al. 2011). Briefly, datasets from each of the five participating eMERGE Phase I sites were combined at the eMERGE Coordinating Center. Sample relatedness was assessed by calculating pairwise kinship

estimates in PLINK with the –genome command; the sample with the lowest call rate of the related pairs (e.g., full sibling, avuncular, parent-offspring) was removed (Zuvich et al. 2011). Population stratification was examined using the principal components analysis implemented in EIGENSTRAT (Price et al. 2006). Hardy-Weinberg equilibrium (HWE) was evaluated after dividing the data into two main groups by ancestry, European and African; a HWE threshold of $p<1x10^{-4}$ was implemented. Call rates and minor allele frequencies were examined for each study site individually to assess batch effects that could skew results (Zuvich et al. 2011).

Statistical methods

Quality control and data analysis were performed using a combination of PLINK (Purcell et al. 2007; Purcell 2009), and R software, and data were plotted using R code obtained from the Getting Genetics Done website (Turner and Bush 2012; R Development Core Team 2012), Stata (StataCorp 2011) and Synthesis-View (Pendergrass et al. 2010). Power calculations were performed using Quanto (Gauderman 2002a). Linear regression was performed assuming an additive genetic model to test for associations between individual SNPs and log-transformed median serum TSH levels. Tests were performed stratified by race/ethnicity, unadjusted and adjusted for age, sex, BMI, and first principal component (PC1) calculated with EIGENSTRAT (Price et al. 2006). Additional tests of association were performed in European Americans stratified by BMI (normal: BMI 18.5-24.9; overweight: BMI≥25) and adjusted for age, sex, and PC1. We also performed formal tests of interaction between SNPs associated with TSH levels as a significance threshold of $p<1x10^{-4}$ and stratified BMI (normal versus overweight) stratified by race/ethnicity in adjusted (age, sex, PC1, and main effects) models. We considered a SNP-BMI interaction significant at a threshold of $p<0.05$. Wilcoxon rank-sum tests were performed to compare median TSH levels at each genotype for normal vs. overweight BMI categories for each SNP.

In addition to GWAS discovery, we sought to replicate and generalize previously reported genetic associations for TSH levels. We considered a SNP replicated in European Americans if the tested SNP was identical to the index SNP, or a proxy in strong linkage disequilibrium (LD) ($r^2>0.7$) with the index SNP in 1000 Genomes CEU reference panel, and the direction of effect was consistent with the previous report after taking into account coding allele differences. We considered a SNP generalized to African Americans if the tested SNP was identical to, or a proxy in strong LD with ($r^2>0.7$), the index SNP in 1000 Genomes CEU reference panel, and the direction of effect was consistent with European Americans. For the replication/generalization analysis, significance was defined at a

threshold of $p<0.05$. Power calculations were performed assuming the genetic effect sizes reported in the literature, the present study sample size, and the present study coded allele frequencies.

Results

All eMERGE participating sites contributed data for European Americans and all sites except Marshfield Clinic contributed data for African Americans (Appendix J). Collectively, European Americans had higher mean TSH levels compared to the African Americans (1.90 μIU/mL vs. 1.45 μIU/mL), had lower BMI (27.51 kg/m^2 vs. 32.16 kg/m^2), included more men (52.19% male vs. 25.07%), and were older (median decade of birth 1930s vs. 1950s) (Table 18). The higher mean TSH level in European Americans compared to African Americans is consistent with previous epidemiologic reports (Vanderpump 2011; Hollowell et al. 2002; Boucai and Surks 2009). The age, BMI, and sex ratio differences between the groups observed here most likely reflect ascertainment differences resulting from the characteristics of the source populations at each eMERGE site, rather than true differences at the overall population level.

Table 18. Population characteristics in euthyroid individuals for serum TSH levels in the eMERGE Network.

		European Americans (n=4,501)	African Americans (n=351)
Female (%)		47.81	74.93
Body mass index, kg/m^2		27.51 (5.55)	32.16 (8.43)
Age at lab, years		65.50 (12.48)	50.59 (18.41)
TSH levels, μIU/mL		1.90 (0.93)	1.45 (0.72)
Decade of birth, # (%)	1910s	608 (13.51)	30 (8.55)
	1920s	865 (19.22)	40(11.40)
	1930s	994 (22.08)	22 (6.27)
	1940s	1246 (27.68)	40 (11.40)
	1950s	612 (13.60)	71 (20.23)
	1960s	89 (1.98)	67 (19.09)
	1970s	48 (1.07)	42 (11.97)
	1980s	38 (0.84)	38 (10.83)
	1990s	1 (0.02)	1 (0.28)

Data are presented as means (standard deviation) unless otherwise noted. Abbreviations: thyroid stimulating hormone (TSH), the Electronic Medical Records and Genomics (eMERGE).

Discovery

We performed standard single SNP tests of association stratified by race/ethnicity and adjusted for sex, age (decade of birth), BMI, and PC1. For European Americans, we identified six SNPs in

PDE8B on chromosome 5 as associated with TSH levels at genome-wide significance ($p<5x10^{-8}$) (Table 19, Figure 5). Our most significant result, rs1382879, was a perfect proxy for previously-identified (Rawal et al. 2012) rs2046045 ($r^2=1.00$) and was in moderate-to-high LD ($r^2>0.30$) with the other significant *PDE8B* SNPs. No novel genotype-phenotype associations were identified at genome-wide significance in this sample of European Americans. However, an additional 111 SNPs were suggestively associated with serum TSH levels ($p<1x10^{-4}$), including seven SNPs in *PDE8B,* ten SNPs near *FOXE1,* three SNPs in *PDE10A,* four SNPs in *THBS4*, and eight SNPs in *NRG1* (Appendix K). The majority of these SNPs are located in noncoding regions of the genome (intronic, upstream, downstream); however, rs3745746 (*CABP5*, $p=4.93x10^{-5}$) is a missense mutation, and rs1443434 (*FOXE1*, $p=6.53x10^{-5}$) is located in the 3′ untranslated region.

Table 19. Genome-wide significant SNP associations for serum TSH levels in eMERGE euthyroid European Americans.

CHR	SNP	GENE	GENE REGION	CODED ALLELE	BETA (SE)	P-VALUE
5	rs1382879	*PDE8B*	intronic	G	0.09 (0.01)	$7.16x10^{-18}$
5	rs2046045	*PDE8B*	intronic	C	0.09 (0.01)	$1.85x10^{-17}$
5	rs989758	*PDE8B*	intronic	T	0.08 (0.01)	$1.33x10^{-14}$
5	rs9687206	*PDE8B*	intronic	G	0.08 (0.01)	$5.52x10^{-14}$
5	rs12515498	*PDE8B*	intronic	C	0.07 (0.01)	$3.27x10^{-10}$
5	rs6885813	*PDE8B*	intronic	A	0.06 (0.01)	$4.05x10^{-8}$

Tests of association using linear regression for 474,366 SNPs assuming an additive genetic model and adjusted for age, sex, principal component (PC1), and body mass index were performed. Significance defined as $p<5x10^{-8}$. Abbreviations: single nucleotide polymorphism (SNP), thyroid stimulating hormone (TSH), Electronic Medical Records and Genomics (eMERGE).

No SNPs were associated with TSH levels in African Americans at the genome-wide significance threshold of $p<5.0x10^{-8}$ (Appendix L). However, 87 SNPs reached a suggestive significance level ($p<1x10^{-4}$); the most significant result was rs1409005 (*POU4F1-AS1*, $p=5.02x10^{-7}$). Similar to the results in the European Americans, the majority of these SNPs were located in noncoding regions except for two missense mutations (*COQ5* rs3742049, $p=6.08x10^{-5}$; *RBM20* rs942077, $p=8.47x10^{-5}$) and one synonymous substitution (*KLK1* rs1054713, $p=4.16x10^{-5}$) (Appendix L).

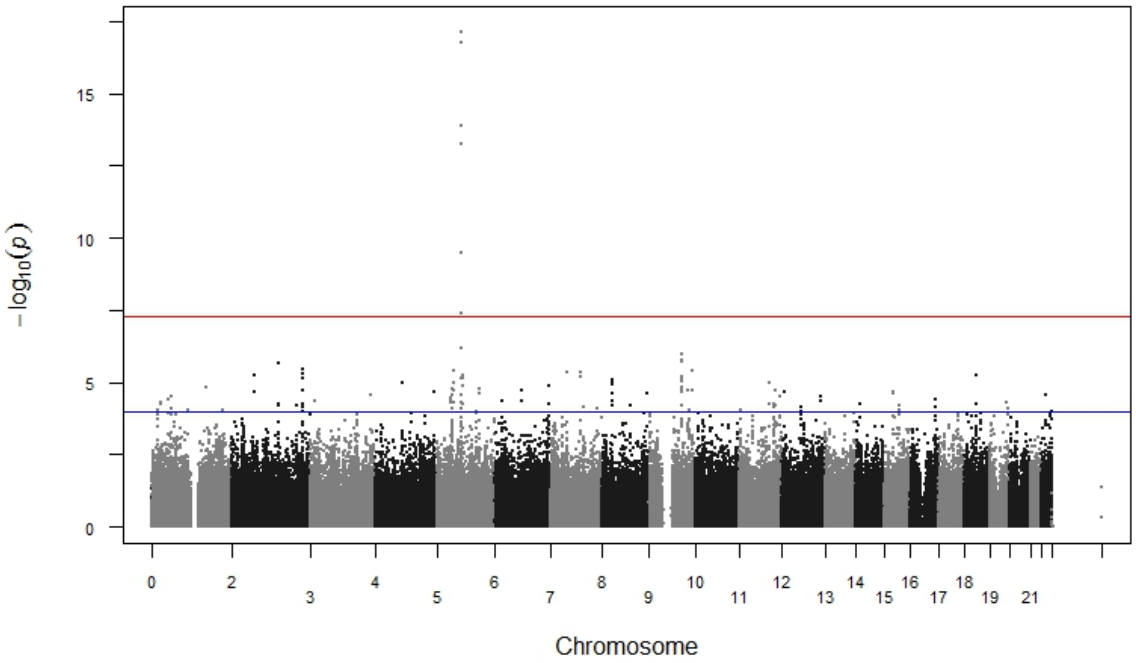


Figure 5. Manhattan plot of tests of association with serum TSH levels in euthyroid European Americans in eMERGE Network.

Data shown are p-values from single SNP tests of association with serum TSH levels in a model adjusted for age, sex, principal component (PC1), and body mass index in euthyroid European Americans in eMERGE Network (n=4,501). The y-axis represents the –log10 (p-value); horizontal lines represent Bonferroni corrected significance level ($p<5x10^{-8}$)(top) and suggestive significance level ($1x10^{-4}$)(bottom). Chromosomes are arranged on the x axis. Abbreviations: thyroid stimulating hormone (TSH), Electronic Medical Records and Genomics (eMERGE), single nucleotide polymorphism (SNP).

Trans-population genetic associations

Given the smaller sample size of African Americans with serum TSH levels, the GWAS was underpowered to detect associations at genome-wide significance with expected small to moderate effect sizes. Therefore, we evaluated the 31 most significant ($p<1x10^{-5}$) associations from the European American dataset for evidence of generalization to the African American dataset at a liberal significance threshold of 0.05 (Figure 6). One SNP, rs813379, was not directly genotyped in African Americans. We observed two SNPs in *PDE8B* associated with serum TSH levels in European Americans (rs2046045: $p=1.85x10^{-17}$ and rs12520862: $p=7.48x10^{-6}$) that were also associated in African Americans (p=0.03 and 0.01, respectively) with consistent directions and magnitude of effect after accounting for the coded allele (Figure 6). We also observed two SNPs upstream of *IGFBP5* (rs1861628 and rs13020935) associated both in European Americans ($p=3.68x10^{-6}$ and $7.02x10^{-6}$, respectively) and African Americans ($1.82x10^{-4}$ and $1.82x10^{-4}$, respectively). These SNPs are in perfect LD in both 1000 Genomes CEU and YRI reference panels ($r^2=1.00$). Interestingly, while the direction of effect was consistent between the two populations, the magnitude of effect was larger in African Americans β= -0.1492, SE=0.04; β=-0.1492, SE=0.04, respectively) compared with European Americans (β = -0.05, SE=0.01; β=-0.05, SE=0.01, respectively). One additional variant, *ABO* rs657152, was significant in both European Americans ($p=4.17x10^{-6}$, β=0.05) and African Americans (p=0.03, β=0.09). Overall, most genetic associations identified in European Americans for serum TSH levels were not significant (p<0.05) in African Americans (25/30; 83.3%); however, the majority of associations (21/30; 70.0%) had genetic effects in the same direction between the two populations (Figure 6).

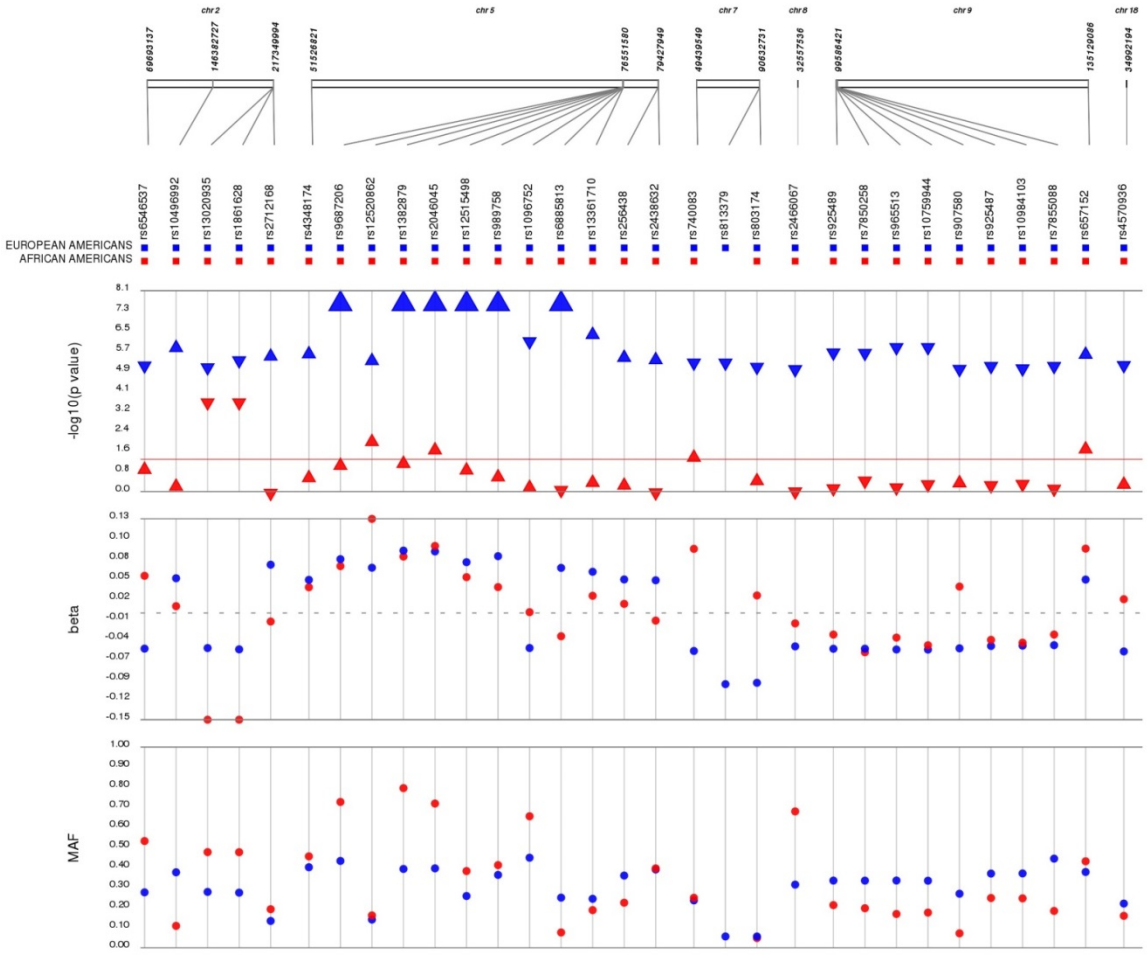


Figure 6. Comparison of most significant associations identified in European Americans with African Americans from the eMERGE Network.

Shown are p-values, coded allele frequencies, and betas for euthyroid European Americans (n=4,501) and African Americans (n=351) in the eMERGE Network for serum TSH level tests of association using SynthesisView. Data shown are comparisons between European Americans (blue markers) and African Americans (red markers) for p-values (data shown are –log10(p-value)), genetic effect magnitudes (beta), and coded allele frequencies (MAF) for the 31 most significant SNPs in European Americans. Red horizontal line on p-value track indicates p=0.05. SNPs are oriented across the top of the figure, arranged by chromosomal location. Large triangles represent p-values at or smaller than $5x10^{-8}$. Direction of the marker for p-values indicates direction of effect for each SNP. Abbreviations: Electronic Medical Records and Genomics (eMERGE), thyroid stimulating hormone (TSH), single nucleotide polymorphism (SNP).

Replication and Generalization

At least 24 SNPs have been associated with serum TSH levels in European descent populations in the literature (Rawal et al. 2012; Porcu et al. 2013; Taylor et al. 2011; Panicker et al. 2010). We considered a SNP replicated if the direction of effect was the same as previously reported and associated at a liberal threshold of $p<0.05$ with serum TSH levels. In European Americans, we replicated 22/25 (88%) SNPs previously associated with serum TSH levels (Table 20). As previously mentioned, the most significant association with TSH levels in European Americans replicated the published reports for *PDE8B* SNPs rs2046045 and rs6885099 (Table 20). Beyond *PDE8B*, we replicated two SNPs on chromosome 1 in *CAPZB* previously implicated as associated with serum TSH levels (Table 20). One SNP, rs12138950, was a perfect proxy for previously-reported *CAPZB* rs10917469 (1000 Genomes CEU $r^2=1.00$, $\beta=-0.05$, $p=8.97x10^{-5}$) (Table 20).

Table 20. Comparison of prior associations with TSH levels to eMERGE European Americans.

SNP	Locus Chr	Gene/ Gene Region	CA	CAF	Prior Association β (SE)	Prior Association P-value	Ref.	SNP/Best Proxy SNP	r²	Current Study CA	Current Study CAF	Current Study β (SE)	Current Study P-value
rs10917469	1	*CAPZB*	G	0.16	-0.16(0.03)	$3.2x10^{-08}$	(Panicker et al.2010)	rs12138950	1.00	C	0.15	-0.05(0.01)	$8.97x10^{-05}$
rs10917477	1	*CAPZB*	A	0.51	-0.06(0.01)	$1.54x10^{-08}$	(Rawal et al.2012)	rs6683419	0.73	G	0.48	0.04(0.01)	$3.56x10^{-04}$
rs10799824	1	*CAPZB*	A	0.16	-0.11(0.01)	$3.60x10^{-21}$	(Porcu et al.2013)	rs12138950	0.95	C	0.15	-0.05(0.01)	$8.97x10^{-05}$
rs334699	1	*NFIA*	A	0.05	-0.14(0.02)	$5.40x10^{-12}$	(Porcu et al.2013)	rs334708	0.79	C	0.08	-0.05(0.02)	$7.20x10^{-03}$
rs13015993	2	*IGFBP5*	A	0.74	0.08(0.01)	$3.24x10^{-15}$	(Porcu et al.2013)	rs1861628	1.00	T	0.27	-0.05(0.01)	$3.68x10^{-06}$
rs10028213	4	*NR3C2*	C	0.82	0.08(0.01)	$2.88x10^{-10}$	(Rawal et al.2012)	rs10519980	1.00	T	0.18	-0.04(0.01)	0.001
rs10032216	4	*NR3C2*	T	0.78	0.09(0.01)	$9.28x10^{-16}$	(Porcu et al.2013)	rs17025017	1.00	A	0.19	-0.04(0.01)	$2.38x10^{-03}$
rs2046045	5	*PDE8B*	T	0.62	-0.12(0.01)	$2.79x10^{-27}$	(Rawal et al.2012;Eriksson et al.2012;Medici et al.2011)	rs2046045	--	C	0.40	0.09(0.01)	$1.85 x10^{-17}$
rs6885099	5	*PDE8B*	A	0.59	-0.14(0.01)	$1.95x10^{-56}$	(Porcu et al.2013)	rs2046045	1.00	C	0.40	0.09(0.01)	$1.85 x10^{-17}$
rs4704397	5	*PDE8B*	A	0.40*	0.21	$1.64x10^{-10}$	(Taylor et al.2011)	rs1382879	0.94	G	0.39	0.09(0.01)	$7.16 x10^{-18}$
rs753760	6	*PDE10A*	C	0.69	0.10(0.01)	$1.21x10^{-24}$	(Porcu et al.2013)	rs2983514	0.93	G	0.33	-0.05(0.01)	$1.36 x10^{-05}$
rs9472138	6	*VEGFA*	T	0.29	-0.08(0.01)	$6.72x10^{-16}$	(Porcu et al.2013)	rs9472138	--	T	0.28	-0.04(0.01)	$6.41 x10^{-04}$
rs11755845	6	*VEGFA*	T	0.27	-0.07(0.01)	$1.68x10^{-10}$	(Porcu et al.2013)	rs11755845	--	T	0.24	-0.02(0.01)	0.04
rs9497965	6	*SASH1*	T	0.42	0.05(0.01)	$2.25 x10^{-08}$	(Porcu et al.2013)	rs9377117	0.54	G	0.30	0.02(0.01)	0.12
rs7825175	8	*NRG1*	A	0.21	-0.07(0.01)	$2.94 x10^{-09}$	(Porcu et al.2013)	rs2466067	0.21	C	0.31	-0.05(0.01)	$8.41 x10^{-06}$
rs657152	9	*ABO*	A	0.34	0.06(0.01)	$4.11 x10^{-10}$	(Porcu et al.2013)	rs657152	--	T	0.38	0.05(0.01)	$4.17 x10^{-06}$
rs1571583	9	*GLIS3*	A	0.25	0.06(0.01)	$2.55 x10^{-08}$	(Porcu et al.2013)	rs1571583	--	T	0.25	0.03(0.01)	0.01
rs17723470	11	*PRDM11*	T	0.28	-0.07(0.01)	$8.83 x10^{-11}$	(Porcu et al.2013)	rs7940871	0.89	T	0.29	-0.04(0.01)	$1.42 x10^{-04}$
rs1537424	14	*MBIP*	T	0.61	-0.05(0.01)	$1.17 x10^{-08}$	(Porcu et al.2013)	rs1537424	--	G	0.43	0.03(0.01)	$2.89 x10^{-03}$
rs11624776	14	*ITPK1*	A	0.66	-0.06(0.01)	$1.79 x10^{-09}$	(Porcu et al.2013)	rs957362	0.31	C	0.22	0.02(0.01)	0.09
rs10519227	15	*FGF7*	A	0.25	-0.07(0.01)	$1.02 x10^{-11}$	(Porcu et al.2013)	rs7168316	1.00	T	0.23	-0.05(0.01)	$2.10 x10^{-05}$
rs17776563	15	*MIR1179*	A	0.32	-0.06(0.01)	$2.89 x10^{-10}$	(Porcu et al.2013)	rs11073790	0.81	T	0.35	-0.01(0.01)	0.24
rs3813582	16	LOC440389/ *MAF*	T	0.67	0.08(0.01)	$8.45 x10^{-18}$	(Rawal et al.2012;Porcu et al.2013)	rs17767383	1.00	A	0.31	-0.04(0.01)	$1.42 x10^{-04}$
rs9915657	17	*SOX9*	T	0.54	-0.06(0.01)	$7.53 x10^{-13}$	(Porcu et al.2013)	rs9915657	--	C	0.46	0.03(0.01)	$9.53 x10^{-04}$
rs4804416	19	*INSR*	T	0.57	-0.06(0.01)	$3.16 x10^{-10}$	(Porcu et al.2013)	rs4804416	--	G	0.44	0.03(0.01)	$7.20 x10^{-04}$

SNP rs number, chromosomal location, nearest gene/gene region, coded allele (CA), coded allele frequency (CAF), and association summary statistics (betas, standard errors, and p-values) are given for each previously reported association with the TSH levels in European Americans. CAF for rs4704397 is the mean CAF for the combined cohorts described in Taylor et al. For SNPs not directly genotyped in this study, the proxy in highest linkage disequilibrium in 1000 Genomes CEU reference panel was identified. Results of adjusted (age, sex, body mass index, and principal component 1) tests of association are given for each previously reported SNP or its proxy in this European American dataset (n=4,501).

In African Americans, 5/24 (25%) SNPs previously associated with TSH levels in European-descent populations generalized at a liberal significance threshold of $p<0.05$ and a consistent direction of effect (Appendix M). *PDE8B* rs2046045, a proxy for rs6885099 (1000 Genomes CEU $r^2=1.00$, YRI $r^2=0.945$), was associated with serum TSH levels in African Americans ($\beta = -0.09$, $p=0.03$) (Appendix M). *NFIA* rs334713, a proxy for rs334699 (1000 Genomes CEU $r^2=1.00$, YRI $r^2=0.774$), was associated with serum TSH levels in eMERGE African Americans ($p=1.50 \times 10^{-3}$) with a similar effect size ($\beta=-0.17$) as previously-reported European-descent populations. Notably, the coded allele frequency of this SNP was greater in African Americans (coded allele frequency = 0.17) compared with either eMERGE European Americans (rs334708 coded allele frequency=0.08)(Table 20) or the previously-reported European descent population (0.05) (Appendix M). Intronic *ABO* rs657152 was significant at $p=0.03$, and the magnitude and direction of effect were similar to previously published European American data (Appendix M). *VEGFA* rs11755845 was significant at $p=0.01$ (Appendix M) with an effect size nearly double that of the previously reported result in European Americans (Appendix M). SNP rs13020935 upstream of *IGFBP5*, a proxy for rs13015993 ($r^2=1.00$), was significant at $p=1.82 \times 10^{-4}$ (Appendix M).

SNPs previously associated with thyroid disease

Next, we investigated SNPs that had previously been associated with a thyroid disease phenotype, specifically: hypothyroidism, thyroid cancer, and Graves disease (Eriksson et al. 2012; Chu et al. 2011; Gudmundsson et al. 2009), since variation in TSH levels may indicate thyroid disease. Six SNPs in the *FOXE1* region, including rs925489, generalized to euthyroid European American subjects (Appendix N). An additional SNP in *FOXE1*, rs965513, previously associated with hypothyroidism (Eriksson et al. 2012; Denny et al. 2011), generalized to serum TSH levels in European Americans ($p=1.09 \times 10^{-6}$, $\beta=-0.05$) (Appendix N). *FOXE1* rs1877432, previously associated with hypothyroidism, generalized to serum TSH levels in African Americans ($p=9.73 \times 10^{-3}$, $\beta=0.11$) (Appendix O). *RHOH/CHRNA9* rs6832151, previously associated with Grave's Disease, generalized to serum TSH levels in African Americans ($p=0.01$, $\beta=-0.10$) (Appendix O). None of the SNPs previously associated with thyroid cancer (Gudmundsson et al. 2009) were associated with serum TSH levels in either European Americans or African Americans at a liberal significance threshold of $p<0.05$ (Appendix N, Appendix O). Broadly, we found little evidence of association with serum TSH levels for SNPs, apart from *FOXE1*, that have been associated with other thyroid-related phenotypes.

Interaction with BMI

Increased BMI is significantly associated with TSH levels and changes in BMI can be a symptom of thyroid disease, with hypothyroid persons gaining weight and hyperthyroid persons losing weight (Knudsen et al. 2005). We observed that the addition of BMI into the linear regression model yielded more significant p-values for the SNPs in *PDE8B* and others, and the results from the stratified analyses differed within each race/ethnicity (Appendix P, Appendix Q). Therefore, we performed formal tests of interaction between BMI and all SNPs (n=118) with $p<1x10^{-4}$ from the age, sex, PC1, and BMI adjusted model in European Americans and considered evidence for an interaction at $p<0.05$. Three SNPs met our significance threshold in European Americans for an interaction with BMI: *NFIA* rs10489909, *NRG1* rs2466067 and rs4298457. An additional *NRG1* SNP was just outside the $p<0.05$ significance threshold for the interaction: rs10954859 (Table 21). The *NRG1* SNPs are in moderate-to-high LD with each other ($r^2>0.70$). We compared median TSH levels by BMI category for each genotype by SNP and observed lower median TSH levels for individuals with the AA genotype for rs10489909 who were of normal BMI than compared to individuals with overweight BMI ($p<0.005$) (Figure 7). We observed similar trends for rs2466067 (CC genotype), rs10954859 (GG genotype), and rs4298457 (GG genotype) ($p<0.05$) which suggests serum TSH levels may be attenuated based on BMI for these homozygous genotypes (Figure 7).

Table 21. Body mass index as a modifier of serum TSH levels genetic associations.

POPULATION	SNP	GENE/REGION	MODIFIER	BETA (SE SNPxBMI)	P (SNPxBMI)
European American	rs10489909	*NFIA*	BMI	0.01(0.004)	6.21E-03
European American	rs2466067	*NRG1*	BMI	0.004(0.002)	0.040
European American	rs4298457	*NRG1*	BMI	0.004(0.002)	0.047
European American	rs10954859	*NRG1*	BMI	0.004(0.002)	0.050
African American	rs6728613	*MYT1L*	BMI	-0.016(0.005)	2.28E-03
African American	rs4073401	*MYT1L*	BMI	-0.016(0.005)	2.28E-03
African American	rs10518306	LOC285419	BMI	-0.026(0.011)	0.020
African American	rs6062344	*TCEA2*	BMI	-0.010(0.005)	0.043
African American	rs6090040	*TCEA2*	BMI	-0.009(0.005)	0.047

Interaction analyses were performed using the SNPs with $p<1x10^{-4}$ significance levels in the model adjusted for age, sex, principal component 1 (PC1), and body mass index (BMI) in European Americans (n=4,501) and African Americans (n=351) in eMERGE serum TSH levels study. The model was stratified by race/ethnicity and by normal/overweight BMI (normal: BMI 18-24.9; overweight: BMI 25+). We considered a SNPxBMI interaction significant at a threshold of $p<0.05$. Displayed are significant interaction results at $p=0.05$.

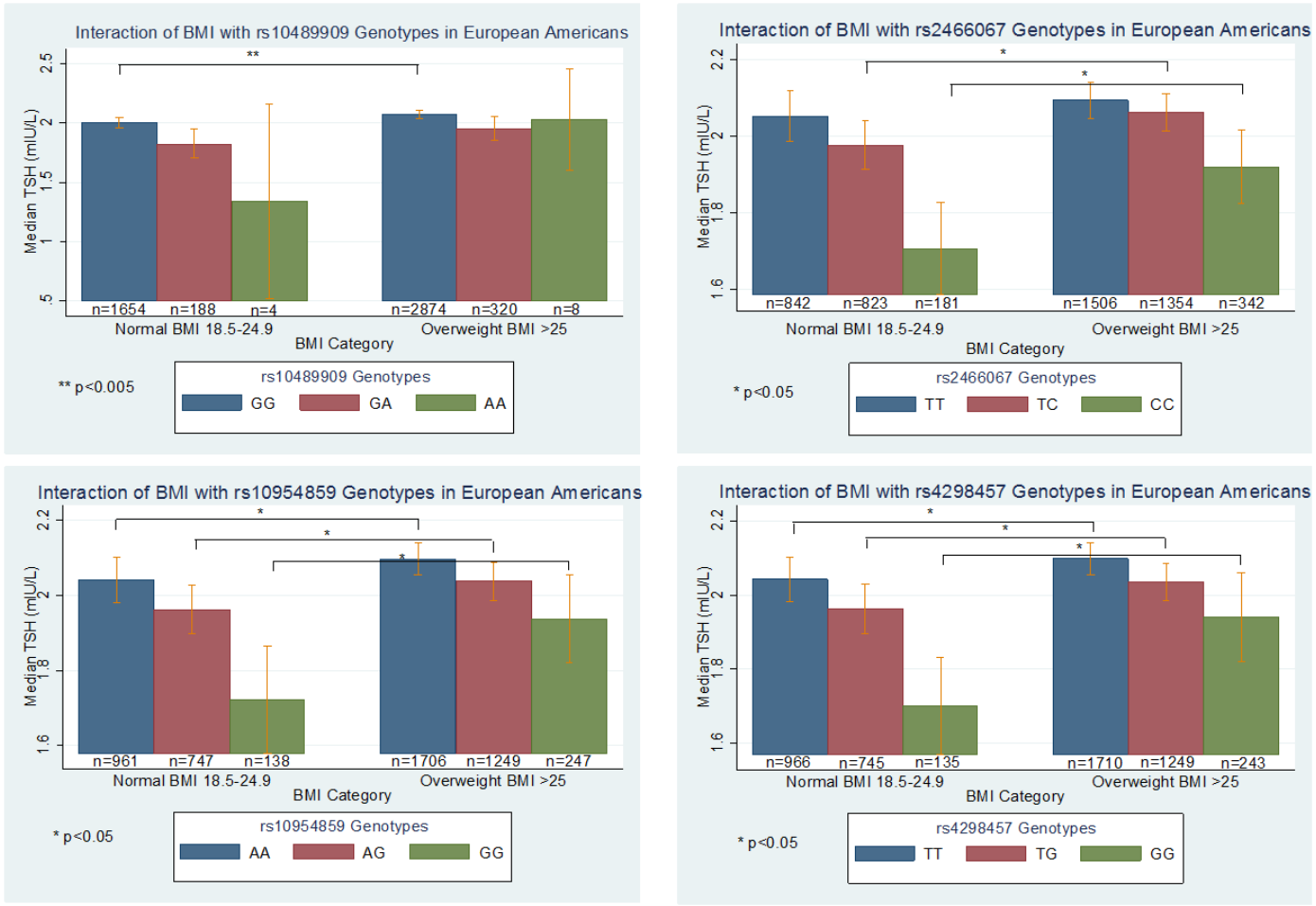


Figure 7. BMI as a modifier of serum TSH levels in eMERGE European Americans.

Interaction analyses were performed using the SNPs with $p<1x10^{-04}$ significance levels in the model adjusted for age, sex, PC1, and body mass index in European Americans (n=4,501). For each significant ($p<0.05$) interaction term, the model was then stratified by normal/overweight BMI (normal BMI =18-24.9; overweight BMI≥25). We considered a SNPxBMI interaction significant at a threshold of $p<0.05$. Shown are p-values from Wilcoxon rank-sum test comparing median TSH values between BMI categories at each genotype.

We also performed tests of interaction in African Americans for BMI and the 87 most significant SNPs ($p<1x10^{-4}$ from the age, sex, PC1, and BMI adjusted model). We observed five SNPs at the $p<0.05$ significance threshold (Table 21, Appendix R). *MYT1L* rs6728613 and rs4073401 are in perfect LD with each other ($r^2=1.00$) and were the most significant in this interaction analysis ($p=2.28x10^{-3}$) (Table 21, Appendix R). While other interaction terms were significant in the African American sample, small sample sizes and low counts made comparisons across genotypes and BMI categories difficult to interpret (Appendix R).

Summary

The eMERGE Network was established in 2007 to determine whether electronic medical records could be used to identify disease susceptibility in diverse patient populations for complex traits/diseases. At each study site, DNA linked to an EMR was genotyped for a GWAS for specific complex diseases (e.g., type II diabetes) and medically relevant quantitative traits (e.g., cardiac conduction). A recent eMERGE Network GWAS demonstrated that these study-specific genotype data can be "reused" for additional GWAS for binary outcomes (hypothyroidism) extracted from the EMR (Denny et al.2011). As an extension of this exercise, we performed a GWAS for an additional medically relevant quantitative trait: thyroid stimulating hormone (TSH) levels, in 4,501 European American and 351 African American euthyroid individuals.

Several studies have shown associations between TSH levels and *PDE8B* (briefly: (Arnaud-Lopez et al. 2008; Medici et al. 2011; Taylor et al. 2011)). *PDE8B* is a phosphodiesterase gene that encodes a cAMP-specific protein expressed in thyroid tissue (Horvath et al. 2010). *PDE8B* upregulates cAMP through interaction with the TSH receptor on thyroid cells (Arnaud-Lopez et al. 2008; Horvath et al. 2010). In this study, we have replicated the results recently obtained by several groups finding association of TSH levels and several SNPs in the *PDE8B* region in European Americans (Medici et al. 2011; Taylor et al. 2011). Variants in *PDE8B* were the only SNPs in this analysis to reach genome-wide significance in European Americans after accounting for multiple testing. In African Americans, rs2046045 (in high/perfect LD with rs6885099 and rs4704397) was nominally significant.

The *FOXE1* region was not as strongly associated with TSH levels as *PDE8B* in European Americans, a result similar to that obtained by Medici *et al* (Medici et al. 2011). *FOXE1* encodes a thyroid transcription factor with a characteristic forkhead motif believed to be important in thyroid

morphogenesis (Cuesta, Zaret, and Santisteban 2007; De and Di 2004). Mutations in *FOXE1* have been implicated in hypothyroidism (Eriksson et al. 2012; Gudmundsson et al. 2009; Denny et al. 2011) and thyroid cancer (Tomaz et al. 2012; Landa et al. 2009). No SNPs in *FOXE1* reached genome-wide significance in this study, though several were associated at the 10^{-6} threshold in European Americans.

Given the relationship between TSH levels and specific clinical outcomes, we hypothesized that serum TSH levels would also be associated with SNPs previously associated with hypothyroidism, Grave's Disease, or thyroid cancer by GWAS or candidate gene studies (Chu et al. 2011; Gudmundsson et al. 2009; Eriksson et al. 2012). Patients with these disorders exhibit abnormal TSH levels and there is a strong autoimmune component to the diseases. No SNPs in previously identified gene regions (*CTLA-4, TSHR, TTF1, HLA,* and *PTPN22*) were significantly associated with TSH levels in either European Americans or African Americans from the eMERGE Network (Appendix N, Appendix O).

Obesity (BMI >30) has been implicated in higher TSH levels and change in an individual's set point (Marzullo et al. 2010; De et al. 2007). We performed additional analyses adjusting for age, sex, PC1, and BMI in both the European American and African American cohorts and stratified analyses by BMI (normal versus overweight). In the European Americans, adjusting for BMI did not appreciably modify the results, though the results in both *PDE8B* and *FOXE1* were more highly significant (Appendix P). These results led us to consider potential SNPxBMI interactions. After performing tests of association for an interaction in the most significant results from the primary analysis, we identified two loci with SNPxBMI interactions in European Americans: *NFIA* and *NRG1*. *NFIA*, a transcription factor, has not previously been associated with thyroid-related traits. *NRG1* encodes neuregulin, a signaling protein recently identified in a study to be associated with thyroid cancer, potentially mediated by regulation of TSH levels (Gudmundsson et al. 2009). Neuregulin is expressed in papillary thyroid carcinomas and has been found to regulate cell proliferation in a rat thyroid cell model (Breuleux 2007). Further studies on the role *NRG1* may play in regulating TSH levels are warranted. In the African American subjects, significant interactions at a liberal threshold ($p<0.05$) were identified, but small sample sizes and low genotype counts per BMI category made comparisons across groups difficult.

We compared results from the African Americans to those of the European Americans in our study and observed several differences. While several SNPs in *PDE8B* reached genome-wide significance in European Americans, none were significant in African Americans, and only two *PDE8B* variants identified in previous GWAS generalized to this population at a liberal significance threshold

of $p<0.05$. Of the 32 most significant SNPs in European Americans, 21 had the same direction of effect and similar effect sizes in African Americans.

A major limitation of this study is sample size. Among both populations, we exclude individuals in eMERGE with an abnormal TSH level given this study sought to identify genetic determinants of the normal distribution for TSH levels. Despite excluding individuals with abnormal TSH values, the mean (standard deviation) observed here for European Americans [1.90 (0.93)] was well within the range of previous TSH level genetic association studies: 1.5 (0.80) to 2.7 (4.1) μIU/mL (Porcu et al. 2013). The addition of the few individuals with abnormal TSH levels would unlikely increase statistical power enough to detect additional genome-wide associations nor substantially impact the overall trait distribution. In comparison, the African American sample size was very small which impacted our ability to generalize previous findings to this population. In eMERGE African Americans, we were only adequately powered (>80%) for one test of association: *PDE8B* rs4704397. This SNP was not directly genotyped in the eMERGE African American dataset, but is in very high LD with genotyped rs2046045 in the 1000 Genomes CEU panel ($r^2=0.94$), but not with the 1000 Genomes YRI panel ($r^2=0.49$). The small sample size coupled with lower linkage disequilibrium resulted in underpowered tests of association for the African American dataset.

We also observed striking differences in minor allele frequencies (MAF) between European Americans and African Americans that may have impacted our ability to replicate and generalize previously associated variants. In European Americans, most of the minor allele frequencies were comparable to those in previously published studies (Appendix S), and we were adequately powered (80%) to replicate 18/25 SNPs previously associated with serum TSH levels at a liberal significance threshold of 0.05 (Appendix S). Of the 18 properly powered tests of association, all of these SNPs replicated in the eMERGE European American dataset. We further considered that population differentiation may have prevented us from generalizing known variants to the eMERGE African Americans. Between populations, differences in environmental pressures can lead to differential changes in allele frequencies. Calculating F_{ST} by both the Weir/Cockerham(Weir and Cockerham 1984) and Wright(Wright 1965) methods using PLATO((The Ritchie Lab 2013)), we found no evidence corroborating this hypothesis (Table 22).

Table 22. F_{ST} Calculations for European and African Americans in eMERGE TSH analysis.

		F_{ST}	
Gene	**SNP**	**Weir/Cockerham**	**Wright**
PDE8B	rs2046045	0.03	0.00
CAPZB	rs10799824	0.00	0.00
VEGFA	rs9472138	0.00	0.00
VEGFA	rs11755845	0.01	0.11
NRG1	rs7825175	0.00	0.00
ABO	rs657152	0.00	0.00
GLIS3	rs1571583	0.00	0.00
PRDM11	rs17723470	0.00	0.00
MBIP	rs1537424	0.03	0.00
ITPK1	rs11624776	0.00	0.00
SOX9	rs9915657	0.00	0.00
INSR	rs4804416	0.01	0.00

Shown are F_{ST} calculations, rounded to the nearest hundredth, for SNPs previously associated with serum TSH levels in European populations and the F_{ST} at each locus for eMERGE European Americans and African Americans. F_{ST} was calculated with PLATO software for both the Weir/Cockerham and Wright methods. Gene shown is the gene closest to the SNP presented.

This study further demonstrates the feasibility of using genotypes linked to EMRs to perform secondary analyses for quantitative traits in complex diseases in diverse populations (Crosslin et al. 2013; Ding et al. 2012). We identified SNPs associated with serum TSH levels and replicated findings from earlier GWAS for TSH levels and thyroid-related traits to the eMERGE European American euthyroid population. We further suggest BMI may modify genetic associations with serum TSH levels. Consistent with other reports, we found few associations with SNPs associated with serum TSH levels that have effects on other thyroid-related traits/diseases. Importantly, we identified suggestive associations with biologically plausible SNPs and generalized several SNPs from previous GWAS to the eMERGE African American euthyroid population, suggesting additional studies in diverse populations are warranted.

CHAPTER VI

IMPLEMENTING PERSONALIZED MEDICINE: EVIDENCE AND ETHICS[4,5]

Introduction

Personalized medicine (PM), defined as individualization of clinical care based, in part, on the genomic background of an individual, (Cornetta and Brown 2013), has more generally come to mean the use of genetic/genomic data to inform clinical care (decision-making for prevention, diagnosis, and treatment) for individual patients (National Human Genome Research Institute 2014b). Genetics and genomics has been used to identify the etiology of unknown genetic conditions (Need et al. 2012) and additional variants responsible for known disorders (Lupski et al. 2010), determine a tumor's susceptibility to chemotherapies (Tessari, Palmieri, and Di 2013), and guide dosage requirements for medication (Scott et al. 2013). Given the public health impact of complex diseases such as type 2 diabetes (T2D), cardiovascular disease (CVD), and asthma, which have genetic and environmental components, leveraging PM to improve health outcomes in complex diseases offers much potential. PM may be used for complex disease to predict which individuals are at greatest risk for disease development based on their genetic composition and environmental exposures, allowing for pre-symptomatic intervention, or to inform pharmacologic therapy choices once a disorder has manifested. However, there are numerous challenges to overcome for PM to be successful for common, complex diseases. The clinical validity, utility and added benefit attributable to including genetic and genomic information in clinical care for complex disorders has not been adequately assessed. The impact on the health care system, including integration of these data into electronic medical records, data access, privacy, and physician decision support, remain important issues for multiple stakeholders. Addressing public understanding of genomics in the context of healthcare is necessary to avoid

[4] Adapted from: Malinowski J & Clayton EW. From pharmacogenetics and cancer to common, complex diseases: are we ready for precision medicine? (In preparation).

[5] Adapted from: Malinowski J & Naylor H. A rapid evidence review for the inclusion of genetic data in clinical care for a common, complex disease. (In preparation).

genetic-based discrimination, genetic determinism, and conflation of genetic ancestry with complex social constructs of race/ethnicity. This chapter will describe the scientific, systemic, and social barriers to successful implementation of PM for complex disease in the clinical setting and consider how these challenges have been addressed for pharmacogenetics and cancer treatment.

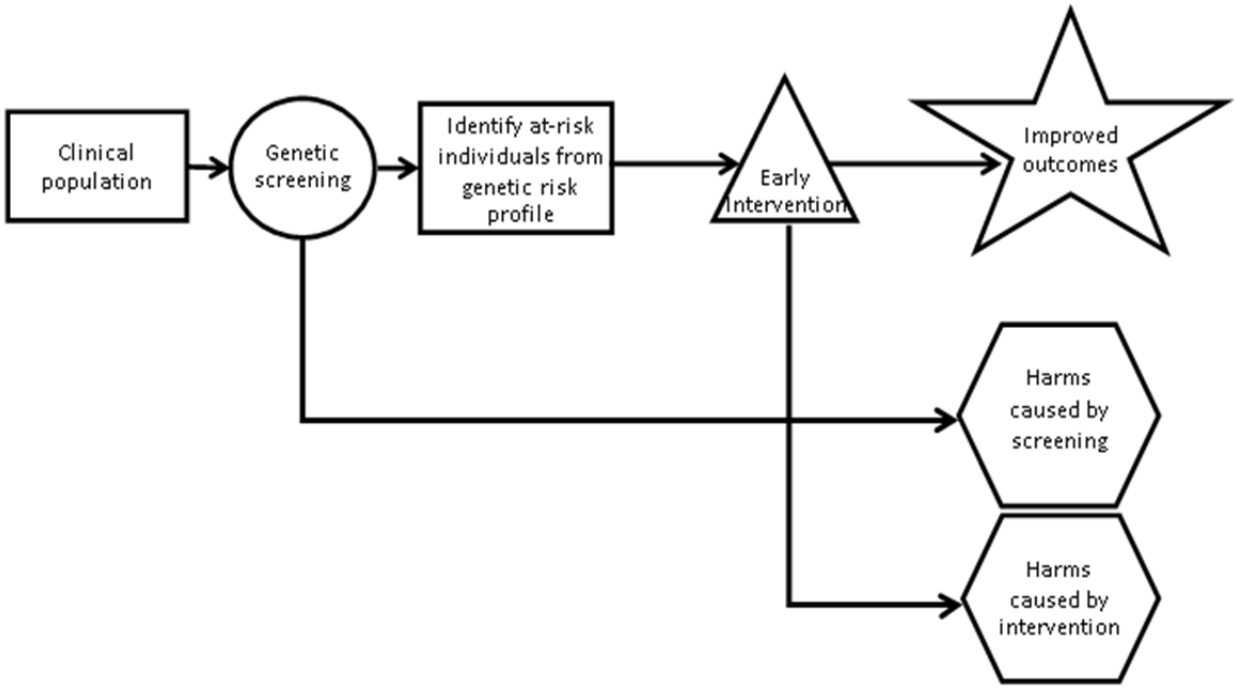


Figure 8. Analytic framework for personalized medicine implementation for complex diseases.

Analytic framework showing how personalized medicine might be used to screen asymptomatic individuals to identify at-risk individuals, allowing for early intervention to prevent disease, leading to improved health outcomes, but with risks from harms caused by both screening and the intervention.

Scientific issues

Challenges to implementation of PM for common, complex diseases in the clinical setting include a deficit of data on the analytic validity, clinical validity, and clinical utility of genetic tests for risk prediction of complex diseases. Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex traits and common diseases but have not generally provided the data necessary to ascertain their benefit in a clinical setting.

Analytic validity

Analytic validity refers to the ability of a genomic test to determine the presence or absence of a particular variant (Table 23) (National Cancer Institute at the National Institutes of Health 2012a). Accuracy and reliability are key measures of analytic validity. Accuracy refers to the ability of a test to measure a specific variant correctly; reliability refers to the ability of a test to yield the same answer when the test is repeated. Despite the importance of this information, little has been published in peer-reviewed literature documenting the analytic validity of genetic tests (Sun et al. 2011). For example, a 2007 review of the literature assessed genetic testing of cytochrome p450 polymorphisms in adults with depression to guide treatment decisions (Thakur et al. 2007). This systematic review included MEDLINE and other databases, such as FDA documents, but found few of the studies compared their results to the "gold standard" of DNA sequencing or other methods such as polymerase chain reaction and restriction fragment length polymorphism or allele specific PCR (Thakur et al. 2007). Of the thirty-seven studies which met their inclusion threshold, nine reports compared clinical genotyping of *CYP2D6* SNPs to a reference standard, but only two studies compared to DNA sequencing (Thakur et al. 2007). This lack of data on the analytic validity of genetic tests has not since been adequately addressed.

Table 23. Glossary of selected terms.

Analytic validity	The ability of a laboratory test to accurately and reliably measure the property it is designed to measure(National Cancer Institute at the National Institutes of Health2012a).
Clinical validity	The accuracy with which a test predicts the presence or absence of a clinical condition or predisposition(National Cancer Institute at the National Institutes of Health2012c).
Clinical utility	The usefulness of the test and value of the information to clinical practice(National Cancer Institute at the National Institutes of Health2012b).
Positive predictive value (PPV)	The proportion of positive results of a given test that is truly positive(Gordis2009).
Negative predictive value (NPV)	The proportion of negative results of a given test that is truly negative(Gordis2009).
Quality-adjusted life years (QALYs)	A measure of disease burden that takes into account quality of and length of time lived(Hyder and Morrow2006).
Disability-adjusted life years (DALYs)	A measure of disease burden that is expressed by the length of time that is lost due to death or disability(Hyder and Morrow2006).
Health-adjusted life expectancy (HALE)	A measure of life expectancy at birth given healthy and sick health states(Hyder and Morrow2006).
Genotype	The genetic constitution of an individual, collectively at all loci or at a single locus(Nussbaum, McInnes, and Willard2007).

The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group was established in 2005 to extend the Analytic Validity, Clinical Validity, Clinical Utility, and Ethical, Legal, & Social Implications (ACCE) Project, a pilot study funded by the Centers for Disease Control and Prevention. The goal of EGAPP is to provide evidence-based evaluation of genomic data for use in clinical practice through a systematic review process, similar to the comprehensive evaluations completed by the US Preventative Services Task Force (USPSTF) for other clinical services (EGAPP Working Group 2014). The EGAPP evaluations methodically review the peer-reviewed literature and other sources (e.g., industry white pages, government documents) to identify the analytic and clinical validity of genetic tests and the clinical utility of these tests in a medical setting. The rigor of the literature and potential bias are also considered (Teutsch et al. 2009). A recent article outlining the challenges and successes of the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group noted the dearth of publicly available data upon which to establish analytic validity (EGAPP Working Group 2014). With a lack of substantive data, the inherent trust in the

analytic validity of genetic tests can lead to erroneous interpretations in the clinical setting, and may lead to negative patient outcomes (Baggerly and Coombes 2011).

Clinical validity

Clinical validity refers to the accuracy with which a test predicts the presence or absence of a clinical condition/disorder (Table 23). For genetic tests, it is the predictive value of the genetic variant for the disorder: what is the probability that someone with a particular variant has the disorder? Several scientific factors determine the clinical validity of genetic information in personalized medicine initiatives. Genome-wide association studies (GWAS) have built upon earlier genetic studies, including linkage analysis and family studies, and uncovered thousands of loci associated with hundreds of phenotypes. Recently, these GWAS findings have been applied to determine risk for hundreds of complex traits and disorders by direct-to-consumer genetic testing companies, such as 23&Me. However, these associations may not be the causal variants responsible for expression of the phenotype or disease, nor may they identify the biologic mechanism responsible for increased disease risk (EGAPP Working Group 2014). This complicates building an accurate and complete risk model for the disease using GWAS findings. Additionally, complex diseases result from the interplay of genetics and environment. The relative contributions each makes to the development of disease is specific for each disorder and have yet to be fully elucidated, affecting the clinical validity by varying the predictive ability of the genetic information (Table 23). Genetic heterogeneity, where multiple variants may cause a disorder, and incomplete penetrance, where there is variation in the expressivity of the disease, further undermine the predictive abilities of genetic tests. Furthermore, intervention in the context of personalized medicine takes place at the individual level, while most genotype-phenotype associations arise from population studies. This changes the role of genetics from deterministic to probablistic, leading to a measure of uncertainty in the relationship between the genetic information and disease risk.

Positive predictive values (PPV), the area under the receiver-operating characteristic curve (AUC), and clinical validity (Table 23) are important, and to-date, mostly missing pieces of data, essential to demonstrate the utility of genetic information in implementation of personalized medicine in the clinical setting. The PPV and other metrics such as the sensitivity, specificity, and negative predictive value (NPV) can be calculated using a 2x2 table (Table 24). For example, the sensitivity of a test is measured by the ratio of true positives and all those who are disease positive (a/a+c); the

specificity is the ratio of true negatives and all those who test negative (d/b+d) (Table 24). The PPV is the ratio of individuals with the genetic variant who have the disorder and the total number of individuals who test positive (including those without the disease) (a/a+b). Consider two scenarios: in the first, the disorder is rare with a 0.5% population prevalence (Appendix T); and in the second, the disorder is more common, with a population prevalence of 8.3%, similar to the prevalence in the general public of type 2 diabetes (T2D)(Centers for Disease Control and Prevention 2012) (Appendix U). If we assume the tests we perform have high sensitivity and specificity (99% each), the PPV of the test will depend on the population prevalence of the disorder. In the first example, the low population prevalence results in a PPV = 0.33, while the increased prevalence in the second example results in a higher PPV = 0.92. This strategy for calculating the PPV of a test for a particular disorder works well when the test provides a dichotomous outcome. For some disorders, such as Huntington's disease, a genetic test can yield a true positive or negative outcome and calculating the PPV of a test for the disorder can be done using a 2x2 table. A PCR-based assay for Huntington's disease has a PPV of close to 1.00 for CAG repeats ≥40 (Saft, Leavitt, and Epplen 2014).

Table 24. Calculating specificity, sensitivity, positive and negative predictive values for genetic tests using a 2x2 table.

Variant	Disease (+)	(-)	Total
(+)	True Positives (a)	False Positives (b)	a+b
(-)	False Negatives (c)	True Negatives (d)	c+d
Total	a+c	b+d	a+b+c+d
	PPV=a/(a+b)	**NPV=d/(c+d)**	

Abbreviations: positive predictive value (PPV), negative predictive value (NPV).

For complex diseases, such as T2D or cardiovascular disease (CVD), as lifetime risk and/or population prevalence increases, the upper limit on the predictive capacity of genetic data may

decrease (Dreyfuss et al. 2012). This is unsurprising considering the other factors that come into play with common, complex diseases: the relative contribution of the environment to the disorder, genetic heterogeneity, incomplete penetrance, complex inheritance patterns, gene-environment interactions, and variable expressivity. Therefore, calculating the PPV of a genetic test for a common, complex disorder using the traditional 2x2 tables may not be the most appropriate method (Janssens et al.2006). The area under the receiver operator characteristic (ROC) curve (AUC) yields the accuracy of a continuous test to discriminate those who will develop the disease from those who will not and may be used for genetic tests as a measure of clinical validity (Wray et al. 2010). The measure of a test that perfectly discriminates between those who will develop disease and those who will not will have an AUC = 1.00, while a test that performs no better than random chance will have an AUC = 0.50 (Figure 9) (Wray et al. 2010). Calculating the maximum AUC to determine the clinical validity of a given genetic test for a common, complex disorder requires the genomic architecture of a disease be understood and the genetic variance completely explained by the variants tested. Complex genomic architecture that is unaccounted for will decrease the AUC (Wray et al. 2010). For any two disorders with the same population prevalence, the AUC will be higher (and the genetic test more predictive of disease status) for the disease with the greater genetic contribution relative to the environmental component (Wray et al. 2010). Similarly, for any two disorders with the same heritability measurement, the AUC will be higher for the disease with the lower population prevalence (Wray et al. 2010). Thus far, the variance explained by genetic associations identified through GWAS is small for common diseases with complex genetic architecture and often undetermined environmental component. Therefore, the effective AUC is likely to be much lower than the maximum AUC for any genetic test for common, complex disorders.

This suggests a potential trade-off in the benefit of genetic testing for complex disease in a clinical setting. Screening all patients who come to a clinic or medical center for a complex disease with moderate-to-high prevalence and complex genetic and environmental architecture will reduce the predictive value of an individual genetic test. In the context of PM used to predict disease risk, institution-wide screening would result in many false positives and the value of the genetic information would differ based on the prevalence of the disease considered. Evaluating only patients who have already been identified as 'at risk' is not ideal, either. This method would lead to decreased negative predictive value (NPV, Table 23) as initial risk scores fail to capture 100% of the at-risk population. Additionally, this method would reduce some of the potential benefit ascribed to PM:

namely, that PM could allow you to identify at-risk individuals *prior* to the development of risk factors/symptoms.

Clinical utility: interventions and patient behavior

The clinical utility of a test broadly refers to the ability of the test (screening or diagnostic) to inform clinical decision making, generally in comparison to the current test or case management (Grosse and Khoury 2006; Bossuyt et al. 2012). For genetic tests used as a screening mechanism to identify at-risk individuals for common diseases, there must be recognized interventions available to ameliorate disease risk. These interventions may be clinical in nature, such as more frequent lab tests, or behavioral, such as smoking cessation. When the genetic test is used as a diagnostic tool, such as in cancer treatments or pharmacogenomics, the intervention is the alternate treatment. For both screening and diagnostic genetic tests, comparing the outcome from the standard treatment absent the genetic

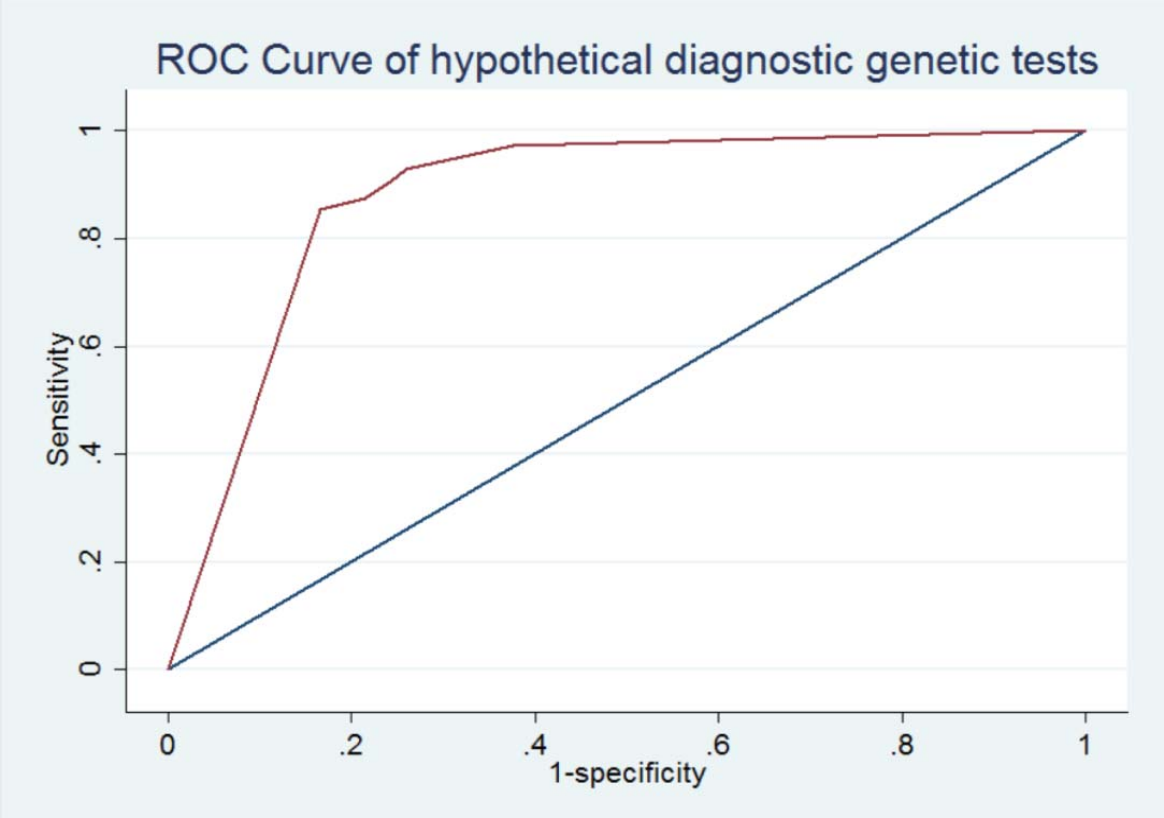


Figure 9. Receiver-operator characteristic curve of hypothetical genetic tests for complex diseases.

Area under the receiver-operator characteristic (ROC) curve (AUC) to determine accuracy of genetic tests. Blue line (bottom) represents an uninformative genetic test with AUC=0.50. Red line (top) represents an informative genetic test with AUC=0.869.

information to the course with the information is essential. A lack of clearly stated, measurable health outcomes makes determining the utility of personalized medicine challenging (Botkin et al. 2010). Borrowing from epidemiology and public health research, health-adjusted life expectancy (HALE), quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) (Table 23) may be of use to researchers addressing personalized medicine outcomes in large populations. Additional quantitative outcomes may include: cost-effectiveness analyses, time to diagnosis of disease from onset of symptoms, number of tests needed to diagnose disease, clinical measurements (e.g., lab values, BMI), symptom severity, and dose of medication required to maintain appropriate lab values. Qualitative data may also contribute meaningfully to health outcomes research in personalized medicine. These may include psychosocial measures such as lifestyle changes and patient/clinician perspectives. These valuable measurements will provide the foundation for clinical utility research in PM; though thousands of genetic variants have been associated with hundreds of clinical traits and diseases, few of them have been evaluated in this context. Electronic health records may facilitate this research by aiding risk model algorithm development, prescription information, and the ability to follow patients over the course of many years. Determining the clinical utility of genetic variants for complex diseases could be accomplished through EGAPP, evidence-based practice centers, or translational scientists. The EGAPP working group has repeatedly determined there is insufficient evidence to recommend genetic testing to improve health outcomes for several diseases; however, they have completed fewer than a dozen studies thus far, and it may be premature to draw broad conclusions from their work (EGAPP Working Group 2014).

Cost-utility analysis (CUA) is one type of cost-effectiveness study that integrates economic costs with health outcomes using QALYs (Phillips et al. 2014). A recent analysis of fifty-nine published CUAs for PM tests found gene expression profile tests for breast cancer to be the most common (Phillips et al. 2014). The majority of tests evaluated by Phillips et al. provided improved health outcomes; however, this was countered by increased costs (Phillips et al. 2014). Less than 25% of the CUAs evaluated demonstrated cost-savings and 8% had increased costs while failing to demonstrate improvements in outcomes (Phillips et al. 2014). An additional challenge to accurately assessing the costs associated with PM programs is that many economic evidence studies rely on statistical modeling with hypothetical cohorts (Lieberthal 2013). A recent review by Lieberthal found most literature on the economics of genomic testing for women with breast cancer relied on modeling with hypothetical cohorts for their analyses (Lieberthal 2013). Cost-effectiveness studies in pharmacogenomics also use

hypothetical cohorts and statistical modeling (Pink et al. 2014; Paulden et al. 2013; You et al. 2012). Utilizing clinical trials data to generate the cost-effectiveness data needed of genetic tests would provide an accurate representation of the costs in real-world scenarios; however, the overall dearth of comparative effectiveness research or clinical utility data for genetic tests performed for risk prediction of common, complex diseases is likely to remain a major barrier to necessary cost-effectiveness studies for PM (Lieberthal 2013; Garber and Tunis 2009).

Despite the lack of data demonstrating clinical utility for inclusion of genetic information in clinical care, several studies have evaluated research participants' lifestyle changes and opinions about the personal utility of genetic information for complex disease risk. Little behavioral change has been observed in a survey of young adults given hypothetical risk scenarios for CVD, T2D, and stroke (Vassy et al. 2013). High-risk genetic results were associated with increased likelihood that individuals would alter their diet and exercise behavior; however, this response was mitigated by poor nutrition and physical activity at baseline (Vassy et al. 2013). Several studies have examined the impact of genetic information of on smoking cessation (Lerman et al. 1997; Audrain et al. 1997; McBride et al. 2002; Ito et al. 2006; Carpenter et al. 2007). Lerman et al. found initial differences in perceived risk of lung cancer and benefits to smoking cessation in smoking participants given risk information about genetic susceptibility to lung cancer compared to participants not given genetic risk information; however, these differences were no longer observed at the two-month follow-up (Lerman et al. 1997). Long-term follow-up in this cohort failed to identify significant differences in actual smoking cessation between the groups, though subjects who had been given their genetic susceptibility were more likely to have attempted quitting (Audrain et al. 1997). More recent studies continue to support a lack of significant changes in long-term smoking behavior attributable to the participant's knowledge of genetic susceptibility to smoking-associated disease (McBride et al. 2002; Carpenter et al. 2007; Ito et al. 2006).

A 2011 study by Roberts et al. examined the perceived clinical utility of *APOE* testing for risk of Alzheimer's disease among the Risk Evaluation and Education for Alzheimer's disease (REVEAL) Study participants (Roberts, Christensen, and Green 2011). Roberts et al. chose Alzheimer's disease as a model for other common, complex disorders, because of similarities in the low predictive value of genetic risk factors identified via GWAS, increased population prevalence of risk alleles compared to rare Mendelian variants, and the availability (at that time) of direct-to-consumer genetic testing which provided a risk assessment for the disorder (Roberts, Christensen, and Green 2011). The authors

identified several purposes driving study participation, including the opportunity to consider long term care insurance, the impact on personal affairs, and the ability to prepare oneself and family emotionally for disease development (Roberts, Christensen, and Green 2011). Despite the lack of medical interventions to reduce Alzheimer's disease risk, participants believed the genetic information to be valuable (Roberts, Christensen, and Green 2011).

Colorectal cancer is a complex disease with both genetic and environmental risk factors, such as lack of physical activity, low vegetable and fruit intake, obese and overweight BMI, and alcohol or tobacco use(Centers for Disease Control and Prevention 2014). Graves et al. evaluated response to genetic SNP testing for CRC risk in male and female primary care patients (Graves et al. 2013). Participants received genetic counseling pre- and post-test; post-test materials included lifetime risk assessments based on their genetic test results, family history, and other risk factors. Subjects were assessed for their emotional distress, comprehension of lifetime risk estimates, lifestyle changes, disclosure of results to family, and contact with physicians or colorectal cancer screenings (Graves et al. 2013). This study found no distress among participants after learning of their genetic test results and limited (28% of study participants) disclosure of the results to physicians (Graves et al. 2013). Participants reported increased physical activity and healthy eating post-test (Graves et al. 2013), though self-report of exercise, diet, and traits like BMI may be subject to bias (Wen and Kowaleski-Jones 2012; Warren et al. 2010).

Others have observed an increase in risk-reducing behaviors following DNA based risk information (Watson et al. 2004; Johnson et al. 2002; Botkin et al. 2003; Schwartz et al. 2002). REVEAL study participants with *APOE* ε4 alleles who are at increased risk for developing Alzheimer's disease were more likely to have indicated lifestyle changes specific to Alzheimer's disease prevention (e.g., diet, exercise) than those without the ε4 alleles (Chao et al. 2008; Vernarelli et al. 2010). These data and others (Bloss et al. 2013; Bunnik, Janssens, and Schermer 2014) suggest individuals may obtain personal utility from learning of their genetic risk for complex disease. Bunnick *et al*. described the personal utility of genetic testing to include reproductive and lifestyle planning and psychological benefit of "knowing" genetic risk (Bunnik, Schermer, and Janssens 2011). A follow up study in the Scripps Genomic Health Initiative group found that participants generally felt no long term distress related to their genetic testing and believed the test to be of high personal utility, though there was little change for those at risk in fat intake or exercise compared to pre-test levels (Bloss et al. 2013).

A 2006 paper that modifies Leventhal's common-sense model (CSM) of self-regulation of health

and illness (Leventhal et al. 1997) may provide a basis for the lack of consistent, observable risk-reducing behaviors when individuals are given genetic risk information (Marteau and Weinman 2006). This self-regulation theory suggests individuals cope with information about health threats based on their pre-existing beliefs and how the new information fits within their belief structure. Extending this framework to a personalized medicine context, patients may decide to act in a clinically meaningful way when provided genetic risk information based on their perception of the health risk, their perception of the likely effectiveness of the suggested behavior change, and their confidence in their ability to perform the behavior (De Wit and Stroebe 2004). The physician and patient perception of the health threat may differ, particularly when the connection between the risk-reducing behavior and the health threat are abstract (Marteau and Weinman 2006). If patients perceive death from cancer to be a greater and more significant threat than death from cardiovascular disease, they may be more willing to modify their behavior to reduce the perceived threat of cancer. This may partly explain the discrepancy among study outcomes.

With complex diseases, the lack of consistent, observable risk-reducing behaviors following genetic testing may also stem from a general inability to integrate multiple pieces of information (French et al. 2002; French et al. 2000). Patients may be less likely to believe that a DNA test can accurately predict disease risk for common diseases with genetic and environmental contributions and/or complex genetic architecture (e.g., genetic heterogeneity, variable expressivity) (Michie et al. 2003; Michie et al. 2002). The weight a patient assigns to the genetic and environmental contribution to disease risk may also play a role in whether or not the patient engages in risk reducing behavior after learning of their genetic risk. If a patient perceives that the genetic component to disease risk outweighs the environmental or modifiable contribution, they may be less likely to engage in risk-reducing behaviors; the CSM suggests this is due to an imbalance between perceived cause of health threat and the related coping procedure (Marteau and Weinman 2006).

For some individuals, belief that both genetics and the environment contributed to the development of familial adenomatous polyposis (FAP) made them less likely to believe a genetic test could accurately predict their risk of FAP (Michie et al. 2002). Comparing individuals who expected to undergo colonoscopy screening for FAP with those who did not plan to, genetic tests were less likely to be perceived as extremely accurate in those who planned to have bowel screenings and those individuals were more likely to attribute the cause of FAP to behavioral factors (Michie et al. 2002). Though both environmental and genetic factors contribute to psychiatric disorders, the stigma

associated with mental illness often prevents affected individuals from seeking professional treatment (Corrigan 2004). The Genes, Disease, and Stigma (GDS) Study and the MacArthur Mental Health Module of the 1996 General Social Survey were analyzed to identify potential associations between participants' beliefs about the role of genetics in mental illness etiology and the perceived effectiveness of treatment for depression and schizophrenia (Phelan, Yang, and Cruz-Rojas 2006). Individuals attributing genetic factors to the cause of the illness were more likely to recommend prescription medication or psychiatric hospitalization for treatment than those who did not believe the illness was caused by genetic factors (Phelan, Yang, and Cruz-Rojas 2006). Perceived effectiveness in treatment did not differ between depression and schizophrenia; however, in the GDS study, attributing genetic factors to the disorder significantly reduced belief in the effectiveness of the treatment (Phelan, Yang, and Cruz-Rojas 2006).

In families with familial hypercholesterolemia, individuals with genetic mutations who received routine clinical diagnosis and genetic testing trended toward stronger belief in the efficacy of cholesterol lowering medication and decreased belief that diet could reduce cholesterol levels (Marteau et al. 2004). However, a recent study evaluating participant trust in genetic risk assessment for T2D reported high levels of trust in the information (Mills, Barry, and Haga 2014). Mills et al. used surveys to assess participant understanding of genetic tests for risk prediction of T2D and attitudes about risk, test results, and method of result delivery (online through testing company website or in person via genetic counselor) (Mills, Barry, and Haga 2014). The majority of participants perceived the benefit of genetic testing for T2D risk to be learning about healthy behaviors that could reduce risk for developing T2D to be very or somewhat important (98.8%) (Mills, Barry, and Haga 2014). Accurate portrayal of the clinical validity and utility of genetic tests will likely improve patient perceptions in the context of assigning disease risk based on genotypes; patient attitudes toward the efficacy of lifestyle-based interventions to minimize disease risk may be more challenging to change. To obtain improved health outcomes with PM for common, complex diseases, it will be necessary to use a multifaceted approach that broadens patients' understanding of complex disease composition, provides concrete examples of disease risk, offers interventions that are acceptable to patients, and educates patients so their perceptions of disease risk and the benefit of risk-reducing behaviors more closely align to physician perspectives.

Added value of genomic information with existing risk models

For complex diseases like CVD or T2D, physicians can use established risk models based on a

combination of variables (e.g., family history, anthropomorphic traits such as body mass index, and lab values) to predict the likelihood of developing the disorder in a future time period. For PM to be widely implemented, clinicians need to demonstrate that adding genomic information to a risk model improves the model's ability to distinguish those who will remain unaffected from those who will become affected, or provides incentive to those who are genetically at risk of developing the disease to make behavioral changes to minimize risk. T2D has an estimated prevalence of 9.3% and is responsible for an estimated $174 billion in direct and indirect costs in the United States (Sheehy, Coursin, and Gabbay 2009). Risk prediction for T2D is not generally a part of routine clinical practice; however, lifestyle interventions have been shown to reduce the development of T2D in high risk individuals (Knowler et al. 2002; Teufel and Ritenbaugh 1998) and the development of complications (Sheehy, Coursin, and Gabbay 2009), supporting the utility of identifying at-risk individuals prior to the development of disease. For healthy individuals, the Cambridge and Framingham risk score algorithms predict risk of developing T2D based on routinely collected clinical data and information, such as parental history of T2D, sex, and body mass index (BMI) (Rahman et al. 2008; Wilson et al. 2007). In a study by Talmud et al., using the prospective Whitehall II cohort, adding the known genetic associations with T2D to the Framingham and Cambridge risk scores did not significantly improve the predictive ability of those models (Talmud et al. 2010). Notably, the predictive ability of the genetic information alone yielded an AUC of 0.54, far below that of the Framingham risk model (AUC=0.78) or Cambridge risk model (AUC=0.72) (Talmud et al. 2010). These data suggest including genetic data in clinical practice for T2D yields limited benefit for the prediction of disease development (Clayton 2009).

Cardiovascular disease is another complex disease with significant morbidity, mortality, and associated economic costs. The Framingham Risk Score (FRS) is a well-known model that includes variables such as age, sex, BMI, smoking status, and cholesterol levels to predict the 10 year risk of developing cardiovascular disease (D'Agostino, Sr. et al. 2008). Similar to the findings in the Talmud et al. paper for T2D, adding genetic variants to the FRS model has not consistently improved discrimination in several studies (Brautbar et al. 2012; Hughes et al. 2012; Talmud et al. 2008). These studies indicate that using genetic variants, which have previously been associated with disease in GWAS, to predict complex disease risk, does not generally prove as predictive as established risk scores based on easily obtainable clinical data. Furthermore, including the genetic component to these risk scores does not significantly improve the predictive ability of the risk score (Mihaescu et al. 2011; Wray et al. 2013). Though the added benefit of genetics in a risk model will likely differ substantially

across complex diseases, it is clear that for some, there is little evidence to suggest that this information is currently clinically useful.

A new study being performed in adult US Air Force primary care patients seeks to determine if health coaching with dissemination of genetic information leads to improved health outcomes for T2D and coronary heart disease (Vorderstrasse et al. 2013). This study employs the modified CSM (Marteau and Weinman 2006) and considers patient response to genetic information using a variety of clinical measures (e.g., BMI, lipid profile) and surveys (Vorderstrasse et al. 2013). Additional studies similar to Vorderstrasse et al., leveraging study techniques employed by social scientists and behavioral psychologists in addition to clinical measurements, would improve the ability of genetic researchers to calculate the added value of genetic information in current disease management.

Evidence review for inclusion of genetic data in clinical care for hypothyroidism

A recent review by the EGAPP working group noted the challenge in implementing thorough reviews with limited resources and time, and the scant evidence for analytic validity, clinical validity, and clinical utility (EGAPP Working Group 2014). To address the lack of clinical validity and utility data, we investigated if a rapid evidence review could be performed in an academic setting to identify analytic evidence that inclusion of genetic data associated with common, complex diseases in clinical care improved health outcomes. Our goal was to establish a first-pass method to evaluate the evidence base for including genetic data in the clinical setting, which could be used to identify gaps in knowledge or to justify more comprehensive evidence reviews (Ganann, Ciliska, and Thomas 2010; Watt et al. 2008).

Until recently, consumers could bypass medical professionals and obtain their genetic data for carrier status and disease risk through direct-to-consumer (DTC) genetic testing providers, e.g., 23&Me. These DTC companies provided consumers' estimated risks of developing more than one hundred disorders (e.g., T2D, CVD, hypothyroidism) and traits (e.g., male pattern baldness, response to the drug clopidogrel) (23&Me 2014a) with a goal of empowering their customers to use genetic data to improve their lives (23&Me 2014b). However, few variants identified through genome-wide association studies (GWAS) have been demonstrated to be actionable clinically as part of a screening mechanism. Candidate gene studies similarly lack accepted, clinically actionable results, though there are exceptions: pre-symptomatic risk assessment using genetic/genomic testing performed for breast cancer (BRCA1/2) (Robson and Offit 2007), screening for Lynch syndrome hereditary colon cancer in

family members of those recently diagnosed with the disease (MLH1, MSH2, MSH6, PMS2) (2009), and long QT interval (KCNQ1, KCNH2, SCN5A) (Napolitano et al. 2005). Well-documented obstacles to expanded implementation of PM are the unknown risks and benefits in the clinical setting, lack of clinical validity and utility data (Teutsch et al. 2009; Palomaki et al. 2009b; Palomaki et al. 2009a; Palomaki et al. 2010; Palomaki et al. 2013), and the challenge of incorporating multiple causes (genetic, environmental, behavioral) in a risk prediction model (Wray et al. 2013).

For this study, we selected a common, complex disease, hypothyroidism and related quantitative trait, serum thyroid stimulating hormone (TSH) levels, both with numerous genetic associations identified through GWAS (see Chapter I). Diagnosis of hypothyroidism involves measuring TSH levels in the blood; elevated TSH levels may indicate hypothyroidism. TSH levels vary across individuals, with higher mean TSH levels observed in non-Hispanic whites than in Mexican Americans or non-Hispanic blacks (Hollowell et al. 2002); however, within an individual, TSH levels are tightly regulated (Hollowell et al. 2002; Arnaud-Lopez et al. 2008; Chiamolera and Wondisford 2009). Environmental and genetic factors influence TSH levels and risk for hypothyroidism. Increased body mass index (BMI), smoking, pregnancy, and physical and emotional stress are all risk factors for elevated TSH levels (Brix et al. 2000; Jorde and Sundsfjord 2006; Nyrnes, Jorde, and Sundsfjord 2006). Several genes have been implicated in hypothyroidism (*FOXE1, PTPN22, VAV3*, and the HLA region) (Eriksson et al. 2012; Denny et al. 2011) and serum TSH levels (*PDE8B, CAPZB, NR3C2*) (Panicker et al. 2010; Rawal et al. 2012; Porcu et al. 2013; Malinowski et al. 2014(in press)). Symptoms for hypothyroidism are generally nonspecific and include fatigue, sensitivity to cold, unexplained weight gain, and depression (Dubbs and Spangler 2014). Left untreated, hypothyroidism may lead to goiters which can affect appearance and breathing/swallowing, heart problems, depression, peripheral neuropathy, infertility, and myxedema (Dubbs and Spangler 2014). Untreated hypothyroidism during pregnancy may lead to adverse maternal and fetal outcomes, including recurrent miscarriage and birth defects, and though screening high-risk pregnant women is advocated, universal screening is controversial (Nathan and Sullivan 2014; Dosiou et al. 2012).

Methods

We began our review by identifying a clinical scenario to guide the process. The clinical scenario focuses on women of childbearing age (18-55) in the general (not high-risk) population being offered a genetic test to assess their risk for developing hypothyroidism. These women would be

asymptomatic, or be unrecognized as having symptoms, of hypothyroidism and therefore, be unlikely to have their TSH levels checked during a routine clinic visit. The proposed clinical utility for testing is to improve health outcomes in women of childbearing age by identifying those who are at risk for developing the disease. Identification of at-risk individuals based on genetic information could lead to 1) regular TSH level testing, potentially minimizing the lag time between development of symptoms and a diagnosis of hypothyroidism and treatment; 2) knowledge of health behaviors, such as smoking, some medication use, and maintaining a healthy BMI, that can be changed to reduce the risk of developing hypothyroidism; 3) reduction of miscarriages and/or birth defects caused by undiagnosed/untreated hypothyroidism prior to conception and during the prenatal period. This review addressed the overarching question: "Does the genotyping of variants previously associated with hypothyroidism in adult, asymptomatic women of reproductive age (18-55), lead to improved health outcomes?"

The ACCE framework used by the CDC EGAPP (Teutsch et al. 2009) working group was used as a reference to perform this rapid review. The disease hypothyroidism was selected as an example of a common, complex disease with numerous genetic associations identified through genome-wide association studies (GWAS) that might be a candidate for personalized medicine initiatives in a clinical setting. The variants selected for this were those used by a direct to consumer genetic testing company to provide their customers risk of developing hypothyroidism (Table 25). The general stages of a systematic evidence review were followed. In collaboration with an information scientist in knowledge management at the Vanderbilt University Medical Center (VUMC) Eskind Biomedical Library (EBL), an overarching research question was developed based on similar questions from previous evidence reviews, generally following the PICOTS (population, intervention, comparator, outcome, timing, setting) method. Key questions were formulated to identify the analytic validity, clinical validity, and clinical utility of genetic tests for those variants. Additional questions were created to identify evidence of improved health outcomes tied to the genetic tests and relevant ethical, legal, or social issues (ELSI) associated with genetic testing for hypothyroidism. A comprehensive PubMed search for citations was developed incorporating Medical Subject Heading (MeSH) terminology and search results exported for abstract and full text reviews. Citations were included in this rapid review if they were included in the PubMed search results as of August 2013 and were written available in English.

Table 25. Variants selected for inclusion in rapid evidence review.

Variant	Gene	GWAS Study	OR	P-value
rs7850258	*FOXE1*	Denny,J.C., 2011	0.74(A)	$3.96x10^{-9}$
rs2476601	*PTPN22*	Eriksson,N., 2012	1.36(A)	$3.9x10^{-13}$
rs3184504	*SH2B3*	Eriksson,N., 2012	0.84(C)	$2.6x10^{-12}$
rs4915077	*VAV3*	Eriksson,N., 2012	1.30(C)	$7.5x10^{-10}$
rs2517532	HLA region	Eriksson,N., 2012	0.86(A)	$1.3x10^{-8}$

Shown are single nucleotide polymorphisms selected for inclusion in a rapid evidence review of genetic data for hypothyroidism in a personalized medicine program in a clinical, but asymptomatic and low-risk, population. Variants were previously used by 23&Me to report risk for developing hypothyroidism to their clients. Listed are the SNP rs number, nearest gene to the SNP, GWAS study first author and year of publication, odds ratio (OR) and allele, and p-value.

Study data were collected and managed using REDCap electronic data capture tools hosted at Vanderbilt University Medical Center (Harris et al. 2009). REDCap (Research Electronic Data Capture) is a "secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources" (Harris et al. 2009). An abstract review was performed independently by two reviewers. After the first 50 search results were reviewed, concordance was evaluated between reviewers to identify key questions that could be misinterpreted and discrepancies were resolved by discussion. Concordance was again evaluated at the halfway point of the abstract review. Citations without abstracts or instances where the reviewers selected "cannot determine" answers for any of the key questions were automatically passed through to the full text review. A full text review was performed on the citations for which full text articles were available to the two reviewers either in print, or electronically. Concordance was evaluated after the first 50 articles had been reviewed to assure interpretation of the questions was consistent. A third reviewer was available for the full text review for discordant reviews. Branching logic was used in REDCap for full text review questions; the first non-affirmative answer for a question provided by the reviewer prompted questions pertaining to the next key question. The survey was ended when no further affirmative answers were provided by the reviewer (Figure 10). Data from the full text review was exported from REDCap (Harris et al. 2009) to Stata (Boston and Sumner 2003).

Studies were excluded from final analysis on the basis of several factors: inclusion of pediatric samples in the study without stratification or statistically adjusting for age, inclusion of male samples

in the study without stratification or statistical adjustment for age, failure to include race/ethnicity information about the sample, small sample size (n<10), or inclusion of thyroid cancer samples in the study without stratification. Additional exclusion criteria included failure to provide effect sizes/odds ratios, positive predictive values (PPV) or area under the receiver-operator characteristics curve (AUC) for the genetic variant. Studies that did not address the key questions in any way were also removed from analysis.

Results

This rapid review was started in August 2013 and completed in March 2014. Specific variants were selected for evaluation as they had been used previously by a direct to consumer genetic testing company to report risk of developing hypothyroidism (23&Me 2014a). A total of 631 citations were obtained from the search query and their abstracts were reviewed using a REDCap database specifically set up for this project; 346 articles were moved forward to full text review (54.9%). No articles were accepted after full text review for further analysis (Table 26). Twenty-five articles were unavailable through open-access publications, EBL electronic subscription, or in-print at EBL (7.2%) (Table 26). One article was eliminated due to inclusion of samples with thyroid cancer and three were removed due to low sample size (n<10) (Table 26). The majority of articles were excluded due to the study's focus not providing information corresponding to one of our key questions (65.0%) (Table 26). Fourteen of the articles fully reviewed provided odds ratios but not positive predictive values or AUC for the genetic variant (4.0%) (Table 27). No articles addressed the ethical, legal, or social issues (ELSI) of genetic testing for hypothyroidism risk in asymptomatic women of childbearing age (18-55). We found no evidence that genetic testing of five SNPs improved health outcomes for this clinical population.

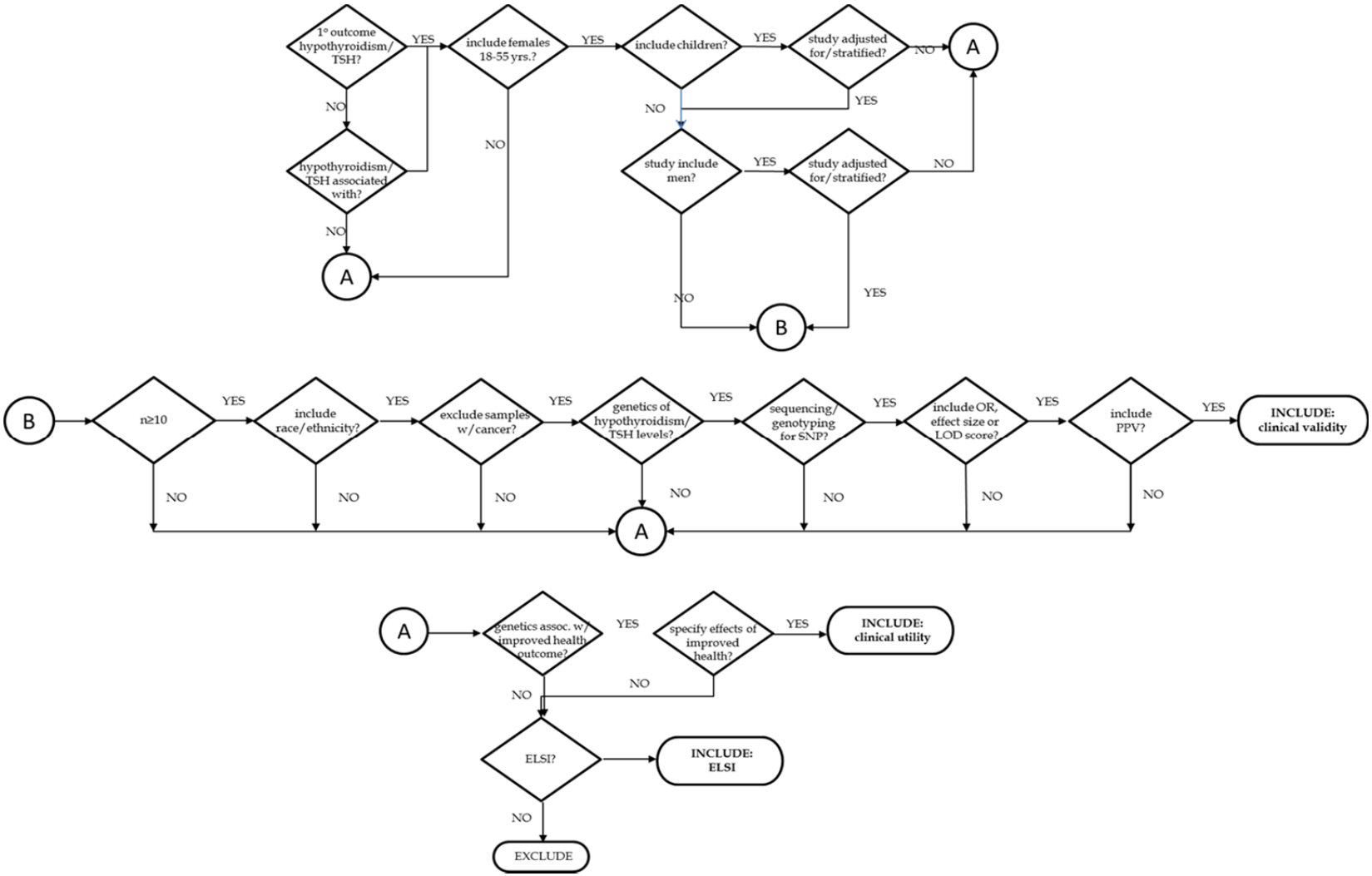


Figure 10. Flowchart for rapid evidence review.

Table 26. Number of studies excluded in full text review.

Reason for removal	Count (n)
Full text unavailable	25
Study included pediatric samples (age ≤17), age not adjusted for or pediatric results reported separately	22
Study included males, sex not adjusted for or male results reported separately	47
Study excluded due to small sample size (n<10)	3
Study excluded due to race/ethnicity not adjusted for or results stratified by race/ethnicity	5
Study excluded because sample included individuals with thyroid cancer, results not stratified by cancer status	1
Study excluded because no effect sizes/odds ratios given	3
Study excluded because no positive predictive values/AUC given	15
Study excluded because study did not address any of the key questions of the rapid evidence review	225

Table 27. List of studies providing ORs/effect sizes for genetic variants associated with hypothyroidism/TSH levels.

First author, year of publication	Type of study	Study outcome	Variants reported	p-value
Petrone, A.; 2001	candidate gene	Hashimoto's thyroiditis	HLA region	<0.05
Terauchi, M.; 2003	candidate gene	Hashimoto's thyroiditis	HLA region	<0.01
Brix, T.H.; 2005	candidate	Hashimoto's thyroiditis	X-inactivation	n.s.
Hansen, P.S.; 2007	candidate gene	TSH levels	*TSHR*	0.007
Panicker, V.; 2008	linkage scan	TSH levels	2q36, 4q32, 9q34	LOD 2.1-3.2
Arnaud-Lopez, L.; 2008	GWAS	TSH levels	*PDE8B*	$1.3x10^{-11}$
Panicker, V.; 2008	candidate gene	TSH levels	*DIO1*	n.s.
Panicker, V.; 2010	GWAS	TSH levels	*CAPZB*	$3.2x10^{-8}$
Volpato, C.B.; 2011	linkage, association	TSH levels	*PDE10A*	LOD 2.66
Denny, J.C.; 2011	GWAS,meta-analysis	hypothyroidism	*FOXE1*	$3.96x10^{-9}$
Eriksson, N.; 2012	GWAS	hypothyroidism	*FOXE1, PTPN22, SH2B3, VAV3*, HLA region	$2.4x10^{-19}$, $2.8x10^{-13}$, $2.6x10^{-12}$, $7.5x10^{-10}$, $1.3x10^{-8}$
Rawal, R., 2012	meta-analysis	TSH levels	*PDE8B, CAPZB, NR3C2*, LOC440389	$2.79x10^{-27}$, $1.54x10^{-8}$, $2.88x10^{-10}$, $5.63x10^{-10}$
Piacentini, S., 2013	candidate gene	hypothyroidism	*GSTO2*	0.009
Porcu, E., 2013	meta-analysis	TSH levels	*	*

Data shown are the studies that provided odds ratios (ORs) or effect sizes and level of significance for an association between the variants reported and either hypothyroidism or serum thyroid stimulating hormone (TSH) levels. *Porcu et al. reported 26 independent associations with TSH. The associations and corresponding p-values can be found in Porcu E. et al., PLoS Genet 2013; 9(2):e1003266 doi: 10.1371/journal/pgen.1003266 Abbreviations: n.s.: not significant; TSH: thyroid stimulating hormone

Summary for rapid evidence review

This rapid review sought to identify analytic evidence that the inclusion of genetics in routine clinical care for common, complex disorders improves health outcomes. We generally followed the ACCE/EGAPP method for an evidence review with several key differences. Only one database was searched (Medline through PubMed) for pertinent articles and the breadth of the search was limited to articles with the clinically relevant patient population to allow for faster applicability in the medical setting. Two reviewers performed the abstract and full text reviews, though an additional reviewer was available as necessary. This review was completed in seven months over two academic semesters, with both reviewers working part-time on this project. Potential bias and the quality of evidence of the studies were not graded. Several potential analytic frameworks were considered for this rapid review: ACCE, Fryback-Thornbury, USPSTF framework for Screening, and the EGAPP framework. A combined ACCE/EGAPP structure was deemed the most appropriate framework to use with genetic studies and has been used successfully for several EGAPP projects, uses an analytic framework to visually address how the genetic test could lead to health outcomes, and incorporates key components of both the USPSTF and Fryback-Thornbury models. Consistent recommendations for laboratories to publish analytic validation studies of their genetic tests in peer-review journals have been ignored and most genetic studies do not provide the analytic evidence (analytic validity, clinical validity, clinical utility) that is needed to evaluate PM policies using current frameworks as described above. If these important data continue to be lacking, it may be necessary to consider alternate methods to evaluate PM initiatives.

Several studies have addressed the lack of appropriate data upon which to base recommendations for or against genetic testing for various health conditions. We found that the analytic evidence supporting genetic testing in asymptomatic women of childbearing age for hypothyroidism risk to be inadequate. Few of the articles that were full text reviewed addressed our key questions. Though our key questions were very specific and failed to identify studies that met the criteria for inclusion, broadening the queries would have undesired consequences due to identification of superfluous studies. Generalizing the population of interest would fail to meet the updated Wilson and Jungner screening criteria proposed by Andermann et al. for a targeted screening population (see Chapter I, Table 1). The five genetic variants of interest were selected based on their use by a DTC genetic testing company to provide risk of developing hypothyroidism to their clients (23&Me 2014a). We considered variants published in other studies as well, based on knowledge of other genetic

associations with hypothyroidism and serum TSH levels, but did not find adequate evidence of clinical validity or utility for any.

Using our review, we identified deficits in the knowledge base that future genetic studies should address. First, the lack of substantive evidence demonstrating analytic and clinical validity must be corrected. As our rapid review considered only articles cited within PubMed, it is possible that a more comprehensive search including other databases, manufacturer's websites, and other grey literature would provide the sensitivity, specificity, and predictive values for these variants for hypothyroidism. However, our results are consistent with others that have documented the same challenges to obtaining this crucial information using more comprehensive approaches. Replication of genetic associations should be undertaken in large, independent cohorts to identify which variants may be false positives and which are robust disease associations; meta-analyses are another method. This is of particular importance as many of the reported disease associations are for variants with relatively small effect sizes and may not be the causal variant affecting the disease process. Though the PPV is one metric used to evaluate the clinical validity of a test, for genetic testing in common, complex diseases, the area under the receiver-operator characteristic curve (AUC) may be a better measurement.

Second, the clinical utility of genetic information in PM for common disorders is unclear and likely varies by disease. Comparing health outcomes using the current screening model to one that includes genetic information is a key part of demonstrating the utility of PM. For the general population, screening practices for hypothyroidism by measuring TSH levels can vary by physician. Though screening of high-risk pregnant women for hypothyroidism is advocated and recent studies have shown universal screening to be cost-effective, universal screening of remains contested. Additional studies with clearly stated and measureable health outcomes will improve researchers' ability to determine the utility of PM. Quantitative measurements such as quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) may be of use in studies evaluating PM initiatives in large populations. The use of electronic health records (EHRs) may make obtaining the necessary clinical and outcomes data easier. Qualitative data, such as surveys of symptom severity or change in patient behavior, can further inform clinical utility studies.

Additionally, it is important to consider how generalizable the results from genetic studies are to populations of diverse ancestry, particularly when there are disease prevalence differences. These disease burden variations may occur from differences in environmental or behavioral exposures to risk factors, may arise through differences in risk allele frequencies across populations, or some

combination of these. Understanding the true population risk is an integral part of implementing PM in the clinical setting.

Despite these negative results, our rapid review methodology serves as an example of deploying an evidence review in an academic setting to systematically identify the essential data required to accurately assess the utility of including genetic data to improve health outcomes for common, complex diseases. We have highlighted the gaps in knowledge preventing researchers from adequately assessing the use of genetic information in the clinical setting for complex disorders. Future genetic studies that include analytic evidence will benefit outcomes research and help policymakers determine the use of genetic data for precision medicine in the clinical setting. Rapid reviews such as this one can be a valuable first-pass to establish the need for a more comprehensive evidence review prior to setting policy.

Ethical, legal, and social issues

Impact on health care system

If PM is to be implemented in the clinical setting for common, complex diseases, the impact on the health system should not be overlooked. Data management, access to data/privacy issues, integration with existing electronic health records (EHR) and physician decision-support mechanisms must be comprehensively investigated. These factors and the costs required to implement such a program should be carefully weighed and cost-benefit analyses performed if widespread adoption of such programs is to expand beyond a few academic medical centers.

Data management

The amount of data generated by whole genome sequencing (WGS), whole exome sequencing (WES), and genotyping arrays with >2 million SNPs is substantial. This leads directly to two issues: the cost and infrastructure required for generating and maintaining the data, and the time and human capital needed to analyze and translate the data for clinical use. Though the technological cost of WES and WGS are decreasing, the cost of data storage has not declined as dramatically (Baker 2010). Each WES run yields several terabytes of raw sequence data and hundreds of GB of stored data and WGS will increase substantially the amount of stored data (Kho et al. 2013). Some medical centers are pursuing technologies such as cloud computing for managing data and data analysis; however, legal regulations such as HIPAA and others may restrict the use of cloud technologies for restricted and

personal health information (Fusaro et al. 2011; Kuo 2011; Schadt et al. 2010; Rosenthal et al. 2010). Similar to the declining cost of WES and WGS, the time required to obtain genetic data has also declined, with WES results in <24 hours and WGS results in a few days (Kho et al. 2013). Analysis of well-characterized variants is not difficult with databases that link genotype and phenotype; however, curation of novel variants is time consuming (Famiglietti et al. 2014). In addition, periodic re-examination of the pathogenicity and function of genetic variants should take place, as demonstrated by Das et al. where variants previously described as pathogenic were reclassified as benign or of unknown significance, and variants previously classified as of unknown significance were upgraded to pathogenic (Das et al. 2014). Though the analysis of genetic data has become more automated, it is unclear to what extent the cost to analyze and translate the data into clinically meaningful information will decline; this is likely to vary substantially by disease and current knowledge of genetic variants.

Data privacy

From the health system's perspective, data security and privacy are not only ethically appropriate, but legally required (Juengst 2014). Controlling data access and ensuring data privacy are key issues that must be addressed for successful implementation of PM, given the sensitive health information contained in patients' EMRs. Current identification of data breaches in EHRs occurs post-event through random audits; new methods are being developed that can predict inappropriate access using historical data (Menon et al. 2014; Fabbri and Lefevre 2013). Maintaining patients' genomic data outside of the EHR with controlled access to the data is one method of minimizing inappropriate access (Hazin et al. 2013). As patients gain access to their EHRs through patient portals, medical centers will need to create policies centered on educating patients of their responsibilities for privacy and security of their health care data (Hazin et al. 2013). A recent mandate (Swain and Patel 2014) that providers allow patients access to their laboratory records may serve as a test for medical centers to navigate data access and privacy requirements when patients play a key role in keeping those data secure. State or local laws may additional limitations on access to genetic data that surpass the requirements of HIPAA and dictate segregation of certain types of health data within EHRs or require additional levels of patient consent before genetic data can be disclosed (National Human Genome Research Institute 2014a). Data security and controlled access will likely remain an integral facet to PM's impact on the health system.

Integration with existing EHRs

Increased use of EHRs can play a substantial role in PM implementation through enhanced communication between patients and clinicians, physician decision-support mechanisms, PM cohort identification, exchange of health information between providers, and by providing the necessary data upon which clinical validity and utility of PM initiatives can be measured (Kohane 2011; Boland et al. 2013; Goldspiel et al. 2014; Middleton et al. 2013; Payne et al. 2013). With the quantity of genomic data generated from WES, WGS, and genotyping arrays, not all data is likely to be inserted into the EHR. Maintaining genomic data in an off-site repository and including only the data that directly informs clinical care is one method to control the burden of data storage, but raises ethical and legal issues (Hazin et al. 2013). A segmented system in which only certain individuals have access to some of the data may pose a challenge to PM practices (Hazin et al. 2013). Determining which clinicians and ancillary health professionals have access to the genomic data and when they may access it, is an important facet to utilizing this information in the clinical setting.

Integrating genomic information and disease risk prediction into EHRs will be an essential component to PM for common, complex diseases and will require both accessible genomic data and physician decision support mechanisms (Hazin et al. 2013). Design and implementation of EHRs for this purpose should involve multiple stakeholders, including patient, physician, and informatics representatives (Hartzler et al. 2013); input from these perspectives may improve patient participation in healthcare choices and physician buy-in to PM initiatives. EHR platforms may require substantial modifications be made so they can be used for PM (Kho et al. 2013). Where genomic data will reside, how it will be accessed through bioinformatics pipelines, and when physician decision-support mechanisms will initiate are key questions for health systems to address. This will require a substantial level of automation for data to be updated routinely and quickly (Schneeweiss 2014). Though adoption of EHR technologies have become more widespread following the Health Information Technology for Economic and Clinical Health (HITECH) Act (Ancker et al. 2013; Blumenthal 2011), technological and logistical issues relating to the management and analysis of genetic data in EHRs remain a barrier for the application of PM for common, complex diseases (Shoenbill et al. 2014).

Physician decision support

Decision-support should be integrated into clinical care without adding to physician burden (Schneeweiss 2014). Alert fatigue, ignoring/overriding decision-support messages, remains a

considerable barrier to PM (Ancker et al. 2014; McCoy et al. 2014); however, successful examples of decision-support systems have emerged in pharmacogenetics (Goldspiel et al. 2014; Pulley et al. 2012; Laerum et al. 2014). More importantly, physicians must feel adequately prepared and comfortable integrating genetic testing into clinical practice (Gray et al. 2014; Hazin et al. 2013; Haga et al. 2012). Decision-support mechanisms with access to corroborating research may help to alleviate physician uncertainty and improve physician confidence integrating genetic data in routine clinical care.

Economic costs

The economic costs involved in PM programs encompass several areas: diagnostic tests and equipment, personnel, computing infrastructure, the costs associated with false-negative and false-positive test results, and the costs associated with risk-reducing interventions. These costs may be balanced or outweighed by improved health outcomes and decreased future health care costs resulting from the PM program. It is currently unclear at what point implementation of PM is a cost-effective strategy (Phillips et al. 2014). Given the public health burden of common, complex diseases such as CVD and T2D, effective PM could improve health care outcomes while reducing health care expenditures. Multiple methods exist for calculating the cost-effectiveness of a test; cost-utility analysis (CUA) is one method that integrates health outcomes using QALYs (Phillips et al. 2014). Current CUAs focus primarily on pharmaceuticals, though CUAs of PM tests for cancer and pharmacogenetics are becoming more common (Phillips et al. 2014).

Impact on society: the patient perspective

Similar to the issues health systems face, the patient perspective should also be considered if PM programs are to be expanded for common, complex diseases. Privacy and data access issues of EHRs from the patient's perspective must be addressed to allay fears of misuse. If PM is expanded as a screening mechanism to identify risk for common disorders, a significant portion of the population will be exposed to complex and abstract genetic concepts that may require patient education or counseling to understand. Public knowledge of genetics and genomics, fear of discrimination based on genetics, misinterpretation that genetic ancestry is a proxy for social constructs of race/ethnicity are challenges that must be addressed for public acceptance of PM initiatives for common diseases.

Data privacy/access

Patient-centric concern over privacy of their health records and control over data mirrors privacy issues from the health system perspective. Patient concern over privacy and data access is considered a barrier to successful implementation of EHRs (Boonstra and Broekhuis 2010; Barrows, Jr. and Clayton 1996). These concerns include data access management to prevent discrimination or unconsented re-use of data for research or commercial purposes (Caine and Hanania 2013). These concerns were identified in a cross-sectional survey of health consumers in New York (Abramson et al. 2014). These survey participants perceived that use of EHRs would not improve security or privacy (Abramson et al. 2014). A 2013 issue of *JAMIA* presented several papers with a socio-legal perspective on privacy with regard to EHRs (reviewed in (Malin, Emam, and O'Keefe 2013)). In general, patients have expressed desire to maintain some level of control over who may access their EHRs and for what purpose (Caine and Hanania 2013). Participants with highly sensitive data in their EHRs were less likely to grant access to the sensitive data. Participants were discriminating in their likelihood to share data; subjects indicated willingness to share a greater percent of their EHRs with primary care physicians, specialists, and emergency physicians and less willing to share with family, administration, researchers, or non-treating recipients (Caine and Hanania 2013).

Health literacy

Overall health literacy, the ability of individuals to obtain, process, understand basic health information, and use that information to make appropriate health decisions is lacking for a significant proportion of the general public (Lea et al. 2011; Kutner et al. 2006). Health literacy contributes to health outcomes in that individuals with limited health literacy, for example, have increased incidence of chronic illness and decreased use of preventative care (Lea et al. 2011; Berkman et al. 2011). One facet of health literacy is genetics and genomics. Public awareness and understanding of genetics and genomics has improved over the past several years, though many misconceptions still abound (Christensen et al. 2010a). Mathematical illiteracy, or innumeracy, will make understanding true inherited risk of disease and population risk difficult, though the extent to which this impacts genetic-based PM has not been evaluated (Lea et al. 2011; Syurina et al. 2011). Genomic health literacy requires understanding that for complex diseases, genetic and environmental factors contribute to disease risk and some competency in numeracy (Hurle et al. 2013). Health literacy and numeracy skills will play a significant role in successful PM; the average patient with an elementary understanding of genetic

inheritance may be unlikely to fully grasp the difference between the role genetic mutations play in Mendelian diseases with the role of genetic variation in conferring varying degrees of risk for complex traits without additional education. Christensen et al. determined that less than 50% of their survey population was able to correctly answer 7 out of 8 basic statements about genetics (Christensen et al. 2010a). This held for black and white, male and female participants, though there were between group differences (Christensen et al. 2010a). The lack of genetic knowledge identified in Christensen et al. is consistent with earlier reports (Lanie et al. 2004; Emery, Kumar, and Smith 1998). Though public awareness of genetics has increased, it appears that genetic knowledge and understanding has not correspondingly improved. This deficit will impede implementation of PM for in the clinical setting if stakeholder support is lacking due to misinterpretations and lack of understanding.

A recent Health Information National Trends Survey (HINTS), a population-based, nationally representative survey of the civilian, non-institutionalized population in the United States, collected self-report questionnaires with questions relating to the genetic contribution to common, complex diseases, how disease risk is interpreted numerically, and awareness of direct to consumer genetic testing (2013). Analysis of the HINTS data showed participants largely expressed belief in multifactorial causes for common, complex diseases. About 10% of respondents believed the etiology of cancer, diabetes, heart disease, and hypertension to be mostly behavioral in nature, though one-fourth of the respondents believed obesity is caused by mostly behavioral factors (Waters, Muff, and Hamilton 2014). The results of this study are encouraging, in that the participants perceived common, complex diseases having both environmental and genetic causes; however, it is unclear if additional studies will come to the same conclusions or if belief in multifactorial causes of complex diseases is associated with risk-reducing behavior change (Waters, Muff, and Hamilton 2014).

Of particular concern is the penchant for persistent notions of genetic or biologic determinism. Biological determinism, the view that biological components are the causal factors of behavioral differences between people (Malott 2007), and genetic determinism, the idea that genes dictate health outcomes without any contribution from environment (Parrott et al. 2012) are key concepts that must be addressed to maximize potential risk-reducing behavior modification for PM. Numerous groups have attempted to quantify the extent of genetic deterministic beliefs (Lynch et al. 2008; Dambrun et al. 2009; Parrott et al. 2004; Keller 2005). These scales have been criticized for mixing conceptual themes (Condit 2011), for example, genetic determinism and racial bias (Keller 2005; Lynch et al. 2008). Genetic determinism has the potential to undermine PM initiatives if individuals believe behavioral change is

unlikely to improve their health condition or reduce risk of future illness (Shiloh, Rashuk-Rosenthal, and Benyamini 2002; Senior, Marteau, and Peters 1999; Nelkin and Lindee 1996). Consistent with prior studies (Marteau and Weinman 2006; Michie et al. 2003; De Wit and Stroebe 2004), a recent review of the literature identified five factors that determined whether or not a genetic test motivates behavioral change: characteristics of the target behavior, the patient's perception of the severity of the disease, the amount of risk the genetic variant contributes to the disease, patient characteristics (e.g., demographic factors, socio-economic status, educational attainment, level of health literacy and numeracy), and how patients perceive the value of the genetic information compared to other data, such as family history or lab values (Condit 2011). Patients may rely on genetic determinism to support unhealthy behaviors, though this may be applied discriminately depending on the factors listed above (Condit 2011).

Vulnerable populations

Given the complex interactions between ancestry and genetics, it is reasonable to acknowledge the concern that advances in genomics may make individuals and groups vulnerable or disadvantaged in new ways (McClellan et al. 2013). PM depends on the identification of genetic risk markers to classify patients into different risk groups. This classification process may leave some patients vulnerable if the risk categories in which they are placed are groups that historically have been discriminated against or if the risk groups are perceived as socially disadvantaged relative to others. Three types of vulnerabilities may result from the implementation of PM in the clinical setting: disease-specific, genome-specific, or race/ethnicity-specific.

For some complex diseases, like T2D or asthma, the risk for vulnerability may be slight from the social perspective, when compared to other complex diseases like Alzheimer's or autoimmune diseases. Individuals identified as "at risk" for stigmatizing diseases may require additional resources to reduce potential vulnerabilities in the PM setting. With PM utilizing genetics as a basis for assigning disease risk, individuals may be at risk of vulnerability if it becomes apparent that some genotypes (Table 23) consistently utilize more healthcare resources than others. This differs from increased resource utilization by individuals who are already sick, in that PM for common, complex diseases would target asymptomatic individuals who are clinically healthy at the time of screening. Rigorous safeguards to patient privacy and EHR data access may minimize harm to the patient. Even if genomic risk information is well protected, however, cultural beliefs about chronic, complex diseases impact the

patient's ability to understand their disease risk and participant fully in their healthcare (Shaw et al. 2009). Methods to assess how the patient perceives the severity and health risk of a disease, such as the explanatory model (Kleinman and Benson 2006) or the CSM (Marteau and Weinman 2006; McAndrew et al. 2008) may yield insight into ways in which vulnerabilities based on disease and disease risk can be minimized.

Vulnerable populations may result from the categorization of at-risk patients by race/ethnicity. This is a sensitive issue, as there are known differences in disease prevalence among ethnic groups that might naturally suggest candidates for risk group assignment and personalized medicine-driven interventions. Genetic researchers agree that labels such as "black" or "white" do not adequately convey the complexities involved in genetic ancestry and have developed sophisticated statistical methods to account for these complexities (Fujimura and Rajagopalan 2011; Fujimura, Duster, and Rajagopalan 2008), though some clinicians may be prone to using race/ethnicity as a proxy for genetic ancestry (Hunt, Truesdell, and Kreiner 2013). Individuals, who self-identify with a given race/ethnicity, are often genetically heterogeneous. For example, admixture in African Americans yields significant variation in the amount of European ancestry across individuals (Zakharia et al. 2009). This heterogeneity may influence an individual's disease risk and should be taken into account when assigning individuals to a risk group based on sociocultural norms (e.g., European, African, white, black) (Fujimura, Duster, and Rajagopalan 2008; Fujimura and Rajagopalan 2011; Rajagopalan and Fujimura 2012). Though grouping individuals based on race/ethnicity allows geneticists to recruit diverse populations for health research, it is necessary to develop mechanisms that aid public understanding of how these terms are used by scientists, so that negative connotations associated with certain groups are not reinforced (Foster and Sharp 2002). Public understanding of genetic ancestry and potential conflation of genetic ancestry with complex social constructs of race/ethnicity remain a concern. Williams and Eberhardt developed a race conceptions scale (RCS) designed to quantify the idea that race is biologically based (Williams and Eberhardt 2008). The RCS has been correlated with Modern Racism Scale Scores, Attitudes Toward Blacks Scale, Social Dominance Orientation, and acceptance of racial disparities (Condit 2011). Similar to the way in which genetic determinism may be used to perpetuate some health behaviors or beliefs (Condit 2011), genetic ancestry may be misinterpreted to mean that health vulnerabilities are inherent to particular groups and could be dismissed as natural phenomena instead of being recognized as genuine health disparities (Isler et al. 2013).

Summary

Personalized medicine has risen in the public consciousness as medical centers advertise their ability to inform clinical care using a patient's genetic information to improve outcomes and on-demand genetic testing for health and ancestry gain popularity. Until recently, adults could obtain an analysis of their DNA using direct-to-consumer genetic testing companies (Public Health Service Food and Drug Administration and Gutierrez 2013). These analyses included risk prediction for hundreds of common, complex diseases and traits, pharmacogenetic results, and disease carrier status (Public Health Service Food and Drug Administration and Gutierrez 2013). However, integrating genetic information into clinical care for complex disease is challenging for a variety of scientific, ethical, and social issues.

The very nature of complex diseases makes them challenging to understand disease risk on an individual level. Complex diseases result from a mixture of genetics and environmental influence, and the combination of these factors differ between diseases. Even when the entirety of genetic factors has been identified, risk prediction that includes genetics may perform poorly, if the environmental influence is significant. Due to incomplete penetrance, some individuals with a susceptible genotype may never develop the disease. Different variants may be responsible for disease development in some individuals, frustrating variant-phenotype associations, and phenotypic heterogeneity can play a role in recognition of at-risk individuals.

A lack of clearly stated, measurable health outcomes makes determining the utility of personalized medicine challenging. In clinical care, quantitative outcomes could be measured in terms of morbidity and mortality, or through more nuanced measurements such as DALYs, QALYs, or HALEs. Patient/clinician perspectives or symptom severity are examples of more qualitative outcomes that may be assessed. These valuable measurements will provide the foundation for clinical utility research in genetics. Though thousands of genetic variants have been associated with hundreds of clinical traits and diseases, few of them have been validated in the clinical setting. Additional studies should be performed to evaluate the clinical validity of variants that are likely to be included in risk models to identify at-risk populations or guide clinical care, as in pharmacogenetics. EHRs may facilitate this research by aiding risk model algorithm development, prescription information, and the ability to follow patients over the course of many years. Determining the clinical validity and utility of genetic variants for complex diseases could be accomplished through EGAPP, evidence-based practice centers, or translational scientists. The EGAPP working group has repeatedly determined there is

insufficient evidence at present to recommend genetic testing to improve health outcomes for several diseases; however, they have completed fewer than a dozen studies thus far, and the future may be more promising if the needed data become available.

Implementation of personalized medicine for complex disease is also complicated by ethical and social issues. The challenge of accurately identifying at-risk individuals without creating vulnerable populations needs to be addressed. For common, complex diseases such as diabetes or cardiovascular disease, the risk of stigmatization may be less than for other diseases, such as Alzheimer's disease or autoimmune diseases. Certain racial/ethnic groups may be more likely to fall into a vulnerable classification due to historical discrimination; the targeting of specific groups for personalized medicine interventions or screenings may lead to increased vulnerability in those groups. Equitable access to screening and intervention should be based on genetic susceptibility and other risk factors for disease, not sociocultural labels.

How personalized medicine is implemented leads to ethical and social issues. Identification of the at-risk population for PM can take several forms and may depend upon the typical age of onset for a disease. A revised Wilson and Jungner screening criteria may be more appropriate for evaluating whether to pursue screening for a given disease using genomic information (Andermann et al. 2008). Vanderbilt's PREDICT program demonstrates another method of identifying individuals for intervention; patients are identified as high-risk to start a particular drug therapy as calculated by a predictive algorithm, or as they come through the cardiac catheterization clinic (Pulley et al. 2012). For common, complex diseases, such as CVD, timing genetic testing for PM will depend on the value conferred by testing, the disease, its natural course, and the proposed intervention. Current implementation of precision medicine for cancer treatment or pharmacogenetics is targeted; non-targeted approaches such as WGS, WES, and genome-wide genotyping arrays can provide additional data, though the utility of this information is unclear and raises other issues (Lawrence et al.2014). For complex diseases with late-adulthood onset, such as age-related macular degeneration, identifying the at-risk patients in early adulthood or middle age may be the most appropriate time. For diseases that occur at earlier ages, such as CVD or T2D, it may be more advantageous to identify the at-risk population in childhood, perhaps maximizing the effects from the intervention, but raising a host of issues that are beyond the scope of this work (Erickson et al. 2014; Ross et al. 2013). These differences suggest that the decision to use PM for complex diseases will require thoughtful deliberation on a disease by disease and intervention by intervention basis. A new method of determining what diseases

are most appropriate for PM initiatives may be required and may rely on evidence-based medicine and cost-utility analyses. A three-pronged method that considers the potential utility, benefits, and harms is one method of identifying which diseases, which individuals, and when a precision medicine approach is best suited. Evidence supporting any precision medicine method should begin in multidisciplinary collaborations between social scientists, clinicians, and geneticists. Legal issues, which are outside the scope of this work, are numerous for precision medicine.

Despite the recent cessation of risk prediction for complex diseases through a direct-to-consumer genetic testing company, the public desire for useful health information and individualized treatment is unlikely to subside. Numerous medical centers have embraced precision medicine to inform cancer care and pharmacogenetics, though evidence that the practice has improved patient care or health outcomes is still emerging. Yet the role that precision medicine can play for common, complex disease has not been fully explored. Numerous scientific, ethical, and social concerns will need to be addressed, and evidence that genetic information improves health outcomes through medical and lifestyle modifications must be determined.

CHAPTER VII

CONCLUSION AND FUTURE DIRECTIONS

Conclusion

Personalized medicine, the use of genomic data to guide clinical decision making for an individual patient, currently takes a variety of forms. For numerous cancers, the result is cancer tumors being tested for specific genetic variants and measuring gene expression levels that may make the cancer more susceptible to a specific chemotherapy regimen. Genetic testing prior to the development of cancer for familial forms of breast cancer or colorectal cancer allows patients to modify risk behaviors or undergo prophylactic surgery or chemotherapy. Importantly, this may also be used to infer risk in related individuals, perhaps providing clinically useful information for entire families. Pharmacogenomic studies look to determine the correct dosing strategy and medication choice for patients to avoid adverse events and maximize therapeutic efficacy. For individuals who have previously been without a clear diagnosis, exome and/or whole genome sequencing may provide insight to the underlying biological mechanism responsible for their disorder and end the diagnostic odyssey.

Despite these successes, the promise of personalized medicine to revolutionize care for common, complex diseases has not materialized. With the exception of Mendelian disorders, some cancers, and pharmacologic interactions, the prognostic capabilities of genetic tests to accurately gauge risk of disease development are low. This results from both an incomplete understanding of the biologic mechanisms responsible for disease development and progression, and a fragmented grasp of the role of environmental factors and interactions play in these disorders. Though cancer and adverse drug events contribute significantly toward the overall health picture of the nation, they are vastly overshadowed by the public health impact from complex diseases, like age-related macular degeneration, type 2 diabetes, and cardiovascular disease. Given the increasing population burden of these disorders, a personalized medicine approach to correctly identify which patients are at greatest risk prior to developing the disorder and then targeting interventions to those individuals in order to prevent, delay, or ameliorate the disorder, could result in substantially improved health outcomes.

This body of work sought to understand the role of genetic variation in complex traits associated with women's health. Though women comprise a majority of the population in the United States, their specific health needs have been largely ignored until the past few decades. This disparity is mirrored in biomedical research in diverse populations. Differences in disease prevalence and severity between men and women or between individuals of European- and non-European ancestry for many complex traits have been documented. How gender, race/ethnicity, and the role of environmental interactions contribute to these differences is unclear. Nevertheless, that these differences exist suggests a potential role for personalized medicine approaches to health in women and diverse populations to improve health outcomes.

The role of the reproductive lifespan in women has been associated with various complex traits and diseases. Beginning with menarche during the pubertal period, until menopause, cyclic hormone patterns and the role of pregnancy impact a woman's risk of developing certain cancers, osteoporosis, and cardiovascular disease. In addition, women are more at risk of most autoimmune disorders, and there are differences in autoimmune disease risk that varies by ethnicity. Despite the importance of these traits, their molecular basis and the biological mechanisms by which they play a role in disease development are not well understood. The importance of these traits, their association with disease, and gap in scientific evidence served as the impetus for studying these traits.

I began by assessing the role of genetic variation on age at menarche and age at natural menopause in African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study in the first of three case studies. Using the Metabochip, a genotyping array with a primary emphasis on fine-mapping GWAS-identified genomic regions associated with cardiovascular traits, I sought to generalize to our sample variants previously identified in European-descent cohorts for these traits. Differences in the timing of these traits by race/ethnicity have been observed consistently and have yet to be fully explained. This was the first study to consider these traits in a large African American cohort at the time of publication, though others have followed. We were able to generalize only one previously identified SNP, rs1361108, for age at menarche, and two SNPs, rs897798 and rs9385399, for age at natural menopause, to our cohort. We failed to identify novel variants associated with age at menarche, after correcting for multiple tests; however, we observed three novel SNPs associated with age at natural menopause in this population. Our ability to generalize variants identified in European cohorts, including the *LIN28B* region, was compromised by Metabochip coverage of some genomic regions and its emphasis on genes involved in lipid metabolism

and cardiovascular traits. In addition, the sample size and minor allele frequency differences between our population and European populations likely impacted our ability to replicate known variants associated with these traits. Despite these limitations, our results demonstrated the ability to use the Metabochip to identify variants associated with reproductive lifespan traits in a diverse population.

As personalized medicine approaches rely substantially on the use of electronic health records (EHRs) to identify groups of patients for interventions, I developed an algorithm to extract age at menarche and age at menopause from the Vanderbilt University Medical Center biorepository, BioVU. My goal was to create an automated method of extracting these important data from EHRs; these data and this algorithm could subsequently be used by researchers who continue to study these traits and the genetic influences on their timing, as well as by clinicians who could use this information to identify individuals for clinical research cohorts or for targeted interventions using ages at these milestones as risk factors for cancers, cardiovascular diseases, or other complex traits. Despite known links between the timing of these events and complex diseases, these data are not consistently nor uniformly collected and placed in EHRs. To address this issue, we created a method using a combination of free text data mining for regular expressions and pattern matching, billing, and procedure codes to identify the age at these events in women and children (age>7) in BioVU. Our method captured these data successfully, with a positive predictive value of 94.0% for the age at menarche algorithm. In addition, we were able to discriminate between natural menopause and all-cause menopause.

Our algorithm identified ages at these reproductive milestones that concur with national estimates. We identified several challenges to accurately extracting this information from BioVU that may be generalizable to other institutions using our method. First, in BioVU, specific ages and dates may be de-identified or date-shifted. De-identification was observed in approximately half of our age at menarche and age at natural menopause results; it is possible that if these data were not de-identified, our algorithm's predictive abilities could change. In addition, the date-shifting that is done for privacy concerns meant that the actual timing of these events could have occurred within a one year window (six months forward or backward of the date), further compromising our ability to accurately identify these ages. Furthermore, both the de-identification and date shifting were applied inconsistently, sometimes within a single record. While our method prioritized an actual date/age over a de-identified or shifted age when reporting results, we found instances where the algorithm failed to find the exact date/age which decreased the predictive capability of our method. These challenges

frustrated our ability to obtain these data; however, our method has been recently used by another institution to extract the data from their biorepository/EHR for contribution to a meta-analysis for timing menarche and menopause in the PAGE Study (data not shown). It is hoped that data extraction methods, such as ours, may be used more frequently to provide researchers with necessary data for studies.

The second case study assessed the role of genetic variation in the development of endometrial cancer. Endometrial cancer (EC) is the most common invasive gynecologic cancer in the United States. Typically occurring in post-menopausal women, EC risks appear to vary by race/ethnicity. Multiple subtypes of EC complicate studies evaluating the genetic risk factors for disease. Endometrioid EC tends to be less aggressive than the clear cell or serous types of EC and is generally estrogen-dependent. The molecular mechanisms for EC development are not fully understood; however, timing of menarche, menopause, and pregnancy, and parity are known to be associated with EC. Previous genetic studies have identified several genes implicated in EC, though results suffer from a lack of consistency. We performed a candidate-gene association study for EC using variants selected due to their association with a variety of cancers. Pleiotropic effects have been observed for other cancers; identifying cancer variants additionally linked to EC may result in better understanding of general cancer mechanisms. After quality control filters, our single-study sample size for this study was drastically reduced. Consequently, though our single-site results have been presented in Chapter IV, we contributed our results to a larger meta-analysis comprised of studies from two consortia: PAGE and the E2C2 (Epidemiology of Endometrial Cancer Consortium). The leading results from our analysis implicate SNPs previously associated with prostate and colorectal cancer. This is interesting, as the PAGE/E2C2 meta-analysis also observed significant associations for SNPs associated with prostate cancer, suggesting a potential shared mechanism for the development of both cancers. Given our small sample size, our single-site results should be interpreted with caution.

A genome-wide association study (GWAS) for thyroid stimulating hormone (TSH) levels was the last case study presented here. Unlike a candidate gene approach (Chapter IV), or a more targeted association study (Chapter II), this method generates hypotheses for additional studies, rather than testing specific hypotheses. Similar to the reproductive traits and endometrial cancer, the heritability of TSH levels is approximately 0.50, though genetic studies have failed to identify variants responsible for more than a few percent. And like reproductive traits and EC, sex and population differences have been observed in both mean TSH levels and prevalence of thyroid disease. Performed in the eMERGE

Network, this GWAS replicated known variants associated with TSH levels in European Americans, including the well-characterized *PDE8B* rs2046045; however, we were unable to generalize most associations to African Americans. Differences in minor allele frequencies between the European and African-descent populations may have impacted our statistical power to replicate in the African American population.

We also considered the role of interactions between genetic(SNPs) and environmental factors (body mass index (BMI)) in this case study. Obesity has been implicated in higher TSH levels and contribute to a change in an individual's normal TSH level (Marzullo et al. 2010; De et al. 2007). We identified two loci, *NRG1* and *NFIA*, with BMIxSNP interactions in European Americans. Nominally significant, the *NRG1* interaction is particularly interesting as it has been associated with thyroid cancer (Gudmundsson et al. 2009) and has been shown to regulate cell proliferation in an animal thyroid cell model (Breuleux 2007). Though interactions were identified for the African Americans, small sample sizes and low genotype counts per BMI category; comparisons across groups are therefore difficult. This study highlights the challenges in assessing genetic and environmental interactions in small samples.

Identifying the genetic variants associated with these women's health-related traits lays the foundation for personalized medicine in a clinical setting; however, identification is merely the first step of many in translating this data into clinical practice for complex diseases. Though identification of genetic variants associated with a given trait may provide insight to the underlying biological mechanisms, these associations are rarely predictive for disease risk in an individual. The complexity in understanding the role of genetics in many common diseases, like cancer or hypothyroidism, results in part from the interactions of genes and environment, and is complicated by genetic heterogeneity and incomplete penetrance. Few genetic studies have considered the positive predictive value of the genetic associations for disease risk. Until recently, several companies provided disease risk for hundreds of common, complex diseases and traits, despite the lack of rigorously tested evidence that GWAS-identified SNPs are substantially predictive of disease risk. It is not unlikely that physicians will encounter patients with this disease risk information in the near future, and be confronted with how to best manage preventative care based upon this genetic data. The EGAPP framework to evaluate genetic evidence for inclusion in clinical care was established in 2007 based off the earlier Analytic Validity, Clinical Validity, Clinical Utility, and Ethical, Legal, & Social Implications (ACCE) Project. EGAPP has performed numerous studies in the last seven years evaluating the evidence that genetic testing

improves health outcomes for several diseases. These studies have consistently found a lack of analytic evidence upon which to base their recommendations. This lack of evidence frustrates clinician's abilities to determine when a personalized medicine approach may be effectively implemented.

To address the lack of data observed by the EGAPP working group and others, we developed a rapid evidence review procedure as a first-pass method that could be implemented in an academic medical center setting to evaluate the analytic evidence supporting the use of genetic data in the clinical setting to improve health outcomes for a complex disease and used hypothyroidism/serum TSH levels as the example. Systematic evidence reviews (SERs) are notoriously time consuming, often taking more than one year to complete, and costly, both economically and in terms of human capital. To address these issues, a streamlined approach to the SER was taken. As in a full SER, an overarching research question and supporting key questions were developed to identify the desired analytic evidence. One database was searched for the literature review (PubMed) and the abstract and full text reviews were performed by two individuals, with access to a third independent party as necessary. The REDCap database was used to capture and house the data for the review process; this tool has a web-based interface and allows for automatic export of the data to common statistical packages.

Though we identified hundreds of potential studies for the review, we found no studies that met our inclusion criteria. Importantly, we found studies lack important clinical validity data, such as positive predictive values or areas under receiver-operator characteristics curves. In addition, we found a lack of evidence that studies considered the clinical utility of knowing the genetic variants improved health outcomes. Finally, ethical, legal, and social issues were not addressed in the studies identified through our rapid review; though this may be a result of searching one database only or key questions that did not sufficiently address that topic. In general, our rapid review results concur with those performed for other diseases/traits and highlight the need for genetic studies to publish this key analytic data.

A lack of analytic evidence is not the only barrier to implementing personalized medicine for complex diseases. The health care system may not be adequately prepared to expand PM in this way due to data management challenges, data privacy issues, and complications arising from EHR integration and physician decision support mechanisms. Economic analyses on the costs to a health care system to implement PM are largely missing. Logistic issues with legal implications include the consequences of a patient changing their medical home, such as data and sample ownership. Institutions that develop highly predictive algorithms for certain complex diseases may feel protective

of their investment and be unwilling to share with other medical centers. Concerns around proprietary testing, analytic models, patents, and licensing will likely play a significant role as personalized medicine becomes more prevalent. The patient and public role in PM requires considering data privacy and access, health literacy, and vulnerable populations. Historic injustices to certain populations have led to distrust of the medical community and continue to inhibit participation in biomedical research. Unwillingness to share information with researchers presents an immediate challenge for PM implementation. Health outcomes research for PM initiatives will rely on a large number of individuals consenting that their data be included in analyses; if patients are less likely to allow sharing of their personal health information with researchers, necessary clinical utility and outcomes data may be difficult to obtain.

Future Directions

Genetic studies

Though the health of women and minorities has been significant focus of governmental agencies in the past several decades, health disparities still abound for many diseases. It is clear from the limitations described in the previous case studies that additional genetic research in women and minorities should remain a goal of researchers. The etiology of different disease trajectories for women and men for some diseases has yet to be fully explored. Self-defined or third-party observed labels of race/ethnicity may be insufficient to serve as proxies for genetic ancestry; known variation in percent African ancestry across African Americans should serve as an example of the complexities involved when assigning disease risk to an individual based on population-level data. However, this data may reflect socio-cultural data that genetic ancestry cannot.

Additional studies in large cohorts with diverse populations will improve our understanding of the genetic contribution to disease. Several studies, such as the Jackson Heart Study (Taylor, Jr. et al. 2005) and the Southern Community Cohort (Signorello et al. 2005), focus their attention on understanding the underlying causes of health disparities for minorities in cancer and cardiovascular disease, and consortia like the PAGE Study have a considerable multi-ethnic component. Despite these and other studies, health disparities remain a significant challenge and it is unclear to what extent genetic ancestry, in combination with other genetic and environmental factors influences disease risk and highlights the need for additional large studies in these populations. This is much more easily said

than done, however. Past studies that led to mistrust of the medical and scientific community have had long-lasting effects on participation in research studies and clinical trials. While social scientists have identified methods to overcome this mistrust, minority participation in clinical trials is dismal, even where disease burden is higher in non-European descent individuals. Collaborations with social scientists and community leaders that lead to a better understanding of the barriers that exist to minority participation in biomedical studies may improve recruitment and retention.

Despite the hundreds of GWAS studies performed for dozens of diseases and complex traits, the vast majority of associations have failed to explain much of the variation observed. The "missing heritability" phenomenon has led to increased exploration of the role of interactions and rare variants, though there is little evidence that these studies will be more successful at explaining the heritability of a specific trait. Larger and larger samples sizes are required to uncover variants with smaller and smaller effect sizes. While the additive nature of variants likely plays a significant role in the "missing heritability," for some diseases it is unlikely researchers will ever be able to amass enough samples to identify all causal variants. Prospective studies and nested case-control studies within prospective studies can reduce the effective sample size needed for some genetic studies; however, interaction analyses will likely be compromised by the number of samples for each possible interaction. Our lack of understanding of the biological mechanisms that lead to disease certainly plays a significant role in the failure to identify the responsible genetic variants. The ENCODE Project, which has identified regulatory regions of the genome, may lead to insight on the biologic relevance to disease of many intergenic or intronic GWAS associations. Whole exome (WES) and whole genome sequencing (WGS) may yield important clues about genetic contribution to disease. As the number of individuals who undergo WES/WGS increases, knowledge about allele frequencies across population will improve, leading to better understanding of which alleles are causal variants (and in which populations) and which are along for the ride. This again leads back to the need for increased participation in research by non-European populations.

Role of research findings in clinical care

Establishing an evidence base

Though genetic research to identify the basic biology underlying disease is necessary, this body of work has examined the role of genetic variation in disease from a context of personalized medicine.

Moving research findings into the clinical sphere is challenging and encompasses numerous disciplines. It is clear that from the studies performed in the past decade and the rapid review described herein, that the analytic evidence that genetic testing fails to improve health outcomes for common, complex diseases currently. I would like to emphasize that this does not suggest that genetic testing will never lead to improved outcomes, and certainly successes in cancer research and pharmacogenetics demonstrate the potential of PM approaches in the clinical setting. I have outlined many of the challenges personalized medicine currently faces—they are by no means insurmountable.

The current lack of analytic evidence prevents researchers from demonstrating clinical validity and clinical utility. This can be addressed by a concerted effort from geneticists to publish the positive predictive value or AUC for their studies. It may be that neither the PPV nor the AUC is the most appropriate statistic to demonstrate clinical validity—PPV works best when the outcome is dichotomous and AUC implies the test captures all of the heritability. Any genetic test for a complex trait is unlikely to meet these criteria. Biostatisticians working alongside geneticists may be able to develop a new method that better describes the clinical validity of genetic variants.

How institutions prioritize which diseases may benefit from a PM approach is complicated. From an evidence-based perspective, those diseases for which the analytic evidence is complete and the disorder fulfills all/most of the updated Wilson and Jungner(Andermann) criteria may prove to be the easiest. Extensions to current PM initiatives in cancer and pharmacogenetics are a natural first step. Other disorders, for which there are well-documented interventions, such as cardiovascular disease and type 2 diabetes, may be alternative first steps. Prioritization based on the public health burden of the disease offers another method to determine which diseases could benefit from this approach.

The economic realities of biomedical research today cannot be overlooked. Economic studies considering the cost-effectiveness of various interventions or screening modalities, including genetic testing, are important and mostly absent from the literature. In light of limited financial resources, collaborations across institutions, the inclusion of non-academic medical centers in studies, partnerships with industry and non-academic entities (e.g., patient advocacy groups) with financial backing may provide opportunities for large-scale studies with reduced cost burden for any one institution. Additionally, data previously collected for other research studies or the information held in EHRs may alleviate the burden of researchers to recruit and test new participants. This leads directly to legal, ethical, and practical challenges which should be carefully considered. The role of patients and the public as research partners should not be underestimated. Examples of motivated patients who

desire to become involved in biomedical research by contributing samples and financially may be seen in Patients Like Me, UBiome, and 23&Me's research arm, 23&We. This approach led to recruitment of more than 3,000 individuals with Parkinson's disease and the identification of two new variants associated with the disorder through 23&Me's Parkinson's Research Community(23&Me). How this method can be implemented in collaboration with a medical institution should be considered.

Complex disease prevention and management

Personalized medicine requires a multidisciplinary approach—geneticists are experts in genetics, not behavioral psychology; roles for ethicists, psychologists, geneticists, epidemiologists, clinicians, and other professionals should be included. Patient motivation to change risk behaviors has been largely unsuccessful with regard to genetic testing, though *BRCA1/2* testing has been shown to increase breast cancer screenings. Whether knowing disease risk based on genetics for other complex diseases will lead to heightened surveillance or reduction in risk behaviors, such as smoking, is unclear, and downstream consequences of these interventions are not likely to be known for many years. Tracking individual patients through the medical system would, in essence, become a prospective study, where health outcomes could be evaluated and nested case-control studies be performed. However, this calls into question additional ELSI and technical issues, such as how to manage patients who leave one medical home for another. Widespread collaborations across institutions nationally, or a single-payer health system could alleviate some of these problems, though thoughtful consideration of these suggestions is beyond the scope of this body of work.

Summary

Our scientific understanding of the basis for many common, complex diseases and traits has improved in the past decades with advances in genomic studies. Despite these advances, real improvements to health outcomes remain generally unrealized and widespread health disparities continue. To address these issues, I have considered the role of genetic variation in the timing of the female reproductive lifespan, risk of endometrial cancer, and serum TSH levels in both European-descent and African American populations, using candidate gene, large-scale association, and genome-wide association studies. I developed a method to extract important reproductive lifespan data out of an EHR for use in research studies and suggest how that could be used in the clinical setting. I have identified variants associated with these traits, considered how prior published studies generalize to

more diverse populations, discovered potential gene-environment interactions. These studies have been undertaken in the context of personalized medicine—how genetic data can be used to inform clinical care. I have noted significant challenges facing implementation of personalized medicine for complex diseases and performed a rapid review that could serve as a template for others wishing to perform similar studies. Evaluation of the role of genetic variants as predictors of complex disease is developing; though the challenges are significant, it is likely they will be overcome in the future.

APPENDICES

Appendix A: Comparison of SNPs in Elks *et al.* (2010) meta-analysis for AM to African American women in the PAGE Study.

Locus		Gene/	Elks et al.				African American women from the PAGE Study					
SNP	Chr	region	Minor Allele	MAF	Beta	P-value	Best Proxy SNP from present study	r^2 in HapMap CEU/YRI	Coded Allele	CAF	Beta (SE)	P-value
rs7759938	6	*LIN28B*	C	0.32	0.12	5.4E-60	rs7759938	-	A	0.46	-0.02(0.04)	0.61
rs2090409	9	*TMEM38B*	A	0.31	-0.09	2.2E-33	rs4452860	0.83/0.82	A	0.67	-0.03(0.04)	0.43
rs1079866	7	*INHBA*	G	0.15	0.08	5.5E-14	rs6947337	0.02/0.001	A	0.28	-0.07(0.04)	0.10
rs466639	1	*RXRG*	T	0.13	-0.08	1.3E-13	rs285482	0.36/0.05	A	0.55	-0.003(0.04)	0.93
rs6438424	3	*3q13.3*	A	0.50	-0.05	1.4E-13	rs9283566	0.51/0.15	A	0.29	0.03(0.04)	0.47
rs1398217	18	*FUSSEL18*	G	0.43	-0.05	2.3E-13	NA	NA	NA	NA	NA	NA
rs12617311	2	*PLCL1*	A	0.32	-0.06	6.0E-13	NA	NA	NA	NA	NA	NA
rs9635759	17	*CA10*	A	0.32	0.06	7.3E-13	NA	NA	NA	NA	NA	NA
rs6589964	11	*BSX*	A	0.48	-0.05	1.9E-12	rs922252	0.24/0.16	A	0.40	0.03(0.04)	0.43
rs10980926	9	*ZNF483*	A	0.36	0.05	4.2E-11	rs6477828	0.12/0.13	A	0.55	-0.01(0.04)	0.76
rs17268785	2	*CCDC85A*	G	0.17	0.06	9.7E-11	NA	NA	NA	NA	NA	NA
rs13187289	5	*PHF15*	G	0.20	0.06	1.9E-10	NA	NA	NA	NA	NA	NA
rs7642134	3	*VGLL3*	A	0.38	-0.05	3.5E-10	rs1825896	0.02/0.05	A	0.15	-0.04(0.05)	0.40
rs17188434	2	*NR4A2*	C	0.07	-0.09	1.1E-09	NA	NA	NA	NA	NA	NA
rs2002675	3	*TRA2B*	G	0.42	0.04	1.2E-09	NA	NA	NA	NA	NA	NA
rs7821178	8	*PXMP3*	A	0.34	-0.05	3.0E-09	NA	NA	NA	NA	NA	NA
rs1659127	16	*MKL2*	A	0.34	0.05	4.0E-09	rs1659127	-	A	0.30	0.03(0.04)	0.46
rs10423674	19	*CRTC1*	A	0.35	0.04	5.9E-09	rs757318	0.63/0.14	A	0.78	-0.03(0.04)	0.49
rs10899489	11	*GAB2*	A	0.15	0.06	8.1E-09	rs7115850	0.96/0.58	C	0.58	-0.05(0.04)	0.23
rs6575793	14	*BEGAIN*	C	0.42	0.04	1.2E-08	NA	NA	NA	NA	NA	NA
rs4929923	11	*TRIM66*	T	0.36	0.04	1.2E-08	rs4929923	-	A	0.46	0.07(0.04)	0.06
rs6439371	3	*TMEM108*	G	0.34	0.04	1.3E-08	NA	NA	NA	NA	NA	NA
rs900145	11	*ARNTL*	C	0.30	0.04	1.6E-08	rs900145	-	A	0.47	-0.03(0.04)	0.41
rs6762477	3	*RBM6*	G	0.44	0.05	1.6E-08	rs2240327	0.69/0.16	A	0.64	-0.06(0.04)	0.15
rs2947411	2	*TMEM18*	A	0.17	0.05	1.7E-08	rs2947411	-	A	0.23	-0.10(0.04)	0.02
rs1361108	6	*C6orf173*	T	0.46	-0.04	1.7E-08	rs9385399	1.00/0.60	A	0.25	-0.12(0.04)	0.01
rs1364063	16	*NFAT5*	C	0.43	0.04	1.8E-08	rs889398	0.93/0.42	A	0.28	-0.05(0.04)	0.25
rs633715	1	*SEC16B*	C	0.20	-0.05	2.1E-08	rs516636	1.00/0.92	A	0.11	0.04(0.06)	0.55
rs4840086	6	*PRDM13*	G	0.42	-0.04	2.4E-08	NA	NA	NA	NA	NA	NA

Locus		Gene/		Elks et al.				African American women from the PAGE Study				
SNP	Chr	region	Minor Allele	MAF	Beta	P-value	Best Proxy SNP from present study	r^2 in HapMap CEU/YRI	Coded Allele	CAF	Beta (SE)	P-value
rs7617480	3	*KLHDC8B*	A	0.22	0.05	2.8E-08	rs13096474	0.76/0.56	A	0.35	-0.05(0.04)	0.21
rs9939609	16	*FTO*	A	0.40	-0.04	3.1E-08	rs9939609	-	A	0.48	0.003(0.04)	0.93
rs852069	20	*PCSK2*	A	0.37	-0.04	3.3E-08	NA	NA	NA	NA	NA	NA
rs757647	5	*KDM3B*	A	0.22	-0.05	5.4E-08	rs757647	-	A	0.41	-0.06(0.04)	0.09
rs9555810	13	*C13orf16*	G	0.28	0.04	5.6E-08	NA	NA	NA	NA	NA	NA
rs16938437	11	*PHF21A*	T	0.09	-0.07	5.9E-08	rs16938437	-	A	0.23	0.02(0.04)	0.63
rs2687729	3	*EEFSEC*	G	0.27	0.04	1.3E-07	rs2811415	0.17/0.83	A	0.29	0.05(0.04)	0.26
rs1862471	19	*OLFM2*	G	0.47	0.04	1.5E-07	NA	NA	NA	NA	NA	NA
rs12472911	2	*LRP1B*	C	0.20	0.05	1.5E-07	NA	NA	NA	NA	NA	NA
rs3914188	3	*ECE2*	G	0.27	-0.04	2.6E-07	NA	NA	NA	NA	NA	NA
rs2243803	18	*SLC14A2*	A	0.40	0.04	3.4E-07	NA	NA	NA	NA	NA	NA
rs3743266	15	*RORA*	C	0.32	-0.04	8.0E-07	rs17270188	0.31/0.02	A	0.92	0.05(0.07)	0.47
rs7359257	15	*IQCH*	A	0.45	0.03	1.9E-06	rs7359257	-	A	0.66	0.06(0.04)	0.15

Comparison of previously reported SNPs associated with AM in European descent women to 4,159 African American women from the PAGE Study in a model minimally adjusted for study site and principal components (Model 1). Beta values from Elks et al. converted from weeks to years. Data presented are for the previously identified SNP. If the previously identified SNP was not directly genotyped in present study, data shown are for the best proxy SNP based on linkage disequilibrium from the International HapMap Project CEU panel. (NA) = no sufficient proxy available on the Metabochip. Abbreviations: single nucleotide polymorphism (SNP), age at menarche (AM), Population Architecture using Genomics and Epidemiology (PAGE), chromosome (Chr), minor allele frequency (MAF), coded allele frequency (CAF).

Appendix B. Comparison of SNPs in Stolk *et al*. meta-analysis for ANM to African American women in the PAGE Study.

Locus		Gene/		Stolk et al.			African American women from the PAGE Study					
SNP	Chr	region	Minor Allele	MAF	Beta	P-value	Best Proxy SNP from present study	r^2 in HapMap CEU/YRI	Coded allele	CAF	Beta (SE)	P-value
rs4246511	1	*RHBDL2*	T	0.27	0.24	9.08E-17	NA	NA	NA	NA	NA	NA
rs1635501	1	*EXO1*	C	0.48	-0.16	8.46E-10	rs1776133	0.91/0.23	A	0.70	-0.05(0.04)	0.17
rs2303369	2	*FNDC4*	T	0.39	-0.18	2.25E-12	rs2303369	-	A	0.36	-0.02(0.03)	0.48
rs10183486	2	*TLK1*	T	0.37	-0.20	2.21E-14	rs4668368	0.86/0.62	A	0.65	0.03(0.03)	0.37
rs7606918	2	*METAP1D*	G	0.16	-0.23	2.89E-08	rs11681005	0.08/0.02	A	0.13	-0.06(0.05)	0.25
rs4693089	4	*HELQ*	G	0.49	0.23	2.38E-19	NA	NA	NA	NA	NA	NA
rs890835	5	*RNF44*	A	0.11	0.18	6.10E-06	NA	NA	NA	NA	NA	NA
rs365132	5	*UIMC1*	T	0.49	0.29	9.11E-32	NA	NA	NA	NA	NA	NA
rs2153157	6	*SYCP2L*	A	0.49	0.17	7.76E-12	rs2153157	-	A	0.70	0.03(0.04)	0.47
rs1046089	6	*PRRC2A*	A	0.35	-0.21	1.63E-16	rs9264532	0.08/0.05	A	0.65	0.02(0.03)	0.50
rs2517388	8	*ASH2L*	G	0.17	0.26	9.31E-15	rs4976896	0.003/0.000	A	0.78	0.01(0.04)	0.77
rs12294104	11	*MPPED2*	T	0.17	0.23	1.46E-11	rs7951733	0.35/-	A	0.99	0.11(0.13)	0.37
rs2277339	12	*PRIM1*	G	0.10	-0.38	2.47E-19	rs12809466	0.01/0.04	A	0.88	0.02(0.05)	0.66
rs3736830	13	*KPNA3*	G	0.16	-0.18	9.41E-08	NA	NA	NA	NA	NA	NA
rs4886238	13	*TDRD3*	A	0.33	0.17	9.53E-11	NA	NA	NA	NA	NA	NA
rs2307449	15	*POLG*	G	0.41	-0.18	3.56E-13	rs12593363	0.91/0.12	A	0.74	0.01(0.04)	0.83
rs10852344	16	*GSPT1*	C	0.42	0.17	1.01E-11	rs8053435	0.04/0.01	A	0.77	-0.01(0.04)	0.77
rs11668344	19	*TMEM150B*	G	0.36	-0.42	1.45E-59	NA	NA	NA	NA	NA	NA
rs12461110	19	*NLRP11*	A	0.36	-0.16	8.74E-10	rs302469	0.03/0.004	A	0.25	-0.05(0.04)	0.22
rs16991615	20	*MCM8*	A	0.07	0.95	1.42E-73	rs16991615	-	A	0.01	-0.17(0.15)	0.25

Comparison of previously reported SNPs from Stolk et al. meta-analysis (Stolk et al.2012) associated with ANM in a combined cohort (discovery and replication) of 53,403 European descent women to 1,860 PAGE Study African American women in a minimally adjusted for study site and principal components for ANM. MAF from Stolk et al. reported for discovery cohort, beta and p-values reported for combined discovery and replication cohorts (Stolk et al.2012). Data presented are for the previously identified SNP. If the previously identified SNP was not directly genotyped in present study, data shown are for the best proxy SNP based on linkage disequilibrium from the International HapMap Project CEU panel. (NA)= no sufficient proxy available on the Metabochip. Abbreviations: single nucleotide polymorphism (SNP), age at natural menopause (ANM), Population Architecture using Genomics and Epidemiology (PAGE), chromosome (Chr), minor allele frequency (MAF), coded allele frequency (CAF).

Appendix C. SNPs associated ($p \leq 1x10^{-04}$) with AM in African American women from the PAGE Study.

						MODEL 1			MODEL 2		
CHR	**SNP**	**GENE**	**GENE REGION**	**CODED ALLELE**	**CAF**	**BETA**	**SE**	**P VALUE**	**BETA**	**SE**	**P VALUE**
11	rs11604207	*RSF1*	upstream	A	0.58	0.38	0.08	1.59E-06	0.38	0.08	1.82E-06
17	rs59622946	*HEXIM2*	flanking	A	0.01	1.01	0.23	1.14E-05	1.04	0.23	4.93E-06
1	rs2753399	*ZFYVE9*	intronic	A	0.08	0.31	0.07	1.16E-05	0.30	0.07	1.63E-05
5	rs40602	*MAST4*	intronic	A	0.57	0.16	0.04	1.26E-05	0.15	0.04	3.65E-05
15	rs7181548	*C15orf27*	intronic	A	0.58	-0.16	0.04	1.96E-05	-0.16	0.04	2.36E-05
8	rs4922116	*LPL*	downstream	A	0.15	-0.22	0.05	2.18E-05	-0.21	0.05	3.10E-05
8	rs1372339	*LPL*	downstream	A	0.84	0.21	0.05	2.19E-05	0.21	0.05	2.89E-05
17	rs116523982	*HEXIM2*	flanking	A	0.01	0.96	0.23	2.51E-05	0.99	0.23	1.11E-05
17	rs3744412	*HEXIM2*	5′ UTR	C	0.01	0.96	0.23	2.52E-05	1.00	0.23	1.12E-05
3	rs11922097	*PPP2R3A*	upstream	A	0.55	0.16	0.04	2.55E-05	0.16	0.04	1.64E-05
17	rs16939893	*HEXIM2*	intronic	A	0.004	1.26	0.30	2.73E-05	1.29	0.30	1.46E-05
6	rs73725617	*PHACTR1*	intronic	A	0.99	-0.64	0.15	3.11E-05	-0.60	0.15	6.83E-05
7	rs11979121	*TFEC*	upstream	A	0.98	-0.52	0.13	4.50E-05	-0.51	0.13	5.41E-05
3	rs1320623	*LSG1*	intronic	A	0.37	0.15	0.04	5.64E-05	0.15	0.04	5.56E-05
12	rs61507607	*CUX2*	intronic	A	0.43	0.15	0.04	5.85E-05	0.13	0.04	3.15E-04
11	rs11224447	*ARHGAP42*	intronic	A	0.07	0.30	0.07	6.11E-05	0.30	0.07	6.11E-05
19	rs1273516	*CYP4F22*	downstream	A	0.40	0.15	0.04	6.33E-05	0.15	0.04	6.29E-05
6	rs9503555	*IRF4*	upstream	A	0.78	0.18	0.04	6.83E-05	0.18	0.04	6.03E-05
15	rs8032832	*FAM174B*	upstream	A	0.35	-0.15	0.04	7.26E-05	-0.16	0.04	3.70E-05
5	rs17730451	*C5orf41*	3′ flanking	A	0.05	0.34	0.09	7.53E-05	0.33	0.09	1.35E-04
7	rs849326	*JAZF1*	upstream	C	0.09	0.25	0.06	9.35E-05	0.24	0.06	1.07E-04

Tests of association at p≤1x10-04 for Model 1 from individual SNP linear regressions adjusted for study site and principal components (Model 1) and study site, principal components, year of birth, and BMI (Model 2) in 4,159 African American women from the PAGE Study are shown. For each significant test of association, the chromosome, rs number, nearest gene, location, coded allele, beta, standard error (SE), and p-value are given. Genes listed are nearest genes to the SNP as measured from the transcription start site for upstream SNPs or the transcription stop site for downstream SNPs. Abbreviations: single nucleotide polymorphism (SNP), age at menarche (AM), Population Architecture using Genomics and Epidemiology (PAGE), chromosome (CHR), coded allele frequency (CAF), standard error (SE).

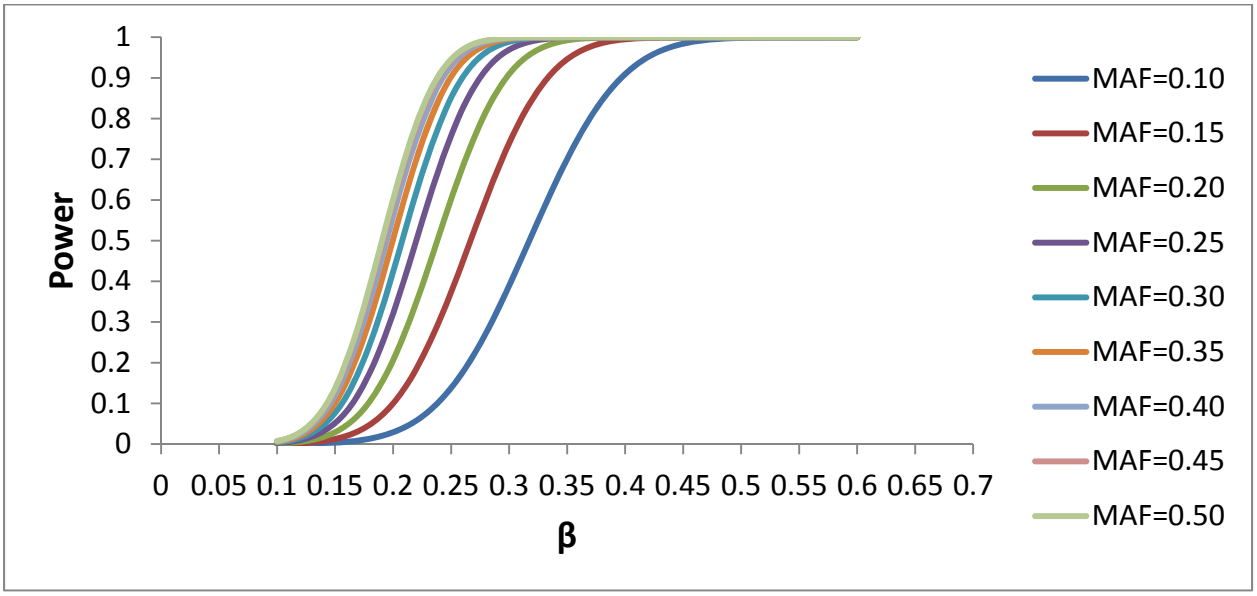
Appendix D. SNP associations with ANM in African American women from the PAGE Study.

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA	SE	P VALUE
19	rs189596789	*LDLR*	upstream	A	0.006	1.09	0.20	4.98E-08
11	rs79972789	*KCNQ1*	intronic	C	0.997	-1.76	0.34	1.90E-07
5	rs181686584	*COL4A3BP*	intronic	A	0.002	2.35	0.46	2.85E-07
6	rs114158228	*CDKAL1*	intronic	A	9E-04	3.60	0.73	7.12E-07
21	rs117876865	*KCNE1*	downstream	A	9E-04	3.58	0.73	8.55E-07
10	rs11195485	*ADRA2A*	downstream	A	0.002	2.89	0.59	9.63E-07
11	rs11224401	*ARHGAP42*	intronic	A	0.997	2.20	0.45	1.13E-06
1	rs78937547	*SEC16B*	downstream	A	0.992	-1.97	0.41	1.89E-06
17	rs75394140	*KCNJ2*	downstream	A	0.002	-0.93	0.21	6.48E-06
11	rs76988592	*KCNJ1*	downstream	A	0.702	-0.93	0.21	7.24E-06
3	rs114451007	*PPARG*	intronic	A	0.253	1.70	0.38	9.30E-06
12	rs10846771	*DHX37*	downstream	A	0.997	-0.16	0.04	9.43E-06
11	rs12804247	*CCDC81*	upstream	A	0.655	0.17	0.04	1.45E-05
1	rs76571116	*SEC16B*	downstream	A	3E-04	-1.54	0.36	1.57E-05
17	rs17634167	*TTLL6*	cds-synon.	A	6E-04	-0.34	0.08	1.62E-05
7	rs117382431	*FKBP6*	downstream	A	0.999	4.38	1.03	2.17E-05
6	rs76294174	*LOC100130357*	intronic	C	3E-04	4.38	1.03	2.17E-05
6	rs74918542	*SCGN*	intronic	A	0.999	-4.38	1.03	2.17E-05
1	rs76078015	*NOS1AP*	intronic	A	9E-04	4.38	1.03	2.17E-05
18	rs117454233	*MC4R*	downstream	A	0.999	-4.38	1.03	2.17E-05
3	rs73025249	*PPARG*	intronic	A	9E-04	4.38	1.03	2.17E-05
3	rs182857216	*ETV5*	intronic	A	0.999	-4.38	1.03	2.17E-05
3	rs73027210	*PPARG*	intronic	A	9E-04	4.38	1.03	2.17E-05
9	rs75220302	*CDKN2A*	downstream	A	0.999	-4..38	1.03	2.18E-05
9	rs74599268	*CDKN2B*	upstream	A	3E-04	4.38	1.03	2.18E-05
9	rs3731245	*CDKN2A*	intronic	A	3E-04	4.38	1.03	2.18E-05
9	rs76774391	*CDKN2B*	upstream	C	3E-04	4.38	1.03	2.18E-05
2	rs117258126	*IRS1*	downstream	A	3E-04	4.38	1.03	2.18E-05
9	rs3808846	*CDKN2B*	5' flanking	A	3E-04	4.38	1.03	2.18E-05
9	rs77706751	*CDKN2B*	upstream	A	6E-04	4.38	1.03	2.18E-05
9	rs3808845	*CDKN2B*	5' flanking	A	3E-04	4.38	1.03	2.18E-05
9	rs76810097	*CDKN2B*	upstream	A	3E-04	4.38	1.03	2.18E-05
9	rs36228836	*CDKN2A*	5' flanking	A	3E-04	4.38	1.03	2.18E-05
9	rs75039118	*ADAMTS13*	intronic	A	0.999	-4.38	1.03	2.19E-05
18	rs75914913	*MC4R*	downstream	A	3E-04	4.38	1.03	2.19E-05
11	rs190060931	*BUD13*	downstream	A	0.999	-4.38	1.03	2.21E-05
2	rs186397905	*IRS1*	downstream	C	3E-04	4.38	1.03	2.21E-05
16	rs9934222	*JPH3*	cds-synon.	A	0.163	-0.19	0.04	2.28E-05
15	rs72751410	*MAP2K5*	intronic	A	0.998	-1.51	0.36	2.30E-05
15	rs72747452	*LOC100506686*	intronic	A	0.002	1.51	0.36	2.30E-05
11	rs180751580	*NUCB2*	missense	C	0.999	-4.36	1.03	2.30E-05
3	rs186437034	*SCN5A*	intronic	A	0.999	-2.46	0.58	2.45E-05
7	rs78912482	*JAZF1*	upstream	A	0.012	0.64	0.15	3.04E-05
1	rs116071515	*SEC16B*	intronic	A	0.002	1.88	0.45	3.06E-05
6	rs1997770	*OFCC1*	downstream	A	0.970	-0.41	0.10	3.55E-05
7	rs118135044	*DGKB*	upstream	A	4E-04	4.22	1.02	3.73E-05
11	rs74402657	*ARFGAP2*	intronic	C	4E-04	2.93	0.72	3.96E-05
1	rs117217277	*SEC16B*	downstream	A	0.999	-2.97	0.72	3.97E-05

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA	SE	P VALUE
1	rs116881786	*SEC16B*	downstream	A	0.999	-2.97	0.72	3.97E-05
1	rs76471454	*SEC16B*	downstream	A	6E-04	2.97	0.72	3.97E-05
1	rs79775735	*SEC16B*	downstream	A	6E-04	2.97	0.72	3.97E-05
1	rs79468804	*SEC16B*	downstream	A	6E-04	2.97	0.72	3.97E-05
1	rs74703854	*SEC16B*	downstream	A	0.999	-2.97	0.72	3.97E-05
1	rs116923068	*SEC16B*	downstream	C	0.999	-2.97	0.72	3.97E-05
1	rs117674205	*SEC16B*	downstream	C	0.999	-2.97	0.72	3.97E-05
1	rs117260315	*SEC16B*	downstream	A	6E-04	2.97	0.72	3.97E-05
1	rs76020919	*SEC16B*	downstream	A	6E-04	2.97	0.72	3.97E-05
11	rs2306034	*LRP4*	UTR-3′	A	4E-04	2.94	0.72	3.99E-05
2	rs189110944	*IRS1*	downstream	A	4E-04	4.17	1.02	4.72E-05
5	rs1976311	*KCNN2*	upstream	C	0.996	-1.02	0.25	4.98E-05
7	rs13245084	*LOC100507421*	intronic	A	4E-04	4.14	1.02	5.07E-05
6	rs115178932	*LRRC16A*	intronic	A	4E-04	4.14	1.02	5.07E-05
1	rs77353590	*SYF2*	downstream	A	0.009	0.74	0.18	5.42E-05
2	rs111826230	*APOB*	upstream	A	0.984	-0.58	0.14	5.47E-05
11	rs193030163	*DDB2*	upstream	C	0.999	-4.11	1.02	5.57E-05
11	rs114702513	*KCNQ1*	intronic	A	0.996	-1.23	0.31	5.60E-05
6	rs117124693	*PHACTR1*	intronic	A	0.999	-4.11	1.02	5.62E-05
6	rs181947983	*SLC17A3*	upstream	A	4E-04	4.11	1.02	5.62E-05
15	rs183951867	*CHRNB4*	upstream	A	9E-04	4.11	1.02	5.62E-05
9	rs191930498	*CDKN2B*	upstream	C	4E-04	4.10	1.02	5.83E-05
17	rs192656758	*CCT6B*	downstream	A	4E-04	4.10	1.02	5.86E-05
7	rs740259	*JAZF1*	5′ flanking	A	4E-04	4.09	1.02	5.97E-05
1	rs114389068	*GPR153*	cds-synon.	A	0.005	0.93	0.23	6.07E-05
11	rs185476610	*KCNQ1*	intronic	A	0.999	-4.08	1.02	6.24E-05
16	rs246192	*NDRG4*	intronic	C	0.256	0.15	0.04	6.25E-05
7	rs192457106	*JAZF1*	intronic	A	0.999	-4.08	1.02	6.35E-05
7	rs73702566	*WBSCR22*	intronic	A	0.999	-4.08	1.02	6.35E-05
6	rs187190790	*TAP2D*	upstream	A	0.999	-4.08	1.02	6.38E-05
7	rs74984879	*DGKB*	upstream	C	0.999	-2.04	0.51	6.40E-05
11	rs184056970	*ARAP1*	intronic	A	4E-04	4.07	1.02	6.53E-05
3	rs76909367	*COLQ*	intronic	A	4E-04	4.06	1.02	6.89E-05
10	rs11187795	*PLCE1*	intronic	A	4E-04	4.06	1.02	6.93E-05
6	rs186129489	*TFAP2D*	intronic	A	4E-04	4.05	1.02	7.12E-05
2	rs73923981	*BRE*	intronic	A	9E-04	4.05	1.02	7.32E-05
15	rs180807356	*ADAMTS7*	upstream	A	0.999	-4.04	1.02	7.52E-05
5	rs10062135	*NPR3*	intronic	A	0.009	0.73	0.19	7.85E-05
12	rs17568045	*C12orf42*	intronic	A	0.993	-0.86	0.22	8.11E-05
1	rs116411856	*WARS2*	upstream	A	0.003	1.32	0.34	8.16E-05
1	rs78696400	*LYPLAL1*	downstream	A	0.985	-0.58	0.15	8.96E-05
15	rs74979292	*C15orf39*	upstream	A	0.002	1.49	0.38	9.29E-05
11	rs144204188	*TRIM66*	intronic	A	0.002	2.79	0.72	9.39E-05
1	rs78411379	*TBX15*	intronic	A	0.999	-2.27	0.58	9.62E-05
15	rs190893945	*ADAMTSL3*	intronic	A	0.998	-1.76	0.45	9.67E-05
9	rs12555547	*CDKN2B*	upstream	C	0.998	-2.30	0.59	9.69E-05
2	rs10932320	*C2orf67*	intronic	A	0.807	-0.17	0.04	9.93E-05

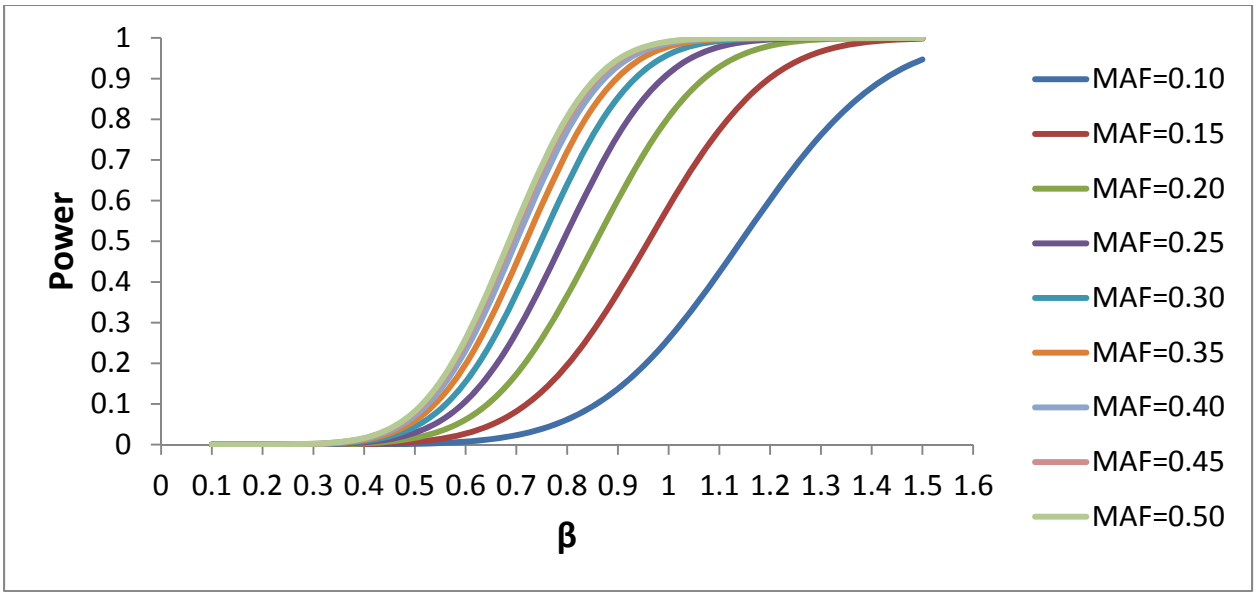
Tests of association at p≤1x10-04 from individual SNP linear regressions adjusted for study site and principal components in 1,860 African American women from the PAGE Study are shown. For each significant test of association, the chromosome, rs number, nearest gene, location, coded allele, beta, standard error, and p-value are given. Genes listed are nearest genes to the SNP as measured from the transcription start site for upstream SNPs or the transcription stop site for downstream SNPs. Abbreviations: single nucleotide polymorphism (SNP), age at natural menopause (ANM), Population Architecture using Genomics and Epidemiology (PAGE), chromosome (CHR), coded allele frequency (CAF), standard error (SE).

Appendix E. Power calculations for AM analysis for women in the PAGE Study.



Shown are power calculations for the age at menarche study of African American women in the PAGE Study calculated using QUANTO. Models were for 4,159 independent individuals and a continuous trait, gene-only hypothesis assuming an additive model of inheritance and a 3.1×10^{-7} two-sided significance level. Abbreviations: age at menarche, AM; Population Architecture using Genomics and Epidemiology, PAGE.

Appendix F. Power calculations for ANM analysis for African American women in the PAGE Study.



Shown are power calculations for the age at natural menopause study of African American women in the PAGE Study calculated using QUANTO. Models were for 1,860 independent individuals and a continuous trait, gene-only hypothesis assuming an additive model of inheritance and a 3.1×10^{-7} two-sided significance level. Abbreviations: age at natural menopause, ANM; Population Architecture using Genomics and Epidemiology, PAGE.

Appendix G. SNPs analyzed in endometrial cancer meta-analysis.

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs7538876	1	*PADI4, PADI6, RCC2, ARHGEF10L*	Basal cell carcinoma	Stacey SN	2008
rs1866967	1	*PTPRU*	Prostate cancer	Tao S	2012
rs903263	1	*PRKACB*	Male breast cancer	Orr N	2012
rs11249433	1	intergenic	Breast cancer	Thomas G	2009
rs4072037	1	*MUC1*	Esophageal & gastric cancer	Abnet CC	2010
rs2808630	1	*CRP*	Lung cancer	Amos CL	2008
rs3790844	1	*NR5A2*	Pancreatic cancer	Petersen GM	2010
rs6687758	1	*DUSP10*	Colorectal cancer	Houlston RS	2010
rs801114	1	*RHOU*	Basal cell carcinoma	Stacey SN	2008
rs1465618	2	*THADA*	Prostate cancer	Eeles RA	2009
rs10490113	2	intergenic	Breast cancer	Murabito JM	2007
rs721048	2	*EHBP1*	Prostate cancer	Gudmundsson J	2008
rs6545977	2	*EHBP1*	Prostate cancer	Eeles RA	2009
rs4254535	2	intergenic	Lung cancer	Broderick P	2009
rs10187424	2	*GGCX, VAMP8, VAMP5, RNF181*	Prostate cancer	Kote-Jarai Z	2011
rs12615966	2	*LOC284998*	Pancreatic cancer	Low SK	2010
rs3789080	2	*ACOXL*	Prostate cancer	Tao S	2012
rs12621278	2	*ITGA6*	Prostate cancer	Eeles RA	2009
rs16867225	2	*CWC22*	Prostate cancer	Tao S	2012
rs13398206	2	*PLCL1*	Prostate cancer	Tao S	2012
rs6435862	2	*BARD1*	Neuroblastoma (high-risk)	Capasso M	2009
rs13387042	2	intergenic	Breast cancer	Stacey SN	2007
rs966423	2	*DIRC3*	Thyroid cancer	Gudmundsson J	2012
rs1656402	2	*EIF4E2*	Lung cancer	Sato Y	2010
rs11892031	2	*UGT1A*	Bladder cancer	Rothman N	2010
rs7584330	2	intergenic	Prostate cancer	Kote-Jarai Z	2011
rs2292884	2	*MLPH*	Prostate cancer	Schumacher FR	2011
rs975334	3	*CNTN4*	Gallbladder cancer	Cha PC	2012
rs4973768	3	*SLC4A7*	Breast cancer	Turnbull C	2010
rs1530057	3	*RBMS3*	Lung cancer	Broderick P	2009
rs2660753	3	intergenic	Prostate cancer	Eeles RA	2008
rs17023900	3	intergenic	Prostate cancer	Cheng I	2012
rs17181170	3	LINC00506	Prostate cancer	Eeles RA	2009
rs10934853	3	*EEFSEC*	Colorectal cancer	Spain SL	2009
rs6763931	3	*ZBTB38*	Prostate cancer	Kote-Jarai Z	2011
rs6788895	3	*SIAH2*	Breast cancer	Elgazzar S	2012
rs2665390	3	*TIPARP*	Ovarian cancer	Goode EL	2010
rs10936599	3	*MYNN*	Colorectal cancer	Houlston RS	2010

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs710521	3	*TP63*	Bladder cancer	Kiemeney LA	2008
rs7626795	3	*IL1RAP*	Lung cancer	Amos CI	2008
rs4927850	3	*TFRC*	Pancreatic cancer	Wu C	2011
rs735172	4	*EVC*	Prostate cancer	Tao S	2012
rs12500426	4	*PDLIM5*	Prostate cancer	Eeles RA	2009
rs17021918	4	*PDLIM5*	Prostate cancer	Eeles RA	2009
rs1789924	4	*ADH1C*	Upper aerodigestive tract cancers	McKay JD	2011
rs971074	4	*ADH7*	Upper aerodigestive tract cancers	McKay JD	2011
rs7679673	4	*TET2*	Prostate cancer	Eeles RA	2009
rs9790517	4	*TET2*	Breast Cancer	Michailidou K	2013
rs10069690	5	*TERT*	Breast cancer	Haiman CA	2011
rs2736100	5	*TERT*	Testicular germ cell cancer	Turnbull C	2010
rs4635969	5	*TERT*	Pancreatic cancer	Diergaarde B	2010
rs4975616	5	*CLPTM1L*	Lung cancer	Broderick P	2009
rs402710	5	*TERT, CLPTM1L*	Lung cancer	McKay JD	2008
rs401681	5	*CLPTM1L*	Lung cancer	Wang Y	2008
rs12653946	5	intergenic	Prostate cancer	Takata R	2010
rs6879627	5	*LOC731559*	Pancreatic cancer	Low SK	2010
rs2121875	5	*FGF10*	Prostate cancer	Kote-Jarai Z	2011
rs4415084	5	intergenic	Breast cancer	Fletcher O	2011
rs7716600	5	*MRPS30*	Breast cancer	Li J	2010
rs981782	5	intergenic	Breast cancer	Easton DF	2007
rs16886165	5	*MAP3K1*	Breast cancer	Thomas G	2009
rs889312	5	*MAP3K1*	Breast cancer	Easton DF	2007
rs10940579	5	*ACTBL2*	Prostate cancer	Tao S	2012
rs10052657	5	*PDE4D*	Esophageal cancer	Wu C	2011
rs7717572	5	*CD180*	Prostate cancer	Tao S	2012
rs6869388	5	*KIAA0825*	Gallbladder cancer	Cha PC	2012
rs4624820	5	*SPRY4*	Testicular germ cell cancer	Rapley EA	2009
rs6556756	5	intergenic	Breast cancer	Murabito JM	2007
rs9502893	6	*FOXQ1*	Pancreatic cancer	Low SK	2010
rs10456809	6	*KIF13A*	Prostate cancer	Tao S	2012
rs2523395	6	*LOC285830*	Prostate cancer	Tao S	2012
rs130067	6	*CCHCR1*	Prostate cancer	Kote-Jarai Z	2011
rs3117582	6	*BAT3,MSH5*	Lung cancer	Wang Y	2008
rs1321311	6	*CDKN1A*	Colorectal cancer	Dunlop MG	2012
rs10498792	6	*PKHD1*	Prostate cancer	Murabito JM	2007
rs763780	6	*IL17F*	Pancreatic cancer	Innocenti F	2011

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs9363918	6	*BAI3*	Pancreatic cancer	Wu C	2011
rs17530068	6	*FAM46A, C6orf37*	Breast cancer	Siddiq A	2012
rs339331	6	*GPRC6A,RFX6*	Prostate cancer	Takata R	2010
rs2180341	6	*ECHDC1,RNF146*	Breast cancer	Gold B	2008
rs3757318	6	*ESR1,C6orf97*	Breast cancer	Turnbull C	2010
rs3734805	6	*ESR1*	Breast cancer	Fletcher O	2011
rs2046210	6	*ESR1, C6orf97*	Breast cancer	Zheng W	2009
rs651164	6	intergenic	Prostate cancer	Eeles RA	2009
rs9364554	6	*SLC22A3*	Prostate cancer	Eeles RA	2008
rs7758229	6	*SLC22A3*	Colorectal cancer	Cui R	2011
rs3016539	6	*PARK2*	Pancreatic cancer	Low SK	2010
rs12155172	7	intergenic	Prostate cancer	Eeles RA	2009
rs10486567	7	*JAZF1*	Prostate cancer	Thomas G	2008
rs7789197	7	*INHBA*	Prostate cancer	Tao S	2012
rs10263639	7	intergenic	Breast cancer	Murabito JM	2007
rs6465657	7	*LMTK2*	Prostate cancer	Eeles RA	2008
rs9649213	7	*BAIAP2L1*	Prostate cancer	Tao S	2012
rs1495741	8	*NAT2*	Bladder cancer	Rothman N	2010
rs1512268	8	*NKX3.1*	Prostate cancer	Eeles RA	2009
rs10503733	8	intergenic	Prostate cancer	Cheng I	2012
rs7832232	8	*RNF5P1*	Pancreatic cancer	Low SK	2010
rs16892766	8	*EIF3H*	Colorectal cancer	Tomlinson IP	2008
rs10088262	8	*FAM91A1*	Pancreatic cancer	Low SK	2010
rs1016343	8	intergenic	Prostate cancer	Eeles RA	2008
rs13252298	8	intergenic	Prostate cancer	Schumacher FR	2011
rs1456315	8	intergenic	Prostate cancer	Takata R	2010
rs13254738	8	*PRNCR1*	Prostate cancer	Cheng I	2012
rs6983561	8	intergenic	Prostate cancer	Cheng I	2012
rs10505483	8	intergenic	Prostate cancer	Cheng I	2012
rs16902094	8	intergenic	Prostate cancer	Gudmundsson J	2009
rs445114	8	intergenic	Prostate cancer	Gudmundsson J	2009
rs13281615	8	intergenic	Breast cancer	Easton DF	2007
rs1562430	8	intergenic	Breast cancer	Turnbull C	2010
rs10505477	8	*ORF DQ515897*	Colorectal cancer	Zanke BW	2007
rs6983267	8	*POU5F1B*	Lung cancer	Spinola M	2007
rs7014346	8	*POU5FIP1, HsG57825, DQ515897*	Colorectal cancer	Tenesa A	2008
rs4242382	8	intergenic	Prostate cancer	Thomas G	2008

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs7837688	8	intergenic	Prostate cancer	Takata R	2010
rs9642880	8	*MYC, BC042052*	Bladder cancer	Kiemeney LA	2008
rs10088218	8	*MYC,THEM75*	Ovarian cancer	Goode EL	2010
rs2294008	8	*PSCA*	Bladder cancer	Wu X	2009
rs7040024	9	*DMRT1*	Testicular cancer	Kanetsky PA	2011
rs755383	9	*DMRT1*	Testicular germ cell cancer	Turnbull C	2010
rs3814113	9	*BNC2, LOC648570, CNTLN*	Ovarian cancer	Song H	2009
rs1412829	9	intergenic	Glioma (high-grade)	Wrensch M	2009
rs1011970	9	*CDKN2A,CDKN2B*	Breast cancer	Turnbull C	2010
rs11141915	9	*DAPK1*	Pancreatic cancer	Kiyotani K	2012
rs965513	9	*FOXE1*	Thyroid cancer	Gudmundsson J	2009
rs817826	9	*RAD23B, KLF4*	Prostate cancer	Xu J	2012
rs7847271	9	*TNC*	Prostate cancer	Tao S	2012
rs505922	9	*ABO*	Pancreatic cancer	Amundadottir L	2009
rs10795668	10	intergenic	Colorectal cancer	Tomlinson IP	2008
rs3123078	10	intergenic	Prostate cancer	Eeles RA	2009
rs10993994	10	*MSMB*	Prostate cancer	Thomas G	2008
rs1926203	10	*ACTA2*	Lung cancer	Broderick P	2009
rs2274223	10	*PLCE1*	Esophageal cancer	Wu C	2011
rs3750817	10	*FGFR2*	Breast cancer	Elgazzar S	2012
rs2981579	10	*FGFR2*	Breast cancer	Thomas G	2009
rs1219648	10	*FGFR2*	Breast cancer	Hunter DJ	2007
rs2981582	10	*FGFR2*	Breast cancer	Easton DF	2007
rs10510102	10	*FGFR2*	Breast cancer	Fletcher O	2011
rs4962416	10	*CTBP2*	Prostate cancer	Thomas G	2008
rs3817198	11	*LSP1*	Breast cancer	Easton DF	2007
rs909116	11	*LSP1*	Breast cancer	Turnbull C	2010
rs7127900	11	*IGF2, IGF2AS, INS, TH*	Prostate cancer	Eeles RA	2009
rs11228565	11	intergenic	Prostate cancer	Gudmundsson J	2009
rs7931342	11	intergenic	Prostate cancer	Eeles RA	2008
rs10896449	11	intergenic	Prostate cancer	Thomas G	2008
rs7130881	11	intergenic	Prostate cancer	Eeles RA	2009
rs614367	11	*MYEOV,CCND1,ORAOV1,FGF19,FGF4,FGF3*	Breast cancer	Turnbull C	2010
rs3802842	11	intergenic	Colorectal cancer	Tenesa A	2008
rs11062040	12	*DCP1B*	Pancreatic cancer	Innocenti F	2011
rs2900174	12	*PRB2*	Pancreatic cancer	Innocenti F	2011
rs2711721	12	*AMIGO2*	Prostate cancer	Tao S	2012

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs4489787	12	*ANP32D*	Prostate cancer	Tao S	2012
rs10875943	12	*PRPH*	Prostate cancer	Kote-Jarai Z	2011
rs11169552	12	*DIP2B, ATF1*	Colorectal cancer	Houlston RS	2010
rs902774	12	*KRT8, EIF4B, TENC1*	Prostate cancer	Schumacher FR	2011
rs1878022	12	*CMKLR1*	Prostate cancer	FitzGerald LM	2011
rs9600079	13	intergenic	Prostate cancer	Takata R	2010
rs9543325	13	*KLF5, KLF12*	Pancreatic cancer	Petersen GM	2010
rs1886449	13	*LOC730242*	Pancreatic cancer	Low SK	2010
rs2039553	13	*NDFIP2*	Pancreatic cancer	Low SK	2010
rs16944141	13	*MIR622*	Prostate cancer	Tao S	2012
rs1926657	13	*ABCC4*	Breast cancer	Murabito JM	2007
rs17450420	13	*SLC10A2*	Esophageal cancer	Wu C	2012
rs1243647	14	*RNASE9*	Prostate cancer	Tao S	2012
rs944289	14	*NKX2-1*	Thyroid cancer	Gudmundsson J	2009
rs4444235	14	*BMP4*	Colorectal cancer	Houlston RS	2008
rs1314913	14	*RAD51B*	Male breast cancer	Orr N	2012
rs3784099	14	*RAD51L1*	Breast cancer	Shu XO	2012
rs999737	14	*RAD51L1*	Breast cancer	Thomas G	2009
rs3850370	14	*SKIIP, SNW1, ALKBH1, NRXN3*	Lung cancer	Hu L	2012
rs4322600	14	*GALC*	Breast cancer	Chen F	2012
rs2400997	14	*MIR656*	Prostate cancer	Tao S	2012
rs748404	15	*TGM5*	Lung cancer	Broderick P	2009
rs1876206	15	*FBN1*	Breast cancer	Murabito JM	2007
rs8034191	15	*CHRNA3,CHRNA5, CHRNB4, PSMA4, LOC123688*	Lung cancer	Hung RJ	2008
rs1051730	15	intergenic	Lung cancer	McKay JD	2008
rs8042374	15	*CHRNA3*	Lung cancer	Wang Y	2008
rs9635542	16	*PPL*	Lung cancer	Wei S	2012
rs8057939	16	*C16orf78*	Prostate cancer	Tao S	2012
rs3803662	16	*TNRC9*	Breast cancer	Stacey SN	2007
rs3112612	16	*TOX3*	Breast cancer	Fletcher O	2011
rs9929218	16	*CDH1*	Colorectal cancer	Houlston RS	2008
rs9934948	16	intergenic	Breast cancer	Shu XO	2012
rs4924935	17	*PRPSAP2*	Pancreatic cancer	Low SK	2010
rs2257205	17	*RNF43*	Pancreatic cancer	Low SK	2010
rs16951095	18	*LAMA1*	Lung cancer	Yoon KA	2010
rs2847281	18	*PTPN2*	Esophageal cancer (squamous cell)	Wu C	2012
rs998124	18	*MIR4319*	Prostate cancer	Tao S	2012

SNP	Chr	Gene	Cancer type	GWAS author	GWAS year
rs4939827	18	*SMAD7*	Colorectal cancer	Broderick P	2007
rs7504990	18	*DCC*	Gallbladder cancer	Cha PC	2012
rs1978503	18	intergenic	Breast cancer	Murabito JM	2007
rs8170	19	*C19orf62,MERIT40*	Ovarian cancer	Bolton KL	2010
rs2363956	19	*ANKLE1*	Ovarian cancer	Bolton KL	2010
rs8102137	19	*CCNE1*	Bladder cancer	Rothman N	2010
rs10411210	19	*RHPN2*	Colorectal cancer	Houlston RS	2008
rs8102476	19	intergenic	Prostate cancer	Gudmundsson J	2009
rs2735839	19	*KLK3*	Prostate cancer	Eeles RA	2008
rs103294	19	intergenic	Prostate cancer	Xu J	2012
rs961253	20	intergenic	Colorectal cancer	Houlston RS	2008
rs4925386	20	*LAMA5*	Colorectal cancer	Houlston RS	2010
rs6010620	20	*RTEL1*	Glioma (high-grade)	Wrensch M	2009
rs4809324	20	*RTEL1*	Glioma (high-grade)	Wrensch M	2009
rs372883	21	*BACH1*	Pancreatic cancer	Wu C	2011
rs458685	21	*GRIK1*	Breast cancer	Murabito JM	2007
rs1209950	21	*ETS2*	Lung cancer	Sato Y	2010
rs9981861	21	*DSCAM*	Breast cancer	Li J	2010
rs1547374	21	*TFF1*	Pancreatic cancer	Wu C	2011
rs5751168	22	*ZNF280B*	Prostate cancer	Tao S	2012
rs6005451	22	*MN1*	Prostate cancer	Tao S	2012
rs738722	22	*CHEK2,HSCB*	Esophageal cancer and gastric cancer	Abnet CC	2010
rs2239815	22	*XBP1*	Esophageal cancer (squamous cell)	Wu C	2012
rs1014971	22	*CBX6,APOBEC3A*	Bladder cancer	Rothman N	2010
rs9623117	22	*TNRC6B*	Prostate cancer	Sun J	2009
rs5759167	22	intergenic	Prostate cancer	Eeles RA	2009
rs5945572	X	*NUDT10, NUDT11, LOC340602*	Prostate cancer	Gudmundsson J	2008

Shown are SNPs associated with cancer in previous GWAS that were genotyped in EAGLE BioVU for association with endometrial cancer. Chromosome location, previous cancer association, first author and year of the first GWAS to publish an association of the SNP with cancer. Abbreviations: single nucleotide polymorphism (SNP), chromosome (Chr), genome-wide association study (GWAS). Adapted from Setiawan VW et al. 2014.

Appendix H. SNP associations with endometrial cancer in EAGLE BioVU.

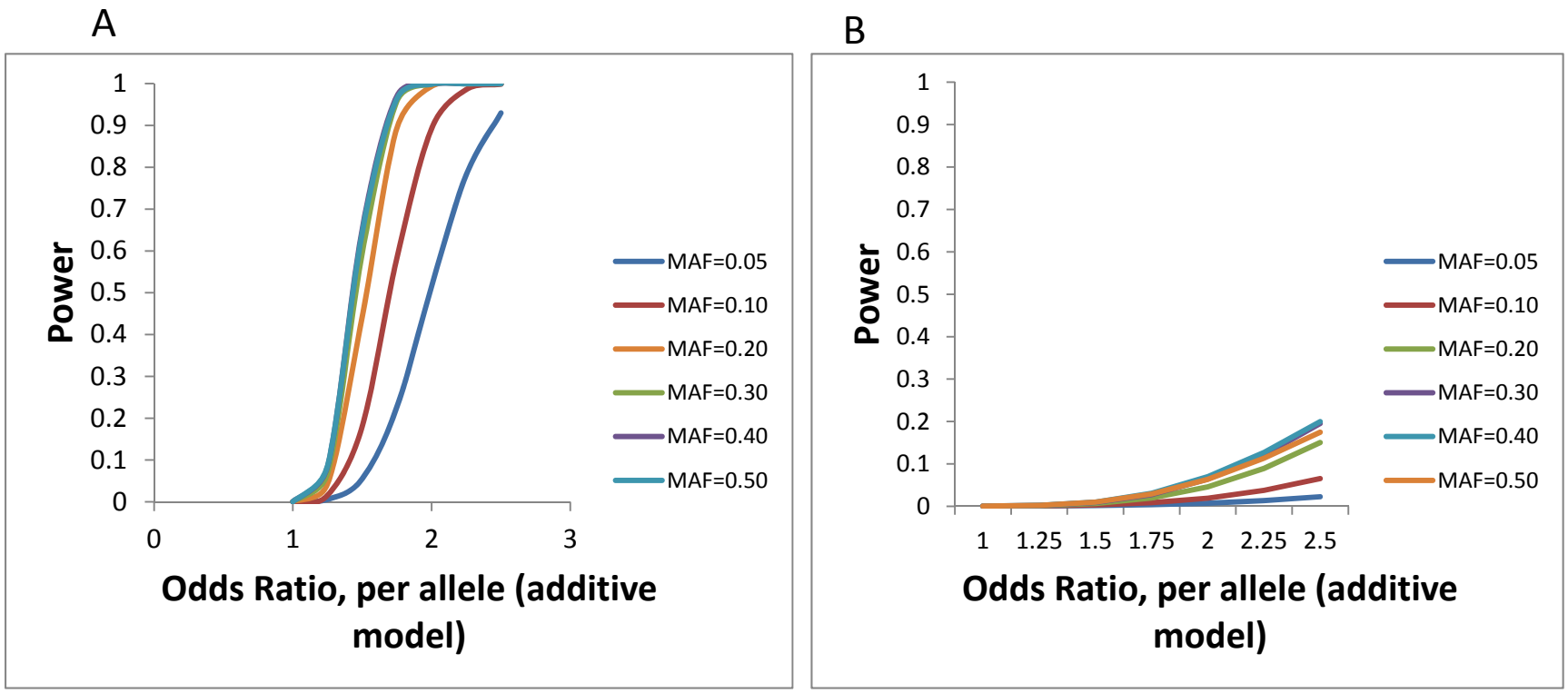
CHR	SNP	GENE	A1	TEST	NMISS	OR	SE	P-value
8	RS10086908	intergenic	T	ADD	92	8.059	0.8012	0.009196
8	RS10090154	intergenic	A	ADD	92	3.555	0.7785	0.1033
19	RS10411210	RHPN2	C	ADD	92	4.339	1.278	0.251
2	RS1045485	CASP8	G	ADD	92	0.5402	0.6592	0.3503
7	RS10486567	JAZF1	G	ADD	92	1.025	0.6064	0.9671
2	RS10490113	intergenic	C	ADD	92	0.207	1.048	0.1327
8	RS10505477	ORF DQ515897	A	ADD	92	0.2604	0.5934	0.02338
15	RS1051730	intergenic	A	ADD	92	1.251	0.4745	0.6369
12	RS10778826	PPFIA2	A	ADD	92	4.407	0.6341	0.01933
10	RS10795668	intergenic	G	ADD	92	2.54	0.666	0.1616
11	RS10896449	intergenic	G	ADD	91	1.319	0.4948	0.5755
10	RS10993994	MSMB	T	ADD	92	0.9444	0.5785	0.9212
11	RS11228565	intergenic	A	ADD	92	0.9837	0.5761	0.9772
1	RS11249433	intergenic	C	ADD	92	1.994	0.4832	0.1531
16	RS11649338	intergenic	C	ADD	92	1.973	0.5418	0.2098
17	RS11649743	HNF1B	G	ADD	92	2.847	0.8802	0.2346
16	RS11861609	CDH13	C	ADD	92	1.347	0.4561	0.5139
7	RS12155172	intergenic	A	ADD	92	1.183	0.5354	0.7539
11	RS12418451	RP11-554A11.8	A	ADD	89	1.442	0.5311	0.4909
4	RS12500426	PDLIM5	A	ADD	92	1.842	0.5397	0.2576
8	RS12543663	intergenic	C	ADD	92	0.4964	0.5742	0.2225
2	RS12621278	ITGA6	A	ADD	92	8.56E+07	7454	0.998
8	RS13254738	PRNCR1	C	ADD	89	0.8486	0.5218	0.7531
8	RS13281615	intergenic	T	ADD	92	0.6143	0.5009	0.3307
2	RS13387042	intergenic	A	ADD	92	0.4301	0.4724	0.07411
11	RS1393350	TYR	A	ADD	92	1.458	0.5436	0.4878
2	RS1465618	THADA	T	ADD	92	7.129	0.8207	0.01669
8	RS1512268	intergenic	T	ADD	89	0.6309	0.5398	0.3935
3	RS1530057	RBMS3	A	ADD	92	0.6356	1.009	0.6534
9	RS1571801	DAB2IP	T	ADD	92	1.6	0.4831	0.3308
5	RS16891982	SLC45A2	G	ADD	89	6.37E+08	1.01E+04	0.9984
8	RS16892766	EIF3H	C	ADD	89	0.3024	1.234	0.3323
8	RS16901979	intergenic	A	ADD	88	6.44E-09	1.05E+04	0.9986
8	RS16902094	intergenic	G	ADD	91	1.218	0.6842	0.7732
4	RS17021918	PDLIM5	C	ADD	92	1.839	0.6172	0.3235
15	RS1876206	FBN1	G	ADD	89	0.4745	0.7257	0.3043
10	RS1926203	ACTA2	T	ADD	89	1.53	0.5566	0.4452
13	RS1926657	ABCC4	T	ADD	92	2.166	0.5809	0.1832
18	RS1978503	intergenic	G	ADD	91	1.092	0.656	0.893
6	RS2046210	ESR1, C6orf97	A	ADD	91	2.623	0.6246	0.1227

CHR	SNP	GENE	A1	TEST	NMISS	OR	SE	P-value
17	RS2075555	COL1A1	T	ADD	92	1.394	0.7214	0.6453
22	RS2284063	PLA2G6	A	ADD	92	1.028	0.5285	0.9584
20	RS2296241	CYP24A1	G	ADD	92	1.023	0.5002	0.9644
19	RS25487	XRCC1	A	ADD	92	1.009	0.5619	0.9876
16	RS258322	CDK10	A	ADD	92	0.9609	0.8933	0.9644
3	RS2660753	intergenic	T	ADD	92	1.72	0.8109	0.5037
19	RS266849	intergenic	A	ADD	91	0.7578	0.5888	0.6376
2	RS2710647	EHBP1	C	ADD	88	1.222	0.5054	0.6917
19	RS2735839	KLK3	G	ADD	92	0.6574	0.6681	0.5302
5	RS2736100	TERT	G	ADD	88	1.843	0.502	0.2234
8	RS2928679	intergenic	A	ADD	92	0.8392	0.4819	0.716
10	RS2981578	FGFR2	G	ADD	92	1.064	0.46	0.8935
10	RS2981582	FGFR2	T	ADD	92	0.8353	0.5096	0.724
6	RS3131379	MSH5-SAPCD1	T	ADD	90	1.191	0.718	0.8073
5	RS31489	CLPTM1L	C	ADD	87	1.198	0.518	0.7275
10	RS3750817	FGFR2	C	ADD	92	1.302	0.4951	0.5943
11	RS3802842	intergenic	C	ADD	92	0.8701	0.6396	0.8278
16	RS3803662	TNRC9	T	ADD	92	1.108	0.5643	0.8552
9	RS3814113	BNC2, LOC648570, CNTLN	T	ADD	92	2.508	0.6238	0.1404
11	RS3817198	LSP1	C	ADD	92	1.548	0.5195	0.4003
16	RS3863435	intergenic	C	ADD	92	1.119	0.4735	0.8122
5	RS401681	CLPTM1L	C	ADD	92	1.739	0.537	0.3027
5	RS402710	TERT, CLPTM1L	C	ADD	92	1.236	0.5136	0.6804
2	RS4254535	intergenic	C	ADD	92	0.7048	0.6259	0.5761
6	RS4324798	intergenic	A	ADD	88	0.9292	0.8394	0.9302
5	RS4415084	intergenic	T	ADD	91	0.5186	0.5361	0.2206
17	RS4430796	HNF1B	G	ADD	92	0.6604	0.5766	0.4718
14	RS4444235	BMP4	C	ADD	92	0.6297	0.5041	0.3588
21	RS458685	GRIK1	C	ADD	92	1.626	0.6899	0.4811
4	RS4588	GC	A	ADD	89	0.643	0.5992	0.461
9	RS4636294	intergenic	A	ADD	89	1.054	0.5592	0.9249
15	RS4779584	intergenic	T	ADD	92	1.223	0.5631	0.7203
16	RS4782780	CDH13	T	ADD	92	1.105	0.4615	0.8292
16	RS4785763	AFG3L1P	A	ADD	90	0.9574	0.5578	0.9378
3	RS4857841	EEFSEC	A	ADD	92	0.7802	0.582	0.6698
18	RS4939827	SMAD7	T	ADD	92	0.9456	0.5161	0.9137
8	RS4961199	intergenic	A	ADD	92	0.9528	0.6984	0.9449
10	RS4962416	CTBP2	C	ADD	92	2.734	0.6047	0.09627
3	RS4973768	SLC4A7	T	ADD	92	0.6317	0.5361	0.3916
5	RS4975616	CLPTM1L	A	ADD	92	2.079	0.6011	0.2234
22	RS5759167	intergenic	G	ADD	92	1.653	0.5844	0.3896

CHR	SNP	GENE	A1	TEST	NMISS	OR	SE	P-value
X	RS5945572	NUDT10, NUDT11, LOC340602	A	ADD	92	0.5652	0.5983	0.3402
X	RS5945619	intergenic	C	ADD	89	0.5694	0.5963	0.345
20	RS6068816	CYP24A1	T	ADD	92	2.637	0.8508	0.2544
8	RS620861	LOC101930033	G	ADD	91	0.3639	0.5966	0.09016
7	RS6465657	LMTK2	C	ADD	92	0.6371	0.5054	0.3724
17	RS6504950	STXBP4	G	ADD	92	0.644	0.5058	0.3842
5	RS6556756	intergenic	G	ADD	92	1.225	0.8238	0.8053
8	RS6983267	POU5F1B	G	ADD	92	0.2593	0.5855	0.02115
8	RS7000448	intergenic	T	ADD	92	0.4986	0.6317	0.2705
8	RS7014346	POU5FIP1, HsG57825, DQ515897	A	ADD	92	0.4448	0.5552	0.1445
9	RS7023329	MTAP	A	ADD	92	0.8699	0.5484	0.7994
4	RS7041	GC	T	ADD	92	0.7994	0.4734	0.6363
11	RS7117034	intergenic	T	ADD	92	0.9837	0.5761	0.9772
9	RS719725	intergenic	A	ADD	92	1.083	0.5294	0.88
12	RS731236	VDR	C	ADD	89	1.388	0.5071	0.5177
15	RS748404	TGM5	T	ADD	90	0.3624	0.5831	0.08175
17	RS7501939	HNF1B	C	ADD	92	1.401	0.5466	0.5371
3	RS7626795	IL1RAP	G	ADD	88	0.2351	1.158	0.2112
4	RS7679673	TET2	C	ADD	88	0.4859	0.5465	0.1866
8	RS7837688	intergenic	T	ADD	92	3.555	0.7785	0.1033
8	RS7841060	PRNCR1	G	ADD	92	0.404	0.7368	0.2186
15	RS8042374	CHRNA3	G	ADD	90	0.8378	0.6759	0.7934
19	RS8102476	intergenic	C	ADD	92	1.956	0.5504	0.2229
5	RS889312	MAP3K1	C	ADD	92	1.27	0.5024	0.634
20	RS910873	PIGU	A	ADD	91	0.8671	0.9447	0.88
6	RS9295740	intergenic	A	ADD	92	1.083	0.5632	0.8873
15	RS931794	HYKK	G	ADD	92	1.525	0.4834	0.383
6	RS9364554	SLC22A3	T	ADD	92	1.691	0.5365	0.3275
20	RS961253	intergenic	A	ADD	92	2.16	0.5243	0.1418
22	RS9623117	TNRC6B	C	ADD	92	2.011	0.528	0.1857
5	RS981782	intergenic	T	ADD	92	1.64	0.4911	0.3138
16	RS9929218	CDH1	G	ADD	92	1.2	0.5147	0.723
14	RS999737	RAD51L1	C	ADD	89	1.643	0.6949	0.475

Data shown are SNP associations for EAGLE BioVU endometrial cancer analysis. Gene shown is the closest gene to the SNP. Abbreviations: chromosome (CHR), single nucleotide polymorphism (SNP), risk allele (A1), number of cases and controls used in each logistic regression (NMISS), odds ratio (OR), standard error (SE).

Appendix I. Power calculations for EAGLE BioVU endometrial cancer study.



Shown are power calculations for the EAGLE BioVU endometrial cancer study calculated using QUANTO. (A) 206 cases, 11 controls/case; (B) 20 cases, 8 controls/case. Models were for unmatched case-control, gene-only hypothesis assuming a log-additive model of inheritance and a 4.2×10^{-4} two-sided significance level.

Appendix J. eMERGE Network site contributions to TSH levels study.

Site	Primary Phenotype	Total # Genotyped (n)	TSH Levels	
			European Americans (n)	African Americans (n)
Marshfield Clinic	Cataracts	4,113	1,157	0
Vanderbilt	Cardiac Conductance	2,712	284	88
Group Health	Dementia	2,532	1,167	64
Mayo Clinic	Peripheral Artery Disease	3,043	1,881	10
Northwestern	Type 2 Diabetes	1,217	12	189
Total		**13,617**	**4,501**	**351**

Primary phenotypes reflects initial GWAS phenotype investigated at each site for the eMERGE Network. Total (n) genotyped are for each site's primary phenotype GWAS. Euthyroid subjects for serum thyroid stimulating hormone (TSH) level analysis are a subset of the total number genotyped in eMERGE for the primary phenotypes. All sites contributed European Americans to the serum TSH level analysis; all sites except Marshfield Clinic contributed African Americans. Data shown are counts (n).

Appendix K. SNP associations for serum TSH levels in eMERGE study European Americans.

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P-VALUE
5	rs1382879	*PDE8B*	intronic	G	0.39	0.09 (0.01)	7.16E-18
5	rs2046045	*PDE8B*	intronic	C	0.40	0.09 (0.01)	1.85E-17
5	rs989758	*PDE8B*	intronic	T	0.36	0.08 (0.01)	1.33E-14
5	rs9687206	*PDE8B*	intronic	G	0.43	0.08 (0.01)	5.52E-14
5	rs12515498	*PDE8B*	intronic	C	0.26	0.07 (0.01)	3.27E-10
5	rs6885813	*PDE8B*	intronic	A	0.25	0.06 (0.01)	4.05E-08
5	rs1096752	*PDE8B*	intronic	A	0.45	-0.05 (0.01)	6.30E-07
5	rs13361710	*PDE8B*	intronic	T	0.24	0.06 (0.01)	6.60E-07
9	rs10759944	*FOXE1*	upstream	A	0.33	-0.05 (0.01)	1.08E-06
9	rs965513	*FOXE1*	upstream	A	0.34	-0.05 (0.01)	1.09E-06
9	rs925489	*FOXE1*	upstream	C	0.34	-0.05 (0.01)	1.79E-06
9	rs7850258	*FOXE1*	upstream	A	0.33	-0.05 (0.01)	1.85E-06
2	rs10496992	-	intergenic	G	0.38	0.05 (0.01)	2.22E-06
2	rs1861628	*IGFBP5*	upstream	T	0.27	-0.05 (0.01)	3.68E-06
5	rs4348174	*ITGA1*	upstream	C	0.40	0.05 (0.01)	3.97E-06
9	rs657152	*ABO*	intronic	T	0.38	0.05 (0.01)	4.18E-06
7	rs740083	*VWC2*	upstream	A	0.24	-0.05 (0.01)	4.56E-06
7	rs813379	*CDK14*	intronic	G	0.06	-0.10 (0.02)	4.57E-06
2	rs2712168	*IGFBP5*	upstream	C	0.13	0.07 (0.01)	4.98E-06
5	rs256438	*THBS4*	intronic	C	0.36	0.05 (0.01)	5.53E-06
18	rs4570936	-	intergenic	T	0.22	-0.05 (0.01)	5.73E-06
2	rs6546537	*AAK1*	intronic	C	0.28	-0.05 (0.01)	5.92E-06
9	rs7855088	*ANP32B*	upstream	C	0.44	-0.05 (0.01)	6.23E-06
9	rs925487	*FOXE1*	downstream	G	0.37	-0.05 (0.01)	6.24E-06
7	rs803174	*CDK14*	intronic	G	0.06	-0.10 (0.02)	6.74E-06
5	rs2438632	*THBS4*	downstream	A	0.39	0.05 (0.01)	6.88E-06
2	rs13020935	*IGFBP5*	upstream	G	0.28	-0.05 (0.01)	7.02E-06
5	rs12520862	*PDE8B*	intronic	T	0.14	0.06 (0.01)	7.48E-06
9	rs10984103	*FOXE1*	downstream	A	0.37	-0.05 (0.01)	7.81E-06
9	rs907580	*FOXE1*	downstream	A	0.27	-0.05 (0.01)	8.20E-06
8	rs2466067	*NRG1*	intronic	C	0.31	-0.05 (0.01)	8.42E-06
9	rs7870926	*ANP32B*	downstream	G	0.50	-0.04 (0.01)	8.67E-06
5	rs7341064	*ITGA1*	upstream	C	0.40	0.04 (0.01)	1.03E-05
8	rs4298457	*NRG1*	intronic	G	0.27	-0.05 (0.01)	1.07E-05
11	rs598599	*MRE11A*	intronic	A	0.28	0.05 (0.01)	1.09E-05
4	rs4693596	*COQ2*	intronic	C	0.38	-0.04 (0.01)	1.10E-05
8	rs10954859	*NRG1*	intronic	G	0.27	-0.05 (0.01)	1.12E-05
5	rs404375	*THBS4*	intronic	G	0.50	-0.04 (0.01)	1.26E-05
6	rs2983525	*PDE10A*	intronic	C	0.27	-0.05 (0.01)	1.34E-05
6	rs2983514	*PDE10A*	intronic	G	0.33	-0.05 (0.01)	1.36E-05

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P-VALUE
1	rs3766122	*SELP*	intronic	C	0.05	-0.10 (0.02)	1.42E-05
9	rs7866436	*C9orf156*	downstream	G	0.37	-0.04 (0.01)	1.52E-05
9	rs7024345	*FOXE1*	upstream	A	0.27	-0.05 (0.01)	1.65E-05
5	rs26367	*FSTL4*	intronic	G	0.10	-0.07 (0.02)	1.65E-05
5	rs10073636	*HCN1*	intronic	T	0.43	0.04 (0.01)	1.73E-05
9	rs13285674	*ASTN2*	upstream	A	0.23	0.05 (0.01)	1.90E-05
9	rs505922	*ABO*	intronic	C	0.36	0.04 (0.01)	1.94E-05
5	rs7445986	*ITGA1*	upstream	T	0.40	0.04 (0.01)	1.95E-05
2	rs10204522	*IGFBP5*	upstream	C	0.10	0.07 (0.02)	1.95E-05
6	rs4054489	*IBTK*	downstream	T	0.18	-0.05 (0.01)	1.95E-05
11	rs1055075	*TTC12*	downstream	T	0.34	-0.04 (0.01)	1.96E-05
4	rs4861534	*DCTD*	downstream	G	0.10	0.07 (0.02)	2.09E-05
15	rs7168316	*C15orf33*	intronic	T	0.23	-0.05 (0.01)	2.10E-05
9	rs7848973	*FOXE1*	upstream	A	0.40	-0.04 (0.01)	2.11E-05
12	rs3136559	*CD69*	upstream	A	0.28	0.05 (0.01)	2.13E-05
2	rs6727435	*AAK1*	intronic	A	0.27	-0.05 (0.01)	2.15E-05
5	rs33613	*FSTL4*	intronic	T	0.09	-0.07 (0.02)	2.35E-05
15	rs12592277	*C15orf33*	intronic	A	0.22	-0.05 (0.01)	2.35E-05
8	rs2466062	*NRG1*	intronic	G	0.30	-0.05 (0.01)	2.36E-05
8	rs3898456	*FAM135B*	intronic	A	0.35	0.04 (0.01)	2.56E-05
3	rs4402960	*IGF2BP2*	intronic	T	0.30	-0.05 (0.01)	2.63E-05
3	rs1470579	*IGF2BP2*	intronic	C	0.31	-0.05 (0.01)	2.67E-05
5	rs13354798	*HCN1*	intronic	C	0.43	0.04 (0.01)	2.75E-05
5	rs9686502	*PDE8B*	intronic	G	0.49	0.04 (0.01)	2.85E-05
22	rs9606756	*PDE8B*	intronic	G	0.12	0.07 (0.02)	2.86E-05
11	rs494442	*KIRREL3*	upstream	T	0.40	-0.04 (0.01)	3.03E-05
12	rs2695148	*ANAPC5*	upstream	T	0.10	-0.07 (0.02)	3.11E-05
1	rs17265852	*NFIA*	intronic	C	0.08	-0.07 (0.02)	3.16E-05
5	rs6414906	*HCN1*	intronic	C	0.43	0.04 (0.01)	3.51E-05
16	rs3813583	*WWOX*	downstream	C	0.38	0.04 (0.01)	4.06E-05
1	rs749378	*GLIS1*	downstream	A	0.27	-0.05 (0.01)	4.09E-05
5	rs6451801	*HCN1*	intronic	A	0.43	0.04 (0.01)	4.10E-05
5	rs13162651	*HCN1*	intronic	C	0.43	0.04 (0.01)	4.11E-05
6	rs12201217	*CDKAL1*	intronic	T	0.38	-0.04 (0.01)	4.30E-05
3	rs370234	*VGLL4*	upstream	T	0.39	-0.04 (0.01)	4.32E-05
12	rs1647253	*ANAPC5*	upstream	A	0.10	-0.07 (0.02)	4.57E-05
8	rs6989877	*NRG1*	downstream	T	0.13	0.06 (0.01)	4.59E-05
6	rs11963665	*FAM46A*	upstream	C	0.20	-0.05 (0.01)	4.63E-05
5	rs6892290	*HCN1*	intronic	G	0.43	0.04 (0.01)	4.76E-05
1	rs6668505	*PTAFR*	intronic	T	0.06	-0.08 (0.02)	4.88E-05
19	rs3745746	*CABP5*	missense	C	0.39	-0.04 (0.01)	4.93E-05

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P-VALUE
5	rs12521494	*PDE8B*	intronic	C	0.23	0.05 (0.01)	5.00E-05
5	rs10064949	*ITGA1*	upstream	C	0.43	0.04 (0.01)	5.07E-05
2	rs1515259	-	intergenic	T	0.45	0.04 (0.01)	5.44E-05
2	rs1012319	*IGFBP5*	upstream	T	0.19	-0.05 (0.01)	5.47E-05
6	rs2983500	*PDE10A*	intronic	T	0.11	-0.06 (0.02)	5.54E-05
18	rs8096947	-	intergenic	A	0.19	-0.05 (0.01)	5.56E-05
2	rs888186	*IGFBP5*	upstream	C	0.10	-0.07 (0.02)	5.63E-05
5	rs4703797	*THBS4*	intronic	G	0.33	0.04 (0.01)	5.65E-05
11	rs529126	*MRE11A*	intronic	A	0.26	0.04 (0.01)	5.65E-05
1	rs11805172	*SESN2*	upstream	G	0.07	-0.08 (0.02)	5.69E-05
14	rs8009673	*ARHGAP5*	upstream	A	0.15	0.06 (0.01)	5.83E-05
11	rs1939422	*C11orf87*	upstream	T	0.36	-0.04 (0.01)	5.94E-05
2	rs1986415	*AOX1*	intronic	A	0.12	0.06 (0.02)	6.05E-05
8	rs2439300	*NRG1*	intronic	A	0.27	-0.04 (0.01)	6.19E-05
8	rs2943179	*CNBD1*	intronic	T	0.22	0.05 (0.01)	6.31E-05
15	rs8035662	*MEGF11*	intronic	A	0.33	-0.04 (0.10)	6.37E-05
11	rs877138	*ANKK1*	upstream	G	0.35	-0.04 (0.01)	6.37E-05
9	rs1443434	*FOXE1*	UTR-3′	G	0.40	-0.04 (0.01)	6.53E-05
2	rs2381866	-	intergenic	C	0.44	0.04 (0.01)	6.68E-05
2	rs888182	*IGFBP5*	upstream	C	0.16	0.05 (0.01)	6.75E-05
16	rs7184757	*WWOX*	intronic	C	0.09	-0.07 (0.02)	7.22E-05
12	rs11172482	*XRCC6BP1*	downstream	C	0.37	-0.04 (0.01)	7.29E-05
7	rs39334	*RELN*	intronic	G	0.37	0.04 (0.01)	7.47E-05
11	rs12278001	*DDX10*	downstream	A	0.06	-0.08 (0.02)	7.53E-05
5	rs12654213	*HCN1*	upstream	G	0.43	0.04 (0.01)	7.69E-05
1	rs10489909	*NFIA*	intronic	A	0.05	-0.09 (0.02)	7.81E-05
7	rs13231383	*TPK1*	upstream	A	0.25	0.04 (0.01)	8.19E-05
5	rs2306344	*PDE8B*	intronic	A	0.31	-0.04 (0.01)	8.23E-05
19	rs11666426	*ZNF665*	intronic	C	0.41	0.04 (0.01)	8.30E-05
1	rs12138950	*CAPZB*	upstream	C	0.15	-0.05 (0.01)	8.97E-05
9	rs424829	*STOM*	upstream	A	0.29	0.04 (0.01)	9.02E-05
15	rs11071858	*MEGF11*	intronic	G	0.41	-0.04 (0.01)	9.33E-05
11	rs12282135	*OR52E2*	upstream	C	0.15	-0.05 (0.01)	9.47E-05
1	rs11118832	*DUSP10*	intronic	C	0.08	-0.07 (0.02)	9.52E-05
1	rs630505	*DENND2D*	intronic	C	0.27	-0.04 (0.01)	9.58E-05
2	rs16856529	*IGFBP5*	upstream	C	0.15	0.05 (0.01)	9.80E-05
12	rs1502816	*XRCC6BP1*	downstream	C	0.38	-0.04 (0.01)	9.95E-05

Tests of association using linear regression, adjusted for age, sex, principal component (PC1), and body mass index (BMI) were performed. Tests of association at p<1x10-4 are listed. Gene listed is the gene in closest proximity to the SNP. Coded allele frequency (CAF) is for the allele frequency in eMERGE European Americans in the serum TSH study (n=4,501).

Appendix L. SNP associations for serum TSH levels in eMERGE study African Americans.

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P VALUE
13	rs1409005	*POU4F1-AS1*	downstream	T	0.20	0.25 (0.05)	5.02E-07
1	rs2378497	*DUSP10*	upstream	G	0.08	0.33 (0.07)	3.53E-06
20	rs6062344	*TCEA2*	intronic	T	0.40	0.18 (0.04)	4.06E-06
16	rs270421	*WWOX*	downstream	C	0.28	0.19 (0.04)	7.75E-06
7	rs2299116	*CREB5*	intronic	A	0.17	0.25 (0.06)	8.16E-06
2	rs6728613	*MYT1L*	intronic	A	0.24	0.20 (0.04)	1.14E-05
10	rs6585018	*PDCD4*	near-5'	G	0.17	-0.22 (0.05)	1.17E-05
14	rs1013757	*TTC6*	downstream	A	0.32	-0.19 (0.04)	1.33E-05
2	rs4073401	*MYT1L*	intronic	T	0.24	0.19 (0.04)	1.33E-05
14	rs12883861	LOC728755	downstream	G	0.20	0.21 (0.05)	1.63E-05
7	rs9784959	*ABCA13*	intronic	A	0.30	-0.18 (0.04)	1.82E-05
16	rs270422	*WWOX*	downstream	A	0.29	0.18 (0.04)	2.17E-05
12	rs261875	*BICD1*	intronic	C	0.32	0.18 (0.04)	2.24E-05
7	rs274614	*GRM3*	intronic	G	0.30	-0.18 (0.04)	2.36E-05
3	rs11711934	*DNAH1*	intronic	C	0.31	-0.17 (0.04)	2.45E-05
2	rs12621889	*KIAA1715*	intronic	T	0.06	0.36 (0.08)	2.68E-05
2	rs12464144	*KIAA1715*	intronic	A	0.06	0.36 (0.08)	2.68E-05
18	rs10163845	*NETO1*	near-5'	A	0.28	-0.18 (0.04)	2.74E-05
19	rs12610504	*ZNF536*	downstream	G	0.19	0.20 (0.05)	3.07E-05
13	rs1274744	-	intergenic	C	0.42	-0.17 (0.04)	3.21E-05
5	rs10060607	*SLC36A3*	intronic	A	0.30	0.18 (0.04)	3.28E-05
18	rs1824304	*FAM59A*	intronic	C	0.37	0.17 (0.04)	3.32E-05
2	rs841452	*HS6ST1*	upstream	C	0.37	0.17 (0.04)	3.52E-05
7	rs11977108	*ABCA13*	intronic	A	0.17	-0.21 (0.05)	3.70E-05
3	rs4678798	*ARPP21*	intronic	A	0.14	0.24 (0.06)	3.71E-05
4	rs6851816	*MLF1IP*	intronic	T	0.50	0.16 (0.04)	3.83E-05
22	rs133201	*LRP5L*	5'-UTR	A	0.09	0.27 (0.06)	4.04E-05
12	rs2593996	*BICD1*	intronic	C	0.50	-0.16 (0.04)	4.09E-05
19	rs1054713	*KLK1*	cds-synon	T	0.26	0.19 (0.05)	4.16E-05
19	rs12609319	*ZNF536*	downstream	T	0.19	0.20 (0.05)	4.23E-05
3	rs1918092	*ARL8B,EDEM1*	downstream	C	0.09	0.30 (0.07)	4.90E-05
12	rs2303478	*ASCL4*	downstream	A	0.28	0.18 (0.04)	5.11E-05
1	rs3738605	*SZRD1*	3'UTR	A	0.12	0.24 (0.06)	5.12E-05
19	rs2659099	*MGC45922*	near-5'	T	0.29	0.18 (0.04)	5.13E-05
3	rs4955261	*CMTM8*	upstream	G	0.39	0.16 (0.04)	5.19E-05
13	rs4772145	*DOCK9*	downstream	T	0.43	0.15 (0.04)	5.23E-05
2	rs13403407	*C2orf43*	intronic	C	0.47	-0.16 (0.04)	5.31E-05
3	rs1513476	*ARPP21*	intronic	C	0.14	0.22 (0.05)	5.46E-05
2	rs17032566	*CAMKMT*	intronic	T	0.07	-0.30 (0.07)	5.52E-05
17	rs1105813	*DNAH2*	intronic	T	0.43	0.16 (0.04)	5.61E-05

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P VALUE
12	rs1563333	*DYNLL1*	intronic	A	0.22	-0.19 (0.05)	5.64E-05
10	rs1907356	*C10orf11*	intronic	T	0.16	-0.21 (0.05)	5.71E-05
10	rs11001788	*C10orf11*	intronic	A	0.16	-0.21 (0.05)	5.71E-05
2	rs12470895	*LOC729164*	ncRNA	T	0.20	0.21 (0.05)	5.86E-05
3	rs646929	*CACNA2D3*	intronic	C	0.08	0.30 (0.07)	5.96E-05
3	rs2335640	*DNAH1*	intronic	C	0.30	-0.17 (0.04)	5.98E-05
12	rs3742049	*COQ5*	missense	T	0.25	0.18 (0.04)	6.08E-05
3	rs17052068	*DNAH1*	intronic	T	0.30	-0.16 (0.04)	6.46E-05
1	rs2819757	*RYR2*	intronic	C	0.18	0.22 (0.05)	6.49E-05
2	rs10804139	*PARD3B*	upstream	A	0.38	-0.16 (0.04)	6.54E-05
18	rs736218	*FAM59A*	intronic	C	0.38	0.16 (0.04)	6.66E-05
9	rs10989120	*MSANTD3-TMEFF1*	intronic	A	0.39	-0.19 (0.05)	7.05E-05
12	rs10744020	*C12orf36*	downstream	C	0.23	0.16 (0.04)	7.05E-05
19	rs2659103	*KLK1*	intronic	T	0.25	0.19 (0.05)	7.29E-05
1	rs10918914	*XCL2*	downstream	G	0.17	0.22 (0.05)	7.39E-05
12	rs261878	*BICD1*	intronic	C	0.32	-0.16 (0.04)	7.42E-05
15	rs12914266	*SQRDL*	intronic	A	0.29	0.17 (0.04)	7.58E-05
7	rs6965055	*C7orf10*	intronic	G	0.39	-0.16 (0.04)	7.65E-05
7	rs7808606	*C7orf10*	intronic	C	0.39	-0.15 (0.04)	7.66E-05
14	rs17322359	*PRKD1*	upstream	T	0.10	0.25 (0.06)	7.74E-05
5	rs11949641	*MSX2*	downstream	A	0.23	0.18 (0.05)	7.89E-05
1	rs12120382	*CHRM3*	upstream	C	0.09	0.29 (0.07)	7.96E-05
2	rs6731363	*LOC729164*	ncRNA	A	0.20	0.20 (0.05)	7.99E-05
4	rs13144021	*NR3C2*	upstream	G	0.14	0.23 (0.06)	8.00E-05
18	rs877128	*MC2R*	intronic	A	0.25	0.18 (0.04)	8.10E-05
10	rs7923004	*BBIP1*	intronic	C	0.18	-0.20 (0.05)	8.19E-05
8	rs6999969	*XKR6*	intronic	C	0.42	-0.16 (0.04)	8.33E-05
11	rs1027388	*LRRC4C*	intronic	A	0.28	-0.17 (0.04)	8.36E-05
1	rs17011253	*DUSP10*	upstream-	C	0.09	0.27 (0.07)	8.38E-05
10	rs942077	*RBM20*	missense	G	0.48	-0.15 (0.04)	8.47E-05
4	rs4370216	-	intergenic	C	0.46	-0.15 (0.04)	8.55E-05
4	rs2333727	*HSFY2*	upstream	C	0.46	-0.15 (0.04)	8.55E-05
7	rs1029357	*SAMD9L*	3'-UTR	G	0.46	0.15 (0.04)	8.57E-05
9	rs1332598	*MSANTD3-TMEFF1*	intronic	A	0.23	-0.19 (0.05)	8.69E-05
5	rs6864667	*SLC12A7*	intronic	G	0.47	0.15 (0.04)	8.98E-05
12	rs4411338	*CCND2*	upstream	C	0.29	0.16 (0.04)	9.05E-05
19	rs171953	*KLK1*	downstream	G	0.48	-0.15 (0.04)	9.07E-05
19	GA035020	*SSC5D*	intronic	T	0.25	0.19 (0.05)	9.11E-05
15	rs2040578	*SV2B*	intronic	G	0.29	0.17 (0.04)	9.13E-05
17	rs1106826	*DNAH2*	intronic	A	0.32	0.17 (0.04)	9.23E-05

CHR	SNP	GENE	GENE REGION	CODED ALLELE	CAF	BETA (SE)	P VALUE
20	rs6090040	*TCEA2*	intronic	C	0.47	0.15 (0.04)	9.35E-05
20	rs4408777	*RGS19*	intronic	G	0.33	0.16 (0.04)	9.39E-05
16	rs2521676	-	intergenic	G	0.39	0.16 (0.04)	9.73E-05
1	rs16845412	-	intergenic	G	0.10	0.27 (0.07)	9.76E-05
4	rs10518306	*LOC285419*	intronic	A	0.06	0.35 (0.09)	9.78E-05
8	rs10098991	-	intergenic	C	0.44	0.16 (0.04)	9.86E-05
16	rs8059691	*EMC8*	intronic	G	0.11	0.23 (0.06)	9.90E-05

Tests of association using linear regression, adjusted for age, sex, principal component (PC1), and body mass index (BMI) were performed. Tests of association at $p<1x10^{-4}$ are listed. Gene listed is the gene in closest proximity to the SNP. Coded allele frequency (CAF) is for the allele frequency in eMERGE African Americans in the serum TSH study (n=351).

Appendix M. Comparisons of reported associations with serum TSH levels in Europeans to eMERGE African Americans.

Locus			Prior Association					Current Study					
SNP	Chr	Gene	CA	CAF	β (SE)	P-value	Ref.	SNP/Best Proxy SNP	r^2	CA	CAF	β (SE)	P-value
rs10917469	1	*CAPZB*	G	0.16	-0.16 (0.03)	3.2E-08	(Panicker et al.2010)	rs12138950	1.00	C	0.24	-0.02 (0.05)	0.64
rs10917477	1	*CAPZB*	A	0.51	-0.06 (0.01)	1.54E-08	(Rawal et al.2012)	rs6683419	0.62	G	0.49	0.01 (0.04)	0.85
rs10799824	1	*CAPZB*	A	0.16	-0.11 (0.01)	3.60E-21	(Porcu et al.2013)	rs10799824	--	A	0.24	-0.03 (0.05)	0.58
rs334699	1	*NFIA*	A	0.05	-0.14 (0.02)	5.40E-12	(Porcu et al.2013)	rs334713	1.00	A	0.17	-0.17 (0.05)	1.50E-03
rs13015993	2	*IGFBP5*	A	0.74	0.08 (0.01)	3.24E-15	(Porcu et al.2013)	rs13020935	1.00	G	0.48	-0.15 (0.04)	1.82E-04
rs10028213	4	*NR3C2*	C	0.82	0.08 (0.01)	2.88E-10	(Rawal et al.2012)	rs10519980	1.00	T	0.33	-0.07 (0.04)	0.11
rs10032216	4	*NR3C2*	T	0.78	0.09 (0.01)	9.28E-16	(Porcu et al.2013)	rs17025017	1.00	A	0.42	-0.07 (0.04)	0.08
rs2046045	5	*PDE8B*	T	0.62	-0.12 (0.01)	2.79E-27	(Rawal et al.2012;Eriksson et al.2012;Medici et al.2011)	rs2046045	--	A	0.28	-0.09 (0.04)	0.03
rs6885099	5	*PDE8B*	A	0.59	-0.14 (0.01)	1.95E-56	(Porcu et al.2013)	rs2046045	1.00	A	0.28	-0.09 (0.04)	0.03
rs4704397	5	*PDE8B*	A	0.41*	0.21	1.64E-10	(Taylor et al.2011)	rs2046045	0.94	A	0.28	-0.09 (0.04)	0.03
rs753760	6	*PDE10A*	C	0.69	0.10 (0.01)	1.21E-24	(Porcu et al.2013)	rs2983514	0.93	G	0.38	-0.01 (0.04)	0.73
rs9472138	6	*VEGFA*	T	0.29	-0.08 (0.01)	6.72E-16	(Porcu et al.2013)	rs9472138	--	T	0.19	-0.10 (0.05)	0.05
rs11755845	6	*VEGFA*	T	0.27	-0.07 (0.01)	1.68E-10	(Porcu et al.2013)	rs11755845	--	T	0.14	-0.13 (0.05)	0.01
rs9497965	6	*SASH1*	T	0.42	0.05 (0.01)	2.25E-08	(Porcu et al.2013)	rs9377117	0.54	G	0.18	0.01 (0.06)	0.85
rs7825175	8	*NRG1*	A	0.21	-0.07 (0.01)	2.94E-09	(Porcu et al.2013)	rs7825175	--	A	0.13	-0.10 (0.06)	0.12
rs657152	9	*ABO*	A	0.34	0.06 (0.01)	4.11E-10	(Porcu et al.2013)	rs657152	--	T	0.43	0.09 (0.04)	0.03
rs1571583	9	*GLIS3*	A	0.25	0.06 (0.01)	2.55E-08	(Porcu et al.2013)	rs1571583	--	T	0.22	0.01 (0.05)	0.79
rs17723470	11	*PRDM11*	T	0.28	-0.07 (0.01)	8.83E-11	(Porcu et al.2013)	rs17723470	--	T	0.11	-0.10 (0.06)	0.11
rs1537424	14	*MBIP*	T	0.61	-0.05 (0.01)	1.17E-08	(Porcu et al.2013)	rs1537424	--	A	0.34	0.04 (0.04)	0.35
rs11624776	14	*ITPK1*	A	0.66	-0.06 (0.01)	1.79E-09	(Porcu et al.2013)	rs11624776	--	C	0.11	0.04 (0.07)	0.57
rs10519227	15	*FGF7*	A	0.25	-0.07 (0.01)	1.02E-11	(Porcu et al.2013)	rs7168316	1.00	T	0.12	-0.03 (0.06)	0.62
rs17776563	15	*MIR1179*	A	0.32	-0.06 (0.01)	2.89E-10	(Porcu et al.2013)	rs13329353	0.96	C	0.45	-0.07 (0.04)	0.09
rs3813582	16	LOC440389/*MAF*	T	0.67	0.08 (0.01)	8.45E-18	(Rawal et al.2012;Porcu et al.2013)	rs17767383	1.00	A	0.25	-0.06 (0.05)	0.18
rs9915657	17	*SOX9*	T	0.54	-0.06 (0.01)	7.53E-13	(Porcu et al.2013)	rs9915657	--	T	0.49	-0.06 (0.04)	0.17
rs4804416	19	*INSR*	T	0.57	-0.06 (0.01)	3.16E-10	(Porcu et al.2013)	rs4804416	--	G	0.26	0.01 (0.05)	0.81

SNP rs number, chromosomal location, nearest gene/gene region, coded allele (CA), coded allele frequency (CAF), and association summary statistics (betas, standard errors, and p-values) are given for each previously reported association with TSH levels in European Americans. CAF highlighted with (*) represents the average CAF in the Taylor et al. study. For SNPs not directly genotyped in this study, the proxy in highest linkage disequilibrium in 1000 Genomes CEU samples was identified. Results of adjusted (age, sex, body mass index, and principal component 1) tests of association are given for each previously reported SNP or its proxy in this African American dataset (n=351).

Appendix N. Comparison of thyroid-related trait associations to eMERGE European Americans.

Locus			Prior Association					Current Study					
SNP	Chr	Gene	CA	CAF	OR	P-value	Ref.	SNP/Best Proxy SNP	r^2	CA	CAF	β(SE)	P-value
Hypothyroidism													
rs6679677	1	*PTPN22*	A	0.09	1.36	2.80E-13	(Eriksson et al.2012)	rs2476601	0.78	A	0.10	0.004 (0.02)	0.81
rs2476601	1	*PTPN22*	A	0.09	1.36	3.9E-13	(Eriksson et al.2012)	rs2476601	--	A	0.10	0.004 (0.02)	0.81
rs4915076	1	*VAV3*	C	0.08	1.30	7.50E-10	(Eriksson et al.2012)	rs4915076	--	C	0.09	0.03 (0.02)	0.06
rs2517532	6	HLA	A	0.40	0.86	1.30E-08	(Eriksson et al.2012)	rs2517532	--	T	0.43	-0.01 (0.01)	0.27
rs1064191	6	HCG22/C6orf15	T	0.46	0.87	2.2E-08	(Eriksson et al.2012)	rs1064191	--	A	0.48	-0.002 (0.01)	0.83
rs925487	9	*FOXE1*/C9orf156	C	0.37	0.86	4.1E-08	(Eriksson et al.2012)	rs925487	--	G	0.37	-0.05 (0.01)	6.24E-06
rs907580	9	*FOXE1*/C9orf156	T	0.26	0.84	1.2E-08	(Eriksson et al.2012)	rs907580	--	A	0.27	-0.05 (0.01)	8.2E-06
rs925489	9	*KRT18P13,FOXE1*	C	0.33	0.78	2.40E-19	(Eriksson et al.2012)	rs925489	--	C	0.34	-0.05 (0.01)	1.79E-06
rs1877432	9	*KRT18P13,FOXE1*	A	0.40	1.16	4.40E-09	(Eriksson et al.2012)	rs1877432	--	A	0.40	0.02 (0.01)	0.03
rs7024345	9	*KRT18P13,FOXE1*	A	0.26	0.84	1E-08	(Eriksson et al.2012)	rs7024345	--	A	0.27	-0.05 (0.01)	1.65E-05
rs7848973	9	*KRT18P13,FOXE1*	A	0.40	0.84	7.10E-11	(Eriksson et al.2012)	rs7848973	--	A	0.40	-0.04 (0.01)	2.11E-05
rs11065987	12	LOC100101246 BRAP	G	0.45	1.18	1.70E-10	(Eriksson et al.2012)	rs11065987	--	G	0.44	0.006 (0.01)	0.53
rs17696736	12	*NAA25*	G	0.46	1.18	2.80E-10	(Eriksson et al.2012)	rs17696736	--	G	0.45	0.003 (0.01)	0.75
rs11066320	12	*PTPN11*	A	0.45	1.17	3.50E-09	(Eriksson et al.2012)	rs11066320	--	A	0.45	0.002 (0.01)	0.82
rs3184504	12	*SH2B3*	C	0.50	0.84	2.60E-12	(Eriksson et al.2012)	rs3184504	--	T	0.50	0.007 (0.01)	0.50
rs11066188	12	C12orf51	A	0.44	1.18	4.1E-10	(Eriksson et al.2012)	rs11066188	--	A	0.43	0.007 (0.01)	0.46
rs653178	12	*ATXN2*	T	0.50	0.84	5.0E-12	(Eriksson et al.2012)	rs653178	--	G	0.50	0.006 (0.01)	0.51
Grave's Disease/Autoimmune Thyroid Disease													
rs3761959	1	*FCRL3*	A	0.40	1.23	1.50E-13	(Chu et al.2011)	rs3761959	--	A	0.44	-0.006 (0.01)	0.53
rs1024161	2	*CTLA4*	T	0.69	1.3	2.34E-17	(Chu et al.2011)	rs1024161	--	T	0.41	0.01 (0.01)	0.15
rs6832151	4	*RHOH,CHRNA9*	G	0.35	1.24	1.08E-13	(Chu et al.2011)	rs6832151	--	G	0.28	0.01 (0.01)	0.24
rs9355610	6	*RNASET2*	G	0.47	1.19	6.85E-10	(Chu et al.2011)	rs9355610	--	A	0.33	-0.008 (0.01)	0.46
rs4947296	6	*MUC21*,C6orf15	C	0.14	1.77	3.51E-51	(Chu et al.2011)	NA	NA	NA	NA	NA	NA
rs2281388	6	*HLA-DPB1*	T	0.32	1.64	1.5E-65	(Chu et al.2011)	NA	NA	NA	NA	NA	NA
rs6457617	6	*HLA-DR-DQ*	T	0.45	1.4	7.38E-33	(Chu et al.2011)	rs6457617	--	T	0.50	0.01 (0.01)	0.14
rs6903608	6	*HLA-DR-DQ*	C	0.38	1.34	5.12E-24	(Chu et al.2011)	rs6903608	--	C	0.31	-0.02 (0.01)	0.13
rs965513	9	*FOXE1*	A	0.34	1.75	1.70E-27	(Eriksson et al.2012)	rs965513	--	A	0.34	-0.05 (0.01)	1.09E-06
rs12101261	14	*TSHR*	T	0.64	1.35	6.64E-24	(Chu et al.2011)	rs12101261	--	T	0.36	0.002 (0.01)	0.87
Thyroid Cancer													
rs966423	2	*DIRC3*	C		1.34	1.30E-09	(Gudmundsson et al.2012)	rs966423	--	C	0.42	-0.02 (0.01)	0.13
rs2439302	8	*NRG1*	G		1.36	2.00E-09	(Gudmundsson et al.2012)	rs7005606	1.00	G	0.46	-0.02 (0.01)	0.02
rs944289	14	*NKX2-1/TTF1*	T	0.57	1.37	2.0E-09	(Gudmundsson et al.2012;Gudmundsson et al.2009)	rs944289	--	C	0.43	0.02 (0.01)	0.01
rs116909374	14	*MBIP*	T		2.09	4.60E-11	(Gudmundsson et al.2012)	rs2553571	0.13	T	0.22	-0.005 (0.01)	0.68

SNP rs number, chromosomal location, nearest gene/gene region, coded allele (CA), coded allele frequency (CAF), and association summary statistics (odds ratio (OR) and p-values) are given for each previously reported association with thyroid-related traits in European Americans. For SNPs not directly genotyped in this study, the proxy in highest linkage disequilibrium in 1000 Genomes CEU samples was identified. Results of adjusted (age, sex, body mass index, and principal component 1) tests of association are given for each SNP in this European American dataset.

Appendix O. Comparison of thyroid-related trait associations to eMERGE African Americans.

	Locus		Prior Association					Current Study					
SNP	Chr	Gene	CA	CAF	OR	P-value	Ref.	SNP/Best Proxy SNP	r^2	CA	CAF	B (SE)	P-value
Hypothyroidism													
rs6679677	1	*PHTF1, RSBN1*	A	0.09	1.36	2.80E-13	(Eriksson et al.2012)	rs1217413	0.60	G	0.05	0.08 (0.10)	0.43
rs2476601	1	*PTPN22*	A	0.09	1.36	3.9E-13	(Eriksson et al.2012)	rs1217413	0.56	G	0.05	0.08 (0.10)	0.43
rs4915076	1	*VAV3*	C	0.08	1.3	8.00E-10	(Eriksson et al.2012)	rs4915076	--	C	0.05	-0.03 (0.08)	0.72
rs2517532	6	*LOC729792 HCG22*	A	0.40	0.86	1.3E-08	(Eriksson et al.2012)	rs2517532	--	T	0.34	-0.02 (0.04)	0.59
rs1064191	6	*HCG22/C6orf15*	T	0.46	0.87	2.2E-08	(Eriksson et al.2012)	rs1064191	--	A	0.46	-0.04 (0.04)	0.36
rs925487	9	*FOXE1/C9orf156*	C	0.37	0.86	4.1E-08	(Eriksson et al.2012)	rs925487	--	G	0.25	-0.04 (0.05)	0.41
rs907580	9	*FOXE1/C9orf156*	T	0.26	0.84	1.2E-08	(Eriksson et al.2012)	rs907580	--	A	0.07	0.04 (0.08)	0.64
rs925489	9	*KRT18P13,FOXE1*	C	0.33	0.78	2.40E-19	(Eriksson et al.2012)	rs925489	--	C	0.21	-0.03 (0.05)	0.53
rs1877432	9	*KRT18P13,FOXE1*	A	0.40	1.16	4.40E-09	(Eriksson et al.2012)	rs1877432	--	A	0.34	0.11 (0.04)	9.73E-03
rs7024345	9	*KRT18P13,FOXE1*	A	0.26	0.84	1E-08	(Eriksson et al.2012)	rs7024345	--	A	0.07	0.03 (0.08)	0.69
rs7848973	9	*KRT18P13,FOXE1*	A	0.40	0.84	7.10E-11	(Eriksson et al.2012)	rs7848973	--	A	0.23	-0.09 (0.04)	0.06
rs11065987	12	*LOC100101246 BRAP*	G	0.45	1.18	1.70E-10	(Eriksson et al.2012)	rs11065987	--	G	0.08	0.12 (0.08)	0.13
rs17696736	12	*NAA25*	G	0.46	1.18	2.80E-10	(Eriksson et al.2012)	rs17696736	--	G	0.09	0.11 (0.08)	0.16
rs11066320	12	*PTPN11*	A	0.45	1.17	3.50E-09	(Eriksson et al.2012)	rs11066320	--	A	0.08	0.08 (0.08)	0.31
rs3184504	12	*SH2B3*	C	0.50	0.84	2.60E-12	(Eriksson et al.2012)	rs3184504	--	T	0.10	0.09 (0.07)	0.20
rs11066188	12	*C12orf51*	A	0.44	1.18	4.1E-10	(Eriksson et al.2012)	rs11066188	--	A	0.08	0.12 (0.08)	0.12
rs653178	12	*ATXN2*	T	0.50	0.84	5.0E-12	(Eriksson et al.2012)	rs653178	--	G	0.10	0.09 (0.07)	0.20
Grave's Disease/Autoimmune Thyroid Disease													
rs3761959	1	*FCRL3*	A	0.40	1.23	1.50E-13	(Chu et al.2011)	rs3761959	--	G	0.39	0.03 (0.04)	0.45
rs1024161	2	*CTLA4*	T	0.69	1.3	2.34E-17	(Chu et al.2011)	rs1024161	--	C	0.48	-0.06 (0.09)	0.13
rs6832151	4	*RHOH,CHRNA9*	G	0.35	1.24	1.08E-13	(Chu et al.2011)	rs6832151	--	G	0.31	-0.10 (0.04)	0.01
rs9355610	6	*RNASET2*	G	0.47	1.19	6.85E-10	(Chu et al.2011)	rs9355610	--	A	0.39	-0.03 (0.04)	0.47
rs4947296	6	*MUC21,C6orf15*	C	0.14	1.77	3.51E-51	(Chu et al.2011)	NA	NA	NA	NA	NA	NA
rs2281388	6	*HLA-DPB1*	T	0.32	1.64	1.5E-65	(Chu et al.2011)	NA	NA	NA	NA	NA	NA
rs6457617	6	*HLA-DR-DQ*	T	0.45	1.4	7.38E-33	(Chu et al.2011)	rs6457617	--	T	0.48	0.02 (0.04)	0.63
rs6903608	6	*HLA-DR-DQ*	C	0.38	1.34	5.12E-24	(Chu et al.2011)	rs6903608	--	C	0.38	0.02 (0.04)	0.61
rs965513	9	*FOXE1*	A	0.34	1.75	1.70E-27	(Eriksson et al.2012)	rs965513	--	A	0.17	-0.03 (0.05)	0.50
rs12101261	14	*TSHR*	T	0.64	1.35	6.64E-24	(Chu et al.2011)	rs12101261	--	T	0.39	-0.01 (0.04)	0.82
Thyroid Cancer													
rs966423	2	*DIRC3*	C		1.34	1.30E-09	(Gudmundsson et al.2012)	rs966423	--	T	0.23	0.03 (0.05)	0.47
rs2439302	8	*NRG1*	G		1.36	2.00E-09	(Gudmundsson et al.2012)	rs4733130	1.00	C	0.23	-0.06 (0.05)	0.25
rs944289	14	*NKX2-1/TTF1*	T	0.57	1.37	2E-09	(Gudmundsson et al.2012; Gudmundsson et al.2009)	rs1169151	0.93	A	0.23	0.03 (0.05)	0.57
rs116909374	14	*MBIP*	T		2.09	4.60E-11	(Gudmundsson et al.2012)	NA	NA	NA	NA	NA	NA

SNP rs number, chromosomal location, nearest gene/gene region, coded allele (CA), coded allele frequency (CAF), and association summary statistics (odds ratio (OR) and p-values) are given for each previously reported association with thyroid-related traits in European Americans. For SNPs not directly genotyped in this study, the proxy in highest linkage disequilibrium in 1000 Genomes CEU samples was identified. Results of adjusted (age, sex, body mass index, and principal component 1) tests of association are given for each SNP in this African American dataset.

Appendix P. Comparison of SNP associations in regression models with and without BMI covariates for serum TSH levels in eMERGE study European Americans.

SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs1382879	7.16E-18	0.09	2.16E-15	0.08
rs2046045	1.85E-17	0.09	2.07E-15	0.08
rs989758	1.33E-14	0.08	1.10E-12	0.07
rs9687206	5.52E-14	0.08	4.98E-11	0.06
rs12515498	3.27E-10	0.07	3.18E-09	0.06
rs6885813	4.05E-08	0.06	1.41E-07	0.06
rs1096752	6.30E-07	-0.05	7.48E-06	-0.04
rs13361710	6.60E-07	0.06	2.20E-06	0.05
rs10759944	1.08E-06	-0.05	1.33E-06	-0.05
rs965513	1.09E-06	-0.05	1.40E-06	-0.05
rs925489	1.79E-06	-0.05	2.52E-06	-0.05
rs7850258	1.85E-06	-0.05	2.87E-06	-0.05
rs10496992	2.22E-06	0.05	1.61E-05	0.04
rs1861628	3.68E-06	-0.05	2.15E-05	-0.04
rs4348174	3.97E-06	0.05	1.13E-05	0.04
rs657152	4.18E-06	0.05	1.06E-07	0.05
rs740083	4.56E-06	-0.05	1.37E-05	-0.05
rs813379	4.57E-06	-0.10	2.66E-06	-0.10
rs2712168	4.98E-06	0.07	2.05E-05	0.06
rs256438	5.53E-06	0.05	1.81E-05	0.04
rs4570936	5.73E-06	-0.05	2.32E-07	-0.06
rs6546537	5.92E-06	-0.05	2.92E-05	-0.04
rs7855088	6.23E-06	-0.05	2.37E-05	-0.04
rs925487	6.24E-06	-0.05	6.99E-06	-0.04
rs803174	6.74E-06	-0.10	3.38E-06	-0.09
rs2438632	6.88E-06	0.05	1.01E-05	0.04
rs13020935	7.02E-06	-0.05	3.45E-05	-0.04
rs12520862	7.48E-06	0.06	9.36E-05	0.05
rs10984103	7.81E-06	-0.05	7.69E-06	-0.04
rs907580	8.20E-06	-0.05	6.14E-06	-0.05
rs2466067	8.42E-06	-0.05	5.86E-07	-0.05
rs7870926	8.67E-06	-0.04	1.47E-05	-0.04
rs7341064	1.03E-05	0.04	2.18E-05	0.04
rs4298457	1.07E-05	-0.05	1.05E-06	-0.05
rs598599	1.09E-05	0.05	8.83E-05	0.04
rs4693596	1.10E-05	-0.04	1.05E-04	-0.04
rs10954859	1.12E-05	-0.05	1.21E-06	-0.05
rs404375	1.26E-05	-0.04	2.38E-05	-0.04
rs2983525	1.34E-05	-0.05	1.04E-06	-0.05

SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs2983514	1.36E-05	-0.05	6.45E-07	-0.05
rs3766122	1.42E-05	-0.10	7.20E-04	-0.07
rs7866436	1.52E-05	-0.04	1.55E-05	-0.04
rs7024345	1.65E-05	-0.05	1.41E-05	-0.05
rs26367	1.65E-05	-0.07	2.05E-05	-0.07
rs10073636	1.73E-05	0.04	1.35E-04	0.04
rs13285674	1.90E-05	0.05	1.86E-05	0.05
rs505922	1.94E-05	0.04	3.55E-07	0.05
rs7445986	1.95E-05	0.04	5.84E-05	0.04
rs10204522	1.95E-05	0.07	3.89E-05	0.07
rs4054489	1.95E-05	-0.05	7.09E-05	-0.05
rs1055075	1.96E-05	-0.04	1.44E-05	-0.04
rs4861534	2.09E-05	0.07	2.15E-05	0.06
rs7168316	2.10E-05	-0.05	2.55E-05	-0.05
rs7848973	2.11E-05	-0.04	2.22E-05	-0.04
rs3136559	2.13E-05	0.05	8.51E-04	0.03
rs6727435	2.15E-05	-0.05	9.24E-05	-0.04
rs33613	2.35E-05	-0.07	2.68E-05	-0.07
rs12592277	2.35E-05	-0.05	2.39E-05	-0.05
rs2466062	2.36E-05	-0.05	1.50E-06	-0.05
rs3898456	2.56E-05	0.04	9.80E-05	0.04
rs4402960	2.63E-05	-0.05	1.03E-05	-0.04
rs1470579	2.67E-05	-0.05	8.35E-06	-0.04
rs13354798	2.75E-05	0.04	1.91E-04	0.03
rs9686502	2.85E-05	0.04	3.83E-04	0.03
rs9606756	2.86E-05	0.07	1.27E-04	0.06
rs494442	3.03E-05	-0.04	1.76E-04	-0.04
rs2695148	3.11E-05	-0.07	9.14E-05	-0.06
rs17265852	3.16E-05	-0.07	3.34E-05	-0.07
rs6414906	3.51E-05	0.04	1.81E-04	0.04
rs3813583	4.06E-05	0.04	5.94E-06	0.04
rs749378	4.09E-05	-0.05	1.03E-04	-0.04
rs6451801	4.10E-05	0.04	2.27E-04	0.03
rs13162651	4.11E-05	0.04	2.14E-04	0.03
rs12201217	4.30E-05	-0.04	1.86E-05	-0.04
rs370234	4.32E-05	-0.04	6.80E-05	-0.04
rs1647253	4.57E-05	-0.07	1.23E-04	-0.06
rs6989877	4.59E-05	0.06	3.06E-06	0.06
rs11963665	4.63E-05	-0.05	2.73E-04	-0.04
rs6892290	4.76E-05	0.04	2.41E-04	0.03
rs6668505	4.88E-05	-0.08	1.85E-04	-0.07
rs3745746	4.93E-05	-0.04	8.12E-05	-0.04

SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs12521494	5.00E-05	0.05	1.05E-03	0.04
rs10064949	5.07E-05	0.04	7.77E-05	0.04
rs1515259	5.44E-05	0.04	3.21E-04	0.03
rs1012319	5.47E-05	-0.05	1.04E-04	-0.05
rs2983500	5.54E-05	-0.06	4.80E-06	-0.07
rs8096947	5.56E-05	-0.05	9.39E-06	-0.05
rs888186	5.63E-05	-0.07	2.03E-04	-0.06
rs4703797	5.65E-05	0.04	7.77E-05	0.04
rs529126	5.65E-05	0.04	1.85E-04	0.04
rs11805172	5.69E-05	-0.08	2.09E-04	-0.07
rs8009673	5.83E-05	0.06	3.14E-05	0.05
rs1939422	5.94E-05	-0.04	2.35E-04	-0.04
rs1986415	6.05E-05	0.06	1.56E-04	0.05
rs2439300	6.19E-05	-0.04	1.91E-05	-0.04
rs2943179	6.31E-05	0.05	4.18E-05	0.05
rs8035662	6.37E-05	-0.04	2.98E-04	-0.03
rs877138	6.37E-05	-0.04	4.11E-05	-0.04
rs1443434	6.53E-05	-0.04	5.97E-05	-0.04
rs2381866	6.68E-05	0.04	9.97E-05	0.04
rs888182	6.75E-05	0.05	3.55E-05	0.05
rs7184757	7.22E-05	-0.07	6.70E-05	-0.07
rs11172482	7.29E-05	-0.04	1.11E-04	-0.04
rs39334	7.47E-05	0.04	2.38E-04	0.04
rs12278001	7.53E-05	-0.08	8.24E-05	-0.08
rs12654213	7.69E-05	0.04	3.64E-04	0.03
rs10489909	7.81E-05	-0.09	8.43E-04	-0.07
rs13231383	8.19E-05	0.04	1.71E-04	0.04
rs2306344	8.23E-05	-0.04	6.54E-05	-0.04
rs11666426	8.30E-05	0.04	8.65E-05	0.04
rs12138950	8.97E-05	-0.05	1.29E-04	-0.05
rs424829	9.02E-05	0.04	3.39E-04	0.04
rs11071858	9.33E-05	-0.04	4.40E-04	-0.03
rs12282135	9.47E-05	-0.05	2.44E-04	-0.05
rs11118832	9.52E-05	-0.07	1.71E-04	-0.07
rs630505	9.58E-05	-0.04	1.38E-04	-0.04
rs16856529	9.80E-05	0.05	5.90E-05	0.05
rs1502816	9.95E-05	-0.04	1.32E-04	-0.04

For each SNP, p-values and betas are given for models that include or exclude body mass index (BMI) as a covariate. All models are linear regressions assuming an additive genetic model adjusted for age, sex, and principal component 1 in this European American dataset (n=4,501).

Appendix Q. Comparison of SNP associations in regression models with and without BMI covariates for serum TSH levels in eMERGE African Americans.

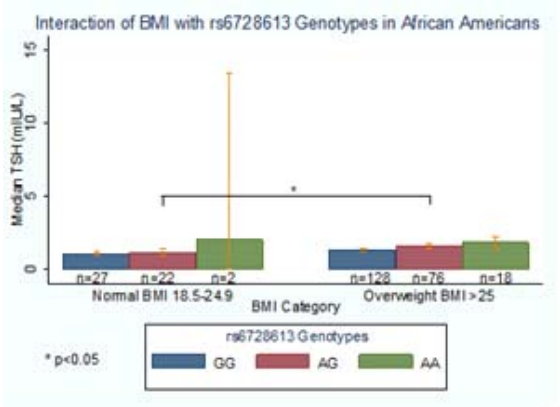
SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs1409005	5.02E-07	0.25	2.60E-05	0.17
rs2378497	3.53E-06	0.33	4.76E-03	0.16
rs6062344	4.06E-06	0.18	3.64E-03	0.09
rs270421	7.75E-06	0.19	3.08E-03	0.10
rs2299116	8.16E-06	0.25	6.13E-05	0.18
rs6728613	1.14E-05	0.20	7.95E-03	0.10
rs6585018	1.17E-05	-0.22	2.58E-04	-0.14
rs1013757	1.33E-05	-0.19	7.32E-04	-0.12
rs4073401	1.33E-05	0.19	8.28E-03	0.09
rs12883861	1.63E-05	0.21	2.22E-04	0.14
rs9784959	1.82E-05	-0.18	1.63E-05	-0.15
rs270422	2.17E-05	0.18	0.01	0.09
rs261875	2.24E-05	0.18	3.56E-05	0.14
rs274614	2.36E-05	-0.18	0.03	-0.08
rs11711934	2.45E-05	-0.17	1.41E-03	-0.11
rs12621889	2.68E-05	0.36	0.01	0.16
rs12464144	2.68E-05	0.36	0.01	0.16
rs10163845	2.74E-05	-0.18	0.04	-0.07
rs12610504	3.07E-05	0.20	2.44E-03	0.12
rs1274744	3.21E-05	-0.17	1.98E-03	-0.10
rs10060607	3.28E-05	0.18	8.25E-05	0.13
rs1824304	3.32E-05	0.17	9.22E-04	0.11
rs841452	3.52E-05	0.17	1.96E-03	0.10
rs11977108	3.70E-05	-0.21	6.20E-04	-0.14
rs4678798	3.71E-05	0.24	8E-05	0.18
rs6851816	3.83E-05	0.16	6.50E-04	0.10
rs133201	4.04E-05	0.27	0.05	0.11
rs2593996	4.09E-05	-0.16	5.52E-05	-0.13
rs1054713	4.16E-05	0.19	9.71E-03	0.10
rs12609319	4.23E-05	0.20	2.76E-03	0.12
rs1918092	4.90E-05	0.30	4.47E-04	0.20
rs2303478	5.11E-05	0.18	0.02	0.08
rs3738605	5.12E-05	0.24	6.33E-03	0.13
rs2659099	5.13E-05	0.18	6.38E-03	0.10
rs4955261	5.19E-05	0.16	3.77E-04	0.11
rs4772145	5.23E-05	0.15	0.03	0.07
rs13403407	5.31E-05	-0.16	4.40E-03	-0.09
rs1513476	5.46E-05	0.22	1.17E-04	0.17
rs17032566	5.52E-05	-0.30	1.45E-03	-0.19

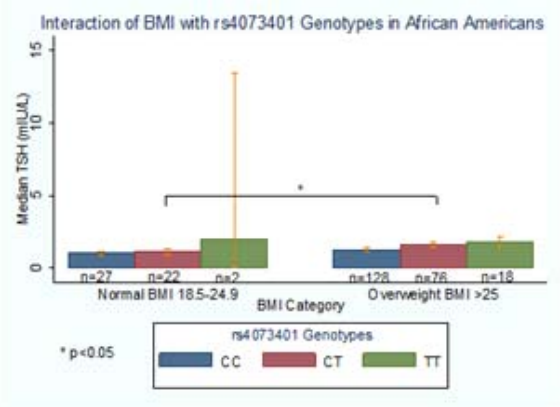
SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs1105813	5.61E-05	0.16	1.07E-03	0.10
rs1563333	5.64E-05	-0.19	2.06E-03	-0.11
rs1907356	5.71E-05	-0.21	2.26E-04	-0.15
rs11001788	5.71E-05	-0.21	2.26E-04	-0.15
rs12470895	5.86E-05	0.21	1.46E-03	0.13
rs646929	5.96E-05	0.30	7.19E-04	0.19
rs2335640	5.98E-05	-0.17	3.31E-03	-0.10
rs3742049	6.08E-05	0.18	3.31E-03	0.11
rs17052068	6.46E-05	-0.16	1.96E-03	-0.10
rs2819757	6.49E-05	0.22	3.77E-03	0.13
rs10804139	6.54E-05	-0.16	8.86E-04	-0.11
rs736218	6.66E-05	0.16	2.02E-03	0.10
rs10989120	7.05E-05	-0.19	0.11	-0.06
rs10744020	7.05E-05	0.16	1.97E-03	0.10
rs2659103	7.29E-05	0.19	0.01	0.10
rs10918914	7.39E-05	0.22	1.29E-04	0.16
rs261878	7.42E-05	-0.16	7.12E-05	-0.13
rs12914266	7.58E-05	0.17	1.16E-03	0.11
rs6965055	7.65E-05	-0.16	1.34E-03	-0.10
rs7808606	7.66E-05	-0.15	9.37E-04	-0.10
rs17322359	7.74E-05	0.25	4.05E-03	0.15
rs11949641	7.89E-05	0.18	2.12E-03	0.11
rs12120382	7.96E-05	0.29	8.34E-04	0.20
rs6731363	7.99E-05	0.20	1.77E-03	0.13
rs13144021	8.00E-05	0.23	1.06E-03	0.15
rs877128	8.10E-05	0.18	2.34E-03	0.11
rs7923004	8.19E-05	-0.20	1.03E-03	-0.13
rs6999969	8.33E-05	-0.16	7.19E-03	-0.09
rs1027388	8.36E-05	-0.17	2E-03	-0.11
rs17011253	8.38E-05	0.27	0.03	0.12
rs942077	8.47E-05	-0.15	1.58E-04	-0.12
rs4370216	8.55E-05	-0.15	1.8E-03	-0.10
rs2333727	8.55E-05	-0.15	1.8E-03	-0.10
rs1029357	8.57E-05	0.15	1.05E-03	0.10
rs1332598	8.69E-05	-0.19	0.11	-0.06
rs6864667	8.98E-05	0.15	2.21E-03	0.09
rs4411338	9.05E-05	0.16	0.02	0.08
rs171953	9.07E-05	-0.15	0.03	-0.07
GA035020	9.11E-05	0.19	0.01	0.10
rs2040578	9.13E-05	0.17	2.72E-04	0.13
rs1106826	9.23E-05	0.17	0.01	0.08
rs6090040	9.35E-05	0.15	0.02	0.08

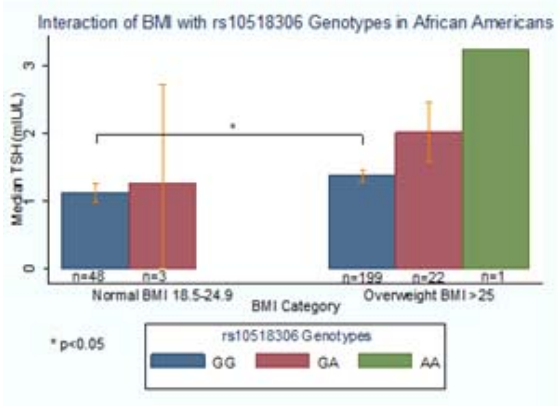
SNP	P BMI	BETA BMI	P NO BMI	BETA NO BMI
rs4408777	9.39E-05	0.16	0.07	0.06
rs2521676	9.73E-05	0.16	8.4E-04	0.11
rs16845412	9.76E-05	0.27	6.14E-03	0.15
rs10518306	9.78E-05	0.35	4.33E-03	0.19
rs10098991	9.86E-05	0.16	0.08	0.06
rs8059691	9.90E-05	0.23	0.01	0.11

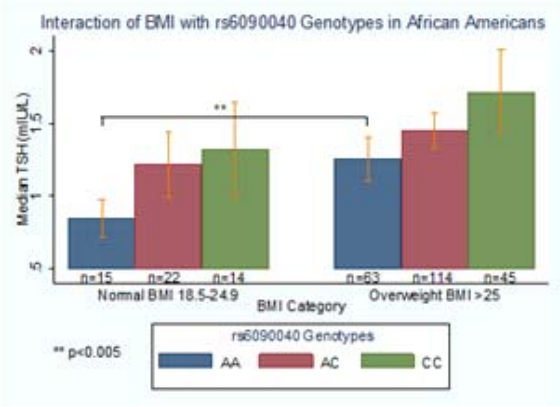
For each SNP, p-values and betas are given for models that include or exclude body mass index (BMI) as a covariate. All models are linear regressions assuming an additive genetic model adjusted for age, sex, and principal component 1 in this African American dataset (n=351).

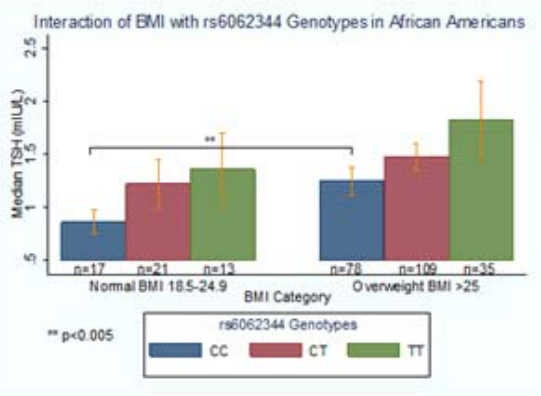
Appendix R. Body mass index as a modifier of serum TSH levels genetic associations in eMERGE African Americans.











Interaction analyses were performed using the SNPs with p<1x10-4 significance levels in the model adjusted for age, sex, principal component 1 (PC1), and body mass index (BMI) in a model stratified for race/ethnicity and by normal/overweight BMI (normal: BMI 18-24.9; overweight: BMI 25+). We considered a SNPxBMI interaction significant at a threshold of p<0.05. Shown are p-values from Wilcoxon rank-sum tests comparing median TSH values between BMI categories at each genotype.

Appendix S. Power calculations for replication/generalization in eMERGE TSH levels study.

SNP	Locus Chr	Gene	CA	CAF	Prior Association β	Prior Association P-value	Ref.	CAF EA	Power EA n=4501	CAF AA	Power AA n=351
rs10917469	1	*CAPZB*	G	0.16	-0.16	3.2E-08	(Panicker et al.2010)	0.15	1.00	0.24	0.53
rs10917477	1	*CAPZB*	A	0.51	-0.06	1.54E-08	(Rawal et al.2012)	0.48	0.74*	0.49	0.14*
rs10799824	1	*CAPZB*	A	0.16	-0.11	3.60E-21	(Porcu et al.2013)	0.15	0.97*	0.24	0.41
rs334699	1	*NFIA*	A	0.05	-0.14	5.40E-12	(Porcu et al.2013)	0.08	0.93*	0.17	0.49
rs13015993	2	*IGFBP5*	A	0.74	0.08	3.24E-15	(Porcu et al.2013)	0.27	0.95	0.48	0.31
rs10028213	4	*NR3C2*	C	0.82	0.08	2.88E-10	(Rawal et al.2012)	0.18	0.88	0.33	0.28
rs10032216	4	*NR3C2*	T	0.78	0.09	9.28E-16	(Porcu et al.2013)	0.19	0.95	0.42	0.37
rs2046045	5	*PDE8B*	T	0.62	-0.12	2.79E-27	(Rawal et al.2012;Eriksson et al.2012;Medici et al.2011)	0.40	1.00	0.28	0.51
rs6885099	5	*PDE8B*	A	0.59	-0.14	1.95E-56	(Porcu et al.2013)	0.40	1.00	0.28	0.64
rs4704397	5	*PDE8B*	A	0.40*	0.21	1.64E-10	(Taylor et al.2011)	0.39	1.00*	0.28	0.92*
rs753760	6	*PDE10A*	C	0.69	0.10	1.21E-24	(Porcu et al.2013)	0.33	1.00*	0.38	0.41*
rs9472138	6	*VEGFA*	T	0.29	-0.08	6.72E-16	(Porcu et al.2013)	0.28	0.96	0.19	0.21
rs11755845	6	*VEGFA*	T	0.27	-0.07	1.68E-10	(Porcu et al.2013)	0.24	0.86	0.14	0.15
rs9497965	6	*SASH1*	T	0.42	0.05	2.25E-08	(Porcu et al.2013)	0.30	0.41*	0.18	0.08*
rs7825175	8	*NRG1*	A	0.21	-0.07	2.94E-09	(Porcu et al.2013)	0.31	0.33*	0.13	0.14
rs657152	9	*ABO*	A	0.34	0.06	4.11E-10	(Porcu et al.2013)	0.38	0.84	0.43	0.19
rs1571583	9	*GLIS3*	A	0.25	0.06	2.55E-08	(Porcu et al.2013)	0.25	0.76	0.22	0.15
rs17723470	11	*PRDM11*	T	0.28	-0.07	8.83E-11	(Porcu et al.2013)	0.29	0.87*	0.11	0.13
rs1537424	14	*MBIP*	T	0.61	-0.05	1.17E-08	(Porcu et al.2013)	0.43	0.71	0.34	0.14
rs11624776	14	*ITPK1*	A	0.66	-0.06	1.79E-09	(Porcu et al.2013)	0.22	0.29*	0.11	0.11
rs10519227	15	*FGF7*	A	0.25	-0.07	1.02E-11	(Porcu et al.2013)	0.23	0.85	0.12	0.13
rs17776563	15	*MIR1179*	A	0.32	-0.06	2.89E-10	(Porcu et al.2013)	0.35	0.75*	0.45	0.18*
rs3813582	16	LOC440389/*MAF*	T	0.67	0.08	8.45E-18	(Rawal et al.2012;Porcu et al.2013)	0.31	0.97	0.25	0.25
rs9915657	17	*SOX9*	T	0.54	-0.06	7.53E-13	(Porcu et al.2013)	0.46	0.86	0.49	0.20
rs4804416	19	*INSR*	T	0.57	-0.06	3.16E-10	(Porcu et al.2013)	0.44	0.86	0.26	0.16

Power calculations for replication/generalization of SNPs previously associated with serum TSH levels to eMERGE euthyroid European Americans (EA) and African Americans (AA). SNP rs number, chromosomal location, nearest gene/gene region, coded allele (CA), coded allele frequency (CAF), and association summary statistics (betas and p-values) are given for each previously reported association with serum TSH levels in European Americans. Starred (*) CAF represents mean CAF from Taylor et al. Power was calculated for each race/ethnicity using Quanto assuming the previously reported effect size, an additive genetic model, a liberal significance threshold of $p<0.05$,the eMERGE minor allele frequencies, and the eMERGE sample sizes. Power calculations labeled with an asterisk indicate proxy SNPs listed in Table 20 (European Americans) and Appendix M (African Americans) as described in the Chapter V.

Appendix T. 2x2 table for calculating the PPV of a hypothetical rare disorder.

Variant	Disease (+)	(-)	Total
(+)	99	199	298
(-)	1	19,701	19,702
Total	100	19,900	20,000
	PPV=99/298=33%		

Data shown are for calculating the positive predictive value (PPV) of a hypothetical rare disorder with a population n=20,000, disease prevalence = 0.5%, and tests sensitivity and specificity of 99%.

Appendix U. 2x2 table for calculating the PPV of a hypothetical common disorder.

Variant	Disease (+)	(-)	Total
(+)	822	72	894
(-)	8	9,098	9,106
Total	830	9,170	10,000
	PPV=822/894=92%		

Data shown are for calculating the positive predictive value (PPV) of a hypothetical common disorder with a population n=10,000, disease prevalence = 8.3%, and tests sensitivity and specificity of 99%.

REFERENCES

2-13-1987. "COmbination Oral Contraceptive Use and the Risk of Endometrial Cancer." *JAMA: The Journal of the American Medical Association* 257(6):796-800. doi:doi: 10.1001/jama.1987.03390060086029.

1989. "The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives. The ARIC Investigators." *Am.J.Epidemiol.* 129(4):687-702.

1998. "Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group." *Control Clin.Trials* 19(1):61-109. doi:S0197245697000780 [pii].

SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics .

2009. "Recommendations From the EGAPP Working Group: Genetic Testing Strategies in Newly Diagnosed Individuals With Colorectal Cancer Aimed at Reducing Morbidity and Mortality From Lynch Syndrome in Relatives." *Genet.Med.* 11(1):35-41. doi:10.1097/GIM.0b013e31818fa2ff [doi].

Health Information National Trends Survey 4 (HINTS4) .

23&Me.,7-1-2013a. *23&Me* .

-----.,*Status of our health-related genetic reports* .

Abramson, E. L., V. Patel, A. Edwards, and R. Kaushal. 2014. "Consumer Perspectives on Personal Health Records: a 4-Community Study." *Am.J.Manag.Care* 20(4):287-96. doi:85477 [pii].

Accurso, F. J., S. M. Rowe, J. P. Clancy, M. P. Boyle, J. M. Dunitz, P. R. Durie, S. D. Sagel, D. B. Hornick, M. W. Konstan, S. H. Donaldson, R. B. Moss, J. M. Pilewski, R. C. Rubenstein, A. Z. Uluer, M. L. Aitken, S. D. Freedman, L. M. Rose, N. Mayer-Hamblett, Q. Dong, J. Zha, A. J. Stone, E. R. Olson, C. L. Ordonez, P. W. Campbell, M. A. Ashlock, and B. W. Ramsey. 11-18-2010. "Effect of VX-770 in Persons With Cystic Fibrosis and the G551D-CFTR Mutation." *N.Engl.J.Med.* 363(21):1991-2003. doi:10.1056/NEJMoa0909825 [doi].

Agency for Healthcare Research and Quality.,*Excellence Centers to Eliminate Ethnic/Racial Disparities (EXCEED): Program Brief* Rockville, MD.

Aksglaede, L., K. Sorensen, J. H. Petersen, N. E. Skakkebaek, and A. Juul. 2009. "Recent Decline in Age at Breast Development: the Copenhagen Puberty Study." *Pediatrics* 123(5):e932-e939. doi:123/5/e932 [pii];10.1542/peds.2008-2491 [doi].

Albertsen, H. M., R. Chettier, P. Farrington, and K. Ward. 2013. "Genome-Wide Association Study Link Novel Loci to Endometriosis." *PLoS One* 8(3):e58257. doi:10.1371/journal.pone.0058257 [doi];PONE-D-12-30930 [pii].

Ali, A. T. 2014. "Reproductive Factors and the Risk of Endometrial Cancer." *Int.J.Gynecol.Cancer* 24(3):384-93. doi:10.1097/IGC.0000000000000075 [doi].

Ancker, J. S., L. M. Kern, A. Edwards, S. Nosal, D. M. Stein, D. Hauser, and R. Kaushal. 6-9-2014. "How Is the Electronic Health Record Being Used? Use of EHR Data to Assess Physician-Level Variability in Technology Use." *J.Am.Med.Inform.Assoc.* doi:amiajnl-2013-002627 [pii];10.1136/amiajnl-2013-002627 [doi].

Ancker, J. S., M. P. Singh, R. Thomas, A. Edwards, A. Snyder, A. Kashyap, and R. Kaushal. 2013. "Predictors of Success for Electronic Health Record Implementation in Small Physician Practices." *Appl.Clin.Inform.* 4(1):12-24. doi:10.4338/ACI-2012-09-RA-0033 [doi].

Andermann, A., I. Blancquaert, S. Beauchamp, and V. Dery. 2008. "Revisiting Wilson and Jungner in the Genomic Age: a Review of Screening Criteria Over the Past 40 Years." *Bull.World.Health.Organ.* 86(4):317-19. doi:S0042-96862008000400018 [pii].

Andersen, S., K. M. Pedersen, N. H. Bruun, and P. Laurberg. 2002. "Narrow Individual Variations in Serum T(4) and T(3) in Normal Subjects: a Clue to the Understanding of Subclinical Thyroid Disease." *J.Clin.Endocrinol.Metab.* 87(3):1068-72. doi:10.1210/jcem.87.3.8165 [doi].

Anderson, C. A., G. Zhu, M. Falchi, S. M. van den Berg, S. A. Treloar, T. D. Spector, N. G. Martin, D. I. Boomsma, P. M. Visscher, and G. W. Montgomery. 2008. "A Genome-Wide Linkage Scan for Age at Menarche in Three Populations of European Descent." *J.Clin.Endocrinol.Metab* 93(10):3965-70. doi:jc.2007-2568 [pii];10.1210/jc.2007-2568 [doi].

Armstrong, D. L., R. Zidovetzki, M. E. Alarcon-Riquelme, B. P. Tsao, L. A. Criswell, R. P. Kimberly, J. B. Harley, K. L. Sivils, T. J. Vyse, P. M. Gaffney, C. D. Langefeld, and C. O. Jacob. 5-29-2014. "GWAS Identifies Novel SLE Susceptibility Genes and Explains the Association of the HLA Region." *Genes.Immun.* doi:gene201423 [pii];10.1038/gene.2014.23 [doi].

Armstrong, T. D., L. D. Crum, R. H. Rieger, T. A. Bennett, and L. J. Edwards. 1999. "Attitudes of African Americans Toward Participation in Medical Research." *J.Appl.Soc.Psych.* 29(3):552-74.

Arnaud-Lopez, L., G. Usala, G. Ceresini, B. D. Mitchell, M. G. Pilia, M. G. Piras, N. Sestu, A. Maschio, F. Busonero, G. Albai, M. Dei, S. Lai, A. Mulas, L. Crisponi, T. Tanaka, S. Bandinelli, J. M. Guralnik, A. Loi, L. Balaci, G. Sole, A. Prinzis, S. Mariotti, A. R. Shuldiner, A. Cao, D. Schlessinger, M. Uda, G. R. Abecasis, R. Nagaraja, S. Sanna, and S. Naitza. 2008. "Phosphodiesterase 8B Gene Variants Are Associated With Serum TSH Levels and Thyroid Function." *Am.J.Hum.Genet.* 82(6):1270-1280. doi:S0002-9297(08)00307-8 [pii];10.1016/j.ajhg.2008.04.019 [doi].

Ashton, K. A., A. Proietto, G. Otton, I. Symonds, M. McEvoy, J. Attia, M. Gilbert, U. Hamann, and R. J. Scott. 2009a. "Estrogen Receptor Polymorphisms and the Risk of Endometrial Cancer." *BJOG.* 116(8):1053-61. doi:BJO2185 [pii];10.1111/j.1471-0528.2009.02185.x [doi].

-----. 2009b. "Polymorphisms in TP53 and MDM2 Combined Are Associated With High Grade Endometrial Cancer." *Gynecol.Oncol.* 113(1):109-14. doi:S0090-8258(09)00003-1 [pii];10.1016/j.ygyno.2008.12.036 [doi].

Ashton, K. A., A. Proietto, G. Otton, I. Symonds, M. McEvoy, J. Attia, and R. J. Scott. 2010. "Toll-Like Receptor (TLR) and Nucleosome-Binding Oligomerization Domain (NOD) Gene Polymorphisms and Endometrial Cancer Risk." *BMC.Cancer* 10:382. doi:1471-2407-10-382 [pii];10.1186/1471-2407-10-382 [doi].

Asthma Disparities Working Group.,*President's Task Force on Environmental Health Risks and Safety Risks to Children: Coordinated Federal Action Plan to Reduce Racial and Ethnic Asthma Disparities* .

Audrain, J., A. Gomez-Caminero, A. R. Robertson, R. Boyd, C. T. Orleans, and C. Lerman. 1997. "Gender and Ethnic Differences in Readiness to Change Smoking Behavior." *Womens Health.* 3(2):139-50.

Bagchi, N., T. R. Brown, and R. F. Parish. 1990. "Thyroid Dysfunction in Adults Over Age 55 Years. A Study in an Urban US Community." *Arch.Intern.Med.* 150(4):785-87.

Baggerly, K. A. and K. R. Coombes. 2011. "What Information Should Be Required to Support Clinical "Omics" Publications?" *Clin.Chem.* 57(5):688-90. doi:clinchem.2010.158618 [pii];10.1373/clinchem.2010.158618 [doi].

Bai, M. F., C. Y. Gao, C. K. Yang, X. P. Wang, J. Liu, D. T. Qi, Y. Zhang, P. Y. Hao, and M. W. Li. 6-17-2014. "Effects of Thyroid Dysfunction on the Severity of Coronary Artery Lesions and Its Prognosis." *J.Cardiol.* doi:S0914-5087(14)00104-X [pii];10.1016/j.jjcc.2014.03.009 [doi].

Baker, M. 2010. "Next-Generation Sequencing: Adjusting to Data Overload." *Nat.Meth.* 7(7):495-99. doi:10.1038/nmeth0710-495.

Barnabe, C., L. Bessette, C. Flanagan, S. Leclercq, A. Steiman, F. Kalache, T. Kung, J. E. Pope, B. Haraoui, J. Hochman, D. Mosher, C. Thorne, and V. Bykerk. 2012. "Sex Differences in Pain Scores and Localization in Inflammatory Arthritis: a Systematic Review and Metaanalysis." *J.Rheumatol.* 39(6):1221-30. doi:jrheum.111393 [pii];10.3899/jrheum.111393 [doi].

Barrows, R. C., Jr. and P. D. Clayton. 1996. "Privacy, Confidentiality, and Electronic Medical Records." *J.Am.Med.Inform.Assoc.* 3(2):139-48.

Bassett, J. H. and G. R. Williams. 2008. "Critical Role of the Hypothalamic-Pituitary-Thyroid Axis in Bone." *Bone* 43(3):418-26. doi:S8756-3282(08)00242-1 [pii];10.1016/j.bone.2008.05.007 [doi].

Bavi, P., Z. Jehan, R. Bu, S. Prabhakaran, N. Al-Sanea, F. Al-Dayel, M. Al-Assiri, T. Al-Halouly, R. Sairafi, S. Uddin, and K. S. Al-Kuraya. 11-12-2013. "ALK Gene Amplification Is Associated With Poor Prognosis in Colorectal Carcinoma." *Br.J.Cancer* 109(10):2735-43. doi:bjc2013641 [pii];10.1038/bjc.2013.641 [doi].

Beesley, J., H. A. Pickett, S. E. Johnatty, A. M. Dunning, X. Chen, J. Li, K. Michailidou, Y. Lu, D. N. Rider, R. T. Palmieri, M. D. Stutz, D. Lambrechts, E. Despierre, S. Lambrechts, I. Vergote, J. Chang-Claude, S. Nickels, A. Vrieling, D. Flesch-Janys, S. Wang-Gohrke, U. Eilber, N. Bogdanova, N. Antonenkova, I. B. Runnebaum, T. Dork, M. T. Goodman, G. Lurie, L. R. Wilkens, R. K. Matsuno, L. A. Kiemeney, K. K. Aben, T. Marees, L. F. Massuger, B. L. Fridley, R. A. Vierkant, E. V. Bandera, S. H. Olson, I. Orlow, L. Rodriguez-Rodriguez, L. S. Cook, N. D. Le, A. Brooks-Wilson, L. E. Kelemen, I. Campbell, S. A. Gayther, S. J. Ramus, A. Gentry-Maharaj, U. Menon, S. Ahmed, C. Baynes, P. D. Pharoah, K. Muir, A. Lophatananon, A. Chaiwerawattana, S. Wiangnon, S. Macgregor, D. F. Easton, R. R. Reddel, E. L. Goode, and G. Chenevix-Trench. 2011. "Functional Polymorphisms in the TERT Promoter Are Associated With Risk of Serous Epithelial Ovarian and Breast Cancers." *PLoS One* 6(9):e24987. doi:10.1371/journal.pone.0024987 [doi];PONE-D-11-15672 [pii].

Berkman, N. D., S. L. Sheridan, K. E. Donahue, D. J. Halpern, A. Viera, K. Crotty, A. Holland, M. Brasure, K. N. Lohr, E. Harden, E. Tant, I. Wallace, and M. Viswanathan. 2011. "Health Literacy Interventions and Outcomes: an Updated Systematic Review." *Evid.Rep.Technol.Assess.(Full.Rep.)*(199):1-941.

Berndt, S. I., S. Gustafsson, R. Magi, A. Ganna, E. Wheeler, M. F. Feitosa, A. E. Justice, K. L. Monda, D. C. Croteau-Chonka, F. R. Day, T. Esko, T. Fall, T. Ferreira, D. Gentilini, A. U. Jackson, J. Luan, J. C. Randall, S. Vedantam, C. J. Willer, T. W. Winkler, A. R. Wood, T. Workalemahu, Y. J. Hu, S. H. Lee, L. Liang, D. Y. Lin, J. L. Min, B. M. Neale, G. Thorleifsson, J. Yang, E. Albrecht, N. Amin, J. L. Bragg-Gresham, G. Cadby, Heijer M. den, N. Eklund, K. Fischer, A. Goel, J. J. Hottenga, J. E. Huffman, I. Jarick, A. Johansson, T. Johnson, S. Kanoni, M. E. Kleber, I. R. Konig, K. Kristiansson, Z. Kutalik, C. Lamina, C. Lecoeur, G. Li, M. Mangino, W. L. McArdle, C. Medina-Gomez, M. Muller-Nurasyid, J. S. Ngwa, I. M. Nolte, L. Paternoster, S. Pechlivanis, M. Perola, M. J. Peters, M. Preuss, L. M. Rose, J. Shi, D. Shungin, A. V. Smith, R. J. Strawbridge, I. Surakka, A. Teumer, M. D. Trip, J. Tyrer, J. V. van Vliet-Ostaptchouk, L. Vandenput, L. L. Waite, J. H. Zhao, D. Absher, F. W. Asselbergs, M. Atalay, A. P. Attwood, A. J. Balmforth, H. Basart, J. Beilby, L. L. Bonnycastle, P. Brambilla, M. Bruinenberg, H. Campbell, D. I. Chasman, P. S. Chines, F. S. Collins, J. M. Connell, W. O. Cookson, Faire U. de, Vegt F. de, M. Dei, M. Dimitriou, S. Edkins, K. Estrada, D. M. Evans, M. Farrall, M. M. Ferrario, J. Ferrieres, L. Franke, F. Frau, P. V. Gejman, H. Grallert, H. Gronberg, V. Gudnason, A. S. Hall, P. Hall, A. L. Hartikainen, C. Hayward, N. L. Heard-Costa, A. C. Heath, J. Hebebrand, G. Homuth, F. B. Hu, S. E. Hunt, E. Hypponen, C. Iribarren, K. B. Jacobs, J. O. Jansson, A. Jula, M. Kahonen, S. Kathiresan, F. Kee, K. T. Khaw, M. Kivimaki, W. Koenig, A. T. Kraja, M. Kumari, K. Kuulasmaa, J. Kuusisto, J. H. Laitinen, T. A. Lakka, C. Langenberg, L. J. Launer, L. Lind, J. Lindstrom, J. Liu, A. Liuzzi, M. L. Lokki, M. Lorentzon, P. A. Madden, P. K. Magnusson, P. Manunta, D. Marek, W. Marz, Leach Mateo, I, B. McKnight, S. E. Medland, E. Mihailov, L. Milani, G. W. Montgomery, V. Mooser, T. W.

Muhleisen, P. B. Munroe, A. W. Musk, N. Narisu, G. Navis, G. Nicholson, E. A. Nohr, K. K. Ong, B. A. Oostra, C. N. Palmer, A. Palotie, J. F. Peden, N. Pedersen, A. Peters, O. Polasek, A. Pouta, P. P. Pramstaller, I. Prokopenko, C. Putter, A. Radhakrishnan, O. Raitakari, A. Rendon, F. Rivadeneira, I. Rudan, T. E. Saaristo, J. G. Sambrook, A. R. Sanders, S. Sanna, J. Saramies, S. Schipf, S. Schreiber, H. Schunkert, S. Y. Shin, S. Signorini, J. Sinisalo, B. Skrobek, N. Soranzo, A. Stancakova, K. Stark, J. C. Stephens, K. Stirrups, R. P. Stolk, M. Stumvoll, A. J. Swift, E. V. Theodoraki, B. Thorand, D. A. Tregouet, E. Tremoli, M. M. Van der Klauw, J. B. van Meurs, S. H. Vermeulen, J. Viikari, J. Virtamo, V. Vitart, G. Waeber, Z. Wang, E. Widen, S. H. Wild, G. Willemsen, B. R. Winkelmann, J. C. Witteman, B. H. Wolffenbuttel, A. Wong, A. F. Wright, M. C. Zillikens, P. Amouyel, B. O. Boehm, E. Boerwinkle, D. I. Boomsma, M. J. Caulfield, S. J. Chanock, L. A. Cupples, D. Cusi, G. V. Dedoussis, J. Erdmann, J. G. Eriksson, P. W. Franks, P. Froguel, C. Gieger, U. Gyllensten, A. Hamsten, T. B. Harris, C. Hengstenberg, A. A. Hicks, A. Hingorani, A. Hinney, A. Hofman, K. G. Hovingh, K. Hveem, T. Illig, M. R. Jarvelin, K. H. Jockel, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, D. Kuh, M. Laakso, T. Lehtimaki, D. F. Levinson, N. G. Martin, A. Metspalu, and A. D. Morris. 2013. "Genome-Wide Meta-Analysis Identifies 11 New Loci for Anthropometric Traits and Provides Insights into Genetic Architecture." *Nat.Genet.* 45(5):501-12. doi:ng.2606 [pii];10.1038/ng.2606 [doi].

Berner, E. S. 2002. "Ethical and Legal Issues in the Use of Clinical Decision Support Systems." *J.Healthc.Inf.Manag.* 16(4):34-37.

Biro, F. M., P. Khoury, and J. A. Morrison. 2006. "Influence of Obesity on Timing of Puberty." *Int.J.Androl* 29(1):272-77. doi:IJA602 [pii];10.1111/j.1365-2605.2005.00602.x [doi].

Blanck, H. M., M. Marcus, P. E. Tolbert, C. Rubin, A. K. Henderson, V. S. Hertzberg, R. H. Zhang, and L. Cameron. 2000. "Age at Menarche and Tanner Stage in Girls Exposed in Utero and Postnatally to Polybrominated Biphenyl." *Epidemiology* 11(6):641-47.

Blavin, F., C. Ramos, A. Shah, and K. Devers, Blavin, F., C. Ramos, A. Shah, and K. Devers. 2013. *Final Report: Lessons from the literature on electronic health record implementation 2013*.

Bloss, C. S., N. E. Wineinger, B. F. Darst, N. J. Schork, and E. J. Topol. 2013. "Impact of Direct-to-Consumer Genomic Testing at Long Term Follow-Up." *J.Med.Genet.* 50(6):393-400. doi:jmedgenet-2012-101207 [pii];10.1136/jmedgenet-2012-101207 [doi].

Blumenthal, D. 2-4-2010. "Launching HITECH." *N.Engl.J.Med.* 362(5):382-85. doi:NEJMp0912825 [pii];10.1056/NEJMp0912825 [doi].

-----. 12-22-2011. "Implementation of the Federal Health Information Technology Initiative." *N.Engl.J.Med.* 365(25):2426-31. doi:10.1056/NEJMsr1112158 [doi].

Blumenthal, D. and M. Tavenner. 8-5-2010. "The "Meaningful Use" Regulation for Electronic Health Records." *N.Engl.J.Med.* 363(6):501-4. doi:NEJMp1006114 [pii];10.1056/NEJMp1006114 [doi].

Bobrow, K. L., M. A. Quigley, J. Green, G. K. Reeves, and V. Beral. 2013. "Persistent Effects of Women's Parity and Breastfeeding Patterns on Their Body Mass Index: Results From the Million Women Study." *Int.J.Obes.(Lond.)* 37(5):712-17. doi:ijo201276 [pii];10.1038/ijo.2012.76 [doi].

Boland, M. R., G. Hripcsak, Y. Shen, W. K. Chung, and C. Weng. 2013. "Defining a Comprehensive Verotype Using Electronic Health Records for Personalized Medicine." *J.Am.Med.Inform.Assoc.* 20(e2):e232-e238. doi:amiajnl-2013-001932 [pii];10.1136/amiajnl-2013-001932 [doi].

Boonstra, A. and M. Broekhuis. 2010. "Barriers to the Acceptance of Electronic Medical Records by Physicians From Systematic Review to Taxonomy and Interventions." *BMC.Health Serv.Res.* 10:231. doi:1472-6963-10-231 [pii];10.1186/1472-6963-10-231 [doi].

Bossuyt, P. M., J. B. Reitsma, K. Linnet, and K. G. Moons. 2012. "Beyond Diagnostic Accuracy: the Clinical Utility of Diagnostic Tests." *Clin.Chem.* 58(12):1636-43. doi:clinchem.2012.182576 [pii];10.1373/clinchem.2012.182576 [doi].

Boston, R. C. and A. E. Sumner. 2003. "STATA: a Statistical Analysis System for Examining Biomedical Data." *Adv.Exp.Med.Biol.* 537:353-69.

Botkin, J. R., K. R. Smith, R. T. Croyle, B. J. Baty, J. E. Wylie, D. Dutson, A. Chan, H. A. Hamann, C. Lerman, J. McDonald, V. Venne, J. H. Ward, and E. Lyon. 4-30-2003. "Genetic Testing for a BRCA1 Mutation: Prophylactic Surgery and Screening Behavior in Women 2 Years Post Testing." *Am.J.Med.Genet.A.* 118A(3):201-9. doi:10.1002/ajmg.a.10102 [doi].

Botkin, J. R., S. M. Teutsch, C. I. Kaye, M. Hayes, J. E. Haddow, L. A. Bradley, K. Szegda, and W. D. Dotson. 2010. "Outcomes of Interest in Evidence-Based Evaluations of Genetic Tests." *Genet.Med.* 12(4):228-35. doi:10.1097/GIM.0b013e3181cdde04 [doi].

Boucai, L. and M. I. Surks. 2009. "Reference Limits of Serum TSH and Free T4 Are Significantly Influenced by Race and Age in an Urban Outpatient Medical Practice." *Clin.Endocrinol.(Oxf.)* 70(5):788-93. doi:CEN3390 [pii];10.1111/j.1365-2265.2008.03390.x [doi].

Boyce, B. F. and L. Xing. 5-15-2008. "Functions of RANKL/RANK/OPG in Bone Modeling and Remodeling." *Arch.Biochem.Biophys.* 473(2):139-46. doi:S0003-9861(08)00159-8 [pii];10.1016/j.abb.2008.03.018 [doi].

Brautbar, A., L. A. Pompeii, A. Dehghan, J. S. Ngwa, V. Nambi, S. S. Virani, F. Rivadeneira, A. G. Uitterlinden, A. Hofman, J. C. Witteman, M. J. Pencina, A. R. Folsom, L. A. Cupples, C. M. Ballantyne, and E. Boerwinkle. 2012. "A Genetic Risk Score Based on Direct Associations With Coronary Heart Disease Improves Coronary Heart Disease Risk Prediction in the Atherosclerosis Risk in Communities (ARIC), but Not in the Rotterdam and Framingham Offspring, Studies." *Atherosclerosis.* 223(2):421-26. doi:S0021-9150(12)00345-0 [pii];10.1016/j.atherosclerosis.2012.05.035 [doi].

Breiman, L., J. H. Friedman, and R. A. Olshen. "Stone, CJ (1984) Classification and Regression Trees." *Wadsworth, Belmont, California*.

Breuleux, M. 2007. "Role of Heregulin in Human Cancer." *Cell Mol.Life.Sci.* 64(18):2358-77. doi:10.1007/s00018-007-7120-0 [doi].

Brix, T. H., P. S. Hansen, K. O. Kyvik, and L. Hegedus. 3-13-2000. "Cigarette Smoking and Risk of Clinically Overt Thyroid Disease: a Population-Based Twin Case-Control Study." *Arch.Intern.Med.* 160(5):661-66.

Brown, L. A., S. E. Kalloger, M. A. Miller, IeM Shih, S. E. McKinney, J. L. Santos, K. Swenerton, P. T. Spellman, J. Gray, C. B. Gilks, and D. G. Huntsman. 2008. "Amplification of 11q13 in Ovarian Carcinoma." *Genes Chromosomes.Cancer* 47(6):481-89. doi:10.1002/gcc.20549 [doi].

Brownstein, J. S., S. N. Murphy, A. B. Goldfine, R. W. Grant, M. Sordo, V. Gainer, J. A. Colecchi, A. Dubey, D. M. Nathan, J. P. Glaser, and I. S. Kohane. 2010. "Rapid Identification of Myocardial Infarction Risk Associated With Diabetes Medications Using Electronic Medical Records." *Diabetes Care* 33(3):526-31. doi:dc09-1506 [pii];10.2337/dc09-1506 [doi].

Buber, J., J. Mathew, A. J. Moss, W. J. Hall, A. Barsheshet, S. McNitt, J. L. Robinson, W. Zareba, M. J. Ackerman, E. S. Kaufman, D. Luria, M. Eldar, J. A. Towbin, M. VINCENT, and I. Goldenberg. 6-21-2011. "Risk of Recurrent Cardiac Events After Onset of Menopause in Women With Congenital Long-QT Syndrome Types 1 and 2." *Circulation* 123(24):2784-91. doi:CIRCULATIONAHA.110.000620 [pii];10.1161/CIRCULATIONAHA.110.000620 [doi].

Bunnik, E. M., A. C. Janssens, and M. H. Schermer. 5-28-2014. "Personal Utility in Genomic Testing: Is There Such a Thing?" *J.Med.Ethics.* doi:medethics-2013-101887 [pii];10.1136/medethics-2013-101887 [doi].

Bunnik, E. M., M. H. Schermer, and A. C. Janssens. 2011. "Personal Genome Testing: Test Characteristics to Clarify the Discourse on Ethical, Legal and Societal Issues." *BMC.Med.Ethics.* 12:11. doi:1472-6939-12-11 [pii];10.1186/1472-6939-12-11 [doi].

Burchard, E. G., E. Ziv, N. Coyle, S. L. Gomez, H. Tang, A. J. Karter, J. L. Mountain, E. J. Perez-Stable, D. Sheppard, and N. Risch. 3-20-2003. "The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice." *N.Engl.J.Med.* 348(12):1170-1175. doi:10.1056/NEJMsb025007 [doi];348/12/1170 [pii].

Bureau of the Census.,11-1-2012. *2010 Census Briefs Age and Sex Composition: 2010. C2010BR-03* Washington, DC: Bureau of the Census.

Burger, H. G., G. E. Hale, D. M. Robertson, and L. Dennerstein. 2007. "A Review of Hormonal Changes During the Menopausal Transition: Focus on Findings From the Melbourne Women's Midlife Health Project." *Hum.Reprod.Update.* 13(6):559-65. doi:dmm020 [pii];10.1093/humupd/dmm020 [doi].

Butts, S. F., M. D. Sammel, C. Greer, T. R. Rebbeck, D. W. Boorman, and E. W. Freeman. 2014. "Cigarettes, Genetic Background, and Menopausal Timing: the Presence of Single Nucleotide Polymorphisms in Cytochrome P450 Genes Is Associated With Increased Risk of Natural Menopause in European-American Smokers." *Menopause.* 21(7):694-701. doi:10.1097/GME.0000000000000140 [doi].

Buyske, S., Y. Wu, C. L. Carty, I. Cheng, T. L. Assimes, L. Dumitrescu, L. A. Hindorff, S. Mitchell, J. L. Ambite, E. Boerwinkle, P. Buzkova, C. S. Carlson, B. Cochran, D. Duggan, C. B. Eaton, M. D. Fesinmeyer, N. Franceschini, J. Haessler, N. Jenny, H. M. Kang, C. Kooperberg, Y. Lin, Marchand L. Le, T. C. Matise, J. G. Robinson, C. Rodriguez, F. R. Schumacher, B. F. Voight, A. Young, T. A. Manolio, K. L. Mohlke, C. A. Haiman, U. Peters, D. C. Crawford, and K. E. North. 2012. "Evaluation of the Metabochip Genotyping Array in African Americans and Implications for Fine Mapping of GWAS-Identified Loci: the PAGE Study." *PLoS.One.* 7(4):e35651. doi:10.1371/journal.pone.0035651 [doi];PONE-D-12-01108 [pii].

Caine, K. and R. Hanania. 1-1-2013. "Patients Want Granular Privacy Control Over Health Information in Electronic Medical Records." *J.Am.Med.Inform.Assoc.* 20(1):7-15. doi:amiajnl-2012-001023 [pii];10.1136/amiajnl-2012-001023 [doi].

Campbell Jenkins, B. W., C. Addison, G. Wilson, J. Liu, M. Fortune, K. Robinson, M. White, and D. Sarpong. 2011. "Association of the Joint Effect of Menopause and Hormone Replacement Therapy and Cancer in African American Women: the Jackson Heart Study." *Int.J.Environ.Res.Public Health* 8(6):2491-504. doi:10.3390/ijerph8062491 [doi];ijerph-08-02491 [pii].

Carpenter, M. J., C. Strange, Y. Jones, M. R. Dickson, C. Carter, M. A. Moseley, and G. E. Gilbert. 2007. "Does Genetic Testing Result in Behavioral Health Change? Changes in Smoking Behavior Following Testing for Alpha-1 Antitrypsin Deficiency." *Ann.Behav.Med.* 33(1):22-28. doi:10.1207/s15324796abm3301_3 [doi].

Carr, M. C. 2003. "The Emergence of the Metabolic Syndrome With Menopause." *J.Clin.Endocrinol.Metab* 88(6):2404-11.

Carter, C. L., E. M. Resnick, M. Mallampalli, and A. Kalbarczyk. 2012. "Sex and Gender Differences in Alzheimer's Disease: Recommendations for Future Research." *J.Womens.Health (Larchmt.)* 21(10):1018-23. doi:10.1089/jwh.2012.3789 [doi].

Castillo, R. S. and A. Kelemen. 2013. "Considerations for a Successful Clinical Decision Support System." *CIN:Computers Informatics Nursing* 31(7):319-26.

Caudle, K. E., T. E. Klein, J. M. Hoffman, D. J. Muller, M. Whirl-Carrillo, L. Gong, E. M. McDonagh, K. Sangkuhl, C. F. Thorn, M. Schwab, J. A. Agundez, R. R. Freimuth, V. Huser, M. T. Lee, O. F. Iwuchukwu, K. R. Crews, S. A. Scott, M. Wadelius, J. J. Swen, R. F. Tyndale, C. M. Stein, D. Roden, M. V. Relling, M. S. Williams, and S. G. Johnson. 2014. "Incorporation of Pharmacogenomics into Routine Clinical Practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process." *Curr.Drug Metab.* 15(2):209-17. doi:CDM-EPUB-58901 [pii].

Caudle, K. E., C. F. Thorn, T. E. Klein, J. J. Swen, H. L. McLeod, R. B. Diasio, and M. Schwab. 2013. "Clinical Pharmacogenetics Implementation Consortium Guidelines for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing." *Clin.Pharmacol.Ther.* 94(6):640-645. doi:clpt2013172 [pii];10.1038/clpt.2013.172 [doi].

Cecchini, R. S., J. P. Costantino, J. A. Cauley, W. M. Cronin, D. L. Wickerham, S. R. Land, J. L. Weissfeld, and N. Wolmark. 2012. "Body Mass Index and the Risk for Developing Invasive Breast Cancer Among High-Risk Women in NSABP P-1 and STAR Breast Cancer Prevention Trials." *Cancer Prev.Res.(Phila)* 5(4):583-92. doi:1940-6207.CAPR-11-0482 [pii];10.1158/1940-6207.CAPR-11-0482 [doi].

Centers for Disease Control and Prevention.,*Diabetes Public Health Resource* .

-----.,*Colorectal (Colon) Cancer What Are the Risk Factors?*

Centers for Disease Control and Prevention and U.S.Department of Health and Human Services.,*National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011.*

Cha, P. C., A. Takahashi, N. Hosono, S. K. Low, N. Kamatani, M. Kubo, and Y. Nakamura. 2011. "A Genome-Wide Association Study Identifies Three Loci Associated With Susceptibility to Uterine Fibroids." *Nat.Genet.* 43(5):447-50. doi:ng.805 [pii];10.1038/ng.805 [doi].

Chakravarthy, U., G. J. McKay, P. T. De Jong, M. Rahu, J. Seland, G. Soubrane, L. Tomazzoli, F. Topouzis, J. R. Vingerling, J. Vioque, I. S. Young, R. Sofat, A. D. Hingorani, and A. E. Fletcher. 2013. "ARMS2 Increases the Risk of Early and Late Age-Related Macular Degeneration in the European Eye Study." *Ophthalmology.* 120(2):342-48. doi:S0161-6420(12)00759-2 [pii];10.1016/j.ophtha.2012.08.004 [doi].

Chao, S., J. S. Roberts, T. M. Marteau, R. Silliman, L. A. Cupples, and R. C. Green. 2008. "Health Behavior Changes After Genetic Risk Assessment for Alzheimer Disease: The REVEAL Study." *Alzheimer.Dis.Assoc.Disord.* 22(1):94-97. doi:10.1097/WAD.0b013e31815a9dcc [doi].

Chen, C. T., L. Fernandez-Rhodes, R. G. Brzyski, C. S. Carlson, Z. Chen, G. Heiss, K. E. North, N. F. Woods, A. Rajkovic, C. Kooperberg, and N. Franceschini. 3-15-2012a. "Replication of Loci Influencing Ages at Menarche and Menopause in Hispanic Women: the Women's Health Initiative SHARe Study." *Hum.Mol.Genet.* 21(6):1419-32. doi:ddr570 [pii];10.1093/hmg/ddr570 [doi].

Chen, C. T., C. T. Liu, G. K. Chen, J. S. Andrews, A. M. Arnold, J. Dreyfus, N. Franceschini, M. E. Garcia, K. F. Kerr, G. Li, K. K. Lohman, S. K. Musani, M. A. Nalls, L. J. Raffel, J. Smith, C. B. Ambrosone, E. V. Bandera, L. Bernstein, A. Britton, R. G. Brzyski, A. Cappola, C. S. Carlson, D. Couper, S. L. Deming, M. O. Goodarzi, G. Heiss, E. M. John, X. Lu, Marchand L. Le, K. Marciante, B. McKnight, R. Millikan, N. L. Nock, A. F. Olshan, M. F. Press, D. Vaiyda, N. F. Woods, H. A. Taylor, W. Zhao, W. Zheng, M. K. Evans, T. B. Harris, B. E. Henderson, S. L. Kardia, C. Kooperberg, Y. Liu, T. H. Mosley, B. Psaty, M. Wellons, B. G. Windham, A. B. Zonderman, L. A. Cupples, E. W. Demerath, C. Haiman,

J. M. Murabito, and A. Rajkovic. 6-15-2014a. "Meta-Analysis of Loci Associated With Age at Natural Menopause in African-American Women." *Hum.Mol.Genet.* 23(12):3327-42. doi:ddu041 [pii];10.1093/hmg/ddu041 [doi].

Chen, D., I. Juko-Pecirep, J. Hammer, E. Ivansson, S. Enroth, I. Gustavsson, L. Feuk, P. K. Magnusson, J. D. McKay, E. Wilander, and U. Gyllensten. 5-1-2013. "Genome-Wide Association Study of Susceptibility Loci for Cervical Cancer." *J.Natl.Cancer Inst.* 105(9):624-33. doi:djt051 [pii];10.1093/jnci/djt051 [doi].

Chen, M. M., M. Crous-Bou, V. W. Setiawan, J. Prescott, S. H. Olson, N. Wentzensen, A. Black, L. Brinton, C. Chen, C. Chen, L. S. Cook, J. Doherty, C. M. Friedenreich, S. E. Hankinson, P. Hartge, B. E. Henderson, D. J. Hunter, Marchand L. Le, X. Liang, J. Lissowska, L. Lu, I. Orlow, S. Petruzella, S. Polidoro, L. Pooler, T. R. Rebbeck, H. Risch, C. Sacerdote, F. Schumacher, X. Sheng, X. O. Shu, N. S. Weiss, L. Xia, D. Van Den Berg, H. P. Yang, H. Yu, S. Chanock, C. Haiman, P. Kraft, and Vivo De, I. 2014b. "Exome-Wide Association Study of Endometrial Cancer in a Multiethnic Population." *PLoS One* 9(5):e97045. doi:10.1371/journal.pone.0097045 [doi];PONE-D-14-08046 [pii].

Chen, X., C. T. Liu, M. Zhang, and H. Zhang. 12-4-2007. "A Forest-Based Approach to Identifying Gene and Gene Gene Interactions." *Proc.Natl.Acad.Sci.U.S.A.* 104(49):19199-203. doi:0709868104 [pii];10.1073/pnas.0709868104 [doi].

Chen, X., Y. B. Xiang, J. R. Long, H. Cai, Q. Cai, J. Cheng, W. Wen, Y. T. Gao, W. Zheng, and X. O. Shu. 7-1-2012b. "Genetic Polymorphisms in Obesity-Related Genes and Endometrial Cancer Risk." *Cancer* 118(13):3356-64. doi:10.1002/cncr.26552 [doi].

Cheng, I., J. M. Kocarnik, L. Dumitrescu, N. M. Lindor, J. Chang-Claude, C. L. Avery, C. P. Caberto, S. A. Love, M. L. Slattery, A. T. Chan, J. A. Baron, L. A. Hindorff, S. L. Park, F. R. Schumacher, M. Hoffmeister, P. Kraft, A. M. Butler, D. J. Duggan, L. Hou, C. S. Carlson, K. R. Monroe, Y. Lin, C. L. Carty, S. Mann, J. Ma, E. L. Giovannucci, C. S. Fuchs, P. A. Newcomb, M. A. Jenkins, J. L. Hopper, R. W. Haile, D. V. Conti, P. T. Campbell, J. D. Potter, B. J. Caan, R. E. Schoen, R. B. Hayes, S. J. Chanock, S. I. Berndt, S. Kury, S. Bezieau, J. L. Ambite, G. Kumaraguruparan, D. M. Richardson, R. J. Goodloe, H. H. Dilks, P. Baker, B. W. Zanke, M. Lemire, S. Gallinger, L. Hsu, S. Jiao, T. A. Harrison, D. Seminara, C. A. Haiman, C. Kooperberg, L. R. Wilkens, C. M. Hutter, E. White, D. C. Crawford, G. Heiss, T. J. Hudson, H. Brenner, W. S. Bush, G. Casey, Marchand L. Le, and U. Peters. 2014. "Pleiotropic Effects of Genetic Risk Variants for Other Cancers on Colorectal Cancer Risk: PAGE, GECCO and CCFR Consortia." *Gut* 63(5):800-807. doi:gutjnl-2013-305189 [pii];10.1136/gutjnl-2013-305189 [doi].

Chiamolera, M. I. and F. E. Wondisford. 2009. "Minireview: Thyrotropin-Releasing Hormone and the Thyroid Hormone Feedback Mechanism." *Endocrinology* 150(3):1091-96. doi:en.2008-1795 [pii];10.1210/en.2008-1795 [doi].

Choi, J. H., J. J. Sheu, B. Guan, N. Jinawath, P. Markowski, T. L. Wang, and IeM Shih. 2-15-2009. "Functional Analysis of 11q13.5 Amplicon Identifies Rsf-1 (HBXAP) As a Gene Involved in Paclitaxel Resistance in Ovarian Cancer." *Cancer Res.* 69(4):1407-15. doi:0008-5472.CAN-08-3602 [pii];10.1158/0008-5472.CAN-08-3602 [doi].

Christensen, K. D., T. E. Jayaratne, J. S. Roberts, S. L. Kardia, and E. M. Petty. 2010a. "Understandings of Basic Genetics in the United States: Results From a National Survey of Black and White Men and Women." *Public.Health Genomics.* 13(7-8):467-76. doi:000293287 [pii];10.1159/000293287 [doi].

Christensen, K. Y., M. Maisonet, C. Rubin, A. Holmes, W. D. Flanders, J. Heron, A. Ness, C. Drews-Botsch, C. Dominguez, M. A. McGeehin, and M. Marcus. 2010b. "Progression Through Puberty in Girls Enrolled in a Contemporary British Cohort." *J.Adolesc.Health* 47(3):282-89. doi:S1054-139X(10)00101-1 [pii];10.1016/j.jadohealth.2010.02.005 [doi].

Chu, X., C. M. Pan, S. X. Zhao, J. Liang, G. Q. Gao, X. M. Zhang, G. Y. Yuan, C. G. Li, L. Q. Xue, M. Shen, W. Liu, F. Xie, S. Y. Yang, H. F. Wang, J. Y. Shi, W. W. Sun, W. H. Du, C. L. Zuo, J. X. Shi, B. L. Liu, C. C. Guo, M. Zhan, Z. H. Gu, X. N. Zhang, F. Sun, Z. Q. Wang, Z. Y. Song, C. Y. Zou, W. H. Sun, T. Guo, H. M. Cao, J. H. Ma, B. Han, P. Li, H. Jiang, Q. H. Huang, L. Liang, L. B. Liu, G. Chen, Q. Su, Y. D. Peng, J. J. Zhao, G. Ning, Z. Chen, J. L. Chen, S. J. Chen, W. Huang, and H. D. Song. 2011. "A Genome-Wide Association Study Identifies Two New Risk Loci for Graves' Disease." *Nat.Genet.* 43(9):897-901. doi:ng.898 [pii];10.1038/ng.898 [doi].

Chumlea, W. C., C. M. Schubert, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and S. S. Sun. 2003. "Age at Menarche and Racial Comparisons in US Girls." *Pediatrics* 111(1):110-113.

Cicchetti, D. and F. A. Rogosch. 2012. "Gene x Environment Interaction and Resilience: Effects of Child Maltreatment and Serotonin, Corticotropin Releasing Hormone, Dopamine, and Oxytocin Genes." *Dev.Psychopathol.* 24(2):411-27. doi:S0954579412000077 [pii];10.1017/S0954579412000077 [doi].

Cirulli, E. T. and D. B. Goldstein. 2010. "Uncovering the Roles of Rare Variants in Common Disease Through Whole-Genome Sequencing." *Nat.Rev.Genet.* 11(6):415-25. doi:nrg2779 [pii];10.1038/nrg2779 [doi].

Clancy, J. P. and M. Jain. 10-1-2012. "Personalized Medicine in Cystic Fibrosis: Dawning of a New Era." *Am.J.Respir.Crit.Care Med.* 186(7):593-97. doi:rccm.201204-0785PP [pii];10.1164/rccm.201204-0785PP [doi].

Clancy, J. P., S. G. Johnson, S. W. Yee, E. M. McDonagh, K. E. Caudle, T. E. Klein, M. Cannavo, and K. M. Giacomini. 2014. "Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for Ivacaftor Therapy in the Context of CFTR Genotype." *Clin.Pharmacol.Ther.* 95(6):592-97. doi:clpt201454 [pii];10.1038/clpt.2014.54 [doi].

Clayton, D. G. 2009. "Prediction and Interaction in Complex Disease Genetics: Experience in Type 1 Diabetes." *PLoS Genet.* 5(7):e1000540. doi:10.1371/journal.pgen.1000540 [doi].

Clayton, E. W., M. Smith, S. M. Fullerton, W. Burke, C. A. McCarty, B. A. Koenig, A. L. McGuire, L. M. Beskow, L. Dressler, A. A. Lemke, E. M. Ramos, and L. L. Rodriguez. 2010. "Confronting Real Time Ethical, Legal, and Social Issues in the Electronic Medical Records and Genomics (EMERGE) Consortium." *Genet.Med.* 12(10):616-20. doi:10.1097/GIM.0b013e3181efdbd0 [doi].

Compton, C. C. 2012. "AJCC Cancer Staging Atlas: A Companion to the Seventh Editions of the AJCC Cancer Staging Manual and Handbook." American Joint Committee on Cancer.

Condit, Celeste M. 2011. "When Do People Deploy Genetic Determinism? A Review Pointing to the Need for MultiΓÇÉfactorial Theories of Public Utilization of Scientific Discourses." *Sociology Compass* 5(7):618-35.

Cooper, G. S. and B. C. Stroehla. 2003. "The Epidemiology of Autoimmune Diseases." *Autoimmun.Rev.* 2(3):119-25. doi:S1568997203000065 [pii].

Cornetta, K. and C. G. Brown. 2013. "Balancing Personalized Medicine and Personalized Care." *Acad.Med.* 88(3):309-13. doi:10.1097/ACM.0b013e3182806345 [doi].

Corrigan, P. 2004. "How Stigma Interferes With Mental Health Care." *Am.Psychol.* 59(7):614-25. doi:2004-19091-003 [pii];10.1037/0003-066X.59.7.614 [doi].

Costa-e-Sousa RH and A. N. Hollenberg. 2012. "Minireview: The Neural Regulation of the Hypothalamic-Pituitary-Thyroid Axis." *Endocrinology* 153(9):4128-35. doi:en.2012-1467 [pii];10.1210/en.2012-1467 [doi].

Cousminer, D. L., E. Stergiakouli, D. J. Berry, W. Ang, M. M. Groen-Blokhuis, A. Korner, N. Siitonen, I. Ntalla, M. Marinelli, J. R. Perry, J. Kettunen, R. Jansen, I. Surakka, N. J. Timpson, S. Ring, G. Mcmahon, C. Power, C. Wang, M. Kahonen, J. Viikari, T. Lehtimaki, C. M. Middeldorp, H. E. Hulshoff Pol, M. Neef, S. Weise, K. Pahkala, H. Niinikoski, E. Zeggini, K. Panoutsopoulou, M. Bustamante, B. W. Penninx, J. Murabito, M. Torrent, G. V. Dedoussis, W. Kiess, D. I. Boomsma, C. E. Pennell, O. T. Raitakari, E. Hypponen, Smith G. Davey, S. Ripatti, M. I. McCarthy, and E. Widen. 8-15-2014. "Genome-Wide Association Study of Sexual Maturation in Males and Females Highlights a Role for Body Mass and Menarche Loci in Male Puberty." *Hum.Mol.Genet.* 23(16):4452-64. doi:ddu150 [pii];10.1093/hmg/ddu150 [doi].

Cramer, D. W., B. L. Harlow, H. Xu, C. Fraer, and R. Barbieri. 1995. "Cross-Sectional and Case-Controlled Analyses of the Association Between Smoking and Early Menopause." *Maturitas* 22(2):79-87. doi:0378-5122(95)00928-E [pii].

Cramer, D. W. and H. Xu. 1996. "Predicting Age at Menopause." *Maturitas* 23(3):319-26. doi:0378512296009929 [pii].

Crawford, D. C., D. R. Crosslin, G. Tromp, I. J. Kullo, H. Kuivaniemi, M. G. Hayes, J. C. Denny, W. S. Bush, J. L. Haines, D. M. Roden, C. A. McCarty, G. P. Jarvik, and M. D. Ritchie. 2014. "EMERGEing Progress in Genomics-the First Seven Years." *Front.Genet.* 5:184. doi:10.3389/fgene.2014.00184 [doi].

Creasman, W. T., M. F. Kohler, F. Odicino, P. Maisonneuve, and P. Boyle. 2004. "Prognosis of Papillary Serous, Clear Cell, and Grade 3 Stage I Carcinoma of the Endometrium." *Gynecol.Oncol.* 95(3):593-96. doi:S0090-8258(04)00645-6 [pii];10.1016/j.ygyno.2004.08.019 [doi].

Creasman, W. T., F. Odicino, P. Maisonneuve, M. A. Quinn, U. Beller, J. L. Benedet, A. P. Heintz, H. Y. Ngan, and S. Pecorelli. 2006. "Carcinoma of the Corpus Uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer." *Int.J.Gynaecol.Obstet.* 95 Suppl 1:S105-S143. doi:S0020-7292(06)60031-3 [pii];10.1016/S0020-7292(06)60031-3 [doi].

Crews, K. R., A. Gaedigk, H. M. Dunnenberger, J. S. Leeder, T. E. Klein, K. E. Caudle, C. E. Haidar, D. D. Shen, J. T. Callaghan, S. Sadhasivam, C. A. Prows, E. D. Kharasch, and T. C. Skaar. 2014. "Clinical Pharmacogenetics Implementation Consortium Guidelines for Cytochrome P450 2D6 Genotype and Codeine Therapy: 2014 Update." *Clin.Pharmacol.Ther.* 95(4):376-82. doi:clpt2013254 [pii];10.1038/clpt.2013.254 [doi].

Crosslin, D. R., A. McDavid, N. Weston, X. Zheng, E. Hart, Andrade M. de, I. J. Kullo, C. A. McCarty, K. F. Doheny, E. Pugh, A. Kho, M. G. Hayes, M. D. Ritchie, A. Saip, D. C. Crawford, P. K. Crane, K. Newton, D. S. Carrell, C. J. Gallego, M. A. Nalls, R. Li, D. B. Mirel, A. Crenshaw, D. J. Couper, T. Tanaka, F. J. van Rooij, M. H. Chen, A. V. Smith, N. A. Zakai, Q. Yango, M. Garcia, Y. Liu, T. Lumley, A. R. Folsom, A. P. Reiner, J. F. Felix, A. Dehghan, J. G. Wilson, J. C. Bis, C. S. Fox, N. L. Glazer, L. A. Cupples, J. Coresh, G. Eiriksdottir, V. Gudnason, S. Bandinelli, T. M. Frayling, A. Chakravarti, C. M. van Duijn, D. Melzer, D. Levy, E. Boerwinkle, A. B. Singleton, D. G. Hernandez, D. L. Longo, J. C. Witteman, B. M. Psaty, L. Ferrucci, T. B. Harris, C. J. O'Donnell, S. K. Ganesh, E. B. Larson, C. S. Carlson, and G. P. Jarvik. 5-15-2013. "Genetic Variation Associated With Circulating Monocyte Count in the EMERGE Network." *Hum.Mol.Genet.* 22(10):2119-27. doi:ddt010 [pii];10.1093/hmg/ddt010 [doi].

Cuesta, I., K. S. Zaret, and P. Santisteban. 2007. "The Forkhead Factor FoxE1 Binds to the Thyroperoxidase Promoter During Thyroid Cell Differentiation and Modifies Compacted Chromatin Structure." *Mol.Cell Biol.* 27(20):7302-14. doi:MCB.00758-07 [pii];10.1128/MCB.00758-07 [doi].

Cui, B., X. Zhu, M. Xu, T. Guo, D. Zhu, G. Chen, X. Li, L. Xu, Y. Bi, Y. Chen, Y. Xu, X. Li, W. Wang, H. Wang, W. Huang, and G. Ning. 2011. "A Genome-Wide Association Study Confirms Previously Reported Loci for Type 2 Diabetes in Han Chinese." *PLoS.One.* 6(7):e22353. doi:10.1371/journal.pone.0022353 [doi];PONE-D-11-06453 [pii].

Culverhouse, R. 2007. "The Use of the Restricted Partition Method With Case-Control Data." *Hum.Hered.* 63(2):93-100. doi:000099181 [pii];10.1159/000099181 [doi].

Cust, A. E., R. Kaaks, C. Friedenreich, F. Bonnet, M. Laville, A. Lukanova, S. Rinaldi, L. Dossus, N. Slimani, E. Lundin, A. Tjonneland, A. Olsen, K. Overvad, F. Clavel-Chapelon, S. Mesrine, V. Joulin, J. Linseisen, S. Rohrmann, T. Pischon, H. Boeing, D. Trichopoulos, A. Trichopoulou, V. Benetou, D. Palli, F. Berrino, R. Tumino, C. Sacerdote, A. Mattiello, J. R. Quiros, M. A. Mendez, M. J. Sanchez, N. Larranaga, M. J. Tormo, E. Ardanaz, H. B. Bueno-de-Mesquita, P. H. Peeters, C. H. van Gils, K. T. Khaw, S. Bingham, N. Allen, T. Key, M. Jenab, and E. Riboli. 2007. "Plasma Adiponectin Levels and Endometrial Cancer Risk in Pre- and Postmenopausal Women." *J.Clin.Endocrinol.Metab.* 92(1):255-63. doi:jc.2006-1371 [pii];10.1210/jc.2006-1371 [doi].

D'Agostino, R. B., Sr., R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel. 2-12-2008. "General Cardiovascular Risk Profile for Use in Primary Care: the Framingham Heart Study." *Circulation* 117(6):743-53. doi:CIRCULATIONAHA.107.699579 [pii];10.1161/CIRCULATIONAHA.107.699579 [doi].

Dambrun, Micha+ l, Rodolphe Kamiejski, Nicolas Haddadi, and Sandra Duarte. 2009. "Why Does Social Dominance Orientation Decrease With University Exposure to the Social Sciences? The Impact of Institutional Socialization and the Mediating Role of ΓÇ£GeneticismΓÇ¥." *European Journal of Social Psychology* 39(1):88-100.

Das, K. J., J. Ingles, R. D. Bagnall, and C. Semsarian. 2014. "Determining Pathogenicity of Genetic Variants in Hypertrophic Cardiomyopathy: Importance of Periodic Reassessment." *Genet.Med.* 16(4):286-93. doi:gim2013138 [pii];10.1038/gim.2013.138 [doi].

Davies, J. C., C. E. Wainwright, G. J. Canny, M. A. Chilvers, M. S. Howenstine, A. Munck, J. G. Mainz, S. Rodriguez, H. Li, K. Yen, C. L. Ordonez, and R. Ahrens. 6-1-2013. "Efficacy and Safety of Ivacaftor in Patients Aged 6 to 11 Years With Cystic Fibrosis With a G551D Mutation." *Am.J.Respir.Crit.Care Med.* 187(11):1219-25. doi:10.1164/rccm.201301-0153OC [doi].

De Castro-Oros, I., S. Pampin, A. Bolado-Carrancio, Cubas A. De, L. Palacios, N. Plana, J. Puzo, E. Martorell, M. Stef, L. Masana, F. Civeira, J. C. Rodriguez-Rey, and M. Pocovi. 2011. "Functional Analysis of LDLR Promoter and 5' UTR Mutations in Subjects With Clinical Diagnosis of Familial Hypercholesterolemia." *Hum.Mutat.* 32(8):868-72. doi:10.1002/humu.21520 [doi].

De Wit, J. and W. Stroebe. 2004. "Health Psychology." Pp. 52-83 in *Social Cognition Models of Health Behaviour.*, edited by A.Kaptein and J.Weinman. Oxford: Blackwell.

De, Felice M. and Lauro R. Di. 2004. "Thyroid Development and Its Disorders: Genetics and Molecular Mechanisms." *Endocr.Rev.* 25(5):722-46. doi:25/5/722 [pii];10.1210/er.2003-0028 [doi].

De, Pergola G., A. Ciampolillo, S. Paolotti, P. Trerotoli, and R. Giorgino. 2007. "Free Triiodothyronine and Thyroid Stimulating Hormone Are Directly Associated With Waist Circumference, Independently of Insulin Resistance, Metabolic Parameters and Blood Pressure in Overweight and Obese Women." *Clin.Endocrinol.(Oxf.)* 67(2):265-69. doi:CEN2874 [pii];10.1111/j.1365-2265.2007.02874.x [doi].

De, R., W. S. Bush, and J. H. Moore. 2014. "Bioinformatics Challenges in Genome-Wide Association Studies (GWAS)." *Methods.Mol.Biol.* 1168:63-81. doi:10.1007/978-1-4939-0847-9_5 [doi].

De, Vivo, I, J. Prescott, V. W. Setiawan, S. H. Olson, N. Wentzensen, J. Attia, A. Black, L. Brinton, C. Chen, C. Chen, L. S. Cook, M. Crous-Bou, J. Doherty, A. M. Dunning, D. F. Easton, C. M. Friedenreich, M. Garcia-Closas, M. M. Gaudet, C. Haiman, S. E. Hankinson, P. Hartge, B. E. Henderson, E. Holliday, P. L. Horn-Ross, D. J. Hunter, Marchand L. Le, X.

Liang, J. Lissowska, J. Long, L. Lu, A. M. Magliocco, M. McEvoy, T. A. O'Mara, I. Orlow, J. N. Painter, L. Pooler, R. Rastogi, T. R. Rebbeck, H. Risch, C. Sacerdote, F. Schumacher, R. J. Scott, X. Sheng, X. O. Shu, A. B. Spurdle, D. Thompson, D. Vanden Berg, N. S. Weiss, L. Xia, Y. B. Xiang, H. P. Yang, H. Yu, W. Zheng, S. Chanock, and P. Kraft. 2014. "Genome-Wide Association Study of Endometrial Cancer in E2C2." *Hum.Genet.* 133(2):211-24. doi:10.1007/s00439-013-1369-1 [doi].

Delahanty, R. J., A. Beeghly-Fadiel, J. R. Long, Y. T. Gao, W. Lu, Y. B. Xiang, Y. Zheng, B. T. Ji, W. Q. Wen, Q. Y. Cai, W. Zheng, and X. O. Shu. 2013. "Evaluation of GWAS-Identified Genetic Variants for Age at Menarche Among Chinese Women." *Hum.Reprod.* 28(4):1135-43. doi:det011 [pii];10.1093/humrep/det011 [doi].

Demerath, E. W., C. T. Liu, N. Franceschini, G. Chen, J. R. Palmer, E. N. Smith, C. T. Chen, C. B. Ambrosone, A. M. Arnold, E. V. Bandera, G. S. Berenson, L. Bernstein, A. Britton, A. R. Cappola, C. S. Carlson, S. J. Chanock, W. Chen, Z. Chen, S. L. Deming, C. E. Elks, M. K. Evans, Z. Gajdos, B. E. Henderson, J. J. Hu, S. Ingles, E. M. John, K. F. Kerr, L. N. Kolonel, Marchand L. Le, X. Lu, R. C. Millikan, S. K. Musani, N. L. Nock, K. North, S. Nyante, M. F. Press, J. L. Rodriquez-Gil, E. A. Ruiz-Narvaez, N. J. Schork, S. R. Srinivasan, N. F. Woods, W. Zheng, R. G. Ziegler, A. Zonderman, G. Heiss, Windham B. Gwen, M. Wellons, S. S. Murray, M. Nalls, T. Pastinen, A. Rajkovic, J. Hirschhorn, Cupples L. Adrienne, C. Kooperberg, J. M. Murabito, and C. A. Haiman. 8-15-2013. "Genome-Wide Association Study of Age at Menarche in African-American Women." *Hum.Mol.Genet.* 22(16):3329-46. doi:ddt181 [pii];10.1093/hmg/ddt181 [doi].

Denny, J. C., D. C. Crawford, M. D. Ritchie, S. J. Bielinski, M. A. Basford, Y. Bradford, H. S. Chai, L. Bastarache, R. Zuvich, P. Peissig, D. Carrell, A. H. Ramirez, J. Pathak, R. A. Wilke, L. Rasmussen, X. Wang, J. A. Pacheco, A. N. Kho, M. G. Hayes, N. Weston, M. Matsumoto, P. A. Kopp, K. M. Newton, G. P. Jarvik, R. Li, T. A. Manolio, I. J. Kullo, C. G. Chute, R. L. Chisholm, E. B. Larson, C. A. McCarty, D. R. Masys, D. M. Roden, and Andrade M. de. 10-7-2011. "Variants Near FOXE1 Are Associated With Hypothyroidism and Other Thyroid Conditions: Using Electronic Medical Records for Genome- and Phenome-Wide Studies." *Am.J.Hum.Genet.* 89(4):529-42. doi:S0002-9297(11)00398-3 [pii];10.1016/j.ajhg.2011.09.008 [doi].

Denny, J. C., M. D. Ritchie, M. A. Basford, J. M. Pulley, L. Bastarache, K. Brown-Gentry, D. Wang, D. R. Masys, D. M. Roden, and D. C. Crawford. 5-1-2010. "PheWAS: Demonstrating the Feasibility of a Phenome-Wide Scan to Discover Gene-Disease Associations." *Bioinformatics.* 26(9):1205-10. doi:btq126 [pii];10.1093/bioinformatics/btq126 [doi].

Diakou, M., G. Miltiadous, S. L. Xenophontos, P. Manoli, M. A. Cariolou, and M. Elisaf. 2011. "Spectrum of LDLR Gene Mutations, Including a Novel Mutation Causing Familial Hypercholesterolaemia, in North-Western Greece." *Eur.J.Intern.Med.* 22(5):e55-e59. doi:S0953-6205(11)00008-2 [pii];10.1016/j.ejim.2011.01.003 [doi].

Dieudonne, M. N., M. Bussiere, Santos E. Dos, M. C. Leneveu, Y. Giudicelli, and R. Pecquery. 6-23-2006. "Adiponectin Mediates Antiproliferative and Apoptotic Responses in Human

MCF7 Breast Cancer Cells." *Biochem.Biophys.Res.Commun.* 345(1):271-79. doi:S0006-291X(06)00919-3 [pii];10.1016/j.bbrc.2006.04.076 [doi].

Ding, K., K. Shameer, H. Jouni, D. R. Masys, G. P. Jarvik, A. N. Kho, M. D. Ritchie, C. A. McCarty, C. G. Chute, T. A. Manolio, and I. J. Kullo. 2012. "Genetic Loci Implicated in Erythroid Differentiation and Cell Cycle Regulation Are Associated With Red Blood Cell Traits." *Mayo.Clin.Proc.* 87(5):461-74. doi:S0025-6196(12)00303-5 [pii];10.1016/j.mayocp.2012.01.016 [doi].

Dishi, M., D. A. Enquobahrie, D. F. Abetew, C. Qiu, C. B. Rudra, and M. A. Williams. 2011. "Age at Menarche, Menstrual Cycle Characteristics and Risk of Gestational Diabetes." *Diabetes Res.Clin.Pract.* 93(3):437-42. doi:S0168-8227(11)00351-2 [pii];10.1016/j.diabres.2011.07.001 [doi].

DiVall, S. A. and S. Radovick. 2008. "Pubertal Development and Menarche." *Ann.N.Y.Acad.Sci.* 1135:19-28. doi:1135/1/19 [pii];10.1196/annals.1429.026 [doi].

Dodson, C. 2011. "Knowlede and Attitudes Concerning Pharmacogenomics Among Healthcare Professionals." *Per.Med.* 8(4):421-28.

Dosiou, C., J. Barnes, A. Schwartz, R. Negro, L. Crapo, and A. Stagnaro-Green. 2012. "Cost-Effectiveness of Universal and Risk-Based Screening for Autoimmune Thyroid Disease in Pregnant Women." *J.Clin.Endocrinol.Metab.* 97(5):1536-46. doi:jc.2011-2884 [pii];10.1210/jc.2011-2884 [doi].

Downing, G. J., S. N. Boyle, K. M. Brinner, and J. A. Osheroff. 2009. "Information Management to Enable Personalized Medicine: Stakeholder Roles in Building Clinical Decision Support." *BMC.Med.Inform.Decis.Mak.* 9:44. doi:1472-6947-9-44 [pii];10.1186/1472-6947-9-44 [doi].

Dreyfuss, J. M., D. Levner, J. E. Galagan, G. M. Church, and M. F. Ramoni. 2012. "How Accurate Can Genetic Predictions Be?" *BMC.Genomics.* 13:340. doi:1471-2164-13-340 [pii];10.1186/1471-2164-13-340 [doi].

Dubbs, S. B. and R. Spangler. 2014. "Hypothyroidism: Causes, Killers, and Life-Saving Treatments." *Emerg.Med.Clin.North Am.* 32(2):303-17. doi:S0733-8627(13)00133-8 [pii];10.1016/j.emc.2013.12.003 [doi].

Dudbridge, F. and A. Gusnanto. 2008. "Estimation of Significance Thresholds for Genomewide Association Scans." *Genet.Epidemiol.* 32(3):227-34. doi:10.1002/gepi.20297 [doi].

Dvornyk, V. and Ul Haq Waqar. 1-18-2012. "Genetics of Age at Menarche: a Systematic Review." *Hum.Reprod.Update.* doi:dmr050 [pii];10.1093/humupd/dmr050 [doi].

Eberle, M. A., P. C. Ng, K. Kuhn, L. Zhou, D. A. Peiffer, L. Galver, K. A. Viaud-Martinez, C. T. Lawley, K. L. Gunderson, R. Shen, and S. S. Murray. 2007. "Power to Detect Risk Alleles Using Genome-Wide Tag SNP Panels." *PLoS Genet.* 3(10):1827-37. doi:07-PLGE-RA-0218 [pii];10.1371/journal.pgen.0030170 [doi].

Eckford, P. D., C. Li, M. Ramjeesingh, and C. E. Bear. 10-26-2012. "Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Potentiator VX-770 (Ivacaftor) Opens the Defective Channel Gate of Mutant CFTR in a Phosphorylation-Dependent but ATP-Independent Manner." *J.Biol.Chem.* 287(44):36639-49. doi:M112.393637 [pii];10.1074/jbc.M112.393637 [doi].

Edwards, A. O., R. Ritter, III, K. J. Abel, A. Manning, C. Panhuysen, and L. A. Farrer. 4-15-2005. "Complement Factor H Polymorphism and Age-Related Macular Degeneration." *Science.* 308(5720):421-24. doi:1110189 [pii];10.1126/science.1110189 [doi].

Eeles, R. A., Z. Kote-Jarai, A. A. Al Olama, G. G. Giles, M. Guy, G. Severi, K. Muir, J. L. Hopper, B. E. Henderson, C. A. Haiman, J. Schleutker, F. C. Hamdy, D. E. Neal, J. L. Donovan, J. L. Stanford, E. A. Ostrander, S. A. Ingles, E. M. John, S. N. Thibodeau, D. Schaid, J. Y. Park, A. Spurdle, J. Clements, J. L. Dickinson, C. Maier, W. Vogel, T. Dork, T. R. Rebbeck, K. A. Cooney, L. Cannon-Albright, P. O. Chappuis, P. Hutter, M. Zeegers, R. Kaneva, H. W. Zhang, Y. J. Lu, W. D. Foulkes, D. R. English, D. A. Leongamornlert, M. Tymrakiewicz, J. Morrison, A. T. Ardern-Jones, A. L. Hall, L. T. O'Brien, R. A. Wilkinson, E. J. Saunders, E. C. Page, E. J. Sawyer, S. M. Edwards, D. P. Dearnaley, A. Horwich, R. A. Huddart, V. S. Khoo, C. C. Parker, As N. Van, C. J. Woodhouse, A. Thompson, T. Christmas, C. Ogden, C. S. Cooper, M. C. Southey, A. Lophatananon, J. F. Liu, L. N. Kolonel, Marchand L. Le, T. Wahlfors, T. L. Tammela, A. Auvinen, S. J. Lewis, A. Cox, L. M. FitzGerald, J. S. Koopmeiners, D. M. Karyadi, E. M. Kwon, M. C. Stern, R. Corral, A. D. Joshi, A. Shahabi, S. K. McDonnell, T. A. Sellers, J. Pow-Sang, S. Chambers, J. Aitken, R. A. Gardiner, J. Batra, M. A. Kedda, F. Lose, A. Polanowski, B. Patterson, J. Serth, A. Meyer, M. Luedeke, K. Stefflova, A. M. Ray, E. M. Lange, J. Farnham, H. Khan, C. Slavov, A. Mitkova, G. Cao, and D. F. Easton. 2009. "Identification of Seven New Prostate Cancer Susceptibility Loci Through a Genome-Wide Association Study." *Nat.Genet.* 41(10):1116-21. doi:ng.450 [pii];10.1038/ng.450 [doi].

EGAPP Working Group. 2014. "The EGAPP Initiative: Lessons Learned." *Genet.Med.* 16(3):217-24. doi:gim2013110 [pii];10.1038/gim.2013.110 [doi].

Ege, M. J., D. P. Strachan, W. O. Cookson, M. F. Moffatt, I. Gut, M. Lathrop, M. Kabesch, J. Genuneit, G. Buchele, B. Sozanska, A. Boznanski, P. Cullinan, E. Horak, C. Bieli, C. Braun-Fahrlander, D. Heederik, and Mutius E. von. 2011. "Gene-Environment Interaction for Childhood Asthma and Exposure to Farming in Central Europe." *J.Allergy.Clin.Immunol.* 127(1):138-44, 144. doi:S0091-6749(10)01630-1 [pii];10.1016/j.jaci.2010.09.041 [doi].

Ehret, G. B., P. B. Munroe, K. M. Rice, M. Bochud, A. D. Johnson, D. I. Chasman, A. V. Smith, M. D. Tobin, G. C. Verwoert, S. J. Hwang, V. Pihur, P. Vollenweider, P. F. O'Reilly, N. Amin, J. L. Bragg-Gresham, A. Teumer, N. L. Glazer, L. Launer, J. H. Zhao, Y. Aulchenko, S. Heath, S. Sober, A. Parsa, J. Luan, P. Arora, A. Dehghan, F. Zhang, G. Lucas, A. A. Hicks, A. U. Jackson, J. F. Peden, T. Tanaka, S. H. Wild, I. Rudan, W. Igl, Y. Milaneschi, A. N. Parker, C. Fava, J. C. Chambers, E. R. Fox, M. Kumari, M. J. Go, P. van der Harst, W. H. Kao, M. Sjogren, D. G. Vinay, M. Alexander, Y. Tabara, S. Shaw-Hawkins, P. H. Whincup, Y. Liu, G. Shi, J. Kuusisto, B. Tayo, M. Seielstad, X. Sim, K. D. Nguyen, T. Lehtimaki, G. Matullo, Y. Wu, T. R. Gaunt, N. C. Onland-Moret, M. N.

Cooper, C. G. Platou, E. Org, R. Hardy, S. Dahgam, J. Palmen, V. Vitart, P. S. Braund, T. Kuznetsova, C. S. Uiterwaal, A. Adeyemo, W. Palmas, H. Campbell, B. Ludwig, M. Tomaszewski, I. Tzoulaki, N. D. Palmer, T. Aspelund, M. Garcia, Y. P. Chang, J. R. O'Connell, N. I. Steinle, D. E. Grobbee, D. E. Arking, S. L. Kardia, A. C. Morrison, D. Hernandez, S. Najjar, W. L. McArdle, D. Hadley, M. J. Brown, J. M. Connell, A. D. Hingorani, I. N. Day, D. A. Lawlor, J. P. Beilby, R. W. Lawrence, R. Clarke, J. C. Hopewell, H. Ongen, A. W. Dreisbach, Y. Li, J. H. Young, J. C. Bis, M. Kahonen, J. Viikari, L. S. Adair, N. R. Lee, M. H. Chen, M. Olden, C. Pattaro, J. A. Bolton, A. Kottgen, S. Bergmann, V. Mooser, N. Chaturvedi, T. M. Frayling, M. Islam, T. H. Jafar, J. Erdmann, S. R. Kulkarni, S. R. Bornstein, J. Grassler, L. Groop, B. F. Voight, J. Kettunen, P. Howard, A. Taylor, S. Guarrera, F. Ricceri, V. Emilsson, A. Plump, I. Barroso, K. T. Khaw, A. B. Weder, S. C. Hunt, Y. V. Sun, R. N. Bergman, F. S. Collins, L. L. Bonnycastle, L. J. Scott, H. M. Stringham, L. Peltonen, M. Perola, E. Vartiainen, S. M. Brand, J. A. Staessen, T. J. Wang, P. R. Burton, M. S. Artigas, Y. Dong, H. Snieder, X. Wang, H. Zhu, K. K. Lohman, M. E. Rudock, S. R. Heckbert, N. L. Smith, K. L. Wiggins, A. Doumatey, D. Shriner, G. Veldre, M. Viigimaa, S. Kinra, D. Prabhakaran, V. Tripathy, C. D. Langefeld, A. Rosengren, D. S. Thelle, A. M. Corsi, A. Singleton, T. Forrester, G. Hilton, C. A. McKenzie, T. Salako, N. Iwai, Y. Kita, T. Ogihara, T. Ohkubo, T. Okamura, H. Ueshima, S. Umemura, S. Eyheramendy, T. Meitinger, H. E. Wichmann, Y. S. Cho, H. L. Kim, J. Y. Lee, J. Scott, J. S. Sehmi, W. Zhang, B. Hedblad, P. Nilsson, G. D. Smith, A. Wong, N. Narisu, A. Stancakova, L. J. Raffel, J. Yao, S. Kathiresan, C. J. O'Donnell, S. M. Schwartz, M. A. Ikram, W. T. Longstreth, Jr., T. H. Mosley, S. Seshadri, N. R. Shrine, L. V. Wain, M. A. Morken, A. J. Swift, J. Laitinen, I. Prokopenko, P. Zitting, J. A. Cooper, S. E. Humphries, J. Danesh, A. Rasheed, A. Goel, A. Hamsten, H. Watkins, S. J. Bakker, W. H. van Gilst, C. S. Janipalli, K. R. Mani, C. S. Yajnik, A. Hofman, F. U. Mattace-Raso, B. A. Oostra, A. Demirkan, A. Isaacs, F. Rivadeneira, E. G. Lakatta, M. Orru, A. Scuteri, M. Ala-Korpela, A. J. Kangas, L. P. Lyytikainen, P. Soininen, T. Tukiainen, P. Wurtz, R. T. Ong, M. Dorr, H. K. Kroemer, U. Volker, H. Volzke, P. Galan, S. Hercberg, M. Lathrop, D. Zelenika, P. Deloukas, M. Mangino, T. D. Spector, and G. Zhai. 10-6-2011. "Genetic Variants in Novel Pathways Influence Blood Pressure and Cardiovascular Disease Risk." *Nature* 478(7367):103-9. doi:nature10405 [pii];10.1038/nature10405 [doi].

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore, and J. H. Nadeau. 2010. "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease." *Nat.Rev.Genet.* 11(6):446-50. doi:nrg2809 [pii];10.1038/nrg2809 [doi].

Elks, C. E., J. R. Perry, P. Sulem, D. I. Chasman, N. Franceschini, C. He, K. L. Lunetta, J. A. Visser, E. M. Byrne, D. L. Cousminer, D. F. Gudbjartsson, T. Esko, B. Feenstra, J. J. Hottenga, D. L. Koller, Z. Kutalik, P. Lin, M. Mangino, M. Marongiu, P. F. McArdle, A. V. Smith, L. Stolk, S. H. van Wingerden, J. H. Zhao, E. Albrecht, T. Corre, E. Ingelsson, C. Hayward, P. K. Magnusson, E. N. Smith, S. Ulivi, N. M. Warrington, L. Zgaga, H. Alavere, N. Amin, T. Aspelund, S. Bandinelli, I. Barroso, G. S. Berenson, S. Bergmann, H. Blackburn, E. Boerwinkle, J. E. Buring, F. Busonero, H. Campbell, S. J. Chanock, W. Chen, M. C. Cornelis, D. Couper, A. D. Coviello, P. d'Adamo, Faire U. de, E. J. de Geus, P. Deloukas, A. Doring, G. D. Smith, D. F. Easton, G. Eiriksdottir, V. Emilsson, J. Eriksson, L. Ferrucci, A. R. Folsom, T. Foroud, M. Garcia, P. Gasparini, F. Geller, C. Gieger, V. Gudnason, P. Hall, S. E. Hankinson, L. Ferreli, A. C. Heath, D. G. Hernandez,

A. Hofman, F. B. Hu, T. Illig, M. R. Jarvelin, A. D. Johnson, D. Karasik, K. T. Khaw, D. P. Kiel, T. O. Kilpelainen, I. Kolcic, P. Kraft, L. J. Launer, J. S. Laven, S. Li, J. Liu, D. Levy, N. G. Martin, W. L. McArdle, M. Melbye, V. Mooser, J. C. Murray, S. S. Murray, M. A. Nalls, P. Navarro, M. Nelis, A. R. Ness, K. Northstone, B. A. Oostra, M. Peacock, L. J. Palmer, A. Palotie, G. Pare, A. N. Parker, N. L. Pedersen, L. Peltonen, C. E. Pennell, P. Pharoah, O. Polasek, A. S. Plump, A. Pouta, E. Porcu, T. Rafnar, J. P. Rice, S. M. Ring, F. Rivadeneira, I. Rudan, C. Sala, V. Salomaa, S. Sanna, D. Schlessinger, N. J. Schork, A. Scuteri, A. V. Segre, A. R. Shuldiner, N. Soranzo, U. Sovio, S. R. Srinivasan, D. P. Strachan, M. L. Tammesoo, E. Tikkanen, D. Toniolo, K. Tsui, L. Tryggvadottir, J. Tyrer, M. Uda, R. M. van Dam, J. B. van Meurs, P. Vollenweider, G. Waeber, N. J. Wareham, D. M. Waterworth, M. N. Weedon, H. E. Wichmann, G. Willemsen, J. F. Wilson, A. F. Wright, L. Young, G. Zhai, W. V. Zhuang, L. J. Bierut, D. I. Boomsma, H. A. Boyd, L. Crisponi, E. W. Demerath, C. M. van Duijn, M. J. Econs, T. B. Harris, D. J. Hunter, R. J. Loos, A. Metspalu, G. W. Montgomery, P. M. Ridker, T. D. Spector, E. A. Streeten, K. Stefansson, U. Thorsteinsdottir, A. G. Uitterlinden, E. Widen, J. M. Murabito, K. K. Ong, and A. Murray. 2010. "Thirty New Loci for Age at Menarche Identified by a Meta-Analysis of Genome-Wide Association Studies." *Nat.Genet.* 42(12):1077-85. doi:ng.714 [pii];10.1038/ng.714 [doi].

Elliott, K. S., E. Zeggini, M. I. McCarthy, J. Gudmundsson, P. Sulem, S. N. Stacey, S. Thorlacius, L. Amundadottir, H. Gronberg, J. Xu, V. Gaborieau, R. A. Eeles, D. E. Neal, J. L. Donovan, F. C. Hamdy, K. Muir, S. J. Hwang, M. R. Spitz, B. Zanke, L. Carvajal-Carmona, K. M. Brown, N. K. Hayward, S. Macgregor, I. P. Tomlinson, M. Lemire, C. I. Amos, J. M. Murabito, W. B. Isaacs, D. F. Easton, P. Brennan, R. B. Barkardottir, D. F. Gudbjartsson, T. Rafnar, D. J. Hunter, S. J. Chanock, K. Stefansson, and J. P. Ioannidis. 2010. "Evaluation of Association of HNF1B Variants With Diverse Cancers: Collaborative Analysis of Data From 19 Genome-Wide Association Studies." *PLoS One* 5(5):e10858. doi:10.1371/journal.pone.0010858 [doi].

Emery, J., S. Kumar, and H. Smith. 1998. "Patient Understanding of Genetic Principles and Their Expectations of Genetic Services Within the NHS: a Qualitative Study." *Community Genet.* 1(2):78-83. doi:16141 [doi];16141 [pii].

Enseleit, F., S. Michels, and F. Ruschitzka. 2010. "Anti-VEGF Therapies and Blood Pressure: More Than Meets the Eye." *Curr.Hypertens.Rep.* 12(1):33-38. doi:10.1007/s11906-009-0082-x [doi].

Erickson, J. A., L. Kuzmich, K. E. Ormond, E. Gordon, M. F. Christman, M. K. Cho, and D. F. Levinson. 2014. "Genetic Testing of Children for Predisposition to Mood Disorders: Anticipating the Clinical Issues." *J.Genet.Couns.* 23(4):566-77. doi:10.1007/s10897-014-9710-y [doi].

Eriksson, N., J. Y. Tung, A. K. Kiefer, D. A. Hinds, U. Francke, J. L. Mountain, and C. B. Do. 2012. "Novel Associations for Hypothyroidism Include Known Autoimmune Risk Loci." *PLoS.One.* 7(4):e34442. doi:10.1371/journal.pone.0034442 [doi];PONE-D-11-14352 [pii].

Evans, K. E., M. Hadjivassiliou, and D. S. Sanders. 2011. "Is It Time to Screen for Adult Coeliac Disease?" *Eur.J.Gastroenterol.Hepatol.* 23(10):833-38. doi:10.1097/MEG.0b013e328348f9aa [doi].

Everett, E., H. Tamimi, B. Greer, E. Swisher, P. Paley, L. Mandel, and B. Goff. 2003. "The Effect of Body Mass Index on Clinical/Pathologic Features, Surgical Morbidity, and Outcome in Patients With Endometrial Cancer." *Gynecol.Oncol.* 90(1):150-157. doi:S0090825803002324 [pii].

Fabbri, D. and K. Lefevre. 1-1-2013. "Explaining Accesses to Electronic Medical Records Using Diagnosis Information." *J.Am.Med.Inform.Assoc.* 20(1):52-60. doi:amiajnl-2012-001018 [pii];10.1136/amiajnl-2012-001018 [doi].

Famiglietti, M. L., A. Estreicher, A. Gos, J. Bolleman, S. Gehant, L. Breuza, A. Bridge, S. Poux, N. Redaschi, L. Bougueleret, and I. Xenarios. 5-21-2014. "Genetic Variations and Diseases in UniProtKB/Swiss-Prot: The Ins and Outs of Expert Manual Curation." *Hum.Mutat.* doi:10.1002/humu.22594 [doi].

Feehan, M., J. Hartman, R. Durante, M. A. Morrison, J. W. Miller, I. K. Kim, and M. M. DeAngelis. 2011. "Identifying Subtypes of Patients With Neovascular Age-Related Macular Degeneration by Genotypic and Cardiovascular Risk Characteristics." *BMC.Med.Genet.* 12:83. doi:1471-2350-12-83 [pii];10.1186/1471-2350-12-83 [doi].

Finlayson, A., I. Barnes, S. Sayeed, B. McIver, V. Beral, and R. Ali. 3-4-2014. "Incidence of Thyroid Cancer in England by Ethnic Group, 2001-2007." *Br.J.Cancer* 110(5):1322-27. doi:bjc20144 [pii];10.1038/bjc.2014.4 [doi].

Food and Drug Administration and U.S.Department of Health and Human Services.,*FDA Report Collection, Analysis, and Availability of Demographic Subgroup Data for FDA-Approved Medical Products August 2013* .

Ford, J. G., M. W. Howerton, S. Bolen, T. L. Gary, G. Y. Lai, J. Tilburt, M. C. Gibbons, C. Baffi, R. F. Wilson, C. J. Feuerstein, P. Tanpitukpongse, N. R. Powe, and E. B. Bass. 2005. "Knowledge and Access to Information on Recruitment of Underrepresented Populations to Cancer Clinical Trials." *Evid.Rep.Technol.Assess.(Summ.)*(122):1-11.

Forsyth, J. K., L. M. Ellman, A. Tanskanen, U. Mustonen, M. O. Huttunen, J. Suvisaari, and T. D. Cannon. 2013. "Genetic Risk for Schizophrenia, Obstetric Complications, and Adolescent School Outcome: Evidence for Gene-Environment Interaction." *Schizophr.Bull.* 39(5):1067-76. doi:sbs098 [pii];10.1093/schbul/sbs098 [doi].

Foster, M. W. and R. R. Sharp. 2002. "Race, Ethnicity, and Genomics: Social Classifications As Proxies of Biological Heterogeneity." *Genome.Res.* 12(6):844-50. doi:10.1101/gr.99202 [doi].

Freimuth, V. S., S. C. Quinn, S. B. Thomas, G. Cole, E. Zook, and T. Duncan. 2001. "African Americans' Views on Research and the Tuskegee Syphilis Study." *Soc.Sci.Med.* 52(5):797-808. doi:S0277953600001787 [pii].

French, D. P., E. L. Gayton, J. Burton, M. Thorogood, and T. M. Marteau. 2002. "Measuring Perceptions of Synergistic Circulatory Disease Risk Due to Smoking and the Oral Contraceptive Pill." *Risk Anal.* 22(6):1139-51.

French, D. P., T. M. Marteau, V. Senior, and J. Weinman. 2000. "Perceptions of Multiple Risk Factors for Heart Attacks." *Psychol.Rep.* 87(2):681-87.

Fujimura, J. H., T. Duster, and R. Rajagopalan. 2008. "Race, Genetics, and Disease: Questions of Evidence, Matters of Consequence." *Soc.Stud.Sci.* 38(5):643-56.

Fujimura, J. H. and R. Rajagopalan. 2011. "Different Differences: the Use of 'Genetic Ancestry' Versus Race in Biomedical Human Genetic Research." *Soc.Stud.Sci.* 41(1):5-30.

Fullerton, S. M., W. A. Wolf, K. B. Brothers, E. W. Clayton, D. C. Crawford, J. C. Denny, P. Greenland, B. A. Koenig, K. A. Leppig, N. M. Lindor, C. A. McCarty, A. L. McGuire, E. R. McPeek Hinz, D. B. Mirel, E. M. Ramos, M. D. Ritchie, M. E. Smith, C. J. Waudby, W. Burke, and G. P. Jarvik. 2012. "Return of Individual Research Results From Genome-Wide Association Studies: Experience of the Electronic Medical Records and Genomics (EMERGE) Network." *Genet.Med.* 14(4):424-31. doi:gim201215 [pii];10.1038/gim.2012.15 [doi].

Fusaro, V. A., P. Patil, E. Gafni, D. P. Wall, and P. J. Tonellato. 2011. "Biomedical Cloud Computing With Amazon Web Services." *PLoS Comput.Biol.* 7(8):e1002147. doi:10.1371/journal.pcbi.1002147 [doi];PCOMPBIOL-D-10-00322 [pii].

Gamble, V. N. 1997. "Under the Shadow of Tuskegee: African Americans and Health Care." *Am.J.Public Health* 87(11):1773-78. doi:10.2105/AJPH.87.11.1773 [doi].

Ganann, R., D. Ciliska, and H. Thomas. 2010. "Expediting Systematic Reviews: Methods and Implications of Rapid Reviews." *Implement.Sci.* 5:56. doi:1748-5908-5-56 [pii];10.1186/1748-5908-5-56 [doi].

Garber, A. M. and S. R. Tunis. 5-7-2009. "Does Comparative-Effectiveness Research Threaten Personalized Medicine?" *N.Engl.J.Med.* 360(19):1925-27. doi:360/19/1925 [pii];10.1056/NEJMp0901355 [doi].

Garcia-Closas, M., F. J. Couch, S. Lindstrom, K. Michailidou, M. K. Schmidt, M. N. Brook, N. Orr, S. K. Rhie, E. Riboli, H. S. Feigelson, Marchand L. Le, J. E. Buring, D. Eccles, P. Miron, P. A. Fasching, H. Brauch, J. Chang-Claude, J. Carpenter, A. K. Godwin, H. Nevanlinna, G. G. Giles, A. Cox, J. L. Hopper, M. K. Bolla, Q. Wang, J. Dennis, E. Dicks, W. J. Howat, N. Schoof, S. E. Bojesen, D. Lambrechts, A. Broeks, I. L. Andrulis, P. Guenel, B. Burwinkel, E. J. Sawyer, A. Hollestelle, O. Fletcher, R. Winqvist, H. Brenner, A. Mannermaa, U. Hamann, A. Meindl, A. Lindblom, W. Zheng, P. Devillee, M. S. Goldberg, J. Lubinski, V. Kristensen, A. Swerdlow, H. Anton-Culver, T. Dork, K. Muir, K. Matsuo, A. H. Wu, P. Radice, S. H. Teo, X. O. Shu, W. Blot, D. Kang, M. Hartman, S. Sangrajrang, C. Y. Shen, M. C. Southey, D. J. Park, F. Hammet, J. Stone, L. J. Veer, E. J. Rutgers, A. Lophatananon, S. Stewart-Brown, P. Siriwanarangsan, J. Peto, M. G. Schrauder, A. B. Ekici, M. W. Beckmann, Santos Silva Dos, I, N. Johnson, H. Warren, I.

Tomlinson, M. J. Kerin, N. Miller, F. Marme, A. Schneeweiss, C. Sohn, T. Truong, P. Laurent-Puig, P. Kerbrat, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, R. L. Milne, J. I. Perez, P. Menendez, H. Muller, V. Arndt, C. Stegmaier, P. Lichtner, M. Lochmann, C. Justenhoven, Y. D. Ko, T. A. Muranen, K. Aittomaki, C. Blomqvist, D. Greco, T. Heikkinen, H. Ito, H. Iwata, Y. Yatabe, N. N. Antonenkova, S. Margolin, V. Kataja, V. M. Kosma, J. M. Hartikainen, R. Balleine, C. C. Tseng, D. V. Berg, D. O. Stram, P. Neven, A. S. Dieudonne, K. Leunen, A. Rudolph, S. Nickels, D. Flesch-Janys, P. Peterlongo, B. Peissel, L. Bernard, J. E. Olson, X. Wang, K. Stevens, G. Severi, L. Baglietto, C. McLean, G. A. Coetzee, Y. Feng, B. E. Henderson, F. Schumacher, N. V. Bogdanova, F. Labreche, M. Dumont, C. H. Yip, N. A. Taib, C. Y. Cheng, M. Shrubsole, J. Long, K. Pylkas, A. Jukkola-Vuorinen, S. Kauppila, J. A. Knight, G. Glendon, A. M. Mulligan, R. A. Tollenaar, C. M. Seynaeve, M. Kriege, M. J. Hooning, A. M. van den Ouweland, C. H. van Deurzen, W. Lu, Y. T. Gao, H. Cai, S. P. Balasubramanian, S. S. Cross, M. W. Reed, L. Signorello, Q. Cai, M. Shah, H. Miao, C. W. Chan, K. S. Chia, A. Jakubowska, K. Jaworska, K. Durda, C. N. Hsiung, P. E. Wu, J. C. Yu, A. Ashworth, M. Jones, D. C. Tessier, A. Gonzalez-Neira, G. Pita, M. R. Alonso, D. Vincent, F. Bacot, C. B. Ambrosone, E. V. Bandera, E. M. John, G. K. Chen, J. J. Hu, J. L. Rodriguez-Gil, L. Bernstein, M. F. Press, R. G. Ziegler, R. M. Millikan, S. L. Deming-Halverson, S. Nyante, S. A. Ingles, Q. Waisfisz, H. Tsimiklis, E. Makalic, D. Schmidt, M. Bui, L. Gibson, B. Muller-Myhsok, R. K. Schmutzler, R. Hein, N. Dahmen, L. Beckmann, K. Aaltonen, K. Czene, A. Irwanto, J. Liu, C. Turnbull, N. Rahman, H. Meijers-Heijboer, A. G. Uitterlinden, F. Rivadeneira, C. Olswold, S. Slager, R. Pilarski, F. Ademuyiwa, I. Konstantopoulou, N. G. Martin, G. W. Montgomery, D. J. Slamon, C. Rauh, M. P. Lux, S. M. Jud, T. Bruning, J. Weaver, P. Sharma, H. Pathak, W. Tapper, S. Gerty, L. Durcan, D. Trichopoulos, R. Tumino, P. H. Peeters, R. Kaaks, D. Campa, F. Canzian, E. Weiderpass, M. Johansson, K. T. Khaw, R. Travis, F. Clavel-Chapelon, L. N. Kolonel, C. Chen, A. Beck, S. E. Hankinson, C. D. Berg, R. N. Hoover, J. Lissowska, J. D. Figueroa, D. I. Chasman, and M. M. Gaudet. 2013. "Genome-Wide Association Studies Identify Four ER Negative-Specific Breast Cancer Risk Loci." *Nat.Genet.* 45(4):392. doi:ng.2561 [pii];10.1038/ng.2561 [doi].

Garcia-Galiano, D., L. Pinilla, and M. Tena-Sempere. 2012. "Sex Steroids and the Control of the Kiss1 System: Developmental Roles and Major Regulatory Actions." *J.Neuroendocrinol.* 24(1):22-33. doi:10.1111/j.1365-2826.2011.02230.x [doi].

Garg, K., M. M. Leitao, Jr., C. A. Wynveen, G. L. Sica, J. Shia, W. Shi, and R. A. Soslow. 2010. "P53 Overexpression in Morphologically Ambiguous Endometrial Carcinomas Correlates With Adverse Clinical Outcomes." *Mod.Pathol.* 23(1):80-92. doi:modpathol2009153 [pii];10.1038/modpathol.2009.153 [doi].

Gasche, Y., Y. Daali, M. Fathi, A. Chiappe, S. Cottini, P. Dayer, and J. Desmeules. 12-30-2004. "Codeine Intoxication Associated With Ultrarapid CYP2D6 Metabolism." *N.Engl.J.Med.* 351(27):2827-31. doi:351/27/2827 [pii];10.1056/NEJMoa041888 [doi].

Gauderman, W. J. 3-1-2002a. "Sample Size Requirements for Association Studies of Gene-Gene Interaction." *Am.J.Epidemiol.* 155(5):478-84.

-----. 1-15-2002b. "Sample Size Requirements for Matched Case-Control Studies of Gene-Environment Interaction." *Stat.Med.* 21(1):35-50. doi:10.1002/sim.973 [pii].

Gold, E. B. 2011. "The Timing of the Age at Which Natural Menopause Occurs." *Obstet.Gynecol.Clin.North Am.* 38(3):425-40. doi:S0889-8545(11)00066-0 [pii];10.1016/j.ogc.2011.05.002 [doi].

Gold, E. B., J. Bromberger, S. Crawford, S. Samuels, G. A. Greendale, S. D. Harlow, and J. Skurnick. 5-1-2001. "Factors Associated With Age at Natural Menopause in a Multiethnic Sample of Midlife Women." *Am.J.Epidemiol.* 153(9):865-74.

Goldspiel, B. R., W. A. Flegel, G. DiPatrizio, T. Sissung, S. D. Adams, S. R. Penzak, L. G. Biesecker, T. A. Fleisher, J. J. Patel, D. Herion, W. D. Figg, J. J. Lertora, and J. W. McKeeby. 2014. "Integrating Pharmacogenetic Information and Clinical Decision Support into the Electronic Health Record." *J.Am.Med.Inform.Assoc.* 21(3):522-28. doi:amiajnl-2013-001873 [pii];10.1136/amiajnl-2013-001873 [doi].

Goodman, K. W. 2007. "Clinical Decision Support Systems." Pp. 126-139 in *Ethical and Legal Issues in Decision Support,* edited by E.S.Berner. New York: Springer.

Gordis, Leon. 2009. *Epidemiology*Saunders.

Gorelick, P. B., Y. Harris, B. Burnett, and F. J. Bonecutter. 1998. "The Recruitment Triangle: Reasons Why African Americans Enroll, Refuse to Enroll, or Voluntarily Withdraw From a Clinical Trial. An Interim Report From the African-American Antiplatelet Stroke Prevention Study (AAASPS)." *J.Natl.Med.Assoc.* 90(3):141-45.

Gough, S. C. and M. J. Simmonds. 2007. "The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action." *Curr.Genomics.* 8(7):453-65. doi:10.2174/138920207783591690 [doi].

Graves, K. D., K. G. Leventhal, R. Nusbaum, Y. Salehizadeh, G. W. Hooker, B. N. Peshkin, M. Butrick, W. Tuong, J. Mathew, D. Goerlitz, M. B. Fishman, P. G. Shields, and M. D. Schwartz. 2013. "Behavioral and Psychosocial Responses to Genomic Testing for Colorectal Cancer Risk." *Genomics.* 102(2):123-30. doi:S0888-7543(13)00064-5 [pii];10.1016/j.ygeno.2013.04.002 [doi].

Gray, S. W., K. Hicks-Courant, A. Cronin, B. J. Rollins, and J. C. Weeks. 5-1-2014. "Physicians' Attitudes About Multiplex Tumor Genomic Testing." *J.Clin.Oncol.* 32(13):1317-23. doi:JCO.2013.52.4298 [pii];10.1200/JCO.2013.52.4298 [doi].

Grosse, S. D. and M. J. Khoury. 2006. "What Is the Clinical Utility of Genetic Testing?" *Genet.Med.* 8(7):448-50. doi:10.109701.gim.0000227935.26763.c6 [doi];00125817-200607000-00008 [pii].

Gruber, S. B. and W. D. Thompson. 1996. "A Population-Based Study of Endometrial Cancer and Familial Risk in Younger Women. Cancer and Steroid Hormone Study Group." *Cancer Epidemiol.Biomarkers Prev.* 5(6):411-17.

Gudmundsson, J., P. Sulem, D. F. Gudbjartsson, J. G. Jonasson, G. Masson, H. He, A. Jonasdottir, A. Sigurdsson, S. N. Stacey, H. Johannsdottir, H. T. Helgadottir, W. Li, R.

Nagy, M. D. Ringel, R. T. Kloos, M. C. de Visser, T. S. Plantinga, Heijer M. den, E. Aguillo, A. Panadero, E. Prats, A. Garcia-Castano, Juan A. De, F. Rivera, G. B. Walters, H. Bjarnason, L. Tryggvadottir, G. I. Eyjolfsson, U. S. Bjornsdottir, H. Holm, I. Olafsson, K. Kristjansson, H. Kristvinsson, O. T. Magnusson, G. Thorleifsson, J. R. Gulcher, A. Kong, L. A. Kiemeney, T. Jonsson, H. Hjartarson, J. I. Mayordomo, R. T. Netea-Maier, A. de la Chapelle, J. Hrafnkelsson, U. Thorsteinsdottir, T. Rafnar, and K. Stefansson. 2012. "Discovery of Common Variants Associated With Low TSH Levels and Thyroid Cancer Risk." *Nat.Genet.* 44(3):319-22. doi:ng.1046 [pii];10.1038/ng.1046 [doi].

Gudmundsson, J., P. Sulem, D. F. Gudbjartsson, J. G. Jonasson, A. Sigurdsson, J. T. Bergthorsson, H. He, T. Blondal, F. Geller, M. Jakobsdottir, D. N. Magnusdottir, S. Matthiasdottir, S. N. Stacey, O. B. Skarphedinsson, H. Helgadottir, W. Li, R. Nagy, E. Aguillo, E. Faure, E. Prats, B. Saez, M. Martinez, G. I. Eyjolfsson, U. S. Bjornsdottir, H. Holm, K. Kristjansson, M. L. Frigge, H. Kristvinsson, J. R. Gulcher, T. Jonsson, T. Rafnar, H. Hjartarsson, J. I. Mayordomo, A. de la Chapelle, J. Hrafnkelsson, U. Thorsteinsdottir, A. Kong, and K. Stefansson. 2009. "Common Variants on 9q22.33 and 14q13.3 Predispose to Thyroid Cancer in European Populations." *Nat.Genet.* 41(4):460-464. doi:ng.339 [pii];10.1038/ng.339 [doi].

Guo, Y., H. Shen, P. Xiao, D. H. Xiong, T. L. Yang, Y. F. Guo, J. R. Long, R. R. Recker, and H. W. Deng. 2006a. "Genomewide Linkage Scan for Quantitative Trait Loci Underlying Variation in Age at Menarche." *J.Clin.Endocrinol.Metab* 91(3):1009-14. doi:jc.2005-2179 [pii];10.1210/jc.2005-2179 [doi].

Guo, Y., D. H. Xiong, T. L. Yang, Y. F. Guo, R. R. Recker, and H. W. Deng. 8-15-2006b. "Polymorphisms of Estrogen-Biosynthesis Genes CYP17 and CYP19 May Influence Age at Menarche: a Genetic Association Study in Caucasian Females." *Hum.Mol.Genet.* 15(16):2401-8. doi:ddl155 [pii];10.1093/hmg/ddl155 [doi].

Haga, S. B., W. Burke, G. S. Ginsburg, R. Mills, and R. Agans. 2012. "Primary Care Physicians' Knowledge of and Experience With Pharmacogenetic Testing." *Clin.Genet.* 82(4):388-94. doi:10.1111/j.1399-0004.2012.01908.x [doi].

Haidopoulos, D., M. Simou, N. Akrivos, A. Rodolakis, G. Vlachos, S. Fotiou, M. Sotiropoulou, N. Thomakos, I. Biliatis, A. Protopappas, and A. Antsaklis. 2010. "Risk Factors in Women 40 Years of Age and Younger With Endometrial Carcinoma." *Acta Obstet.Gynecol.Scand.* 89(10):1326-30. doi:10.3109/00016349.2010.515666 [doi].

Haines, J. L., M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins, K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E. A. Postel, and M. A. Pericak-Vance. 4-15-2005. "Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration." *Science.* 308(5720):419-21. doi:1110359 [pii];10.1126/science.1110359 [doi].

Hale, G. E., X. Zhao, C. L. Hughes, H. G. Burger, D. M. Robertson, and I. S. Fraser. 2007. "Endocrine Features of Menstrual Cycles in Middle and Late Reproductive Age and the Menopausal Transition Classified According to the Staging of Reproductive Aging

Workshop (STRAW) Staging System." *J.Clin.Endocrinol.Metab.* 92(8):3060-3067. doi:jc.2007-0066 [pii];10.1210/jc.2007-0066 [doi].

Han, Y., L. B. Signorello, S. S. Strom, R. A. Kittles, B. A. Rybicki, J. L. Stanford, P. J. Goodman, S. I. Berndt, J. Carpten, G. Casey, L. Chu, D. V. Conti, K. A. Rand, W. R. Diver, A. J. Hennis, E. M. John, A. S. Kibel, E. A. Klein, S. Kolb, L. L. Marchand, M. C. Leske, A. B. Murphy, C. Neslund-Dudas, J. Y. Park, C. Pettaway, T. R. Rebbeck, S. M. Gapstur, S. L. Zheng, S. Y. Wu, J. S. Witte, J. Xu, W. Isaacs, S. A. Ingles, A. Hsing, D. F. Easton, R. A. Eeles, F. R. Schumacher, S. Chanock, B. Nemesure, W. J. Blot, D. O. Stram, B. E. Henderson, and C. A. Haiman. 7-8-2014. "Generalizability of Established Prostate Cancer Risk Variants in Men of African Ancestry." *Int.J.Cancer.* doi:10.1002/ijc.29066 [doi].

Hansen, P. S., T. H. Brix, T. I. Sorensen, K. O. Kyvik, and L. Hegedus. 2004. "Major Genetic Influence on the Regulation of the Pituitary-Thyroid Axis: a Study of Healthy Danish Twins." *J.Clin.Endocrinol.Metab.* 89(3):1181-87. doi:10.1210/jc.2003-031641 [doi].

Harlow, S. D., M. Gass, J. E. Hall, R. Lobo, P. Maki, R. W. Rebar, S. Sherman, P. M. Sluss, and T. J. de Villiers. 2012. "Executive Summary of the Stages of Reproductive Aging Workshop + 10: Addressing the Unfinished Agenda of Staging Reproductive Aging." *J.Clin.Endocrinol.Metab.* 97(4):1159-68. doi:jc.2011-3362 [pii];10.1210/jc.2011-3362 [doi].

Harman, S. M. 2014. "Menopausal Hormone Treatment Cardiovascular Disease: Another Look at an Unresolved Conundrum." *Fertil.Steril.* 101(4):887-97. doi:S0015-0282(14)00195-2 [pii];10.1016/j.fertnstert.2014.02.042 [doi].

Harris, P. A., R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde. 2009. "Research Electronic Data Capture (REDCap)--a Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support." *J.Biomed.Inform.* 42(2):377-81. doi:S1532-0464(08)00122-6 [pii];10.1016/j.jbi.2008.08.010 [doi].

Hartge, P. 2009. "Genetics of Reproductive Lifespan." *Nat.Genet.* 41(6):637-38. doi:ng0609-637 [pii];10.1038/ng0609-637 [doi].

Hartzler, A., C. A. McCarty, L. V. Rasmussen, M. S. Williams, M. Brilliant, E. A. Bowton, E. W. Clayton, W. A. Faucett, K. Ferryman, J. R. Field, S. M. Fullerton, C. R. Horowitz, B. A. Koenig, J. B. McCormick, J. D. Ralston, S. C. Sanderson, M. E. Smith, and S. B. Trinidad. 2013. "Stakeholder Engagement: a Key Component of Integrating Genomic Information into Electronic Health Records." *Genet.Med.* 15(10):792-801. doi:gim2013127 [pii];10.1038/gim.2013.127 [doi].

Hayes, M. G., M. Urbanek, M. F. Hivert, L. L. Armstrong, J. Morrison, C. Guo, L. P. Lowe, D. A. Scheftner, A. Pluzhnikov, D. M. Levine, C. P. McHugh, C. M. Ackerman, L. Bouchard, D. Brisson, B. T. Layden, D. Mirel, K. F. Doheny, M. V. Leya, R. N. Lown-Hecht, A. R. Dyer, B. E. Metzger, T. E. Reddy, N. J. Cox, and W. L. Lowe, Jr. 2013. "Identification of HKDC1 and BACE2 As Genes Influencing Glycemic Traits During Pregnancy Through Genome-Wide Association Studies." *Diabetes* 62(9):3282-91. doi:db12-1692 [pii];10.2337/db12-1692 [doi].

Hazin, R., K. B. Brothers, B. A. Malin, B. A. Koenig, S. C. Sanderson, M. A. Rothstein, M. S. Williams, E. W. Clayton, and I. J. Kullo. 2013. "Ethical, Legal, and Social Implications of Incorporating Genomic Information into Electronic Health Records." *Genet.Med.* 15(10):810-816. doi:gim2013117 [pii];10.1038/gim.2013.117 [doi].

He, C., P. Kraft, D. I. Chasman, J. E. Buring, C. Chen, S. E. Hankinson, G. Pare, S. Chanock, P. M. Ridker, and D. J. Hunter. 2010. "A Large-Scale Candidate Gene Association Study of Age at Menarche and Age at Natural Menopause." *Hum.Genet.* 128(5):515-27. doi:10.1007/s00439-010-0878-4 [doi].

He, C., P. Kraft, C. Chen, J. E. Buring, G. Pare, S. E. Hankinson, S. J. Chanock, P. M. Ridker, D. J. Hunter, and D. I. Chasman. 2009a. "Genome-Wide Association Studies Identify Loci Associated With Age at Menarche and Age at Natural Menopause." *Nat.Genet.* 41(6):724-28. doi:ng.385 [pii];10.1038/ng.385 [doi].

He, C. and J. M. Murabito. Forthcoming. "Genome-Wide Association Studies of Age at Menarche and Age at Natural Menopause." *Molecular and Cellular Endocrinology*.

He, L. N., R. R. Recker, H. W. Deng, and V. Dvornyk. 1-20-2009b. "A Polymorphism of Apolipoprotein E (APOE) Gene Is Associated With Age at Natural Menopause in Caucasian Females." *Maturitas* 62(1):37-41. doi:S0378-5122(08)00266-1 [pii];10.1016/j.maturitas.2008.10.011 [doi].

He, L. N., D. H. Xiong, Y. J. Liu, F. Zhang, R. R. Recker, and H. W. Deng. 2007. "Association Study of the Oestrogen Signalling Pathway Genes in Relation to Age at Natural Menopause." *J.Genet.* 86(3):269-76.

Henriet, P., H. P. Gaide Chevronnay, and E. Marbaix. 7-25-2012. "The Endocrine and Paracrine Control of Menstruation." *Mol.Cell Endocrinol.* 358(2):197-207. doi:S0303-7207(11)00455-2 [pii];10.1016/j.mce.2011.07.042 [doi].

Herman-Giddens, M. E. 2006. "Recent Data on Pubertal Milestones in United States Children: the Secular Trend Toward Earlier Development." *Int.J.Androl* 29(1):241-46. doi:IJA575 [pii];10.1111/j.1365-2605.2005.00575.x [doi].

Hofbauer, L. C. and M. Schoppet. 7-28-2004. "Clinical Implications of the Osteoprotegerin/RANKL/RANK System for Bone and Vascular Diseases." *JAMA* 292(4):490-495. doi:10.1001/jama.292.4.490 [doi];292/4/490 [pii].

Hoffman, J. M., C. E. Haidar, M. R. Wilkinson, K. R. Crews, D. K. Baker, N. M. Kornegay, W. Yang, C. H. Pui, U. M. Reiss, A. H. Gaur, S. C. Howard, W. E. Evans, U. Broeckel, and M. V. Relling. 2014. "PG4KDS: a Model for the Clinical Implementation of Pre-Emptive Pharmacogenetics." *Am.J.Med.Genet.C.Semin.Med.Genet.* 166C(1):45-55. doi:10.1002/ajmg.c.31391 [doi].

Hoffmann, T. J., Y. Zhan, M. N. Kvale, S. E. Hesselson, J. Gollub, C. Iribarren, Y. Lu, G. Mei, M. M. Purdy, C. Quesenberry, S. Rowell, M. H. Shapero, D. Smethurst, C. P. Somkin, S. K. Van Den Eeden, L. Walter, T. Webster, R. A. Whitmer, A. Finn, C. Schaefer, P. Y. Kwok,

and N. Risch. 2011. "Design and Coverage of High Throughput Genotyping Arrays Optimized for Individuals of East Asian, African American, and Latino Race/Ethnicity Using Imputation and a Novel Hybrid SNP Selection Algorithm." *Genomics.* 98(6):422-30. doi:S0888-7543(11)00196-0 [pii];10.1016/j.ygeno.2011.08.007 [doi].

Hognas, E., A. Kauppila, E. Pukkala, and J. S. Tapanainen. 2014. "Cancer Risk in Women With 10 or More Deliveries." *Obstet.Gynecol.* 123(4):811-16. doi:10.1097/AOG.0000000000000182 [doi];00006250-201404000-00013 [pii].

Hollowell, J. G., N. W. Staehling, W. D. Flanders, W. H. Hannon, E. W. Gunter, C. A. Spencer, and L. E. Braverman. 2002. "Serum TSH, T(4), and Thyroid Antibodies in the United States Population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III)." *J.Clin.Endocrinol.Metab.* 87(2):489-99.

Horvath, A., F. Faucz, G. P. Finkielstain, M. E. Nikita, A. Rothenbuhler, M. Almeida, V. Mericq, and C. A. Stratakis. 2010. "Haplotype Analysis of the Promoter Region of Phosphodiesterase Type 8B (PDE8B) in Correlation With Inactivating PDE8B Mutation and the Serum Thyroid-Stimulating Hormone Levels." *Thyroid* 20(4):363-67. doi:10.1089/thy.2009.0260 [doi].

Hughes, M. F., O. Saarela, J. Stritzke, F. Kee, K. Silander, N. Klopp, J. Kontto, J. Karvanen, C. Willenborg, V. Salomaa, J. Virtamo, P. Amouyel, D. Arveiler, J. Ferrieres, P. G. Wiklund, J. Baumert, B. Thorand, P. Diemert, D. A. Tregouet, C. Hengstenberg, A. Peters, A. Evans, W. Koenig, J. Erdmann, N. J. Samani, K. Kuulasmaa, and H. Schunkert. 2012. "Genetic Markers Enhance Coronary Risk Prediction in Men: the MORGAM Prospective Cohorts." *PLoS One* 7(7):e40922. doi:10.1371/journal.pone.0040922 [doi];PONE-D-11-08927 [pii].

Hui, J., Y. Xu, K. Yang, M. Liu, D. Wei, D. Wei, Y. Zhang, X. H. Shi, F. Yang, N. Wang, Y. Zhang, X. Wang, S. Liang, X. Chen, L. Sun, X. Zhu, L. Zhu, Y. Yang, L. Tang, Y. Zhang, Z. Yang, and J. Wang. 2014. "Study of Genetic Variants of 8q21 and 8q24 Associated With Prostate Cancer in Jing-Jin Residents in Northern China." *Clin.Lab.* 60(4):645-52.

Hunt, L. M., N. D. Truesdell, and M. J. Kreiner. 2013. "Genes, Race, and Culture in Clinical Care: Racial Profiling in the Management of Chronic Illness." *Med.Anthropol.Q.* 27(2):253-71. doi:10.1111/maq.12026 [doi].

Hurle, B., T. Citrin, J. F. Jenkins, K. A. Kaphingst, N. Lamb, J. E. Roseman, and V. L. Bonham. 2013. "What Does It Mean to Be Genomically Literate?: National Human Genome Research Institute Meeting Report." *Genet.Med.* 15(8):658-63. doi:gim201314 [pii];10.1038/gim.2013.14 [doi].

Hyder, A. A. and R. H. Morrow. 2006. "International Public Health Diseases, Programs, Systems, and Policies." Pp. 12-20 in *Measures of Health and Disease in Populations,* edited by M.H.Merson, R.E.Black, and A.J.Mills. Sudbury: Jones and Bartlett.

Irvine, K., K. R. Laws, T. M. Gale, and T. K. Kondel. 2012. "Greater Cognitive Deterioration in Women Than Men With Alzheimer's Disease: a Meta Analysis." *J.Clin.Exp.Neuropsychol.* 34(9):989-98. doi:10.1080/13803395.2012.712676 [doi].

Isler, M. R., K. Sutton, R. J. Cadigan, and G. Corbie-Smith. 2013. "Community Perceptions of Genomic Research: Implications for Addressing Health Disparities." *N.C.Med.J.* 74(6):470-476. doi:74602 [pii].

Ito, H., K. Matsuo, K. Wakai, T. Saito, H. Kumimoto, K. Okuma, K. Tajima, and N. Hamajima. 2006. "An Intervention Study of Smoking Cessation With Feedback on Genetic Cancer Susceptibility in Japan." *Prev.Med.* 42(2):102-8. doi:S0091-7435(05)00167-2 [pii];10.1016/j.ypmed.2005.10.006 [doi].

James, M. A., W. Wen, Y. Wang, L. A. Byers, J. V. Heymach, K. R. Coombes, L. Girard, J. Minna, and M. You. 2012. "Functional Characterization of CLPTM1L As a Lung Cancer Risk Candidate Gene in the 5p15.33 Locus." *PLoS One* 7(6):e36116. doi:10.1371/journal.pone.0036116 [doi];PONE-D-11-16033 [pii].

Janssens, A. C., Y. S. Aulchenko, S. Elefante, G. J. Borsboom, E. W. Steyerberg, and C. M. van Duijn. 2006. "Predictive Testing for Complex Diseases Using Multiple Genes: Fact or Fiction?" *Genet.Med.* 8(7):395-400. doi:10.109701.gim.0000229689.18263.f4 [doi];00125817-200607000-00001 [pii].

Jarde, T., F. Caldefie-Chezet, N. Goncalves-Mendes, F. Mishellany, C. Buechler, F. Penault-Llorca, and M. P. Vasson. 2009. "Involvement of Adiponectin and Leptin in Breast Cancer: Clinical and in Vitro Studies." *Endocr.Relat.Cancer* 16(4):1197-210. doi:ERC-09-0043 [pii];10.1677/ERC-09-0043 [doi].

Jasen, P. 2011. "Menopause and Historical Constructions of Cancer Risk." *Can.Bull.Med.Hist* 28(1):43-70.

Jemal, A., F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. 2011. "Global Cancer Statistics." *CA Cancer J.Clin.* 61(2):69-90. doi:caac.20107 [pii];10.3322/caac.20107 [doi].

Jha, A. K., C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, A. Shields, S. Rosenbaum, and D. Blumenthal. 4-16-2009. "Use of Electronic Health Records in U.S. Hospitals." *N.Engl.J.Med.* 360(16):1628-38. doi:NEJMsa0900592 [pii];10.1056/NEJMsa0900592 [doi].

Jia, J., A. D. Bosley, A. Thompson, J. W. Hoskins, A. Cheuk, I. Collins, H. Parikh, Z. Xiao, K. Ylaya, M. Dzyadyk, W. Cozen, B. Y. Hernandez, C. F. Lynch, J. Loncarek, S. F. Altekruse, L. Zhang, C. J. Westlake, V. M. Factor, S. Thorgeirsson, W. R. Bamlet, S. M. Hewitt, G. M. Petersen, T. Andresson, and L. T. Amundadottir. 5-15-2014. "CLPTM1L Promotes Growth and Enhances Aneuploidy in Pancreatic Cancer Cells." *Cancer Res.* 74(10):2785-95. doi:0008-5472.CAN-13-3176 [pii];10.1158/0008-5472.CAN-13-3176 [doi].

Jick, H., A. M. Walker, and K. J. Rothman. 1980. "The Epidemic of Endometrial Cancer: a Commentary." *Am.J.Public.Health* 70(3):264-67.

Jick, S. S., A. M. Walker, and H. Jick. 1993a. "Estrogens, Progesterone, and Endometrial Cancer." *Epidemiology.* 4(1):20-24.

-----. 1993b. "Oral Contraceptives and Endometrial Cancer." *Obstet.Gynecol.* 82(6):931-35.

Johnson, A. D., R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, and P. I. de Bakker. 12-15-2008. "SNAP: a Web-Based Tool for Identification and Annotation of Proxy SNPs Using HapMap." *Bioinformatics.* 24(24):2938-39. doi:btn564 [pii];10.1093/bioinformatics/btn564 [doi].

Johnson, K. A., J. D. Trimbath, G. M. Petersen, C. A. Griffin, and F. M. Giardiello. 2002. "Impact of Genetic Counseling and Testing on Colorectal Cancer Screening Behavior." *Genet.Test.* 6(4):303-6. doi:10.1089/10906570260471831 [doi].

Jorde, R. and J. Sundsfjord. 2006. "Serum TSH Levels in Smokers and Non-Smokers. The 5th Tromso Study." *Exp.Clin.Endocrinol.Diabetes* 114(7):343-47. doi:10.1055/s-2006-924264 [doi].

Juengst, E. T. 2014. "TMI! Ethical Challenges in Managing and Using Large Patient Data Sets." *N.C.Med.J.* 75(3):214-17. doi:75314 [pii].

Jurisicova, A., A. Taniuchi, H. Li, Y. Shang, M. Antenos, J. Detmar, J. Xu, T. Matikainen, Hernandez A. Benito, G. Nunez, and R. F. Casper. 2007. "Maternal Exposure to Polycyclic Aromatic Hydrocarbons Diminishes Murine Ovarian Reserve Via Induction of Harakiri." *J.Clin.Invest.* 117(12):3971-78. doi:10.1172/JCI28493 [doi].

Jurvansuu, J. M. and A. Goldman. 2011. "Obesity Risk Gene TMEM18 Encodes a Sequence-Specific DNA-Binding Protein." *PLoS.One.* 6(9):e25317. doi:10.1371/journal.pone.0025317 [doi];PONE-D-11-12025 [pii].

Kallen, A. N. and L. Pal. 2011. "Cardiovascular Disease and Ovarian Function." *Curr.Opin.Obstet.Gynecol.* 23(4):258-67. doi:10.1097/GCO.0b013e3283488a21 [doi].

Kalogiannidis, I., S. Petousis, M. Bobos, C. Margioula-Siarkou, M. Topalidou, A. Papanikolaou, I. Vergote, and T. Agorastos. 7-15-2014. "HER-2/Neu Is an Independent Prognostic Factor in Type I Endometrial Adenocarcinoma." *Arch.Gynecol.Obstet.* doi:10.1007/s00404-014-3333-2 [doi].

Kaplowitz, P. 2006. "Pubertal Development in Girls: Secular Trends." *Curr.Opin.Obstet.Gynecol.* 18(5):487-91. doi:10.1097/01.gco.0000242949.02373.09 [doi];00001703-200610000-00003 [pii].

Katanoda, K. and D. Qiu. 2006. "International Comparisons of Cumulative Risk of Uterine Cancer, From Cancer Incidence in Five Continents Vol. VIII." *Jpn.J.Clin.Oncol.* 36(7):474-75. doi:36/7/474 [pii];10.1093/jjco/hyl070 [doi].

Keller, Johannes. 2005. "In Genes We Trust: the Biological Component of Psychological Essentialism and Its Relationship to Mechanisms of Motivated Social Cognition." *Journal of Personality and Social Psychology* 88(4):686.

Kho, A. N., L. V. Rasmussen, J. J. Connolly, P. L. Peissig, J. Starren, H. Hakonarson, and M. G. Hayes. 2013. "Practical Challenges in Integrating Genomic Data into the Electronic Health Record." *Genet.Med.* 15(10):772-78. doi:gim2013131 [pii];10.1038/gim.2013.131 [doi].

Kim, H. S., S. Mendiratta, J. Kim, C. V. Pecot, J. E. Larsen, I. Zubovych, B. Y. Seo, J. Kim, B. Eskiocak, H. Chung, E. McMillan, S. Wu, Brabander J. De, K. Komurov, J. E. Toombs, S. Wei, M. Peyton, N. Williams, A. F. Gazdar, B. A. Posner, R. A. Brekken, A. K. Sood, R. J. Deberardinis, M. G. Roth, J. D. Minna, and M. A. White. 10-24-2013. "Systematic Identification of Molecular Subtype-Selective Vulnerabilities in Non-Small-Cell Lung Cancer." *Cell* 155(3):552-66. doi:S0092-8674(13)01215-4 [pii];10.1016/j.cell.2013.09.041 [doi].

Kim, N. W., M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich, and J. W. Shay. 12-23-1994. "Specific Association of Human Telomerase Activity With Immortal Cells and Cancer." *Science.* 266(5193):2011-15.

Kitchener, H. C. and E. L. Trimble. 2009. "Endometrial Cancer State of the Science Meeting." *Int.J.Gynecol.Cancer* 19(1):134-40. doi:10.1111/IGC.0b013e3181995f90 [doi];00009577-200901000-00024 [pii].

Klein, R., C. F. Chou, B. E. Klein, X. Zhang, S. M. Meuer, and J. B. Saaddine. 2011. "Prevalence of Age-Related Macular Degeneration in the US Population." *Arch.Ophthalmol.* 129(1):75-80. doi:129/1/75 [pii];10.1001/archophthalmol.2010.318 [doi].

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh. 4-15-2005. "Complement Factor H Polymorphism in Age-Related Macular Degeneration." *Science.* 308(5720):385-89. doi:1109557 [pii];10.1126/science.1109557 [doi].

Kleinman, A. and P. Benson. 2006. "Anthropology in the Clinic: the Problem of Cultural Competency and How to Fix It." *PLoS Med.* 3(10):e294. doi:06-PLME-ES-0159R1 [pii];10.1371/journal.pmed.0030294 [doi].

Knowler, W. C., E. Barrett-Connor, S. E. Fowler, R. F. Hamman, J. M. Lachin, E. A. Walker, and D. M. Nathan. 2-7-2002. "Reduction in the Incidence of Type 2 Diabetes With Lifestyle Intervention or Metformin." *N.Engl.J.Med.* 346(6):393-403. doi:10.1056/NEJMoa012512 [doi];346/6/393 [pii].

Knudsen, N., P. Laurberg, L. B. Rasmussen, I. Bulow, H. Perrild, L. Ovesen, and T. Jorgensen. 2005. "Small Differences in Thyroid Function May Be Important for Body Mass Index and the Occurrence of Obesity in the Population." *J.Clin.Endocrinol.Metab.* 90(7):4019-24. doi:jc.2004-2225 [pii];10.1210/jc.2004-2225 [doi].

Kohane, I. S. 2011. "Using Electronic Health Records to Drive Discovery in Disease Genomics." *Nat.Rev.Genet.* 12(6):417-28. doi:nrg2999 [pii];10.1038/nrg2999 [doi].

Korner, A., K. Pazaitou-Panayiotou, T. Kelesidis, I. Kelesidis, C. J. Williams, A. Kaprara, J. Bullen, A. Neuwirth, S. Tseleni, N. Mitsiades, W. Kiess, and C. S. Mantzoros. 2007. "Total and High-Molecular-Weight Adiponectin in Breast Cancer: in Vitro and in Vivo Studies." *J.Clin.Endocrinol.Metab.* 92(3):1041-48. doi:jc.2006-1858 [pii];10.1210/jc.2006-1858 [doi].

Kote-Jarai, Z., A. A. Olama, G. G. Giles, G. Severi, J. Schleutker, M. Weischer, D. Campa, E. Riboli, T. Key, H. Gronberg, D. J. Hunter, P. Kraft, M. J. Thun, S. Ingles, S. Chanock, D. Albanes, R. B. Hayes, D. E. Neal, F. C. Hamdy, J. L. Donovan, P. Pharoah, F. Schumacher, B. E. Henderson, J. L. Stanford, E. A. Ostrander, K. D. Sorensen, T. Dork, G. Andriole, J. L. Dickinson, C. Cybulski, J. Lubinski, A. Spurdle, J. A. Clements, S. Chambers, J. Aitken, R. A. Gardiner, S. N. Thibodeau, D. Schaid, E. M. John, C. Maier, W. Vogel, K. A. Cooney, J. Y. Park, L. Cannon-Albright, H. Brenner, T. Habuchi, H. W. Zhang, Y. J. Lu, R. Kaneva, K. Muir, S. Benlloch, D. A. Leongamornlert, E. J. Saunders, M. Tymrakiewicz, N. Mahmud, M. Guy, L. T. O'Brien, R. A. Wilkinson, A. L. Hall, E. J. Sawyer, T. Dadaev, J. Morrison, D. P. Dearnaley, A. Horwich, R. A. Huddart, V. S. Khoo, C. C. Parker, As N. Van, C. J. Woodhouse, A. Thompson, T. Christmas, C. Ogden, C. S. Cooper, A. Lophatonanon, M. C. Southey, J. L. Hopper, D. R. English, T. Wahlfors, T. L. Tammela, P. Klarskov, B. G. Nordestgaard, M. A. Roder, A. Tybjaerg-Hansen, S. E. Bojesen, R. Travis, F. Canzian, R. Kaaks, F. Wiklund, M. Aly, S. Lindstrom, W. R. Diver, S. Gapstur, M. C. Stern, R. Corral, J. Virtamo, A. Cox, C. A. Haiman, Marchand L. Le, L. Fitzgerald, S. Kolb, E. M. Kwon, D. M. Karyadi, T. F. Orntoft, M. Borre, A. Meyer, J. Serth, M. Yeager, S. I. Berndt, J. R. Marthick, B. Patterson, D. Wokolorczyk, J. Batra, F. Lose, S. K. McDonnell, A. D. Joshi, A. Shahabi, A. E. Rinckleb, A. Ray, T. A. Sellers, H. Y. Lin, R. A. Stephenson, J. Farnham, H. Muller, D. Rothenbacher, N. Tsuchiya, S. Narita, G. W. Cao, C. Slavov, V. Mitev, D. F. Easton, and R. A. Eeles. 2011. "Seven Prostate Cancer Susceptibility Loci Identified by a Multi-Stage Genome-Wide Association Study." *Nat.Genet.* 43(8):785-91. doi:ng.882 [pii];10.1038/ng.882 [doi].

Kuhn, E., R. C. Wu, B. Guan, G. Wu, J. Zhang, Y. Wang, L. Song, X. Yuan, L. Wei, R. B. Roden, K. T. Kuo, K. Nakayama, B. Clarke, P. Shaw, N. Olvera, R. J. Kurman, D. A. Levine, T. L. Wang, and IeM Shih. 10-3-2012. "Identification of Molecular Pathway Aberrations in Uterine Serous Carcinoma by Genome-Wide Analyses." *J.Natl.Cancer Inst.* 104(19):1503-13. doi:djs345 [pii];10.1093/jnci/djs345 [doi].

Kukafka, Rita, Jessica S. Ancker, Connie Chan, John Chelico, Sharib Khan, Selasie Mortoti, Karthik Natarajan, Kempton Presley, and Kayann Stephens. 2007. "Redesigning Electronic Health Record Systems to Support Public Health." *Journal of Biomedical Informatics* 40(4):398-409.

Kuo, A. M. 2011. "Opportunities and Challenges of Cloud Computing to Improve Health Care Services." *J.Med.Internet.Res.* 13(3):e67. doi:v13i3e67 [pii];10.2196/jmir.1867 [doi].

Kutner, M., E. Greenberg, Y. Jin, and C. Paulsen.,*The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy* .

Laerum, H., S. Bremer, S. Bergan, and T. Grunfeld. 2014. "A Taste of Individualized Medicine: Physicians' Reactions to Automated Genetic Interpretations." *J.Am.Med.Inform.Assoc.* 21(e1):e143-e146. doi:amiajnl-2012-001587 [pii];10.1136/amiajnl-2012-001587 [doi].

Landa, I., S. Ruiz-Llorente, C. Montero-Conde, L. Inglada-Perez, F. Schiavi, S. Leskela, G. Pita, R. Milne, J. Maravall, I. Ramos, V. Andia, P. Rodriguez-Poyo, A. Jara-Albarran, A. Meoro, Peso C. del, L. Arribas, P. Iglesias, J. Caballero, J. Serrano, A. Pico, F. Pomares, G. Gimenez, P. Lopez-Mondejar, R. Castello, I. Merante-Boschin, M. R. Pelizzo, D. Mauricio, G. Opocher, C. Rodriguez-Antona, A. Gonzalez-Neira, X. Matias-Guiu, P. Santisteban, and M. Robledo. 2009. "The Variant Rs1867277 in FOXE1 Gene Confers Thyroid Cancer Susceptibility Through the Recruitment of USF1/USF2 Transcription Factors." *PLoS Genet.* 5(9):e1000637. doi:10.1371/journal.pgen.1000637 [doi].

Lanie, A. D., T. E. Jayaratne, J. P. Sheldon, S. L. Kardia, E. S. Anderson, M. Feldbaum, and E. M. Petty. 2004. "Exploring the Public Understanding of Basic Genetic Concepts." *J.Genet.Couns.* 13(4):305-20.

Laurberg, P., S. Andersen, Pedersen Bulow, I, and A. Carle. 2005. "Hypothyroidism in the Elderly: Pathophysiology, Diagnosis and Treatment." *Drugs.Aging.* 22(1):23-38. doi:2212 [pii].

Lawrence, L., M. Sincan, T. Markello, D. R. Adams, F. Gill, R. Godfrey, G. Golas, C. Groden, D. Landis, M. Nehrebecky, G. Park, A. Soldatos, C. Tifft, C. Toro, C. Wahl, L. Wolfe, W. A. Gahl, and C. F. Boerkoel. 5-1-2014. "The Implications of Familial Incidental Findings From Exome Sequencing: the NIH Undiagnosed Diseases Program Experience." *Genet.Med.* doi:gim201429 [pii];10.1038/gim.2014.29 [doi].

Lawrence, R. C., C. G. Helmick, F. C. Arnett, R. A. Deyo, D. T. Felson, E. H. Giannini, S. P. Heyse, R. Hirsch, M. C. Hochberg, G. G. Hunder, M. H. Liang, S. R. Pillemer, V. D. Steen, and F. Wolfe. 1998. "Estimates of the Prevalence of Arthritis and Selected Musculoskeletal Disorders in the United States." *Arthritis.Rheum.* 41(5):778-99. doi:10.1002/1529-0131(199805)41:5<778::AID-ART4>3.0.CO;2-V [doi].

Lea, D. H., K. A. Kaphingst, D. Bowen, I. Lipkus, and D. W. Hadley. 2011. "Communicating Genetic and Genomic Information: Health Literacy and Numeracy Considerations." *Public.Health Genomics.* 14(4-5):279-89. doi:000294191 [pii];10.1159/000294191 [doi].

Lee, E., C. Hsu, C. A. Haiman, P. Razavi, P. L. Horn-Ross, D. Van Den Berg, L. Bernstein, Marchand L. Le, B. E. Henderson, V. W. Setiawan, and G. Ursin. 2010. "Genetic Variation in the Progesterone Receptor Gene and Risk of Endometrial Cancer: a Haplotype-Based Approach." *Carcinogenesis* 31(8):1392-99. doi:bgq113 [pii];10.1093/carcin/bgq113 [doi].

Lerman, C., K. Gold, J. Audrain, T. H. Lin, N. R. Boyd, C. T. Orleans, B. Wilfond, G. Louben, and N. Caporaso. 1997. "Incorporating Biomarkers of Exposure and Genetic Susceptibility into Smoking Cessation Treatment: Effects on Smoking-Related Cognitions, Emotions, and Behavior Change." *Health.Psychol.* 16(1):87-99.

Leventhal, H., Y. Benyamini, S. Brownlee, M. Diefenbach, E. A. Leventhal, L. Patrick-Miller, and C. Robitaille. 1997. "Perceptions of Health and Illness." Pp. 19-45 in *Illness Representations: Theoretical Foundations,* edited by J.Weinman and K.Petrie. London: Harwood Publishers.

Lieberthal, R. D. 2013. "Economics of Genomic Testing for Women With Breast Cancer." *Am.J.Manag.Care* 19(12):1024-31. doi:85331 [pii].

Lin, H. Y., W. Wang, Y. H Lui, S. J. Soong, T. P. York, L. Myers, and J. J. Hu. 2008. "Comparison of Multivariate Adaptive Regression Splines and Logistic Regression in Detecting SNP-SNP Interactions and Their Application in Prostate Cancer." *J.Hum.Genet.* 53:802-11.

Lin, W. Y. and W. C. Lee. 2012. "Improving Power of Genome-Wide Association Studies With Weighted False Discovery Rate Control and Prioritized Subset Analysis." *PLoS One* 7(4):e33716. doi:10.1371/journal.pone.0033716 [doi];PONE-D-11-18994 [pii].

Liu, F., A. W. Hsing, X. Wang, Q. Shao, J. Qi, Y. Ye, Z. Wang, H. Chen, X. Gao, G. Wang, L. W. Chu, Q. Ding, J. OuYang, X. Gao, Y. Huang, Y. Chen, Y. T. Gao, Z. F. Zhang, J. Rao, R. Shi, Q. Wu, M. Wang, Z. Zhang, Y. Zhang, H. Jiang, J. Zheng, Y. Hu, L. Guo, X. Lin, S. Tao, G. Jin, J. Sun, D. Lu, S. L. Zheng, Y. Sun, Z. Mo, and J. Xu. 2011. "Systematic Confirmation Study of Reported Prostate Cancer Risk-Associated Single Nucleotide Polymorphisms in Chinese Men." *Cancer Sci.* 102(10):1916-20. doi:10.1111/j.1349-7006.2011.02036.x [doi].

Liu, W., C. C. Xie, Y. Zhu, T. Li, J. Sun, Y. Cheng, C. M. Ewing, S. Dalrymple, A. R. Turner, J. Sun, J. T. Isaacs, B. L. Chang, S. L. Zheng, W. B. Isaacs, and J. Xu. 2008. "Homozygous Deletions and Recurrent Amplifications Implicate New Genes Involved in Prostate Cancer." *Neoplasia.* 10(8):897-907.

Long, J., W. Zheng, Y. B. Xiang, F. Lose, D. Thompson, I. Tomlinson, H. Yu, N. Wentzensen, D. Lambrechts, T. Dork, N. Dubrowinskaja, M. T. Goodman, H. B. Salvesen, P. A. Fasching, R. J. Scott, R. Delahanty, Y. Zheng, T. O'Mara, C. S. Healey, S. Hodgson, H. Risch, H. P. Yang, F. Amant, N. Turmanov, A. Schwake, G. Lurie, J. Trovik, M. W. Beckmann, K. Ashton, B. T. Ji, P. P. Bao, K. Howarth, L. Lu, J. Lissowska, L. Coenegrachts, D. Kaidarova, M. Durst, P. J. Thompson, C. Krakstad, A. B. Ekici, G. Otton, J. Shi, B. Zhang, M. Gorman, L. Brinton, A. Coosemans, R. K. Matsuno, M. K. Halle, A. Hein, A. Proietto, H. Cai, W. Lu, A. Dunning, D. Easton, Y. T. Gao, Q. Cai, A. B. Spurdle, and X. O. Shu. 2012. "Genome-Wide Association Study Identifies a Possible Susceptibility Locus for Endometrial Cancer." *Cancer Epidemiol.Biomarkers Prev.* 21(6):980-987. doi:1055-9965.EPI-11-1160 [pii];10.1158/1055-9965.EPI-11-1160 [doi].

Long, J. R., X. O. Shu, Q. Cai, H. Cai, Y. T. Gao, F. Jin, and W. Zheng. 10-20-2006. "Polymorphisms of the CYP1B1 Gene May Be Associated With the Onset of Natural Menopause in Chinese Women." *Maturitas* 55(3):238-46. doi:S0378-5122(06)00096-X [pii];10.1016/j.maturitas.2006.03.005 [doi].

Lotsch, J., M. Rohrbacher, H. Schmidt, A. Doehring, J. Brockmoller, and G. Geisslinger. 2009. "Can Extremely Low or High Morphine Formation From Codeine Be Predicted Prior to

Therapy Initiation?" *Pain.* 144(1-2):119-24. doi:S0304-3959(09)00193-6 [pii];10.1016/j.pain.2009.03.023 [doi].

Low, S. K., A. Takahashi, K. Ashikawa, J. Inazawa, Y. Miki, M. Kubo, Y. Nakamura, and T. Katagiri. 2013. "Genome-Wide Association Study of Breast Cancer in the Japanese Population." *PLoS One* 8(10):e76463. doi:10.1371/journal.pone.0076463 [doi];PONE-D-13-29227 [pii].

Lu, Y., P. Liu, R. R. Recker, H. W. Deng, and V. Dvornyk. 2010. "TNFRSF11A and TNFSF11 Are Associated With Age at Menarche and Natural Menopause in White Women." *Menopause.* 17(5):1048-54. doi:10.1097/gme.0b013e3181d5d523 [doi].

Lucenteforte, E., R. Talamini, M. Montella, Maso L. Dal, C. Pelucchi, S. Franceschi, Vecchia C. La, and E. Negri. 2009. "Family History of Cancer and the Risk of Endometrial Cancer." *Eur.J.Cancer Prev.* 18(2):95-99. doi:10.1097/CEJ.0b013e328305a0c9 [doi];00008469-200904000-00001 [pii].

Lumachi, F., A. Brunello, M. Maruzzo, U. Basso, and S. M. Basso. 12-17-2012. "Treatment of Estrogen Receptor-Positive Breast Cancer." *Curr.Med.Chem.* doi:CMC-EPUB-20121217-1 [pii].

Lupski, J. R., J. G. Reid, C. Gonzaga-Jauregui, Deiros D. Rio, D. C. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D. A. Wheeler, A. L. McGuire, F. Zhang, P. Stankiewicz, J. J. Halperin, C. Yang, C. Gehman, D. Guo, R. K. Irikat, W. Tom, N. J. Fantin, D. M. Muzny, and R. A. Gibbs. 4-1-2010. "Whole-Genome Sequencing in a Patient With Charcot-Marie-Tooth Neuropathy." *N.Engl.J.Med.* 362(13):1181-91. doi:NEJMoa0908094 [pii];10.1056/NEJMoa0908094 [doi].

Lynch, John, Jennifer Bevan, Paul Achter, Tina Harris, and Celeste M. Condit. 2008. "A Preliminary Study of How Multiple Exposures to Messages About Genetics Impact on Lay Attitudes Towards Racial and Genetic Discrimination." *New Genetics and Society* 27(1):43-56.

Ma, J., N. Ledbetter, and L. Glenn. 2013. "Testing Women With Endometrial Cancer for Lynch Syndrome: Should We Test All?" *J.Adv.Pract.Oncol.* 4(5):322-30.

Ma, L., A. G. Clark, and A. Keinan. 2013. "Gene-Based Testing of Interactions in Association Studies of Quantitative Traits." *PLoS Genet.* 9(2):e1003321. doi:10.1371/journal.pgen.1003321 [doi];PGENETICS-D-12-01978 [pii].

Madison, T., D. Schottenfeld, S. A. James, A. G. Schwartz, and S. B. Gruber. 2004. "Endometrial Cancer: Socioeconomic Status and Racial/Ethnic Differences in Stage at Diagnosis, Treatment, and Survival." *Am.J.Public.Health* 94(12):2104-11. doi:94/12/2104 [pii].

Maeda, D., X. Chen, B. Guan, S. Nakagawa, T. Yano, Y. Taketani, M. Fukayama, T. L. Wang, and IeM Shih. 2011. "Rsf-1 (HBXAP) Expression Is Associated With Advanced Stage and Lymph Node Metastasis in Ovarian Clear Cell Carcinoma." *Int.J.Gynecol.Pathol.* 30(1):30-35. doi:10.1097/PGP.0b013e3181e9a319 [doi].

Malin, B. A., K. E. Emam, and C. M. O'Keefe. 1-1-2013. "Biomedical Data Privacy: Problems, Perspectives, and Recent Advances." *J.Am.Med.Inform.Assoc.* 20(1):2-6. doi:amiajnl-2012-001509 [pii];10.1136/amiajnl-2012-001509 [doi].

Malott, R. W. 2007. "Are Women, People of Color, Asians, and Southern Europeans Inherently Inferior to North-European Males? A History of Biological Determinism--A Cultural, Spiritual and Intellectual Disgrace--And the Implications for Understanding "Mental Illness"." *Behavior and Social Issues* 16(2):134-69.

Mani, S., A. Ozdas, C. Aliferis, H. A. Varol, Q. Chen, R. Carnevale, Y. Chen, J. Romano-Keeler, H. Nian, and J. H. Weitkamp. 2014. "Medical Decision Support Using Machine Learning for Early Detection of Late-Onset Neonatal Sepsis." *J.Am.Med.Inform.Assoc.* 21(2):326-36. doi:amiajnl-2013-001854 [pii];10.1136/amiajnl-2013-001854 [doi].

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher. 10-8-2009. "Finding the Missing Heritability of Complex Diseases." *Nature.* 461(7265):747-53. doi:nature08494 [pii];10.1038/nature08494 [doi].

Manson, J. E. 2014. "Current Recommendations: What Is the Clinician to Do?" *Fertil.Steril.* 101(4):916-21. doi:S0015-0282(14)00196-4 [pii];10.1016/j.fertnstert.2014.02.043 [doi].

Markatseli, A. E., E. Hatzi, C. Pamporaki, I. Bouba, N. Xita, S. Tigas, I. Georgiou, and A. Tsatsoulis. 2014. "Association of the (TAAAA)n Repeat Polymorphism of SHBG Gene With the Age at Menopause in Greek Postmenopausal Women." *Maturitas.* 78(2):113-16. doi:S0378-5122(14)00095-4 [pii];10.1016/j.maturitas.2014.03.009 [doi].

Marshall, W. A. and J. M. Tanner. 1969. "Variations in Pattern of Pubertal Changes in Girls." *Arch.Dis.Child.* 44(235):291-303.

Marteau, T., V. Senior, S. E. Humphries, M. Bobrow, T. Cranston, M. A. Crook, L. Day, M. Fernandez, R. Horne, A. Iversen, Z. Jackson, J. Lynas, H. Middleton-Price, R. Savine, J. Sikorski, M. Watson, J. Weinman, A. S. Wierzbicki, and R. Wray. 7-30-2004. "Psychological Impact of Genetic Testing for Familial Hypercholesterolemia Within a Previously Aware Population: a Randomized Controlled Trial." *Am.J.Med.Genet.A.* 128A(3):285-93. doi:10.1002/ajmg.a.30102 [doi].

Marteau, T. M. and J. Weinman. 2006. "Self-Regulation and the Behavioural Response to DNA Risk Information: a Theoretical Analysis and Framework for Future Research." *Soc.Sci.Med.* 62(6):1360-1368. doi:S0277-9536(05)00422-3 [pii];10.1016/j.socscimed.2005.08.005 [doi].

Marzullo, P., A. Minocci, M. A. Tagliaferri, G. Guzzaloni, Blasio A. Di, Medici C. De, G. Aimaretti, and A. Liuzzi. 2010. "Investigations of Thyroid Hormones and Antibodies in Obesity: Leptin Levels Are Associated With Thyroid Autoimmunity Independent of

Bioanthropometric, Hormonal, and Weight-Related Determinants." *J.Clin.Endocrinol.Metab.* 95(8):3965-72. doi:jc.2009-2798 [pii];10.1210/jc.2009-2798 [doi].

Matikainen, T., G. I. Perez, A. Jurisicova, J. K. Pru, J. J. Schlezinger, H. Y. Ryu, J. Laine, T. Sakai, S. J. Korsmeyer, R. F. Casper, D. H. Sherr, and J. L. Tilly. 2001. "Aromatic Hydrocarbon Receptor-Driven Bax Gene Expression Is Required for Premature Ovarian Failure Caused by Biohazardous Environmental Chemicals." *Nat.Genet.* 28(4):355-60. doi:10.1038/ng575 [doi];ng575 [pii].

Matise, T. C., J. L. Ambite, S. Buyske, C. S. Carlson, S. A. Cole, D. C. Crawford, C. A. Haiman, G. Heiss, C. Kooperberg, L. L. Marchand, T. A. Manolio, K. E. North, U. Peters, M. D. Ritchie, L. A. Hindorff, and J. L. Haines. 10-1-2011. "The Next PAGE in Understanding Complex Traits: Design for the Analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study." *Am.J.Epidemiol.* 174(7):849-59. doi:kwr160 [pii];10.1093/aje/kwr160 [doi].

Mattison, D. R., M. S. Nightingale, and K. Shiromizu. 1983. "Effects of Toxic Substances on Female Reproduction." *Environ.Health Perspect.* 48:43-52.

Matzuk, M. M. and D. J. Lamb. 2008. "The Biology of Infertility: Research Advances and Clinical Challenges." *Nat.Med.* 14(11):1197-213. doi:nm.f.1895 [pii];10.1038/nm.f.1895 [doi].

McAndrew, L. M., T. J. Musumeci-Szabo, P. A. Mora, L. Vileikyte, E. Burns, E. A. Halm, E. A. Leventhal, and H. Leventhal. 2008. "Using the Common Sense Model to Design Interventions for the Prevention and Management of Chronic Illness Threats: From Description to Process." *Br.J.Health Psychol.* 13(Pt 2):195-204. doi:296500 [pii];10.1348/135910708X295604 [doi].

McBride, C. M., G. Bepler, I. M. Lipkus, P. Lyna, G. Samsa, J. Albright, S. Datta, and B. K. Rimer. 2002. "Incorporating Genetic Susceptibility Feedback into a Smoking Cessation Program for African-American Smokers With Low Income." *Cancer Epidemiol.Biomarkers.Prev.* 11(6):521-28.

McCarthy, D., K. Mueller, and J. Wrenn, McCarthy, D., K. Mueller, and J. Wrenn. 2009. *Commonwealth Fund Case Study Organized Health Care Delivery System 2009 Kaiser Permanente: Bridging the Quality Divide with Integrated Practice, Group Accountability, and Health Information Technology.*

McCarty, C. A., R. L. Chisholm, C. G. Chute, I. J. Kullo, G. P. Jarvik, E. B. Larson, R. Li, D. R. Masys, M. D. Ritchie, D. M. Roden, J. P. Struewing, and W. A. Wolf. 2011. "The EMERGE Network: a Consortium of Biorepositories Linked to Electronic Medical Records Data for Conducting Genomic Studies." *BMC.Med.Genomics.* 4:13. doi:1755-8794-4-13 [pii];10.1186/1755-8794-4-13 [doi].

McClellan, K. A., E. Kleiderman, L. Black, K. Bouchard, M. Dorval, J. Simard, B. M. Knoppers, and D. Avard. 2013. "Exploring Resources for Intrafamilial Communication of Cancer Genetic Risk: We Still Need to Talk." *Eur.J.Hum.Genet.* 21(9):903-10. doi:ejhg2012286 [pii];10.1038/ejhg.2012.286 [doi].

McCoy, A. B., E. J. Thomas, M. Krousel-Wood, and D. F. Sittig. 2014. "Clinical Decision Support Alert Appropriateness: a Review and Proposal for Improvement." *Ochsner.J.* 14(2):195-202.

McDowell, M. A., D. J. Brody, and J. P. Hughes. 2007. "Has Age at Menarche Changed? Results From the National Health and Nutrition Examination Survey (NHANES) 1999-2004." *J.Adolesc.Health* 40(3):227-31. doi:S1054-139X(06)00394-6 [pii];10.1016/j.jadohealth.2006.10.002 [doi].

Means, J. H. 1940. "Hypothyroidism: Diagnosis and Treatment." *Bull.N.Y.Acad.Med.* 16(1):14-19.

Medici, M., W. M. van der Deure, M. Verbiest, S. H. Vermeulen, P. S. Hansen, L. A. Kiemeney, A. R. Hermus, M. M. Breteler, A. Hofman, L. Hegedus, K. O. Kyvik, Heijer M. den, A. G. Uitterlinden, T. J. Visser, and R. P. Peeters. 2011. "A Large-Scale Association Analysis of 68 Thyroid Hormone Pathway Genes With Serum TSH and FT4 Levels." *Eur.J.Endocrinol.* 164(5):781-88. doi:EJE-10-1130 [pii];10.1530/EJE-10-1130 [doi].

Meikle, A. W., J. D. Stringham, M. G. Woodward, and J. C. Nelson. 1988. "Hereditary and Environmental Influences on the Variation of Thyroid Hormones in Normal Male Twins." *J.Clin.Endocrinol.Metab.* 66(3):588-92. doi:10.1210/jcem-66-3-588 [doi].

Meng, F. T., Y. L. Wang, J. Liu, J. Zhao, R. Y. Liu, and J. N. Zhou. 7-27-2011a. "ApoE Genotypes Are Associated With Age at Natural Menopause in Chinese Females." *Age (Dordr.)*. doi:10.1007/s11357-011-9287-4 [doi].

-----. 7-27-2011b. "ApoE Genotypes Are Associated With Age at Natural Menopause in Chinese Females." *Age (Dordr.)*. doi:10.1007/s11357-011-9287-4 [doi].

Menon, A. K., X. Jiang, J. Kim, J. Vaidya, and L. Ohno-Machado. 4-1-2014. "Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering." *Mach.Learn.* 95(1):87-101. doi:10.1007/s10994-013-5376-1 [doi].

Michailidou, K., P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, R. L. Milne, M. K. Schmidt, J. Chang-Claude, S. E. Bojesen, M. K. Bolla, Q. Wang, E. Dicks, A. Lee, C. Turnbull, N. Rahman, O. Fletcher, J. Peto, L. Gibson, Santos Silva Dos, I, H. Nevanlinna, T. A. Muranen, K. Aittomaki, C. Blomqvist, K. Czene, A. Irwanto, J. Liu, Q. Waisfisz, H. Meijers-Heijboer, M. Adank, R. B. van der Luijt, R. Hein, N. Dahmen, L. Beckman, A. Meindl, R. K. Schmutzler, B. Muller-Myhsok, P. Lichtner, J. L. Hopper, M. C. Southey, E. Makalic, D. F. Schmidt, A. G. Uitterlinden, A. Hofman, D. J. Hunter, S. J. Chanock, D. Vincent, F. Bacot, D. C. Tessier, S. Canisius, L. F. Wessels, C. A. Haiman, M. Shah, R. Luben, J. Brown, C. Luccarini, N. Schoof, K. Humphreys, J. Li, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, F. J. Couch, X. Wang, C. Vachon, K. N. Stevens, D. Lambrechts, M. Moisse, R. Paridaens, M. R. Christiaens, A. Rudolph, S. Nickels, D. Flesch-Janys, N. Johnson, Z. Aitken, K. Aaltonen, T. Heikkinen, A. Broeks, L. J. Veer, C. E. van der Schoot, P. Guenel, T. Truong, P. Laurent-Puig, F. Menegaux, F. Marme, A. Schneeweiss, C. Sohn, B. Burwinkel, M. P. Zamora, J. I. Perez, G. Pita, M. R. Alonso, A. Cox, I. W. Brock, S. S. Cross, M. W. Reed, E. J. Sawyer, I. Tomlinson, M. J. Kerin, N. Miller, B. E. Henderson, F. Schumacher, Marchand L. Le, I. L. Andrulis, J. A. Knight, G. Glendon, A.

M. Mulligan, A. Lindblom, S. Margolin, M. J. Hooning, A. Hollestelle, A. M. van den Ouweland, A. Jager, Q. M. Bui, J. Stone, G. S. Dite, C. Apicella, H. Tsimiklis, G. G. Giles, G. Severi, L. Baglietto, P. A. Fasching, L. Haeberle, A. B. Ekici, M. W. Beckmann, H. Brenner, H. Muller, V. Arndt, C. Stegmaier, A. Swerdlow, A. Ashworth, N. Orr, M. Jones, J. Figueroa, J. Lissowska, L. Brinton, M. S. Goldberg, F. Labreche, M. Dumont, R. Winqvist, K. Pylkas, A. Jukkola-Vuorinen, M. Grip, H. Brauch, U. Hamann, T. Bruning, P. Radice, P. Peterlongo, S. Manoukian, B. Bonanni, P. Devilee, R. A. Tollenaar, C. Seynaeve, C. J. van Asperen, A. Jakubowska, J. Lubinski, K. Jaworska, K. Durda, A. Mannermaa, V. Kataja, V. M. Kosma, J. M. Hartikainen, N. V. Bogdanova, N. N. Antonenkova, T. Dork, V. N. Kristensen, H. Anton-Culver, S. Slager, A. E. Toland, S. Edge, F. Fostira, D. Kang, K. Y. Yoo, D. Y. Noh, K. Matsuo, H. Ito, H. Iwata, A. Sueta, A. H. Wu, C. C. Tseng, D. Van Den Berg, D. O. Stram, X. O. Shu, W. Lu, Y. T. Gao, H. Cai, S. H. Teo, C. H. Yip, S. Y. Phuah, B. K. Cornes, M. Hartman, H. Miao, W. Y. Lim, J. H. Sng, K. Muir, A. Lophatananon, S. Stewart-Brown, P. Siriwanarangsan, C. Y. Shen, C. N. Hsiung, P. E. Wu, S. L. Ding, S. Sangrajrang, V. Gaborieau, P. Brennan, J. McKay, W. J. Blot, L. B. Signorello, Q. Cai, W. Zheng, S. Deming-Halverson, M. Shrubsole, J. Long, J. Simard, M. Garcia-Closas, P. D. Pharoah, G. Chenevix-Trench, A. M. Dunning, J. Benitez, and D. F. Easton. 2013. "Large-Scale Genotyping Identifies 41 New Loci Associated With Breast Cancer Risk." *Nat.Genet.* 45(4):353-2. doi:ng.2563 [pii];10.1038/ng.2563 [doi].

Michie, S., J. A. Smith, V. Senior, and T. M. Marteau. 6-15-2003. "Understanding Why Negative Genetic Test Results Sometimes Fail to Reassure." *Am.J.Med.Genet.A.* 119A(3):340-347. doi:10.1002/ajmg.a.20200 [doi].

Michie, S., J. Weinman, J. Miller, V. Collins, J. Halliday, and T. M. Marteau. 2002. "Predictive Genetic Testing: High Risk Expectations in the Face of Low Risk Information." *J.Behav.Med.* 25(1):33-50.

Michnovicz, J. J., R. J. Hershcopf, H. Naganuma, H. L. Bradlow, and J. Fishman. 11-20-1986. "Increased 2-Hydroxylation of Estradiol As a Possible Mechanism for the Anti-Estrogenic Effect of Cigarette Smoking." *N.Engl.J.Med.* 315(21):1305-9. doi:10.1056/NEJM198611203152101 [doi].

Middleton, B., M. Bloomrosen, M. A. Dente, B. Hashmat, R. Koppel, J. M. Overhage, T. H. Payne, S. T. Rosenbloom, C. Weaver, and J. Zhang. 2013. "Enhancing Patient Safety and Quality of Care by Improving the Usability of Electronic Health Record Systems: Recommendations From AMIA." *J.Am.Med.Inform.Assoc.* 20(e1):e2-e8. doi:amiajnl-2012-001458 [pii];10.1136/amiajnl-2012-001458 [doi].

Mihaescu, R., R. Moonesinghe, M. J. Khoury, and A. C. Janssens. 2011. "Predictive Genetic Testing for the Identification of High-Risk Groups: a Simulation Study on the Impact of Predictive Ability." *Genome.Med.* 3(7):51. doi:gm267 [pii];10.1186/gm267 [doi].

Mills, R., W. Barry, and S. B. Haga. 2014. "Public Trust in Genomic Risk Assessment for Type 2 Diabetes Mellitus." *J.Genet.Couns.* 23(3):401-8. doi:10.1007/s10897-013-9674-3 [doi].

Milne, F. H., D. S. Judge, D. B. Preen, and P. Weinstein. 2011. "Early Life Environment, Life History and Risk of Endometrial Cancer." *Med.Hypotheses* 77(4):626-32. doi:S0306-9877(11)00318-5 [pii];10.1016/j.mehy.2011.07.001 [doi].

Mitchell, E. S., F. M. Farin, P. L. Stapleton, J. M. Tsai, E. Y. Tao, K. Smith-Dijulio, and N. F. Woods. 2008. "Association of Estrogen-Related Polymorphisms With Age at Menarche, Age at Final Menstrual Period, and Stages of the Menopausal Transition." *Menopause.* 15(1):105-11. doi:10.1097/gme.0b013e31804d2406 [doi].

Modugno, F., R. B. Ness, C. Chen, and N. S. Weiss. 2005. "Inflammation and Endometrial Cancer: a Hypothesis." *Cancer Epidemiol.Biomarkers Prev.* 14(12):2840-2847. doi:14/12/2840 [pii];10.1158/1055-9965.EPI-05-0493 [doi].

Murabito, J. M., Q. Yang, C. S. Fox, and L. A. Cupples. 2005. "Genome-Wide Linkage Analysis to Age at Natural Menopause in a Community-Based Sample: the Framingham Heart Study." *Fertil.Steril.* 84(6):1674-79. doi:S0015-0282(05)02933-X [pii];10.1016/j.fertnstert.2005.05.046 [doi].

Murphy, K., P. Travers, and M. Walport. 2008. "Janeway's Immunobiology." Pp. 611-654 in *Autoimmunity and Transplantation*. New York: Garland Science.

Murray, A., C. E. Bennett, J. R. Perry, M. N. Weedon, P. A. Jacobs, D. H. Morris, N. Orr, M. J. Schoemaker, M. Jones, A. Ashworth, and A. J. Swerdlow. 1-1-2011. "Common Genetic Variants Are Significant Risk Factors for Early Menopause: Results From the Breakthrough Generations Study." *Hum.Mol.Genet.* 20(1):186-92. doi:ddq417 [pii];10.1093/hmg/ddq417 [doi].

Murray, J. F. 6-15-2012. "Personalized Medicine: Been There, Done That, Always Needs Work!" *Am.J.Respir.Crit.Care Med.* 185(12):1251-52. doi:185/12/1251 [pii];10.1164/rccm.201203-0523ED [doi].

Must, A., S. M. Phillips, E. N. Naumova, M. Blum, S. Harris, B. Dawson-Hughes, and W. M. Rand. 4-1-2002. "Recall of Early Menstrual History and Menarcheal Body Size: After 30 Years, How Well Do Women Remember?" *Am.J.Epidemiol.* 155(7):672-79.

Napolitano, C., S. G. Priori, P. J. Schwartz, R. Bloise, E. Ronchetti, J. Nastoli, G. Bottelli, M. Cerrone, and S. Leonardi. 12-21-2005. "Genetic Testing in the Long QT Syndrome: Development and Validation of an Efficient Approach to Genotyping in Clinical Practice." *JAMA.* 294(23):2975-80. doi:294/23/2975 [pii];10.1001/jama.294.23.2975 [doi].

Narod, S. A. 11-2-2011a. "Alcohol and Risk of Breast Cancer." *JAMA* 306(17):1920-1921. doi:306/17/1920 [pii];10.1001/jama.2011.1589 [doi].

-----. 2011b. "Early-Onset Breast Cancer: What Do We Know About the Risk Factors?: A Countercurrents Series." *Curr.Oncol.* 18(5):204-5.

Nathan, B. M. and M. R. Palmert. 2005. "Regulation and Disorders of Pubertal Timing." *Endocrinol.Metab.Clin.North Am.* 34(3):617-41, ix. doi:S0889-8529(05)00038-1 [pii];10.1016/j.ecl.2005.04.015 [doi].

Nathan, N. and S. D. Sullivan. 2014. "Thyroid Disorders During Pregnancy." *Endocrinol.Metab.Clin.North Am.* 43(2):573-97. doi:S0889-8529(14)00009-7 [pii];10.1016/j.ecl.2014.02.008 [doi].

National Cancer Institute at the National Institutes of Health.,*Levels of Evidence for Cancer Genetic Studies (PDQ) Analytic Validity* .

-----.,*Levels of Evidence for Cancer Genetics Studies (PDQ) Clinical Utility* .

-----.,*Levels of Evidence for Cancer Genetics Studies (PDQ) Clinical Validity* .

-----.,*BRCA1 and BRCA2: Cancer Risk and Genetic Testing* .

-----.,*Endometrial Cancer* .

-----.,*General Information about non-small cell lung cancer* .

National Center for Immunization and Respiratory Diseases.,*PCV13 (Pneumococcal Conjugate) Vaccine* .

National Endocrine and Metabolic Diseases Information Service (NEMDIS), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and National Institutes of Health.,*Thyroid Tests: NIH Publication No. 14-6284* .

National Eye Institute at the National Institutes of Health.,*Facts About Age-Related Macular Degeneration* .

National Human Genome Research Institute.,*Genetic Discrimination* .

-----.,*Talking Glossary of Genetic Terms* .

National Institutes of Health.,*National Institutes of Health Revitalization Act of 1993 Subtitle B--Clinical Research Equity Regarding Women and Minorities* .

National Library of Medicine.,*Genetics Home Reference: What is Pharmacogenomics* .

Need, A. C., V. Shashi, Y. Hitomi, K. Schoch, K. V. Shianna, M. T. McDonald, M. H. Meisler, and D. B. Goldstein. 2012. "Clinical Application of Exome Sequencing in Undiagnosed Genetic Conditions." *J.Med.Genet.* 49(6):353-61. doi:jmedgenet-2012-100819 [pii];10.1136/jmedgenet-2012-100819 [doi].

Nelkin, D. and M. S. Lindee. 1996. ""Genes Made Me Do It": the Appeal of Biological Explanations." *Politics.Life.Sci.* 15(1):95-97.

Nelson, M. R., S. L. Kardia, R. E. Ferrell, and C. F. Sing. 2001. "A Combinatorial Partitioning Method to Identify Multilocus Genotypic Partitions That Predict Quantitative Trait Variation." *Genome.Res.* 11(3):458-70. doi:10.1101/gr.172901 [doi].

Newton, K. M., P. L. Peissig, A. N. Kho, S. J. Bielinski, R. L. Berg, V. Choudhary, M. Basford, C. G. Chute, I. J. Kullo, R. Li, J. A. Pacheco, L. V. Rasmussen, L. Spangler, and J. C. Denny. 2013. "Validation of Electronic Medical Record-Based Phenotyping Algorithms: Results and Lessons Learned From the EMERGE Network." *J.Am.Med.Inform.Assoc.* 20(e1):e147-e154. doi:amiajnl-2012-000896 [pii];10.1136/amiajnl-2012-000896 [doi].

Nussbaum, R. L., R. R. McInnes, and H. F. Willard. 2007. "Thompson & Thompson Genetics In Medicine." Pp. 115 in *Patterns of Single-Gene Inheritance*. Philadelphia: Saunders Elsevier.

Nyholt, D. R., S. K. Low, C. A. Anderson, J. N. Painter, S. Uno, A. P. Morris, S. Macgregor, S. D. Gordon, A. K. Henders, N. G. Martin, J. Attia, E. G. Holliday, M. McEvoy, R. J. Scott, S. H. Kennedy, S. A. Treloar, S. A. Missmer, S. Adachi, K. Tanaka, Y. Nakamura, K. T. Zondervan, H. Zembutsu, and G. W. Montgomery. 2012. "Genome-Wide Association Meta-Analysis Identifies New Endometriosis Risk Loci." *Nat.Genet.* 44(12):1355-59. doi:ng.2445 [pii];10.1038/ng.2445 [doi].

Nyrnes, A., R. Jorde, and J. Sundsfjord. 2006. "Serum TSH Is Positively Associated With BMI." *Int.J.Obes.(Lond.)* 30(1):100-105. doi:0803112 [pii];10.1038/sj.ijo.0803112 [doi].

O'Hara, A. J. and D. W. Bell. 2012. "The Genomics and Genetics of Endometrial Cancer." *Adv.Genomics Genet.* 2012(2):33-47. doi:10.2147/AGG.S28953 [doi].

O'Mara, T. A., P. Fahey, K. Ferguson, L. Marquart, D. Lambrechts, E. Despierre, I. Vergote, F. Amant, P. Hall, J. Liu, K. Czene, T. R. Rebbeck, S. Ahmed, A. M. Dunning, C. S. Gregory, M. Shah, P. M. Webb, and A. B. Spurdle. 2011a. "Progesterone Receptor Gene Variants and Risk of Endometrial Cancer." *Carcinogenesis.* 32(3):331-35. doi:bgq263 [pii];10.1093/carcin/bgq263 [doi].

O'Mara, T. A., K. Ferguson, P. Fahey, L. Marquart, H. P. Yang, J. Lissowska, S. Chanock, M. Garcia-Closas, D. J. Thompson, C. S. Healey, A. M. Dunning, D. F. Easton, P. M. Webb, and A. B. Spurdle. 2011b. "CHEK2, MGMT, SULT1E1 and SULT1A1 Polymorphisms and Endometrial Cancer Risk." *Twin.Res.Hum.Genet.* 14(4):328-32. doi:10.1375/twin.14.4.328 [doi];10.1375/twin.14.4.328 [pii].

Ober, C., D. A. Loisel, and Y. Gilad. 2008. "Sex-Specific Genetic Architecture of Human Disease." *Nat.Rev.Genet.* 9(12):911-22. doi:nrg2415 [pii];10.1038/nrg2415 [doi].

Ober, C. and D. Vercelli. 2011. "Gene-Environment Interactions in Human Disease: Nuisance or Opportunity?" *Trends.Genet.* 27(3):107-15. doi:S0168-9525(10)00231-3 [pii];10.1016/j.tig.2010.12.004 [doi].

Office of Minority Health and U.S.Department of Health and Human Services.,*Diabetes and African Americans* .

-----.,*Diabetes data/statistics* .

-----.,*Heart disease and African Americans* .

Okamura, H., H. Tsukamura, S. Ohkura, Y. Uenoyama, Y. Wakabayashi, and K. Maeda. 2013. "Kisspeptin and GnRH Pulse Generation." *Adv.Exp.Med.Biol.* 784:297-323. doi:10.1007/978-1-4614-6199-9_14 [doi].

Okayasu, I., Y. Hara, K. Nakamura, and N. R. Rose. 1994. "Racial and Age-Related Differences in Incidence and Severity of Focal Autoimmune Thyroiditis." *Am.J.Clin.Pathol.* 101(6):698-702.

Okeke, T., U. Anyaehie, and C. Ezenyeaku. 2013. "Premature Menopause." *Ann.Med.Health Sci.Res.* 3(1):90-95. doi:10.4103/2141-9248.109458 [doi];AMHSR-3-90 [pii].

Olson, S. H., C. Chen, Vivo De, I, J. A. Doherty, V. Hartmuller, P. L. Horn-Ross, J. V. Lacey, Jr., S. M. Lynch, L. Sansbury, V. W. Setiawan, L. J. Schouten, and X. O. Shu. 2009. "Maximizing Resources to Study an Uncommon Cancer: E2C2--Epidemiology of Endometrial Cancer Consortium." *Cancer Causes Control* 20(4):491-96. doi:10.1007/s10552-008-9290-y [doi].

Ong, K. K., C. E. Elks, S. Li, J. H. Zhao, J. Luan, L. B. Andersen, S. A. Bingham, S. Brage, G. D. Smith, U. Ekelund, C. J. Gillson, B. Glaser, J. Golding, R. Hardy, K. T. Khaw, D. Kuh, R. Luben, M. Marcus, M. A. McGeehin, A. R. Ness, K. Northstone, S. M. Ring, C. Rubin, M. A. Sims, K. Song, D. P. Strachan, P. Vollenweider, G. Waeber, D. M. Waterworth, A. Wong, P. Deloukas, I. Barroso, V. Mooser, R. J. Loos, and N. J. Wareham. 2009. "Genetic Variation in LIN28B Is Associated With the Timing of Puberty." *Nat.Genet.* 41(6):729-33. doi:ng.382 [pii];10.1038/ng.382 [doi].

Online Mendelian Inheritance in Man, OMI.,*Cystic Fibrosis Transmembrance Conductance Regulator; CFTR: MIM 602421* Baltimore, MD: Johns Hopkins University.

Opdahl, S., M. D. Alsaker, I. Janszky, P. R. Romundstad, and L. J. Vatten. 8-23-2011. "Joint Effects of Nulliparity and Other Breast Cancer Risk Factors." *Br.J.Cancer* 105(5):731-36. doi:bjc2011286 [pii];10.1038/bjc.2011.286 [doi].

Pabalan, N., M. R. Pineda, H. Jarjanazi, D. M. Christofolini, C. P. Barbosa, and B. Bianco. 7-20-2014. "Association of the +331G/A Progesterone Receptor Gene (PgR) Polymorphism With Risk of Endometrial Cancer in Caucasian Women: a Meta-Analysis." *Arch.Gynecol.Obstet.* doi:10.1007/s00404-014-3344-z [doi].

Painter, J. N., C. A. Anderson, D. R. Nyholt, S. Macgregor, J. Lin, S. H. Lee, A. Lambert, Z. Z. Zhao, F. Roseman, Q. Guo, S. D. Gordon, L. Wallace, A. K. Henders, P. M. Visscher, P. Kraft, N. G. Martin, A. P. Morris, S. A. Treloar, S. H. Kennedy, S. A. Missmer, G. W. Montgomery, and K. T. Zondervan. 2011. "Genome-Wide Association Study Identifies a Locus at 7p15.2 Associated With Endometriosis." *Nat.Genet.* 43(1):51-54. doi:ng.731 [pii];10.1038/ng.731 [doi].

Palmer, J. R., L. Rosenberg, L. A. Wise, N. J. Horton, and L. L. Adams-Campbell. 2003. "Onset of Natural Menopause in African American Women." *Am.J.Public Health* 93(2):299-306.

Palomaki, G. E., L. A. Bradley, M. P. Douglas, K. Kolor, and W. D. Dotson. 2009a. "Can UGT1A1 Genotyping Reduce Morbidity and Mortality in Patients With Metastatic Colorectal Cancer Treated With Irinotecan? An Evidence-Based Review." *Genet.Med.* 11(1):21-34. doi:10.1097/GIM.0b013e31818efd77 [doi].

Palomaki, G. E., M. R. McClain, S. Melillo, H. L. Hampel, and S. N. Thibodeau. 2009b. "EGAPP Supplementary Evidence Review: DNA Testing Strategies Aimed at Reducing Morbidity and Mortality From Lynch Syndrome." *Genet.Med.* 11(1):42-65. doi:10.1097/GIM.0b013e31818fa2db [doi].

Palomaki, G. E., S. Melillo, M. Marrone, and M. P. Douglas. 2013. "Use of Genomic Panels to Determine Risk of Developing Type 2 Diabetes in the General Population: a Targeted Evidence-Based Review." *Genet.Med.* 15(8):600-611. doi:gim20138 [pii];10.1038/gim.2013.8 [doi].

Palomaki, G. E., S. Melillo, L. Neveux, M. P. Douglas, W. D. Dotson, A. C. Janssens, E. A. Balkite, and L. A. Bradley. 2010. "Use of Genomic Profiling to Assess Risk for Cardiovascular Disease and Identify Individualized Prevention Strategies--a Targeted Evidence-Based Review." *Genet.Med.* 12(12):772-84. doi:10.1097/GIM.0b013e3181f8728d [doi].

Pan, F., P. Xiao, Y. Guo, Y. J. Liu, H. Y. Deng, R. R. Recker, and H. W. Deng. 2008. "Chromosomal Regions 22q13 and 3p25 May Harbor Quantitative Trait Loci Influencing Both Age at Menarche and Bone Mineral Density." *Hum.Genet.* 123(4):419-27. doi:10.1007/s00439-008-0490-z [doi].

Pan, Q. 2013. "Multiple Hypotheses Testing Procedures in Clinical Trials and Genomic Studies." *Front.Public.Health* 1:63. doi:10.3389/fpubh.2013.00063 [doi].

Panicker, V. 2011. "Genetics of Thyroid Function and Disease." *Clin.Biochem.Rev.* 32(4):165-75.

Panicker, V., S. G. Wilson, T. D. Spector, S. J. Brown, M. Falchi, J. B. Richards, G. L. Surdulescu, E. M. Lim, S. J. Fletcher, and J. P. Walsh. 2008a. "Heritability of Serum TSH, Free T4 and Free T3 Concentrations: a Study of a Large UK Twin Cohort." *Clin.Endocrinol.(Oxf.)* 68(4):652-59. doi:CEN3079 [pii];10.1111/j.1365-2265.2007.03079.x [doi].

Panicker, V., S. G. Wilson, T. D. Spector, S. J. Brown, B. S. Kato, P. W. Reed, M. Falchi, J. B. Richards, G. L. Surdulescu, E. M. Lim, S. J. Fletcher, and J. P. Walsh. 2008b. "Genetic Loci Linked to Pituitary-Thyroid Axis Set Points: a Genome-Wide Scan of a Large Twin Cohort." *J.Clin.Endocrinol.Metab.* 93(9):3519-23. doi:jc.2007-2650 [pii];10.1210/jc.2007-2650 [doi].

Panicker, V., S. G. Wilson, J. P. Walsh, J. B. Richards, S. J. Brown, J. P. Beilby, A. P. Bremner, G. L. Surdulescu, E. Qweitin, I. Gillham-Nasenya, N. Soranzo, E. M. Lim, S. J. Fletcher, and T. D. Spector. 9-10-2010. "A Locus on Chromosome 1p36 Is Associated With Thyrotropin

and Thyroid Function As Identified by Genome-Wide Association Study." *Am.J.Hum.Genet.* 87(3):430-435. doi:S0002-9297(10)00418-0 [pii];10.1016/j.ajhg.2010.08.005 [doi].

Parrott, R., M. L. Kahl, K. Ndiaye, and T. Traeder. 2012. "Health Communication, Genetic Determinism, and Perceived Control: the Roles of Beliefs About Susceptibility and Severity Versus Disease Essentialism." *J.Health Commun.* 17(7):762-78. doi:10.1080/10810730.2012.677301 [doi].

Parrott, Roxanne, Kami Silk, Judith Weiner, Celeste Condit, Tina Harris, and Jay Bernhardt. 2004. "Deriving Lay Models of Uncertainty About Genes' Role in Illness Causation to Guide Communication About Human Genetics." *Journal of Communication* 54(1):105-22.

Paulden, M., J. Franek, B. Pham, P. L. Bedard, M. Trudeau, and M. Krahn. 2013. "Cost-Effectiveness of the 21-Gene Assay for Guiding Adjuvant Chemotherapy Decisions in Early Breast Cancer." *Value.Health* 16(5):729-39. doi:S1098-3015(13)01698-7 [pii];10.1016/j.jval.2013.03.1625 [doi].

Payne, T. H., D. W. Bates, E. S. Berner, E. V. Bernstam, H. D. Covvey, M. E. Frisse, T. Graf, R. A. Greenes, E. P. Hoffer, G. Kuperman, H. P. Lehmann, L. Liang, B. Middleton, G. S. Omenn, and J. Ozbolt. 2013. "Healthcare Information Technology and Economics." *J.Am.Med.Inform.Assoc.* 20(2):212-17. doi:amiajnl-2012-000821 [pii];10.1136/amiajnl-2012-000821 [doi].

Pendergrass, S. A., K. Brown-Gentry, S. M. Dudek, E. S. Torstenson, J. L. Ambite, C. L. Avery, S. Buyske, C. Cai, M. D. Fesinmeyer, C. Haiman, G. Heiss, L. A. Hindorff, C. N. Hsu, R. D. Jackson, C. Kooperberg, Marchand L. Le, Y. Lin, T. C. Matise, L. Moreland, K. Monroe, A. P. Reiner, R. Wallace, L. R. Wilkens, D. C. Crawford, and M. D. Ritchie. 2011. "The Use of Phenome-Wide Association Studies (PheWAS) for Exploration of Novel Genotype-Phenotype Relationships and Pleiotropy Discovery." *Genet.Epidemiol.* 35(5):410-422. doi:10.1002/gepi.20589 [doi].

Pendergrass, S. A., S. M. Dudek, D. C. Crawford, and M. D. Ritchie. 2010. "Synthesis-View: Visualization and Interpretation of SNP Association Results for Multi-Cohort, Multi-Phenotype Data and Meta-Analysis." *BioData.Min* 3:10. doi:1756-0381-3-10 [pii];10.1186/1756-0381-3-10 [doi].

Perry, J. R., Y. H. Hsu, D. I. Chasman, A. D. Johnson, C. Elks, E. Albrecht, I. L. Andrulis, J. Beesley, G. S. Berenson, S. Bergmann, S. E. Bojesen, M. K. Bolla, J. Brown, J. E. Buring, H. Campbell, J. Chang-Claude, G. Chenevix-Trench, T. Corre, F. J. Couch, A. Cox, K. Czene, A. P. d'Adamo, G. Davies, I. J. Deary, J. Dennis, D. F. Easton, E. G. Engelhardt, J. G. Eriksson, T. Esko, P. A. Fasching, J. D. Figueroa, H. Flyger, A. Fraser, M. Garcia-Closas, P. Gasparini, C. Gieger, G. Giles, P. Guenel, S. Hagg, P. Hall, C. Hayward, J. Hopper, E. Ingelsson, S. L. Kardia, K. Kasiman, J. A. Knight, J. Lahti, D. A. Lawlor, P. K. Magnusson, S. Margolin, J. A. Marsh, A. Metspalu, J. E. Olson, C. E. Pennell, O. Polasek, I. Rahman, P. M. Ridker, A. Robino, I. Rudan, A. Rudolph, A. Salumets, M. K. Schmidt, M. J. Schoemaker, E. N. Smith, J. A. Smith, M. Southey, D. Stockl, A. J. Swerdlow, D. J. Thompson, T. Truong, S. Ulivi, M. Waldenberger, Q. Wang, S. Wild, J. F. Wilson, A. F.

Wright, L. Zgaga, K. K. Ong, J. M. Murabito, D. Karasik, and A. Murray. 5-1-2014. "DNA Mismatch Repair Gene MSH6 Implicated in Determining Age at Natural Menopause." *Hum.Mol.Genet.* 23(9):2490-2497. doi:ddt620 [pii];10.1093/hmg/ddt620 [doi].

Perry, J. R., L. Stolk, N. Franceschini, K. L. Lunetta, G. Zhai, P. F. McArdle, A. V. Smith, T. Aspelund, S. Bandinelli, E. Boerwinkle, L. Cherkas, G. Eiriksdottir, K. Estrada, L. Ferrucci, A. R. Folsom, M. Garcia, V. Gudnason, A. Hofman, D. Karasik, D. P. Kiel, L. J. Launer, Meurs J. van, M. A. Nalls, F. Rivadeneira, A. R. Shuldiner, A. Singleton, N. Soranzo, T. Tanaka, J. A. Visser, M. N. Weedon, S. G. Wilson, V. Zhuang, E. A. Streeten, T. B. Harris, A. Murray, T. D. Spector, E. W. Demerath, A. G. Uitterlinden, and J. M. Murabito. 2009. "Meta-Analysis of Genome-Wide Association Data Identifies Two Loci Influencing Age at Menarche." *Nat.Genet.* 41(6):648-50. doi:ng.386 [pii];10.1038/ng.386 [doi].

Peshkin, B. N., M. D. Schwartz, C. Isaacs, C. Hughes, D. Main, and C. Lerman. 2002. "Utilization of Breast Cancer Screening in a Clinically Based Sample of Women After BRCA1/2 Testing." *Cancer Epidemiol.Biomarkers.Prev.* 11(10 Pt 1):1115-18.

Petros, M. 2012. "Revisiting the Wilson-Jungner Criteria: How Can Supplemental Criteria Guide Public Health in the Era of Genetic Screening?" *Genet.Med.* 14(1):129-34. doi:gim0b013e31823331d0 [pii];10.1038/gim.0b013e31823331d0 [doi].

Phelan, J. C., L. H. Yang, and R. Cruz-Rojas. 2006. "Effects of Attributing Serious Mental Illnesses to Genetic Causes on Orientations to Treatment." *Psychiatr.Serv.* 57(3):382-87. doi:57/3/382 [pii];10.1176/appi.ps.57.3.382 [doi].

Phillips, K. A., Sakowski J. Ann, J. Trosman, M. P. Douglas, S. Y. Liang, and P. Neumann. 2014. "The Economic Value of Personalized Medicine Tests: What We Know and What We Need to Know." *Genet.Med.* 16(3):251-57. doi:gim2013122 [pii];10.1038/gim.2013.122 [doi].

Pink, J., M. Pirmohamed, S. Lane, and D. A. Hughes. 2014. "Cost-Effectiveness of Pharmacogenetics-Guided Warfarin Therapy Vs. Alternative Anticoagulation in Atrial Fibrillation." *Clin.Pharmacol.Ther.* 95(2):199-207. doi:clpt2013190 [pii];10.1038/clpt.2013.190 [doi].

Porcu, E., M. Medici, G. Pistis, C. B. Volpato, S. G. Wilson, A. R. Cappola, S. D. Bos, J. Deelen, Heijer M. den, R. M. Freathy, J. Lahti, C. Liu, L. M. Lopez, I. M. Nolte, J. R. O'Connell, T. Tanaka, S. Trompet, A. Arnold, S. Bandinelli, M. Beekman, S. Bohringer, S. J. Brown, B. M. Buckley, C. Camaschella, A. J. de Craen, G. Davies, M. C. de Visser, I. Ford, T. Forsen, T. M. Frayling, L. Fugazzola, M. Gogele, A. T. Hattersley, A. R. Hermus, A. Hofman, J. J. Houwing-Duistermaat, R. A. Jensen, E. Kajantie, M. Kloppenburg, E. M. Lim, C. Masciullo, S. Mariotti, C. Minelli, B. D. Mitchell, R. Nagaraja, R. T. Netea-Maier, A. Palotie, L. Persani, M. G. Piras, B. M. Psaty, K. Raikkonen, J. B. Richards, F. Rivadeneira, C. Sala, M. M. Sabra, N. Sattar, B. M. Shields, N. Soranzo, J. M. Starr, D. J. Stott, F. C. Sweep, G. Usala, M. M. Van der Klauw, Heemst D. van, Mullem A. van, S. H. Vermeulen, W. E. Visser, J. P. Walsh, R. G. Westendorp, E. Widen, G. Zhai, F. Cucca, I. J. Deary, J. G. Eriksson, L. Ferrucci, C. S. Fox, J. W. Jukema, L. A. Kiemeney, P. P.

Pramstaller, D. Schlessinger, A. R. Shuldiner, E. P. Slagboom, A. G. Uitterlinden, B. Vaidya, T. J. Visser, B. H. Wolffenbuttel, I. Meulenbelt, J. I. Rotter, T. D. Spector, A. A. Hicks, D. Toniolo, S. Sanna, R. P. Peeters, and S. Naitza. 2013. "A Meta-Analysis of Thyroid-Related Traits Reveals Novel Loci and Gender-Specific Differences in the Regulation of Thyroid Function." *PLoS Genet.* 9(2):e1003266. doi:10.1371/journal.pgen.1003266 [doi];PGENETICS-D-12-01354 [pii].

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nat.Genet.* 38(8):904-9. doi:ng1847 [pii];10.1038/ng1847 [doi].

Pruim, R. J. et al. 2010. "LocusZoom: Regional Visualization of Genome-Wide Association Scan Results." 26.

Public Health Service Food and Drug Administration and A. Gutierrez.,*Inspections, Compliance, Enforcement, and Criminal Investigations 23&Me, Inc. 11/22/13* .

Pulley, J. M., J. C. Denny, J. F. Peterson, G. R. Bernard, C. L. Vnencak-Jones, A. H. Ramirez, J. T. Delaney, E. Bowton, K. Brothers, K. Johnson, D. C. Crawford, J. Schildcrout, D. R. Masys, H. H. Dilks, R. A. Wilke, E. W. Clayton, E. Shultz, M. Laposata, J. McPherson, J. N. Jirjis, and D. M. Roden. 2012. "Operational Implementation of Prospective Genotyping for Personalized Medicine: the Design of the Vanderbilt PREDICT Project." *Clin.Pharmacol.Ther.* 92(1):87-95. doi:clpt2011371 [pii];10.1038/clpt.2011.371 [doi].

Purcell, S. PLINK V. 1.07. electronic resource/URL . 10-10-2009.
Ref Type: Electronic Citation

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. "PLINK: a Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *Am.J.Hum.Genet.* 81(3):559-75. doi:S0002-9297(07)61352-4 [pii];10.1086/519795 [doi].

R Development Core Team. R: A Language and Environment for Statistical Computing. 2012. R Foundation for Statistical Computing.
Ref Type: Computer Program

Rafnar, T., P. Sulem, S. N. Stacey, F. Geller, J. Gudmundsson, A. Sigurdsson, M. Jakobsdottir, H. Helgadottir, S. Thorlacius, K. K. Aben, T. Blondal, T. E. Thorgeirsson, G. Thorleifsson, K. Kristjansson, K. Thorisdottir, R. Ragnarsson, B. Sigurgeirsson, H. Skuladottir, T. Gudbjartsson, H. J. Isaksson, G. V. Einarsson, K. R. Benediktsdottir, B. A. Agnarsson, K. Olafsson, A. Salvarsdottir, H. Bjarnason, M. Asgeirsdottir, K. T. Kristinsson, S. Matthiasdottir, S. G. Sveinsdottir, S. Polidoro, V. Hoiom, R. Botella-Estrada, K. Hemminki, P. Rudnai, D. T. Bishop, M. Campagna, E. Kellen, M. P. Zeegers, Verdier P. de, A. Ferrer, D. Isla, M. J. Vidal, R. Andres, B. Saez, P. Juberias, J. Banzo, S. Navarrete, A. Tres, D. Kan, A. Lindblom, E. Gurzau, K. Koppova, Vegt F. de, J. A. Schalken, H. F. van der Heijden, H. J. Smit, R. A. Termeer, E. Oosterwijk, Hooij O. van, E. Nagore, S. Porru, G. Steineck, J. Hansson, F. Buntinx, W. J. Catalona, G. Matullo, P. Vineis, A. E. Kiltie, J. I. Mayordomo, R. Kumar, L. A. Kiemeney, M. L. Frigge, T. Jonsson, H.

Saemundsson, R. B. Barkardottir, E. Jonsson, S. Jonsson, J. H. Olafsson, J. R. Gulcher, G. Masson, D. F. Gudbjartsson, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2009. "Sequence Variants at the TERT-CLPTM1L Locus Associate With Many Cancer Types." *Nat.Genet.* 41(2):221-27. doi:ng.296 [pii];10.1038/ng.296 [doi].

Rahman, M., R. K. Simmons, A. H. Harding, N. J. Wareham, and S. J. Griffin. 2008. "A Simple Risk Score Identifies Individuals at High Risk of Developing Type 2 Diabetes: a Prospective Cohort Study." *Fam.Pract.* 25(3):191-96. doi:cmn024 [pii];10.1093/fampra/cmn024 [doi].

Rajagopalan, R. and J. H. Fujimura. 2012. "Will Personalized Medicine Challenge or Reify Categories of Race and Ethnicity?" *Virtual.Mentor.* 14(8):657-63. doi:virtualmentor.2012.14.8.msoc1-1208 [pii];10.1001/virtualmentor.2012.14.8.msoc1-1208 [doi].

Ramirez, A. H., J. S. Schildcrout, D. L. Blakemore, D. R. Masys, J. M. Pulley, M. A. Basford, D. M. Roden, and J. C. Denny. 2011. "Modulators of Normal Electrocardiographic Intervals Identified in a Large Electronic Medical Record." *Heart Rhythm.* 8(2):271-77. doi:S1547-5271(10)01139-2 [pii];10.1016/j.hrthm.2010.10.034 [doi].

Ramsey, B. W., J. Davies, N. G. McElvaney, E. Tullis, S. C. Bell, P. Drevinek, M. Griese, E. F. McKone, C. E. Wainwright, M. W. Konstan, R. Moss, F. Ratjen, I. Sermet-Gaudelus, S. M. Rowe, Q. Dong, S. Rodriguez, K. Yen, C. Ordonez, and J. S. Elborn. 11-3-2011. "A CFTR Potentiator in Patients With Cystic Fibrosis and the G551D Mutation." *N.Engl.J.Med.* 365(18):1663-72. doi:10.1056/NEJMoa1105185 [doi].

Rasmussen-Torvik, L. J., J. A. Pacheco, R. A. Wilke, W. K. Thompson, M. D. Ritchie, A. N. Kho, A. Muthalagu, M. G. Hayes, L. L. Armstrong, D. A. Scheftner, J. T. Wilkins, R. L. Zuvich, D. Crosslin, D. M. Roden, J. C. Denny, G. P. Jarvik, C. S. Carlson, I. J. Kullo, S. J. Bielinski, C. A. McCarty, R. Li, T. A. Manolio, D. C. Crawford, and R. L. Chisholm. 2012. "High Density GWAS for LDL Cholesterol in African Americans Using Electronic Medical Records Reveals a Strong Protective Variant in APOE." *Clin.Transl.Sci.* 5(5):394-99. doi:10.1111/j.1752-8062.2012.00446.x [doi].

Rasmussen-Torvik, L. J., J. S. Pankow, D. R. Jacobs, Jr., J. Steinberger, A. Moran, and A. R. Sinaiko. 2009. "The Association of SNPs in ADIPOQ, ADIPOR1, and ADIPOR2 With Insulin Sensitivity in a Cohort of Adolescents and Their Parents." *Hum.Genet.* 125(1):21-28. doi:10.1007/s00439-008-0595-4 [doi].

Rawal, R., A. Teumer, H. Volzke, H. Wallaschofski, T. Ittermann, B. O. Asvold, T. Bjoro, K. H. Greiser, D. Tiller, K. Werdan, H. E. Meyer Zu Schwabedissen, A. Doering, T. Illig, C. Gieger, C. Meisinger, and G. Homuth. 7-15-2012. "Meta-Analysis of Two Genome-Wide Association Studies Identifies Four Genetic Loci Associated With Thyroid Function." *Hum.Mol.Genet.* 21(14):3275-82. doi:dds136 [pii];10.1093/hmg/dds136 [doi].

Real, L. M., A. Ruiz, J. Gayan, A. Gonzalez-Perez, M. E. Saez, R. Ramirez-Lorca, F. J. Moron, J. Velasco, R. Marginet-Flinch, E. Musulen, J. M. Carrasco, C. Moreno-Rey, E. Vazquez, M. Chaves-Conde, J. A. Moreno-Nogueira, M. Hidalgo-Pascual, E. Ferrero-Herrero, S.

Castellvi-Bel, A. Castells, C. Fernandez-Rozadilla, C. Ruiz-Ponte, A. Carracedo, B. Gonzalez, S. Alonso, and M. Perucho. 2014. "A Colorectal Cancer Susceptibility New Variant at 4q26 in the Spanish Population Identified by Genome-Wide Association Analysis." *PLoS One* 9(6):e101178. doi:10.1371/journal.pone.0101178 [doi];PONE-D-14-15882 [pii].

Rees, S. D., M. Z. Hydrie, A. S. Shera, S. Kumar, J. P. O'Hare, A. H. Barnett, A. Basit, and M. A. Kelly. 2011. "Replication of 13 Genome-Wide Association (GWA)-Validated Risk Variants for Type 2 Diabetes in Pakistani Populations." *Diabetologia* 54(6):1368-74. doi:10.1007/s00125-011-2063-2 [doi].

Reich, D. E. and E. S. Lander. 2001. "On the Allelic Spectrum of Human Disease." *Trends.Genet.* 17(9):502-10. doi:S0168-9525(01)02410-6 [pii].

Reinhardt, D., B. T. Helfand, P. R. Cooper, K. A. Roehl, W. J. Catalona, and S. Loeb. 12-15-2013. "Prostate Cancer Risk Alleles Are Associated With Prostate Cancer Volume and Prostate Size." *J.Urol.* doi:S0022-5347(13)06128-4 [pii];10.1016/j.juro.2013.12.030 [doi].

Ripatti, S., E. Tikkanen, M. Orho-Melander, A. S. Havulinna, K. Silander, A. Sharma, C. Guiducci, M. Perola, A. Jula, J. Sinisalo, M. L. Lokki, M. S. Nieminen, O. Melander, V. Salomaa, L. Peltonen, and S. Kathiresan. 10-23-2010. "A Multilocus Genetic Risk Score for Coronary Heart Disease: Case-Control and Prospective Cohort Analyses." *Lancet* 376(9750):1393-400. doi:S0140-6736(10)61267-6 [pii];10.1016/S0140-6736(10)61267-6 [doi].

Ritchie, M. D., J. C. Denny, R. L. Zuvich, D. C. Crawford, J. S. Schildcrout, L. Bastarache, A. H. Ramirez, J. D. Mosley, J. M. Pulley, M. A. Basford, Y. Bradford, L. V. Rasmussen, J. Pathak, C. G. Chute, I. J. Kullo, C. A. McCarty, R. L. Chisholm, A. N. Kho, C. S. Carlson, E. B. Larson, G. P. Jarvik, N. Sotoodehnia, T. A. Manolio, R. Li, D. R. Masys, J. L. Haines, and D. M. Roden. 4-2-2013. "Genome- and Phenome-Wide Analyses of Cardiac Conduction Identifies Markers of Arrhythmia Risk." *Circulation* 127(13):1377-85. doi:CIRCULATIONAHA.112.000604 [pii];10.1161/CIRCULATIONAHA.112.000604 [doi].

Ritchie, M. D., L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl, and J. H. Moore. 2001. "Multifactor-Dimensionality Reduction Reveals High-Order Interactions Among Estrogen-Metabolism Genes in Sporadic Breast Cancer." *Am.J.Hum.Genet.* 69(1):138-47. doi:S0002-9297(07)61453-0 [pii];10.1086/321276 [doi].

Robbins, C., J. B. Torres, S. Hooker, C. Bonilla, W. Hernandez, A. Candreva, C. Ahaghotu, R. Kittles, and J. Carpten. 2007. "Confirmation Study of Prostate Cancer Risk Variants at 8q24 in African Americans Identifies a Novel Risk Locus." *Genome.Res.* 17(12):1717-22. doi:gr.6782707 [pii];10.1101/gr.6782707 [doi].

Roberts, J. S., K. D. Christensen, and R. C. Green. 2011. "Using Alzheimer's Disease As a Model for Genetic Risk Disclosure: Implications for Personal Genomics." *Clin.Genet.* 80(5):407-14. doi:10.1111/j.1399-0004.2011.01739.x [doi].

Robson, M. and K. Offit. 7-12-2007. "Clinical Practice. Management of an Inherited Predisposition to Breast Cancer." *N.Engl.J.Med.* 357(2):154-62. doi:357/2/154 [pii];10.1056/NEJMcp071286 [doi].

Rosenthal, A., P. Mork, M. H. Li, J. Stanford, D. Koester, and P. Reynolds. 2010. "Cloud Computing: a New Business Paradigm for Biomedical Information Sharing." *J.Biomed.Inform.* 43(2):342-53. doi:S1532-0464(09)00115-4 [pii];10.1016/j.jbi.2009.08.014 [doi].

Ross, L. F., H. M. Saal, K. L. David, and R. R. Anderson. 2013. "Technical Report: Ethical and Policy Issues in Genetic Testing and Screening of Children." *Genet.Med.* 15(3):234-45. doi:gim2012176 [pii];10.1038/gim.2012.176 [doi].

Rossolatos, D. and K. J. Aitchison. 1-28-2014. "Genomics for Clinical Utility: the Future Is Near." *Genome.Med.* 6(1):3. doi:gm522 [pii];10.1186/gm522 [doi].

Rothenbuhler, A., D. Fradin, S. Heath, H. Lefevre, C. Bouvattier, M. Lathrop, and P. Bougneres. 2006. "Weight-Adjusted Genome Scan Analysis for Mapping Quantitative Trait Loci for Menarchal Age." *J.Clin.Endocrinol.Metab* 91(9):3534-37. doi:jc.2006-0150 [pii];10.1210/jc.2006-0150 [doi].

Rothman, N., M. Garcia-Closas, N. Chatterjee, N. Malats, X. Wu, J. D. Figueroa, F. X. Real, D. Van Den Berg, G. Matullo, D. Baris, M. Thun, L. A. Kiemeney, P. Vineis, Vivo De, I, D. Albanes, M. P. Purdue, T. Rafnar, M. A. Hildebrandt, A. E. Kiltie, O. Cussenot, K. Golka, R. Kumar, J. A. Taylor, J. I. Mayordomo, K. B. Jacobs, M. Kogevinas, A. Hutchinson, Z. Wang, Y. P. Fu, L. Prokunina-Olsson, L. Burdett, M. Yeager, W. Wheeler, A. Tardon, C. Serra, A. Carrato, R. Garcia-Closas, J. Lloreta, A. Johnson, M. Schwenn, M. R. Karagas, A. Schned, G. Andriole, Jr., R. Grubb, III, A. Black, E. J. Jacobs, W. R. Diver, S. M. Gapstur, S. J. Weinstein, J. Virtamo, V. K. Cortessis, M. Gago-Dominguez, M. C. Pike, M. C. Stern, J. M. Yuan, D. J. Hunter, M. McGrath, C. P. Dinney, B. Czerniak, M. Chen, H. Yang, S. H. Vermeulen, K. K. Aben, J. A. Witjes, R. R. Makkinje, P. Sulem, S. Besenbacher, K. Stefansson, E. Riboli, P. Brennan, S. Panico, C. Navarro, N. E. Allen, H. B. Bueno-de-Mesquita, D. Trichopoulos, N. Caporaso, M. T. Landi, F. Canzian, B. Ljungberg, A. Tjonneland, F. Clavel-Chapelon, D. T. Bishop, M. T. Teo, M. A. Knowles, S. Guarrera, S. Polidoro, F. Ricceri, C. Sacerdote, A. Allione, G. Cancel-Tassin, S. Selinski, J. G. Hengstler, H. Dietrich, T. Fletcher, P. Rudnai, E. Gurzau, K. Koppova, S. C. Bolick, A. Godfrey, Z. Xu, J. I. Sanz-Velez, D. Garcia-Prats, M. Sanchez, G. Valdivia, S. Porru, S. Benhamou, R. N. Hoover, J. F. Fraumeni, Jr., D. T. Silverman, and S. J. Chanock. 2010. "A Multi-Stage Genome-Wide Association Study of Bladder Cancer Identifies Multiple Susceptibility Loci." *Nat.Genet.* 42(11):978-84. doi:ng.687 [pii];10.1038/ng.687 [doi].

Rowe, S. M., S. Miller, and E. J. Sorscher. 5-12-2005. "Cystic Fibrosis." *N.Engl.J.Med.* 352(19):1992-2001. doi:352/19/1992 [pii];10.1056/NEJMra043184 [doi].

Ruiz-Alonso, M., D. Blesa, and C. Simon. 2012. "The Genomics of the Human Endometrium." *Biochim.Biophys.Acta* 1822(12):1931-42. doi:S0925-4439(12)00112-3 [pii];10.1016/j.bbadis.2012.05.004 [doi].

Saft, C., B. R. Leavitt, and J. T. Epplen. 2014. "Clinical Utility Gene Card for: Huntington's Disease." *Eur.J.Hum.Genet.* 22(5). doi:ejhg2013206 [pii];10.1038/ejhg.2013.206 [doi].

Saif-Ali, R., I. S. Ismail, Z. Al-Hamodi, H. M. Al-Mekhlafi, L. C. Siang, A. M. Alabsi, and S. Muniandy. 2011a. "KCNQ1 Haplotypes Associate With Type 2 Diabetes in Malaysian Chinese Subjects." *Int.J.Mol.Sci.* 12(9):5705-18. doi:10.3390/ijms12095705 [doi];ijms-12-05705 [pii].

Saif-Ali, R., S. Muniandy, Z. Al-Hamodi, C. S. Lee, K. A. Ahmed, A. M. Al-Mekhlafi, and I. S. Ismail. 2011b. "KCNQ1 Variants Associate With Type 2 Diabetes in Malaysian Malay Subjects." *Ann.Acad.Med.Singapore* 40(11):488-92.

Salpeter, S. R., J. M. Walsh, E. Greyber, and E. E. Salpeter. 2006. "Brief Report: Coronary Heart Disease Events Associated With Hormone Therapy in Younger and Older Women. A Meta-Analysis." *J.Gen.Intern.Med.* 21(4):363-66. doi:JGI389 [pii];10.1111/j.1525-1497.2006.00389.x [doi].

Salsberry, P. J., P. B. Reagan, and K. Pajer. 2009. "Growth Differences by Age of Menarche in African American and White Girls." *Nurs.Res.* 58(6):382-90. doi:10.1097/NNR.0b013e3181b4b921 [doi].

Samollow, P. B., G. Perez, C. M. Kammerer, D. Finegold, P. W. Zwartjes, L. M. Havill, A. G. Comuzzie, M. C. Mahaney, H. H. Goring, J. Blangero, T. P. Foley, and M. M. Barmada. 2004. "Genetic and Environmental Influences on Thyroid Hormone Variation in Mexican Americans." *J.Clin.Endocrinol.Metab.* 89(7):3276-84. doi:10.1210/jc.2003-031706 [doi];89/7/3276 [pii].

San Francisco, I. F., P. A. Rojas, V. Torres-Estay, S. Smalley, J. Cerda-Infante, V. P. Montecinos, C. Hurtado, and A. S. Godoy. 2014. "Association of RNASEL and 8q24 Variants With the Presence and Aggressiveness of Hereditary and Sporadic Prostate Cancer in a Hispanic Population." *J.Cell Mol.Med.* 18(1):125-33. doi:10.1111/jcmm.12171 [doi].

Santoro, N. and J. F. Randolph, Jr. 2011. "Reproductive Hormones and the Menopause Transition." *Obstet.Gynecol.Clin.North Am.* 38(3):455-66. doi:S0889-8545(11)00068-4 [pii];10.1016/j.ogc.2011.05.004 [doi].

Schadt, E. E., M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan. 2010. "Computational Solutions to Large-Scale Data Management and Analysis." *Nat.Rev.Genet.* 11(9):647-57. doi:nrg2857 [pii];10.1038/nrg2857 [doi].

Schildkraut, J. M., N. Risch, and W. D. Thompson. 1989. "Evaluating Genetic Association Among Ovarian, Breast, and Endometrial Cancer: Evidence for a Breast/Ovarian Cancer Relationship." *Am.J.Hum.Genet.* 45(4):521-29.

Schneeweiss, S. 6-5-2014. "Learning From Big Health Care Data." *N.Engl.J.Med.* 370(23):2161-63. doi:10.1056/NEJMp1401111 [doi].

Schunkert, H., I. R. Konig, S. Kathiresan, M. P. Reilly, T. L. Assimes, H. Holm, M. Preuss, A. F. Stewart, M. Barbalic, C. Gieger, D. Absher, Z. Aherrahrou, H. Allayee, D. Altshuler, S. S. Anand, K. Andersen, J. L. Anderson, D. Ardissino, S. G. Ball, A. J. Balmforth, T. A. Barnes, D. M. Becker, L. C. Becker, K. Berger, J. C. Bis, S. M. Boekholdt, E. Boerwinkle, P. S. Braund, M. J. Brown, M. S. Burnett, I. Buysschaert, J. F. Carlquist, L. Chen, S. Cichon, V. Codd, R. W. Davies, G. Dedoussis, A. Dehghan, S. Demissie, J. M. Devaney, P. Diemert, R. Do, A. Doering, S. Eifert, N. E. Mokhtari, S. G. Ellis, R. Elosua, J. C. Engert, S. E. Epstein, Faire U. de, M. Fischer, A. R. Folsom, J. Freyer, B. Gigante, D. Girelli, S. Gretarsdottir, V. Gudnason, J. R. Gulcher, E. Halperin, N. Hammond, S. L. Hazen, A. Hofman, B. D. Horne, T. Illig, C. Iribarren, G. T. Jones, J. W. Jukema, M. A. Kaiser, L. M. Kaplan, J. J. Kastelein, K. T. Khaw, J. W. Knowles, G. Kolovou, A. Kong, R. Laaksonen, D. Lambrechts, K. Leander, G. Lettre, M. Li, W. Lieb, C. Loley, A. J. Lotery, P. M. Mannucci, S. Maouche, N. Martinelli, P. P. McKeown, C. Meisinger, T. Meitinger, O. Melander, P. A. Merlini, V. Mooser, T. Morgan, T. W. Muhleisen, J. B. Muhlestein, T. Munzel, K. Musunuru, J. Nahrstaedt, C. P. Nelson, M. M. Nothen, O. Olivieri, R. S. Patel, C. C. Patterson, A. Peters, F. Peyvandi, L. Qu, A. A. Quyyumi, D. J. Rader, L. S. Rallidis, C. Rice, F. R. Rosendaal, D. Rubin, V. Salomaa, M. L. Sampietro, M. S. Sandhu, E. Schadt, A. Schafer, A. Schillert, S. Schreiber, J. Schrezenmeir, S. M. Schwartz, D. S. Siscovick, M. Sivananthan, S. Sivapalaratnam, A. Smith, T. B. Smith, J. D. Snoep, N. Soranzo, J. A. Spertus, K. Stark, K. Stirrups, M. Stoll, W. H. Tang, S. Tennstedt, G. Thorgeirsson, G. Thorleifsson, M. Tomaszewski, A. G. Uitterlinden, A. M. van Rij, B. F. Voight, N. J. Wareham, G. A. Wells, H. E. Wichmann, P. S. Wild, C. Willenborg, J. C. Witteman, B. J. Wright, S. Ye, T. Zeller, A. Ziegler, F. Cambien, A. H. Goodall, L. A. Cupples, T. Quertermous, W. Marz, C. Hengstenberg, S. Blankenberg, W. H. Ouwehand, A. S. Hall, P. Deloukas, J. R. Thompson, K. Stefansson, R. Roberts, U. Thorsteinsdottir, C. J. O'Donnell, R. McPherson, J. Erdmann, and N. J. Samani. 2011. "Large-Scale Association Analysis Identifies 13 New Susceptibility Loci for Coronary Artery Disease." *Nat.Genet.* 43(4):333-38. doi:ng.784 [pii];10.1038/ng.784 [doi].

Schwartz, M. D., B. N. Peshkin, C. Hughes, D. Main, C. Isaacs, and C. Lerman. 1-15-2002. "Impact of BRCA1/BRCA2 Mutation Testing on Psychologic Distress in a Clinic-Based Sample." *J.Clin.Oncol.* 20(2):514-20.

Scott, S. A., K. Sangkuhl, C. M. Stein, J. S. Hulot, J. L. Mega, D. M. Roden, T. E. Klein, M. S. Sabatine, J. A. Johnson, and A. R. Shuldiner. 2013. "Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C19 Genotype and Clopidogrel Therapy: 2013 Update." *Clin.Pharmacol.Ther.* 94(3):317-23. doi:clpt2013105 [pii];10.1038/clpt.2013.105 [doi].

Sehgal, R., K. Sheahan, P. R. O'Connell, A. M. Hanly, S. T. Martin, and D. C. Winter. 2014. "Lynch Syndrome: an Updated Review." *Genes.(Basel.)* 5(3):497-507. doi:genes5030497 [pii];10.3390/genes5030497 [doi].

Senior, V., T. M. Marteau, and T. J. Peters. 1999. "Will Genetic Testing for Predisposition for Disease Result in Fatalism? A Qualitative Study of Parents Responses to Neonatal Screening for Familial Hypercholesterolaemia." *Soc.Sci.Med.* 48(12):1857-60. doi:S0277953699000994 [pii].

Setiawan, V. W., J. Haessler, F. Schumacher, M. L. Cote, E. Deelman, M. D. Fesinmeyer, B. E. Henderson, R. D. Jackson, J. S. Vockler, L. R. Wilkens, S. Yasmeen, C. A. Haiman, U. Peters, Marchand L. Le, and C. Kooperberg. 2012. "HNF1B and Endometrial Cancer Risk: Results From the PAGE Study." *PLoS.One.* 7(1):e30390. doi:10.1371/journal.pone.0030390 [doi];PONE-D-11-22303 [pii].

Setiawan, V. W., M. C. Pike, L. N. Kolonel, A. M. Nomura, M. T. Goodman, and B. E. Henderson. 2-1-2007. "Racial/Ethnic Differences in Endometrial Cancer Risk: the Multiethnic Cohort Study." *Am.J.Epidemiol.* 165(3):262-70. doi:kwk010 [pii];10.1093/aje/kwk010 [doi].

Setiawan, V. W., F. Schumacher, J. Prescott, J. Haessler, J. Malinowski, N. Wentzensen, H. Yang, S. Chanock, L. Brinton, P. Hartge, J. Lissowska, S. L. Park, I. Cheng, W. S. Bush, D. C. Crawford, G. Ursin, P. Horn-Ross, L. Bernstein, L. Lu, H. Risch, H. Yu, L. C. Sakoda, J. Doherty, C. Chen, R. Jackson, S. Yasmeen, M. Cote, J. M. Kocarnik, U. Peters, P. Kraft, Vivo De, I, C. A. Haiman, C. Kooperberg, and Marchand L. Le. 5-15-2014. "Cross-Cancer Pleiotropic Analysis of Endometrial Cancer: PAGE and E2C2 Consortia." *Carcinogenesis.* doi:carcin/bgu107 [pii];10.1093/carcin/bgu107 [doi].

Shaw, S. J., C. Huebner, J. Armin, K. Orzech, and J. Vivian. 2009. "The Role of Culture in Health Literacy and Chronic Disease Screening and Management." *J.Immigr.Minor.Health* 11(6):460-467. doi:10.1007/s10903-008-9135-5 [doi].

Sheehy, A. M., D. B. Coursin, and R. A. Gabbay. 2009. "Back to Wilson and Jungner: 10 Good Reasons to Screen for Type 2 Diabetes Mellitus." *Mayo.Clin.Proc.* 84(1):38-42. doi:S0025-6196(11)60806-9 [pii];10.1016/S0025-6196(11)60806-9 [doi].

Shete, S., F. J. Hosking, L. B. Robertson, S. E. Dobbins, M. Sanson, B. Malmer, M. Simon, Y. Marie, B. Boisselier, J. Y. Delattre, K. Hoang-Xuan, Hallani S. El, A. Idbaih, D. Zelenika, U. Andersson, R. Henriksson, A. T. Bergenheim, M. Feychting, S. Lonn, A. Ahlbom, J. Schramm, M. Linnebank, K. Hemminki, R. Kumar, S. J. Hepworth, A. Price, G. Armstrong, Y. Liu, X. Gu, R. Yu, C. Lau, M. Schoemaker, K. Muir, A. Swerdlow, M. Lathrop, M. Bondy, and R. S. Houlston. 2009. "Genome-Wide Association Study Identifies Five Susceptibility Loci for Glioma." *Nat.Genet.* 41(8):899-904. doi:ng.407 [pii];10.1038/ng.407 [doi].

Shi, Y., L. Li, Z. Hu, S. Li, S. Wang, J. Liu, C. Wu, L. He, J. Zhou, Z. Li, T. Hu, Y. Chen, Y. Jia, S. Wang, L. Wu, X. Cheng, Z. Yang, R. Yang, X. Li, K. Huang, Q. Zhang, H. Zhou, F. Tang, Z. Chen, J. Shen, J. Jiang, H. Ding, H. Xing, S. Zhang, P. Qu, X. Song, Z. Lin, D. Deng, L. Xi, W. Lv, X. Han, G. Tao, L. Yan, Z. Han, Z. Li, X. Miao, S. Pan, Y. Shen, H. Wang, D. Liu, E. Gong, Z. Li, L. Zhou, X. Luan, C. Wang, Q. Song, S. Wu, H. Xu, J. Shen, F. Qiang, G. Ma, L. Liu, X. Chen, J. Liu, J. Wu, Y. Shen, Y. Wen, M. Chu, J. Yu, X. Hu, Y. Fan, H. He, Y. Jiang, Z. Lei, C. Liu, J. Chen, Y. Zhang, C. Yi, S. Chen, W. Li, D. Wang, Z. Wang, W. Di, K. Shen, D. Lin, H. Shen, Y. Feng, X. Xie, and D. Ma. 2013. "A Genome-Wide Association Study Identifies Two New Cervical Cancer Susceptibility Loci at 4q12 and 17q12." *Nat.Genet.* 45(8):918-22. doi:ng.2687 [pii];10.1038/ng.2687 [doi].

Shiloh, S., D. Rashuk-Rosenthal, and Y. Benyamini. 2002. "Illness Causal Attributions: an Exploratory Study of Their Structure and Associations With Other Illness Cognitions and Perceptions of Control." *J.Behav.Med.* 25(4):373-94.

Shoenbill, K., N. Fost, U. Tachinardi, and E. A. Mendonca. 2014. "Genetic Data and Electronic Health Records: a Discussion of Ethical, Logistical and Technological Considerations." *J.Am.Med.Inform.Assoc.* 21(1):171-80. doi:amiajnl-2013-001694 [pii];10.1136/amiajnl-2013-001694 [doi].

Sievert, L. L., L. Murphy, L. A. Morrison, A. M. Reza, and D. E. Brown. 2013. "Age at Menopause and Determinants of Hysterectomy and Menopause in a Multi-Ethnic Community: the Hilo Women's Health Study." *Maturitas.* 76(4):334-41. doi:S0378-5122(13)00259-4 [pii];10.1016/j.maturitas.2013.08.007 [doi].

Signorello, L. B., M. K. Hargreaves, M. D. Steinwandel, W. Zheng, Q. Cai, D. G. Schlundt, M. S. Buchowski, C. W. Arnold, J. K. McLaughlin, and W. J. Blot. 2005. "Southern Community Cohort Study: Establishing a Cohort to Investigate Health Disparities." *J.Natl.Med.Assoc.* 97(7):972-79.

Simons, L. A., J. Simons, Y. Friedlander, and J. McCallum. 2012. "Childbearing History and Late-Life Mortality: the Dubbo Study of Australian Elderly." *Age.Ageing* 41(4):523-28. doi:afs016 [pii];10.1093/ageing/afs016 [doi].

Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, and W. L. McGuire. 1-9-1987. "Human Breast Cancer: Correlation of Relapse and Survival With Amplification of the HER-2/Neu Oncogene." *Science.* 235(4785):177-82.

Slamon, D. J., B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton. 3-15-2001. "Use of Chemotherapy Plus a Monoclonal Antibody Against HER2 for Metastatic Breast Cancer That Overexpresses HER2." *N.Engl.J.Med.* 344(11):783-92. doi:10.1056/NEJM200103153441101 [doi].

Solovieff, N., C. Cotsapas, P. H. Lee, S. M. Purcell, and J. W. Smoller. 2013. "Pleiotropy in Complex Traits: Challenges and Strategies." *Nat.Rev.Genet.* 14(7):483-95. doi:nrg3461 [pii];10.1038/nrg3461 [doi].

Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. Luan, R. Magi, J. C. Randall, S. Vedantam, T. W. Winkler, L. Qi, T. Workalemahu, I. M. Heid, V. Steinthorsdottir, H. M. Stringham, M. N. Weedon, E. Wheeler, A. R. Wood, T. Ferreira, R. J. Weyant, A. V. Segre, K. Estrada, L. Liang, J. Nemesh, J. H. Park, S. Gustafsson, T. O. Kilpelainen, J. Yang, N. Bouatia-Naji, T. Esko, M. F. Feitosa, Z. Kutalik, M. Mangino, S. Raychaudhuri, A. Scherag, A. V. Smith, R. Welch, J. H. Zhao, K. K. Aben, D. M. Absher, N. Amin, A. L. Dixon, E. Fisher, N. L. Glazer, M. E. Goddard, N. L. Heard-Costa, V. Hoesel, J. J. Hottenga, A. Johansson, T. Johnson, S. Ketkar, C. Lamina, S. Li, M. F. Moffatt, R. H. Myers, N. Narisu, J. R. Perry, M. J. Peters, M. Preuss, S. Ripatti, F. Rivadeneira, C. Sandholt, L. J. Scott, N. J. Timpson, J. P. Tyrer, Wingerden S. van, R. M. Watanabe, C. C. White, F. Wiklund, C. Barlassina, D. I.

Chasman, M. N. Cooper, J. O. Jansson, R. W. Lawrence, N. Pellikka, I. Prokopenko, J. Shi, E. Thiering, H. Alavere, M. T. Alibrandi, P. Almgren, A. M. Arnold, T. Aspelund, L. D. Atwood, B. Balkau, A. J. Balmforth, A. J. Bennett, Y. Ben-Shlomo, R. N. Bergman, S. Bergmann, H. Biebermann, A. I. Blakemore, T. Boes, L. L. Bonnycastle, S. R. Bornstein, M. J. Brown, T. A. Buchanan, F. Busonero, H. Campbell, F. P. Cappuccio, C. Cavalcanti-Proenca, Y. D. Chen, C. M. Chen, P. S. Chines, R. Clarke, L. Coin, J. Connell, I. N. Day, Heijer M. den, J. Duan, S. Ebrahim, P. Elliott, R. Elosua, G. Eiriksdottir, M. R. Erdos, J. G. Eriksson, M. F. Facheris, S. B. Felix, P. Fischer-Posovszky, A. R. Folsom, N. Friedrich, N. B. Freimer, M. Fu, S. Gaget, P. V. Gejman, E. J. Geus, C. Gieger, A. P. Gjesing, A. Goel, P. Goyette, H. Grallert, J. Grassler, D. M. Greenawalt, C. J. Groves, V. Gudnason, C. Guiducci, A. L. Hartikainen, N. Hassanali, A. S. Hall, A. S. Havulinna, C. Hayward, A. C. Heath, C. Hengstenberg, A. A. Hicks, A. Hinney, A. Hofman, G. Homuth, J. Hui, W. Igl, C. Iribarren, B. Isomaa, K. B. Jacobs, I. Jarick, E. Jewell, U. John, T. Jorgensen, P. Jousilahti, A. Jula, M. Kaakinen, E. Kajantie, L. M. Kaplan, S. Kathiresan, J. Kettunen, L. Kinnunen, J. W. Knowles, I. Kolcic, I. R. Konig, S. Koskinen, P. Kovacs, J. Kuusisto, P. Kraft, K. Kvaloy, J. Laitinen, O. Lantieri, C. Lanzani, L. J. Launer, C. Lecoeur, T. Lehtimaki, G. Lettre, J. Liu, M. L. Lokki, M. Lorentzon, R. N. Luben, B. Ludwig, P. Manunta, D. Marek, M. Marre, N. G. Martin, W. L. McArdle, A. McCarthy, B. McKnight, T. Meitinger, O. Melander, D. Meyre, K. Midthjell, G. W. Montgomery, M. A. Morken, A. P. Morris, R. Mulic, J. S. Ngwa, M. Nelis, M. J. Neville, D. R. Nyholt, C. J. O'Donnell, S. O'Rahilly, K. K. Ong, B. Oostra, G. Pare, A. N. Parker, M. Perola, I. Pichler, K. H. Pietilainen, C. G. Platou, O. Polasek, A. Pouta, S. Rafelt, O. Raitakari, N. W. Rayner, M. Ridderstrale, W. Rief, A. Ruokonen, N. R. Robertson, P. Rzehak, V. Salomaa, A. R. Sanders, M. S. Sandhu, S. Sanna, J. Saramies, M. J. Savolainen, S. Scherag, S. Schipf, S. Schreiber, H. Schunkert, K. Silander, J. Sinisalo, D. S. Siscovick, J. H. Smit, N. Soranzo, U. Sovio, J. Stephens, I. Surakka, A. J. Swift, M. L. Tammesoo, J. C. Tardif, M. Teder-Laving, T. M. Teslovich, J. R. Thompson, B. Thomson, A. Tonjes, T. Tuomi, J. B. van Meurs, and G. J. van Ommen. 2010. "Association Analyses of 249,796 Individuals Reveal 18 New Loci Associated With Body Mass Index." *Nat.Genet.* 42(11):937-48. doi:ng.686 [pii];10.1038/ng.686 [doi].

Spencer, C. A., J. G. Hollowell, M. Kazarosyan, and L. E. Braverman. 2007. "National Health and Nutrition Examination Survey III Thyroid-Stimulating Hormone (TSH)-Thyroperoxidase Antibody Relationships Demonstrate That TSH Upper Reference Limits May Be Skewed by Occult Thyroid Dysfunction." *J.Clin.Endocrinol.Metab.* 92(11):4236-40. doi:jc.2007-0287 [pii];10.1210/jc.2007-0287 [doi].

Spurdle, A. B., D. J. Thompson, S. Ahmed, K. Ferguson, C. S. Healey, T. O'Mara, L. C. Walker, S. B. Montgomery, E. T. Dermitzakis, P. Fahey, G. W. Montgomery, P. M. Webb, P. A. Fasching, M. W. Beckmann, A. B. Ekici, A. Hein, D. Lambrechts, L. Coenegrachts, I. Vergote, F. Amant, H. B. Salvesen, J. Trovik, T. S. Njolstad, H. Helland, R. J. Scott, K. Ashton, T. Proietto, G. Otton, I. Tomlinson, M. Gorman, K. Howarth, S. Hodgson, M. Garcia-Closas, N. Wentzensen, H. Yang, S. Chanock, P. Hall, K. Czene, J. Liu, J. Li, X. O. Shu, W. Zheng, J. Long, Y. B. Xiang, M. Shah, J. Morrison, K. Michailidou, P. D. Pharoah, A. M. Dunning, and D. F. Easton. 2011. "Genome-Wide Association Study Identifies a Common Variant Associated With Risk of Endometrial Cancer." *Nat.Genet.* 43(5):451-54. doi:ng.812 [pii];10.1038/ng.812 [doi].

StataCorp. Stata Statistical Software: Release 12. 2011. College Station, TX, StataCorp LP.
Ref Type: Generic

Stearns, F. W. 2010. "One Hundred Years of Pleiotropy: a Retrospective." *Genetics.* 186(3):767-73. doi:186/3/767 [pii];10.1534/genetics.110.122549 [doi].

Steenland, K., D. Loomis, C. Shy, and N. Simonsen. 1996. "Review of Occupational Lung Carcinogens." *Am.J.Ind.Med.* 29(5):474-90. doi:10.1002/(SICI)1097-0274(199605)29:5<474::AID-AJIM6>3.0.CO;2-M [pii];10.1002/(SICI)1097-0274(199605)29:5<474::AID-AJIM6>3.0.CO;2-M [doi].

Stevens, L. A., J. P. Palma, K. K. Pandher, and C. A. Longhurst. 2013. "Immunization Registries in the EMR Era." *Online.J.Public.Health Inform.* 5(2):211. doi:10.5210/ojphi.v5i2.4696 [doi];ojphi-05-211 [pii].

Stolk, L., J. R. Perry, D. I. Chasman, C. He, M. Mangino, P. Sulem, M. Barbalic, L. Broer, E. M. Byrne, F. Ernst, T. Esko, N. Franceschini, D. F. Gudbjartsson, J. J. Hottenga, P. Kraft, P. F. McArdle, E. Porcu, S. Y. Shin, A. V. Smith, Wingerden S. van, G. Zhai, W. V. Zhuang, E. Albrecht, B. Z. Alizadeh, T. Aspelund, S. Bandinelli, L. B. Lauc, J. S. Beckmann, M. Boban, E. Boerwinkle, F. J. Broekmans, A. Burri, H. Campbell, S. J. Chanock, C. Chen, M. C. Cornelis, T. Corre, A. D. Coviello, P. d'Adamo, G. Davies, Faire U. de, E. J. de Geus, I. J. Deary, G. V. Dedoussis, P. Deloukas, S. Ebrahim, G. Eiriksdottir, V. Emilsson, J. G. Eriksson, B. C. Fauser, L. Ferreli, L. Ferrucci, K. Fischer, A. R. Folsom, M. E. Garcia, P. Gasparini, C. Gieger, N. Glazer, D. E. Grobbee, P. Hall, T. Haller, S. E. Hankinson, M. Hass, C. Hayward, A. C. Heath, A. Hofman, E. Ingelsson, A. C. Janssens, A. D. Johnson, D. Karasik, S. L. Kardia, J. Keyzer, D. P. Kiel, I. Kolcic, Z. Kutalik, J. Lahti, S. Lai, T. Laisk, J. S. Laven, D. A. Lawlor, J. Liu, L. M. Lopez, Y. V. Louwers, P. K. Magnusson, M. Marongiu, N. G. Martin, I. M. Klaric, C. Masciullo, B. McKnight, S. E. Medland, D. Melzer, V. Mooser, P. Navarro, A. B. Newman, D. R. Nyholt, N. C. Onland-Moret, A. Palotie, G. Pare, A. N. Parker, N. L. Pedersen, P. H. Peeters, G. Pistis, A. S. Plump, O. Polasek, V. J. Pop, B. M. Psaty, K. Raikkonen, E. Rehnberg, J. I. Rotter, I. Rudan, C. Sala, A. Salumets, A. Scuteri, A. Singleton, J. A. Smith, H. Snieder, N. Soranzo, S. N. Stacey, J. M. Starr, M. G. Stathopoulou, K. Stirrups, R. P. Stolk, U. Styrkarsdottir, Y. V. Sun, A. Tenesa, B. Thorand, D. Toniolo, L. Tryggvadottir, K. Tsui, S. Ulivi, R. M. van Dam, Y. T. van der Schouw, C. H. van Gils, Nierop P. van, J. M. Vink, P. M. Visscher, M. Voorhuis, G. Waeber, H. Wallaschofski, H. E. Wichmann, E. Widen, C. J. Wijnands-van Gent, G. Willemsen, J. F. Wilson, B. H. Wolffenbuttel, A. F. Wright, L. M. Yerges-Armstrong, T. Zemunik, L. Zgaga, M. C. Zillikens, M. Zygmunt, T. L. Study, A. M. Arnold, D. I. Boomsma, J. E. Buring, L. Crisponi, E. W. Demerath, V. Gudnason, T. B. Harris, F. B. Hu, D. J. Hunter, L. J. Launer, A. Metspalu, G. W. Montgomery, B. A. Oostra, P. M. Ridker, S. Sanna, D. Schlessinger, T. D. Spector, K. Stefansson, E. A. Streeten, U. Thorsteinsdottir, M. Uda, A. G. Uitterlinden, C. M. van Duijn, H. Volzke, A. Murray, J. M. Murabito, J. A. Visser, and K. L. Lunetta. 1-22-2012. "Meta-Analyses Identify 13 Loci Associated With Age at Menopause and Highlight DNA Repair and Immune Pathways." *Nat.Genet.* doi:ng.1051 [pii];10.1038/ng.1051 [doi].

Stolk, L., G. Zhai, J. B. van Meurs, M. M. Verbiest, J. A. Visser, K. Estrada, F. Rivadeneira, F. M. Williams, L. Cherkas, P. Deloukas, N. Soranzo, J. J. de Keyzer, V. J. Pop, P. Lips, C. E.

Lebrun, Y. T. van der Schouw, D. E. Grobbee, J. Witteman, A. Hofman, H. A. Pols, J. S. Laven, T. D. Spector, and A. G. Uitterlinden. 2009. "Loci at Chromosomes 13, 19 and 20 Influence Age at Natural Menopause." *Nat.Genet.* 41(6):645-47. doi:ng.387 [pii];10.1038/ng.387 [doi].

Stratton-Loeffler, M. J., J. C. Lo, R. L. Hui, A. Coates, J. R. Minkoff, and A. Budayr. 2012. "Treatment of Vitamin D Deficiency Within a Large Integrated Health Care Delivery System." *J.Manag.Care Pharm.* 18(7):497-505. doi:2012(18)7: 497-505 [pii].

Sulem, P., D. F. Gudbjartsson, T. Rafnar, H. Holm, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, P. Alexandersen, B. Feenstra, H. A. Boyd, K. K. Aben, A. L. Verbeek, N. Roeleveld, A. Jonasdottir, U. Styrkarsdottir, V. Steinthorsdottir, A. Karason, S. N. Stacey, J. Gudmundsson, M. Jakobsdottir, G. Thorleifsson, G. Hardarson, J. Gulcher, A. Kong, L. A. Kiemeney, M. Melbye, C. Christiansen, L. Tryggvadottir, U. Thorsteinsdottir, and K. Stefansson. 2009. "Genome-Wide Association Study Identifies Sequence Variants on 6q21 Associated With Age at Menarche." *Nat.Genet.* 41(6):734-38. doi:ng.383 [pii];10.1038/ng.383 [doi].

Sun, F., W. Bruening, E. Erinoff, and K. M. Schoelles. 2011. doi:NBK56750 [bookaccession].

Sun, S. S., C. M. Schubert, W. C. Chumlea, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and A. S. Ryan. 2002. "National Estimates of the Timing of Sexual Maturation and Racial Differences Among US Children." *Pediatrics.* 110(5):911-19.

Sun, S. S., C. M. Schubert, R. Liang, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and W. C. Chumlea. 2005. "Is Sexual Maturity Occurring Earlier Among U.S. Children?" *J.Adolesc.Health* 37(5):345-55. doi:S1054-139X(05)00053-4 [pii];10.1016/j.jadohealth.2004.10.009 [doi].

Swain, M. and V. Patel.,*Patient Access to Test Results among Clinical Laboratories* .

Syurina, E. V., I. Brankovic, N. Probst-Hensch, and A. Brand. 2011. "Genome-Based Health Literacy: a New Challenge for Public Health Genomics." *Public.Health Genomics.* 14(4-5):201-10. doi:000324238 [pii];10.1159/000324238 [doi].

Talmud, P. J., J. A. Cooper, J. Palmen, R. Lovering, F. Drenos, A. D. Hingorani, and S. E. Humphries. 2008. "Chromosome 9p21.3 Coronary Heart Disease Locus Genotype and Prospective Risk of CHD in Healthy Middle-Aged Men." *Clin.Chem.* 54(3):467-74. doi:clinchem.2007.095489 [pii];10.1373/clinchem.2007.095489 [doi].

Talmud, P. J., A. D. Hingorani, J. A. Cooper, M. G. Marmot, E. J. Brunner, M. Kumari, M. Kivimaki, and S. E. Humphries. 2010. "Utility of Genetic and Non-Genetic Risk Factors in Prediction of Type 2 Diabetes: Whitehall II Prospective Cohort Study." *BMJ.* 340:b4838.

Tanner, J. M. and R. H. Whitehouse. 1976. "Clinical Longitudinal Standards for Height, Weight, Height Velocity, Weight Velocity, and Stages of Puberty." *Arch.Dis.Child.* 51(3):170-179.

Tashiro, H., C. Isacson, R. Levine, R. J. Kurman, K. R. Cho, and L. Hedrick. 1997. "P53 Gene Mutations Are Common in Uterine Serous Carcinoma and Occur Early in Their Pathogenesis." *Am.J.Pathol.* 150(1):177-85.

Taylor, C. 1994. "Gender Equity in Research." *J.Womens Health* 3(3):143-53.

Taylor, H. A., Jr., J. G. Wilson, D. W. Jones, D. F. Sarpong, A. Srinivasan, R. J. Garrison, C. Nelson, and S. B. Wyatt. 2005. "Toward Resolution of Cardiovascular Health Disparities in African Americans: Design and Methods of the Jackson Heart Study." *Ethn.Dis.* 15(4 Suppl 6):S6-17.

Taylor, P. N., V. Panicker, A. Sayers, B. Shields, A. Iqbal, A. P. Bremner, J. P. Beilby, P. J. Leedman, A. T. Hattersley, B. Vaidya, T. Frayling, J. EVANS, J. H. Tobias, N. J. Timpson, J. P. Walsh, and C. M. Dayan. 2011. "A Meta-Analysis of the Associations Between Common Variation in the PDE8B Gene and Thyroid Hormone Parameters, Including Assessment of Longitudinal Stability of Associations Over Time and Effect of Thyroid Hormone Replacement." *Eur.J.Endocrinol.* 164(5):773-80. doi:EJE-10-0938 [pii];10.1530/EJE-10-0938 [doi].

Tessari, A., D. Palmieri, and Cosimo S. Di. 2013. "Overview of Diagnostic/Targeted Treatment Combinations in Personalized Medicine for Breast Cancer Patients." *Pharmgenomics.Pers.Med.* 7:1-19. doi:10.2147/PGPM.S53304 [doi];pgpm-7-001 [pii].

Teufel, N. I. and C. K. Ritenbaugh. 1998. "Development of a Primary Prevention Program: Insight Gained in the Zuni Diabetes Prevention Program." *Clin.Pediatr.(Phila)* 37(2):131-41.

Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, W. D. Dotson, M. P. Douglas, and A. O. Berg. 2009. "The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group." *Genet.Med.* 11(1):3-14. doi:10.1097/GIM.0b013e318184137c [doi].

Thakur, M., I. Grossman, D. C. McCrory, L. A. Orlando, D. C. Steffens, K. E. Cline, R. N. Gray, J. Farmer, G. DeJesus, C. O'Brien, G. Samsa, D. B. Goldstein, and D. B. Matchar. 2007. "Review of Evidence for Genetic Testing for CYP450 Polymorphisms in Management of Patients With Nonpsychotic Depression With Selective Serotonin Reuptake Inhibitors." *Genet.Med.* 9(12):826-35. doi:10.1097GIM.0b013e31815bf98f [doi];00125817-200712000-00005 [pii].

The Ritchie Lab.,*PLATO Downloads* .

The University of Arizona Cancer Center.,*Breakthroughs: Personalized Cancer Care* .

Thomas, S. B. and S. C. Quinn. 1991. "The Tuskegee Syphilis Study, 1932 to 1972: Implications for HIV Education and AIDS Risk Education Programs in the Black Community." *Am.J.Public Health* 81(11):1498-505. doi:10.2105/AJPH.81.11.1498 [doi].

Tomaz, R. A., I. Sousa, J. G. Silva, C. Santos, M. R. Teixeira, V. Leite, and B. M. Cavaco. 2012. "FOXE1 Polymorphisms Are Associated With Familial and Sporadic Nonmedullary Thyroid Cancer Susceptibility." *Clin.Endocrinol.(Oxf.)* 77(6):926-33. doi:10.1111/j.1365-2265.2012.04505.x [doi].

Turati, F., C. Galeone, M. Rota, C. Pelucchi, E. Negri, V. Bagnardi, G. Corrao, P. Boffetta, and Vecchia C. La. 2014. "Alcohol and Liver Cancer: a Systematic Review and Meta-Analysis of Prospective Studies." *Ann.Oncol.* 25(8):1526-35. doi:mdu020 [pii];10.1093/annonc/mdu020 [doi].

Turner, S and W Bush.,*Getting Genetics Done* .

Turner, S., L. L. Armstrong, Y. Bradford, C. S. Carlson, D. C. Crawford, A. T. Crenshaw, Andrade M. de, K. F. Doheny, J. L. Haines, G. Hayes, G. Jarvik, L. Jiang, I. J. Kullo, R. Li, H. Ling, T. A. Manolio, M. Matsumoto, C. A. McCarty, A. N. McDavid, D. B. Mirel, J. E. Paschall, E. W. Pugh, L. V. Rasmussen, R. A. Wilke, R. L. Zuvich, and M. D. Ritchie. 2011. "Quality Control Procedures for Genome-Wide Association Studies." *Curr.Protoc.Hum.Genet.* Chapter 1:Unit1. doi:10.1002/0471142905.hg0119s68 [doi].

Turner, S. D., S. M. Dudek, and M. D. Ritchie. 2010. "ATHENA: A Knowledge-Based Hybrid Backpropagation-Grammatical Evolution Neural Network Algorithm for Discovering Epistasis Among Quantitative Trait Loci." *BioData.Min.* 3(1):5. doi:1756-0381-3-5 [pii];10.1186/1756-0381-3-5 [doi].

Tuskegee University Centers of Excellence Bioethics Center.,*Impact on Health Care* .

U.S.Department of Commerce.,*United States Census Bureau State & County Quick Facts* .

U.S.Preventative Services Task Force. 2008. "Screening for Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement." *Ann.Intern.Med.* 149(9):627-37.

-----.,*Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women, Topic Page* .

University of California, Irvine.,*Proactive, Personalized Cancer Medicine at UC Irvine* .

Unoki, H., A. Takahashi, T. Kawaguchi, K. Hara, M. Horikoshi, G. Andersen, D. P. Ng, J. Holmkvist, K. Borch-Johnsen, T. Jorgensen, A. Sandbaek, T. Lauritzen, T. Hansen, S. Nurbaya, T. Tsunoda, M. Kubo, T. Babazono, H. Hirose, M. Hayashi, Y. Iwamoto, A. Kashiwagi, K. Kaku, R. Kawamori, E. S. Tai, O. Pedersen, N. Kamatani, T. Kadowaki, R. Kikkawa, Y. Nakamura, and S. Maeda. 2008. "SNPs in KCNQ1 Are Associated With Susceptibility to Type 2 Diabetes in East Asian and European Populations." *Nat.Genet.* 40(9):1098-102. doi:ng.208 [pii];10.1038/ng.208 [doi].

UyBico, S. J., S. Pavel, and C. P. Gross. 2007. "Recruiting Vulnerable Populations into Research: a Systematic Review of Recruitment Interventions." *J.Gen.Intern.Med.* 22(6):852-63. doi:10.1007/s11606-007-0126-3 [doi].

van Asselt, K. M., H. S. Kok, H. Putter, C. Wijmenga, P. H. Peeters, Y. T. van der Schouw, D. E. Grobbee, E. R. Te Velde, S. Mosselman, and P. L. Pearson. 2004. "Linkage Analysis of Extremely Discordant and Concordant Sibling Pairs Identifies Quantitative Trait Loci Influencing Variation in Human Menopausal Age." *Am.J.Hum.Genet.* 74(3):444-53. doi:10.1086/382136 [doi];S0002-9297(07)61862-X [pii].

Van, Goor F., S. Hadida, P. D. Grootenhuis, B. Burton, D. Cao, T. Neuberger, A. Turnbull, A. Singh, J. Joubran, A. Hazlewood, J. Zhou, J. McCartney, V. Arumugam, C. Decker, J. Yang, C. Young, E. R. Olson, J. J. Wine, R. A. Frizzell, M. Ashlock, and P. Negulescu. 11-3-2009. "Rescue of CF Airway Epithelial Cell Function in Vitro by a CFTR Potentiator, VX-770." *Proc.Natl.Acad.Sci.U.S.A.* 106(44):18825-30. doi:0904709106 [pii];10.1073/pnas.0904709106 [doi].

Vanderbilt Ingram Cancer Center.,*Personalized Cancer Therapy* .

Vanderpump, M. P. 2011. "The Epidemiology of Thyroid Disease." *Br.Med.Bull.* 99:39-51. doi:ldr030 [pii];10.1093/bmb/ldr030 [doi].

Vassy, J. L., K. Donelan, M. F. Hivert, R. C. Green, and R. W. Grant. 2013. "Genetic Susceptibility Testing for Chronic Disease and Intention for Behavior Change in Healthy Young Adults." *J.Community Genet.* 4(2):263-71. doi:10.1007/s12687-013-0140-6 [doi].

Vernarelli, J. A., J. S. Roberts, S. Hiraki, C. A. Chen, L. A. Cupples, and R. C. Green. 2010. "Effect of Alzheimer Disease Genetic Risk Disclosure on Dietary Supplement Use." *Am.J.Clin.Nutr.* 91(5):1402-7. doi:ajcn.2009.28981 [pii];10.3945/ajcn.2009.28981 [doi].

Vidali, S., J. Knuever, J. Lerchner, M. Giesen, T. Biro, M. Klinger, B. Kofler, W. Funk, B. Poeggeler, and R. Paus. 2014. "Hypothalamic-Pituitary-Thyroid Axis Hormones Stimulate Mitochondrial Function and Biogenesis in Human Hair Follicles." *J.Invest.Dermatol.* 134(1):33-42. doi:jid2013286 [pii];10.1038/jid.2013.286 [doi].

Voight, B. F., H. M. Kang, J. Ding, C. D. Palmer, C. Sidore, P. S. Chines, N. P. Burtt, C. Fuchsberger, Y. Li, J. Erdmann, T. M. Frayling, I. M. Heid, A. U. Jackson, T. Johnson, T. O. Kilpelainen, C. M. Lindgren, A. P. Morris, I. Prokopenko, J. C. Randall, R. Saxena, N. Soranzo, E. K. Speliotes, T. M. Teslovich, E. Wheeler, J. Maguire, M. Parkin, S. Potter, N. W. Rayner, N. Robertson, K. Stirrups, W. Winckler, S. Sanna, A. Mulas, R. Nagaraja, F. Cucca, I. Barroso, P. Deloukas, R. J. Loos, S. Kathiresan, P. B. Munroe, C. Newton-Cheh, A. Pfeufer, N. J. Samani, H. Schunkert, J. N. Hirschhorn, D. Altshuler, M. I. McCarthy, G. R. Abecasis, and M. Boehnke. 2012. "The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits." *PLoS.Genet.* 8(8):e1002793. doi:10.1371/journal.pgen.1002793 [doi];PGENETICS-D-11-02644 [pii].

Voight, B. F., L. J. Scott, V. Steinthorsdottir, A. P. Morris, C. Dina, R. P. Welch, E. Zeggini, C. Huth, Y. S. Aulchenko, G. Thorleifsson, L. J. McCulloch, T. Ferreira, H. Grallert, N. Amin, G. Wu, C. J. Willer, S. Raychaudhuri, S. A. McCarroll, C. Langenberg, O. M. Hofmann, J. Dupuis, L. Qi, A. V. Segre, Hoek M. van, P. Navarro, K. Ardlie, B. Balkau, R. Benediktsson, A. J. Bennett, R. Blagieva, E. Boerwinkle, L. L. Bonnycastle, Bostrom K. Bengtsson, B. Bravenboer, S. Bumpstead, N. P. Burtt, G. Charpentier, P. S. Chines, M.

Cornelis, D. J. Couper, G. Crawford, A. S. Doney, K. S. Elliott, A. L. Elliott, M. R. Erdos, C. S. Fox, C. S. Franklin, M. Ganser, C. Gieger, N. Grarup, T. Green, S. Griffin, C. J. Groves, C. Guiducci, S. Hadjadj, N. Hassanali, C. Herder, B. Isomaa, A. U. Jackson, P. R. Johnson, T. Jorgensen, W. H. Kao, N. Klopp, A. Kong, P. Kraft, J. Kuusisto, T. Lauritzen, M. Li, A. Lieverse, C. M. Lindgren, V. Lyssenko, M. Marre, T. Meitinger, K. Midthjell, M. A. Morken, N. Narisu, P. Nilsson, K. R. Owen, F. Payne, J. R. Perry, A. K. Petersen, C. Platou, C. Proenca, I. Prokopenko, W. Rathmann, N. W. Rayner, N. R. Robertson, G. Rocheleau, M. Roden, M. J. Sampson, R. Saxena, B. M. Shields, P. Shrader, G. Sigurdsson, T. Sparso, K. Strassburger, H. M. Stringham, Q. Sun, A. J. Swift, B. Thorand, J. Tichet, T. Tuomi, R. M. van Dam, T. W. van Haeften, Herpt T. van, J. V. van Vliet-Ostaptchouk, G. B. Walters, M. N. Weedon, C. Wijmenga, J. Witteman, R. N. Bergman, S. Cauchi, F. S. Collins, A. L. Gloyn, U. Gyllensten, T. Hansen, W. A. Hide, G. A. Hitman, A. Hofman, D. J. Hunter, K. Hveem, M. Laakso, K. L. Mohlke, A. D. Morris, C. N. Palmer, P. P. Pramstaller, I. Rudan, E. Sijbrands, L. D. Stein, J. Tuomilehto, A. Uitterlinden, M. Walker, N. J. Wareham, R. M. Watanabe, G. R. Abecasis, B. O. Boehm, H. Campbell, M. J. Daly, A. T. Hattersley, F. B. Hu, J. B. Meigs, J. S. Pankow, O. Pedersen, H. E. Wichmann, I. Barroso, J. C. Florez, T. M. Frayling, L. Groop, R. Sladek, U. Thorsteinsdottir, J. F. Wilson, T. Illig, P. Froguel, C. M. van Duijn, K. Stefansson, D. Altshuler, M. Boehnke, and M. I. McCarthy. 2010. "Twelve Type 2 Diabetes Susceptibility Loci Identified Through Large-Scale Association Analysis." *Nat.Genet.* 42(7):579-89. doi:ng.609 [pii];10.1038/ng.609 [doi].

Voorhuis, M., F. J. Broekmans, B. C. Fauser, N. C. Onland-Moret, and Y. T. van der Schouw. 2011. "Genes Involved in Initial Follicle Recruitment May Be Associated With Age at Menopause." *J.Clin.Endocrinol.Metab* 96(3):E473-E479. doi:jc.2010-1799 [pii];10.1210/jc.2010-1799 [doi].

Vorderstrasse, A. A., G. S. Ginsburg, W. E. Kraus, M. C. Maldonado, and R. Q. Wolever. 2013. "Health Coaching and Genomics-Potential Avenues to Elicit Behavior Change in Those at Risk for Chronic Disease: Protocol for Personalized Medicine Effectiveness Study in Air Force Primary Care." *Glob.Adv.Health Med.* 2(3):26-38. doi:10.7453/gahmj.2013.035 [doi];gahmj.2013.035 [pii].

Wang, J. J., G. H. Buitendijk, E. Rochtchina, K. E. Lee, B. E. Klein, C. M. van Duijn, V. M. Flood, S. M. Meuer, J. Attia, C. Myers, E. G. Holliday, A. G. Tan, W. T. Smith, S. K. Iyengar, P. T. De Jong, A. Hofman, J. R. Vingerling, P. Mitchell, R. Klein, and C. C. Klaver. 2014a. "Genetic Susceptibility, Dietary Antioxidants, and Long-Term Incidence of Age-Related Macular Degeneration in Two Populations." *Ophthalmology.* 121(3):667-75. doi:S0161-6420(13)00942-1 [pii];10.1016/j.ophtha.2013.10.017 [doi].

Wang, M., M. Zhang, X. Chen, and H. Zhang. 11-1-2009. "Detecting Genes and Gene-Gene Interactions for Age-Related Macular Degeneration With a Forest-Based Approach." *Stat.Biopharm.Res.* 1(4):424-30. doi:10.1198/sbr.2009.0046 [doi].

Wang, V. M., R. B. Rosen, C. B. Meyerle, S. K. Kurup, D. Ardeljan, E. Agron, K. Tai, M. Pomykala, E. Y. Chew, C. C. Chan, and J. Tuo. 2012. "Suggestive Association Between PLA2G12A Single Nucleotide Polymorphism Rs2285714 and Response to Anti-Vascular

Endothelial Growth Factor Therapy in Patients With Exudative Age-Related Macular Degeneration." *Mol.Vis.* 18:2578-85.

Wang, Y., P. Broderick, E. Webb, X. Wu, J. Vijayakrishnan, A. Matakidou, M. Qureshi, Q. Dong, X. Gu, W. V. Chen, M. R. Spitz, T. Eisen, C. I. Amos, and R. S. Houlston. 2008. "Common 5p15.33 and 6p21.33 Variants Influence Lung Cancer Risk." *Nat.Genet.* 40(12):1407-9. doi:ng.273 [pii];10.1038/ng.273 [doi].

Wang, Y., Y. Wang, J. Li, J. Cragun, K. Hatch, S. K. Chambers, and W. Zheng. 2013. "Lynch Syndrome Related Endometrial Cancer: Clinical Significance Beyond the Endometrium." *J.Hematol.Oncol.* 6:22. doi:1756-8722-6-22 [pii];10.1186/1756-8722-6-22 [doi].

Wang, Z., B. Zhu, M. Zhang, H. Parikh, J. Jia, C. C. Chung, J. N. Sampson, J. W. Hoskins, A. Hutchinson, L. Burdette, A. Ibrahim, C. Hautman, P. S. Raj, C. C. Abnet, A. A. Adjei, A. Ahlbom, D. Albanes, N. E. Allen, C. B. Ambrosone, M. Aldrich, P. Amiano, C. Amos, U. Andersson, G. Andriole, Jr., I. L. Andrulis, C. Arici, A. A. Arslan, M. A. Austin, D. Baris, D. A. Barkauskas, B. A. Bassig, L. E. Beane Freeman, C. D. Berg, S. I. Berndt, P. A. Bertazzi, R. B. Biritwum, A. Black, W. Blot, H. Boeing, P. Boffetta, K. Bolton, M. C. Boutron-Ruault, P. M. Bracci, P. Brennan, L. A. Brinton, M. Brotzman, H. B. Bueno-de-Mesquita, J. E. Buring, M. A. Butler, Q. Cai, G. Cancel-Tassin, F. Canzian, G. Cao, N. E. Caporaso, A. Carrato, T. Carreon, A. Carta, G. C. Chang, I. S. Chang, J. Chang-Claude, X. Che, C. J. Chen, C. Y. Chen, C. H. Chen, C. Chen, K. Y. Chen, Y. M. Chen, A. P. Chokkalingam, L. W. Chu, F. Clavel-Chapelon, G. A. Colditz, J. S. Colt, D. Conti, M. B. Cook, V. K. Cortessis, E. D. Crawford, O. Cussenot, F. G. Davis, Vivo De, I, X. Deng, T. Ding, C. P. Dinney, A. L. Di Stefano, W. R. Diver, E. J. Duell, J. W. Elena, J. H. Fan, H. S. Feigelson, M. Feychting, J. D. Figueroa, A. M. Flanagan, J. F. Fraumeni, Jr., N. D. Freedman, B. L. Fridley, C. S. Fuchs, M. Gago-Dominguez, S. Gallinger, Y. T. Gao, S. M. Gapstur, M. Garcia-Closas, R. Garcia-Closas, J. M. Gastier-Foster, J. M. Gaziano, D. S. Gerhard, C. A. Giffen, G. G. Giles, E. M. Gillanders, E. L. Giovannucci, M. Goggins, N. Gokgoz, A. M. Goldstein, C. Gonzalez, R. Gorlick, M. H. Greene, M. Gross, H. B. Grossman, R. Grubb, III, J. Gu, P. Guan, C. A. Haiman, G. Hallmans, S. E. Hankinson, C. C. Harris, P. Hartge, C. Hattinger, R. B. Hayes, Q. He, L. Helman, B. E. Henderson, R. Henriksson, J. Hoffman-Bolton, C. Hohensee, E. A. Holly, Y. C. Hong, R. N. Hoover, H. D. Hosgood, III, C. F. Hsiao, A. W. Hsing, C. A. Hsiung, N. Hu, W. Hu, Z. Hu, M. S. Huang, D. J. Hunter, P. D. Inskip, H. Ito, E. J. Jacobs, K. B. Jacobs, M. Jenab, B. T. Ji, C. Johansen, M. Johansson, A. Johnson, R. Kaaks, A. M. Kamat, A. Kamineni, M. Karagas, C. Khanna, K. T. Khaw, C. Kim, I. S. Kim, Y. H. Kim, Y. C. Kim, Y. T. Kim, C. H. Kang, Y. J. Jung, C. M. Kitahara, A. P. Klein, R. Klein, M. Kogevinas, W. P. Koh, T. Kohno, L. N. Kolonel, C. Kooperberg, C. P. Kratz, V. Krogh, H. Kunitoh, R. C. Kurtz, N. Kurucu, Q. Lan, M. Lathrop, C. C. Lau, F. Lecanda, K. M. Lee, M. P. Lee, Marchand L. Le, S. P. Lerner, D. Li, L. M. Liao, W. Y. Lim, D. Lin, J. Lin, S. Lindstrom, M. S. Linet, J. Lissowska, J. Liu, B. Ljungberg, J. Lloreta, D. Lu, J. Ma, N. Malats, S. Mannisto, N. Marina, G. Mastrangelo, K. Matsuo, K. A. McGlynn, R. McKean-Cowdin, L. H. McNeill, R. R. McWilliams, B. S. Melin, P. S. Meltzer, J. E. Mensah, X. Miao, D. S. Michaud, A. M. Mondul, L. E. Moore, K. Muir, S. Niwa, S. H. Olson, N. Orr, S. Panico, J. Y. Park, A. V. Patel, A. Patino-Garcia, S. Pavanello, P. H. Peeters, B. Peplonska, U. Peters, G. M. Petersen, P. Picci, M. C. Pike, S. Porru, J. Prescott, X. Pu, M. P. Purdue, Y. L. Qiao, P.

Rajaraman, E. Riboli, H. A. Risch, R. J. Rodabough, N. Rothman, A. M. Ruder, J. S. Ryu, M. Sanson, A. Schned, F. R. Schumacher, A. G. Schwartz, K. L. Schwartz, M. Schwenn, K. Scotlandi, A. Seow, C. Serra, M. Serra, H. D. Sesso, and G. Severi. 7-15-2014b. "Imputation and Subset Based Association Analysis Across Different Cancer Types Identifies Multiple Independent Risk Loci in the TERT-CLPTM1L Region on Chromosome 5p15.33." *Hum.Mol.Genet.* doi:ddu363 [pii];10.1093/hmg/ddu363 [doi].

Warren, Andersen S., A. Trentham-Dietz, R. E. Gangnon, J. M. Hampton, H. G. Skinner, C. D. Engelman, B. E. Klein, L. J. Titus, K. M. Egan, and P. A. Newcomb. 2014. "Breast Cancer Susceptibility Loci in Association With Age at Menarche, Age at Natural Menopause and the Reproductive Lifespan." *Cancer Epidemiol.* 38(1):62-65. doi:S1877-7821(13)00187-2 [pii];10.1016/j.canep.2013.12.001 [doi].

Warren, J., J. Kennelly, D. Warren, C. R. Elley, K. C. Wai, M. Manukia, J. Davy, T. Mabotuwana, and E. Robinson. 2012. "Using the General Practice EMR for Improving Blood Pressure Medication Adherence." *Stud.Health Technol.Inform.* 178:228-34.

Warren, J. M., U. Ekelund, H. Besson, A. Mezzani, N. Geladas, and L. Vanhees. 2010. "Assessment of Physical Activity - a Review of Methodologies With Reference to Epidemiological Research: a Report of the Exercise Physiology Section of the European Association of Cardiovascular Prevention and Rehabilitation." *Eur.J.Cardiovasc.Prev.Rehabil.* 17(2):127-39. doi:10.1097/HJR.0b013e32832ed875 [doi].

Waters, E. A., J. Muff, and J. G. Hamilton. 5-15-2014. "Multifactorial Beliefs About the Role of Genetics and Behavior in Common Health Conditions: Prevalence and Associations With Participant Characteristics and Engagement in Health Behaviors." *Genet.Med.* doi:gim201449 [pii];10.1038/gim.2014.49 [doi].

Watson, M., C. Foster, R. Eeles, D. Eccles, S. Ashley, R. Davidson, J. Mackay, P. J. Morrison, P. Hopwood, and D. G. Evans. 11-15-2004. "Psychosocial Impact of Breast/Ovarian (BRCA1/2) Cancer-Predictive Genetic Testing in a UK Multi-Centre Clinical Cohort." *Br.J.Cancer* 91(10):1787-94. doi:6602207 [pii];10.1038/sj.bjc.6602207 [doi].

Watt, A., A. Cameron, L. Sturm, T. Lathlean, W. Babidge, S. Blamey, K. Facey, D. Hailey, I. Norderhaug, and G. Maddern. 2008. "Rapid Versus Full Systematic Reviews: Validity in Clinical Practice?" *ANZ.J.Surg.* 78(11):1037-40. doi:ANS4730 [pii];10.1111/j.1445-2197.2008.04730.x [doi].

Wei, Z. 2012. "Hidden Markov Models for Controlling False Discovery Rate in Genome-Wide Association Analysis." *Methods.Mol.Biol.* 802:337-44. doi:10.1007/978-1-61779-400-1_22 [doi].

Weir, B. S. and C. C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population Structure." *Evolution* 38(6):1358-70.

Weiss, G., J. H. Skurnick, L. T. Goldsmith, N. F. Santoro, and S. J. Park. 12-22-2004. "Menopause and Hypothalamic-Pituitary Sensitivity to Estrogen." *JAMA.* 292(24):2991-96. doi:292/24/2991 [pii];10.1001/jama.292.24.2991 [doi].

Wen, M. and L. Kowaleski-Jones. 2012. "Sex and Ethnic Differences in Validity of Self-Reported Adult Height, Weight and Body Mass Index." *Ethn.Dis.* 22(1):72-78.

Weyand, C. M., D. Schmidt, U. Wagner, and J. J. Goronzy. 1998. "The Influence of Sex on the Phenotype of Rheumatoid Arthritis." *Arthritis.Rheum.* 41(5):817-22. doi:10.1002/1529-0131(199805)41:5<817::AID-ART7>3.0.CO;2-S [doi].

Wiley, L. K., A. Shah, H. Xu, and W. S. Bush. 2013. "ICD-9 Tobacco Use Codes Are Effective Identifiers of Smoking Status." *J.Am.Med.Inform.Assoc.* 20(4):652-58. doi:amiajnl-2012-001557 [pii];10.1136/amiajnl-2012-001557 [doi].

Willer, C. J., Y. Li, and G. R. Abecasis. 9-1-2010. "METAL: Fast and Efficient Meta-Analysis of Genomewide Association Scans." *Bioinformatics.* 26(17):2190-2191. doi:btq340 [pii];10.1093/bioinformatics/btq340 [doi].

Williams, J. E. 2009. "Clinical Trials and Translational Research Advisory Committee Meeting." *Eliminating Disparities in Clinical Trials (EDICT) Project* 3-7-2009).

Williams, M. J. and J. L. Eberhardt. 2008. "Biological Conceptions of Race and the Motivation to Cross Racial Boundaries." *J.Pers.Soc.Psychol.* 94(6):1033-47. doi:2008-06135-008 [pii];10.1037/0022-3514.94.6.1033 [doi].

Wilson, J. M. and Y. G. Jungner. 1968. "[Principles and Practice of Mass Screening for Disease]." *Bol.Oficina Sanit.Panam.* 65(4):281-393.

Wilson, P. W., J. B. Meigs, L. Sullivan, C. S. Fox, D. M. Nathan, and R. B. D'Agostino, Sr. 5-28-2007. "Prediction of Incident Diabetes Mellitus in Middle-Aged Adults: the Framingham Offspring Study." *Arch.Intern.Med.* 167(10):1068-74. doi:167/10/1068 [pii];10.1001/archinte.167.10.1068 [doi].

Wolff, M. S. and P. J. Landrigan. 2002. "Organochlorine Chemicals and Children's Health." *J.Pediatr.* 140(1):10-13. doi:S0022-3476(02)99361-3 [pii];10.1067/mpd.2002.121690 [doi].

Wray, N. R., J. Yang, M. E. Goddard, and P. M. Visscher. 2010. "The Genetic Interpretation of Area Under the ROC Curve in Genomic Profiling." *PLoS Genet.* 6(2):e1000864. doi:10.1371/journal.pgen.1000864 [doi].

Wray, N. R., J. Yang, B. J. Hayes, A. L. Price, M. E. Goddard, and P. M. Visscher. 2013. "Pitfalls of Predicting Complex Traits From SNPs." *Nat.Rev.Genet.* 14(7):507-15. doi:nrg3457 [pii];10.1038/nrg3457 [doi].

Wright, S. 1965. "The Interpretation of Population Structure by F-Statistics With Special Regard to Systems of Mating." *Evolution* 19(3):395-420.

Wu, T., P. Mendola, and G. M. Buck. 2002. "Ethnic Differences in the Presence of Secondary Sex Characteristics and Menarche Among US Girls: the Third National Health and Nutrition Examination Survey, 1988-1994." *Pediatrics* 110(4):752-57.

Xu, H. L., W. H. Xu, Q. Cai, M. Feng, J. Long, W. Zheng, Y. B. Xiang, and X. O. Shu. 2009a. "Polymorphisms and Haplotypes in the Caspase-3, Caspase-7, and Caspase-8 Genes and Risk for Endometrial Cancer: a Population-Based, Case-Control Study in a Chinese Population." *Cancer Epidemiol.Biomarkers Prev.* 18(7):2114-22. doi:1055-9965.EPI-09-0152 [pii];10.1158/1055-9965.EPI-09-0152 [doi].

Xu, J., A. S. Kibel, J. J. Hu, A. R. Turner, K. Pruett, S. L. Zheng, J. Sun, S. D. Isaacs, K. E. Wiley, S. T. Kim, F. C. Hsu, W. Wu, F. M. Torti, P. C. Walsh, B. L. Chang, and W. B. Isaacs. 2009b. "Prostate Cancer Risk Associated Loci in African Americans." *Cancer Epidemiol.Biomarkers.Prev.* 18(7):2145-49. doi:1055-9965.EPI-09-0091 [pii];10.1158/1055-9965.EPI-09-0091 [doi].

Xu, J., X. Lin, H. Zhu, Z. Zhang, and B. Yang. 2013. "Genetic Variation of the CYP17 and Susceptibility to Endometrial Cancer: a Meta-Analysis." *Mol.Biol.Rep.* 40(8):5085-91. doi:10.1007/s11033-013-2609-0 [doi].

Xu, W. H., J. R. Long, W. Zheng, Z. X. Ruan, Q. Cai, J. R. Cheng, G. M. Zhao, Y. B. Xiang, and X. O. Shu. 2009c. "Association of Thymidylate Synthase Gene With Endometrial Cancer Risk in a Chinese Population." *Cancer Epidemiol.Biomarkers Prev.* 18(2):579-84. doi:1055-9965.EPI-08-0831 [pii];10.1158/1055-9965.EPI-08-0831 [doi].

Yang, H. P., Bosquet J. Gonzalez, Q. Li, E. A. Platz, L. A. Brinton, M. E. Sherman, J. V. Lacey, Jr., M. M. Gaudet, L. A. Burdette, J. D. Figueroa, J. G. Ciampa, J. Lissowska, B. Peplonska, S. J. Chanock, and M. Garcia-Closas. 2010. "Common Genetic Variation in the Sex Hormone Metabolic Pathway and Endometrial Cancer Risk: Pathway-Based Evaluation of Candidate Genes." *Carcinogenesis.* 31(5):827-33. doi:bgp328 [pii];10.1093/carcin/bgp328 [doi].

Yang, Y., D. M. Muzny, J. G. Reid, M. N. Bainbridge, A. Willis, P. A. Ward, A. Braxton, J. Beuten, F. Xia, Z. Niu, M. Hardison, R. Person, M. R. Bekheirnia, M. S. Leduc, A. Kirby, P. Pham, J. Scull, M. Wang, Y. Ding, S. E. Plon, J. R. Lupski, A. L. Beaudet, R. A. Gibbs, and C. M. Eng. 10-17-2013. "Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders." *N.Engl.J.Med.* 369(16):1502-11. doi:10.1056/NEJMoa1306555 [doi].

Yasuda, K., K. Miyake, Y. Horikawa, K. Hara, H. Osawa, H. Furuta, Y. Hirota, H. Mori, A. Jonsson, Y. Sato, K. Yamagata, Y. Hinokio, H. Y. Wang, T. Tanahashi, N. Nakamura, Y. Oka, N. Iwasaki, Y. Iwamoto, Y. Yamada, Y. Seino, H. Maegawa, A. Kashiwagi, J. Takeda, E. Maeda, H. D. Shin, Y. M. Cho, K. S. Park, H. K. Lee, M. C. Ng, R. C. Ma, W. Y. So, J. C. Chan, V. Lyssenko, T. Tuomi, P. Nilsson, L. Groop, N. Kamatani, A. Sekine, Y. Nakamura, K. Yamamoto, T. Yoshida, K. Tokunaga, M. Itakura, H. Makino, K. Nanjo, T. Kadowaki, and M. Kasuga. 2008. "Variants in KCNQ1 Are Associated With Susceptibility to Type 2 Diabetes Mellitus." *Nat.Genet.* 40(9):1092-97. doi:ng.207 [pii];10.1038/ng.207 [doi].

You, J. H., K. K. Tsui, R. S. Wong, and G. Cheng. 2012. "Cost-Effectiveness of Dabigatran Versus Genotype-Guided Management of Warfarin Therapy for Stroke Prevention in Patients With Atrial Fibrillation." *PLoS One* 7(6):e39640. doi:10.1371/journal.pone.0039640 [doi];PONE-D-12-07526 [pii].

Zablocki, R. W., A. J. Schork, R. A. Levine, O. A. Andreassen, A. M. Dale, and W. K. Thompson. 8-1-2014. "Covariate-Modulated Local False Discovery Rate for Genome-Wide Association Studies." *Bioinformatics.* 30(15):2098-104. doi:btu145 [pii];10.1093/bioinformatics/btu145 [doi].

Zakharia, F., A. Basu, D. Absher, T. L. Assimes, A. S. Go, M. A. Hlatky, C. Iribarren, J. W. Knowles, J. Li, B. Narasimhan, S. Sidney, A. Southwick, R. M. MYERS, T. Quertermous, N. Risch, and H. Tang. 2009. "Characterizing the Admixed African Ancestry of African Americans." *Genome.Biol.* 10(12):R141. doi:gb-2009-10-12-r141 [pii];10.1186/gb-2009-10-12-r141 [doi].

Zhang, B., Y. Li, L. Li, M. Chen, C. Zhang, X. B. Zuo, F. S. Zhou, B. Liang, J. Zhu, P. Li, Z. L. Huang, H. Xuan, W. Li, and Z. D. Chen. 2014. "Association Study of Susceptibility Loci With Specific Breast Cancer Subtypes in Chinese Women." *Breast.Cancer Res.Treat.* 146(3):503-14. doi:10.1007/s10549-014-3041-4 [doi].

Zhang, K., S. Pollack, A. Ghods, C. Dicken, B. Isaac, G. Adel, G. Zeitlian, and N. Santoro. 2008. "Onset of Ovulation After Menarche in Girls: a Longitudinal Study." *J.Clin.Endocrinol.Metab.* 93(4):1186-94. doi:jc.2007-1846 [pii];10.1210/jc.2007-1846 [doi].

Zhang, Q., Q. Long, and J. Ott. 2014. "AprioriGWAS, a New Pattern Mining Strategy for Detecting Genetic Variants Associated With Disease Through Interaction Effects." *PLoS Comput.Biol.* 10(6):e1003627. doi:10.1371/journal.pcbi.1003627 [doi];PCOMPBIOL-D-13-00743 [pii].

Zhao, F., Y. Xu, K. Yang, M. Liu, D. Wei, Y. Zhang, X. Shi, F. Yang, X. Wang, S. Liang, C. Zhao, X. Chen, L. Sun, X. Zhu, N. Wang, J. Hui, Y. Zhang, L. Zhu, Y. Yang, L. Tang, J. Wang, and Z. Yang. 2014. "THADA Gene Polymorphism and Prostate Cancer Risk: a Meta-Analysis." *Oncol.Res.Treat.* 37(3):106-10. doi:000360206 [pii];10.1159/000360206 [doi].

Zhao, J., D. H. Xiong, Y. Guo, T. L. Yang, R. R. Recker, and H. W. Deng. 2007. "Polymorphism in the Insulin-Like Growth Factor 1 Gene Is Associated With Age at Menarche in Caucasian Females." *Hum.Reprod.* 22(6):1789-94. doi:dem052 [pii];10.1093/humrep/dem052 [doi].

Zhu, B. T. and A. H. Conney. 1998. "Functional Role of Estrogen Metabolism in Target Cells: Review and Perspectives." *Carcinogenesis.* 19(1):1-27.

Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander. 1-24-2012. "The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability." *Proc.Natl.Acad.Sci.U.S.A.* 109(4):1193-98. doi:1119675109 [pii];10.1073/pnas.1119675109 [doi].

Zuvich, R. L., L. L. Armstrong, S. J. Bielinski, Y. Bradford, C. S. Carlson, D. C. Crawford, A. T. Crenshaw, Andrade M. de, K. F. Doheny, J. L. Haines, M. G. Hayes, G. P. Jarvik, L. Jiang, I. J. Kullo, R. Li, H. Ling, T. A. Manolio, M. E. Matsumoto, C. A. McCarty, A. N. McDavid, D. B. Mirel, L. M. Olson, J. E. Paschall, E. W. Pugh, L. V. Rasmussen, L. J. Rasmussen-Torvik, S. D. Turner, R. A. Wilke, and M. D. Ritchie. 2011. "Pitfalls of Merging GWAS Data: Lessons Learned in the EMERGE Network and Quality Control

Procedures to Maintain High Data Quality." *Genet.Epidemiol.* 35(8):887-98. doi:10.1002/gepi.20639 [doi].