Pleiotropy: Epidemiologic analyses and implications for return of results

Jonathan Mielke Kocarnik

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Ulrike Peters, Chair

Stephanie Malia Fullerton

Carolyn M Hutter

Program Authorized to Offer Degree:

School of Public Health

University of Washington

Abstract

Pleiotropy: Epidemiologic analyses and implications for return of results

Jonathan Mielke Kocarnik

Chair of the Supervisory Committee:
Ulrike Peters, Research Associate Professor
Department of Epidemiology

Pleiotropy refers to a single gene with multiple phenotypic effects. Genetic variants associated with multiple diseases can point to shared biologic mechanisms or other similarities. The first half of this dissertation comprises two epidemiologic analyses evaluating common genetic variants for pleiotropic associations. The first investigates whether levels of the inflammatory biomarker C-reactive protein are associated with genetic variants previously associated with various inflammation-related diseases such as stroke, obesity, or type 2 diabetes. The second evaluates whether genetic variants previously associated with various cancers are also associated with melanoma. Results from both analyses provide evidence supporting both known and potentially novel pleiotropic associations. The second half of this dissertation focuses on how the presence of pleiotropy impacts the return of individual genetic results to research

participants. Researchers are generating increasingly large amounts of genetic information on research participants, uncovering genetic variants with a spectrum of health implications. While many genetic variants are of little or unknown clinical significance, some genetic variants are highly predictive of disease, and can inform screening or other actions that can prevent or reduce disease. A few variants are sufficiently important to suggest an ethical duty on the part of the researcher to return that genetic information to a participant. Conversely, other variants may impart information on risk that is not actionable, and returning such information may inflict more harms than benefits. Recent guidelines have proposed various criteria for prioritizing which results are appropriate or not appropriate for return, focusing on weighing the clinical validity and clinical utility of a particular genetic result. However, these guidelines do not currently take into account that genetic variants may have multiple meanings, which might have contradicting recommendations for whether that same variant should be returned. This dissertation reviews the concept and extent of pleiotropy, reviews current recommendations for returning results, and explores the ways in which the existence of pleiotropic variants impacts these guidelines. This dissertation proposes a potential framework for considering pleiotropic relationships in return recommendations, and suggests that future guidelines will need to account for pleiotropy if they are to be effective.

Acknowledgements

I am extremely thankful of the many people who provided assistance with this project, in ways great and small, direct and indirect. Their help was most appreciated, and I extend a sincere thank you to the following people:

- Ulrike Peters, for taking me in when I began this program, including me in some extremely interesting projects, and mentoring me through many years of questions, drafts, presentations, and meetings;
- Stephanie Malia Fullerton, for engaging in countless discussions and debates with me on ethical issues of all kinds, and for pushing my growth in all things ELSI;
- Carolyn Hutter, for her patience and willingness to answer any questions I might have on epidemiology, biostatistics, Stata, working in research, or other minutia;
- Holly Tabor, for serving as the Graduate Student Representative on my committee, and offering up her insight and expertise in the area of returning results;
- Emily White, for providing me funding support for four years of training, coursework, conferences, and other resources on her Cancer Prevention Training Grant;
- Charles Kooperberg and Ulrike Peters, for funding support and for including me in PAGE, which has been a wonderful experience;
- The many friends, professors, and colleagues I have met through the Institute for Public Health Genetics at the University of Washington, some of the most committed, interdisciplinary, and interesting individuals around;
- The faculty, students and staff at the Fred Hutchinson Cancer Research Center, which have made it a truly great place to do research and training;

- The many collaborators I've have had the privilege of working with, both in the Population Architecture using Genomics and Epidemiology study (PAGE) and the Genetics of Colorectal Cancer Consortium (GECCO), who together demonstrate the power and benefits of interdisciplinary collaboration;
- My parents, siblings, family, and friends, who have supported me in so many ways throughout the years;
- My wife Beverly, for everything, including our son Walter, who brings increasing amounts of joy into our lives.
- The many other people who have helped me reach this point, and continue to urge me onwards, to whom I am humbly grateful.

Pleiotropy: epidemiologic analyses and implications for return of results

Table of Contents:

Introduction p. 1 – 2

Specific aims p. 3

Part A – Pleiotropy analyses

I. Analysis I - Inflammation
   a. Manuscript p. 4 – 22
   b. References p. 22 – 26
   c. Tables p. 27 – 32

II. Analysis II – Melanoma
   a. Manuscript p. 33 – 46
   b. References p. 47 – 50
   c. Tables and Figures p. 50 – 53

Part B – Impact of pleiotropy on the return of results to research participants

I. Abstract p. 54 – 55
II. Introduction p. 56 – 58
III. Background on incidental findings and return of results p. 58 – 62
IV. Description and estimation of pleiotropy p. 62 – 68
V. Implications of pleiotropy on return of results p. 69 – 76
VI. Framework and policy recommendations p. 76 – 82
VII. Other implications and Conclusion p. 82 – 84
VIII. References p. 85 – 87
IX. Tables and Figures p. 88 – 92

Appendices

Inflammation Appendices 1 – 4 p. 93 – 106

Melanoma Appendices 1 – 4 p. 107 – 120

Introduction

Pleiotropy is broadly defined as a single mutation, gene, or locus affecting multiple phenotypic traits or disorders. Pleiotropy is also the overarching theme of this dissertation, and is explored in an interdisciplinary fashion. The first half of this dissertation applies the concept of pleiotropy to genetic epidemiology, conducting two separate studies evaluating genetic variants for pleiotropic effects. The second half of this dissertation applies the concept of pleiotropy to the return of result discussion, providing several suggestions for how current guidelines will need to be altered so as appropriately make recommendations for pleiotropic results with multiple meanings.

Genome-wide association studies (GWAS) have been highly successful in identifying associations between single nucleotide polymorphisms (SNPs) and various phenotypes of interest. Building on this success, researchers are employing strategies to generate additional information beyond baseline GWAS data, and search for additional factors that can account for the missing heritability of many phenotypes. One such strategy involves searching for pleiotropic effects among GWAS findings. By evaluating whether genetic variation important for known phenotypes are also associated with other phenotypes, researchers seek to generate novel biological insights beyond the single-gene single-phenotype relationship. Pleiotropic variants may provide insights into biological pathways shared between similar or dissimilar phenotypes, and enhance the understanding of disease processes. Such analyses may additionally lead to the identification of novel loci, contributing to a fuller understanding of the genetic contributions to disease. This dissertation performs two genetic epidemiologic analyses to evaluate genetic variants for pleiotropic effects.

Once identified through such studies, pleiotropic relationships can impact the social science realm as well. The question on whether to return genetic results, particularly in the research setting, is an increasingly important topic. An important aspect of this debate is determining which of the many possible genetic results are appropriate to return to individual participants. While several guidelines have been developed to help inform the threshold for which genetic results to return, this guidance often relies upon single genotype-phenotype evaluations. However, the existence of pleiotropy majorly impacts this discussion, since a single result may have multiple meanings and consequences beyond that particular relationship. Some of these consequences may favor return, while other related or unrelated consequences may not. As such, any current recommendations which rely on single-genotype single-phenotype associations will likely be incomplete without considering current and future pleiotropic information as well. The continued generation of pleiotropic knowledge suggests that pleiotropy will be an increasingly important consideration in this debate. To this end, this dissertation explores the many ways in which the presence of pleiotropy may have implications for determining which results are appropriate to return.

Specific Aims

The goals of this dissertation are 1) to conduct two genetic epidemiologic analysis projects evaluating the potential pleiotropic effects of known GWAS SNPs with related phenotypes of interest, and 2) to evaluate how the presence of pleiotropy impacts the discussion on which genetic results to return to research participants. As such, this dissertation proposes the following specific aims:

1 – Perform pleiotropy analysis projects within the PAGE consortium.

a. Evaluate for evidence of pleiotropy for SNPs previously associated with cardiovascular disease phenotypes such as type 2 diabetes, obesity, stroke, and serum lipids, by assessing their association with levels of serum C-reactive protein.
b. Evaluate for evidence of pleiotropy for SNPs previously associated with other cancers, by assessing their association with melanoma.

2 – Evaluate the impact of pleiotropy on the return of results discussion.

a. Review the return of results literature in order to identify current policy guidelines and recommendations
b. Review the pleiotropy literature in order to estimate the amount of pleiotropy which may be expected given current knowledge
c. Perform a critical review of how the presence of pleiotropy impacts current return of results guidelines and recommendations

Part A, Analysis I - Inflammation

Title:

Multi-ancestral analysis of inflammation-related genetic variants and C-reactive protein in the Population Architecture using Genomics and Epidemiology (PAGE) Study

Authors:

Jonathan M Kocarnik[1], Sarah A Pendergrass[2], Cara L Carty[1], James S Pankow[3], Fredrick R Schumacher[4], Iona Cheng[5], Peter Durda[6], José Luis Ambite[8], Ewa Deelman[8], Nancy R Cook[9], Simin Liu[10], Jean Wactawski-Wende[11], Carolyn Hutter[12], Kristin Brown-Gentry[13], Sarah Wilson[13], Nathan Pankratz[14], Ching-Ping Hong[3], Shelley A Cole[15], V Saroja Voruganti[15], Petra Bůžková[16], Neal W Jorgensen[16], Nancy S Jenny[6], Lynne R Wilkens[17], Christopher A Haiman[4], Laurence N Kolonel[17], Andrea LaCroix[1], Kari North[18], Rebecca Jackson[19], Loic Le Marchand[17], Lucia A Hindorff[7], Dana C Crawford[13], Myron Gross[20], Ulrike Peters[1]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.
[2]Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA.
[3]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.
[4]Keck School of Medicine, University of Southern California, Los Angeles, CA.
[5]Cancer Prevention Institute of California, Fremont, CA, USA
[6]Department of Pathology, University of Vermont, Burlington, VT.
[7]Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD.
[8]Information Sciences Institute, University of Southern California, Marina del Rey, CA.
[9]Department of Medicine, Brigham & Women's Hospital, Boston, MA.
[10]Department of Epidemiology, University of California Los Angeles, Los Angeles, CA.
[11]Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY.
[12]Epidemiology and Genomics Research Program, DCCPS, NCI, Bethesda, MD.
[13]Center for Human Genetics Research, Vanderbilt University, Nashville, TN.
[14]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.
[15]Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.
[16]Department of Biostatistics, University of Washington, Seattle, WA.
[17]Cancer Research Center, University of Hawaii, Honolulu, HI.
[18]Department of Epidemiology, University of North Carolina, Chapel Hill, NC.
[19]Department of Internal Medicine, The Ohio State University, Columbus, OH.
[20]Masonic Cancer Center, University of Minnesota, Minneapolis, MN.

Funding / Acknowledgments:

(a) 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), and their respective NHGRI ARRA supplements. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at http://www.pagestudy.org.

(b) The data and materials included in this report result from a collaboration between the following studies:

The "Epidemiologic Architecture for Genes Linked to Environment (EAGLE)" is funded through the NHGRI PAGE program (U01HG004798-01 and its NHGRI ARRA supplement). The study participants derive from the National Health and Nutrition Examination Surveys (NHANES), and these studies are supported by the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (R37CA54281, R01 CA63, P01CA33619, U01CA136792, and U01CA98758).

Funding support for the “Epidemiology of putative genetic variants: The Women’s Health Initiative” study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/ WHI_investigators_shortlist.pdf.

Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01HG004803 and its NHGRI ARRA supplement). The following studies contributed to this manuscript and are funded by the following agencies: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following National Institutes of Health, National Heart, Lung and Blood Institute contracts: N01-HC-95095; N01-HC-48047; N01-HC-48048; N01-HC-48049; N01-HC-48050; N01-HC-45134; N01-HC-05187; and N01-HC-45205. The Cardiovascular Health Study (CHS) is supported by contracts HHSN268201200036C, N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-

55222, N01-HC-75150, N01-HC-45133, and grant HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute on Aging (NIA). The Strong Heart Study (SHS) is supported by NHLBI grants U01 HL65520, U01 HL41642, U01 HL41652, U01 HL41654, and U01 HL65521. The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service.

Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01 and its NHGRI ARRA supplement). The National Institutes of Mental Health also contributes to the support for the Coordinating Center.

Funding for work by Jonathan Kocarnik was supported by grant R25CA94880 from NCI. Funding for performing CRP assays in a portion of WHI subjects was provided by NHLBI BAA award HHSN268200960011C.

The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions.

Abstract:

Introduction: C-reactive protein (CRP) is a marker of inflammation. Genome-wide association studies (GWAS), mostly conducted in European-descent populations, have identified several single nucleotide polymorphisms (SNPs) associated with CRP concentrations. We aimed to replicate these previous findings in European Americans as well as assess if these associations generalize to four additional race/ethnicity groups: African American, American Indian, Asian/Pacific Islander and Hispanic. We also evaluated whether SNPs previously associated with other inflammation-related phenotypes or outcomes demonstrate a pleiotropic association with CRP.

Methods: We analyzed 266 SNPs among 40,473 participants from 7 studies collaborating in the Population Architecture using Genetics and Epidemiology (PAGE) study. SNPs were selected

based on previous GWAS findings for CRP and inflammation-related traits, such as cardiovascular disease, type 2 diabetes, or obesity. Linear regression models were used to estimate the association between each SNP and high-sensitivity CRP in each study, which were then combined using fixed-effect meta-analysis.

Results: Across all race/ethnicities, we identified 18 SNPs in 8 loci significantly associated with serum CRP concentrations using a Bonferroni-corrected p-value cutoff of 1.9×10^{-4}. These included: 5 SNPs in the *CRP* locus; 4 in *APOE/APOC1/TOMM40*; 2 each in *GCKR, HNF4A,* and *CELSR2/PSRC1/SORT1*; and 1 each in *HNF1A, IL6R,* and *LEPR*. The two SNPs in the *CELSR2/PSRC1/SORT1* locus had not previously been associated with CRP: rs599839 (Overall $p = 9.2 \times 10^{-7}$) and rs646776 (Overall $p = 1.65 \times 10^{-5}$). In race/ethnicity-stratified analyses, 6 SNPs in 3 loci demonstrated a statistically significant association with CRP in more than one racial/ethnic group: 3 SNPs in the *APOE/APOC1/TOMM40* locus, 2 SNPs in *CRP*, and 1 in *IL6R*. We observed 4 SNPs in which the association in European Americans generalized to African Americans and 5 SNPs that generalized to Hispanics. We also observed 1 SNP that only demonstrated an association in African Americans.

Conclusion: We replicated several findings and generalized some of them to non-European-descent populations. Our data also suggest that two potentially pleiotropic SNPs in the *CELSR2/PSRC1/SORT1* locus, previously associated with coronary artery disease and LDL cholesterol, are also associated with CRP. Our findings demonstrate the benefit of evaluating genotype-phenotype associations in multiple race/ethnicity groups, and of looking for pleiotropic relationships among SNPs previously associated with related phenotypes.

Introduction:

C-reactive protein (CRP) is an acute-phase reactant protein produced by the liver. Circulating levels rise sharply following inflammatory stimulation from infection or injury, then fall rapidly following stimulus resolution. In situations of chronic underlying disease, however, CRP remains slightly raised over time, serving as a biomarker to characterize systemic inflammation (1). Elevated CRP levels have been associated with a large number of outcomes and traits, such as cardiovascular events (1), atherosclerosis (2), stroke (3-5), type 2 diabetes (6), metabolic syndrome (7-9), fitness level and body composition (10, 11), and cancer (12, 13). As such, serum CRP is an important biomarker for the development of disease among apparently healthy individuals (1).

Heritability estimates for CRP are between 25 and 56% (14-16). Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in 21 loci associated with CRP concentrations (17-19). Environmental factors associated with increased CRP levels include female sex (14, 20), hormone replacement therapy (HRT) use (21), and obesity (1, 22), while factors associated with decreased CRP levels include exercise (23) and the use of aspirin, NSAIDs, and statins (1).

CRP levels have previously been shown to vary according to race/ethnicity. Compared to those of European ancestry, those of African ancestry often have elevated CRP levels (20, 24) and those of Asian ancestry often have lower CRP levels (25, 26). Many initial genetic studies of CRP were performed primarily on populations of European descent, and studies examining CRP-associated variants in other race/ethnicity populations have had moderate success in generalizing some, but not all, of these loci (24, 27-29). Generalization and fine-mapping of loci

in non-European ancestry populations is thus an important step in further elucidating genetic contributions to CRP concentrations in multiple race/ethnicity groups.

Many of the genetic variants associated with serum CRP levels demonstrate pleiotropic effects, and have also been associated with various other outcomes related to inflammation. SNP rs1205 in the CRP gene (*CRP*), for example, has been associated not only with CRP (17, 27), but also with heart rate variability (30) and colon cancer (12). Given the multiple phenotypes associated with CRP, we hypothesize that genetic variants associated with inflammation-related phenotypes, such as cardiovascular disease, type 2 diabetes, or obesity, might also be associated with CRP. Identifying additional pleiotropic associations could aid in elucidating shared biological pathways and relationships.

We seek to replicate previously reported associations between SNPs and serum CRP, as well as generalize these associations to four additional race/ethnicity groups: African Americans, Hispanics, Asian/Pacific Islanders, and American Indians. We also evaluate several SNPs previously associated with inflammation-related phenotypes for an association with CRP, both overall and stratified by race/ethnicity groups. The size and diversity of our study population provides an opportunity to replicate previous findings, generalize existing associations to additional race/ethnicity groups, and discover novel pleiotropic associations with this inflammatory biomarker.

Methods:

Study description

The data for this analysis were generated as part of the Population Architecture using Genomics and Epidemiology study (PAGE), which has been previously described in detail (31).

Briefly, PAGE is a consortium of large, well-characterized population-based studies, investigating the epidemiologic architecture of genetic variants associated with complex diseases across various race/ethnicity groups.

Study populations

We included 40,473 participants in this study: 7,228 from Epidemiologic Architecture for Genes Linked to Environment (EAGLE), based on data from three National Health and Nutrition Examination Surveys (NHANES); 1,593 from the Multiethnic Cohort Study (MEC); 9,901 from the Women's Health Initiative (WHI); and 21,751 from Causal Variants Across the Life Course (CALiCo), itself a consortium of five cohort studies. Participants from CALiCo included: 11,303 from Atherosclerosis Risk in Communities (ARIC); 3,249 from Coronary Artery Risk Development in Young Adults (CARDIA); 4,070 from the Cardiovascular Health Study (CHS); and 3,129 from the Strong Heart Study (SHS).

Studies included one or more of five race/ethnicity groups, totaling 24,958 European American, 8,471 African American, 2,935 Hispanic, 980 Asian/Pacific Islander, and 3,129 American Indian participants. Several of these studies have previously been utilized to evaluate genetic associations with CRP, including participants in ARIC (18), CHS (18, 32, 33), and CARDIA (34). Within EAGLE, we included NHANES 1999-2002 data, while NHANES III data were excluded due to non-high-sensitivity CRP measurements. Each study was approved by a local institutional review board, and all participants gave informed consent.

Measurements

Baseline high-sensitivity C-reactive protein (hsCRP) measurements were available in each study as part of previous investigations, with assay method and instrument used to measure

CRP varying by study (see Appendix 1). Because the distribution of CRP is skewed, the values were natural-log transformed for analyses. Demographic and other epidemiologic information was obtained according to the enrollment protocols of each study.

SNP Selection and Genotyping

Using information available in the literature in 2008, investigators in the PAGE project identified 266 SNPs previously associated with various phenotypes of interest, focusing on variants related to cardiovascular disease traits, lipids, body mass index (BMI), and type 2 diabetes. These included 16 SNPs associated with CRP/inflammation, 21 with cardiovascular disease or myocardial infarction, 26 with BMI/obesity, 51 with type 2 diabetes or glucose levels, 82 with HDL/LDL/total cholesterol, 29 with triglycerides, 9 with stroke, and 33 with other related phenotypes (see Appendix 2; previous trait association refers to the first reported association at the time of SNP selection). Genotyping methods for PAGE have been previously described (31). For each SNP, quality control thresholds included SNP and sample call rates > 90%, concordance of blinded replicates > 98%, and no clear evidence of Hardy-Weinberg disequilibrium ($p > 0.001$).

Due to funding constraints, each PAGE study chose a custom subset of SNPs to genotype depending on the phenotypes they had available. As such, the number of SNPs available varied in each study, with 30 SNPs genotyped in ARIC, 7 in CARDIA, 113 in CHS, 196 in EAGLE, 103 in MEC, 19 in SHS, and 94 in WHI (see Table 2 and Appendix 2). While this meant that not all SNPs were genotyped in every study, an effort was made to ensure overlap (56% of SNPs were genotyped in more than one study). Since race/ethnicity group availability differed by study, this also meant that not all SNPs were available for analyses in each race/ethnicity group.

Nearly all SNPs were available for the European American (n = 265), African American (n = 266), and Hispanic groups (n = 261), while fewer SNPs were available for the Asian/Pacific Islander (n = 150) and American Indian groups (n = 19). These overlap issues reduced the overall sample size available for any given SNP- and race/ethnicity-specific association, though numbers were still large for most analyses (see Table 2 and Appendix 3). To avoid small sample size issues in the analysis, for each SNP only study-specific race/ethnicity-specific results with at least 100 genotyped subjects were included in meta-analyses. This restriction meant the exclusion of the WHI American Indian results (≤ 86 participants for each SNP), which reduced the overall number of SNPs available for analysis in the American Indian group from 96 to 19.

Statistical analysis

Within each study, linear regression models were used to test the association between natural log-transformed CRP and each SNP of interest, coded additively (with 0, 1, and 2 representing the number of copies of the coded allele). Models were run separately for each of the five race/ethnicity groups: European American, African American, Hispanic, Asian/Pacific Islander, and American Indian. MEC performed analyses separately for Hawaiian and Japanese ancestry participants. Race/ethnicity-specific models were adjusted for age, sex, center, and/or principal components, as appropriate for each study. In WHI, sampling weights were used in the regression analyses to account for sample selection criteria. Additional models were stratified by sex.

Inverse-variance weighted fixed effect meta-analyses were used to estimate both the overall and the race/ethnicity stratified associations for each SNP. Because studies differed in SNPs genotyped and race/ethnicity of participants, the overall number of observations available

for each SNP-specific meta-analysis varied. Overall combined-race/ethnicity meta-analysis numbers were large for most SNPs (mean number of observations 12,361, range 802 – 36,299). In the race/ethnicity-stratified meta-analyses, the average number of observations across available SNPs was 7,722 for European Americans, 2,367 for African Americans, 1,766 for Hispanics, 617 for Asian/Pacific Islanders, and 3,129 for American Indians (see Table 2 and Appendix 3). Additional meta-analyses were performed stratified by sex. Meta-regression was used to evaluate the heterogeneity between race/ethnicity groups for a given SNP, as well as between males and females. Meta-analyses were performed using Stata version 12 (35). We used a conservative Bonferroni-corrected p-value to adjust for multiple comparisons, accounting for the maximum number of tested SNPs ($0.05/266$ SNPs $= 1.9 \times 10^{-4}$) to determine the statistical significance threshold. Since not all SNPs were available in each race/ethnicity group, this correction may have been overly stringent for some groups.

Results:

European Americans and African Americans represented the largest race/ethnicity groups in our analysis (see Table 1). Median CRP concentration varied between study population and race/ethnicity: the lowest levels were observed in MEC and Asian/Pacific Islanders, and the highest levels were observed in WHI and American Indians. Mean age was lowest in CARDIA and highest in CHS. Overall there were more women than men (64.8%) due to the focus on women in WHI. Mean BMI values varied by race/ethnicity, with the highest values in African Americans and the lowest values in Asian/Pacific Islanders. Extensive efforts were made to harmonize variables across PAGE studies, though some study heterogeneity may remain due to differences in study populations.

Among the 266 SNPs evaluated for a potentially pleiotropic association with CRP in the race/ethnicity-combined analyses, we identified 18 SNPs in 8 loci associated with CRP concentrations at a p-value less than 1.9×10^{-4} (see Table 3). These included: 5 SNPs in the *CRP* locus; 4 in *APOE/APOC1/TOMM40*; 2 each in *GCKR, HNF4A,* and *CELSR2/PSRC1/SORT1*; and 1 each in *HNF1A, IL6R,* and *LEPR*. Several of these SNPs were correlated with each other, with the amount of correlation varying by race/ethnicity (see Appendix 4**)**. Of the 18 SNPs (in 8 loci) reaching statistical significance, 15 SNPs (in 7 loci) had previously been associated with CRP, while rs599839 (near *PSRC1*) and rs646776 (near *CELSR2*) have not. The SNP rs6857 (in *PVRL2*) has not itself been previously associated with CRP, though it is in close proximity to the known *APOE/APOC1/TOMM40* locus.

In race/ethnicity-stratified analyses, 16 SNPs reached statistical significance in European Americans, along with 5 SNPs each in African Americans and Hispanics. No SNPs reached statistical significance among the Asian/Pacific Islander or American Indian groups, though not all SNPs were evaluated and power was reduced in these smaller groups. Notably, for the 18 SNPs that were significant in the race/ethnicity-combined analysis, only 9 were available in the Asian/Pacific Islander group and none were available in the American Indian group.

For generalizability, 6 SNPs in 3 loci demonstrated a statistically significant association with CRP in more than one race/ethnicity group: 3 SNPs in the *APOE/APOC1/TOMM40* locus, 2 SNPs in *CRP*, and 1 in *IL6R* (see Table 3). For the three SNPs where a statistically significant effect was seen in European Americans, African Americans, and Hispanics, two showed similar effect size estimates across race/ethnicity groups: SNP rs2228145 in *IL6R* (β = -0.10, -0.13, -0.18, respectively; p-het = 0.16), and SNP rs429358 in *APOE* (β = -0.24, -0.23, -0.36, respectively; p-het = 0.23). Two SNPs also followed this trend in European Americans and

Hispanics: SNP rs2075650 in *TOMM40* (β = -0.22 and -0.21, respectively; p-het = 0.76), and SNP rs6857 in *PVRL2* (β = -0.23 and -0.28, respectively; p-het = 0.50). Two SNPs in *CRP* showed a potential exception to this trend, where a larger effect was seen for African Americans than for European Americans (rs1800947, β = -0.61 vs. -0.30; p-het = 0.02) or than for European Americans and Hispanics (rs1205, β = -0.27 vs. -0.17 and -0.22, respectively; p-het = 0.02). In general, however, race/ethnicity-specific effect estimates for a given SNP were in the same direction and of similar magnitude across race/ethnicity groups regardless of the statistical significance of the association. We did observe several exceptions to this trend, where the effect estimate for one group was null or in the opposite direction compared to all other groups, though these exceptions could be due to chance.

We also evaluated whether any of these associations differed by sex (see Table 4). Most of the SNPs which were statistically significant overall remained so in sex-stratified analyses. Two SNPs in *GCKR*, rs1260326 and rs780094, demonstrated a statistically significant difference in effect between males and females (p-het = 1.54×10^{-8}), with a stronger effect seen in females than males. Several other SNPs were also suggestive for a potential difference in effect by sex: rs1417938 in *CRP* (p-het = 0.03), rs4420638 and rs429358 in *APOE/APOC1/TOMM40* (p-het = 0.03), and rs2650000 in *HNF1A* (p-het = 0.01). For these SNPs, the direction of the effect was generally similar in both sexes, while effect estimates were slightly larger and p-values were smaller among the females than the males, as expected given the larger sample size. For the two SNPs in *GCKR*, the p-values were actually slightly smaller in the female-stratified results than in the overall results, despite almost half the sample size.

Discussion

Our multi-ethnic meta-analysis observed 18 SNPs in 8 loci statistically significantly associated with serum CRP concentrations at $p < 1.9x10^{-4}$. Six of these SNPs in 3 loci demonstrated statistically significant associations in multiple ancestral groups. Five SNPs in 2 loci had not previously been associated with serum CRP concentration in non-European American populations. Among the 18 significant SNPs, 2 SNPs (rs599839 and rs646776, CEU $r^2 = 0.9$) in 1 locus (*CELSR2/PSRC1/SORT1*) had not previously been associated with serum CRP levels. Three themes emerged from our results: 1) the general consistency of effect at a particular SNP across race/ethnicity groups; 2) variation in association for different SNPs in the same locus can be useful for fine-mapping regions of interest; 3) and the demonstration of pleiotropic effects for specific SNPs.

Generalization

First, for those SNPs which demonstrated a statistically significant association in multiple race/ethnicity groups, the direction and magnitude of the effect was fairly consistent. Several SNPs showed a similar statistically significant effect in European Americans, African Americans, and Hispanics (rs2228145 in *IL6R* and rs429358 in *APOE*) or European Americans and Hispanics (rs2075650 in *TOMM40* and rs6857 in *PVRL2*). This trend of similar direction and magnitude across race/ethnicity groups generally held even where the race/ethnicity-specific associations did not reach statistical significance across all groups, such as SNP rs7310409 in *HNF1A* (p-het 0.589) and SNP rs599839 in *PSRC1* (p-het = 0.962). In general, most SNPs demonstrated similar effect estimates across race/ethnicity groups. When combined with

previous findings, these results suggest a shared genetic influence between race/ethnicity groups at these SNPs.

However, there were several notable exceptions where large differences were seen between race/ethnicity groups. Two SNPs in *CRP*, for example, demonstrated statistically significant associations with CRP that were larger in African Americans than European Americans (rs1800947) or than European Americans and Hispanics (rs1205). For several other SNPs, such as rs2075650 and rs4420638 in *TOMM40* (p-het < 1.64×10^{-9}) and rs3093058 in *CRP* (p-het = 1.19×10^{-4}), there appear to be large differences between race/ethnicity groups, with the SNP demonstrating an effect in one or two groups but not the others. Such differences could be indicative of population differences in linkage disequilibrium (LD), since tagging variants in European-ancestry populations may not be representative in other groups (24, 36). As such, this trend may suggest that the functional SNP is likely to be within the LD block shared by those race/ethnicity groups with similar effect estimates for that tagSNP, while for the "outlying" race/ethnicity group the tagSNP may not be well correlated with the functional SNP. Alternatively, differences in generalizability could also be due to a reduced ability to detect an association in some race/ethnicity groups in our study due to smaller sample sizes, differences in allele frequency, smaller effect sizes, or lower correlation with functional variants.

Fine-mapping

A related second theme that emerges is that differences in SNP generalizability may be informative of the genetic architecture of loci with multiple SNPs associated with CRP. Previous studies have shown that both average CRP levels and SNP-CRP associations can vary by race/ethnicity group (24, 28). Additionally, while some polymorphisms have demonstrated

an association across race/ethnicity groups (rs1205, (27)), others have only shown an association in two race/ethnicity groups, such as European Americans and Hispanics (rs4131568), or only in a single group, such as African Americans (rs3093058, (24, 27)). In our study, some SNPs within the *CRP* and *APOE/APOC1/TOMM40* loci generalized while others did not, providing evidence of potential differences in genetic effect among race/ethnicity groups. Since LD structure also varies by ancestry, we can utilize this information to identify smaller regions that are more likely to be functionally relevant.

We identified three SNPs in the *APOE/APOC1/TOMM40* locus associated with serum CRP that generalized to other race/ethnicity groups. SNPs rs2075650 and rs6857 generalized only to Hispanics, while rs429358 demonstrated an association in European Americans, African Americans and Hispanics. SNP rs429358, in the third exon of the *APOE* gene, is one of two non-synonymous SNPs that define the major ε2, ε3, and ε4 haplotypes of this region (37). Our results align with previous studies finding this haplotype associated with serum CRP in European Americans (38), African Americans (39), Hispanics (40, 41), and Asian/Pacific Islanders (29, 42). The magnitude and direction of the effect were fairly consistent across race/ethnicity groups, supporting a jointly consistent association across multiple race/ethnicity groups. Previous studies have suggested that SNPs at the *APOE/APOC1/TOMM40* locus appear to represent a single signal best represented by rs429358, with other SNPs appearing associated due to correlation (37, 43). Given the consistency of effect seen in rs429358 across race/ethnicity groups compared to the other SNPs we evaluated in this region, our results appear to support this view, and extend this SNP finding to African Americans and Hispanics.

We also observed race/ethnicity-specific differences in association with CRP among SNPs in the *CRP* locus. Of the five SNPs demonstrating an overall association, only rs1205 and

rs1800947 were statistically significant in multiple race/ethnicity groups, while the other three were only statistically significant in a single group. One of these SNPs, rs3093058, was only statistically significant in African Americans. While some of these SNPs may not have generalized to other groups due to smaller sample size, ancestry-related differences in genetic contributions to CRP concentrations at this locus are also possible. For example, for many SNPs in this region the allele frequencies in African Americans were different than other race/ethnicity groups, as previous studies have also highlighted (34). In such a situation, investigating SNPs in multiple populations can be useful for evaluating which SNPs might be associated with CRP in multiple populations (such as rs1205), and which may be race/ethnicity-specific (such as rs3093058). Other SNPs may have a larger effect in one race/ethnicity group than another (such as rs1800947). If confirmed, this type of information could potentially be important for ensuring that genetic information is appropriately interpreted and targeted where there are differences in effect by race/ethnicity.

Pleiotropy

The third theme is that our results show evidence for potentially pleiotropic effects. As previously stated, most of the SNPs evaluated in this study have previously been associated with other inflammation-related phenotypes, such as HDL and LDL cholesterol, triglycerides, and coronary artery disease. Studies have also demonstrated associations between some of these SNPs and phenotypes as diverse as colon and rectal cancer (rs1205 (12)), cervical cancer survival (rs1417938 (44)), and fasting glucose concentration (rs780094 (45)), though only this last trait reached genome-wide significance. Our results support several of these SNPs also being associated with serum CRP concentrations, suggesting that several of these SNPs and loci appear to have pleiotropic effects. This knowledge could be useful for exploring potentially

shared disease etiology or mechanisms between phenotypes associated with a given pleiotropic SNP or locus.

We identified a potentially novel pleiotropic association between two correlated SNPs (CEU $r^2 = 0.9$) in the *CELSR2/PSRC1/SORT1* locus and serum CRP levels, as these SNPs have previously been associated with several other inflammation-related phenotypes but not CRP. SNP rs599839 has been associated with LDL cholesterol (46-51), coronary artery disease (47-49, 52), myocardial infarction (53), triglyceride metabolism (47), and coronary heart disease (51), while SNP rs646776 has been associated with LDL cholesterol (46) and progranulin levels (54). Variations at rs646776 have also been strongly associated with transcript concentrations of the *CELSR2, PSRC1,* and *SORT1* genes (46). Several studies have also found a significant association between variation in rs599839 and sortilin mRNA expression in the liver (46, 49, 55). Sortilin, the gene product of *SORT1*, acts as a multiligand receptor and influences the uptake of LDL particles into cells, with studies suggesting that the G allele of rs599839 offers a protective effect against coronary artery disease mediated through LDL cholesterol lowering (49). Our finding that SNPs in this region are also associated with CRP may provide additional insights on the inflammation pathway, and how these SNPs might impact multiple outcomes. Our potentially novel finding in this region also demonstrates the value of evaluating SNPs previously associated with related phenotypes for additional pleiotropic relationships.

Strengths of this study stem from the inclusion of well-characterized study populations with diverse race/ethnicity groups. We had large sample sizes overall, with large numbers for European Americans, African American, and Hispanics. While we had fewer Asian/Pacific Islander and American Indian participants, we were still able to evaluate several SNPs in these groups. One limiting factor was that not all participants were genotyped for all SNPs, since SNP

panels and genotyping platforms varied between participating studies. This reduced the overall sample size available for any given SNP- and race/ethnicity-specific association, though numbers were still large for most analyses, particularly in European Americans, African Americans, and Hispanics. In Asian/Pacific Islanders and American Indians, for many SNPs the combination of smaller numbers or unavailable genotype information reduced or eliminated our ability to determine whether CRP associations generalized to these other populations. Furthermore, race/ethnicity differences in coded allele frequency may have reduced our ability to detect an association in various groups. For some SNPs, these differences in allele frequency may at least partly explain why an association with CRP in one group did not generalize to another. For these reasons, a lack of generalization in our study should not be interpreted as proof that these SNPs are not associated with CRP in other groups, particularly for the Asian/Pacific Islander or American Indian groups where our ability to detect an association was reduced.

While our analysis benefited from an a priori selection of interesting SNPs from previous GWAS, it is not a comprehensive analysis across all known loci given the rapid progress made since SNP selection. However, this limitation is lessened by our inclusion of the initial GWAS findings, which tend to have strong effect sizes, and hence tend to explain a larger fraction of the genetic variation than more recent GWAS findings. We applied a Bonferroni-corrected p-value to determine the statistical significance of a given association. While we adjusted for 266 independent tests, a potential limitation is that we did not adjust for testing in multiple race/ethnicity groups. However, the Bonferroni adjustment is conservative in nature, and we tested GWAS findings for CRP or related traits. Many, though not all, of our findings greatly surpassed our corrected p-value and reached more stringent genome-wide significance levels ($p < 5x10^{-8}$).

In conclusion, our results support and extend previous observations of associations between CRP and genetic variation in several loci. We generalized several associations between SNPs and serum CRP levels previously identified in European Americans to African Americans or Hispanics, and we also identified a potentially novel CRP locus, previously associated with coronary artery disease and LDL cholesterol. Our findings demonstrate the benefit of evaluating genotype-phenotype associations in multiple race/ethnicity groups, and of looking for pleiotropic relationships among SNPs previously associated with related phenotypes. Additional follow-up and fine-mapping of these loci may lead to better characterization of the functional variants in these regions.

References

1. Ridker PM. C-reactive protein: eighty years from discovery to emergence as a major risk marker for cardiovascular disease. Clinical chemistry. 2009;55(2):209-15. Epub 2008/12/20.
2. Ridker PM, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. JAMA : the journal of the American Medical Association. 2001;285(19):2481-5. Epub 2001/05/23.
3. Ford ES, Giles WH. Serum C-reactive protein and self-reported stroke: findings from the Third National Health and Nutrition Examination Survey. Arteriosclerosis, thrombosis, and vascular biology. 2000;20(4):1052-6. Epub 2000/04/15.
4. Rost NS, Wolf PA, Kase CS, Kelly-Hayes M, Silbershatz H, Massaro JM, et al. Plasma concentration of C-reactive protein and risk of ischemic stroke and transient ischemic attack: the Framingham study. Stroke; a journal of cerebral circulation. 2001;32(11):2575-9. Epub 2001/11/03.
5. Everett BM, Kurth T, Buring JE, Ridker PM. The relative strength of C-reactive protein and lipid levels as determinants of ischemic stroke compared with coronary heart disease in women. Journal of the American College of Cardiology. 2006;48(11):2235-42. Epub 2006/12/13.
6. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. JAMA : the journal of the American Medical Association. 2001;286(3):327-34. Epub 2001/07/24.
7. Laaksonen DE, Niskanen L, Nyyssonen K, Punnonen K, Tuomainen TP, Valkonen VP, et al. C-reactive protein and the development of the metabolic syndrome and diabetes in middle-aged men. Diabetologia. 2004;47(8):1403-10. Epub 2004/08/17.

8. Sattar N, Gaw A, Scherbakova O, Ford I, O'Reilly DS, Haffner SM, et al. Metabolic syndrome with and without C-reactive protein as a predictor of coronary heart disease and diabetes in the West of Scotland Coronary Prevention Study. Circulation. 2003;108(4):414-9. Epub 2003/07/16.
9. Festa A, D'Agostino R, Jr., Tracy RP, Haffner SM, Insulin Resistance Atherosclerosis S. Elevated levels of acute-phase proteins and plasminogen activator inhibitor-1 predict the development of type 2 diabetes: the insulin resistance atherosclerosis study. Diabetes. 2002;51(4):1131-7. Epub 2002/03/28.
10. Mora S, Lee IM, Buring JE, Ridker PM. Association of physical activity and body mass index with novel and traditional cardiovascular biomarkers in women. JAMA : the journal of the American Medical Association. 2006;295(12):1412-9. Epub 2006/03/23.
11. Mora S, Cook N, Buring JE, Ridker PM, Lee IM. Physical activity and reduced risk of cardiovascular events: potential mediating mechanisms. Circulation. 2007;116(19):2110-8. Epub 2007/10/31.
12. Slattery ML, Curtin K, Poole EM, Duggan DJ, Samowitz WS, Peters U, et al. Genetic variation in C-reactive protein in relation to colon and rectal cancer risk and survival. International journal of cancer Journal international du cancer. 2011;128(11):2726-34. Epub 2010/10/16.
13. Groblewska M, Mroczko B, Sosnowska D, Szmitkowski M. Interleukin 6 and C-reactive protein in esophageal cancer. Clinica chimica acta; international journal of clinical chemistry. 2012;413(19-20):1583-90. Epub 2012/05/23.
14. Wessel J, Moratorio G, Rao F, Mahata M, Zhang L, Greene W, et al. C-reactive protein, an 'intermediate phenotype' for inflammation: human twin studies reveal heritability, association with blood pressure and the metabolic syndrome, and the influence of common polymorphism at catecholaminergic/beta-adrenergic pathway loci. Journal of hypertension. 2007;25(2):329-43. Epub 2007/01/11.
15. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, et al. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis. 2001;154(3):681-9. Epub 2001/03/21.
16. Dupuis J, Larson MG, Vasan RS, Massaro JM, Wilson PW, Lipinska I, et al. Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q. Atherosclerosis. 2005;182(2):307-14. Epub 2005/09/15.
17. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, et al. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. American journal of human genetics. 2008;82(5):1185-92. Epub 2008/04/29.
18. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation. 2011;123(7):731-8. Epub 2011/02/09.
19. Reiner AP, Beleza S, Franceschini N, Auer PL, Robinson JG, Kooperberg C, et al. Genome-wide Association and Population Genetic Analysis of C-Reactive Protein in African American and Hispanic American Women. American journal of human genetics. 2012;91(3):502-12. Epub 2012/09/04.
20. Khera A, McGuire DK, Murphy SA, Stanek HG, Das SR, Vongpatanasin W, et al. Race and gender differences in C-reactive protein levels. Journal of the American College of Cardiology. 2005;46(3):464-9. Epub 2005/08/02.

21. Dreon DM, Slavin JL, Phinney SD. Oral contraceptive use and increased plasma concentration of C-reactive protein. Life sciences. 2003;73(10):1245-52. Epub 2003/07/10.
22. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. The Journal of clinical investigation. 2003;111(12):1805-12. Epub 2003/06/19.
23. Ford ES. Does exercise reduce inflammation? Physical activity and C-reactive protein among U.S. adults. Epidemiology. 2002;13(5):561-8. Epub 2002/08/23.
24. Doumatey AP, Chen G, Tekola Ayele F, Zhou J, Erdos M, Shriner D, et al. C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. Human molecular genetics. 2012;21(13):3063-72. Epub 2012/04/12.
25. Saito I, Sato S, Nakamura M, Kokubo Y, Mannami T, Adachi H, et al. A low level of C-reactive protein in Japanese adults and its association with cardiovascular risk factors: the Japan NCVC-Collaborative Inflammation Cohort (JNIC) study. Atherosclerosis. 2007;194(1):238-44. Epub 2006/09/12.
26. Lakoski SG, Cushman M, Palmas W, Blumenthal R, D'Agostino RB, Jr., Herrington DM. The relationship between blood pressure and C-reactive protein in the Multi-Ethnic Study of Atherosclerosis (MESA). Journal of the American College of Cardiology. 2005;46(10):1869-74. Epub 2005/11/16.
27. Crawford DC, Sanders CL, Qin X, Smith JD, Shephard C, Wong M, et al. Genetic variation is associated with C-reactive protein levels in the Third National Health and Nutrition Examination Survey. Circulation. 2006;114(23):2458-65. Epub 2006/11/15.
28. Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, Hosono N, et al. Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Human molecular genetics. 2011;20(6):1224-31. Epub 2011/01/05.
29. Wu Y, McDade TW, Kuzawa CW, Borja J, Li Y, Adair LS, et al. Genome-wide association with C-reactive protein levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. Inflammation. 2012;35(2):574-83. Epub 2011/06/08.
30. Su S, Lampert R, Zhao J, Bremner JD, Miller A, Snieder H, et al. Pleiotropy of C-reactive protein gene polymorphisms with C-reactive protein levels and heart rate variability in healthy male twins. The American journal of cardiology. 2009;104(12):1748-54. Epub 2009/12/08.
31. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, et al. The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. American journal of epidemiology. 2011;174(7):849-59. Epub 2011/08/13.
32. Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, et al. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. American journal of human genetics. 2008;82(5):1193-201. Epub 2008/04/29.
33. Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, et al. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. JAMA : the journal of the American Medical Association. 2006;296(22):2703-11. Epub 2006/12/14.

34. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, et al. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. American journal of human genetics. 2005;77(1):64-77. Epub 2005/05/18.
35. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.
36. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nature reviews Genetics. 2010;11(5):356-66. Epub 2010/04/17.
37. Yu CE, Seltman H, Peskind ER, Galloway N, Zhou PX, Rosenthal E, et al. Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: patterns of linkage disequilibrium and disease/marker association. Genomics. 2007;89(6):655-65. Epub 2007/04/17.
38. Judson R, Brain C, Dain B, Windemuth A, Ruano G, Reed C. New and confirmatory evidence of an association between APOE genotype and baseline C-reactive protein in dyslipidemic individuals. Atherosclerosis. 2004;177(2):345-51. Epub 2004/11/09.
39. Chasman DI, Kozlowski P, Zee RY, Kwiatkowski DJ, Ridker PM. Qualitative and quantitative effects of APOE genetic variation on plasma C-reactive protein, LDL-cholesterol, and apoE protein. Genes and immunity. 2006;7(3):211-9. Epub 2006/03/03.
40. Aiello AE, Nguyen HO, Haan MN. C-reactive protein mediates the effect of apolipoprotein E on cytomegalovirus infection. The Journal of infectious diseases. 2008;197(1):34-41. Epub 2008/01/04.
41. Haan MN, Aiello AE, West NA, Jagust WJ. C-reactive protein and rate of dementia in carriers and non carriers of Apolipoprotein APOE4 genotype. Neurobiology of aging. 2008;29(12):1774-82. Epub 2007/06/02.
42. Austin MA, Zhang C, Humphries SE, Chandler WL, Talmud PJ, Edwards KL, et al. Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. Annals of human genetics. 2004;68(Pt 3):179-88. Epub 2004/06/08.
43. Curocichin G, Wu Y, McDade TW, Kuzawa CW, Borja JB, Qin L, et al. Single-nucleotide polymorphisms at five loci are associated with C-reactive protein levels in a cohort of Filipino young adults. Journal of human genetics. 2011;56(12):823-7. Epub 2011/09/23.
44. Polterauer S, Grimm C, Zeillinger R, Heinze G, Tempfer C, Reinthaller A, et al. Association of C-reactive protein (CRP) gene polymorphisms, serum CRP levels and cervical cancer prognosis. Anticancer research. 2011;31(6):2259-64. Epub 2011/07/09.
45. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes. 2008;57(11):3112-21. Epub 2008/08/06.
46. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nature genetics. 2008;40(2):189-97. Epub 2008/01/15.
47. Kleber ME, Renner W, Grammer TB, Linsel-Nitschke P, Boehm BO, Winkelmann BR, et al. Association of the single nucleotide polymorphism rs599839 in the vicinity of the sortilin 1 gene with LDL and triglyceride metabolism, coronary heart disease and myocardial infarction. The Ludwigshafen Risk and Cardiovascular Health Study. Atherosclerosis. 2010;209(2):492-7. Epub 2009/10/20.

48. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature genetics. 2008;40(2):161-9. Epub 2008/01/15.
49. Linsel-Nitschke P, Heeren J, Aherrahrou Z, Bruse P, Gieger C, Illig T, et al. Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. Atherosclerosis. 2010;208(1):183-9. Epub 2009/08/08.
50. Samani NJ, Braund PS, Erdmann J, Gotz A, Tomaszewski M, Linsel-Nitschke P, et al. The novel genetic variant predisposing to coronary artery disease in the region of the PSRC1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. Journal of molecular medicine. 2008;86(11):1233-41. Epub 2008/07/24.
51. Angelakopoulou A, Shah T, Sofat R, Shah S, Berry DJ, Cooper J, et al. Comparative analysis of genome-wide association studies signals for lipids, diabetes, and coronary heart disease: Cardiovascular Biomarker Genetics Collaboration. European heart journal. 2012;33(3):393-407. Epub 2011/08/02.
52. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. The New England journal of medicine. 2007;357(5):443-53. Epub 2007/07/20.
53. Wang AZ, Li L, Zhang B, Shen GQ, Wang QK. Association of SNP rs17465637 on chromosome 1q41 and rs599839 on 1p13.3 with myocardial infarction in an American caucasian population. Annals of human genetics. 2011;75(4):475-82. Epub 2011/04/06.
54. Carrasquillo MM, Nicholson AM, Finch N, Gibbs JR, Baker M, Rutherford NJ, et al. Genome-wide screen identifies rs646776 near sortilin as a regulator of progranulin levels in human plasma. American journal of human genetics. 2010;87(6):890-7. Epub 2010/11/23.
55. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. PLoS biology. 2008;6(5):e107. Epub 2008/05/09.

Analysis I – Inflammation Table 1

Table 1 - Demographic and epidemiologic characteristics of study populations used in this analysis.

Measure / Unit	Race/Ethnicity	WHI	EAGLE	MEC	ARIC	CARDIA	CHS	SHS	Total
Sample size	European American	6889	4001	277	8774	1754	3263	.	24958
n	African American	1809	1350	481	2529	1495	807	.	8471
	Hispanic	778	1877	280	.	.	.	.	2935
	Asian/Pacific Islande	425	.	555	.	.	.	.	980
	American Indian	.	.	.	.	.	.	3129	3129
	Overall	9901	7228	1593	11303	3249	4070	3129	40473
CRP (mg/dL)	European American	2.9 (1.2 - 4.3)	2.3 (0.9 - 4.9)	1.0 (0.4 - 2.2)	2.3 (1.0 - 5.0)	1.0 (0.5 - 2.6)	1.7 (0.9 - 3.0)	.	
Median (IQR)	African American	2.8 (0.9 - 6.7)	2.8 (1.0 - 6.7)	1.3 (0.5 - 3.2)	3.6 (1.4 - 7.7)	2.0 (0.7 - 5.2)	2.5 (1.3 - 5.7)	.	
	Hispanic	3.1 (1.5 - 6.0)	2.5 (1.0 - 5.4)	1.2 (0.5 - 2.5)	.	.	.	.	
	Asian/Pacific Islande	1.0 (0.5 - 2.1)	.	0.5 (0.2 - 1.0)	.	.	.	.	
	American Indian	.	.	.	.	.	.	3.9 (1.9 - 8.3)	
	Overall	2.9 (1.3 - 6.0)	2.5 (1.0 - 5.4)	0.8 (0.3 - 2.1)	2.5 (1.1 - 5.5)	1.4 (0.6 - 3.8)	1.8 (0.9 - 3.3)	3.9 (1.9 - 8.3)	
Age (years)	European American	67.3 (6.9)	52.4 (19.5)	59.8 (8.2)	63.1 (5.6)	40.7 (3.4)	72.3 (5.4)	.	
Mean (SD)	African American	62.0 (7.2)	47.5 (17.3)	62.9 (7.4)	61.8 (5.7)	39.6 (3.8)	72.9 (5.6)	.	
	Hispanic	61.0 (7.0)	45.8 (17.7)	61.5 (6.7)	.	.	.	.	
	Asian/Pacific Islande	65.6 (7.4)	.	59.8 (8.5)	.	.	.	.	
	American Indian	.	.	.	.	.	.	39.3 (16.4)	
	Overall	65.7 (7.4)	49.8 (18.9)	61.0 (7.9)	62.8 (5.7)	40.2 (3.6)	72.4 (5.4)	39.3 (16.4)	
Female	European American	100.0%	51.1%	36.6%	53.5%	53.0%	60.1%	.	
%	African American	100.0%	52.4%	19.8%	64.1%	58.3%	62.60%	.	
	Hispanic	100.0%	51.7%	30.4%	.	.	.	.	
	Asian/Pacific Islande	100.0%	.	42.5%	.	.	.	.	
	American Indian	.	.	.	.	.	.	61.5%	
	Overall	100.0%	51.5%	32.5%	55.9%	55.4%	61.1%	61.5%	
BMI (SD)	European American	28.4 (6.4)	27.8 (5.9)	26.5 (5.2)	28.3 (5.3)	27.1 (5.5)	26.3 (4.4)	.	
kg/m^2	African American	32.0 (7.5)	29.7 (7.3)	27.8 (4.9)	30.6 (6.4)	30.5 (7.0)	28.5 (5.5)	.	
	Hispanic	29.2 (6.4)	28.6 (5.6)	27.6 (4.2)	.	.	.	.	
	Asian/Pacific Islande	24.8 (4.4)	.	25.4 (4.1)	.	.	.	.	
	American Indian	.	.	.	.	.	.	32.35 (7.9)	
	Overall	29.0 (6.7)	28.4 (6.2)	26.7 (4.7)	28.8 (5.6)	28.7 (6.5)	26.7 (4.7)	32.35 (7.9)	
Smoking	European American	7.5 / 42.6%	25.8 / 28.9%	8.7 / 52.7%	14.1 / 45.8%	16.1 / 23.4%	11.3 / 40.9%	.	
% Current /	African American	10.0 / 39.4%	34.6 / 15.7%	24.2 / 44.9%	17.6 / 35.0%	28.3 / 11.7%	16.0 / 35.2%	.	
% Former	Hispanic	5.5 / 28.0%	20.5 / 23.5%	15.2 / 44.0%	.	.	.	.	
	Asian/Pacific Islande	3.3 / 25.2%	.	10.5 / 43.3%	.	.	.	.	
	American Indian	.	.	.	.	.	.	34 / 23.3%	
	Overall	7.6 / 40.1%	26.1 / 25.0%	15.1 / 45.6%	14.8 / 43.4%	21.7 / 18.0%	12.2 / 39.7%	34 / 23.3%	
HRT	European American	34.2 / 17.8%	31.9% *	37.3 / 33.3%	29.5 / ** %	***	28.9 / 13.1%	.	
% Current /	African American	25.3 / 14.0%	15.3% *	14.0 / 36.6%	17.3 / ** %	***	17.2 / 7.3%	.	
% Former	Hispanic	39.1 / 12.0%	16.1% *	28.2 / 25.9%	.	.	.	.	
	Asian/Pacific Islande	48.9 / 14.1%	.	31.5 / 31.1%	.	.	.	.	
	American Indian	.	.	.	.	.	.	9.8 / 11.4%	
	Overall	33.7 / 16.5%	24.6% *	28.9 / 31.7%	26.4 / ** %	***	16.2 / 7.3%	9.8 / 11.4%	

* EAGLE did not have a way of identifying former hormone therapy usage

** ARIC does not have reliable information for former HRT use

*** CARDIA participants were in their 30's and 40's at time of CRP measurement, so no HRT information is available

Analysis I – Inflammation Table 2

Table 2 - Number of SNPs and partipants available for meta-analyses across studies, stratified by race/ethnicity group and overall

Race/ethnicity gorup		European American	African American	Hispanic	Asian/ Pacific Islander	American Indian	Overall
# SNPs available in each study							
	WHI *	94	92	94	91	.	94
	EAGLE	196	196	196	.	.	196
	MEC	103	103	103	103	.	103
	ARIC	31	31	.	.	.	31
	CARDIA	7	7	.	.	.	7
	CHS	111	118	.	.	.	118
	SHS	.	.	.	.	19	19
	Overall	265	266	261	150	19	266
# observations per SNP meta-analysis							
	Mean	7722	2367	1766	617	3129	12361
	Minumum	233	433	243	388	3129	802
	Median	6820	1850	1857	526	3129	9818
	Maximum	24319	8158	2908	951	3129	36299

* Some SNPs excluded because of < 100 participants (2 in African Americans, 3 in Asian/Pacific Islanders, 94 in American Indians)

Analysis I – Inflammation Table 3

Table 3 - Association between serum C-reactive protein and SNPs previously associated with inflammation-related phenotypes.

SNP	Allele, position, gene									Within-group	Between-group
Locus	Previously associated trait	Race/ethnicity group	n	# Studies	CAF	Beta	SE	P-value	p-het	p-het
rs1205	T	European American	12400	2	0.33	**-0.17**	**0.01**	**1.03E-31**	0.02	
CRP	3' UTR, *CRP*	African American	3669	2	0.20	**-0.27**	**0.04**	**8.09E-15**	0.61	
	C-reactive Protein	Hispanic	1842	1	0.35	**-0.22**	**0.04**	**5.37E-09**	.	
		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	17911	2		**-0.19**	**0.01**	**1.15E-50**		0.02
rs1800947	G	European American	12404	2	0.06	**-0.30**	**0.03**	**3.10E-25**	0.77	
CRP	Exonic, *CRP*	African American	3676	2	0.01	**-0.61**	**0.13**	**1.28E-06**	0.77	
	C-reactive Protein	Hispanic	1842	1	0.02	-0.36	0.13	6.72E-03	.	
		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	17922	2		**-0.32**	**0.03**	**1.13E-30**		0.05
rs1417938	A	European American	3943	1	0.30	**0.14**	**0.03**	**5.57E-07**	.	
CRP	Intronic, *CRP*	African American	1328	1	0.11	0.20	0.08	1.17E-02	.	
	C-reactive Protein	Hispanic	1845	1	0.36	0.14	0.04	2.71E-04	.	
		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	7116	1		**0.15**	**0.02**	**2.97E-11**		0.79
rs4131568	T	European American	11314	2	0.35	**0.08**	**0.02**	**1.24E-07**	0.31	
CRP	Upstream, *CRP*	African American	3461	2	0.07	-0.08	0.06	1.54E-01	0.96	
	C-reactive protein	Hispanic	1857	1	0.39	0.08	0.04	3.77E-02	.	
		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	16632	2		**0.07**	**0.01**	**2.45E-07**		0.02
rs3093058	T	European American	12420	2	0.99	-0.13	0.22	5.57E-01	0.13	
CRP	Upstream, *CRP*	African American	3680	2	0.59	**-0.21**	**0.04**	**6.85E-09**	**3.01E-33**	
	C-reactive Protein	Hispanic	1842	1	0.01	0.67	0.20	1.02E-03	.	
		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	17942	2		**-0.18**	**0.03**	**2.45E-07**		**1.19E-04**
rs4420638	G	European American	21378	4	0.18	**-0.24**	**0.02**	**1.01E-56**	0.12	
APOE/	Downstream, *APOC1*	African American	6076	4	0.20	-0.03	0.03	2.66E-01	0.35	
APOC/	LDL Cholesterol	Hispanic	2637	2	0.10	-0.18	0.05	5.58E-04	0.04	
TOMM40		Asian/Pacific Islander	422	1	0.12	-0.06	0.14	6.76E-01	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	30513	5		**-0.19**	**0.01**	**1.26E-49**		**1.64E-09**
rs2075650	G	European American	15926	4	0.14	**-0.22**	**0.02**	**1.83E-38**	0.54	
APOE/	Intronic, *TOMM40*	African American	4729	4	0.14	-0.02	0.03	6.47E-01	0.29	
APOC/	LDL Cholesterol	Hispanic	2119	2	0.10	**-0.21**	**0.05**	**1.00E-04**	0.12	
TOMM40		Asian/Pacific Islander	528	2	0.18	-0.03	0.04	5.28E-01	0.85	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	23302	5		**-0.16**	**0.01**	**2.59E-32**		**1.60E-09**
rs429358	C	European American	6993	2	0.13	**-0.24**	**0.04**	**2.41E-10**	0.61	
APOE/	Exonic, *APOE*	African American	2220	2	0.20	**-0.23**	**0.04**	**5.07E-08**	0.05	
APOC/	Cardiovascular Disease	Hispanic	1031	2	0.11	**-0.36**	**0.07**	**5.08E-08**	0.02	
TOMM40		Asian/Pacific Islander	943	2	0.11	-0.19	0.05	2.25E-04	0.84	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	11187	2		**-0.24**	**0.02**	**2.29E-25**		0.23
rs6857	T	European American	3955	1	0.16	**-0.23**	**0.04**	**2.07E-10**	.	
APOE/	3' UTR, *PVRL2*	African American	1331	1	0.06	-0.23	0.11	3.04E-02	.	
APOC/	Cardiovascular Disease	Hispanic	1852	1	0.10	**-0.28**	**0.06**	**6.12E-06**	.	
TOMM40		Asian/Pacific Islander	0	.	.	.	.	.	.	
		American Indian	0	.	.	.	.	.	.	
		OVERALL	7138	1		**-0.24**	**0.03**	**7.30E-16**		0.79

Table 3 - Continued											
SNP	Allele, position, gene									Within-group	Between-group
Locus	Previously associated trait		Race/ethnicity	n	# Studies	CAF	Beta	SE	P-value	P-het	P-het
rs1260326	T		European American	22661	5	0.42	**0.10**	**0.01**	**2.38E-17**	0.16	
GCKR	Exonic, *GCKR*		African American	6506	5	0.15	0.06	0.03	2.06E-02	0.42	
	Triglycerides		Hispanic	2873	3	0.33	0.08	0.03	5.78E-03	0.45	
			Asian/Pacific Islander	951	2	0.54	0.09	0.03	2.04E-03	0.11	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	32991	6		**0.09**	**0.01**	**6.40E-22**		0.72
rs780094	T		European American	17567	5	0.40	**0.10**	**0.01**	**1.53E-16**	0.32	
GCKR	Intronic, *GCKR*		African American	6175	5	0.19	0.03	0.03	2.26E-01	0.23	
	C-reactive Protein		Hispanic	2130	2	0.34	0.07	0.03	2.62E-02	0.26	
			Asian/Pacific Islander	528	1	0.54	0.05	0.03	1.51E-01	0.63	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	26400	6		**0.08**	**0.01**	**5.65E-17**		0.09
rs2650000	A		European American	16506	5	0.35	**-0.12**	**0.01**	**2.62E-23**	0.18	
HNF1A	Upstream, *HNF1A*		African American	5983	5	0.12	-0.09	0.03	5.24E-03	0.49	
	LDL Cholesterol		Hispanic	2125	2	0.36	-0.11	0.03	9.45E-04	0.08	
			Asian/Pacific Islander	500	1	0.45	-0.03	0.03	3.61E-01	0.25	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	25114	6		**-0.11**	**0.01**	**4.34E-26**		0.08
rs7310409	A		European American	3956	1	0.40	**-0.18**	**0.03**	**1.57E-10**	.	
HNF1A	Intronic, *HNF1A*		African American	1327	1	0.32	-0.15	0.06	7.94E-03	.	
	C-reactive protein		Hispanic	1855	1	0.41	-0.13	0.04	1.13E-03	.	
			Asian/Pacific Islander	0	.	.	.	.	.	.	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	7138	1		**-0.16**	**0.02**	**3.33E-14**		0.59
rs2228145	C		European American	17184	4	0.40	**-0.10**	**0.01**	**1.47E-18**	0.09	
IL6R	Exonic, *IL6R*		African American	5942	4	0.14	**-0.13**	**0.03**	**6.72E-05**	0.27	
	C-reactive Protein		Hispanic	1841	1	0.52	**-0.18**	**0.04**	**2.18E-06**	.	
			Asian/Pacific Islander	0	.	.	.	.	.	.	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	24967	5		**-0.11**	**0.01**	**3.66E-26**		0.16
rs1892534	T		European American	11344	2	0.39	**-0.08**	**0.02**	**5.80E-08**	0.13	
LEPR	Downstream, *LEPR*		African American	3471	2	0.46	-0.08	0.03	4.45E-03	0.49	
	C-reactive protein		Hispanic	1851	1	0.50	-0.12	0.04	1.99E-03	.	
			Asian/Pacific Islander	0	.	.	.	.	.	.	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	16666	2		**-0.09**	**0.01**	**1.02E-11**		0.67
rs1800961	T		European American	17660	5	0.031	**-0.15**	**0.03**	**4.75E-06**	0.39	
HNF4A	Exonic, *HNF4A*		African American	6222	5	0.007	0.00	0.10	9.86E-01	0.27	
	HDL Cholesterol		Hispanic	2117	2	0.039	-0.06	0.08	4.30E-01	0.99	
			Asian/Pacific Islander	523	1	0.001	-0.18	0.22	4.05E-01	0.89	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	26522	6		**-0.13**	**0.03**	**9.39E-06**		0.45
rs599839	G		European American	22427	5	0.23	**0.05**	**0.01**	**1.01E-04**	0.68	
CELSR2/	Downstream, *PSRC1*		African American	6596	5	0.71	0.04	0.02	5.95E-02	0.52	
PSRC1/	Coronary Artery Disease		Hispanic	2811	3	0.22	0.06	0.03	6.96E-02	0.25	
SORT1			Asian/Pacific Islander	946	2	0.08	0.05	0.05	3.24E-01	0.34	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	32780	6		**0.05**	**0.01**	**9.20E-07**		0.96
rs646776	C		European American	22652	5	0.22	0.05	0.01	2.02E-04	0.66	
CELSR2/	Downstream, *CELSR2*		African American	6479	5	0.35	0.03	0.02	1.97E-01	0.65	
PSRC1/	LDL Cholesterol		Hispanic	2889	3	0.19	0.04	0.03	2.72E-01	0.18	
SORT1			Asian/Pacific Islander	951	2	0.07	0.08	0.06	1.82E-01	0.05	
			American Indian	0	.	.	.	.	.	.	
			OVERALL	32971	6		**0.04**	**0.01**	**1.65E-05**		0.78

Analysis I – Inflammation Table 4

Table 4 - Association between serum C-reactive protein and SNPs previously associated with inflammation-related phenotypes, stratified by gender.

	SNP		Male				Female				Between sex
Locus	Coded allele	Race/ethnicity	N	Beta	SE	P-value	N	Beta	SE	P-value	P-het
CRP	rs1205	European American	**5838**	**-0.19**	**0.02**	**2.80E-19**	**6562**	**-0.16**	**0.02**	**1.62E-14**	0.29
	T	African American	**1469**	**-0.29**	**0.06**	**4.14E-07**	**2200**	**-0.27**	**0.04**	**1.75E-09**	0.78
		Hispanic	893	-0.16	0.05	4.68E-03	**949**	**-0.29**	**0.05**	**6.33E-08**	0.07
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**8200**	**-0.19**	**0.02**	**6.29E-26**	**9711**	**-0.19**	**0.02**	**5.12E-27**	1.00
CRP	rs1800947	European American	**5837**	**-0.28**	**0.04**	**1.16E-12**	**6567**	**-0.31**	**0.04**	**3.52E-14**	0.60
	G	African American	1475	-0.54	0.20	5.85E-03	**2201**	**-0.65**	**0.16**	**8.04E-05**	0.67
		Hispanic	894	-0.40	0.21	6.29E-02	948	-0.33	0.17	4.83E-02	0.80
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**8206**	**-0.30**	**0.04**	**1.08E-14**	**9716**	**-0.33**	**0.04**	**1.37E-17**	0.60
CRP	rs1417938	European American	1924	-0.08	0.04	4.29E-02	**2019**	**-0.20**	**0.04**	**1.29E-06**	0.03
	A	African American	631	-0.26	0.12	2.38E-02	697	-0.15	0.11	1.67E-01	0.50
		Hispanic	894	-0.10	0.06	6.37E-02	951	-0.18	0.05	9.67E-04	0.30
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	3449	-0.10	0.03	1.24E-03	**3667**	**-0.19**	**0.03**	**2.01E-09**	0.03
CRP	rs4131568	European American	5360	0.05	0.02	1.20E-02	**5954**	**0.10**	**0.02**	**2.93E-06**	0.08
	T	African American	1397	0.01	0.09	9.56E-01	2064	-0.13	0.07	6.27E-02	0.22
		Hispanic	893	0.05	0.05	3.32E-01	964	0.11	0.05	4.88E-02	0.40
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	7650	0.05	0.02	8.22E-03	**8982**	**0.09**	**0.02**	**1.18E-05**	0.16
CRP	rs3093058	European American	5846	-0.11	0.29	7.12E-01	6574	-0.09	0.32	7.75E-01	0.96
	T	African American	1475	-0.16	0.06	6.08E-03	**2205**	**-0.24**	**0.05**	**1.91E-07**	0.31
		Hispanic	893	0.90	0.29	1.86E-03	949	0.38	0.28	1.77E-01	0.20
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	8214	-0.12	0.06	3.57E-02	**9728**	**-0.22**	**0.04**	**8.51E-07**	0.17
APOE/	rs4420638	European American	**6633**	**-0.27**	**0.02**	**8.89E-30**	**14745**	**-0.23**	**0.02**	**1.31E-30**	0.16
APOC1/	G	African American	1704	-0.04	0.05	4.26E-01	4372	-0.03	0.03	4.47E-01	0.86
TOMM40		Hispanic	897	-0.15	0.09	8.92E-02	1740	-0.20	0.06	2.25E-03	0.64
		Asian/Pacific Islander	0	.	.	.	422	-0.06	0.14	6.76E-01	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**9234**	**-0.23**	**0.02**	**6.99E-27**	**21365**	**-0.17**	**0.02**	**2.01E-26**	0.03
APOE/	rs2075650	European American	**7277**	**-0.23**	**0.02**	**5.36E-21**	**8649**	**-0.22**	**0.02**	**2.03E-20**	0.72
APOC1/	G	African American	2057	-0.03	0.05	5.86E-01	2672	-0.01	0.05	7.69E-01	0.78
TOMM40		Hispanic	1078	-0.23	0.07	9.68E-04	1041	-0.19	0.08	2.16E-02	0.71
		Asian/Pacific Islander	306	-0.02	0.06	7.84E-01	222	-0.09	0.06	1.57E-01	0.41
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**10718**	**-0.17**	**0.02**	**2.55E-18**	**12584**	**-0.17**	**0.02**	**5.64E-18**	1.00
APOE/	rs429358	European American	163	-0.16	0.10	9.30E-02	**6830**	**-0.25**	**0.04**	**8.07E-10**	0.40
APOC1/	C	African American	363	-0.14	0.06	3.38E-02	**1857**	**-0.30**	**0.06**	**7.94E-08**	0.06
TOMM40		Hispanic	182	-0.22	0.11	5.33E-02	**849**	**-0.42**	**0.08**	**1.82E-07**	0.14
		Asian/Pacific Islander	304	-0.16	0.08	4.04E-02	639	-0.21	0.06	1.33E-03	0.62
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**1012**	**-0.16**	**0.04**	**1.18E-04**	**10260**	**-0.27**	**0.03**	**2.96E-23**	0.03
APOE/	rs6857	European American	**1932**	**-0.24**	**0.05**	**1.08E-06**	**2023**	**-0.24**	**0.05**	**9.14E-06**	1.00
APOC1/	T	African American	636	-0.35	0.15	1.98E-02	695	-0.12	0.15	4.26E-01	0.28
TOMM40		Hispanic	889	-0.30	0.09	4.90E-04	963	-0.27	0.09	2.66E-03	0.81
		Asian/Pacific Islander	0	.	.	.	0	.	.	.	.
		American Indian	0	.	.	.	0	.	.	.	.
		Overall	**1525**	**-0.26**	**0.04**	**1.88E-10**	**3681**	**-0.24**	**0.04**	**8.71E-08**	0.72

Table 4 - Continued													
	SNP		Male					Female				Between sex	
Locus	Coded allele	Race/ethnicity	N	Beta	SE	P-value		N	Beta	SE	P-value	P-het	
GCKR	rs1260326	European American	7237	0.04	0.02	1.78E-02		**15424**	**0.13**	**0.01**	**1.68E-18**	5.70E-05	
	T	African American	2047	0.07	0.04	9.54E-02		4459	0.06	0.04	8.13E-02	0.86	
		Hispanic	1074	0.02	0.05	6.72E-01		1799	0.13	0.04	1.14E-03	0.09	
		Asian/Pacific Islander	306	0.06	0.05	1.97E-01		645	0.12	0.04	2.45E-03	0.35	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	10664	0.04	0.01	2.10E-03		**22327**	**0.12**	**0.01**	**8.96E-23**	1.54E-08	
GCKR	rs780094	European American	8045	0.05	0.02	3.00E-03		**9522**	**0.14**	**0.02**	**9.76E-17**	1.46E-03	
	T	African American	2660	0.02	0.04	5.94E-01		3515	0.05	0.04	1.61E-01	0.60	
		Hispanic	1085	0.03	0.05	5.36E-01		1045	0.13	0.05	8.84E-03	0.16	
		Asian/Pacific Islander	305	0.02	0.05	6.09E-01		223	0.08	0.05	1.15E-01	0.40	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	12095	0.04	0.01	2.85E-03		**14305**	**0.12**	**0.01**	**4.74E-18**	1.54E-08	
HNF1A	rs2650000	European American	**7592**	**-0.15**	**0.02**	**5.52E-18**		**8914**	**-0.10**	**0.02**	**1.78E-08**	0.08	
	A	African American	2588	-0.04	0.04	4.23E-01		3395	-0.14	0.04	1.36E-03	0.08	
		Hispanic	1082	-0.12	0.04	8.76E-03		1043	-0.11	0.05	2.03E-02	0.98	
		Asian/Pacific Islander	283	-0.08	0.05	1.04E-01		217	0.01	0.05	7.76E-01	0.20	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	**11545**	**-0.13**	**0.01**	**1.48E-18**		**13569**	**-0.09**	**0.01**	**1.16E-10**	0.01	
HNF1A	rs7310409	European American	**1931**	**-0.20**	**0.04**	**5.15E-08**		2025	-0.15	0.04	2.76E-04	0.16	
	A	African American	633	-0.16	0.08	4.96E-02		694	-0.12	0.08	1.10E-01	0.76	
		Hispanic	892	-0.17	0.06	1.72E-03		963	-0.09	0.05	1.18E-01	0.32	
		Asian/Pacific Islander	0	.	.	.		0	.	.	.	.	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	**3456**	**-0.19**	**0.03**	**5.44E-11**		**3682**	**-0.13**	**0.03**	**3.06E-05**	0.08	
IL6R	rs2228145	European American	**7841**	**-0.12**	**0.02**	**2.81E-12**		**9343**	**-0.09**	**0.02**	**8.55E-09**	0.29	
	C	African American	2373	-0.15	0.05	1.47E-03		3569	-0.10	0.04	1.11E-02	0.44	
		Hispanic	**886**	**-0.22**	**0.05**	**4.57E-05**		955	-0.14	0.05	7.81E-03	0.26	
		Asian/Pacific Islander	0	.	.	.		0	.	.	.	.	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	**11100**	**-0.13**	**0.02**	**1.87E-17**		**13867**	**-0.10**	**0.01**	**1.21E-11**	0.18	
LEPR	rs1892534	European American	5377	-0.06	0.02	6.69E-03		**5967**	**-0.10**	**0.02**	**2.06E-06**	0.16	
	T	African American	1402	-0.07	0.05	1.59E-01		2069	-0.09	0.04	1.19E-02	0.76	
		Hispanic	893	-0.08	0.05	1.26E-01		958	-0.15	0.05	4.82E-03	0.32	
		Asian/Pacific Islander	0	.	.	.		0	.	.	.		
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	7672	-0.06	0.02	7.12E-04		**8994**	**-0.11**	**0.02**	**2.01E-09**	0.08	
HNF4A	rs1800961	European American	8093	-0.12	0.05	1.03E-02		9567	-0.16	0.05	3.11E-04	0.57	
	T	African American	2678	-0.20	0.14	1.53E-01		3544	0.26	0.16	9.26E-02	0.03	
		Hispanic	1079	-0.24	0.11	3.28E-02		1038	0.12	0.12	3.15E-01	0.03	
		Asian/Pacific Islander	301	-0.23	0.24	3.36E-01		222	.	.	.	.	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	12151	-0.15	0.04	3.25E-04		14371	-0.10	0.04	1.38E-02	0.38	
CELSR2/	rs599839	European American	7119	0.05	0.02	1.06E-02		15308	0.05	0.02	1.34E-03	1.00	
PSRC1/	G	African American	2093	0.00	0.03	9.45E-01		4503	0.06	0.03	2.16E-02	0.16	
SORT1		Hispanic	1041	0.10	0.05	5.83E-02		1770	0.03	0.04	4.30E-01	0.27	
		Asian/Pacific Islander	304	0.00	0.07	9.97E-01		642	0.13	0.09	1.26E-01	0.25	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	10557	0.04	0.02	8.19E-03		**22309**	**0.06**	**0.01**	**1.25E-05**	0.37	
CELSR2/	rs646776	European American	7248	0.05	0.02	1.67E-02		15404	0.05	0.02	1.70E-03	1.00	
PSRC1/	C	African American	2034	0.04	0.03	2.03E-01		4445	0.02	0.03	4.43E-01	0.64	
SORT1		Hispanic	1076	0.08	0.06	1.51E-01		1813	0.01	0.04	8.16E-01	0.33	
		Asian/Pacific Islander	306	-0.01	0.08	8.85E-01		645	0.22	0.09	1.72E-02	0.06	
		American Indian	0	.	.	.		0	.	.	.	.	
		Overall	10664	0.05	0.02	3.61E-03		22393	0.05	0.01	4.90E-04	1.00	

Part A, Analysis II - Melanoma

Title:

Replication of melanoma GWAS hits and exploration of pleiotropic effects of cancer GWAS hits with melanoma risk in the Population Architecture using Genetics and Epidemiology (PAGE) study

Authors:

Jonathan M Kocarnik[1], S Lani Park[2], Jiali Han[3], Logan Dumitrescu[4], Lynne R Wilkens[2], Fredrick R Schumacher[5], Laurence Kolonel[2], Chris Carlson[1], David Duggan[6], Carolyn Hutter[7], Abrar A Qureshi[8], Fengju Song[3], Lucia Hindorff[9], Dana Crawford[4], Charles Kooperberg[1], Loic Le Marchand[2], Ulrike Peters[1]

[1] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.
[2] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI.
[3] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.
[4] Center for Human Genetics Research, Vanderbilt University, Nashville, TN.
[5] Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA.
[6] Translational Genomics Research Institute, Phoenix, AZ.
[7] Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, NCI, NIH, Bethesda, MD.
[8] Department of Dermatology, Brigham and Women's Hospital, Boston, MA.
[9] Office of Population Genomics, NHGRI, NIH, Bethesda, MD.

Abstract

Introduction: Genome-wide association studies (GWAS) have identified several regions of the genome that show pleiotropic associations with multiple types of cancer. We evaluated whether single nucleotide polymorphisms (SNPs) previously associated with various cancers are also associated with melanoma risk. We additionally sought to replicate existing melanoma GWAS SNPs, and evaluate if associations with melanoma vary by sex.

Methods: We evaluated 189 SNPs identified be associated with different cancer sites in GWAS in 2,213 melanoma cases and 21,309 controls. Three studies from the Population Architecture using Genetics and Epidemiology (PAGE) study contributed: the Biorepository of Vanderbilt University (BioVU), the Multiethnic Cohort (MEC), and the Women's Health Initiative (WHI). We also collaborated with the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). Study-specific logistic regression models were combined using fixed-effect meta-analysis to evaluate the association between each SNP and melanoma. All analyses were restricted to participants of European ancestry. We used a Bonferroni-corrected p-value of $2.6x10^{-4}$ (0.05/189) to determine the statistical significance of an association.

Results: We replicated an association with melanoma ($p<2.6x10^{-4}$) for 9 of 10 SNPs: 7 of 8 melanoma GWAS SNPs and two SNPs in the *TERT-CLPTM1L* locus with known pleiotropic effects in opposing directions. None of the other 179 non-melanoma cancer SNPs examined showed statistically significant evidence for association with melanoma risk ($p>2.6x10^{-4}$). Sex-stratified analyses in the PAGE studies showed a male-specific association between the prostate cancer GWAS SNP rs12418451 (near *TPCN2*) and melanoma ($p = 1.92x10^{-4}$). Both this SNP (rs12418451, p-het = 0.005) and melanoma GWAS SNP rs16891982 (*SLC45A2*, p-het = 0.02) had larger effects in males than females.

Discussion: Our analysis provides confirmatory evidence of pleiotropic effects in opposite directions at the *TERT-CLPTM1L* locus. We also provide novel evidence of potential sex-specific associations with melanoma for two SNPs in ion transport genes previously associated with pigmentation. One of these SNPs, in *TPCN2*, was previously associated with prostate cancer but not melanoma. If confirmed, pleiotropic relationships may be useful for elucidating shared molecular pathways between diseases.

Introduction

As the most serious form of skin cancer, melanoma is a considerable public health burden. In 2012, there were an estimated 76,250 new diagnoses and 9,180 deaths from melanoma in the United States alone (1). Ultraviolet radiation exposure is the largest environmental risk factor for melanoma, with an estimated 44 – 90% of melanoma attributable to sun exposure (2). Other risk factors include artificial UV sources such as tanning beds (3), larger numbers of nevi, pigmentation traits (light versus dark hair, eye, and skin color), race/ethnicity (European versus non-European ancestry), skin response to UV exposure (burn versus tan), older age, and male sex (2). Anatomic location of melanoma also tends to vary by sex, arising most commonly on the back, abdomen, and chest in males, and on the lower leg, hip, and thigh in females (2). Females also appear to have lower risk of metastases and longer melanoma-specific survival than males (4).

Genetic factors are known to play a role in melanoma. Several high-penetrance loci for melanoma risk have been identified in melanoma-prone families, particularly cyclin-dependent kinase inhibitor 2A (*CDKN2A,* coding for p16 and p14ARF) and cyclin-dependent kinase 4 (*CDK4*) (5). However these high-penetrance familial syndromes are only a small proportion of overall population melanoma burden, with approximately 10% of melanoma cases occurring in multiplex families (6). In sporadic disease, genome-wide association studies (GWAS) have successfully identified at least 8 susceptibility loci for melanoma, with the melanocortin-1 receptor (*MC1R*) gene being the best replicated (5). Our study aims to replicate existing GWAS findings in order to further evaluate their association with melanoma risk.

Several cancer susceptibility loci that have been identified in GWAS, such as the 8q24 and *TERT-CLPTM1L* loci, have also been associated with numerous other cancers (7, 8). Variants in the *TERT-CLPTM1L* region, for example, have been associated with basal cell carcinoma, melanoma, and glioma, as well as lung, bladder, prostate, pancreatic, and cervical cancers (8-10). This provides evidence of pleiotropy, where a single genotype is associated with multiple phenotypes. The existence of pleiotropic effects at these loci suggests that there may be common mechanisms of carcinogenesis or disease susceptibility pathways across these phenotypes. Such information can be useful for elucidating pathogenic mechanisms, improving disease classification, or targeting therapeutic intervention. While identifying pleiotropy is important, the extent of pleiotropy has not been comprehensively explored. This study aims to evaluate SNPs associated with various cancers in previous GWAS for an additional pleiotropic association with melanoma. As incidence, mortality, survival, and anatomic location of melanoma have been shown to vary by sex, this study also evaluates whether any of these genetic associations may vary by sex as well.

Methods

Study populations

We analyzed 2,213 melanoma cases and 21,309 melanoma-free controls from five study populations. Three of these studies collaborated through their participation in the Population Architecture using Genetics and Epidemiology (PAGE) study (11): the Multiethnic Cohort (MEC) (12); the Women's Health Initiative (WHI) (13, 14); and Epidemiological Architecture for Genes Linked to Environment (EAGLE), which is based on data from 3 National Health and Nutrition Examination Surveys (NHANES) and the Biorepository of Vanderbilt University

(BioVU) (15). Two non-PAGE studies also contributed: the Nurses' Health Study (NHS) (16, 17) and the Health Professionals Follow-up Study (HPFS). NHS and WHI are female-only studies, HPFS is a male-only study, and BioVU and MEC have both male and female participants. Additional details on each of these studies are provided in Appendix 1.

In the PAGE studies, invasive melanoma cases were defined as incident cases of melanoma in participants without a previous cancer diagnosis (except for non-melanoma skin cancer). Melanoma in situ cases were excluded. Controls were participants without any history of melanoma. Additional demographic and epidemiologic information was obtained according to individual study protocols (see Appendix 1). Due to low sample size in other race/ethnicity groups, we restricted our analysis to participants of European ancestry (self-reported Caucasian, clustering with HapMap CEU).

In NHS and HPFS, participants were gathered from previously conducted GWAS on various disease outcomes (NHS: breast cancer, coronary heart disease, type 2 diabetes, kidney stone, pancreatic cancer, glaucoma; HPFS: coronary heart disease, type 2 diabetes, kidney stone, advanced prostate cancer, glaucoma). Controls from these studies (and cases and controls from the kidney stone GWAS) were used for a melanoma GWAS. In this analysis, participants with a melanoma diagnosis were used as cases and participants without a melanoma diagnosis were used as controls. Additional participants with a melanoma diagnosis in both cohorts who were not included in these previous GWAS were also included. Eligible cases had pathologically confirmed invasive melanoma diagnosed between baseline and the 2008 follow-up cycle. Controls had no reported melanoma diagnosis.

SNP selection and genotyping

In PAGE, a panel of 189 SNPs associated with various cancer outcomes was selected and genotyped. These SNPs were chosen based on published GWAS literature as of 2010 (11), as well as SNPs reported as associated with cancer in the National Human Genome Research Institute GWAS catalog (18). Each PAGE study genotyped a subset of this panel in order to maximize replication and generalization opportunities according to the characteristics of their study population. The risk allele for each SNP was determined based on prior literature, and was defined as the allele associated with an increased risk of cancer. See Appendix 2 for the risk allele and originally associated cancer site for each SNP.

Standard quality assurance and quality control measures were utilized to ensure genotyping quality. In PAGE, samples and SNPs were included based on call rates (≥ 90%), concordance of blinded replicates (> 98%), and no strong evidence of Hardy-Weinberg equilibrium ($p < 0.001$). Each laboratory also genotyped 360 HapMap samples to serve as cross-laboratory and cross-platform quality control samples (19).

In NHS and HPFS, participants had been previously genotyped in nested case-control GWAS of various outcomes. For the melanoma GWAS, >2.5 million SNPs were imputed based on NCBI build 35 of phase II HapMap CEU data using MACH. Only SNPs with an imputation quality $r2>0.95$ in each study were included. Genotype information for the panel of 189 SNPs assembled by PAGE was available from this existing GWAS data.

Statistical analyses

For each study we estimated the association between individual SNPs and risk of melanoma using unconditional logistic regression. SNPs were coded additively with 0, 1, 2 referring to the number of purported risk alleles (or the dosage for imputed SNPs). The risk

allele was defined as the allele that increased the risk of cancer in the initial GWAS publication. Models were adjusted for age (HPFS, NHS, and WHI) or age and sex (BioVU and MEC). In NHS and HPFS, models were also adjusted for the top five GWAS-derived eigenvectors, using EIGENSTRAT (20), to account for population substructure. Since the three PAGE studies used ancestry informative markers (AIMs (21)) instead of GWAS-derived markers, and participants were restricted to those of European ancestry, we did not adjust for principle components in these three studies.

Study-specific regression estimates were combined across studies using inverse-variance weighted, fixed-effect meta-analysis. NHS and HPFS study-specific results had already been combined in a meta-analysis as part of a melanoma GWAS prior to transmission to PAGE, and so were received and treated as a single dataset for meta-analysis with the PAGE study results. We calculated the heterogeneity p-values based on Cochran's Q statistic. Analyses were performed using Stata version 12 (22). Because of multiple testing, we used a Bonferroni-corrected p-value ($0.05 / 189 = 2.6 \times 10^{-4}$) to test the statistical significance of the overall association for each SNP with melanoma. In order to evaluate for potential sex-specific genetic effects, we also evaluated the association between each SNP and melanoma risk stratified by sex. Because NHS- and HPFS-specific results were not available separately, only PAGE studies were used for sex-stratified analyses. We performed meta-regression to obtain p-heterogeneity values for the difference between sex-specific regression estimates.

Results

Demographic and epidemiologic characteristics of the study populations are provided in Table 1. Since NHS and WHI are female-only studies, we had roughly twice as many females than males in the overall analysis. Melanoma cases tended to be younger than controls.

In total we evaluated 189 cancer GWAS findings, including 10 previously associated with melanoma (8 from melanoma GWAS and 2 from lung cancer GWAS that were later associated with melanoma). Of these 10 SNPs, 9 replicated ($p < 3.8x10^{-5}$), while one showed suggestive evidence ($p = 0.02$; Table 2 and Figure 1). Five of these SNPs showed a modest increase in melanoma risk (OR = 1.17 – 1.31), while rs258322 (*CKD10,* OR = 1.55) and rs16891982 (*SLC45A2,* OR = 3.11) showed larger effects. Of these top findings previously associated with melanoma, two were originally found in GWAS of lung cancer: rs4975616 and rs401681, both in the *TERT/CLPTM1L* locus. Though we modeled the alleles previously associated with an increased risk of lung cancer, both of these SNPs demonstrated a modest reduction in melanoma risk (OR = 0.87), consistent with previous studies (8, 23). In addition to association with melanoma risk, four of our top SNPs have also previously been associated with various pigmentation phenotypes: rs4785763 (near *AFG3L1P*), rs16891982 (*SLC45A2*), rs1393350 (*TYR*), and rs7023329 (near *MTAP*).

Among the non-melanoma GWAS SNPs, no SNP was below our Bonferroni-corrected statistical significance threshold of $2.6x10^{-4}$, though 11 SNPs had p-values below 0.05. These SNPs were previously associated with a handful of other cancers (Table 3). Eight of these eleven SNPs showed an increased risk for melanoma in the same direction as the previously associated cancer (OR = 1.09 – 1.23). The other three of these SNPs showed a decreased risk of

melanoma: two different *TERT/CLPTM1L* SNPs (rs402710 and rs31389) previously associated with lung cancer (OR = 0.87 – 0.89) and one *ABO* SNP previously associated with pancreatic cancer (OR = 0.90). Due to multiple testing, some (or all) of these findings could be due to chance (expect 0.05*189=9.45), though some SNPs are correlated and may not represent independent tests. Results for all SNPs are provided in Appendix 3.

In the sex-stratified analyses, 6 SNPs in males and 5 SNPs in females passed our Bonferroni significance threshold (Supplemental Table 3). Each of these SNPs had also demonstrated an association in the overall analysis except for one: rs12418451, near *TPCN2*. This SNP reached statistical significance in males ($p = 1.92x10^{-4}$) but not in females or overall (p-heterogeneity = 0.005, Table 4), and showed a larger effect in males than females (OR = 1.29 and 1.01, respectively). Four other nearby but uncorrelated ($r^2 < 0.13$ in CEU) SNPs in this region also showed a trend of stronger effects in males than females, though only one suggested a statistically significant difference (rs11228565, p-heterogeneity = 0.046). Another SNP, rs16891982 (*SLC45A2*), also showed a potential difference in effect by sex (p-heterogeneity = 0.02), which again was stronger in males than females (OR = 5.50 and 2.37, respectively). Sex-stratified results for all SNPs are provided in Appendix 4.

Discussion

We replicated an association with melanoma for seven melanoma GWAS SNPs, several of which have been previously associated with pigmentation traits. We also replicated a known association with melanoma for two SNPs in the *TERT-CLPTM1L* region previously shown to demonstrate pleiotropic effects in opposing directions, with decreased risk for melanoma but increased risk for other cancers. Additionally, we identified one SNP in *TPCN2* which

demonstrated a statistically significant association in male-specific analyses, showing potential sex-specific pleiotropic associations with both melanoma and prostate cancer.

Of the 7 melanoma GWAS SNPs we found significantly associated with melanoma, four had previously also shown potentially pleiotropic associations with pigmentation traits such as hair, eye, or skin color. While pigmentation traits are known to influence melanoma risk, the relationship between genetic variants, pigmentation phenotypes, and melanoma risk are not clear. Many variants associated with pigmentation have also been associated with melanoma risk, but not all (24). Additionally, several pigmentation-related variants (in *MC1R* and *TYR*) appear to contribute to skin cancer etiology independent of pigmentation phenotypes (25, 26). Biologically, pigmentation traits are largely determined by the levels of two forms of melanin: the darker black/brown eumelanin and the lighter red/yellow pheomelanin (25). Eumelanin provides better protection from DNA-damage-inducing ultraviolet radiation than pheomelanin, and thus lighter pigmentary phenotypes are thought to comparatively increase melanoma risk due to decreased UV shielding. However, a recent mouse study suggested that pigmentary *MC1R* variants may provide a risk for melanoma independent of ultraviolet radiation exposure (27). This study hypothesized that the process of pheomelanin synthesis itself produces oxidative DNA damage, which can lead towards melanoma (27). Since other genes involved in pigmentary processes can also impact the levels of pheomelanin synthesis (25), these genes could feasibly also contribute to UV-independent melanoma risk through this pathway. As such, some of the SNPs we found significantly associated with melanoma may demonstrate potentially independent pleiotropic relationships with pigmentation phenotypes. Others have previously hypothesized that multiple pathways may lead to melanoma, and various genetic factors may impact one or both of these pathways.

Two of the SNPs we found significantly associated with melanoma are located in the *TERT-CLPTM1L* locus, which contains variants associated with a number of different cancers. The pleiotropic effects of variants in this region have been well established (8-10), and our findings are consistent with previous reports associating cancer risk variants in this region with decreased risk of melanoma (8, 23, 28). One of these SNPs, rs401681, has been associated with an increased risk of lung cancer (29), basal cell carcinoma (28), bladder cancer, prostate cancer, and cervical cancer (8). This same SNP has also been associated with a decreased risk of melanoma (8, 28) and pancreatic cancer (30). The other SNP in this region significant in our study, rs4975616, has also been previously associated with an increased risk of lung cancer (29, 31) and a decreased risk of melanoma (23). While not reaching our Bonferroni cutoff, two other SNPs in this region were also associated with decreased risk of melanoma in our study: rs402710 ($p = 7.74x10^{-4}$) and rs31489 ($p = 4.18x10^{-3}$). Both have previously been associated with increased risk of lung cancer (9). These four SNPs are all located within the *CLPTM1L* gene and are in relatively high linkage disequilibrium with each other ($r^2 > 0.57$; from 1000 genome pilot CEU data using SNAP (32)). Two nearby SNPs within the *TERT* gene were not associated with melanoma ($p > 0.39$), and were not correlated with any of the four *CLPTM1L* SNPs ($r^2 < 0.07$). Taken together, our findings provide further evidence of pleiotropic effects in opposite directions in the *TERT-CLPTM1L* region, where variants associated with increased risk for lung and other cancers are simultaneously associated with reduced melanoma risk.

The action of the *TERT* and *CLPTM1L* genes provides biological plausibility for the opposing pleiotropy seen for different cancers. *TERT* (telomerase reverse transcriptase) has a well-known role in telomere and tumor biology (9, 10), while expression of *CLPTM1L* (cleft lip and palate transmembrane protein 1-like; also referred to as cisplatin resistance related (CRR9))

has been shown to sensitize ovarian cancer cells to cisplatin-induced apoptosis (8, 33). Since both genes are within a 62-kb linkage disequilibrium block, currently identified SNPs could be tagging for nearby unknown functional variants that impact either gene. Variants in the *TERT-CLPTM1L* region such as rs401681 may lead to a more rapid shortening of the telomeres over time (8). While shorter telomeres have been associated with increased risk of several cancers, such as basal cell carcinoma (34) and lung cancer (35), longer telomeres have been shown to increase risk for melanoma (34) and nevi (36). These observations are consistent with the opposing effects seen in the genetic findings. The opposing pleiotropy seen in this region has led to hypotheses that no common molecular pathway leads to the development of all cancer types, and that some pathways friendly to the development of one cancer may be hostile to others (10).

In our sex-stratified analyses we identified two SNPs which demonstrated a stronger genetic effect in males than in females: rs12418452 (p-heterogeneity = 0.005) and rs16891982 (p-het = 0.02). Both of these SNPs are located in or near solute carrier genes involved in melanosome function, though they transport different molecules. SNP rs16891982 demonstrated an association that met Bonferroni-corrected significance in both males and females ($p = 9.5x10^{-8}$ and $4.7x10^{-7}$ respectively), but showed differences in the strength of the association (OR = 5.50 and 2.37, respectively). This SNP, located in *SLC45A2*, has previously been associated with both melanoma (24, 37, 38) and pigmentation traits such as skin and hair color (39). *SLC45A2*, also known as *MATP*, encodes an ion transporter protein in the melanosome, and is also the pathogenic gene for oculocutaneous albinism type 4. These and other melanosome ion transporter proteins have demonstrated the functional importance of ion and small molecule transport to melanogenesis and the pigmentation pathway (25, 40). Ion exchange is also predicted to impact melanogenesis by playing an important role in regulating melanosome pH

levels (41). As such, variants in other ion transport genes similar to *SLC45A2* may also be expected to impact pigmentation and melanoma risk.

The other SNP, rs12418451, demonstrated a significant association with melanoma in males (OR = 1.29, $p = 1.9x10^{-4}$), but not in females (OR = 1.01, p = 0.92) or the overall analysis (OR = 1.11, $p = 4.4x10^{-3}$). This SNP has previously been associated with prostate cancer (42), and is located ~77kb downstream of *TPCN2* and ~126kb upstream of *MYEOV*. The proximity of this SNP to these other genes provides biological plausibility for an association with melanoma. The nearby *TPCN2* (two-pore segment channel 2) encodes a putative cation-selective ion channel that releases Ca^{2+} from acidic organelles (43), which similarly to *SLC45A2* may impact melanogenesis through pH regulation (41). Indeed, two other variants in *TPCN2* have been associated with pigmentation traits (blond versus brown hair color (44)). A later study did not find either of these two *TPCN2* SNPs to be associated with melanoma ($p > 0.12$), though they did not stratify by sex (26). This SNP is also ~126 kb upstream of *MYEOV*, an oncogene that includes variants implicated for multiple cancers, including multiple myeloma, breast cancer, colon cancer, and esophageal squamous cell carcinoma (45). A proxy of rs12418451 is also one of three independent loci in this region associated with prostate cancer (46). Another study evaluating this region for prostate cancer identified an interaction between rs12418451 and rs784411 in *CEP152*, a centrosomal protein shown to function as a regulator of genomic integrity and cellular response to DNA damage (45). In our study, a second SNP in this region (rs7117034, ~117kb downstream of *TPCN2*) was also marginally associated with melanoma risk overall ($p = 3.7x10^{-4}$). While this SNP demonstrated a similar trend of a stronger effect in males than females (OR = 1.26 and 1.18, respectively), this difference was not statistically significant (p-het = 0.60). Together, our findings identify a potentially novel pleiotropic finding for a sex-

specific association between rs12418452 and melanoma and highlight a new locus with plausible biologic function.

The strengths of this study stem from the collaboration of five large studies, which provide sizable samples to evaluate the association of melanoma with cancer GWAS SNPs. A potential limitation is that three of these studies were conducted only in males (HPFS) or females (NHS, WHI), which reduced the sample size for sex-specific analyses. Since not all SNPs were available in all studies, sample sizes also varied by SNP in both the overall and sex-specific analyses depending on which studies had that particular SNP available. These differences in sample size may have reduced our ability to detect an association with melanoma for some SNPs. However, 97% of SNPs were available in at least two studies (and 77% in at least three), and most overall analyses were large (mean number of participants available per SNP 15,278, range 1,925 – 22,141).

In summary, we successfully replicated most of the previous melanoma findings we evaluated, several of which had previously shown pleiotropic relationships with pigmentation traits. Variants in the *TERT-CLPTM1L* locus demonstrated pleiotropic effects in opposite directions, where the allele previously associated with increased risk of lung and other cancers demonstrated an association with decreased risk of melanoma in our study. Additionally, we were able to provide some evidence for two SNPs near solute-carrier genes that showed potential differences in effect by sex, with larger effects in males than females. One of these SNPs, in *TPCN2*, demonstrated a potentially pleiotropic effect on melanoma as well as the previously-associated risk of prostate cancer. While these findings are suggestive, they provide interesting candidates for follow-up studies.

References

1. Howlader N NA, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations). National Cancer Institute, Bethesda, MD [based on November 2011 SEER data submission, posted to the SEER website, April 2012]; Available from: http://seer.cancer.gov/csr/1975_2009_pops09/.
2. Schottenfeld D, Joseph F. Fraumeni J. Cancer Epidemiology and Prevention. 3 ed. New York, NY: Oxford University Press, USA; 2006. 1416 p.
3. Boniol M, Autier P, Boyle P, Gandini S. Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis. BMJ. 2012;345:e4757. Epub 2012/07/27.
4. Joosse A, de Vries E, Eckel R, Nijsten T, Eggermont AM, Holzel D, et al. Gender differences in melanoma survival: female patients have a decreased risk of metastasis. The Journal of investigative dermatology. 2011;131(3):719-26. Epub 2010/12/15.
5. Gerstenblith MR, Shi J, Landi MT. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. Pigment cell & melanoma research. 2010;23(5):587-606. Epub 2010/06/16.
6. Falchi M, Bataille V, Hayward NK, Duffy DL, Bishop JA, Pastinen T, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. Nature genetics. 2009;41(8):915-9. Epub 2009/07/07.
7. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. Journal of the National Cancer Institute. 2008;100(13):962-6. Epub 2008/06/26.
8. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nature genetics. 2009;41(2):221-7. Epub 2009/01/20.
9. Baird DM. Variation at the TERT locus and predisposition for cancer. Expert reviews in molecular medicine. 2010;12:e16. Epub 2010/05/19.
10. Mocellin S, Verdi D, Pooley KA, Landi MT, Egan KM, Baird DM, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. Journal of the National Cancer Institute. 2012;104(11):840-54. Epub 2012/04/24.
11. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, et al. The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. American journal of epidemiology. 2011;174(7):849-59. Epub 2011/08/13.
12. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. American journal of epidemiology. 2000;151(4):346-57. Epub 2000/03/01.
13. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. Annals of epidemiology. 2003;13(9 Suppl):S18-77. Epub 2003/10/25.
14. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Controlled clinical trials. 1998;19(1):61-109. Epub 1998/03/11.
15. Bush W, Boston J, Pendergrass S, dumitrescu L, Goodloe R, Brown-Gentry K, et al. Enabling high-throughput genotype-phenotype associations in the Epidemioloic Architecture for

Genes Linked to Environment (EAGLE) project as part of the Populations Architecture using Genomics and Epidemiology (PAGE) study. *Pacific Symposium on Biocomputing*. 2013(18):373-84.
16. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. The American journal of nursing. 1978;78(6):1039-40. Epub 1978/06/01.
17. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. Nature reviews Cancer. 2005;5(5):388-96. Epub 2005/05/03.
18. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(23):9362-7. Epub 2009/05/29.
19. International HapMap C. The International HapMap Project. Nature. 2003;426(6968):789-96. Epub 2003/12/20.
20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006;38(8):904-9. Epub 2006/07/25.
21. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Human mutation. 2009;30(1):69-78. Epub 2008/08/08.
22. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.
23. Law MH, Montgomery GW, Brown KM, Martin NG, Mann GJ, Hayward NK, et al. Meta-analysis combining new and existing data sets confirms that the TERT-CLPTM1L locus influences melanoma risk. The Journal of investigative dermatology. 2012;132(2):485-7. Epub 2011/10/14.
24. Duffy DL, Zhao ZZ, Sturm RA, Hayward NK, Martin NG, Montgomery GW. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. The Journal of investigative dermatology. 2010;130(2):520-8. Epub 2009/08/28.
25. Scherer D, Kumar R. Genetics of pigmentation in skin cancer--a review. Mutation research. 2010;705(2):141-53. Epub 2010/07/06.
26. Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nature genetics. 2008;40(7):886-91. Epub 2008/05/20.
27. Mitra D, Luo X, Morgan A, Wang J, Hoang MP, Lo J, et al. An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. Nature. 2012;491(7424):449-53. Epub 2012/11/06.
28. Stacey SN, Sulem P, Masson G, Gudjonsson SA, Thorleifsson G, Jakobsdottir M, et al. New common variants affecting susceptibility to basal cell carcinoma. Nature genetics. 2009;41(8):909-14. Epub 2009/07/07.
29. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nature genetics. 2008;40(12):1407-9. Epub 2008/11/04.
30. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on

chromosomes 13q22.1, 1q32.1 and 5p15.33. Nature genetics. 2010;42(3):224-8. Epub 2010/01/27.
31. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer research. 2009;69(16):6633-41. Epub 2009/08/06.
32. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24):2938-9. Epub 2008/11/01.
33. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochemical and biophysical research communications. 2001;280(4):1148-54. Epub 2001/02/13.
34. Han J, Qureshi AA, Prescott J, Guo Q, Ye L, Hunter DJ, et al. A prospective study of telomere length and the risk of skin cancer. The Journal of investigative dermatology. 2009;129(2):415-21. Epub 2008/08/01.
35. Wu X, Amos CI, Zhu Y, Zhao H, Grossman BH, Shay JW, et al. Telomere dysfunction: a potential cancer predisposition factor. Journal of the National Cancer Institute. 2003;95(16):1211-8. Epub 2003/08/21.
36. Bataille V, Kato BS, Falchi M, Gardner J, Kimura M, Lens M, et al. Nevus size and number are associated with telomere length and represent potential markers of a decreased senescence in vivo. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2007;16(7):1499-502. Epub 2007/07/14.
37. Guedj M, Bourillon A, Combadieres C, Rodero M, Dieude P, Descamps V, et al. Variants of the MATP/SLC45A2 gene are protective for melanoma in the French population. Human mutation. 2008;29(9):1154-60. Epub 2008/08/08.
38. Fernandez LP, Milne RL, Pita G, Aviles JA, Lazaro P, Benitez J, et al. SLC45A2: a novel malignant melanoma-associated gene. Human mutation. 2008;29(9):1161-7. Epub 2008/06/20.
39. Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, et al. A genomewide association study of skin pigmentation in a South Asian population. American journal of human genetics. 2007;81(6):1119-32. Epub 2007/11/14.
40. Sturm RA. Molecular genetics of human pigmentation diversity. Human molecular genetics. 2009;18(R1):R9-17. Epub 2009/03/20.
41. Kondo T, Hearing VJ. Update on the regulation of mammalian melanocyte function and skin pigmentation. Expert review of dermatology. 2011;6(1):97-108. Epub 2011/05/17.
42. Zheng SL, Stevens VL, Wiklund F, Isaacs SD, Sun J, Smith S, et al. Two independent prostate cancer risk-associated Loci at 11q13. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2009;18(6):1815-20. Epub 2009/06/10.
43. Calcraft PJ, Ruas M, Pan Z, Cheng X, Arredouani A, Hao X, et al. NAADP mobilizes calcium from acidic organelles through two-pore channels. Nature. 2009;459(7246):596-600. Epub 2009/04/24.
44. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. Nature genetics. 2008;40(7):835-7. Epub 2008/05/20.

45. Tao S, Wang Z, Feng J, Hsu FC, Jin G, Kim ST, et al. A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants. Carcinogenesis. 2012;33(3):598-603. Epub 2012/01/06.
46. Chung CC, Ciampa J, Yeager M, Jacobs KB, Berndt SI, Hayes RB, et al. Fine mapping of a region of chromosome 11q13 reveals multiple independent loci associated with risk of prostate cancer. Human molecular genetics. 2011;20(14):2869-78. Epub 2011/05/03.

Part A, Analysis II – Melanoma Table 1

Table 1 - Demographic characteristics of study populations utilized

		EAGLE-BioVU		MEC		NHS/HPFS		WHI		Total	
		Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
# Participants		742	8063	240	2032	494	5628	737	5586	2213	21309
Sex	Males	445	4351	149	1059	177	2251	0	0	771	7661
	Females	297	3712	91	973	317	3377	737	5586	1442	13648
Age	Median (SE)			37 (9.5)	69 (8.5)			68 (7.3)	75 (7.4)		

Part A, Analysis II – Melanoma Table 2

Table 2 - SNPs demonstrating a statistically significant association with melanoma in overall meta-analyses (Bonferroni-corrected p < 2.6x10-4).

SNP	Gene	Chromosome / Risk allele	Original GWAS finding	Beta	SE	P-value	OR	95% CI	n	# Studies	P-heterogeneity
rs258322	*CDK10*	16 / A	Melanoma	**0.44**	**0.05**	**8.81E-19**	1.55	(1.41 - 1.71)	22082	5	0.60
rs4785763	*AFG3L1P* (near)	16 / A	Melanoma	**0.27**	**0.04**	**1.29E-14**	1.31	(1.22 - 1.40)	21993	5	0.80
rs16891982	*SLC45A2*	5 / C	Melanoma	**1.13**	**0.15**	**7.39E-14**	3.11	(2.31 - 4.18)	15949	3	0.43
rs1393350	*TYR*	11 / A	Melanoma	**0.23**	**0.04**	**5.21E-10**	1.25	(1.17 - 1.35)	22009	5	0.99
rs4636294	*MTAP* (near)	9 / A	Melanoma	**0.17**	**0.03**	**5.04E-07**	1.19	(1.11 - 1.27)	22053	5	0.12
rs7023329	*MTAP* (near)	9 / A	Melanoma	**0.16**	**0.03**	**2.03E-06**	1.17	(1.10 - 1.25)	22114	5	0.28
rs4975616	*TERT/CLPTM1L*	5 / A	Lung cancer	**-0.14**	**0.03**	**2.04E-05**	0.87	(0.81 - 0.93)	22135	5	0.65
rs910873	*PIGU*	5 / A	Melanoma	**0.27**	**0.06**	**2.46E-05**	1.31	(1.15 - 1.48)	22059	3	1.00
rs401681	*TERT/CLPTM1L*	5 / C	Lung cancer	**-0.14**	**0.03**	**3.26E-05**	0.87	(0.81 - 0.93)	22109	5	0.65

Part A, Analysis II – Melanoma Table 3

Table 3 - SNPs demonstrating a marginal association with melanoma in overall meta-analyses (0.05 > p > 2.4e-4)

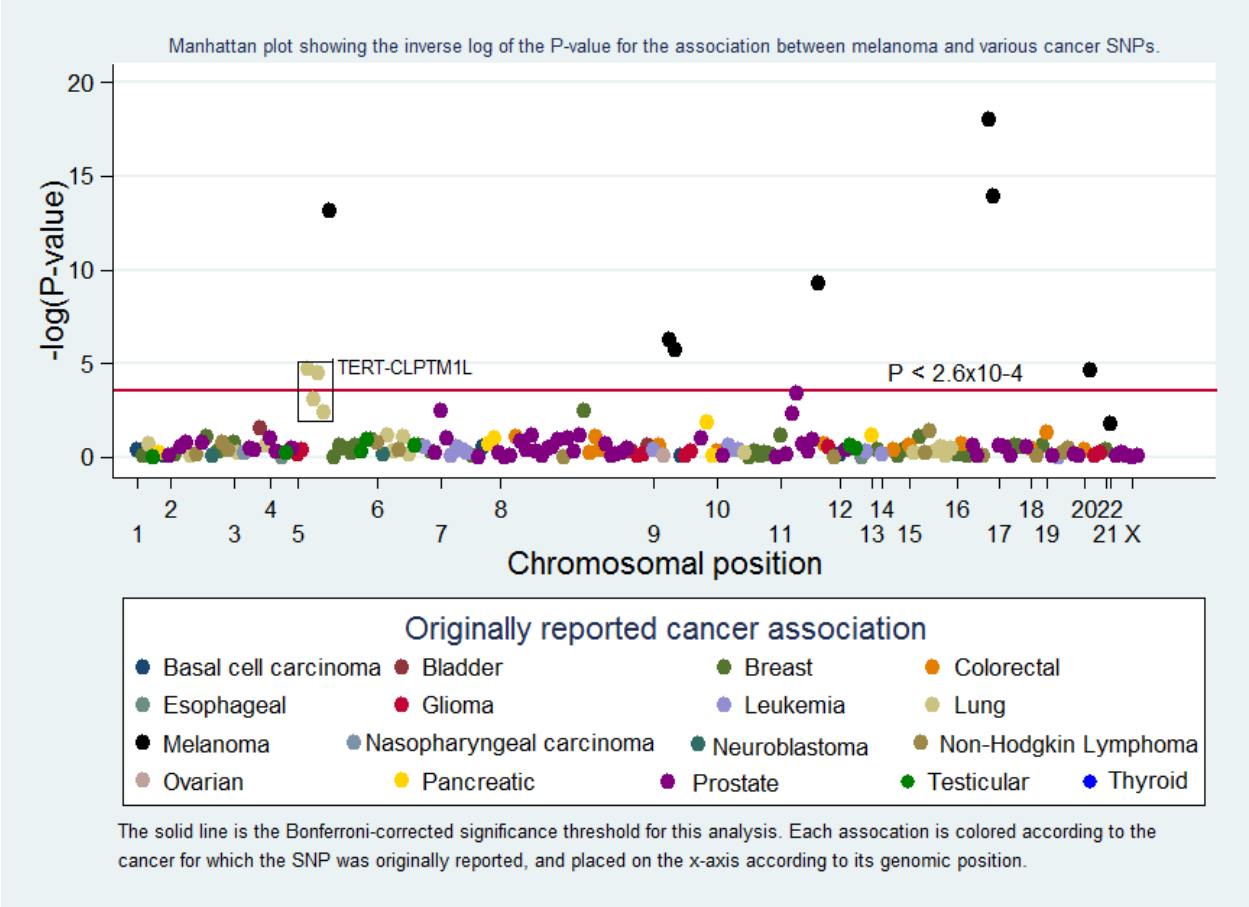
SNP	Gene	Chromosome / Coded allele	Previous trait association	OR	95% CI	P-value	# Studies	n	P-heterogeneity
rs7117034	*TPCN2, MYEOV* (near)	11 / T	Prostate cancer	1.23	(1.10 - 1.37)	3.67E-04	3	16797	0.55
rs402710	*CLPTM1L* (intronic)	5 / C	Lung cancer	0.87	(0.81 - 0.94)	7.74E-04	4	22113	0.04
rs12155172	*ABCB5 (near)*	7 / A	Prostate cancer	1.19	(1.06 - 1.33)	3.38E-03	3	16825	0.61
rs13281615	*MYC, POU5F1B* (near)	8 / T	Breast cancer	1.11	(1.03 - 1.18)	3.57E-03	5	22138	0.15
rs31489	*CLPTM1L* (intronic)	5 / C	Lung cancer	0.89	(0.82 - 0.96)	4.18E-03	3	20126	0.33
rs12418451	*TPCN2, MYEOV* (near)	11 / A	Prostate cancer	1.11	(1.03 - 1.19)	4.39E-03	5	22053	0.24
rs505922	*ABO* (intronic)	9 / C	Pancreatic cancer	0.90	(0.82 - 0.98)	0.015	4	13339	0.25
rs2284063	*PLA2G6* (intronic)	22 / G	Melanoma	1.09	(1.02 - 1.17)	0.017	5	22087	0.29
rs710521	*LEPREL1, TP63* (near)	3 / A	Urinary bladder cancer	1.13	(1.01 - 1.27)	0.028	4	13302	0.55
rs7176508	*RPLP1, GEMIN8P1* (near)	15 / A	Chronic lymphocytic leukemia	1.10	(1.00 - 1.21)	0.044	2	5273	0.47
rs10411210	*RHPN2* (intronic)	19 / C	Colorectal cancer	1.14	(1.00 - 1.30)	0.046	3	15971	0.68

Part A, Analysis II – Melanoma Table 4

Table 4 - SNPs reaching a statistically significnant association with melanoma in sex-stratified analyses (p < 2.6x10-4) that also demonstrate a difference in effect by sex (P-het < 0.05). For additional SNPs near TPCN2/MYEOV are also shown.

									P-heterogeneity	
SNP	Gene, Chromosome / coded allele, Previous trait	Sex	OR	95% CI	P-value	# Stu	n	Between-studies	Between-sexes
rs16891982	*SLC45A2*	Overall	**3.11**	**(2.31 - 4.18)**	**7.39E-14**	3	15949	0.43	
	5 / C	Male	**5.50**	**(2.94 - 10.28)**	**9.53E-08**	2	5789	0.34	**0.02**
	Melanoma	Female	**2.37**	**(1.69 - 3.31)**	**4.67E-07**	3	10160	0.45	
rs12418451	*TPCN2, MYEOV* (near)	Overall	1.11	(1.03 - 1.19)	4.39E-03	4	15931	0.24	
	11 / A	Male	**1.29**	**(1.13 - 1.48)**	**1.92E-04**	2	5785	0.52	**0.005**
	Prostate cancer	Female	1.01	(0.90 - 1.12)	0.92	3	10146	0.69	
rs11228565	*TPCN2, MYEOV* (near)	Overall	1.06	(0.96 - 1.17)	0.22	2	14026	4.38E-04	
	11 / A	Male	1.21	(1.03 - 1.42)	0.02	1	4756	.	**0.046**
	Prostate cancer	Female	0.99	(0.87 - 1.12)	0.84	2	9270	3.53E-03	
rs10896449	*TPCN2, MYEOV* (near)	Overall	1.05	(0.99 - 1.13)	0.12	4	15970	0.41	
	11 / G	Male	1.12	(0.99 - 1.27)	0.07	2	5793	0.97	0.31
	Prostate cancer	Female	1.04	(0.94 - 1.14)	0.46	3	10177	0.22	
rs7117034	*TPCN2, MYEOV* (near)	Overall	1.23	(1.10 - 1.37)	3.67E-04	2	10675	0.55	
	11 / T	Male	1.26	(1.09 - 1.45)	2.07E-03	2	5775	0.20	0.60
	Prostate cancer	Female	1.18	(0.99 - 1.42)	0.07	2	4900	0.44	
rs7931342	*TPCN2, MYEOV* (near)	Overall	1.03	(0.95 - 1.12)	0.47	3	7214	0.49	
	11 / G	Male	1.09	(0.82 - 1.44)	0.55	1	1023	.	0.74
	Prostate cancer	Female	1.04	(0.92 - 1.17)	0.53	3	6191	0.18	

Part A, Analysis II – Melanoma Figure 1



The solid line is the Bonferroni-corrected significance threshold for this analysis. Each assocation is colored according to the cancer for which the SNP was originally reported, and placed on the x-axis according to its genomic position.

Part B – Impact of pleiotropy on the return of results to research participants

Title: Implications of pleiotropy for the return of results

Abstract

Recent improvements in genotyping technology, knowledge of the relationships between genetics and disease, and advances in genetic medicine have led to debate over whether genetic results obtained in research settings should be returned to individual participants. Several groups have put forward guidelines that attempt to sort genetic results into distinct categories, each carrying recommendations on whether or not they should be returned. These categories are primarily determined by whether genetic results meet threshold criteria for clinical validity and clinical utility, and those with the strongest "net benefit" are prioritized for return. Such a categorization scheme is useful for evaluating single genotype-phenotype relationships to determine if the result should be returned. However, these guidelines do not address the situation where a single result can have multiple meanings. The term pleiotropy describes a single gene which has multiple phenotypic effects. Pleiotropic genetic variants may have phenotypic relationships that are categorized differently, potentially leading to conflicting return recommendations for the same genetic result. Current guidelines do not address such potential conflicts, and may break down for pleiotropic results.

This chapter explores the ways in which the existence of pleiotropy impacts the return of results policy discussion. To provide background, this chapter describes current return of result guidelines, defines and provides pertinent examples of pleiotropy, and offers an estimate of the extent of pleiotropy that might be expected. Following this, several examples of how pleiotropy

impacts the return of results are provided. Key among these is that the existence of pleiotropic relationships means that returning any result may directly or indirectly also return additional information. Because of this, deciding to return any information must include consideration of all additional pleiotropic relationships. While focused on return in the research setting, potential impacts of pleiotropy in the clinical and consumer settings are also discussed.

An example framework for weighing the multiple phenotypic relationships of a pleiotropic variant is provided. In this framework, relationships of strong likely benefit or harm may outweigh other pleiotropic relationships, with considerations made for participant preferences. While simplistic, this provides a beginning point for future augmentation. Several policy recommendations are made as well. Key among these are suggestions for consent processes that inform participants of the possibility of pleiotropic relationships, ascertain participant preferences in cases of conflicting return recommendation classifications, and allow participants to choose which results they want returned. Current guidelines will need to be amended to incorporate the possibility of pleiotropic variants into their classification scheme, which will otherwise impact their ability to make appropriate recommendations for whether a result should be returned.

Introduction

Research studies are generating increasing amounts of genetic information on research participants, with more and more studies incorporating exome and whole genome sequencing into their research. Since many studies utilize agnostic approaches, assessing a larger proportion of the genome increases the opportunities to identify genetic mutations important to the research study. However, casting a wider net also captures genetic variants previously associated with other diseases, bringing genetic information of potential clinical relevance into the research arena. When such variants are not part of the primary research goals of the study, these results are often described as secondary, ancillary, or incidental findings (1). Estimates of the extent of this so-called 'incidentalome' (2) in a single individual range from a handful (3, 4) to tens of thousands (5). Recent debate has focused on how researchers should handle such incidental findings, including whether any of these results should be returned to research participants (6).

Several sets of guidelines have proposed category-based schemes which sort and filter incidental findings into categories of return (7), which are sometimes labeled as 'bins' (8) or tiers (9) (see Figures 1, 2, and 3). These guidelines recommend that some categories of incidental findings should be reported while results in other categories should be withheld. This helps to prioritize returning the incidental findings in categories that meet certain threshold criteria. The primary criteria used to categorize a given finding are the clinical validity and clinical utility of the genotype-phenotype relationship for that particular result. Other considerations can include participant preferences (9); logistical issues such as lab certification, funding timelines, or what informed consent documentation was signed (10); and contextual factors such as participant age and reproductive history (7). For a given incidental finding, each of these criteria assesses a particular aspect of the genotype-phenotype relationship. Whether a particular category is

recommended to be returned may differ depending on the criteria emphasized by the recommending body.

Built around the evaluation of individual genotype-phenotype relationships, however, such 'binning' approaches are inadequately equipped to appropriately deal with pleiotropic genetic results. Pleiotropy refers to a single genotype affecting multiple phenotypes. When an incidental finding is pleiotropic, that genetic result may have genotype-phenotype relationships that fit into more than one bin, rather than fit neatly into any one single bin as assumed by current guidelines. This can have many potential ramifications, such as the way in which incidental findings are evaluated for whether they are appropriate for returning to research participants, and for the way in which pleiotropic incidental findings might be returned. While many current guidelines seek to prioritize reporting for only those incidental findings which are clinically actionable, pleiotropic relationships may force researchers to simultaneously return secondary information they otherwise would not have chosen to report on its own, either due to inconclusive evidence or because of potentially harmful information.

While focused here on the return of incidental findings in research settings, the return of primary findings in clinical settings will be impacted by pleiotropy as well. Targeted clinical genotyping results can still carry pleiotropic risk information outside the purposes of the test, which will need to be appropriately considered and managed. Results from direct-to-consumer scans are no less immune to the possibilities raised by pleiotropy. In each of these settings, such pleiotropic information impacts how results should be returned.

As such, this dissertation argues that current guidelines should be revised so as to appropriately deal with the potential pleiotropic relationships of genetic results. To this end,

several emerging examples of pleiotropy in the literature are offered, along with rough estimates of the biological extent of pleiotropy. With this background, it should become clear that the issue of incidental findings in research will not be fully resolved by simply creating more stringent filters or stronger criteria for which findings are returned. Instead, it is suggested that consent documents include warnings that genetic information can carry multiple meanings and solicit participant preferences for how to deal with pleiotropic risk information. As genetic knowledge improves, deciding on what incidental findings to return to research participants must move beyond evaluating single genotype-phenotype associations and move to more holistic approaches where the possibility of pleiotropic relationships are also considered.

Current guidelines for the return of results

As knowledge of the genetic underpinnings of disease improves, the question of whether or not genetic results should be returned to research participants has garnered increasing debate. Since genetic information can carry a broad spectrum of health implications, and because the number of genetic results associated with disease continues to grow, many guidelines have attempted to group genetic results into broad categories that are more easily approachable than addressing each result individually. Each of these categories carries a particular recommendation for how to deal with any genetic result contained in that category. Genetic results can be evaluated according to set criteria and assigned to one of these categories for return recommendations. There are generally three proposed categories: one for results that should be returned, one for results that may or may not be returned, and one for results that should not be returned (7). The ethical basis for the recommendations associated with these three broad categories stems from the concepts of beneficence, non-maleficence, and respect for persons (7, 11-13). Each of these categories attempts to provide recommendations in a way that

maximizes benefits and minimizes harms, while at the same time respecting participants without overburdening researchers.

The first category, which recommends that results should be returned, generally contains a small handful of results which must meet a high threshold for inclusion. Genetic results in this category pose a sufficiently serious risk of disease, which could be averted or reduced through screening or treatment, that their discovery incurs an ethical obligation to disclose that information—often termed a "duty to warn" (7) or "duty to rescue" (14). For genetic variants which fall in this first category, the benefits of return strongly outweigh most other considerations, leading to recommendations that they be returned. Outside of these few results, the majority of variants currently fall into the third category. Genetic results in this category are not known to have any clinical implications, and are mostly variants of uncertain significance. Since these variants provide little or no information of consequence, returning them is unlikely to provide any foreseeable benefit. On the other hand, returning such information could create harms such as unnecessary anxiety or false reassurances (11). For genetic variants in this third category, the potential harms outweigh most other considerations, leading to recommendations that they not be returned (7). Thus results in both the first and last categories have characteristics which make it clear whether or not they should be returned. For genetic results in the middle category, however, determining whether a result should be returned is often less straightforward. Variants in this middle category thus require a more detailed evaluation in weighing the possible benefits and harms of returning results. As such, most guidelines recommend that return of variants in this middle category be optional, with the decision dependent on factors such as participant preferences or the balance of clinical validity and clinical utility.

There are several characteristics often used to characterize the potential benefits and harms of particular genetic variants. The most widely emphasized of these is clinical utility (or actionability), which refers to whether the information provided by the genetic information can be utilized to treat or prevent some phenotypic harm (13). Results are more likely to be recommended for return if they can be acted upon in a way that is known to decrease risk or severity of disease. A strong measure of clinical utility is often needed for guidelines to require return (first category), while results with lower levels of clinical utility may be suggested to be optionally returned (middle category). A second major characteristic of a genetic result is clinical validity, which refers to the validity and strength of the association between the genetic result and a particular phenotype (13). High risk incidental information may be more or less likely to be returned than low risk incidental information, depending on the set of guidelines and the clinical utility of that variant. Most guidelines assume strong evidence of genotype-phenotype association and large effect size (high penetrance) in order for return to be appropriate. Though separate concepts, clinical utility and clinical validity are often considered together when categorizing results.

Another consideration is analytic validity, or the accuracy and reliability with which a genetic result is generated. Most guidelines suggest that results that are not analytically valid should not be returned, though some guidelines are more particular than others in what is necessary to meet this standard (such as whether or not the result was generated in a CLIA-certified setting (10)). Depending on the guidelines, additional considerations can also be assessed when determining whether or not a particular result is appropriate for return. For example, some guidelines also weigh personal meaning and utility (13), reproductive utility (7),

what informed consent was given (9, 10), the research context (14), or the timing of grant funding (10).

Several groups have released guidelines that use the above and other criteria to make recommendations for whether or not particular genetic results are appropriate to return in research settings (11). Some of these guidelines have been updated as genetic technology and knowledge improves, but others have not. The National Heart, Lung, and Blood Institute (NHLBI), for example, released an initial set of guidelines in 2004 (15), which were then revised in 2010 (10). A flowchart of the more recent recommendations is shown in Figure 1, which shows that a large number of hurdles must be surmounted before results can be returned, including the need for informed consent. A 2008 working group on incidental findings made similar recommendations for returning results in three broad categories of expected benefit, and also emphasized the potential importance of results for reproductive decision-making, as seen in Figure 2 (7). A later group expanded this concept by breaking the ambiguous middle category into smaller 'bins' representing various degrees of risk information, as seen in Figure 3 (8). Other authors have also followed this approach, calling for a qualified disclosure policy for different levels of return (9). A summation of additional recommendations from various government agencies can be seen in Figure 4 (7).

Of note, however, is that for each of these guidelines, the criteria used to place a particular result into one of these categories relies on a single genotype-phenotype relationship. Cases of pleiotropy, in which the same genetic result could have different phenotypic implications, are not addressed in these guidelines. In such cases applying the same evaluation criteria can lead to different, and potentially conflicting, return recommendations for the same genetic result. As such, pleiotropy produces possible problems where these guidelines provide

conflicting recommendations for different genotype-phenotype relationships for the same genetic result. This may be particularly problematic for guidelines where participant preferences are utilized for deciding which results to return, as this may also lead to conflicting participant preferences for different categories of return. In order to remain effective, future guidelines will need to be adapted to take pleiotropic variants into account.

Definition and extent of pleiotropy

Broadly defined, pleiotropy refers to the phenomenon of a single variant, gene, or genetic locus affecting multiple phenotypic traits. First coined over one hundred years ago, this broad definition reveals a concept with many different potential meanings (16, 17). That the definition of pleiotropy itself demonstrates pleiotropic characteristics is fitting. For the purposes of this dissertation, that there are multiple definitions of pleiotropy matters less than the fact that pleiotropy exists. Whatever the definition, it is recognized that some genotypes will have multiple phenotypic outcomes. Furthermore, several lines of evidence suggest that pleiotropy is widespread, and will be increasingly recognized as genetic knowledge improves.

The question of how much pleiotropy exists relies somewhat on the definition of pleiotropy used. Some have suggested a theory of universal pleiotropy, where any gene has the potential to affect all traits in some way (17). More recent hypotheses have focused on modular pleiotropy, where genes may have extensive pleiotropic effects within their immediate network, but may be of limited influence outside that network (16). Protein interaction networks in multiple animal models suggest that each gene affects four to seven proteins on average (16). These animal models also suggest an L-shaped distribution of pleiotropic effects, where a few

variants affect a large number of traits and the rest affect one or a handful of traits (18). Of note, however, is that in these models nearly all genes have some degree of pleiotropy (16).

In humans, a recent systematic review quantified the amount of pleiotropy observed among SNPs and genes associated with common complex diseases and traits in the literature (19). This study looked for single nucleotide polymorphisms (SNPs) associated with multiple traits at a genome-wide significance level ($p < 5x10^{-8}$), as reported in the National Human Genome Research Institute's Catalog of Published Genome-wide Association Studies (GWAS) (20). The study found 233 genes and 77 SNPs that showed pleiotropic effects, accounting for 16.9% of genes and 4.6% of SNPs in the catalog, respectively (19). Pleiotropic SNPs were more likely to be exonic and less likely to be intergenic, but many pleiotropic SNPs were not located in genes or tied to predicted function. Since this analysis depended on SNPs already associated with disease, this is expected to be a conservative estimate. With over 20,000 genes in the human genome, these 233 genes would represent only ~1% of genes, which seems low given the L-shaped distribution of pleiotropic effects reported in animal models. Furthermore, extrapolating their GWAS catalog findings to the genome suggests that upwards of 3,380 genes might be pleiotropic (16.9% of 20,000). While rough approximations, such examples provide evidence that abundant pleiotropy can already be seen, and that additional pleiotropic relationships can be expected to be uncovered.

While debate continues on the definition and extent of pleiotropy (17), for the purposes of this dissertation there are two key points: that many, perhaps most, genes in the genome possess the potential for multiple pleiotropic relationships, and that some genes may influence a large number of phenotypic traits. Predicting which genes will have pleiotropic effects is difficult, since additional pleiotropic disease relationships cannot be established until an initial

association has been discovered. Therefore treating a genetic variant as if it only affects one trait can be useful, but provides incomplete information. It may be true at the time of evaluation, but may be likely to be associated with additional phenotypes in the future. Since such pleiotropic relationships are expected to grow in number as genetic knowledge improves, several currently known examples of pleiotropy can provide informative illustrations of potential impacts on the return of results.

Forms and examples of pleiotropy

That multiple distinctive traits can appear to cluster together has long been recognized in medical syndromes. These correlated phenotypes could be due to a mutation with multiple independent effects, or could be due to a mutation that begets a cascade of multiple effects. While such differences in molecular mechanism have led to distinctions in the various definitions of pleiotropy (17), the important point is that in either case multiple phenotypes are influenced by a single genotype. This constellation of phenotypes may be well-recognized when classified together as a medical syndrome, but these are pleiotropic effects nonetheless. Outside of defined syndromes, various phenotypic outcomes may appear together often enough that they are considered simultaneously.

Variants in the *BRCA1* and *BRCA2* genes, for example, can substantially raise a person's risk for both breast and ovarian cancer. While these are distinct phenotypes, the association with each is sufficiently known that both are usually mentioned in conjunction with variants in these genes. This is an inherent recognition of pleiotropy, but is not often described as such. There are additional cancers associated with variants in these genes, though the risk of disease is not as strong and thus are often less discussed than the breast and ovarian cancer risks. For example,

BRCA1 mutations can increase risk for cervical, uterine, pancreatic, and colon cancer, while *BRCA2* mutations can increase risk for melanoma, pancreatic, stomach, gallbladder, bile duct, testicular, and prostate cancer (21). These additional pleiotropic relationships may be less well known, but they may be just as relevant to clinical management or a particular participant's preferences. Where there are multiple pleiotropic relationships, then, there may be an inherent desire to focus on those deemed most important. These genes provide an example of genetic variants with an array of pleiotropic effects which range from small to large increases in risk.

However, pleiotropy can also have multiple molecular effects that influence the risk of diseases which may not usually be considered related or part of the same syndrome, and it is in these situations where return of results is most impacted. In such a situation, a single genotypic change could yield different proteins with different effects, or could yield a single protein that has different effects in different cell types. In either case, these pleiotropic changes can lead to increases or decreases in disease risk for different diseases, some of which may be unanticipated. Such changes in disease risk could be of similar magnitude for multiple disorders (such as *BRCA1/2*), or could be stronger for one disease than another (such as *APOE*, described below). Alternatively, pleiotropy could also lead to opposing changes in risk, where risk for one disease may be increased but risk for another is decreased (such as *TERT-CLPTM1L*, described below). Each of these situations could have consequences on whether a genotypic result is appropriate for return.

One of the most instructive examples of pleiotropy which impacts the return of results discussion involves genetic variants of the *APOE* gene (22). Variants in this gene have been associated with a variety of phenotypes, with variants conferring an increased risk for elevated lipid levels, coronary heart disease (CHD), and Alzheimer's disease (AD) (23, 24). While

variants in *APOE* alter risk of CHD by up to 20% and influences LDL cholesterol levels by up to 30% (25), the effect on AD is much more substantial, with odds ratios up to 14.9 (24). The genetic information on increased lipid levels and CHD risk is clinically actionable, in that the recipient could potentially take preventive measures to reduce risk of future disease (such as lifestyle modifications, screening adjustments, or pharmaceutical interventions). The increased risk of AD, however, is not considered clinically actionable information given the current paucity of preventive or treatment options for the disease. Additionally, it is possible that the specter of neurodegenerative disease may be more psychologically damaging or socially stigmatizing than the possibility of hyperlipidemia. This example shows how a single genotype can inform disease risk for seemingly unrelated phenotypes, which may have very different potential for action or harm.

Some genetic variants demonstrate pleiotropy by having small effects on many similar types of disease, such as variants in the 8q24/*MYC* and *TERT-CLPTM1L* loci for various cancers. In the 8q24/*MYC* locus, more than 14 SNPs have been associated with a number of different cancers (26). It is hypothesized that variants in this region act as enhancers for the *MYC* oncogene, and this common mechanism has pleiotropic effects on cancer risk through tissue-specific effects (27). While pleiotropic effects are seen when taking the 8q24/*MYC* locus as a whole, smaller independent risk regions are important for specific cancers, impacting which relationships might be appropriate to return. Another major pleiotropy locus is the *TERT-CLPTM1L* locus on 5p15.33, which has been associated with at least seven different cancers (28). Interestingly, some variants in this region have pleiotropic effects in opposite directions (28, 29). SNP rs401681, for example, has been associated with increased risk of lung cancer (30), basal cell carcinoma (31), bladder cancer, prostate cancer, and cervical cancer (28), but also

a decreased risk of melanoma (28, 31) and pancreatic cancer (32). Believed to impact telomere length (28), variants in this region lead to similar biological effects in different tissues that carry pleiotropic risk of various cancers. While most of the individual genotype-phenotype relationships have a small effect for a single cancer (odds ratios from 0.75 to 1.61 (33)), taken together these relationships could be important for overall risk of developing any type of cancer. Both of these regions are good examples of highly pleiotropic genetic variants, some of which demonstrate opposite directions of effect.

While the above examples show pleiotropic relationships within the same class of disease (e.g. cancer), the same genotype can also have pleiotropic effects on different diseases which might not be expected to share any genetic cause. For example, variants in several genes, particularly *HNF1B,* have been associated with both an increased susceptibility to prostate cancer and a decreased risk of type 2 diabetes (34). In another example, multiple loci were identified with pleiotropic effects for both inflammatory bowel disease and type 1 diabetes, some with opposite directions of effect (35). When considering the return of results, this suggests that unknown pleiotropic effects cannot be assumed to be in the same class of disease. Rather, it should be recognized that future research may identify unexpected disease relationships.

There is also a temporal aspect to pleiotropy. Evolutionary scientists theorize that some genotypes may produce multiple phenotypes with opposing effects on fitness (36). Termed antagonistic pleiotropy, in such cases a genotype may provide some phenotypic benefit to fitness early in life, but also produce some phenotypic traits which negatively impacts fitness following reproductive age. For example, it has been suggested that carriers of genetic variants in the *APOE* gene have improved memory and other cognitive domains in early life, but increased risk of cognitive decline and AD later in life (37). Thus pleiotropy complicates the return of results

not only because multiple relationships may exist for a given genotype, but also because the contextual relationships between these genotype-phenotype associations may differ in important ways over time.

Whether a genotype has multiple phenotypic implications can also depend on other factors inside and outside the genotype of an individual. For example, variants in the *ATM* gene can lead to the autosomal recessive disorder ataxia-telangiectasia, and affected individuals have a highly increased risk of cancer, particularly lymphoid and epithelial cancers (38). While these pleiotropic effects are only seen in persons homozygous for *ATM* variants, unaffected carriers are also at a moderate increased risk of breast cancer (39, 40). Variants in this gene, therefore, can have different amounts of pleiotropy and different phenotypic implications depending on the genotype and characteristics of the individual. While a heterozygous *ATM*-variant carrier might be at increased risk of breast cancer, whether there was also a pleiotropic risk of having offspring affected with ataxia-telangiectasia would depend on whether the other biological parent was also a carrier. This pleiotropic risk information may also be irrelevant when the individual is past reproductive age. Reporting variants for the purposes of identifying breast cancer risk may also provide pleiotropic risk information with potential reproductive implications, where this secondary risk information may be of greater impact to the offspring than the unaffected carrier. This example further demonstrates the complexity of pleiotropic variants, where implications are likely to vary by individual. The relevance of additional pleiotropic relationships may vary depending on additional factors inherent or external to the individual tested (e.g. carrier status of spouse), and some phenotypic implications may be more important for family members than to the individual themselves (e.g. offspring).

The impact of pleiotropy on the return of results

While knowledge of pleiotropy and its extent in the human genome is still being generated, there are several relevant takeaways when considering the potential impact on the return of research results. First, many genetic variants currently known to be associated with risk of some phenotype are likely to be associated with additional phenotypes. These additional phenotypic relationships could be in the same or opposing directions, and these additional pleiotropic relationships could be with phenotypes either similar or unrelated to the original phenotype. Second, for some genetic variants, the number of pleiotropic relationships could be quite large. This may mean that evaluating or reporting all phenotypic effects for a given variant may be difficult or infeasible. Third, given current knowledge it is likely that additional pleiotropic relationships will be identified after the decision for whether and how to return results has been made. Therefore all parties must recognize that additional information may be available in the future which may have impacted previous decision-making. Fourth, an important measure of the extent of pleiotropy is not only the number of genes which are pleiotropic, but also the number of pleiotropic relationships that a gene might have. Both of these measures will be important considerations when determining how to return results. With these points in mind, there are several ways in which pleiotropy impacts the return of results.

First, pleiotropic relationships impact the way in which a result should be returned. If returning the raw genetic information to the participant (e.g., rs3093058 A/T), this opens the possibility of the participant using this information in future applications. This could allow the participant to search for additional associations with their particular genotype, either at the time of return or at some future point. This would allow the participant to learn about additional pleiotropic relationships with that particular genotype that may not have been known at the time

of return. Alternatively, if returning only the interpretation of the genotype (e.g., XX% increased risk of Alzheimer's disease), the participant would likely not have the ability to learn about future pleiotropic relationships with the genotype behind that interpretation (unless the risk was clearly tied to a particular genotype). Any future knowledge would need to be discovered by and returned through the researcher. Thus, pleiotropy also impacts what type of genetic information is returned to a research participant. This can be the raw genetic information, the genotype-phenotype relationship, or the phenotypic interpretation. The argument for which should be returned may further depend on the variant considered, and whether participant autonomy or provider beneficence is emphasized. If a potentially harmful pleiotropic relationship is recognized to be potentially harmful, researchers may decide to return only the interpretation that is clinically actionable. If instead no pleiotropic information is known, or if that information is not thought to be harmful, researchers may decide to return raw genetic information and any phenotypic interpretations which are deemed to be validated and appropriate (with the knowledge that future pleiotropic associations are possible).

Second, the various pleiotropic relationships of a given variant may or may not meet the same criteria for whether a genetic result should be returned. Depending on how the result is returned, pleiotropic relationships may force the return of additional results which otherwise would not have been recommended to be returned on their own. Secondary phenotypic relationships, for example, could be analytically valid but have little in the way of clinical utility. This could be due to the variant changing the risk of a secondary disease by an amount so small as to be clinically negligible, or could substantially raise the risk of a disease with no screening or treatment options. Learning of a result that negligibly reduces risk may misleadingly encourage negative lifestyle behaviors. Learning of a result with no actionability could result in

psychosocial harms, particularly when associated with potentially stigmatizing conditions such as neurodegenerative diseases. In both situations, these secondary pleiotropic relationships would not meet the clinical utility requirement often at the heart of return of result guidelines, but might still be returned due to their pleiotropic relationship with a result that did meet these requirements. Alternatively, secondary pleiotropic relationships may have inconclusive evidence of an association, raising the possibility of that relationship being a false positive or false negative finding. Such a secondary disease risk association could be the result of a single small published paper, and yet be accessible to a wide audience that could misinterpret or misattribute importance to inconclusive scientific literature or annotation databases. Popular media coverage would likely sensationalize this incomplete information (e.g., "the ___ gene"), potentially exacerbating the issue. Current guidelines do not require the evaluation or consideration of these secondary pleiotropic relationships in determining whether a result should be returned.

Third, pleiotropic relationships may degrade or invalidate the ability of participants to choose the level of results they want returned. Some guidelines have suggested a process where a participant may choose which type of result they want returned, such as only those results which are clinically actionable, or only those which are not potentially stigmatizing, or only those for the disease of interest in the study. Querying participant preferences have particularly been recommended for the ambiguous middle category of results. While this provides additional autonomy to the participant, such distinctions are incomplete and arbitrary when the same genotype has pleiotropic associations that reside in multiple decision categories. Such pleiotropic relationships complicate returning results, since a given result might have pleiotropic relationships that reside in categories with different preferences for return. For a pleiotropic genetic result with one phenotypic relationship the participant wants to know about and another

that they do not want to know about, this could put returners in the unfortunate position of trying to gauge which participant preference to violate. Furthermore, in some cases pleiotropic relationships could mean that the participant preferences for results in the middle category of optional return may conflict with the "duty to warn" return recommendations present in the first category. Current guidelines are not sufficiently detailed to deal with such nuances.

Fourth, whether a particular genetic result meets the threshold for being returned may differ when considering only one genotype-phenotype relationship at a time compared to all the pleiotropic associations for that genotype together. For example, a genetic variant that raises the risk of a single phenotype a small amount may not meet return criteria. If that same genetic variant also raises the risk for a number of other phenotypes, this overall risk increase across multiple phenotypes may improve the likelihood of improving participant lifestyle behaviors in a way that reduces disease risk. While no single genotype-phenotype would meet threshold criteria that would lead to returning that genetic information, the balance of multiple phenotypic risk relationships could push that variant to be returned. Similarly, such a holistic approach may also be important when considering the return of risk information for particular phenotypes. Recent epidemiologic studies have attempted to measure the overall impact of multiple genetic variants on particular phenotypes in an aggregate fashion. This results in a 'risk score' which provides the overall risk association given a range of genetic information. In such situations it might be that no one particular genotype used in the risk score would fulfill the criteria for return, but the overall aggregate score would. If one of the variants included in the risk score had additional pleiotropic relationships, those additional pleiotropic relationships would potentially be returned as well, either directly or indirectly. *APOE* variants could be included in a risk score

for lipid levels, for example, potentially also giving AD risk information. Current guidelines do not offer guidance beyond single genotype-phenotype relationships.

Fifth, pleiotropic genes with effects in opposite directions of risk for different phenotypes may further alter the decision-making and timing of result return. In cases where a genotype has pleiotropic effects in opposing directions, for example, a participant may wish to know genetic information that decreases risk and not information that increases risk (or vice versa). Current guidelines do not address cases of opposing pleiotropy where the same genetic variant both increases and decreases risk for different phenotypes. In cases of antagonistic pleiotropy, the timing of returning a result may also be important. A participant may want to know genetic information that could be important early in development, but not want to learn about risk for diseases later in life. *APOE* variants, for example, increase risk for Alzheimer's Disease in older adults, but are also hypothesized to be beneficial to brain development and cognition early in life (41). A participant may want to know only one of these associations, or may only want to know about them while they are still able to make reproductive decisions. Alternatively, a parent may want to know genetic information about such a variant for their child, but the child may have wished not to know had they been given the choice (a preference that might also change depending on the developmental age of the individual). Using *BRCA1/2* as another example, a woman desiring children may decide to act on one pleiotropic risk (mastectomy) but delay acting on the other (oophorectomy). Current guidelines do not address the possibility of antagonistic pleiotropy. While some guidelines recognize that the same result might have divergent meanings or participant preferences at different life stages (7), this is more in relation to reproductive concerns than due to pleiotropy. Rather, most guidelines only consider the timing of returning

results from the standpoint of how long a researcher can return results, such as only through the duration of grant funding (10).

Pleiotropy in non-research settings

While this project focuses on the issue of pleiotropy impacting the return of results in a research setting, these issues will arise in other arenas as well. In clinical care, even targeted genotyping panels performed for a particular purpose can result in incidental findings due to pleiotropic effects. For example, a recent systematic review evaluated how much extra information might be discovered when clinically genotyping 34 pharmacogenomic-related genes, which could be done to guide drug selection and dosing for a patient (42). From a search of the literature, they found 372 putative incidental genotype-phenotype associations that could be revealed by genotyping these genes. Several of these incidental pleiotropic findings were replicated and carried substantial effects (OR ≥ 2 or ≤ 0.5). Therefore despite being targeted for a specific purpose, these clinical tests revealed additional information due to pleiotropic relationships. As mentioned above, the number of pleiotropic effects seen for such tests can be expected to increase.

Beyond targeted testing, clinical sequencing is increasingly being utilized for cancer and other diseases. Such broad tests are more likely to discover incidental findings outside any particular targets of the test. Many of these results might have pleiotropic meanings, and will need to be addressed accordingly. Current guidelines are set up to prioritize such incidental findings to only obligate the return of those that are of high clinical utility. This stems from a paternalistic medical model of only reporting information that can impact clinical decision-making. At this time, the list of what results are necessary to return is considered fairly small

and manageable (8), and bioinformatics approaches have been proposed for categorizing incidental findings accordingly (3).

In the clinical situation, additional pleiotropic relationships with a small group of findings may be of far less clinical importance than the primary finding, or may be considered part of the same syndrome. Additionally, any such return is likely to lead to clinical care, in which case any additional associations with the primary result could be addressed (though they also may not be, since they are likely of lesser importance and clinical time is often short and triaged). In these cases, the net benefit of returning that primary result will likely be thought to outweigh any potential harm that arises due to additionally informing the participant of pleiotropic associations. Because of this, these results will likely be returned regardless of any additional pleiotropic relationships, which may or may not also be reported. However, as genetic knowledge improves, such lists expand, and additional pleiotropic relationships are discovered, this balance may not remain as clear.

Outside the clinical and research arenas, the direct-to-consumer (DTC) genetic testing community is generating a small but vocal group of stakeholders who view obtaining genetic information as empowering the individual. This may lead to the return of variants which may be of lower clinical utility, have lower susceptibility or reduced penetrance, or may not be interpretable (such as variants of unknown significance). In these situations, consumers may not necessarily be referred to, or partake in, any genetic counseling or clinical care as a result of their genetic information (though resources are likely to be offered). These consumers may be more likely to search for phenotype relationships with their genotype information, either though resources provided by researchers or on their own. They may then go on to find out information they would have otherwise preferred to avoid as a result. Even when the results returned are

initially limited to those the participant wishes to know about, this group is likely to uncover future pleiotropic relationships. Additionally, several of these DTC companies, such as 23andMe, are now using customer data for research purposes. As DTC becomes more widely adopted, this may drive an expectation on the part of traditional research participants to receive their genetic results. While such expectations should not drive research practice, it is possible that researchers may eventually find it necessary to incorporate returning genetic information in order to successfully recruit participants into their studies. If research participants begin demanding more and more access to their genetic information, the potential for pleiotropic information being returned should be communicated.

Potential framework

Because of pleiotropy, when evaluating whether a given genotypic result should be returned, all currently known phenotypic relationships with that genotype must also be considered. These additional relationships must also be evaluated to determine whether they should also be returned, both on their own and within the context of multiple pleiotropic relationships. This complicates the return issue, since in cases of pleiotropy returning one result may directly or indirectly lead to returning another result. This may not be an issue when both phenotypic relationships meet or do not meet given return criteria, but will cause increasing problems when one or more phenotypic relationships with a given genotype meet return criteria and one or more other phenotypic relationships do not. This will be particularly worrisome when secondary pleiotropic relationships do not meet return criteria because of potentially harmful information. In some cases, such secondary pleiotropic relationships may create an argument not to return a primary result which would otherwise meet return criteria. Thus researchers will need to expand the currently limited guidelines to incorporate the presence of pleiotropic

relationships, and weigh single genotype-phenotype relationships not only by themselves, but also within the overall balance of all phenotypes associated with that genotype.

One potential framework for doing so can be constructed by utilizing existing return of result guidelines outlined by the NHLBI working group (10) and others (7). Focusing on clinical validity and utility, each genotype-phenotype relationship can be individually evaluated to have an expected net effect of reporting a result. As such, reporting a given genotype-phenotype relationship could have a strong net benefit (high clinical validity and utility), a possible net benefit (high clinical validity, low utility), a neutral effect (low clinical validity or utility), or lead to net harm (high clinical validity, low utility, stigmatizing condition or vulnerable population). Each particular phenotypic relationship of a pleiotropic genotype can be assigned to one of these categories, after which recommendations for whether or not to return these pleiotropic results may then be based on maximizing benefits and minimizing harms across the sum total of these assignments. Where strong benefits are expected (associated with return of either the primary finding or a pleiotropic secondary finding), results should be returned. Where substantial harm is likely (again, of either result), results should not be returned. In areas of optional return, the overall net benefits must be measured against the net harms to determine whether return is appropriate.

A two-way example of such an approach is shown in Figure 5. In this example, the goal is to maximize benefits and minimize harms to the participant while weighing the relative contributions of two pleiotropic associations (A and B) with the same genetic result. Thus to maximize potential benefits, unless returning association A is expected to lead to harm, the expected strong net benefit from phenotypic association B means that on balance the result should be returned (green boxes, upper left). On the other hand, to minimize harms, unless

returning association A is expected to lead to strong net benefit, the likely harm expected from returning association B means that on balance the result should not be returned (red boxes, lower right). For each of these situations, the stronger argument for one phenotypic association trumps the return recommendation that might have been individually attached to the other phenotypic association. The middle categories, without strong arguments for or against return from either association, are more reliant on particular characteristics of the two associations to determine whether on balance there is an expected net benefit that justifies return. Participant preferences regarding one category or another could be utilized as part of this process as well.

Admittedly, such a pair-wise matrix oversimplifies these complex issues. Many pleiotropic results are likely to have more than two pleiotropic relationships, in which case such a table starts to break down due to multi-dimensionality. With additional pleiotropic effects, result recommendations could rely on trump categories (strong benefit, likely harm) or a more complete weighing of the contributions of each relationship. Despite its simplicity, however, such a model provides a good starting point for future discussion, and shows how genetic results might have different return recommendations depending on their other pleiotropic relationships. This model assumes that the genetic result itself is returned along with phenotypic interpretation, but this could be altered where appropriate. In the corners where one association provides strong benefit and the other likely harm, for example, a case could be made for only returning the interpretation expected to lead to benefit.

In this proposed framework, the combined set of evaluations should be considered individually for each result, preferably in consultation with a genetic counselor, researcher, clinician, or other expert. Ideally, participants provide their preference in consent processes for what results they would prefer to receive, and how they wish conflicting preferences to be

weighed. It may be that in some circumstances the participant's strongest preference to know or not know particular categories of results will overrule the ability of researchers to return (or not return) results they otherwise would deem appropriate (or inappropriate) to return. Due to pleiotropy, the decision-making on whether one result is appropriate to be returned will need to take into account the context of additional phenotypic relationships. This will become increasingly difficult as increasing numbers of pleiotropic relationships are discovered; both as the number of genotypes which demonstrate pleiotropy increases, as well as the number of phenotypic relationships for a given genotype grows. Beyond practical frameworks, policy changes will also be necessary to approach this issue.

Policy recommendations

From the above examples, it should be clear that the existence of pleiotropy introduces several additional levels of complexity to the determination of whether (and when and how) genetic results should be returned to research participants. Current guidelines do not take these additional associations into account, nor do they offer recommendations for how to incorporate these extra relationships into these current guideline structures. Pleiotropic relationships cause serious problems for some guidelines, which will necessitate major revisions in order to ensure appropriate return of genetic information. While additional discussion and consideration will be needed to incorporate changes into existing guidelines, we suggest the following recommendations as ways to begin dealing with this increasingly important situation of pleiotropic results. Key aspects of these suggestions are balancing beneficence and non-maleficence, by trying to maximize benefits and minimize harms while still emphasizing participant autonomy and empowerment.

First, informed consent documentation should include information so that research participants are aware that one genetic result can have multiple implications. These additional implications could be known or unknown at the time of return, and may or may not be recognized, replicated, or relevant. It should also state that research is constantly finding additional genotype-phenotype relationships, some of which could be revealed directly or indirectly against the preferences of the participant. Some of these novel findings could be important for the same phenotype that was formerly reported, while others may be unrelated. There is always the possibility that novel phenotypic findings for the returned genotype may include information that may be psychologically harmful or socially stigmatizing, either to the individual or to family members. Any new risk information for a given result may take time to validate, and may not be of immediate or future clinical relevance or usefulness. Therefore it is important to recognize the rapid pace of scientific and popular media reporting of genetic findings compared to the slow pace of validating, interpreting, and translating genetic discoveries into clinical practice. Potentially novel pleiotropic information may not hold up to further investigation, so exercising caution for interpretation and action is necessary. Overall, it is important that educational information is provided to the participant so that they can make informed decisions regarding potentially pleiotropic genetic information. This is not meant to scare the participant away from research, but rather to hopefully reduce the impact or severity of any unanticipated information should it arise.

Second, participants should be able to choose the type of information they wish to receive, and decide which preferences takes precedence when pleiotropy leads to conflicting recommendations. This could involve choosing between informative hierarchical examples of conflicting pleiotropic relationships. For example, in cases where participant preferences or the

usual return guideline categories conflict due to pleiotropic relationships, the participant should be able to rank whether they would rather find out something they did not want to know, or not find out something they did want to know. This would likely be difficult in the abstract, and so concrete examples will need to be offered. Researchers will need to determine what options to provide the participant, such as whether to report all pleiotropic relationships or only those that meet some minimal threshold criteria (such as analytic validity, potential clinical utility, effect size, etc.). When a pleiotropic genetic result meets that criteria it should be returned unless the participant has had the opportunity to state a preference for not wanting to know about a type of result that may also be represented by the particular genotype. The types of results can be determined by the researcher, or categorized according to existing guidelines focusing on clinical validity and clinical utility.

Third, participants should also be to choose the way in which they want results returned to them, which could mean raw sequence data with phenotypic interpretation, aggregate risk information without genotype information, or some other form of report. In essence, this will be a choice for how much autonomy they desire over their health information. Participants who receive genotype information should be provided with material on resources through which they may be able to get additional information on potential genotype-phenotype relationships, should they desire. This could include documentation on online databases such as NCBI, PubMed, or SNPedia, with the caveat that such databases have shortcomings. In-person resources such as referrals to genetic counselors or other care providers could also be offered. A third option may be emerging portals which are designed to provide phenotype interpretation of genetic research results, such as My46 (my46.org), or coordination with private companies such as 23andMe (23andme.com) or Knome (knome.com). Additional web interfaces may provide other resources

for individuals who have their genetic information to discover future interpretations of that information. Providing such resources allows the participant the option to discover additional information, including future pleiotropic relationships with reported genotypes, without further impacting researcher time or resources. This would also allow for future benefits of currently reported genetic information to extend beyond study-specific funding timelines. Such a practice could be standardized across studies or put into place by funding agencies in a way that would not burden researchers. This would grant research participants the autonomy to decide when to find out about any potential additional pleiotropic relationships with their genetic results.

Future directions

Regardless of how pleiotropic information is discovered, it remains to be determined whether returning pleiotropic risk information is of greater impact than singular genotype-phenotype associations (and whether the magnitude of this risk impacts any resultant behavior). For some genetic variants, additional pleiotropic relationships may have very small effect sizes, effectively making them clinically irrelevant. Others may severely impact risk for multiple phenotypes, in opposite directions, or at different ages. Knowing how best to communicate this information will be important not only for research settings, but also for translating genetic information into clinical and public health settings. For example, when told of two phenotypic relationships with the same genotype, which do people focus on, and does behavior change for either? Does returning both results at the same time, or reporting at separate time points, impact participant outcomes? Does providing multiple phenotypic results oversaturate the return process, and mean that information becomes lost? Are additional resources actually needed to deal with extra phenotypic interpretations, or are people already equipped to deal with multiple meanings simultaneously? Do incidental pleiotropic results enhance or distract from action on

primary results? Are highly pleiotropic genotypes more important to return than single genotype-phenotype relationships? These and other questions will be of growing importance as more pleiotropic relationships are found, and future research will be needed to address them.

Conclusion

Given the potentially large number of results which could be returned, current return of result guidelines seek to prioritize the results which meet high thresholds of clinical utility. While this restriction may be useful for risk variants which only affect a single phenotype, such an approach will be incomplete when the variant has additional pleiotropic relationships. Because of these secondary pleiotropic relationships with other phenotypic traits, the determination that a primary genotype-phenotype relationship is appropriate to return may directly or indirectly cause the return of another result which may not meet the same high standards of clinical utility. While this limitation could be considered an acceptable compromise for the return of the primary result of interest, to this point it does not appear to be accounted for by current criteria. This leaves these guidelines incomplete, as there are many ways in which a pleiotropic variant might impact whether a genetic result is appropriate for return.

For most current guidelines, at least some results fall into categories for which the recommendation on whether or not results should be returned relies in part upon participant preferences for which they wish to receive. These are often the results that have characteristics such as reduced clinical utility, low or little clinical validity, of potential psychosocial harm, or of uncertain personal utility. While allowing participant autonomy is useful in these situations, these decisions cannot truly be informed when pleiotropic relationships may lead to conflicting preferences. We suggest that in such cases, the participants not only provide their preference for

which categories of information they wish to have returned, but also their preferred hierarchy for which preference takes precedence in cases where pleiotropy yields information in multiple categories for the same genetic result.

Given the likelihood of pleiotropic relationships, even returning results with the most stringent criteria and the best of intentions may inadvertently lead to participant harm. Even when additional pleiotropic relationships are seemingly irrelevant, by reporting one result researchers will be also reporting additional results they may not have wanted to report. This is the important takeaway that each of the parties involved in research need to recognize when it comes to returning results, and that needs to be discussed and incorporated into future guidelines. This dissertation has provides a potential starting framework for evaluating multiple pleiotropic relationships simultaneously when considering returning a genetic result to a participant. This dissertation has also offered several policy considerations for ensuring that participant autonomy is emphasized throughout the research process: during the informed consent process, at the time of result return, and at future points after the research is concluded. These suggestions are admittedly simple for such a complex issue, but they nevertheless should provide a place to begin the conversation. Much more discussion will be needed in order to craft proper guidelines for returning potentially pleiotropic research results, an issue that can be reasonably expected to grow in importance over the coming years. To be successful, future return of result guidelines must take the complexities introduced by pleiotropic genetic relationships into account, and take a more holistic approach to evaluating the appropriateness of returning results.

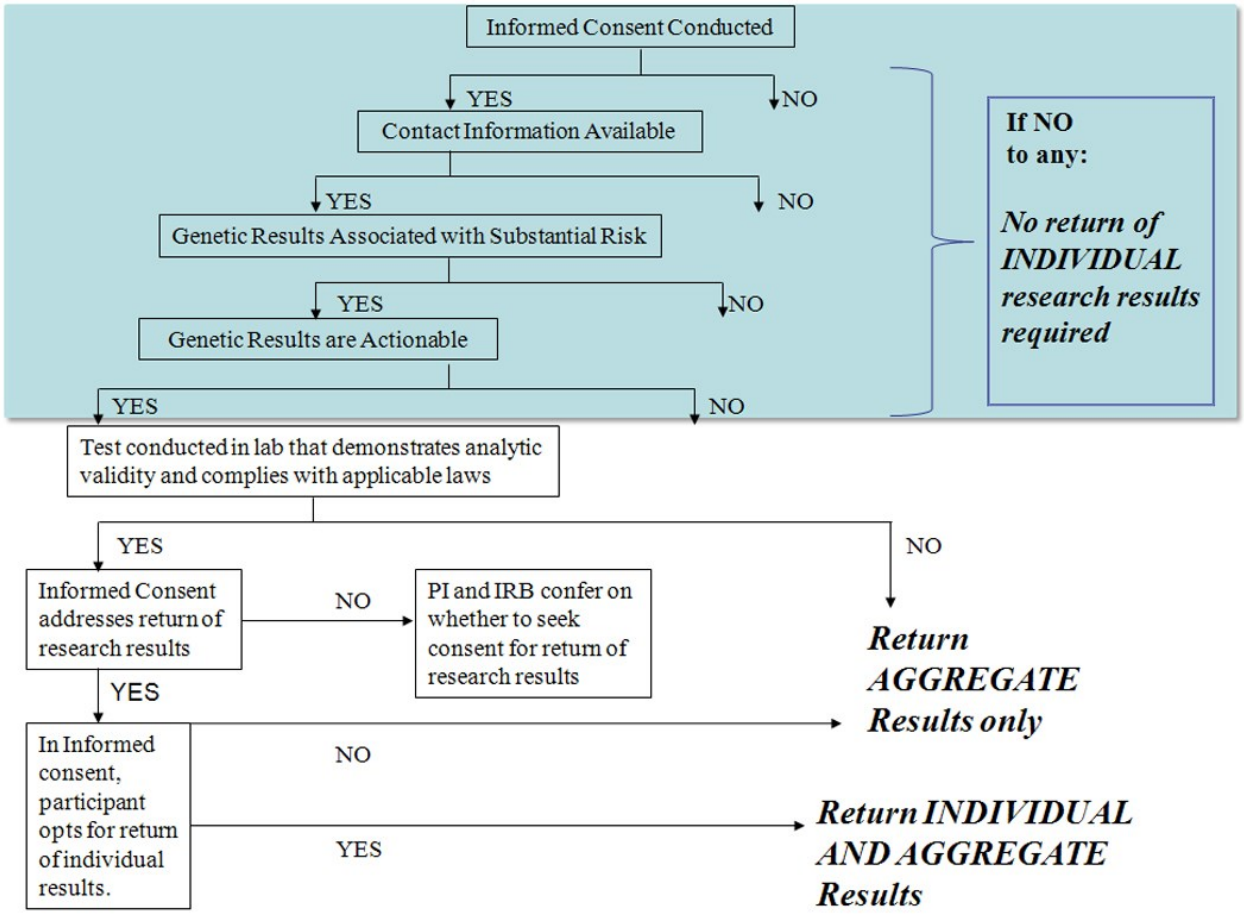
References

1. Cho MK. Understanding incidental findings in the context of genetics and genomics. The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics. 2008;36(2):280-5, 12. Epub 2008/06/13.
2. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA : the journal of the American Medical Association. 2006;296(2):212-5. Epub 2006/07/13.
3. Berg JS, Adams M, Nassar N, Bizon C, Lee K, Schmitt CP, et al. An informatics approach to analyzing the incidentalome. Genetics in medicine : official journal of the American College of Medical Genetics. 2013;15(1):36-44. Epub 2012/09/22.
4. Johnston JJ, Rubinstein WS, Facio FM, Ng D, Singh LN, Teer JK, et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. American journal of human genetics. 2012;91(1):97-108. Epub 2012/06/19.
5. Cassa CA, Savage SK, Taylor PL, Green RC, McGuire AL, Mandl KD. Disclosing pathogenic genetic variants to research participants: quantifying an emerging ethical responsibility. Genome research. 2012;22(3):421-8. Epub 2011/12/08.
6. Wolf SM. Introduction: the challenge of incidental findings. The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics. 2008;36(2):216-8. Epub 2008/06/13.
7. Wolf SM, Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton EW, et al. Managing incidental findings in human subjects research: analysis and recommendations. The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics. 2008;36(2):219-48, 1. Epub 2008/06/13.
8. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. Genetics in medicine : official journal of the American College of Medical Genetics. 2011;13(6):499-504. Epub 2011/05/12.
9. Bredenoord AL, Onland-Moret NC, Van Delden JJ. Feedback of individual genetic results to research participants: in favor of a qualified disclosure policy. Human mutation. 2011;32(8):861-7. Epub 2011/05/04.
10. Fabsitz RR, McGuire A, Sharp RR, Puggal M, Beskow LM, Biesecker LG, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. Circulation Cardiovascular genetics. 2010;3(6):574-80. Epub 2010/12/16.
11. Dressler LG. Disclosure of research results from cancer genomic studies: state of the science. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(13):4270-6. Epub 2009/06/25.
12. Bredenoord AL, Kroes HY, Cuppen E, Parker M, van Delden JJ. Disclosure of individual genetic data to research participants: the debate reconsidered. Trends in genetics : TIG. 2011;27(2):41-7. Epub 2010/12/31.
13. Ravitsky V, Wilfond BS. Disclosing individual genetic results to research participants. The American journal of bioethics : AJOB. 2006;6(6):8-17. Epub 2006/11/07.
14. Beskow LM, Burke W. Offering individual genetic research results: context matters. Science translational medicine. 2010;2(38):38cm20. Epub 2010/07/02.
15. Bookman EB, Langehorne AA, Eckfeldt JH, Glass KC, Jarvik GP, Klag M, et al. Reporting genetic results in research studies: summary and recommendations of an NHLBI

working group. American journal of medical genetics Part A. 2006;140(10):1033-40. Epub 2006/04/01.
16. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics. 2010;186(3):767-73. Epub 2010/11/11.
17. Paaby AB, Rockman MV. The many faces of pleiotropy. Trends in genetics : TIG. 2013;29(2):66-73. Epub 2012/11/13.
18. Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature reviews Genetics. 2011;12(3):204-13. Epub 2011/02/19.
19. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. American journal of human genetics. 2011;89(5):607-18. Epub 2011/11/15.
20. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(23):9362-7. Epub 2009/05/29.
21. BRCA1 and BRCA2: Cancer risk and genetic testing. National Cancer Institute [updated 05/29/2009; cited 2013 3/3]; Available from: http://www.cancer.gov/cancertopics/factsheet/Risk/BRCA.
22. Wachbroit R. The question not asked: the challenge of pleiotropic genetic tests. Kennedy Institute of Ethics journal. 1998;8(2):131-44. Epub 2001/10/20.
23. Lahoz C, Schaefer EJ, Cupples LA, Wilson PW, Levy D, Osgood D, et al. Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. Atherosclerosis. 2001;154(3):529-37. Epub 2001/03/21.
24. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nature reviews Neurology. 2013;9(2):106-18. Epub 2013/01/09.
25. Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. JAMA : the journal of the American Medical Association. 2007;298(11):1300-11. Epub 2007/09/20.
26. Grisanzio C, Freedman ML. Chromosome 8q24-Associated Cancers and MYC. Genes & cancer. 2010;1(6):555-9. Epub 2011/07/23.
27. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(21):9742-6. Epub 2010/05/11.
28. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nature genetics. 2009;41(2):221-7. Epub 2009/01/20.
29. Baird DM. Variation at the TERT locus and predisposition for cancer. Expert reviews in molecular medicine. 2010;12:e16. Epub 2010/05/19.
30. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nature genetics. 2008;40(12):1407-9. Epub 2008/11/04.
31. Stacey SN, Sulem P, Masson G, Gudjonsson SA, Thorleifsson G, Jakobsdottir M, et al. New common variants affecting susceptibility to basal cell carcinoma. Nature genetics. 2009;41(8):909-14. Epub 2009/07/07.

32. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nature genetics. 2010;42(3):224-8. Epub 2010/01/27.
33. Mocellin S, Verdi D, Pooley KA, Landi MT, Egan KM, Baird DM, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. Journal of the National Cancer Institute. 2012;104(11):840-54. Epub 2012/04/24.
34. Berndt SI, Sampson J, Yeager M, Jacobs KB, Wang Z, Hutchinson A, et al. Large-scale fine mapping of the HNF1B locus and prostate cancer risk. Human molecular genetics. 2011;20(16):3322-9. Epub 2011/05/18.
35. Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Human molecular genetics. 2010;19(10):2059-67. Epub 2010/02/24.
36. Carter AJ, Nguyen AQ. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. BMC medical genetics. 2011;12:160. Epub 2011/12/14.
37. Tuminello ER, Han SD. The apolipoprotein e antagonistic pleiotropy hypothesis: review and recommendations. International journal of Alzheimer's disease. 2011;2011:726197. Epub 2011/03/23.
38. Ataxia Telangiectasia. National Cancer Institute; [updated 01/26/2006; cited 2013 03/03]; Available from: http://www.cancer.gov/cancertopics/factsheet/Risk/ataxia.
39. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nature genetics. 2006;38(8):873-5. Epub 2006/07/13.
40. Ahmed M, Rahman N. ATM and breast cancer susceptibility. Oncogene. 2006;25(43):5906-11. Epub 2006/09/26.
41. Rusted JM, Evans SL, King SL, Dowell N, Tabet N, Tofts PS. APOE e4 polymorphism in young adults is associated with improved attention and indexed by distinct neural signatures. NeuroImage. 2013;65:364-73. Epub 2012/10/16.
42. Westbrook MJ, Wright MF, Van Driest SL, McGregor TL, Denny JC, Zuvich RL, et al. Mapping the incidentalome: estimating incidental findings generated through clinical pharmacogenomics testing. Genetics in medicine : official journal of the American College of Medical Genetics. 2012. Epub 2012/12/01.

Part B - Figure 1 – Decision flowchart for whether individual genetic results should be returned to research participants, as laid out by the 2010 NHLBI working group (10).



Part B – Figure 2 – Classification of incidental findings into three categories, each carrying different recommended actions for whether or not results should be returned, as proposed by Wolf et al. (7).

Table 5

Recommended Classification of Incidental Findings

Category	Relevant IFs	Recommended Action
Strong Net Benefit	• information revealing a **condition likely to be life-threatening** • information revealing a **condition likely to be grave that can be avoided or ameliorated** • genetic information revealing significant risk of a condition likely to be life-threatening • genetic information that can be used to avoid or ameliorate a condition likely to be grave • genetic information that can be used in reproductive decision-making: (1) to avoid significant risk for offspring of a condition likely to be life-threatening or grave or (2) to ameliorate a condition likely to be life-threatening or grave	• **Disclose** to research participant as an incidental finding, unless s/he elected not to know.
Possible Net Benefit	• information revealing a nonfatal **condition that is likely to be grave or serious but that cannot be avoided or ameliorated, when a research participant is likely to deem that information important** • genetic information revealing significant risk of a condition likely to be grave or serious, when that risk cannot be modified but a research participant is likely to deem that information important • genetic information that is likely to be deemed important by a research participant and can be used in reproductive decision-making: (1) to avoid significant risk for offspring of a condition likely to be serious or (2) to ameliorate a condition likely to be serious	• **May disclose** to research participant as an incidental findings, unless s/he elected not to know.
Unlikely Net Benefit	• information revealing a **condition that is not likely to be of serious health or reproductive importance** • information whose **likely health or reproductive importance cannot be ascertained**	• **Do not disclose** to research participant as an incidental finding.

Part B - Figure 3 – Proposed 'binning' approach for determining which incidental genetic results should or should not be returned, as offered by Berg et al. (8).

	Criteria:	*Clinical Utility*	*Clinical Validity*			*Unknown Clinical Implications*
Genes	Bins:	**Bin 1 Medically actionable incidental information**	**Bin 2A Low risk incidental information**	**Bin 2B Medium risk incidental information**	**Bin 2C High risk incidental information**	**Bin 3**
	Examples:	*BRCA1/2* *MLH1, MSH2* *FBN1* *NF1*	PGx variants and common risk SNPs	*APOE* Carrier status for recessive Mendelian disorders	Huntington Prion diseases ALS (SOD1)	**All other loci**
	Estimated number of genes/loci:	10s	10s (eventually 100s – 1000s)	1000s	10s	~20,000
Alleles that would be reportable (YES) or not reportable (NO) in a clinical context						
Variants	Known deleterious	YES	YES/NO [1]	YES/NO [1]	YES/NO [1]	N/A [2]
	Presumed deleterious	YES	N/A [3]	YES/NO [1]	YES/NO [1]	NO [4]
	VUS	NO	N/A [3]	NO	NO	NO [4]
	Presumed benign	NO	N/A [3]	NO	NO	NO
	Known benign	NO	NO	NO	NO	NO

N/A: not applicable; VUS: Variant of uncertain significance

[1] Reporting through decision making with an appropriate provider if elected by the patient.

[2] By definition, variants in genes with unknown implications could not be considered deleterious.

[3] By definition, SNPs or PGx variants will either be present or absent.

[4] Variants in genes with unknown clinical implications would not be reported; however, they may serve as an important substrate for research, potentially uncovering new disease genes.

Proposed system for "binning" of incidental WGS results

Part B - Figure 4 – Summation of recommendations for returning research results from several government agencies, as presented in 2008 by Wolf et al. (7).

Table 3

Comparison of Recommendations on Returning Individual Research Results

National Bioethics Advisory Commission (NBAC)*	Return results only if: (a) "the findings are scientifically valid and confirmed" (b) "the findings have significant implications for the subjects' health concerns" and (c) "a course of action to ameliorate or treat these concerns is readily available."
Centers for Disease Control (CDC)**	Criteria for returning individual results in population-based genetic research: "When the risks identified in the study are both valid and associated with a proven intervention for risk reduction, disclosure may be appropriate."
National Heart, Lung, and Blood Institute (NHLBI)***	Criteria for returning individual genetic results: (1) "The risk for the disease should be significant, i.e. relative risk>2.0. Variants with greater penetrance or associated with younger age of onset should receive priority." Note: "Genetic test results should not be reported to study participants and their physicians as clinically valid tests unless the test(s) was performed in a CLIA certified laboratory. If the test was performed in a non-CLIA certified laboratory, a CLIA certified laboratory should be sought to confirm results by redrawing a sample and performing the test within the CLIA certified laboratory. Results reported by a research laboratory should be identified as 'research' results." (2) "The disease should have important health implications, i.e. fatal or substantial morbidity or should have significant reproductive implications" and (3) "Proven therapeutic or preventive interventions should be available."
National Research Council & Institute of Medicine (NRC & IOM)****	In human embryonic stem cell research, the duty to report individual research results "depends in large part on the reliability of the findings and the significance of the information to human health." "CLIA regulations do not permit the return of research results to patients or subjects if the test were not conducted in a CLIA-approved laboratory."
National Human Genome Research Institute (NHGRI)*****	Upon their request, "[r]esearch participants should have access to experimental research data except when…[t]he research results are of unproven clinical validity, and the IRB has judged that there is no benefit to the research subjects."

* National Bioethics Advisory Commission (NBAC), *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance* (Rockville, MD: 1999), 1: at 72.

** L. M. Beskow et al., "Informed Consent for Population-Based Research Involving Genetics," *JAMA* 286, no. 18 (2001): 2315-2321, at 2320.

*** National Heart, Lung, and Blood Institute, *NHLBI Working Group on Reporting Genetic Results in Research Studies, Meeting Summary*, Bethesda, MD, July 12, 2004, *available at* <http://www.nhlbi.nih.gov/meetings/workshops/gene-results.htm> (last visited January 8, 2008).

**** National Research Council and Institute of Medicine Committee on Guidelines for Human Embryonic Stem Cell Research, *Guidelines for Human Embryonic Stem Cell Research* (Washington, D.C.: National Academies Press, 2005): at 89-90.

***** National Human Genome Research Institute, *Federal Policy Recommendations Including HIPAA, available at* <http://www.genome.gov/11510216> (last visited January 8, 2008).

Part B - Figure 5 – Proposed framework on whether or not to return a pleiotropic research result, based on the pairwise comparison of the expected net benefit or harm of returning results in two different phenotypes (A and B) associated with the same pleiotropic genotype.

Expected result of returning genotype	Association A Strong benefit	Association A Possible benefit	Association A Unlikely benefit	Association A Likely harm
Association B Strong benefit	Return	Return	Return	Likely return
Association B Possible benefit	Return	Likely return	May return	No return
Association B Unlikely benefit	Return	May return	Unlikely return	No return
Association B Likely harm	Likely return	No return	No return	No return

Appendices

Analysis I – Inflammation Appendix 1

Appendix 1 - Assays and instruments used by each study to measure C-reactive protein levels in participants						
Study	Assay					
ARIC	Latex particle enhanced immunoturbidimetric assay - hsCRP					
CHS	High sensitvity ELSIA and Latex-enhanced nephelometry					
CARDIA	High Sensitivity CRP -particle enhanced immunonephelometry					
SHS	ELISA developed in-house using purified CRP and anti-CRP antibodies from Calbiochem					
EAGLE	Latex-enhanced nephelometry (N High Sensitivity CRP assay)					
WHI	Immunoturbidimetric assay with reagents and calibrators (Denka Seiken Co Ltd, Niigata, Japan)					
	Dade Behring N Latex High Sensitivity hsCRP mono assay					
MEC	CRP High Sensitivity latex partical enhanced immunoturbidimetric assay (POINTE Scientific)					
	Instrument					
ARIC						
CHS	BN II nephelometer (Dade Behring Inc.)					
CARDIA	BNII nephelometer from Seimens					
SHS						
EAGLE	BN II nephelometer (Dade Behring Inc.)					
WHI	Hitachi 911; Roche Diagnostics, Indianapolis, Indiana					
	Behring Nephelometer II (Dade Behring Inc., Newark, Delaware)					
MEC	Cobas MiraPlus CC clinical autoanalyzer (Roche Diagnostics)					
	Unit of measurement	Limit of detection	Measured in single lab?	Repeated measurements for same visit?	Repeated measurements for subsequent visits?	Coefficients of variation for repeated measures
ARIC	mg/dL	0.05 mg/dL	Yes		No	19.2%, 421 pairs
CHS	mg/L	0.18 mg/L	Yes	Limited	Yes	8.9% (assay), 5% (instrument)
CARDIA	ug/mL		Yes	No	Yes	average 1.6 - 6.7%
SHS	mg/L		Yes			8%
EAGLE	mg/dL	0.01 mg/dL	Yes	No	No	range 4.9 - 7.8%
WHI	mg/L		Yes	No	No	1.61%
	mg/L		No	No	No	
MEC	mg/dL	0.1 mg/dL	Yes	No	No	3 - 5%

Analysis I – Inflammation Appendix 2

Appendix 2 - SNPs evaluated for an association with C-reactive protein. Previous trait association refers to the trait for which the SNP was originally found associated, as of the literature in 2008. For each study, 1 indicates that the SNP was genotyped and included in this analysis.

SNP	Gene	Coded alle	Previous trait associ	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs1049092	ARMS2/LC	T	Age-related macular	1	.	.	.	1	.	.	.
rs1061170	CFH	C	Age-related macular	1	.	.	.	1	.	.	.
rs1049876		C	Body mass index	1	.	.	.	1	.	.	.
rs1091346	SEC16B/R	C	Body mass index	1	.	.	.	.	1	.	.
rs1121980	FTO	A	Body mass index	1	.	.	.	.	1	.	.
rs1297013	MC4R	A	Body mass index	3	.	.	1	1	.	.	.
rs1333026	13q21.32	A	Body Mass Index	3	.	.	1	.	.	.	1
rs1778231	MCR4	C	Body Mass Index	6	1	.	1	1	1	.	1
rs29941	CHST8,	A	Body mass index	1	.	.	.	.	1	.	.
rs6013029		T	Body Mass Index	2	.	.	.	1	.	.	1
rs6265	LGR4/LIN7	T	Body mass index	2	.	.	.	1	1	.	.
rs6548238	TMEM18	T	Body mass index	4	.	.	1	1	.	.	1
rs6602024	PFKP	A	Body mass index	1	.	.	.	1	.	.	.
rs925946	LGR4/LIN7	T	Body mass index	1	.	.	.	.	1	.	.
rs1083873	MTCH2	G	Body Mass Index	5	.	.	1	1	1	.	1
rs1093839	GNPDA2	G	Body Mass Index	3	.	.	.	1	1	.	1
rs1108475	KCTD15	A	Body Mass Index	5	.	.	1	1	1	.	1
rs2815752	NEGR1	G	Body Mass Index	5	.	.	1	1	1	.	1
rs3751812	FTO	T	Body Mass Index	4	.	.	1	.	1	.	1
rs7498665	SH2B1	G	Body Mass Index	5	.	.	1	1	1	.	1
rs7566605	INSIG2	C	Body Mass Index	1	.	.	.	1	.	.	.
rs9939609	FTO	A	Body Mass Index	3	.	.	1	.	1	.	.
rs7138803	BCDIN3D/	A	Body mass index, W	1	.	.	.	.	1	.	.
rs1038304	ESR1	G	Bone mineral densit	1	.	.	.	1	.	.	.
rs1107748	SOST	C	Bone mineral densit	1	.	.	.	1	.	.	.
rs1513670		T	Bone mineral densit	1	.	.	.	1	.	.	.
rs1999805	ESR1	G	Bone mineral densit	1	.	.	.	1	.	.	.
rs3018362	TNFRSF11	A	Bone mineral densit	1	.	.	.	1	.	.	.
rs3130340	MHC	C	Bone mineral densit	1	.	.	.	1	.	.	.
rs3736228	LRP5	T	Bone mineral densit	1	.	.	.	1	.	.	.
rs3771362	PLCL1	T	Bone mineral densit	1	.	.	.	1	.	.	.
rs4355801	TNRFSF11	G	Bone mineral densit	1	.	.	.	1	.	.	.
rs4870044	ESR1	T	Bone mineral densit	1	.	.	.	1	.	.	.
rs6469804	OPG	G	Bone mineral densit	1	.	.	.	1	.	.	.
rs6929137	ESR1	A	Bone mineral densit	1	.	.	.	1	.	.	.
rs6993813	OPG	C	Bone mineral densit	1	.	.	.	1	.	.	.
rs7220711	SOST	G	Bone mineral densit	1	.	.	.	1	.	.	.
rs7524102		G	Bone mineral densit	1	.	.	.	1	.	.	.
rs7595412	PLCL1	G	Bone mineral densit	1	.	.	.	1	.	.	.
SNP	Gene	Coded alle	Previous trait associ	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs9479055	ESR1	C	Bone mineral densit	1	.	.	.	1	.	.	.
rs9594738	RANKL	T	Bone mineral densit	1	.	.	.	1	.	.	.
rs9594759	RANKL	T	Bone mineral densit	1	.	.	.	1	.	.	.
rs2981582	FGFR2	A	Breast Cancer	1	.	.	.	.	.	.	1
rs429358	APOE	C	Cardiovascular Disea	2	.	.	.	.	1	.	1
rs499818	Chr6	A	Cardiovascular Disea	3	.	.	1	.	.	.	1
rs6857	PVRL2	T	Cardiovascular Disea	1	.	.	.	1	.	.	.
rs7412	APOE	T	Cardiovascular Disea	2	.	.	.	.	1	.	1
rs1333049	CDKN2A/C	C	Coronary Artery Dise	4	.	.	1	1	.	.	1
rs1722821	SMAD3	C	Coronary Artery Dise	2	.	.	.	1	.	.	1

rs1746563	MIA3	A	Coronary Artery Dise	1	.	.	.	.	.	.	1
rs1767213	FMN2	C	Coronary Artery Dise	2	1	.	.	1	.	.	.
rs2943634	2q36.3	A	Coronary Artery Dise	1	.	.	.	.	.	.	1
rs501120	CXCL12	C	Coronary Artery Dise	1	.	.	.	.	.	.	1
rs599839	CELSR2/PS	G	Coronary Artery Dise	6	1	.	1	1	1	.	1
rs6922269	MTHFD1L	A	Coronary Artery Dise	1	.	.	.	.	.	.	1
rs8055236	CDH13	T	Coronary Artery Dise	3	.	.	1	1	.	.	.
rs2144300	GALNT2	C	Coronary Heart Disea	5	.	.	1	1	1	.	1
rs2383206	9p21	A	Coronary Heart Disea	2	.	.	.	1	.	.	1
rs2549513	Chr16	C	Coronary Heart Disea	3	.	.	1	.	.	.	1
rs268	LPL	G	Coronary Heart Disea	1	.	.	.	.	.	.	1
rs1077821	12q23	T	C-reactive Protein	2	1	.	.	1	.	.	.
rs1205	CRP	T	C-reactive Protein	2	1	.	.	1	.	.	.
rs1417938	CRP	A	C-reactive Protein	1	.	.	.	1	.	.	.
rs1554606	IL6	G	C-reactive Protein	2	1	.	.	1	.	.	.
rs1800795	IL6	G	C-reactive Protein	1	.	.	.	1	.	.	.
rs1800947	CRP	G	C-reactive Protein	2	1	.	.	1	.	.	.
rs1892534	LEPR	T	C-reactive Protein	2	1	.	.	1	.	.	.
rs2228145	IL6R	C	C-reactive Protein	5	1	1	1	1	.	.	.
rs3093058	CRP	T	C-reactive Protein	2	1	.	.	1	.	.	.
rs4131568	CRP	T	C-reactive Protein	2	1	.	.	1	.	.	.
rs7310409	HNF1A	A	C-reactive Protein	1	.	.	.	1	.	.	.
rs780094	GCKR	T	C-reactive Protein	6	1	1	1	1	1	.	.
rs1024405		G	Glucose	1	.	.	.	1	.	.	.
rs1736674	ADIPOQ	C	Glucose	1	.	.	.	1	.	.	.
rs560887	G6PC2	T	Glucose	1	.	.	.	1	.	.	.
rs563694	G6PC2	C	Glucose	1	.	.	.	1	.	.	.
rs1088512	ADRA2A	T	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs1155847	SLC30A8	G	Glucose/HOMA-B	1	.	.	.	.	1	.	.
SNP	Gene	Coded alle	Previous trait associa	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs1160592	CRY2	C	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs174550	FADS1	C	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs4506565	TCF7L2	T	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs7034200	GLIS3	C	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs7944584	MADD	T	Glucose/HOMA-B	1	.	.	.	.	1	.	.
rs2877716	ADCY5	T	Glucose/OGTT/T2D	1	.	.	.	.	1	.	.
rs1689097	SLC2A9	T	Gout	1	.	.	1	.	.	.	.
rs2231142	ABCG2	T	Gout	3	.	.	1	1	.	.	.
rs1046801	LIPC	T	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1050366	LPL	A	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1259677	CETP	G	HDL Cholesterol	2	.	.	.	1	1	.	.
rs1267891	LPL	G	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1323432	GRIN3A	G	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1566439	CETP	C	HDL Cholesterol	4	.	.	1	1	1	.	.
rs173539	CETP	T	HDL Cholesterol	1	.	.	.	1	.	.	.
rs174547	FADS1	C	HDL Cholesterol	3	.	.	.	1	1	.	1
rs1800588	LIPC	T	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1800777	CETP	A	HDL Cholesterol	1	.	.	.	1	.	.	.
rs1800961	HNF4A	T	HDL Cholesterol	6	1	1	1	1	1	.	.
rs1864163	CETP	A	HDL Cholesterol	3	1	.	.	1	1	.	.
rs1883025	ABCA1	T	HDL Cholesterol	2	.	.	.	1	1	.	.
rs2083637	LPL	G	HDL Cholesterol	3	.	.	1	1	.	.	.
rs2156552	LIPG	A	HDL Cholesterol	5	.	.	1	1	1	.	1
rs2197089	LPL	G	HDL Cholesterol	5	.	.	1	1	1	.	1
rs2217332	HERPUD1	A	HDL Cholesterol	3	.	.	1	1	.	.	.

rs2271293	LCAT	A	HDL Cholesterol	3	.	.	1	.	.	.	1
rs2338104	MMAB/M	C	HDL Cholesterol	4	.	.	1	1	1	.	.
rs255049	LCAT	C	HDL Cholesterol	1	.	.	.	1	.	.	.
rs261332	LIPC	A	HDL Cholesterol	2	.	.	.	1	1	.	.
rs2892768	APOA1/C3	G	HDL Cholesterol	6	1	1	1	1	.	.	1
rs2967605	ANGPTL4	T	HDL Cholesterol	3	.	.	1	1	.	.	.
rs3135506	APOA1/C3	C	HDL Cholesterol	3	.	.	1	1	.	.	.
rs328	LPL	G	HDL Cholesterol	5	.	.	1	1	1	.	1
rs3764261	CETP	A	HDL Cholesterol	5	.	.	1	1	1	.	1
rs3890182	ABCA1	A	HDL Cholesterol	5	.	.	1	1	1	.	1
rs3905000	ABCA1	A	HDL Cholesterol	3	.	.	1	1	.	.	.
rs4149268	ABCA1	T	HDL Cholesterol	3	.	.	.	1	1	.	1
rs4149274	ABCA1	A	HDL Cholesterol	1	.	.	.	1	.	.	.
rs471364	TTC39B	C	HDL Cholesterol	4	.	.	1	1	1	.	.
SNP	Gene	Coded alle	Previous trait associ	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs4939883	LIPG	T	HDL Cholesterol	3	.	.	1	1	.	.	.
rs6586891	LPL	C	HDL Cholesterol	5	.	.	1	1	1	.	1
rs6754295	APOB	G	HDL Cholesterol	2	.	.	1	.	.	.	.
rs711752	CETP	A	HDL Cholesterol	1	.	.	.	1	.	.	.
rs7120118	NR1H3	C	HDL Cholesterol	3	.	.	1	1	.	.	.
rs7205804	CETP	A	HDL Cholesterol	1	.	.	.	.	.	.	1
rs7395662	MADD-FO	A	HDL Cholesterol	3	.	.	1	1	.	.	.
rs7679	PLTP	C	HDL Cholesterol	1	.	.	.	1	.	.	.
rs9282541	ABCA1	T	HDL Cholesterol	1	.	.	.	1	.	.	.
rs964184	APOA1	G	HDL Cholesterol	4	1	.	.	1	1	.	1
rs9891572		T	HDL Cholesterol	3	.	.	1	1	.	.	.
rs9989419	CETP	A	HDL Cholesterol	5	.	.	1	1	1	.	1
rs2494250	MCP/1	G	Inflammation	1	.	.	.	.	.	.	1
rs4128725	MCP/1	C	Inflammation	1	.	.	.	.	.	.	1
rs6732914	GCG	C	Inflammation	1	.	.	.	.	.	.	1
rs1118579	PANK1	A	Insulin	1	.	.	.	1	.	.	.
rs35767	IGF1	T	insulin/HOMA-IR/IG	1	.	.	.	.	1	.	.
rs1048677		A	Ischemic stroke	3	.	.	1	.	.	.	1
rs1799963	Prothroml	A	Ischemic stroke	1	.	.	.	.	.	.	1
rs1801133	MTHFR/C6	A	Ischemic stroke	3	.	.	1	.	.	.	1
rs6025	F5	T	Ischemic stroke	2	1	.	.	.	.	.	1
rs7506045	IMPA2	T	Ischemic stroke	3	.	.	1	.	.	.	1
rs783396	AIM1	A	Ischemic stroke	3	.	.	1	.	.	.	1
rs9536591		C	Ischemic stroke	3	.	.	1	.	.	.	1
rs4821480	MHY9	G	Kidney disease	3	.	.	1	1	.	.	.
rs102275	FADS1-FA	C	LDL Cholesterol	3	.	.	1	1	.	.	.
rs1040196	NCAN	C	LDL Cholesterol	2	.	.	.	1	1	.	.
rs1040227	APOE/C1/	G	LDL Cholesterol	4	.	.	1	1	1	.	.
rs1120651	PCSK9	C	LDL Cholesterol	4	.	.	1	1	.	.	1
rs1159114	PCSK9	T	LDL Cholesterol	6	1	.	1	1	1	.	1
rs1265426	HMGCR	T	LDL Cholesterol	5	.	.	1	1	1	.	1
rs1269538	B4GALT4	G	LDL Cholesterol	2	.	.	.	1	1	.	.
rs1274037	CELSR2	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs1501908	TIMD4	G	LDL Cholesterol	2	.	.	.	1	1	.	.
rs1529729	LDLR	C	LDL Cholesterol	3	.	.	1	1	.	.	.
rs1535	FADS1-FA	G	LDL Cholesterol	3	.	.	1	1	.	.	.
rs1699614	CILP2/PBX	T	LDL Cholesterol	5	.	.	1	1	1	.	1
rs174537	FADS1-FA	T	LDL Cholesterol	3	.	.	1	1	.	.	.
rs174570	FADS2/3	T	LDL Cholesterol	4	.	.	1	1	1	.	.

SNP	Gene	Coded alle	Previous trait associa	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs2075650	TOMM40	G	LDL Cholesterol	5	1	.	1	1	1	.	.
rs2228671	LDLR	T	LDL Cholesterol	5	1	.	1	1	1	.	.
rs2650000	HNF1A	A	LDL Cholesterol	6	1	1	1	1	1	.	.
rs3846662	HMGCR	G	LDL Cholesterol	3	.	.	1	1	.	.	.
rs3846663	HMGCR	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs4420638	APOE/C1/	G	LDL Cholesterol	5	1	.	1	1	.	.	1
rs4591370	APOB	A	LDL Cholesterol	1	.	.	.	1	.	.	.
rs4605275	BCAM	T	LDL Cholesterol	2	.	.	1	.	.	.	.
rs4803750	BCL3	G	LDL Cholesterol	4	.	.	1	1	1	.	.
rs488507	APOB	G	LDL Cholesterol	1	.	.	.	1	.	.	.
rs4970834	CELSR2/PS	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs515135	APOB	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs562338	APOB	A	LDL Cholesterol	5	.	.	1	1	1	.	1
rs602633	CELSR2/PS	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs6102059	MAFB	T	LDL Cholesterol	2	.	.	.	1	1	.	.
rs646776	CELSR2/PS	C	LDL Cholesterol	6	1	.	1	1	1	.	1
rs6511720	LDLR	T	LDL Cholesterol	7	1	1	1	1	1	.	1
rs6544713	ABCGB	T	LDL Cholesterol	3	.	.	.	1	1	.	1
rs6756629	ABCG5	A	LDL Cholesterol	3	.	.	1	1	.	.	.
rs693	APOB	A	LDL Cholesterol	6	.	1	1	1	1	.	1
rs754523	APOB	G	LDL Cholesterol	4	.	.	1	1	.	.	1
rs7575840	APOB	T	LDL Cholesterol	1	.	.	.	1	.	.	.
rs1075727	CDKN2A/C	A	Myocardial infarctio	4	.	.	1	1	.	.	1
rs2383207	CDKN2A/C	G	Myocardial infarctio	4	.	.	1	1	.	.	1
rs2383208	CDKN2A/C	G	Myocardial infarctio	5	.	.	1	1	1	1	1
rs4804611	ZNF627	G	Myocardial infarctio	1	.	.	.	.	.	.	1
rs4977574	CKN2A, CD	A	Myocardial infarctio	1	.	.	.	.	.	.	1
rs1106683	7q32.3	A	Obesity	1	.	.	.	.	.	.	1
rs8050136	FTO	A	Obesity	6	1	.	1	1	1	1	1
rs9930506	FTO	G	Obesity	2	.	.	.	.	1	.	1
rs9941349	FTO	T	Obesity	1	.	.	.	.	1	.	.
rs1421085	FTO	C	Obesity, Waist circu	1	.	.	.	.	1	.	.
rs7935346		A	Osteoporotic fractur	1	.	.	.	1	.	.	.
rs1242579	NINJ2	A	Stroke	1	.	.	.	.	.	.	1
rs2200733	PITX2/TIF/	T	Stroke	3	.	.	1	.	.	.	1
rs2304130	NCAN	G	Total cholesterol	3	.	.	1	1	.	.	.
rs6987702	TRIB1	C	Total cholesterol	3	.	.	1	1	.	.	.
rs1009663	LPL	T	Triglycerides	2	.	.	1	.	.	.	.
rs1088935	ANGPTL3	C	Triglycerides	4	.	.	1	1	1	.	.
SNP	Gene	Coded alle	Previous trait associa	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs1112712	RBKS/GCK	G	Triglycerides	1	.	.	.	1	.	.	.
rs1213033	ANGPTL3	T	Triglycerides	3	.	.	1	1	.	.	.
rs1214069	LOC40079(	T	Triglycerides	1	.	.	.	1	.	.	.
rs1227200	APOA1/C3	A	Triglycerides	3	.	.	1	1	.	.	.
rs1228603	APOA5/A4	T	Triglycerides	2	1	.	.	1	.	.	.
rs1260326	GCKR	T	Triglycerides	6	1	.	1	1	1	.	1
rs1471233		T	Triglycerides	2	.	.	.	1	1	.	.
rs1558861	APOA1/C3	C	Triglycerides	1	.	.	.	1	.	.	.
rs1714573	MLXIPL	T	Triglycerides	5	.	.	1	1	1	.	1
rs1721652	NCAN	T	Triglycerides	1	.	.	.	1	.	.	.
rs1732151	TRIB1	G	Triglycerides	1	.	.	.	1	.	.	.
rs1741091	LPL	A	Triglycerides	1	.	.	.	1	.	.	.
rs1748195	ANGPTL3	G	Triglycerides	3	.	.	.	1	1	.	1
rs1748275	LPL	T	Triglycerides	1	.	.	.	1	.	.	.
rs2000571	APOA1/C3	A	Triglycerides	4	1	.	1	1	.	.	.

rs2074755	BAZ1B	C	Triglycerides	1	.	.	.	1	.	.	.
rs2075292	APOA1	G	Triglycerides	3	.	.	1	1	.	.	.
rs2954029	TRIB1	T	Triglycerides	4	1	.	.	1	1	.	1
rs4740635		C	Triglycerides	1	.	.	.	1	.	.	.
rs4775041	LIPC	C	Triglycerides	5	.	.	1	1	1	.	1
rs486394	APOA1/C3	C	Triglycerides	1	.	.	.	1	.	.	.
rs6589566	APOA1/C3	G	Triglycerides	1	.	.	.	1	.	.	.
rs673548	APOB	A	Triglycerides	3	.	.	1	1	.	.	.
rs714052	MLXIPL	G	Triglycerides	3	.	.	1	1	.	.	.
rs7557067	APOB	G	Triglycerides	1	.	.	.	1	.	.	.
rs7819412	XKR6	G	Triglycerides	1	.	.	.	1	.	.	.
rs7861175		C	Triglycerides	1	.	.	.	1	.	.	.
rs1001013	WFS1	A	Type 2 Diabetes	3	.	.	1	.	.	1	1
rs1049772	Chr2	A	Type 2 Diabetes	1	.	.	.	.	.	.	1
rs1081166	CDKN2A/(	C	Type 2 Diabetes	4	.	.	1	1	.	1	1
rs1083096	MTNR1B	G	Type 2 Diabetes	3	.	.	.	1	1	.	1
rs1092393	NOTCH2	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs1094639	CDKAL1	C	Type 2 Diabetes	2	.	.	1	.	.	1	.
rs1111875	HHEX/IDE	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs1113132	EXT2	C	Type 2 diabetes	1	.	.	.	1	.	.	.
rs1244069	C2CDB4	C	Type 2 Diabetes	1	.	.	.	.	1	.	.
rs1251809	LOC72901,	G	Type 2 Diabetes	1	.	.	.	.	1	.	.
rs1277979	CDC123/C	G	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs1326663	SLC30A8	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
SNP	Gene	Coded alle	Previous trait associa	# Studies	ARIC	CARDIA	CHS	EAGLE	MEC	SHS	WHI
rs1470579	IGF2BP2	C	Type 2 diabetes	3	.	.	1	1	.	.	.
rs1801282	PPARG	G	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs2074196	KCNQ1	T	Type 2 Diabetes	1	.	.	.	.	1	.	.
rs2237892	KCNQ1	T	Type 2 Diabetes	3	.	.	1	.	1	.	.
rs2237895	KCNQ1	C	Type 2 Diabetes	2	.	.	.	1	1	.	.
rs2237897	KCNQ1	T	Type 2 Diabetes	2	.	.	.	1	1	.	.
rs2943641	LOC64673,	T	Type 2 Diabetes	1	.	.	.	.	1	.	.
rs4402960	IGF2BP2	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs4607103	ADAMTS9	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs4689388	WFS1, PPI	G	Type 2 Diabetes	1	.	.	.	.	1	.	.
rs4712523	CDKAL1	G	Type 2 Diabetes	4	.	.	1	1	.	.	1
rs4760920	TSPAN8	C	Type 2 Diabetes	3	.	.	1	.	.	.	1
rs5215	KCNJ11	C	Type 2 Diabetes	4	1	.	1	.	.	1	1
rs5219	KCNJ11	T	Type 2 Diabetes	4	.	.	1	1	.	1	1
rs564398	CDKN2A/(	C	Type 2 Diabetes	1	.	.	.	.	.	.	1
rs7480010	LOC38776	G	Type 2 diabetes	3	.	.	1	1	.	.	.
rs7578597	THADA	C	Type 2 Diabetes	4	.	.	1	1	.	1	1
rs7754840	CDKAL1	C	Type 2 Diabetes	2	.	.	.	1	1	.	.
rs7901695	TCF7L2	C	Type 2 Diabetes	4	.	.	1	1	.	1	1
rs7903146	TCF7L2	T	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs7923837	HHEX	A	Type 2 diabetes	3	.	.	1	1	.	.	.
rs7961581	TSPAN8/L	C	Type 2 Diabetes	4	.	.	1	1	1	1	.
rs864745	JAZF1	C	Type 2 Diabetes	5	.	.	1	1	1	1	1
rs1170806	ADCY5	G	Type 2 Diabetes; Glu	1	.	.	.	.	1	.	.
rs2191349	DGKB-TMI	T	Type 2 Diabetes; Glu	1	.	.	.	.	1	.	.
rs340874	PROX1	A	Type 2 Diabetes; Glu	1	.	.	.	.	1	.	.
rs6855911	GLUT9	G	Uric acid	1	.	.	.	1	.	.	.
rs7442295	SLC2A9	G	Uric acid	1	.	.	.	1	.	.	.
rs1014699	NRXN3	G	Waist circumference	1	.	.	.	.	1	.	.
rs545854	MSRA	C	Waist circumference	1	.	.	.	.	1	.	.
rs987237	TFAP2B	G	Waist circumference	1	.	.	.	.	1	.	.
rs2605100	LYPLAL1	A	Waist/hip in women	1	.	.	.	.	1	.	.
				Total	30	7	113	196	103	19	94

Analysis I – Inflammation Appendix 3

Appendix 3 - Number of participants available for SNP analyses across studies for each race/ethnicity group. *Results excluded due to <100 participants.							
SNP	European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs10010131	6820	2597	777	422	3129		13745
rs10096633	3244	590	.	.	.		3834
rs10146997	255	450	251	527	.		1483
rs102275	7199	1920	1856	.	.		10975
rs10244051	3966	1330	1853	.	.		7149
rs1038304	3957	1328	1856	.	.		7141
rs10401969	4223	1780	2117	529	.		8649
rs10402271	7436	2370	2100	527	.		12433
rs10468017	3949	1333	1862	.	.		7144
rs10486776	10049	2386	776	422	*		13633
rs10490924	3958	1328	1871	.	.		7157
rs10497721	6817	1799	777	422	*		9815
rs10498767	3915	1322	1841	.	.		7078
rs10503669	3946	1320	1867	.	.		7133
rs1061170	3827	1280	1832	.	.		6939
rs10757278	13948	3923	2642	422	*		20935
rs10778213	18056	5412	1858	.	.		25326
rs10811661	13961	3922	2620	422	3129		24054
rs10830963	7943	2996	2503	632	*		14074
rs10838738	14279	4181	2907	942	*		22309
rs10885122	258	454	257	527	.		1496
rs10889353	7432	2375	2100	528	.		12435
rs10913469	259	454	257	525	.		1495
rs10923931	14231	4378	2891	951	3129		25580
rs10938397	10992	4339	2869	932	*		19132
rs10946398	3193	800	.	.	3129		7122
rs1106683	6819	1799	777	422	*		9817
rs1107748	3958	1326	1856	.	.		7140
rs11084753	14210	4378	2891	950	*		22429
rs1111875	14264	4154	2876	949	3129		25372
rs11127129	3955	1329	1855	.	.		7139
rs1113132	3964	1330	1859	.	.		7153
rs11185790	3961	1331	1858	.	.		7150
rs11206510	13994	3938	2649	422	*		21003
rs1121980	255	454	255	525	.		1489
rs11558471	259	456	256	528	.		1499
rs11591147	22665	6717	2874	943	*		33199
rs11605924	259	454	257	529	.		1499
rs11708067	259	454	255	527	.		1495
rs1205	12400	3669	1842	.	.		17911
rs12130333	7178	1916	1859	.	.		10953

SNP	European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs12140698	3962	1329	1861	.	.		7152
rs12272004	7185	1914	1861	.	.		10960
rs12286037	12408	3669	1862	.	.		17939
rs12425791	4178	1614	725	388	*		6905
rs12440695	255	453	255	527	.		1490
rs12518099	259	452	255	527	.		1493
rs12596776	4170	1758	2096	526	.		8550
rs1260326	22661	6506	2873	951	*		32991
rs12654264	14205	4361	2870	943	*		22379
rs12678919	3966	1336	1861	.	.		7163
rs12695382	4228	1789	2119	527	.		8663
rs12740374	3959	1321	1860	.	.		7140
rs12779790	14216	4378	2887	948	3129		25558
rs12970134	7198	1914	1856	.	.		10968
rs1323432	3941	1320	1842	.	.		7103
rs13266634	14269	2358	2886	523	3129		23165
rs1333026	10051	2389	777	421	*		13638
rs1333049	13969	3923	2635	422	*		20949
rs1417938	3943	1328	1845	.	.		7116
rs1421085	252	444	245	516	.		1457
rs1470579	7187	1908	1855	.	.		10950
rs1471233	4215	1785	2116	525	.		8641
rs1501908	4205	2568	2113	524	.		9410
rs1513670	3960	1331	1859	.	.		7150
rs1529729	7170	1908	1838	.	.		10916
rs1535	7199	1922	1856	.	.		10977
rs1554606	12292	3667	1585	.	.		17544
rs1558861	3967	1330	1860	.	.		7157
rs1566439	7440	2357	2106	526	.		12429
rs16890979	.	802	.	.	.		802
rs16996148	14220	4375	2891	948	*		22434
rs17145738	14237	4385	2905	948	*		22475
rs17216525	3971	1336	1862	.	.		7169
rs17228212	10783	3129	2637	422	*		16971
rs17321515	3951	1328	1854	.	.		7133
rs173539	3652	1175	1696	.	.		6523
rs17366743	3937	1321	1844	.	.		7102
rs17410914	3966	1331	1858	.	.		7155
rs174537	7209	1917	1853	.	.		10979

SNP	European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs174547	11002	4389	2893	526	*		18810
rs174550	255	454	252	523	.		1484
rs174570	7453	2372	2107	528	.		12460
rs17465637	6821	1798	777	422	*		9818
rs1748195	10991	3577	2906	948	*		18422
rs17482753	3961	1327	1857	.	.		7145
rs17672135	12365	3677	1856	.	.		17898
rs17782313	21563	6504	2900	946	*		31913
rs1799963	6812	1799	777	422	*		9810
rs1800588	3946	1319	1842	.	.		7107
rs1800777	3970	1330	1861	.	.		7161
rs1800795	3940	1325	1851	.	.		7116
rs1800947	12404	3676	1842	.	.		17922
rs1800961	17660	6222	2117	523	.		26522
rs1801133	10050	2387	777	422	*		13636
rs1801282	14212	4371	2874	948	3129		25534
rs1864163	12672	4126	2113	525	.		19436
rs1883025	4211	1769	2114	526	.		8620
rs1892534	11344	3471	1851	.	.		16666
rs1999805	3955	1323	1857	.	.		7135
rs2000571	15678	4271	1859	.	.		21808
rs2074196	259	454	256	529	.		1498
rs2074755	3961	1331	1855	.	.		7147
rs2075292	7205	1916	1853	.	.		10974
rs2075650	15926	4729	2119	528	.		23302
rs2083637	7197	1919	1860	.	.		10976
rs2144300	14137	4370	2908	944	*		22359
rs2156552	14226	4384	2892	946	*		22448
rs2191349	258	452	256	528	.		1494
rs2197089	14221	4383	2898	947	*		22449
rs2200733	10065	2389	777	422	*		13653
rs2217332	7187	1916	1840	.	.		10943
rs2228145	17184	5942	1841	.	.		24967
rs2228671	15900	4724	2101	523	.		23248
rs2231142	7203	1918	1872	.	.		10993
rs2237892	3494	1033	253	523	.		5303
rs2237895	4193	1779	2122	519	.		8613
rs2237897	3961	1670	1992	529	.		8152
rs2271293	10059	2388	777	422	*		13646
rs2304130	7206	1921	1859	.	.		10986
rs2338104	7422	2585	2127	526	.		12660

SNP		European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs2383206		10766	3126	2633	422	*		16947
rs2383207		13981	1926	2625	*	*		18532
rs2383208		14265	4157	2891	946	3129		25388
rs2494250		6820	1799	776	422	*		9817
rs2549513		9995	2379	777	422	*		13573
rs255049		3966	1331	1859	.	.		7156
rs2605100		257	454	254	529	.		1494
rs261332		3987	1675	2016	511	.		8189
rs2650000		16506	5983	2125	500	.		25114
rs268		6818	1799	777	422	*		9816
rs2815752		14257	4174	2895	946	*		22272
rs2877716		233	433	243	515	.		1424
rs28927680		24095	7694	2632	422	*		34843
rs2943634		6783	1791	774	421	*		9769
rs2943641		258	456	255	528	.		1497
rs2954029		19480	5931	2889	949	*		29249
rs2967605		7135	2136	1853	.	.		11124
rs2981582		6820	1799	777	422	*		9818
rs29941		251	437	247	512	.		1447
rs3018362		3955	1331	1854	.	.		7140
rs3093058		12420	3680	1842	.	.		17942
rs3130340		3960	1331	1858	.	.		7149
rs3135506		7142	2122	1841	.	.		11105
rs328		14216	4368	2875	948	*		22407
rs340874		259	455	257	529	.		1500
rs35767		256	456	256	526	.		1494
rs3736228		3945	1329	1868	.	.		7142
rs3751812		10312	2844	1032	948	*		15136
rs3764261		14222	4157	2889	945	*		22213
rs3771362		3954	1327	1856	.	.		7137
rs3846662		7175	1920	1856	.	.		10951
rs3846663		3782	1242	1761	.	.		6785
rs3890182		14213	4372	2878	946	*		22409
rs3905000		7203	1920	1871	.	.		10994
rs4128725		6821	1799	777	422	*		9819
rs4131568		11314	3461	1857	.	.		16632
rs4149268		11016	3577	2892	946	*		18431
rs4149274		3960	1327	1856	.	.		7143
rs429358		6993	2220	1031	943	*		11187

SNP		European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs4355801		3967	1335	1872	.	.		7174
rs4402960		14210	4371	2876	946	3129		25532
rs4420638		21378	6076	2637	422	*		30513
rs4506565		255	449	255	521	.		1480
rs4591370		3966	1331	1862	.	.		7159
rs4605275		3211	565	.	.	.		3776
rs4607103		14225	4364	2889	948	3129		25555
rs4689388		252	454	257	527	.		1490
rs4712523		14013	3717	2649	422	*		20801
rs471364		7467	2368	2119	523	.		12477
rs4740635		3963	1330	1861	.	.		7154
rs4760920		7383	2201	725	388	*		10697
rs4775041		14200	4377	2885	949	*		22411
rs4803750		7441	2372	2111	528	.		12452
rs4804611		6805	1795	777	422	*		9799
rs4821480		7166	2138	1865	.	.		11169
rs486394		3944	1323	1848	.	.		7115
rs4870044		3951	1331	1858	.	.		7140
rs488507		3930	1326	1853	.	.		7109
rs4939883		7200	1919	1875	.	.		10994
rs4970834		3812	1312	1588	.	.		6712
rs4977574		4181	1611	723	388	*		6903
rs499818		10046	2387	776	422	*		13631
rs501120		6813	1789	777	422	*		9801
rs515135		3961	1324	1869	.	.		7154
rs5215		18429	4736	776	421	3129		27491
rs5219		13945	3924	2645	418	3129		24061
rs545854		257	455	254	529	.		1495
rs560887		3665	1272	1805	.	.		6742
rs562338		14224	4378	2900	947	*		22449
rs563694		3804	1310	1579	.	.		6693
rs564398		6813	1766	767	422	*		9768
rs599839		22427	6596	2811	946	*		32780
rs6013029		10769	3121	2631	422	*		16943
rs6025		15292	4156	777	422	*		20647
rs602633		3961	1326	1854	.	.		7141
rs6102059		4195	1775	2115	521	.		8606
rs6265		4133	1779	2126	523	.		8561
rs646776		22652	6479	2889	951	*		32971
rs6469804		3943	1326	1857	.	.		7126
rs6511720		24319	8158	2875	947	*		36299

SNP		European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs6544713		11045	4383	2906	947	*		19281
rs6548238		13985	3932	2649	422	*		20988
rs6586891		14206	4376	2895	950	*		22427
rs6589566		3964	1329	1855	.	.		7148
rs6602024		3959	1327	1854	.	.		7140
rs6732914		6815	1796	776	422	*		9809
rs673548		7177	1907	1839	.	.		10923
rs6754295		3244	590		.	.		3834
rs6756629		7208	1923	1861	.	.		10992
rs6855911		3932	1326	1865	.	.		7123
rs6857		3955	1331	1852	.	.		7138
rs6922269		6814	1796	777	421	*		9808
rs6929137		3961	1329	1854	.	.		7144
rs693		15953	5630	2873	943	*		25399
rs6987702		7188	1919	1856	.	.		10963
rs6993813		3959	1331	1854	.	.		7144
rs7034200		259	453	257	527	.		1496
rs711752		3955	1327	1856	.	.		7138
rs7120118		7053	1903	1586	.	.		10542
rs7138803		258	453	257	525	.		1493
rs714052		7189	1902	1857	.	.		10948
rs7205804		4173	1613	724	388	*		6898
rs7220711		3947	1319	1849	.	.		7115
rs7310409		3956	1327	1855	.	.		7138
rs7395662		7200	1920	1857	.	.		10977
rs7412		6991	2239	1031	945	*		11206
rs7442295		3944	1333	1855	.	.		7132
rs7480010		7210	1919	1857	.	.		10986
rs7498665		14244	4175	2899	945	*		22263
rs7506045		10064	2386	777	421	*		13648
rs7524102		3964	1329	1861	.	.		7154
rs754523		13946	3708	2630	421	*		20705
rs7557067		3968	1335	1862	.	.		7165
rs7566605		3962	1329	1874	.	.		7165
rs7575840		3964	1330	1859	.	.		7153
rs7578597		14018	3719	2635	422	3129		23923
rs7595412		3959	1330	1855	.	.		7144
rs7679		3964	2131	1874	.	.		7969
rs7754840		4206	1778	2115	527	.		8626
rs780094		17567	6175	2130	528	.		26400
rs7819412		3915	1308	1860	.	.		7083

SNP		European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
rs783396		10044	2386	771	422	*		13623
rs7861175		3967	1331	1860	.	.		7158
rs7901695		13937	3683	2621	422	3129		23792
rs7903146		14259	4162	2872	947	3129		25369
rs7923837		7087	1901	1837	.	.		10825
rs7935346		3943	1326	1850	.	.		7119
rs7944584		257	456	257	528	.		1498
rs7961581		7361	2575	2111	526	3129		15702
rs8050136		21659	6303	2869	948	3129		34908
rs8055236		7175	1909	1853	.	.		10937
rs864745		14241	4369	2896	948	3129		25583
rs925946		255	451	252	520	.		1478
rs9282541		3969	1332	1860	.	.		7161
rs9479055		3962	1326	1860	.	.		7148
rs9536591		10049	2384	776	422	*		13631
rs9594738		3958	1326	1852	.	.		7136
rs9594759		3956	1331	1852	.	.		7139
rs964184		18193	5636	2890	949	*		27668
rs987237		258	450	251	525	.		1484
rs9891572		7199	1907	1851	.	.		10957
rs9930506		7049	2251	1029	948	*		11277
rs9939609		3454	1254	252	525	.		5485
rs9941349		258	453	253	524	.		1488
rs9989419		14243	4124	2877	944	*		22188
		European American	African American	Hispanic	Asian/Pacific Islander	American Indian		OVERALL
Minumum		233	433	243	388	3129		802
Maximum		24319	8158	2908	951	3129		36299
Mean		7722	2367	1766	617	3129		12361
Median		6820	1850	1857	526	3129		9818
# SNPs		265	266	261	150	19		266

Analysis I CRP – Inflammation Appendix 4

Table 4 - Correlation (r2) between SNPs significantly associated with CRP.
Values are obtained from 1000G data for CEU, YRI, and HCB/JPT using SNAP.

CRP					**APOE/APOC1/TOMM40**				
	European Americans					European Americans			
	rs1205	rs1800947	rs1417938	rs4131568		rs6857	rs7412	rs429358	rs2075650
rs1800947	0.142				rs4420638	0.627			0.526
rs1417938	0.253	0.036			rs2075650	0.844	0.015		
rs4131568	0.119	0.04	0.854		rs429358		0.004		
rs3093058					rs7412				
	African Americans					African Americans			
	rs1205	rs1800947	rs1417938	rs4131568		rs6857	rs7412	rs429358	rs2075650
rs1800947					rs4420638	0.04		0.016	0.027
rs1417938	0.005				rs2075650	0.004	0.06	0.01	
rs4131568					rs429358	0	0.017		
rs3093058	0.001	0.025	0.009		rs7412				
	Asian/Pacific Islanders					Asian/Pacific Islanders			
	rs1205	rs1800947	rs1417938	rs4131568		rs6857	rs7412	rs429358	rs2075650
rs1800947	0.013				rs4420638	0.507		0.115	0.457
rs1417938	0.125	0.002			rs2075650	0.905	0.004	0.073	
rs4131568	0.093	0.001	0.741		rs429358	0.08			
rs3093058					rs7412				

GCKR		**HNF1A**	
	rs1260326 and rs780094		rs2650000 and rs7310409
European Americans	0.933	European Americans	0.47
African Americans	0.468	African Americans	0.069
Asian/Pacific Islande	0.832	Asian/Pacific Islande	0.651

CELSR2/PSRC1/SORT1	
	rs599839 and rs646776
European Americans	0.895
African Americans	0.059
Asian/Pacific Islande	0.702

Analysis II – Melanoma Appendix 1

Additional information on each of the five participating studies is provided below.

Epidemiologic Architecture for Genes Linked to Environment (EAGLE) accessing BioVU (1). BioVU is a biorepository of DNA samples extracted from blood drawn for routine clinical care. DNA samples are linked to a de-identified version of the patient's electronic medical records for research purposes unless the patient opts-out out of the biorepository via the consent to treatment form. The Vanderbilt electronic medical record (EMR) began accumulating clinical data in the early 1990s. Biological sample collection for BioVU began in 2007, with an accrual rate of ~700 samples per week. Updating of EMR information and genotyping of additional samples is ongoing. Genotyping was conducted by the Vanderbilt DNA Resources Core with the use of the mid-throughput Sequenom genotyping platform and TaqMan assays and the ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Eight major cancers, including melanoma, were defined using tumor registry entries, billing codes (ICD9 codes), and procedure codes as previously described (2). Two cancer-free controls matched by sex and race/ethnicity and frequency matched by age (roughly ± five years) were identified per cancer case. When appropriate, controls matched to cases of other cancers are used as additional controls to improve power.

Multiethnic Cohort Study (MEC) (3). MEC was initiated in 1993 to investigate the impact of dietary and environmental factors on major chronic diseases, particularly cancer, in ethnically diverse populations in Hawai'i and California. The study recruited 96,810 men and 118,441 women aged 45 to 75 years between 1993 and 1996. Incident cancer cases, occurring since January 1995, and controls were contacted for blood or saliva samples. The median interval between cancer diagnosis and blood draw was 14 months (interquartile range, 10-19) among cases, with a participation rate of 74%. A sample of cohort participants was randomly selected to serve as controls at the onset of the nested case-control study (participation rate 66%). The selection was stratified by sex and race/ethnicity. Melanoma cases are identified through the Rapid Reporting System of the Hawai'i Tumor Registry and through quarterly linkage to the Los Angeles County Cancer Surveillance Program. Both registries are members of SEER. Controls were matched on sex, year of birth (± 1 year), ethnicity (White, African American, Hawaiian, Japanese, Latino), date of blood draw (± 6 months), time of day of blood draw (AM or PM), fasting status (0-6, 6-8, 8-10, or 10+ hours), study (HI-GS, HI-PPG, LA-GS), alive at case diagnosis, and type of urine (first morning, overnight, none). If no controls matched, then a relaxed criteria was used: date of blood draw (± 1 year), year of birth (± 5 years). Controls were preferentially selected from those participants with biomarker data available, such as lipids and glucose. A case-control set was created for each of the cancer types investigated as part of collaboration with PAGE (breast cancer, endometrial cancer, lung cancer, ovarian cancer, melanoma, Non-Hodgkin lymphoma, and colorectal cancer). Where eligible, controls from other cancer control sets (for example, colorectal cancer) were added to the control set for the cancer of interest (here melanoma) to improve power.

Women's Health Initiative (WHI) (5). The Women's Health Initiative (WHI) is a long-term health study of 161,808 post-menopausal women aged 50 to 79 years at 40 clinical centers throughout the U.S. WHI comprises a Clinical Trial (CT) arm, an Observational Study (OS) arm, and several extension studies. The details of WHI have been previously described (5-6), and are available online (https://cleo.whi.org/SitePages/Home.aspx). Melanoma cases occurred through August 2009, and were excluded if they had baseline history of melanoma or a previous incident cancer. Participants with two or more incident cancers were only included as a case for the first cancer type. Controls were required to be free of prevalent caner (including non-melanoma skin cancer) and incident cancer (excluding non-melanoma skin cancer). Controls were matched to cases on age (± 1 year for Whites, ± 2 years for Blacks, ± 3 years for other race/ethnicities), enrollment date (± 365 days), race/ethnicity (White, Black, Hispanic, American Indian, Asian/Pacific Islander), and randomization arms (OS flag, HRT assignments, CaD assignments). Matching was done individually for each of several cancer types in sequential order. Invasive breast cancer was matched first, followed by endometrial cancer, lung cancer, ovarian cancer, melanoma, Non-Hodgkin lymphoma, and colorectal cancer. After each outcome match, those controls selected were reused for the next outcome as the top priority to match from. Control selection was done in a time-forward manner. Each matching factor was given the same weight. Cases and controls are matched 1:1. The matching algorithm was allowed to select the closest match based on a criterion to minimize an overall distance measure (7). SAS code was available to implement this matching scheme. In order to increase sample size for this analysis, cases for other cancers (for example, colorectal cancer) were also used as additional controls for other cancers (here melanoma).

Nurses' Health Study (NHS) (8-9). The Nurses' Health Study (NHS) was established in 1976,

when 121,700 female registered nurses between the ages of 30 and 55 residing in 11 large US states completed and returned the initial self-administered questionnaire on their medical histories and baseline health-related exposures, forming the basis for the NHS cohort. Updated information has been obtained by questionnaires every 2 years, including exposure information on risk factors and outcome data with appropriate follow-up of reported disease events. Overall follow-up has been very high, with ~90% of participants continuing to complete questionnaires, even after more than 20 years. Information on melanoma development was first collected in the 1984 questionnaire. Eligible cases in this study consisted of women with incident skin cancer from the subcohort who gave a blood specimen in 1989–1990 (n = 32,826), and who had melanoma diagnosed any time after baseline up to the 2008 follow-up cycle. Eligible controls were participants from this subcohort without a melanoma diagnosis. All subjects were drawn from among the US non-Hispanic Caucasian women in this study. Continued below.

Health Professionals Follow-up Study (HPFS) (10). The HPFS is a parallel prospective study to the Nurses' Health Study (NHS). The HPFS cohort comprises 51,529 men who, in 1986, responded to a mailed questionnaire. The participants are U.S. male dentists, optometrists, osteopaths, podiatrists, pharmacists, and veterinarians born between 1910 and 1946. Biennial

questionnaires collect disease and health-related information. Follow-up has been excellent, with 94% of the men responding to date. Information on melanoma development was first collected in the 1986 questionnaire. Between 1993 and 1994, 18,159 study participants provided blood samples by overnight courier. Eligible cases in this study consisted of men with incident skin cancer from the subcohort who gave a blood specimen, and who had melanoma diagnosed any time after baseline up to the 2008 follow-up cycle. Eligible controls were participants from this subcohort without a melanoma diagnosis. All subjects were drawn from among the US non-Hispanic Caucasian men in this study. Continued below.

NHS and HPFS: Both cohorts have previously conducted several GWASs on different disease outcomes (NHS: breast cancer, coronary heart disease, type 2 diabetes, kidney stone, pancreatic cancer, and glaucoma; HPFS: coronary heart disease, type 2 diabetes, kidney stone, advanced prostate cancer, and glaucoma). For the purposes of a melanoma GWAS, controls were compiled from each of these prior GWAS (except for the kidney stone GWAS, in which both cases and controls were used). From this data, participants with melanoma diagnosis were the cases and participants without melanoma diagnosis were the controls. Also included were any melanoma cases in both NHS and HPFS who were not included in these previous GWAS. Genotyping of cases and controls occurred in the previous GWAS studies. For the breast cancer GWAS genotyping in NHS was performed on the Illumina HumanHap550 array, as part of the National Cancer Institute's Cancer Genetic Markers of Susceptibility (CGEMS) Project. For the coronary heart disease and type 2 diabetes GWASs, genotyping was performed using the Affymetrix 6.0 array. For the glaucoma GWAS, genotyping was performed using the Illumina HumanHap660 array. For the kidney stone, advanced prostate cancer, and melanoma GWASs, genotyping was performed using the Illumina HumanHap610 array. Based on the genotyped SNPs and haplotype information in the NCBI build 35 of phase II Hapmap CEU data, genotypes were imputed for >2.5 million SNPs using the program MACH (11). Only SNPs with imputation quality $R^2 > 0.95$ in each study were included in the final analysis. A total of 1,579,307 SNPs were included in the final melanoma meta-analysis of the NHS and HPFS. Betas from each study were combined in a meta-analysis with weights proportional to the inverse variance of the beta in each study.

1. Roden DM, Pulley JM, Basford MA, et al. Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. *Clin Pharmacol Ther*. 2008;84(3):362-369.

2. Bush W, Boston J, Pendergrass SA, et al. Enabling high-throughput genotype-phenotype associations in the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) project as part of the Population Architecture using Genomics and Epidemiology (PAGE) study. *Pacific Symposium on Biocomputing* 2013.

3. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol*. 2000;151(4):346-357.

5. Hays J, Hunt JR, Hubbell FA, et al. The women's health initiative recruitment methods and results. *Annals of Epidemiology*. 2003;13(9, Supplement):S18-S77.

6. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials*. 1998;19(1):61-109.

7. Bergstralh EJ & Kosanke JL. Computerized matching of cases to controls. Technical Report #56, Department of Health Sciences Research, Mayo Clinic, Rochester MN. April 1995.

8. Colditz GA & Hanksinson SE. The Nurses' Health Study: Lifestyle and health among women. *Nat Rev Cancer*. 2005;5:388-96.

9. Belanger CF, Hennekens CH, Rosner B*, et al.* The nurses' health study. Am J Nurs 1978;78(6):1039-40.

10. Rimm EB, Stampfer MJ, Colditz GA*, et al.* Validity of self-reported waist and hip circumferences in men and women. Epidemiology 1990;1(6):466-73.

11. Li Y, Willer CJ, Ding J, Scheet P and Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 34:816-834.

Analysis II – Melanoma Appendix 2

Appendix 2 - Full list of 189 SNPs evaluated for an association with melanoma, providing the allele associated with increased risk in the original cancer GWAS publication, as well as the chromosomal location of the SNP and gene.

SNP	k alle	Original cancer association	Mapped Gene	Region	Position	Author Year Journal	Manuscript title
rs10086908	T	Prostate cancer	*Intergenic*	8q24.21	128081119	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with pros
rs10090154	A	Prostate cancer	*Intergenic*	8q24.21	128601319	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with pros
rs1016343	T	Prostate cancer	*Intergenic*	8q24.21	128162479	Eeles 2008 Nat Genet	Multiple newly identified loci associated
rs10220831	T	Non-Hodgkin lymphoma (CLL)	*Intergenic*	15q23	67779168	Di Bernardo 2008 NatGenet	A genome-wide association study identifi
rs10263639	C	Breast cancer	*Intergenic*	7q11.22	66696702	Murabito 2007 BMC Med Genet	A genome-wide association study of brea
rs1036935	T	Non-Hodgkin lymphoma (CLL)	*Intergenic*	18q21.1	46097532	Crowther-Swanepoel 2010 Nat Genet	Common variants at 2q37.3, 8q24.21, 15q2
rs10411210	C	Colorectal cancer	*RHPN2*	19q13.11	38224140	Houlston 2008 Nat Genet	Meta-analysis of genome-wide associatio
rs1045485	G	Breast cancer	*CASP8*	2q33.1	201857834	Cox 2007 Nat Genet	A common coding variant in CASP8 is asso
rs10464870	C	Glioma	*Intergenic*	8q24.21	130547005	Shete 2009 Nat Genet	Genome-wide association study identifie
rs10483813	T	Breast cancer	*RAD51L1*	14q24.1	68101037	Thomas 2009 Nat Genet	A multistage genome-wide association st
rs10486567	G	Prostate cancer	*JAZF1*	7p15.2	27943088	Eeles 2008 Nat Genet	Multiple newly identified loci associated
rs10490113	C	Breast cancer	*Intergenic*	2p16.1	59352851	Murabito 2007 BMC Med Genet	A genome-wide association study of brea
rs10505477	A	Colorectal cancer	*Intergenic*	8q24.21	128476625	Zanke 2007 Nat Genet	Genome-wide association scan identifies
rs1051730	A	Lung cancer	*CHRNA3*	15q25.1	76681394	Landi 2009 Am J Hum Genet	A Genome-wide Association Study of Lun
rs10778826	A	Prostate cancer	*PPFIA2*	12q21.31	80626985	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs10795668	G	Colorectal cancer	*Intergenic*	10p14	8741255	Tomlinson 2008 Nat Genet	A genome-wide association study identifi
rs10821936	C	Leukemia (ALL)	*ARID5B*	10q21.2	63393583	Trevino 2009 Nat Genet	Germline genomic variants associated wit
rs10896449	G	Prostate cancer	*Intergenic*	11q13.3	68751243	Thomas 2008 Nat Genet	Multiple loci identified in a genome-wide
rs10941679	G	Breast cancer	*Intergenic*	5p12	44742255	Stacey 2007 Nat Genet	Common variants on chromosomes 2q35 a
rs10974944	G	Leukemia (Myeloid)	*JAK2*	9p24.1	5060831	Kilpivaara 2009 Nat Genet	A germline JAK2 SNP is associated with pr
rs10993994	T	Prostate cancer	*Intergenic*	10q11.23	51219502	Eeles 2008 Nat Genet	Multiple newly identified loci associated
rs10994982	A	Leukemia (ALL)	*ARID5B*	10q21.2	63380110	Trevino 2009 Nat Genet	Germline genomic variants associated wit
rs11083846	A	Non-Hodgkin lymphoma (CLL)	*PRKD2*	19q13.32	51899494	Di Bernardo 2008 NatGenet	A genome-wide association study identifi
rs11155133	G	Leukemia (ALL)	*Intergenic*	6q24.1	141211518	Trevino 2009 Nat Genet	Germline genomic variants associated wit
rs11170164	A	Basal cell carcinoma	*KRT5*	12q13.13	51199935	Stacey 2009 Nat Genet	New common variants affecting susceptib
rs11228565	A	Prostate cancer	*Intergenic*	11q13.3	68735156	Gudmundsson 2009 Nat Genet	Genome-wide association and replication
rs11249433	C	Breast cancer	*EMBP1*	1p11.2	120982136	Thomas 2009 Nat Genet	A multistage genome-wide association st
rs11649338	C	Breast cancer	*Intergenic*	16q23.1	73028921	Kibriya 2009 Breast Cancer Res Treat	A pilot genome-wide association study of
rs11649743	G	Prostate cancer	*HNF1B*	17q12	33149092	Sun 2008 Nat Genet	Evidence for two independent prostate ca
rs11668878	T	Non-Hodgkin lymphoma (CLL)	*Intergenic*	19q13.32	51960213	Crowther-Swanepoel 2010 Nat Genet	Common variants at 2q37.3, 8q24.21, 15q2
rs11861609	C	Prostate cancer	*CDH13*	16q23.3	81942167	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs12155172	A	Prostate cancer	*Intergenic*	7p15.3	20961016	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs1219648	G	Breast cancer	*FGFR2*	10q26.13	123336180	Hunter 2007 Nat Genet	A genome-wide association study identifi
rs1229984	C	Esophageal cancer	*ADH1B*	4q23	100458342	McKay 2011 PLoS Genet	A genome-wide association study of uppe
rs12418451	A	Prostate cancer	*Intergenic*	11q13.3	68691995	Hsu 2009 Cancer Res	A novel prostate cancer susceptibility locu
rs12500426	A	Prostate cancer	*PDLIM5*	4q22.3	95733632	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs12543663	C	Prostate cancer	*Intergenic*	8q24.21	127993841	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with pros
rs12621278	A	Prostate cancer	*ITGA6*	2q31.1	173019799	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs13252298	A	Prostate cancer	*Intergenic*	8q24.21	128164338	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with pros
rs13254738	C	Prostate cancer	*Intergenic*	8q24.21	128173525	Haiman 2007 Nat Gent	A common genetic risk factor for colorecta
rs13281615	T	Breast cancer	*Intergenic*	8q24.21	128424800	Easton 2007 Nature	Genome-wide association study identifie
rs13387042	A	Breast cancer	*Intergenic*	2q35	217614077	Thomas 2009 Nat Genet	A multistage genome-wide association st
rs13397985	G	Non-Hodgkin lymphoma (CLL)	*SP140*	2q37.1	230799467	Di Bernardo 2008 NatGenet	A genome-wide association study identifi
rs1393350	A	Melanoma	*TYR*	11q14.3	88650694	Bishop 2009 Nat Genet	Genome-wide association study identifie
rs1412829	C	Glioma (high-grade)	*CDKN2BAS1*	9p21.3	22033926	Wrensch 2009 Nat Genet	Variants in the CDKN2B and RTEL1 regions
rs1447295	A	Prostate cancer	*Intergenic*	8q24.21	128554220	Gudmundsson 2007 Nat Genet	Genome-wide association study identifie
rs1465618	T	Prostate cancer	*THADA*	2p21	43407453	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs1512268	T	Prostate cancer	*Intergenic*	8p21.2	23582408	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs1530057	A	Lung cancer	*RBMS3*	3p24.1	29550467	Broderick 2009 Cancer Res	Deciphering the Impact of Common Gene
rs1571801	T	Prostate cancer	*DAB2IP*	9q33.2	123467194	Dugan 2007 J Natl Cancer Inst	Two genome-wide association studies of
rs157935	T	Basal cell carcinoma	*Intergenic*	7q32.3	130236093	Stacey 2009 Nat Genet	New common variants affecting susceptib
rs167020	A	Pancreatic cancer	*Intergenic*	7q36.3	155312494	Amundadottir 2009 Nat Genet	Genome-wide association study identifie
rs16886165	G	Breast cancer	*Intergenic*	5q11.2	56058840	Thomas 2009 Nat Genet	A multistage genome-wide association st
rs16891982	C	Melanoma	*SLC45A2*	5p13.2	33987450	Stokowski 2007 Am J Hum Genet	A genomewide association study of skin p
rs16892766	C	Colorectal cancer	*Intergenic*	8q23.3	117699864	Tomlinson 2008 Nat Genet	A genome-wide association study identifi
rs16901979	A	Prostate cancer	*Intergenic*	8q24.21	128194098	Gudmundsson 2009 Nat Genet	Genome-wide association and replication
rs16902094	G	Prostate cancer	*Intergenic*	8q24.21	128389528	Gudmundsson 2009 Nat Genet	Genome-wide association and replication
rs17021918	C	Prostate cancer	*PDLIM5*	4q22.3	95781900	Eeles 2009 Nat Genet	Identification of seven new prostate canc
rs172310	A	Pancreatic cancer	*Intergenic*	7q36.3	155308388	Amundadottir 2009 Nat Genet	Genome-wide association study identifie
rs17483466	G	Non-Hodgkin lymphoma (CLL)	*ACOXL*	2q13	111513929	Di Bernardo 2008 NatGenet	A genome-wide association study identifi
rs1859962	G	Prostate cancer	*Intergenic*	17q24.3	66620348	Eeles 2008 Nat Genet	Multiple newly identified loci associated
rs1876206	G	Breast cancer	*FBN1*	15q21.1	46687878	Murabito 2007 BMC Med Genet	A genome-wide association study of brea
rs189897	A	Nasopharyngeal carcinoma	*ITGA9*	3p22.2	37493549	Ng 2009 J Hum Genet	A genome-wide association study identifi
rs1926203	T	Lung cancer	*ACTA2*	10q23.31	90717314	Broderick 2009 Cancer Res	Deciphering the Impact of Common Gene
rs1926657	T	Breast cancer	*ABCC4*	13q32.1	94672957	Murabito 2007 BMC Med Genet	A genome-wide association study of brea

SNP	sk alle	Original cancer association	Mapped Gene	Region	Position	Author Year Journal	Manuscript title
rs1978503	G	Breast cancer	*Intergenic*	18q21.2	51815280	Murabito 2007 BMC Med Genet	A genome-wide association study of breast and pros
rs2046210	A	Breast cancer	*Intergenic*	6q25.1	151990059	Zheng 2009 Nat Genet	Genome-wide association study identifies a new br
rs2075555	T	Breast cancer	*COL1A1*	17q21.33	45629290	Murabito 2007 BMC Med Genet	A genome-wide association study of breast and pros
rs2089222	A	Leukemia (ALL)	*MAP1LC3B2*	12q24.22	115487041	Trevino 2009 Nat Genet	Germline genomic variants associated with childhoc
rs210138	G	Testicular germ cell tumor	*BAK1*	6p21.31	33650516	Rapley 2009 Nat Genet	A genome-wide association study of testicular germ
rs2151280	C	Basal cell carcinoma	*CDNK2BAS1*	9p21.3	22024719	Stacey 2009 Nat Genet	New common variants affecting susceptibility to bas
rs2167364	C	Leukemia (ALL)	*DDC*	7p12.1	50533321	Trevino 2009 Nat Genet	Germline genomic variants associated with childhoc
rs2180341	G	Breast cancer	*RNF146*	6q22.33	127642323	Gold 2008 Proc Natl Acad Sci	Genome-wide association study provides evidence
rs2191566	G	Leukemia (ALL)	*ZNF230*	19q13.31	49203229	Trevino 2009 Nat Genet	Germline genomic variants associated with childhoc
rs2239633	G	Leukemia (ALL)	*Intergenic*	14q11.2	22658897	Papaemmanuil 2009 Nat Genet	Loci on 7p12.2, 10q21.2 and 14q11.2 are associated w
rs2242041	G	Leukemia (ALL)	*DDC*	7p12.1	50496943	Trevino 2009 Nat Genet	Germline genomic variants associated with childhoc
rs2284063	G	Melanoma	*PLA2G6*	22q13.1	36874244	Bishop 2009 Nat Genet	Genome-wide association study identifies three loc
rs2294008	T	Bladder cancer	*PSCA*	8q24.3	143758933	Rothman 2010 Nat Genet	A multi-stage genome-wide association study of bla
rs2456449	G	Non-Hodgkin lymphoma (CLL)	*Intergenic*	8q24.21	128262163	Crowther-Swanepoel 2010 Nat Genet	Common variants at 2q37.3, 8q24.21, 15q21.3 and 16
rs258322	A	Melanoma	*CDK10*	16q24.3	88283404	Bishop 2009 Nat Genet	Genome-wide association study identifies three loc
rs2660753	T	Prostate cancer	*Intergenic*	3p12.1	87193364	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs266849	A	Prostate cancer	*Intergenic*	19q13.33	56040902	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs2710647	C	Prostate cancer	*EHBP1*	2p15	63067474	Eeles 2009 Nat Genet	Identification of seven new prostate cancer suscept
rs2735839	G	Prostate cancer	*Intergenic*	19q13.33	56056435	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs2736100	G	Glioma	*TERT*	5p15.33	1339516	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs2808630	C	Lung cancer	*Intergenic*	1q23.2	157947492	Amos 2008 Nat Genet	Genome-wide association scan of tag SNPs identifie
rs2853676	A	Glioma	*TERT*	5p15.33	1341547	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs2928679	A	Prostate cancer	*Intergenic*	8p21.2	23494920	Thomas 2008 Nat Genet	Multiple loci identified in a genome-wide associatic
rs2981578	G	Breast cancer	*FGFR2*	10q26.13	123330301	Turnbull 2010 Nat Genet	Genome-wide association study identifies five new
rs2981579	T	Breast cancer	*FGFR2*	10q26.13	123327325	Thomas 2009 Nat Genet	A multistage genome-wide association study in brea
rs2981582	T	Breast cancer	*FGFR2*	10q26.13	123342307	Easton 2007 Nature	Genome-wide association study identifies novel bre
rs305061	T	Non-Hodgkin lymphoma (CLL)	*Intergenic*	16q24.1	84533160	Crowther-Swanepoel 2010 Nat Genet	Common variants at 2q37.3, 8q24.21, 15q21.3 and 16
rs3117582	C	Lung cancer	*Intergenic*	6p21.33	31728499	Landi 2009 Am J Hum Genet	A Genome-wide Association Study of Lung Cancer Id
rs3131379	T	Lung cancer	*MSH5*	6p21.33	31829012	Harley 2008 Nat Genet	Genome-wide association scan in women with syste
rs31489	C	Lung cancer	*CLPTM1L*	5p15.33	1395714	Landi 2009 Am J Hum Genet	A Genome-wide Association Study of Lung Cancer Id
rs3750817	C	Breast cancer	*FGFR2*	10q26.13	123322567	Prentice 2009 Cancer Epi Biomarkers Pr	Variation in the *FGFR2* Gene and the Effects of Post
rs3790844	T	Pancreatic cancer	*NR5A2*	1q32.1	198274055	Petersen 2010 Nat Genet	A genome-wide association study identifies pancrea
rs3802842	C	Colorectal cancer	*C11orf93*	11q23.1	110676919	Tenesa 2008 Nat Genet	Genome-wide association scan identifies a colorecta
rs3803662	T	Breast cancer	*Intergenic*	16q12.1	51143842	Thomas 2009 Nat Genet	A multistage genome-wide association study in brea
rs3814113	T	Ovarian cancer	*Intergenic*	9p22.2	16905021	Song 2009 Nat Genet	A genome-wide association study identifies a new c
rs3817198	C	Breast cancer	*LSP1*	11p15.5	1865582	Easton 2007 Nature	Genome-wide association study identifies novel bre
rs401681	C	Lung cancer	*CLPTM1L*	5p15.33	1375087	Wang 2008 Nat Genet	Common 5p15.33 and 6p21.33 variants influence lun
rs402710	C	Lung cancer	*CLPTM1L*	5p15.33	1373722	McKay 2008 Nat Genet	Lung cancer susceptibility locus at 5p15.33
rs4132601	C	Leukemia (ALL)	*IKZF1*	7p12.2	50438098	Papaemmanuil 2009 Nat Genet	Loci on 7p12.2, 10q21.2 and 14q11.2 are associated w
rs4242382	A	Prostate cancer	*Intergenic*	8q24.21	128586755	Thomas 2008 Nat Genet	Multiple loci identified in a genome-wide associatic
rs4254535	C	Lung cancer	*Intergenic*	2p13.3	69051892	Broderick 2009 Cancer Res	Deciphering the Impact of Common Genetic Variatic
rs4295627	G	Glioma	*Intergenic*	8q24.21	130754639	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs4324715	T	Testicular cancer	*Intergenic*	5q31.3	141649691	Kanetsky 2009 Nat Genet	Common variation in KITLG and at 5q31.3 predispose
rs4324798	A	Lung cancer	*Intergenic*	6p22.1	28884096	Landi 2009 Am J Hum Genet	A Genome-wide Association Study of Lung Cancer Id
rs4415084	T	Breast cancer	*Intergenic*	5p12	44698272	Stacey 2007 Nat Genet	Common variants on chromosomes 2q35 and 16q12
rs4430796	G	Prostate cancer	*HNF1B*	17q12	33172153	Gudmundsson 2009 Nat Genet	Genome-wide association and replication studies id
rs4444235	C	Colorectal cancer	*Intergenic*	14q22.2	53480669	Houlston 2008 Nat Genet	Meta-analysis of genome-wide association data ide
rs445114	T	Prostate cancer	*Intergenic*	8q24.21	128392363	Gudmundsson 2009 Nat Genet	Genome-wide association and replication studies id
rs4474514	A	Testicular cancer	*KITLG*	12q21.32	87478090	Kanetsky 2009 Nat Genet	Common variation in KITLG and at 5q31.3 predispose
rs458685	C	Breast cancer	*GRIK1*	21q21.3	30099382	Murabito 2007 BMC Med Genet	A genome-wide association study of breast and pros
rs4624820	A	Testicular germ cell tumor	*Intergenic*	5q31.3	141661972	Rapley 2009 Nat Genet	A genome-wide association study of testicular germ
rs4636294	A	Melanoma	*Intergenic*	9p21.3	21737803	Falchi 2009 Nat Genet	Genome-wide association study identifies variants a
rs4657482	A	Testicular germ cell tumor	*UCK2*	1q24.1	164098273	Rapley 2009 Nat Genet	A genome-wide association study of testicular germ
rs4699052	C	Testicular germ cell tumor	*Intergenic*	4q24	104357239	Rapley 2009 Nat Genet	A genome-wide association study of testicular germ
rs4779584	T	Colorectal cancer	*Intergenic*	15q13.3	30782048	Tomlinson 2008 Nat Genet	A genome-wide association study identifies colorec
rs4782780	T	Prostate cancer	*CDH13*	16q23.3	81960548	Eeles 2009 Nat Genet	Identification of seven new prostate cancer suscept
rs4785763	A	Melanoma	*AFG3L1P*	16q24.3	88594437	Bishop 2009 Nat Genet	Genome-wide association study identifies three loc
rs4809324	C	Glioma (high-grade)	*RTEL1*	20q13.33	61788664	Wrensch 2009 Nat Genet	Variants in the CDKN2B and RTEL1 regions are associ
rs4857841	A	Prostate cancer	*EEFSEC*	3q21.3	129529333	Gudmundsson 2009 Nat Genet	Genome-wide association and replication studies id
rs4939827	T	Colorectal cancer	*SMAD7*	18q21.1	44707461	Tenesa 2008 Nat Genet	Genome-wide association scan identifies a colorecta
rs4961199	A	Prostate cancer	*Intergenic*	8q21.3	87650060	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs4962416	C	Prostate cancer	*CTBP2*	10q26.13	126686862	Thomas 2008 Nat Genet	Multiple loci identified in a genome-wide associatic
rs4973768	T	Breast cancer	*SLC4A7*	3p24.1	27391017	Ahmed 2009 Nat Genet	Newly discovered breast cancer susceptibility loci o
rs4975616	A	Lung cancer	*Intergenic*	5p15.33	1368660	Broderick 2009 Cancer Res	Deciphering the Impact of Common Genetic Variatic
rs4977756	G	Glioma	*CDKN2BAS1*	9p21.3	22058652	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs498872	T	Glioma	*PHLDB1*	11q23.3	117982577	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs505922	C	Pancreatic cancer	*ABO*	9q34.2	135139050	Amundadottir 2009 Nat Genet	Genome-wide association study identifies variants i
rs5759167	G	Prostate cancer	*Intergenic*	22q13.2	41830156	Eeles 2009 Nat Genet	Identification of seven new prostate cancer suscept
rs5945572	A	Prostate cancer	*Intergenic*	Xp11.22	51246423	Gudmundsson 2008 Nat Genet	Common sequence variants on 2p15 and Xp11.22 co

SNP	sk alle	Original cancer association	Mapped Gene	Region	Position	Author Year Journal	Manuscript title
rs5945619	C	Prostate cancer	*Intergenic*	Xp11.22	51258412	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs6001749	G	Prostate cancer	*TNRC6B*	22q13.1	38805272	Sun 2009 Cancer Res	Sequence variants at 22q13 are associated with pros
rs6010620	G	Glioma	*RTEL1*	20q13.33	61780283	Shete 2009 Nat Genet	Genome-wide association study identifies five susc
rs620861	G	Prostate cancer	*Intergenic*	8q24.21	128404855	Eeles 2009 Nat Genet	Identification of seven new prostate cancer suscept
rs630014	C	Pancreatic cancer	*ABO*	9q34.2	135139543	Amundadottir 2009 Nat Genet	Genome-wide association study identifies variants i
rs6435862	G	Neuroblastoma (high-risk)	*BARD1*	2q35	215380791	Capasso 2009 Nat Genet	Common variants in BARD1 influence susceptibility
rs6457327	C	NHL (Follicular lymphoma)	*Intergenic*	6p21.33	31182009	Skibola 2009 Nat Genet	Genome-wide association study of follicular lympho
rs6465657	C	Prostate cancer	*LMTK2*	7q21.3	97654263	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs6504950	G	Breast cancer	*STXBP4*	17q22	50411470	Ahmed 2009 Nat Genet	Newly discovered breast cancer susceptibility loci o
rs6556756	G	Breast cancer	*Intergenic*	5q34	163821858	Murabito 2007 BMC Med Genet	A genome-wide association study of breast and pros
rs671	A	Esophageal cancer	*ALDH2*	12q24.12	110726149	Cui 2009 Gastroenterology	Functional variants in ADH1B and ALDH2 coupled wit
rs6939340	G	Neuroblastoma (high-risk)	*FLJ22536*	6p22.3	22247983	Maris 2008 N Engl J Med	Chromosome 6p22 Locus Associated with Clinically A
rs6983267	G	Colorectal cancer	*Intergenic*	8q24.21	128482487	Tomlinson 2008 Nat Genet	A genome-wide association study identifies colorec
rs6983561	C	Prostate cancer	*Intergenic*	8q24.21	128176062	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with prostate cance
rs7000448	T	Prostate cancer	*Intergenic*	8q24.21	128510352	Al Olama 2009 Nat Genet	Multiple loci on 8q24 associated with prostate cance
rs7014346	A	Colorectal cancer	*Intergenic*	8q24.21	128493974	Tenesa 2008 Nat Genet	Genome-wide association scan identifies a colorecta
rs7023329	A	Melanoma	*MTAP*	9p21.3	21806528	Bishop 2009 Nat Genet	Genome-wide association study identifies three loc
rs7089424	C	Leukemia (ALL)	*ARID5B*	10q21.2	63422165	Papaemmanuil 2009 Nat Genet	Loci on 7p12.2, 10q21.2 and 14q11.2 are associated w
rs710521	A	Urinary bladder cancer	*Intergenic*	3q28	191128627	Rothman 2010 Nat Genet	A multi-stage genome-wide association study of bla
rs7117034	T	Prostate cancer	*Intergenic*	11q13.3	68731718	Gudmundsson 2009 Nat Genet	Genome-wide association and replication studies id
rs7127900	A	Prostate cancer	*Intergenic*	11p15.5	2190150	Thomas 2008 Nat Genet	Multiple loci identified in a genome-wide associatio
rs7176508	A	Non-Hodgkin lymphoma (CLL)	*Intergenic*	15q23	67806044	Di Bernardo 2008 NatGenet	A genome-wide association study identifies six susc
rs719725	A	Colorectal cancer	*Intergenic*	9p24	6355683	Zanke 2007 Nat Genet	Genome-wide association scan identifies a colorecta
rs721048	A	Prostate cancer	*EHBP1*	2p15	62985235	Gudmundsson 2008 Nat Genet	Common sequence variants on 2p15 and Xp11.22 cor
rs735665	A	Non-Hodgkin lymphoma (CLL)	*Intergenic*	11q24.1	122866607	Conde 2010 Nat Genet	Genome-wide association study of follicular lympho
rs748404	T	Lung cancer	*Intergenic*	15q15.2	41346523	Broderick 2009 Cancer Res	Deciphering the Impact of Common Genetic Variatio
rs7501939	C	Prostate cancer	*HNF1B*	17q12	33175269	Dugan 2007 J Natl Cancer Inst	Two genome-wide association studies of aggressive
rs7538876	A	Basal cell carcinoma	*PADI6*	1p36.13	17594950	Stacey 2008 Nat Genet	Common variants on 1p36 and 1q42 are associated w
rs757978	A	Non-Hodgkin lymphoma (CLL)	*FARP2*	2q37.3	242019774	Crowther-Swanepoel 2010 Nat Genet	Common variants at 2q37.3, 8q24.21, 15q21.3 and 16
rs7626795	G	Lung cancer	*IL1RAP*	3q28	191833155	Amos 2008 Nat Genet	Genome-wide association scan of tag SNPs identifie
rs7679673	C	Prostate cancer	*Intergenic*	4q24	106280983	Eeles 2009 Nat Genet	Identification of seven new prostate cancer suscept
rs7809758	G	Leukemia (ALL)	*DDC*	7p12.1	50540827	Papaemmanuil 2009 Nat Genet	Loci on 7p12.2, 10q21.2 and 14q11.2 are associated w
rs7837688	T	Prostate cancer	*Intergenic*	8q24.21	128608542	Takata 2010 Nat Genet	Genome-wide association study identifies five new
rs7841060	G	Prostate cancer	*Intergenic*	8q24.21	128165659	Yeager 2009 Nat Genet	Identification of a new prostate cancer susceptibilit
rs7931342	G	Prostate cancer	*Intergenic*	11q13.3	68751073	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs801114	G	Basal cell carcinoma	*Intergenic*	1q42.13	227064458	Stacey 2008 Nat Genet	Common variants on 1p36 and 1q42 are associated w
rs8034191	C	Lung cancer	*AGPHD1*	15q25.1	76593078	Broderick 2009 Cancer Res	Deciphering the Impact of Common Genetic Variatio
rs8042374	G	Lung cancer	*CHRNA3*	15q25.1	76695087	Wang 2008 Nat Genet	Common 5p15.33 and 6p21.33 variants influence lun
rs8102476	C	Prostate cancer	*Intergenic*	19q13.2	43427453	Gudmundsson 2009 Nat Genet	Genome-wide association and replication studies id
rs872071	G	Non-Hodgkin lymphoma (CLL)	*IRF4*	6p25.3	356064	Di Bernardo 2008 NatGenet	A genome-wide association study identifies six susc
rs889312	C	Breast cancer	*Intergenic*	5q11.2	56067641	Easton 2007 Nature	Genome-wide association study identifies novel bre
rs910873	A	Melanoma	*PIGU*	20q11.22	32635433	Brown 2008 Nat Genet	Common sequence variants on 20q11.22 confer mela
rs9295740	A	Lung cancer	*Intergenic*	6p22.1	27797481	Wang 2008 Nat Genet	Common 5p15.33 and 6p21.33 variants influence lun
rs931794	G	Lung cancer	*AGPHD1*	15q25.1	76613235	Amos 2008 Nat Genet	Genome-wide association scan of tag SNPs identifie
rs9364554	T	Prostate cancer	*SLC22A3*	6q25.3	160753654	Eeles 2008 Nat Genet	Multiple newly identified loci associated with prost
rs944289	T	Thyroid cancer	*Intergenic*	14q13.3	35718997	Gudmundsson 2009 Nat Genet	Common variants on 9q22.33 and 14q13.3 predispos
rs9543325	C	Pancreatic cancer	*Intergenic*	13q22.1	72814629	Petersen 2010 Nat Genet	A genome-wide association study identifies pancrea
rs961253	A	Colorectal cancer	*Intergenic*	20p12.3	6352281	Houlston 2008 Nat Genet	Meta-analysis of genome-wide association data ide
rs9623117	C	Prostate cancer	*Intergenic*	22q13.1	38782065	Sun 2009 Cancer Res	Sequence variants at 22q13 are associated with pros
rs9642880	T	Urinary bladder cancer	*Intergenic*	8q24.21	128787250	Rothman 2010 Nat Genet	A multi-stage genome-wide association study of bla
rs965513	A	Thyroid cancer	*Intergenic*	9q22.33	99595930	Gudmundsson 2009 Nat Genet	Common variants on 9q22.33 and 14q13.3 predispos
rs981782	T	Breast cancer	*HCN1*	5p12	45321475	Easton 2007 Nature	Genome-wide association study identifies novel bre
rs9929218	G	Colorectal cancer	*CDH1*	16q22.1	67378447	Houlston 2008 Nat Genet	Meta-analysis of genome-wide association data ide
rs995030	G	Testicular germ cell tumor	*KITLG*	12q21.32	87414802	Rapley 2009 Nat Genet	A genome-wide association study of testicular germ
rs999737	C	Breast cancer	*RAD51B*	14q24.1	68104435	Thomas 2009 Nat Genet	A multistage genome-wide association study in brea

Analysis II – Melanoma Appendix 3

Supplemental Table 2 - Results for the association of each SNP and melanoma.

SNP	Coded allele	Previous trait association	Beta	SE	P-value	# studies	n	OR	95% CI	P-heterogeneity
rs10086908	T	Prostate cancer	-0.03	0.04	0.44	4	16840	0.97	(0.88 - 1.05)	0.13
rs10090154	A	Prostate cancer	-0.04	0.09	0.62	2	10605	0.96	(0.81 - 1.13)	0.39
rs1016343	T	Prostate cancer	0.12	0.07	0.07	2	7188	1.13	(0.98 - 1.29)	0.19
rs10220831	T	Non-Hodgkin lymphoma (CLL)	0.04	0.06	0.55	3	7987	1.04	(0.92 - 1.17)	0.34
rs10263639	C	Breast cancer	0.01	0.06	0.84	3	11392	1.01	(0.89 - 1.15)	0.28
rs1036935	T	Non-Hodgkin lymphoma (CLL)	0.01	0.06	0.91	3	11399	1.01	(0.90 - 1.12)	0.09
rs10411210	C	Colorectal cancer	0.13	0.07	0.05	3	15971	1.14	(1.00 - 1.3)	0.68
rs1045485	G	Breast cancer	-0.10	0.06	0.08	5	21018	0.91	(0.81 - 1.01)	0.01
rs10464870	C	Glioma	0.01	0.07	0.86	2	7208	1.01	(0.88 - 1.16)	0.24
rs10483813	T	Breast cancer	-0.01	0.07	0.86	3	8057	0.99	(0.86 - 1.13)	0.05
rs10486567	G	Prostate cancer	0.07	0.04	0.10	5	22121	1.08	(0.98 - 1.17)	0.03
rs10490113	C	Breast cancer	0.02	0.06	0.73	4	20122	1.02	(0.91 - 1.14)	0.99
rs10505477	A	Colorectal cancer	0.03	0.05	0.55	4	16838	1.03	(0.93 - 1.12)	0.83
rs1051730	A	Lung cancer	-0.01	0.04	0.80	5	22127	0.99	(0.92 - 1.06)	0.64
rs10778826	A	Prostate cancer	0.03	0.03	0.39	5	22134	1.03	(0.96 - 1.1)	0.73
rs10795668	G	Colorectal cancer	0.02	0.04	0.53	5	22021	1.02	(0.95 - 1.1)	0.27
rs10821936	C	Leukemia (ALL)	0.04	0.05	0.44	4	13333	1.04	(0.94 - 1.13)	0.88
rs10896449	G	Prostate cancer	0.05	0.03	0.12	5	22092	1.05	(0.98 - 1.13)	0.41
rs10941679	G	Breast cancer	-0.08	0.06	0.24	2	7207	0.93	(0.81 - 1.05)	0.62
rs10974944	G	Leukemia (Myeloid)	0.04	0.05	0.38	4	13341	1.04	(0.94 - 1.15)	0.22
rs10993994	T	Prostate cancer	-0.01	0.04	0.84	3	16015	0.99	(0.91 - 1.07)	0.02
rs10994982	A	Leukemia (ALL)	0.05	0.04	0.23	4	13307	1.05	(0.96 - 1.15)	0.41
rs11083846	A	Non-Hodgkin lymphoma (CLL)	-0.03	0.07	0.63	2	7214	0.97	(0.85 - 1.1)	0.81
rs11155133	G	Leukemia (ALL)	-0.31	0.29	0.29	2	7219	0.73	(0.41 - 1.3)	0.62
rs11170164	A	Basal cell carcinoma	-0.10	0.23	0.66	1	1925	0.91	(0.58 - 1.41)	.
rs11228565	A	Prostate cancer	0.06	0.05	0.22	2	14026	1.06	(0.96 - 1.17)	0.00
rs11249433	C	Breast cancer	-0.01	0.04	0.88	3	15612	0.99	(0.91 - 1.08)	0.53
rs11649338	C	Breast cancer	-0.01	0.04	0.88	4	20188	0.99	(0.92 - 1.07)	0.16
rs11649743	G	Prostate cancer	0.05	0.04	0.23	5	22085	1.05	(0.96 - 1.15)	0.41
rs11668878	T	Non-Hodgkin lymphoma (CLL)	-0.12	0.12	0.35	1	5275	0.89	(0.69 - 1.14)	.
rs11861609	C	Prostate cancer	-0.05	0.04	0.25	4	16772	0.96	(0.88 - 1.03)	0.65
rs12155172	A	Prostate cancer	0.17	0.06	3.38E-03	2	10703	1.19	(1.05 - 1.33)	0.61
rs1219648	G	Breast cancer	0.02	0.04	0.59	4	13328	1.02	(0.93 - 1.12)	0.04
rs1229984	C	Esophageal cancer	0.00	0.12	0.97	2	7214	1.00	(0.79 - 1.27)	0.06
rs12418451	A	Prostate cancer	0.10	0.04	4.39E-03	5	22053	1.11	(1.03 - 1.19)	0.24
rs12500426	A	Prostate cancer	-0.06	0.03	0.09	5	22077	0.94	(0.88 - 1.01)	0.75
rs12543663	C	Prostate cancer	0.08	0.05	0.15	2	10609	1.08	(0.97 - 1.2)	0.55
rs12621278	A	Prostate cancer	0.10	0.07	0.17	5	22098	1.11	(0.95 - 1.28)	0.86
rs13252298	A	Prostate cancer	0.04	0.06	0.50	3	8054	1.04	(0.92 - 1.18)	0.73
rs13254738	C	Prostate cancer	0.04	0.04	0.32	3	15917	1.04	(0.96 - 1.13)	0.04
rs13281615	T	Breast cancer	0.10	0.03	3.57E-03	5	22138	1.11	(1.03 - 1.18)	0.15
rs13387042	A	Breast cancer	-0.02	0.03	0.54	5	22109	0.98	(0.91 - 1.05)	0.72
rs13397985	G	Non-Hodgkin lymphoma (CLL)	-0.07	0.05	0.18	4	13331	0.93	(0.83 - 1.04)	0.02
rs1393350	A	Melanoma	0.23	0.04	5.21E-10	5	22009	1.25	(1.16 - 1.35)	0.99
rs1412829	C	Glioma (high-grade)	0.01	0.04	0.82	4	13305	1.01	(0.92 - 1.1)	0.52
rs1447295	A	Prostate cancer	-0.02	0.07	0.79	4	13324	0.98	(0.85 - 1.13)	0.59
rs1465618	T	Prostate cancer	-0.01	0.04	0.79	5	22136	0.99	(0.91 - 1.07)	0.64
rs1512268	T	Prostate cancer	0.00	0.03	0.92	5	22062	1.00	(0.93 - 1.07)	0.46
rs1530057	A	Lung cancer	-0.05	0.08	0.55	4	20196	0.95	(0.81 - 1.11)	0.24
rs1571801	T	Prostate cancer	0.08	0.05	0.11	4	16718	1.08	(0.98 - 1.19)	0.21
rs157935	T	Basal cell carcinoma	0.05	0.05	0.29	4	13331	1.05	(0.95 - 1.15)	0.91
rs167020	A	Pancreatic cancer	0.10	0.06	0.10	2	7217	1.11	(0.98 - 1.24)	0.76
rs16886165	G	Breast cancer	-0.03	0.06	0.59	4	13339	0.97	(0.86 - 1.09)	0.27
rs16891982	C	Melanoma	1.13	0.15	7.39E-14	3	15949	3.11	(2.30 - 4.18)	0.43

SNP	Coded allele	Previous trait association	Beta	SE	P-value	# studies	n	OR	95% CI	P-heterogeneity
rs16892766	C	Colorectal cancer	0.11	0.06	0.08	5	22078	1.11	(0.98 - 1.25)	0.34
rs16901979	A	Prostate cancer	-0.19	0.13	0.13	3	14852	0.82	(0.64 - 1.06)	0.10
rs16902094	G	Prostate cancer	-0.09	0.06	0.10	3	16008	0.91	(0.81 - 1.02)	0.38
rs17021918	C	Prostate cancer	-0.02	0.04	0.51	5	22134	0.98	(0.91 - 1.05)	0.87
rs172310	A	Pancreatic cancer	0.08	0.06	0.20	2	7136	1.08	(0.95 - 1.21)	0.89
rs17483466	G	Non-Hodgkin lymphoma (CLL)	-0.02	0.05	0.70	4	13333	0.98	(0.89 - 1.08)	0.16
rs1859962	G	Prostate cancer	0.05	0.04	0.29	4	13338	1.05	(0.96 - 1.14)	0.46
rs1876206	G	Breast cancer	-0.09	0.05	0.08	4	20127	0.91	(0.82 - 1.01)	0.35
rs189897	A	Nasopharyngeal carcinoma	0.04	0.07	0.59	2	7217	1.04	(0.90 - 1.2)	0.81
rs1926203	T	Lung cancer	0.02	0.04	0.57	2	13891	1.03	(0.94 - 1.12)	0.70
rs1926657	T	Breast cancer	-0.04	0.05	0.39	4	20077	0.96	(0.87 - 1.05)	0.40
rs1978503	G	Breast cancer	-0.05	0.05	0.25	4	20185	0.95	(0.86 - 1.04)	0.19
rs2046210	A	Breast cancer	0.02	0.04	0.50	5	22033	1.02	(0.95 - 1.1)	0.40
rs2075555	T	Breast cancer	-0.10	0.08	0.24	1	8734	0.91	(0.76 - 1.07)	.
rs2089222	A	Leukemia (ALL)	-0.12	0.16	0.46	2	7220	0.89	(0.65 - 1.21)	0.05
rs210138	G	Testicular germ cell tumor	-0.07	0.06	0.22	4	13308	0.93	(0.83 - 1.04)	0.33
rs2151280	C	Basal cell carcinoma	-0.01	0.04	0.83	4	13301	0.99	(0.91 - 1.08)	0.18
rs2167364	C	Leukemia (ALL)	-0.04	0.05	0.38	4	13334	0.96	(0.87 - 1.05)	0.26
rs2180341	G	Breast cancer	0.06	0.05	0.24	3	11396	1.06	(0.95 - 1.18)	0.78
rs2191566	G	Leukemia (ALL)	0.00	0.06	0.95	2	7187	1.00	(0.88 - 1.12)	0.93
rs2239633	G	Leukemia (ALL)	-0.02	0.05	0.73	2	7207	0.98	(0.88 - 1.09)	0.45
rs2242041	G	Leukemia (ALL)	-0.08	0.08	0.31	4	13332	0.92	(0.79 - 1.08)	0.71
rs2284063	G	Melanoma	0.08	0.04	0.02	5	22087	1.09	(1.01 - 1.17)	0.29
rs2294008	T	Bladder cancer	-0.07	0.06	0.23	2	7203	0.94	(0.83 - 1.04)	0.90
rs2456449	G	Non-Hodgkin lymphoma (CLL)	0.00	0.05	0.93	3	11390	1.00	(0.91 - 1.1)	0.63
rs258322	A	Melanoma	0.44	0.05	8.81E-19	5	22082	1.55	(1.40 - 1.71)	0.60
rs2660753	T	Prostate cancer	0.05	0.05	0.34	5	22119	1.05	(0.94 - 1.16)	0.03
rs266849	A	Prostate cancer	0.02	0.05	0.64	3	15895	1.02	(0.92 - 1.13)	0.57
rs2710647	C	Prostate cancer	-0.05	0.03	0.18	5	22018	0.96	(0.89 - 1.02)	0.28
rs2735839	G	Prostate cancer	0.01	0.05	0.82	3	16007	1.01	(0.91 - 1.13)	0.89
rs2736100	G	Glioma	-0.01	0.03	0.73	5	21811	0.99	(0.92 - 1.06)	0.70
rs2808630	C	Lung cancer	0.06	0.05	0.19	4	13335	1.06	(0.96 - 1.17)	0.93
rs2853676	A	Glioma	0.05	0.06	0.39	2	7172	1.05	(0.93 - 1.19)	0.85
rs2928679	A	Prostate cancer	0.02	0.03	0.58	5	22083	1.02	(0.95 - 1.09)	0.25
rs2981578	G	Breast cancer	-0.01	0.04	0.84	3	15982	0.99	(0.91 - 1.07)	0.05
rs2981579	T	Breast cancer	0.04	0.05	0.46	4	13341	1.04	(0.94 - 1.15)	0.30
rs2981582	T	Breast cancer	-0.02	0.03	0.65	5	22114	0.98	(0.92 - 1.05)	0.17
rs305061	T	Non-Hodgkin lymphoma (CLL)	0.01	0.05	0.79	3	11396	1.01	(0.92 - 1.12)	0.51
rs3117582	C	Lung cancer	0.13	0.07	0.08	3	11399	1.14	(0.98 - 1.31)	0.46
rs3131379	T	Lung cancer	0.02	0.05	0.74	5	22050	1.02	(0.91 - 1.13)	0.07
rs31489	C	Lung cancer	-0.12	0.04	4.18E-03	2	14004	0.89	(0.81 - 0.96)	0.33
rs3750817	C	Breast cancer	0.00	0.04	0.96	2	14077	1.00	(0.92 - 1.09)	0.66
rs3790844	T	Pancreatic cancer	-0.03	0.05	0.59	3	11400	0.97	(0.87 - 1.08)	0.01
rs3802842	C	Colorectal cancer	0.05	0.04	0.21	5	22122	1.05	(0.97 - 1.13)	0.67
rs3803662	T	Breast cancer	-0.01	0.04	0.70	5	22141	0.99	(0.91 - 1.06)	0.96
rs3814113	T	Ovarian cancer	0.01	0.04	0.87	3	15981	1.01	(0.92 - 1.09)	0.22
rs3817198	C	Breast cancer	0.07	0.04	0.07	5	22109	1.07	(0.99 - 1.15)	0.79
rs401681	C	Lung cancer	-0.14	0.03	3.26E-05	5	22109	0.87	(0.81 - 0.93)	0.65
rs402710	C	Lung cancer	-0.14	0.04	7.74E-04	3	15991	0.87	(0.80 - 0.94)	0.04
rs4132601	C	Leukemia (ALL)	-0.01	0.05	0.87	4	13337	0.99	(0.90 - 1.09)	0.26
rs4242382	A	Prostate cancer	-0.02	0.07	0.74	4	13298	0.98	(0.84 - 1.12)	0.84
rs4254535	C	Lung cancer	-0.01	0.05	0.78	2	14076	0.99	(0.90 - 1.08)	0.11
rs4295627	G	Glioma	-0.02	0.07	0.75	2	7218	0.98	(0.84 - 1.13)	0.72
rs4324715	T	Testicular cancer	-0.03	0.04	0.51	4	13290	0.97	(0.89 - 1.06)	0.12
rs4324798	A	Lung cancer	0.04	0.06	0.51	5	21797	1.04	(0.92 - 1.17)	0.06
rs4415084	T	Breast cancer	0.00	0.04	1.00	2	13928	1.00	(0.92 - 1.09)	0.74
rs4430796	G	Prostate cancer	0.04	0.04	0.30	3	15981	1.04	(0.96 - 1.12)	0.22

SNP	Coded allele	Previous trait association	Beta	SE	P-value	# studies	n	OR	95% CI	P-heterogeneity
rs4444235	C	Colorectal cancer	0.03	0.04	0.41	3	16010	1.03	(0.95 - 1.11)	0.59
rs445114	T	Prostate cancer	0.04	0.06	0.48	2	7197	1.04	(0.93 - 1.17)	0.14
rs4474514	A	Testicular cancer	0.05	0.05	0.36	4	13327	1.05	(0.94 - 1.17)	0.56
rs458685	C	Breast cancer	0.05	0.06	0.42	3	14920	1.05	(0.93 - 1.17)	0.45
rs4624820	A	Testicular germ cell tumor	-0.07	0.04	0.12	4	13324	0.94	(0.85 - 1.02)	0.25
rs4636294	A	Melanoma	0.17	0.03	5.04E-07	5	22053	1.19	(1.10 - 1.27)	0.12
rs4657482	A	Testicular germ cell tumor	0.00	0.06	0.94	2	7208	1.00	(0.88 - 1.12)	0.86
rs4699052	C	Testicular germ cell tumor	-0.03	0.06	0.55	2	7200	0.97	(0.86 - 1.08)	0.89
rs4779584	T	Colorectal cancer	-0.05	0.04	0.22	5	22130	0.95	(0.87 - 1.03)	0.31
rs4782780	T	Prostate cancer	-0.01	0.04	0.83	3	14923	0.99	(0.90 - 1.08)	0.21
rs4785763	A	Melanoma	0.27	0.04	1.29E-14	5	21993	1.31	(1.22 - 1.4)	0.80
rs4809324	C	Glioma (high-grade)	0.05	0.10	0.62	1	5276	1.05	(0.86 - 1.28)	.
rs4857841	A	Prostate cancer	-0.03	0.04	0.43	5	22093	0.97	(0.90 - 1.05)	0.58
rs4939827	T	Colorectal cancer	-0.03	0.03	0.33	5	22105	0.97	(0.90 - 1.03)	0.24
rs4961199	A	Prostate cancer	-0.01	0.05	0.89	5	22124	0.99	(0.90 - 1.09)	0.48
rs4962416	C	Prostate cancer	0.00	0.04	0.97	5	22125	1.00	(0.93 - 1.08)	0.13
rs4973768	T	Breast cancer	-0.05	0.03	0.17	5	22119	0.95	(0.89 - 1.02)	0.31
rs4975616	A	Lung cancer	-0.14	0.03	2.04E-05	5	22135	0.87	(0.81 - 0.93)	0.65
rs4977756	G	Glioma	-0.03	0.04	0.48	4	13343	0.97	(0.89 - 1.05)	0.60
rs498872	T	Glioma	0.05	0.05	0.30	4	13323	1.06	(0.95 - 1.17)	0.47
rs505922	C	Pancreatic cancer	-0.11	0.05	0.02	4	13339	0.90	(0.81 - 0.98)	0.25
rs5759167	G	Prostate cancer	-0.01	0.05	0.80	2	10681	0.99	(0.89 - 1.09)	0.004
rs5945572	A	Prostate cancer	0.00	0.03	0.96	3	16014	1.00	(0.93 - 1.07)	0.53
rs5945619	C	Prostate cancer	0.00	0.03	0.90	3	15941	1.00	(0.93 - 1.08)	0.55
rs6001749	G	Prostate cancer	0.04	0.07	0.58	3	8030	1.04	(0.90 - 1.19)	0.17
rs6010620	G	Glioma	0.01	0.06	0.91	2	7187	1.01	(0.88 - 1.14)	0.57
rs620861	G	Prostate cancer	0.09	0.05	0.07	2	10739	1.10	(0.99 - 1.21)	0.36
rs630014	C	Pancreatic cancer	-0.01	0.04	0.82	4	13324	0.99	(0.90 - 1.08)	0.45
rs6435862	G	Neuroblastoma (high-risk)	-0.01	0.05	0.80	4	13254	0.99	(0.89 - 1.09)	0.63
rs6457327	C	NHL (Follicular lymphoma)	-0.03	0.04	0.45	4	13337	0.97	(0.88 - 1.05)	0.32
rs6465657	C	Prostate cancer	0.00	0.03	0.97	5	22127	1.00	(0.93 - 1.07)	0.42
rs6504950	G	Breast cancer	-0.04	0.04	0.30	5	22104	0.96	(0.89 - 1.04)	0.64
rs6556756	G	Breast cancer	-0.09	0.06	0.12	4	20197	0.91	(0.81 - 1.03)	0.94
rs671	A	Esophageal cancer	-12.59	600.10	0.98	1	7221	0.00	.	.
rs6939340	G	Neuroblastoma (high-risk)	0.03	0.06	0.65	2	7177	1.03	(0.91 - 1.14)	0.24
rs6983267	G	Colorectal cancer	0.05	0.03	0.08	5	22097	1.05	(0.99 - 1.11)	0.52
rs6983561	C	Prostate cancer	-0.16	0.16	0.31	2	7216	0.85	(0.62 - 1.16)	0.52
rs7000448	T	Prostate cancer	0.05	0.04	0.20	4	16833	1.05	(0.97 - 1.14)	0.07
rs7014346	A	Colorectal cancer	0.03	0.03	0.44	5	22097	1.03	(0.95 - 1.1)	0.53
rs7023329	A	Melanoma	0.16	0.03	2.03E-06	5	22114	1.17	(1.09 - 1.25)	0.28
rs7089424	C	Leukemia (ALL)	0.05	0.06	0.39	2	7212	1.05	(0.93 - 1.18)	0.71
rs710521	A	Urinary bladder cancer	0.13	0.06	0.03	4	13302	1.13	(1.01 - 1.27)	0.55
rs7117034	T	Prostate cancer	0.21	0.06	3.67E-04	2	10675	1.23	(1.09 - 1.37)	0.55
rs7127900	A	Prostate cancer	-0.03	0.07	0.67	2	7193	0.97	(0.84 - 1.11)	0.76
rs7176508	A	Non-Hodgkin lymphoma (CLL)	0.09	0.05	0.04	3	11395	1.10	(1.00 - 1.21)	0.47
rs719725	A	Colorectal cancer	0.04	0.04	0.24	5	22041	1.04	(0.97 - 1.12)	0.36
rs721048	A	Prostate cancer	0.08	0.07	0.28	2	7201	1.08	(0.93 - 1.24)	0.08
rs735665	A	Non-Hodgkin lymphoma (CLL)	0.00	0.05	0.99	4	13323	1.00	(0.89 - 1.11)	0.29
rs748404	T	Lung cancer	0.03	0.04	0.54	4	20099	1.03	(0.94 - 1.12)	0.38
rs7501939	C	Prostate cancer	0.00	0.04	0.92	3	15985	1.00	(0.92 - 1.09)	0.96
rs7538876	A	Basal cell carcinoma	0.04	0.04	0.39	4	13327	1.04	(0.95 - 1.13)	0.85
rs757978	A	Non-Hodgkin lymphoma (CLL)	-0.06	0.08	0.42	3	11399	0.94	(0.80 - 1.09)	0.07
rs7626795	G	Lung cancer	-0.06	0.05	0.24	5	22071	0.94	(0.84 - 1.04)	0.08
rs7679673	C	Prostate cancer	-0.04	0.04	0.32	3	15908	0.96	(0.88 - 1.04)	0.42
rs7809758	G	Leukemia (ALL)	-0.02	0.04	0.63	4	13337	0.98	(0.89 - 1.07)	0.22
rs7837688	T	Prostate cancer	-0.07	0.07	0.32	3	16018	0.94	(0.81 - 1.07)	0.55

SNP	Coded allele	Previous trait association	Beta	SE	P-value	# studies	n	OR	95% CI	P-heterogeneity
rs7841060	G	Prostate cancer	0.01	0.05	0.78	3	15970	1.01	(0.92 - 1.11)	0.03
rs7931342	G	Prostate cancer	0.03	0.04	0.47	4	13336	1.03	(0.94 - 1.12)	0.49
rs801114	G	Basal cell carcinoma	0.00	0.05	0.91	4	13340	1.00	(0.91 - 1.09)	0.82
rs8034191	C	Lung cancer	-0.03	0.04	0.49	4	13335	0.97	(0.88 - 1.06)	0.09
rs8042374	G	Lung cancer	-0.04	0.04	0.34	5	22032	0.96	(0.88 - 1.04)	0.07
rs8102476	C	Prostate cancer	0.01	0.05	0.83	2	10743	1.01	(0.91 - 1.11)	0.18
rs872071	G	Non-Hodgkin lymphoma (CLL)	0.08	0.06	0.17	2	7210	1.08	(0.96 - 1.2)	0.95
rs889312	C	Breast cancer	0.04	0.04	0.25	5	22098	1.04	(0.97 - 1.12)	0.20
rs910873	A	Melanoma	0.27	0.06	2.46E-05	3	15937	1.31	(1.15 - 1.48)	1.00
rs9295740	A	Lung cancer	-0.08	0.04	0.07	5	22134	0.92	(0.84 - 1.01)	0.21
rs931794	G	Lung cancer	-0.04	0.04	0.28	5	22041	0.96	(0.89 - 1.03)	0.17
rs9364554	T	Prostate cancer	0.02	0.04	0.60	5	22117	1.02	(0.94 - 1.1)	0.09
rs944289	T	Thyroid cancer	0.01	0.05	0.83	4	13259	1.01	(0.92 - 1.11)	0.25
rs9543325	C	Pancreatic cancer	-0.09	0.05	0.07	3	11397	0.92	(0.83 - 1.01)	0.87
rs961253	A	Colorectal cancer	0.03	0.04	0.45	5	22119	1.03	(0.95 - 1.1)	0.74
rs9623117	C	Prostate cancer	0.01	0.04	0.80	4	20166	1.01	(0.93 - 1.1)	0.91
rs9642880	T	Urinary bladder cancer	-0.03	0.05	0.51	3	11396	0.97	(0.88 - 1.06)	0.99
rs965513	A	Thyroid cancer	-0.01	0.05	0.77	4	13330	0.99	(0.90 - 1.08)	0.51
rs981782	T	Breast cancer	-0.04	0.04	0.36	3	16004	0.97	(0.89 - 1.04)	0.43
rs9929218	G	Colorectal cancer	-0.05	0.04	0.20	5	22140	0.95	(0.88 - 1.02)	0.75
rs995030	G	Testicular germ cell tumor	0.06	0.05	0.25	4	13327	1.06	(0.95 - 1.18)	0.52
rs999737	C	Breast cancer	-0.03	0.04	0.40	4	20098	0.97	(0.89 - 1.05)	0.57

Analysis II – Melanoma Appendix 4

Supplemental Table 5 - Results for sex-stratifed analyses of each SNP and melanoma.

		Male						Female						Between sex
SNP	Previous trait association	OR	95% CI	P-value	Studi	n	P-het	OR	95% CI	P-value	tudi	n	P-het	P-het
rs10086908	Prostate cancer	0.88	(0.77 - 1.01)	0.06	2	5806	0.55	0.98	(0.82 - 1.16)	0.80	2	4912	0.22	0.30
rs10090154	Prostate cancer	0.93	(0.74 - 1.17)	0.54	2	5707	0.21	1.00	(0.77 - 1.28)	1.00	2	4898	0.89	0.68
rs1016343	Prostate cancer	1.37	(0.98 - 1.91)	0.06	1	1019	.	1.08	(0.93 - 1.25)	0.28	2	6169	0.68	0.19
rs10220831	Non-Hodgkin lymphoma (CLL)	1.13	(0.82 - 1.56)	0.44	1	975	.	0.71	(0.48 - 1.05)	0.09	1	890	.	0.07
rs10263639	Breast cancer	.	.	.	0		.	1.08	(0.90 - 1.28)	0.38	1	5270	.	.
rs1036935	Non-Hodgkin lymphoma (CLL)	.	.	.	0		.	1.10	(0.94 - 1.27)	0.22	1	5277	.	.
rs10411210	Colorectal cancer	1.03	(0.83 - 1.28)	0.75	2	5797	0.58	1.20	(1.02 - 1.42)	0.03	3	10174	0.84	0.27
rs1045485	Breast cancer	0.79	(0.66 - 0.94)	0.01	2	5814	0.20	0.96	(0.79 - 1.17)	0.70	3	9082	0.01	0.14
rs10464870	Glioma	1.09	(0.76 - 1.53)	0.64	1	1016	.	1.00	(0.86 - 1.16)	1.00	2	6192	0.22	0.68
rs10483813	Breast cancer	0.88	(0.62 - 1.23)	0.45	1	1023	.	0.69	(0.47 - 1.02)	0.06	1	912	.	0.36
rs10486567	Prostate cancer	1.13	(0.96 - 1.31)	0.13	2	5800	0.26	1.03	(0.91 - 1.16)	0.62	3	10199	0.04	0.37
rs10490113	Breast cancer	1.00	(0.80 - 1.25)	1.00	1	4747	.	1.02	(0.87 - 1.2)	0.77	2	9253	0.77	0.88
rs10505477	Colorectal cancer	1.05	(0.92 - 1.2)	0.41	2	5804	0.91	0.97	(0.83 - 1.13)	0.68	2	4912	0.55	0.42
rs1051730	Lung cancer	1.01	(0.88 - 1.15)	0.87	2	5816	0.63	1.01	(0.91 - 1.12)	0.77	3	10189	0.75	1.00
rs10778826	Prostate cancer	1.13	(0.99 - 1.28)	0.07	2	5817	0.44	1.01	(0.91 - 1.11)	0.87	3	10195	0.53	0.16
rs10795668	Colorectal cancer	1.00	(0.87 - 1.15)	1.00	2	5743	0.05	1.01	(0.90 - 1.12)	0.89	3	10156	0.83	0.91
rs10821936	Leukemia (ALL)	1.04	(0.77 - 1.4)	0.77	1	1025	.	1.05	(0.93 - 1.19)	0.40	2	6186	0.70	0.95
rs10896449	Prostate cancer	1.12	(0.99 - 1.27)	0.07	2	5793	0.97	1.04	(0.94 - 1.14)	0.46	3	10177	0.22	0.31
rs10941679	Breast cancer	0.86	(0.62 - 1.19)	0.38	1	1026	.	0.94	(0.82 - 1.08)	0.39	2	6181	0.91	0.61
rs10974944	Leukemia (Myeloid)	1.09	(0.80 - 1.49)	0.57	1	1027	.	0.99	(0.86 - 1.13)	0.89	2	6192	0.12	0.57
rs10993994	Prostate cancer	0.89	(0.77 - 1.01)	0.07	2	5817	0.001	1.06	(0.96 - 1.17)	0.24	3	10198	0.97	0.04
rs10994982	Leukemia (ALL)	0.91	(0.68 - 1.21)	0.54	1	1006	.	1.09	(0.96 - 1.23)	0.15	2	6179	0.51	0.24
rs11083846	Non-Hodgkin lymphoma (CLL)	0.99	(0.71 - 1.36)	0.93	1	1025	.	0.96	(0.83 - 1.11)	0.61	2	6189	0.88	0.87
rs11155133	Leukemia (ALL)	0.58	(0.13 - 2.41)	0.45	1	1024	.	0.77	(0.41 - 1.44)	0.42	2	6195	0.75	0.73
rs11170164	Basal cell carcinoma	0.82	(0.45 - 1.48)	0.51	1	1017	.	1.06	(0.54 - 2.05)	0.87	1	908	.	0.58
rs11228565	Prostate cancer	1.21	(1.02 - 1.42)	0.02	1	4756	.	0.99	(0.87 - 1.12)	0.84	2	9270	0.00	0.05
rs11249433	Breast cancer	1.00	(0.87 - 1.13)	0.95	2	5818	0.74	0.99	(0.89 - 1.1)	0.90	3	9794	0.21	0.91
rs11649338	Breast cancer	1.05	(0.90 - 1.21)	0.55	1	4787	.	0.91	(0.81 - 1.01)	0.07	2	9279	0.79	0.10
rs11649743	Prostate cancer	1.26	(1.05 - 1.49)	0.01	2	5779	0.98	1.02	(0.90 - 1.15)	0.76	3	10184	0.92	0.05

			Male							Female						Between sex
SNP	Previous trait association		OR	95% CI	P-value	Studi	n	P-het		OR	95% CI	P-value	tudi	n	P-het	P-het
rs11668878	Non-Hodgkin lymphoma (CLL)		.	.	.	0		.		0.89	(0.69 - 1.14)	0.35	1	5275	.	.
rs11861609	Prostate cancer		0.96	(0.83 - 1.09)	0.48	2	5773	0.45		0.98	(0.84 - 1.14)	0.80	2	4877	0.55	0.78
rs12155172	Prostate cancer		1.26	(1.09 - 1.46)	1.66E-03	2	5806	0.66		1.08	(0.89 - 1.29)	0.43	2	4897	0.81	0.16
rs1219648	Breast cancer		1.50	(1.12 - 1.99)	0.01	1	1016	.		1.00	(0.88 - 1.13)	0.97	2	6190	0.48	0.01
rs1229984	Esophageal cancer		2.50	(1.10 - 5.63)	0.03	1	1024	.		0.90	(0.70 - 1.15)	0.40	2	6190	0.91	0.02
rs12418451	Prostate cancer		1.29	(1.12 - 1.48)	1.92E-04	2	5785	0.52		1.01	(0.90 - 1.12)	0.92	3	10146	0.69	0.005
rs12500426	Prostate cancer		0.89	(0.77 - 1.01)	0.06	2	5815	0.38		1.00	(0.91 - 1.11)	0.95	3	10140	0.25	0.16
rs12543663	Prostate cancer		1.15	(1.00 - 1.31)	0.04	2	5707	0.69		0.98	(0.83 - 1.16)	0.83	2	4902	0.14	0.16
rs12621278	Prostate cancer		1.14	(0.86 - 1.5)	0.37	2	5805	0.45		1.09	(0.88 - 1.35)	0.41	3	10171	0.78	0.82
rs13252298	Prostate cancer		1.23	(0.88 - 1.7)	0.22	1	1019	.		0.89	(0.61 - 1.3)	0.56	1	913	.	0.21
rs13254738	Prostate cancer		1.06	(0.92 - 1.21)	0.38	2	5783	0.00		1.03	(0.93 - 1.14)	0.56	3	10134	0.18	0.73
rs13281615	Breast cancer		1.15	(1.00 - 1.3)	0.04	2	5818	0.22		1.08	(0.97 - 1.19)	0.12	3	10198	0.18	0.49
rs13387042	Breast cancer		1.00	(0.88 - 1.13)	0.98	2	5806	0.31		0.98	(0.89 - 1.08)	0.73	3	10181	0.62	0.80
rs13397985	Non-Hodgkin lymphoma (CLL)		1.06	(0.73 - 1.52)	0.75	1	1025	.		1.03	(0.88 - 1.19)	0.71	2	6184	0.04	0.88
rs1393350	Melanoma		1.25	(1.09 - 1.43)	1.15E-03	2	5767	0.53		1.24	(1.12 - 1.38)	4.04E-05	3	10120	0.80	1.00
rs1412829	Glioma (high-grade)		1.11	(0.83 - 1.47)	0.48	1	1004	.		0.96	(0.85 - 1.08)	0.48	2	6179	0.95	0.39
rs1447295	Prostate cancer		1.04	(0.64 - 1.66)	0.88	1	1022	.		0.91	(0.74 - 1.11)	0.33	2	6180	0.75	0.59
rs1465618	Prostate cancer		0.97	(0.82 - 1.13)	0.68	2	5816	0.69		1.05	(0.93 - 1.18)	0.39	3	10198	0.66	0.42
rs1512268	Prostate cancer		0.92	(0.81 - 1.05)	0.21	2	5788	0.19		1.08	(0.98 - 1.19)	0.12	3	10152	0.55	0.04
rs1530057	Lung cancer		1.15	(0.85 - 1.53)	0.36	1	4792	.		0.90	(0.71 - 1.13)	0.37	2	9282	0.34	0.21
rs1571801	Prostate cancer		1.05	(0.90 - 1.21)	0.55	2	5696	0.01		1.25	(1.05 - 1.47)	0.01	2	4900	0.18	0.09
rs157935	Basal cell carcinoma		1.02	(0.75 - 1.39)	0.88	1	1019	.		1.07	(0.94 - 1.22)	0.28	2	6190	0.49	0.77
rs167020	Pancreatic cancer		1.20	(0.88 - 1.63)	0.24	1	1028	.		1.09	(0.95 - 1.24)	0.19	2	6189	0.85	0.59
rs16886165	Breast cancer		0.59	(0.38 - 0.92)	0.02	1	1024	.		0.99	(0.84 - 1.17)	0.93	2	6193	0.66	0.03
rs16891982	Melanoma		5.50	(2.93 - 10.28)	9.53E-08	2	5789	0.34		2.37	(1.69 - 3.31)	4.67E-07	3	10160	0.45	0.02
rs16892766	Colorectal cancer		1.21	(0.96 - 1.51)	0.10	2	5793	0.27		0.98	(0.82 - 1.18)	0.84	3	10163	0.81	0.14
rs16901979	Prostate cancer		0.68	(0.42 - 1.07)	0.09	1	4766	.		0.66	(0.38 - 1.14)	0.14	1	3964	.	0.96
rs16902094	Prostate cancer		0.98	(0.81 - 1.17)	0.81	2	5813	0.40		0.87	(0.75 - 1.01)	0.06	3	10195	0.37	0.33
rs17021918	Prostate cancer		0.93	(0.81 - 1.06)	0.26	2	5816	0.76		1.01	(0.91 - 1.12)	0.82	3	10196	0.26	0.30
rs172310	Pancreatic cancer		1.09	(0.80 - 1.47)	0.57	1	1025	.		1.08	(0.94 - 1.22)	0.25	2	6111	0.78	0.90
rs17483466	Non-Hodgkin lymphoma (CLL)		0.92	(0.65 - 1.3)	0.65	1	1020	.		1.06	(0.92 - 1.23)	0.39	2	6191	0.21	0.47
rs1859962	Prostate cancer		0.99	(0.74 - 1.31)	0.92	1	1027	.		1.11	(0.99 - 1.25)	0.07	2	6189	0.27	0.43
rs1876206	Breast cancer		0.86	(0.69 - 1.07)	0.17	1	4768	.		0.94	(0.80 - 1.09)	0.39	2	9237	0.17	0.51
rs189897	Nasopharyngeal carcinoma		1.15	(0.81 - 1.63)	0.44	1	1025	.		1.02	(0.87 - 1.19)	0.78	2	6192	0.31	0.54
rs1926203	Lung cancer		1.05	(0.90 - 1.22)	0.54	1	4702	.		1.01	(0.91 - 1.12)	0.82	2	9189	0.42	0.67
rs1926657	Breast cancer		0.93	(0.76 - 1.13)	0.45	1	4760	.		0.93	(0.81 - 1.07)	0.32	2	9195	0.23	0.93
rs1978503	Breast cancer		1.03	(0.85 - 1.23)	0.76	1	4783	.		0.94	(0.82 - 1.07)	0.34	2	9280	0.13	0.41
rs2046210	Breast cancer		0.95	(0.83 - 1.09)	0.47	2	5730	0.97		1.05	(0.94 - 1.16)	0.36	3	10181	0.36	0.25
rs2075555	Breast cancer		0.98	(0.79 - 1.22)	0.89	1	4757	.		0.80	(0.61 - 1.05)	0.11	1	3977	.	0.26
rs2089222	Leukemia (ALL)		0.49	(0.19 - 1.23)	0.13	1	1025	.		0.96	(0.69 - 1.34)	0.83	2	6195	0.18	0.18
rs210138	Testicular germ cell tumor		0.81	(0.55 - 1.17)	0.26	1	1009	.		0.93	(0.79 - 1.08)	0.35	2	6177	0.17	0.50
rs2151280	Basal cell carcinoma		0.85	(0.63 - 1.14)	0.29	1	1005	.		0.94	(0.83 - 1.06)	0.31	2	6174	0.28	0.54
rs2167364	Leukemia (ALL)		1.06	(0.78 - 1.43)	0.70	1	1024	.		0.92	(0.80 - 1.04)	0.17	2	6188	0.20	0.35
rs2180341	Breast cancer		.	.	.	0		.		1.05	(0.90 - 1.21)	0.52	1	5274	.	.
rs2191566	Leukemia (ALL)		0.87	(0.63 - 1.18)	0.37	1	1026	.		1.02	(0.89 - 1.16)	0.76	2	6161	0.28	0.36
rs2239633	Leukemia (ALL)		0.76	(0.57 - 1.01)	0.06	1	1025	.		1.02	(0.91 - 1.15)	0.70	2	6182	0.38	0.06
rs2242041	Leukemia (ALL)		1.14	(0.71 - 1.84)	0.58	1	1020	.		0.86	(0.69 - 1.06)	0.16	2	6190	0.91	0.29
rs2284063	Melanoma		1.07	(0.93 - 1.22)	0.34	2	5796	0.54		1.06	(0.95 - 1.17)	0.28	3	10169	0.19	1.00
rs2294008	Bladder cancer		0.75	(0.56 - 1)	0.05	1	1015	.		0.97	(0.86 - 1.09)	0.62	2	6188	0.13	0.11
rs2456449	Non-Hodgkin lymphoma (CLL)		.	.	.	0		.		1.03	(0.90 - 1.17)	0.69	1	5268	.	.
rs258322	Melanoma		1.53	(1.27 - 1.84)	4.55E-06	2	5787	0.15		1.61	(1.39 - 1.85)	2.82E-11	3	10173	0.44	0.66
rs2660753	Prostate cancer		1.31	(1.09 - 1.58)	4.00E-03	2	5804	0.45		0.98	(0.84 - 1.15)	0.84	3	10193	0.45	0.02
rs266849	Prostate cancer		0.95	(0.81 - 1.12)	0.56	2	5731	0.95		1.07	(0.94 - 1.21)	0.31	3	10164	0.13	0.27
rs2710647	Prostate cancer		0.92	(0.81 - 1.05)	0.21	2	5749	0.36		0.95	(0.86 - 1.04)	0.28	3	10147	0.33	0.70
rs2735839	Prostate cancer		1.00	(0.84 - 1.19)	0.97	2	5807	0.51		1.02	(0.89 - 1.16)	0.80	3	10200	0.90	0.86
rs2736100	Glioma		0.91	(0.80 - 1.03)	0.15	2	5692	0.71		1.00	(0.90 - 1.11)	0.96	3	9997	0.71	0.25
rs2808630	Lung cancer		1.03	(0.74 - 1.41)	0.87	1	1018	.		1.06	(0.93 - 1.21)	0.34	2	6195	0.70	0.86
rs2853676	Glioma		1.02	(0.74 - 1.4)	0.89	1	999	.		1.06	(0.93 - 1.21)	0.38	2	6173	0.94	0.82
rs2928679	Prostate cancer		1.09	(0.96 - 1.24)	0.17	2	5790	0.21		0.96	(0.86 - 1.05)	0.37	3	10171	0.77	0.10
rs2981578	Breast cancer		1.00	(0.88 - 1.13)	0.95	2	5808	0.02		0.99	(0.89 - 1.09)	0.84	3	10174	0.52	0.90
rs2981579	Breast cancer		1.34	(1.00 - 1.79)	0.04	1	1027	.		1.00	(0.89 - 1.13)	0.94	2	6192	0.80	0.06
rs2981582	Breast cancer		1.03	(0.90 - 1.17)	0.67	2	5804	0.03		0.97	(0.87 - 1.07)	0.52	3	10188	0.68	0.49
rs305061	Non-Hodgkin lymphoma (CLL)		.	.	.	0		.		0.98	(0.86 - 1.12)	0.80	1	5274	.	.
rs3117582	Lung cancer		.	.	.	0		.		1.08	(0.88 - 1.31)	0.45	1	5277	.	.
rs3131379	Lung cancer		0.80	(0.64 - 1)	0.05	2	5769	0.19		1.05	(0.89 - 1.22)	0.54	3	10159	0.85	0.05
rs31489	Lung cancer		0.91	(0.78 - 1.05)	0.19	1	4764	.		0.88	(0.79 - 0.97)	0.01	2	9240	0.11	0.73
rs3750817	Breast cancer		0.98	(0.84 - 1.13)	0.73	1	4794	.		1.02	(0.91 - 1.13)	0.77	2	9283	0.83	0.56
rs3790844	Pancreatic cancer		.	.	.	0		.		1.13	(0.96 - 1.31)	0.12	1	5278	.	.
rs3802842	Colorectal cancer		0.99	(0.86 - 1.14)	0.90	2	5813	0.79		1.07	(0.96 - 1.19)	0.19	3	10187	0.65	0.35
rs3803662	Breast cancer		1.05	(0.91 - 1.2)	0.53	2	5815	0.85		0.96	(0.86 - 1.07)	0.49	3	10204	0.84	0.39
rs3814113	Ovarian cancer		0.93	(0.81 - 1.06)	0.26	2	5807	0.61		1.06	(0.95 - 1.17)	0.29	3	10174	0.08	0.10

																		Between
				Male						Female								sex
SNP	Previous trait association			OR	95% CI	P-value	Studi	n	P-het	OR	95% CI	P-value	tudi	n	P-het			P-het
rs3817198	Breast cancer			0.92	(0.80 - 1.05)	0.20	2	5795	0.69	1.14	(1.02 - 1.26)	0.01	3	10192	0.21			0.01
rs401681	Lung cancer			0.88	(0.77 - 1)	0.05	2	5800	0.73	0.88	(0.79 - 0.96)	0.01	3	10187	0.35			1.00
rs402710	Lung cancer			0.83	(0.73 - 0.95)	0.01	2	5802	0.56	0.90	(0.81 - 0.99)	0.04	3	10189	0.05			0.42
rs4132601	Leukemia (ALL)			0.98	(0.71 - 1.35)	0.89	1	1026	.	0.93	(0.81 - 1.06)	0.28	2	6189	0.45			0.77
rs4242382	Prostate cancer			1.13	(0.71 - 1.78)	0.60	1	998	.	0.95	(0.77 - 1.16)	0.61	2	6178	0.88			0.50
rs4254535	Lung cancer			0.93	(0.78 - 1.08)	0.34	1	4793	.	1.02	(0.90 - 1.14)	0.77	2	9283	0.19			0.32
rs4295627	Glioma			1.04	(0.71 - 1.52)	0.82	1	1024	.	0.96	(0.82 - 1.13)	0.66	2	6194	0.92			0.70
rs4324715	Testicular cancer			0.95	(0.72 - 1.25)	0.74	1	1000	.	0.92	(0.81 - 1.03)	0.16	2	6168	0.07			0.84
rs4324798	Lung cancer			0.85	(0.67 - 1.09)	0.20	2	5681	0.66	1.01	(0.85 - 1.2)	0.89	3	9994	0.22			0.26
rs4415084	Breast cancer			1.04	(0.90 - 1.2)	0.56	1	4692	.	0.98	(0.88 - 1.08)	0.68	2	9236	0.89			0.49
rs4430796	Prostate cancer			0.97	(0.85 - 1.1)	0.59	2	5795	0.51	1.09	(0.98 - 1.2)	0.08	3	10186	0.64			0.12
rs4444235	Colorectal cancer			1.05	(0.92 - 1.19)	0.47	2	5813	0.58	1.03	(0.93 - 1.13)	0.61	3	10197	0.16			0.70
rs445114	Prostate cancer			1.12	(0.83 - 1.5)	0.45	1	1012	.	1.03	(0.90 - 1.16)	0.67	2	6185	0.13			0.62
rs4474514	Testicular cancer			0.98	(0.69 - 1.38)	0.91	1	1020	.	1.01	(0.87 - 1.16)	0.91	2	6185	1.00			0.87
rs458685	Breast cancer			1.07	(0.89 - 1.29)	0.46	1	4792	.	1.10	(0.87 - 1.37)	0.42	1	4006	.			0.89
rs4624820	Testicular germ cell tumor			0.83	(0.62 - 1.09)	0.18	1	1019	.	0.90	(0.79 - 1.01)	0.07	2	6183	0.10			0.60
rs4636294	Melanoma			1.25	(1.10 - 1.42)	4.45E-04	2	5781	0.53	1.23	(1.11 - 1.35)	3.20E-05	3	10150	0.71			0.80
rs4657482	Testicular germ cell tumor			1.11	(0.82 - 1.48)	0.50	1	1026	.	0.98	(0.86 - 1.11)	0.73	2	6182	0.60			0.46
rs4699052	Testicular germ cell tumor			1.06	(0.79 - 1.42)	0.68	1	1021	.	0.95	(0.84 - 1.07)	0.42	2	6179	0.61			0.50
rs4779584	Colorectal cancer			1.03	(0.87 - 1.2)	0.75	2	5812	0.25	0.92	(0.80 - 1.04)	0.17	3	10196	0.23			0.23
rs4782780	Prostate cancer			0.92	(0.79 - 1.06)	0.23	1	4794	.	1.00	(0.84 - 1.18)	0.98	1	4007	.			0.43
rs4785763	Melanoma			1.29	(1.13 - 1.47)	1.23E-04	2	5732	0.66	1.29	(1.16 - 1.42)	5.59E-07	3	10139	0.43			0.91
rs4809324	Glioma (high-grade)			.	.	.	0		.	1.05	(0.86 - 1.28)	0.62	1	5276	.			.
rs4857841	Prostate cancer			0.94	(0.81 - 1.08)	0.36	2	5804	0.92	1.04	(0.93 - 1.16)	0.45	3	10167	0.76			0.23
rs4939827	Colorectal cancer			0.90	(0.78 - 1.02)	0.09	2	5793	0.76	0.96	(0.87 - 1.06)	0.46	3	10190	0.52			0.37
rs4961199	Prostate cancer			0.89	(0.74 - 1.06)	0.19	2	5807	0.38	1.02	(0.89 - 1.16)	0.75	3	10195	0.91			0.22
rs4962416	Prostate cancer			0.93	(0.81 - 1.08)	0.35	2	5804	0.81	1.04	(0.93 - 1.15)	0.52	3	10199	0.10			0.25
rs4973768	Breast cancer			0.90	(0.79 - 1.02)	0.09	2	5806	0.31	0.98	(0.89 - 1.08)	0.71	3	10191	0.43			0.25
rs4975616	Lung cancer			0.88	(0.77 - 0.99)	0.04	2	5819	0.72	0.88	(0.79 - 0.97)	0.01	3	10194	0.41			1.00
rs4977756	Glioma			0.94	(0.70 - 1.26)	0.69	1	1027	.	0.93	(0.82 - 1.05)	0.25	2	6194	0.95			0.95
rs498872	Glioma			1.19	(0.88 - 1.59)	0.25	1	1019	.	1.00	(0.88 - 1.14)	1.00	2	6182	0.11			0.29
rs505922	Pancreatic cancer			0.72	(0.52 - 0.97)	0.03	1	1027	.	0.94	(0.83 - 1.06)	0.34	2	6190	0.53			0.11
rs5759167	Prostate cancer			0.95	(0.83 - 1.08)	0.46	2	5787	0.08	1.04	(0.88 - 1.21)	0.66	2	4894	0.02			0.45
rs5945572	Prostate cancer			1.00	(0.90 - 1.09)	0.92	2	5820	0.28	1.00	(0.90 - 1.11)	0.98	3	10194	0.91			1.00
rs5945619	Prostate cancer			1.00	(0.91 - 1.1)	0.98	2	5785	0.28	1.01	(0.91 - 1.12)	0.87	3	10156	0.94			0.89
rs6001749	Prostate cancer			1.11	(0.79 - 1.56)	0.55	1	1008	.	1.38	(0.91 - 2.08)	0.12	1	900	.			0.42
rs6010620	Glioma			0.99	(0.72 - 1.36)	0.96	1	1009	.	1.01	(0.87 - 1.16)	0.92	2	6178	0.40			0.91
rs620861	Prostate cancer			1.11	(0.97 - 1.26)	0.12	2	5817	0.84	1.08	(0.92 - 1.26)	0.35	2	4922	0.25			0.85
rs630014	Pancreatic cancer			0.86	(0.65 - 1.13)	0.28	1	1012	.	1.03	(0.91 - 1.16)	0.58	2	6190	0.65			0.24
rs6435862	Neuroblastoma (high-risk)			1.12	(0.82 - 1.53)	0.47	1	952	.	1.01	(0.88 - 1.15)	0.92	2	6180	0.55			0.57
rs6457327	NHL (Follicular lymphoma)			0.87	(0.65 - 1.17)	0.36	1	1024	.	1.03	(0.90 - 1.16)	0.68	2	6191	0.54			0.32
rs6465657	Prostate cancer			1.07	(0.94 - 1.21)	0.29	2	5813	0.62	0.92	(0.83 - 1.02)	0.10	3	10192	0.89			0.05
rs6504950	Breast cancer			1.05	(0.91 - 1.21)	0.50	2	5792	0.65	0.91	(0.81 - 1.01)	0.06	3	10190	0.92			0.08
rs6556756	Breast cancer			0.94	(0.75 - 1.18)	0.61	1	4791	.	0.90	(0.76 - 1.05)	0.17	2	9284	0.90			0.71
rs671	Esophageal cancer			0.00	.	0.99	1	1029	.	0.00	.	0.99	1	6192	.			1.00
rs6939340	Neuroblastoma (high-risk)			0.76	(0.57 - 1.02)	0.07	1	1000	.	1.08	(0.95 - 1.21)	0.22	2	6177	0.48			0.04
rs6983267	Colorectal cancer			1.07	(0.94 - 1.22)	0.28	2	5792	0.42	1.07	(0.96 - 1.17)	0.18	3	10183	0.47			1.00
rs6983561	Prostate cancer			0.82	(0.37 - 1.81)	0.62	1	1026	.	0.86	(0.61 - 1.2)	0.37	2	6190	0.35			0.91
rs7000448	Prostate cancer			1.06	(0.93 - 1.21)	0.36	2	5803	0.13	1.21	(1.03 - 1.41)	0.02	2	4908	0.94			0.22
rs7014346	Colorectal cancer			1.02	(0.89 - 1.16)	0.78	2	5791	0.88	1.02	(0.92 - 1.13)	0.63	3	10184	0.16			1.00
rs7023329	Melanoma			1.28	(1.12 - 1.45)	0.00	2	5806	0.63	1.18	(1.07 - 1.3)	5.63E-04	3	10186	0.87			0.37
rs7089424	Leukemia (ALL)			1.10	(0.82 - 1.47)	0.51	1	1024	.	1.04	(0.92 - 1.18)	0.51	2	6188	0.88			0.71
rs710521	Urinary bladder cancer			1.09	(0.78 - 1.51)	0.61	1	1016	.	1.12	(0.97 - 1.28)	0.10	2	6164	0.17			0.91
rs7117034	Prostate cancer			1.26	(1.08 - 1.45)	2.07E-03	2	5775	0.20	1.18	(0.98 - 1.42)	0.07	2	4900	0.44			0.60
rs7127900	Prostate cancer			0.96	(0.66 - 1.39)	0.83	1	1011	.	0.97	(0.83 - 1.13)	0.69	2	6182	0.64			0.96
rs7176508	Non-Hodgkin lymphoma (CLL)			.	.	.	0		.	1.13	(1.00 - 1.29)	0.05	1	5273	.			.
rs719725	Colorectal cancer			0.91	(0.79 - 1.04)	0.15	2	5753	0.13	1.10	(0.99 - 1.22)	0.06	3	10166	0.79			0.03
rs721048	Prostate cancer			1.31	(0.92 - 1.86)	0.13	1	1026	.	1.04	(0.89 - 1.21)	0.62	2	6175	0.22			0.24
rs735665	Non-Hodgkin lymphoma (CLL)			1.09	(0.76 - 1.55)	0.64	1	1016	.	0.99	(0.85 - 1.15)	0.91	2	6185	0.08			0.65
rs748404	Lung cancer			1.02	(0.85 - 1.21)	0.83	1	4748	.	0.98	(0.86 - 1.1)	0.69	2	9229	0.65			0.71
rs7501939	Prostate cancer			1.00	(0.88 - 1.14)	0.98	2	5802	0.99	1.00	(0.91 - 1.11)	0.93	3	10183	0.92			1.00
rs7538876	Basal cell carcinoma			1.11	(0.81 - 1.49)	0.51	1	1022	.	1.06	(0.93 - 1.19)	0.37	2	6183	0.84			0.76
rs757978	Non-Hodgkin lymphoma (CLL)			.	.	.	0		.	1.06	(0.86 - 1.3)	0.56	1	5277	.			.
rs7626795	Lung cancer			0.81	(0.66 - 0.99)	0.04	2	5793	0.24	0.98	(0.85 - 1.14)	0.83	3	10156	0.14			0.12
rs7679673	Prostate cancer			0.99	(0.87 - 1.13)	0.89	2	5765	0.21	0.94	(0.85 - 1.04)	0.25	3	10143	0.93			0.56
rs7809758	Leukemia (ALL)			0.99	(0.73 - 1.33)	0.94	1	1025	.	0.94	(0.83 - 1.06)	0.32	2	6190	0.06			0.76
rs7837688	Prostate cancer			0.84	(0.66 - 1.06)	0.15	2	5819	0.35	0.99	(0.84 - 1.16)	0.87	3	10199	0.84			0.27
rs7841060	Prostate cancer			0.92	(0.78 - 1.08)	0.31	2	5787	0.005	1.07	(0.94 - 1.2)	0.27	3	10183	0.87			0.13

																		Between
				Male							**Female**							sex
SNP	Previous trait association			OR	95% CI	P-value	Studi	n	P-het		OR	95% CI	P-value	tudi	n	P-het		P-het
rs7931342	Prostate cancer			1.09	(0.82 - 1.44)	0.55	1	1023	.		1.04	(0.92 - 1.17)	0.53	2	6191	0.18		0.74
rs801114	Basal cell carcinoma			0.94	(0.70 - 1.26)	0.68	1	1027	.		1.02	(0.90 - 1.16)	0.71	2	6191	0.94		0.62
rs8034191	Lung cancer			0.87	(0.64 - 1.17)	0.36	1	1023	.		1.08	(0.95 - 1.21)	0.24	2	6190	0.94		0.19
rs8042374	Lung cancer			0.88	(0.75 - 1.02)	0.10	2	5755	0.06		0.92	(0.82 - 1.04)	0.18	3	10155	0.56		0.62
rs8102476	Prostate cancer			1.05	(0.92 - 1.19)	0.46	2	5817	0.54		0.95	(0.81 - 1.11)	0.54	2	4926	0.23		0.32
rs872071	Non-Hodgkin lymphoma (CLL)			1.08	(0.81 - 1.43)	0.60	1	1026	.		1.08	(0.95 - 1.21)	0.22	2	6184	0.98		0.95
rs889312	Breast cancer			1.00	(0.87 - 1.15)	0.99	2	5786	0.03		1.03	(0.92 - 1.15)	0.57	3	10190	0.87		0.73
rs910873	Melanoma			1.49	(1.22 - 1.81)	7.58E-05	2	5761	0.91		1.21	(1.02 - 1.42)	0.02	3	10176	0.49		0.10
rs9295740	Lung cancer			0.87	(0.72 - 1.03)	0.10	2	5820	0.41		0.88	(0.77 - 1)	0.05	3	10192	0.92		0.93
rs931794	Lung cancer			0.93	(0.81 - 1.07)	0.30	2	5783	0.14		1.02	(0.92 - 1.13)	0.67	3	10136	0.69		0.30
rs9364554	Prostate cancer			1.01	(0.88 - 1.16)	0.87	2	5813	0.02		1.03	(0.92 - 1.15)	0.58	3	10182	0.51		0.82
rs944289	Thyroid cancer			0.96	(0.72 - 1.27)	0.75	1	1029	.		0.96	(0.85 - 1.08)	0.51	2	6108	0.74		0.95
rs9543325	Pancreatic cancer			.	.	.	0		.		0.92	(0.81 - 1.05)	0.23	1	5275	.		.
rs961253	Colorectal cancer			1.03	(0.90 - 1.18)	0.63	2	5806	0.58		1.05	(0.95 - 1.17)	0.30	3	10191	0.19		0.82
rs9623117	Prostate cancer			1.03	(0.86 - 1.21)	0.76	1	4763	.		1.02	(0.90 - 1.15)	0.74	2	9281	0.99		0.92
rs9642880	Urinary bladder cancer			.	.	.	0		.		0.97	(0.85 - 1.1)	0.64	1	5274	.		.
rs965513	Thyroid cancer			1.21	(0.90 - 1.61)	0.19	1	1022	.		0.93	(0.81 - 1.05)	0.25	2	6186	0.78		0.11
rs981782	Breast cancer			0.95	(0.83 - 1.08)	0.41	2	5804	0.45		0.97	(0.88 - 1.07)	0.60	3	10200	0.28		0.80
rs9929218	Colorectal cancer			0.95	(0.82 - 1.08)	0.43	2	5822	0.18		0.93	(0.83 - 1.03)	0.17	3	10196	0.57		0.82
rs995030	Testicular germ cell tumor			0.99	(0.69 - 1.41)	0.96	1	1018	.		1.02	(0.87 - 1.18)	0.82	2	6187	0.79		0.88
rs999737	Breast cancer			0.88	(0.75 - 1.04)	0.13	1	4758	.		1.03	(0.91 - 1.16)	0.62	2	9218	0.98		0.13