Genetic Insights into Aging and Age-Related Diseases among Varied Pedigree Structures

By

Jennifer L. Bragg-Gresham

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2013

Doctoral Committee:

Professor Sharon L. R. Kardia, Co-Chair
Professor Goncalo R. Abecasis, Co-Chair
Professor Carlos F. Mendes de Leon
Professor Peter X. Song

ACKNOWLEDGEMENTS

First, I owe Sharon Kardia and Goncalo Abecasis many thanks for their guidance over the last three years. I would like to thank Goncalo for giving me the opportunity to work as full-time staff while pursuing my degree. It allowed me to return to school, while still maintaining my family. More than that, it gave me the experience I needed in analyzing complicated genetic data sets, coordinating large consortia projects, and leading junior-level students.

I am also very thankful to Sharon for taking me on as a graduate student mentee. When I was recommended to Sharon as a new student, she was so busy with her research and other graduate students that I know she was skeptical about taking on another responsibility, so we set up a time to meet each other. I remember my relief after we chatted for about an hour and she concluded that I seemed like I'd be "pretty low-maintenance" and she decided to take me on. Sharon, I hope I have actually been pretty "low-maintenance" for you. Thank you also for helping me keep sight of epidemiologic principles in genetic analyses.

Next I would like to acknowledge my committee members, Carlos Mendes de Leon and Peter Song. Although we didn't meet often, your advice has been very useful. Carlos, thank you for helping me keep my analysis in-line with the community views on aging research. Peter, thank you for helping me explore the usefulness of including the longitudinal nature of the

SardiNIA data in terms of predicting biological age. A very special thank you, also, to Yan Zhou for writing R code to handle the SardiNIA pedigree structure within the longitudinal data.

There are two special mentors that deserve a lot of credit for helping me get to where I am today. After completed my BS in Biostatistics back in 1999, I decided to take a position at a local not-for-profit research organization focusing on improving clinical treatment and outcomes for hemodialysis patients. Philip Held, who founded the organization in 1996 asked me a question during my interview that I'm so happy I answered correctly. "Jennifer, are you prepared to work hard?" I did work hard and enjoyed it. The second mentor is Hal Morgenstern, who if a professor of Epidemiology here at UM. I had the pleasure of working with Hal before returning to complete my education and he really opened my eyes to better ways of thinking about study design and analysis. I also greatly appreciate his support in my decision to apply to the department and pursue my PhD in Epidemiology. Although Hal was not formally on my committee, I found him very happy to discuss and direct me in my analyses.

I cannot leave out a special thank you to my friends and my family who have put up with either not seeing me as often as we would have liked or putting up with one tired and often grumpy mommy. Thank you to my wonderful husband for taking on more responsibilities within the family and always being a shoulder to lean on or a quick wit to make me laugh.

Thank you also to co-workers at the Center for Statistical Genetics for listening to me present my aging work for the past two years. In particular, Serena Sanna, for giving me a crash-course in UNIX scripting, teaching me the more efficient ways to run genome-wide association studies, and always being willing to help me, even from Sardinia. Also, Xiaowei Zhan, who very patiently helped me get my MACH imputation running and completed during one of the busiest

computational times of the year. Without that data, this dissertation would not have been as interesting.

Finally, this dissertation would not have been possible without contributions from my collaborators. I am honored to have been part of the core group for the SardiNIA Study on Aging. Thank you to Dr. David Schlessinger at the National Institute on Aging, and Francesco Cucca at Consiglio Nazionale delle Ricerche in SardiNIA. I also am very thankful to the clinicians and staff members who worked so hard to collect and process the rich data included in the SardiNIA Study.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Genetic Insights into Aging and Age-Related Diseases among Varied Pedigree Structures

By

Jennifer L. Bragg-Gresham

Co-chairs: Sharon L. R. Kardia and Goncalo R. Abecasis

Older age is associated with decline in many physiological functions that eventually lead to increased susceptibility to diseases. The rate of aging varies among individuals and may be influenced by genes. This dissertation has three aims: 1) define a measure of aging using physiologic traits and determine its heritability among various pedigree structures, 2) investigate genetic associations with the age trait using genome-wide association study analyses, and 3) focus on decline in kidney function by examining its association with known kidney loci and clinical risk factors within the SardiNIA Study on Aging.

Using data from five genetic cohorts (30,000+ individuals) with varying pedigree structure (SardiNIA Study on Aging – large pedigrees, Framingham Heart Study offspring – small

pedigrees, Atherosclerotic Risk in Communities Study – unrelated, and two twins studies: Twingene and TwinsUK), two aging traits were developed. Both traits employ three physiologic health measures (kidney function, systolic blood pressure, and waist circumference) to estimate an individual's biologic age and contrast it with actual age. Linear mixed modeling was employed to estimate a predicted age, while Cox models were used to estimate a risk age equivalent to the age of the census population with the same mortality risk.

Using Merlin software (Abecasis, 2002), both age traits were found to be heritable in all cohorts with h^2 estimates of 0.25 to 0.68, depending on pedigree structure. Meta-analysis revealed a genome-wide significant association ($p < 5 \times 10^{-8}$) within the *LRP1B* gene on chromosome 2. *LRP1B* produces low density lipoprotein receptor-related protein 1B and has been previously associated with tumor suppression, metabolic traits, and successful aging.

Association analyses were also conducted using a genetic risk score for CKD, adjusting for clinical factors, with CKD prevalence and change in kidney function. One additional allele in the genetic risk score for CKD was significantly associated with CKD prevalence (OR=1.07, p=0.001), decline in eGFR (β=-0.23, p=0.004), and eGFR decline > 1 SD (OR=1.05, p=0.04).

These findings further our understanding of the genetics of aging and the CKD risk score results suggest a possible clinical utility. Putting the findings into clinical use should be evaluated.

CHAPTER I

Introduction

The human population, not just its individuals, is aging. It is projected that by the year 2016, for the first time in history, the proportion of individuals over the age of 65 years will surpass the proportion of children under the age of 5 years (Dobriansky, et.al., 2007). This population shift has happened for a few reasons. Life expectancy has increased. There has been a shift from high reproductive rates to low reproductive rates. There has been a shift from the predominance of infectious disease to non-communicable disease and chronic conditions as leading cause of death (Dobriansky, et.al., 2007).

In 2012 the 10 leading causes of death in the US were: heart disease (597,689 deaths), cancer (574,743 deaths), chronic lower respiratory diseases (138,080 deaths), stroke/cerebrovascular diseases (129,476 deaths), accidents/unintentional injuries (120,859 deaths, Alzheimer's disease (83,494 deaths), diabetes (69,071 deaths), nephritis/nephrotic syndrome/nephrosis (50,476 deaths), influenza and pneumonia (50,097 deaths), and intentional self-harm/suicide (38,364 deaths) (CDC, 2013). Seven out of these ten leading

causes of death are chronic conditions, typically found among the elderly, while influenza and pneumonia, although infectious diseases, are also much more likely to be deadly among the elderly. These US numbers are very reflective of the world-wide causes of death for middle and high income countries (World Health Organization, 2011). Overall, five of the ten leading causes of death world-wide are chronic conditions and include: ischaemic heart disease (7.25 million deaths, 12.8%), stroke and other cerebrovascular disease (6.15 million deaths, 10.8%), chronic obstructive pulmonary disease (3.28 million deaths, 5.8%), trachea, bronchus, lung cancers (1.39 million deaths, 2.4%), and diabetes mellitus (1.26 million deaths, 2.2%); leaving the remaining five causes to communicable diseases (lower respiratory infections, diarrheal diseases, HIV/AIDS, and tuberculosis) or accidents (road traffic accidents).

While the current rates of these chronic conditions are high, they are projected to get even worse. The American Heart Association (AHA) projects that in the next 20 years, more than 40% of the US population is expected to have some form of cardiovascular disease. This increase will triple the total direct medical costs of caring for hypertension, coronary heart disease, heart failure, stroke, and other forms of cardiovascular disease from the current $273 billion to more than $800 billion (Heidenreich PA, et.al, 2011). It is also predicted that the prevalence of cardiovascular disease will increase by approximately 10% over the next 20 years if there are no changes in prevention practices and treatment trends. The increase is likely to be even greater if some risk factors, such as diabetes and obesity, continue to increase rapidly.

The good news, though, is that heart disease is largely preventable and many believe the healthcare system needs to focus on prevention and early intervention (Heidenreich PA,

et.al, 2011). Recent work from the Coronary Artery Risk Development in Young Adults (CARDIA) study suggests cardiovascular disease prevention should begin earlier in life, as many risk-factor levels in individuals younger than 30 years old were predictive of subclinical atherosclerosis 15 years later. The data also suggests that modest improvements in risk factors earlier in life have a larger impact than more substantial reductions later in life (Lee DH, et.al, 2003).

A large contributor in many of the chronic conditions that are plaguing our society is obesity. Recent research has shown that obesity accelerates the aging of adipose tissue, a process only now beginning to come to light at the molecular level (Ahima, 2009). Evident of this has come from experiments conducted in mice which suggest that obesity increases the formation of reactive oxygen species in fat cells, shortens telomeres—and ultimately results in activation of the p53 tumor suppressor, inflammation and the promotion of insulin resistance.

Data from the National Health and Nutrition Examination Survey, 2009–2010 (Flegal, 2012 and Ogden, 2012) show that more than 2 in 3 adults are considered to be overweight or obese, more than 1 in 3 adults are considered to be obese, with more than 1 in 20 adults considered to have extreme obesity. About one-third of children and adolescents ages 6 to 19 are considered to be overweight or obese, and more than 1 in 6 children and adolescents ages 6 to 19 are considered to be obese.

The health risks of being overweight or obese include: type 2 diabetes, heart disease, high blood pressure, nonalcoholic fatty liver disease (excess fat and inflammation in the liver of people who drink little or no alcohol), osteoarthritis (a health problem causing pain, swelling, and stiffness in one or more joints), some types of cancer (breast, colon, endometrial, and

kidney), and stroke (Weight-control Information Network, 2012). Since the early 1960s, the prevalence of obesity among adults has more than doubled, increasing from 13.4 to 35.7 percent in U.S. adults age 20 and older (Flegal, 2012 and Ogden, 2010). Obesity prevalence remained mostly stable from 1999 to 2010, but has increased slightly, yet in a statistically significant way, among men overall, as well as among black women and Mexican American women. Among children and adolescents, the prevalence of obesity also increased in the 1980s and 1990s but is now mostly stable at about 17 percent (Ogden, 2012).

With rates of high blood pressure and diabetes on the rise, it is no surprise that kidney disease is also on the list of the leading causes of death in the US. In the US kidney disease affects more than 20 million people, or approximately 7% of the US population (USRD, 2011). Rates of chronic kidney disease and end stage renal disease are increasing as our population of elderly is increasing. From 1991-2004, the prevalence of CKD has increased from 10% - 13% and the number of patients treated with dialysis or transplantation has increased from 209K to 472K. In the period 1991-2001 ESRD incidence increased 43% (USRDS, 2011).

How can we reduce the burden of these chronic diseases on our society and improve the health of our elderly individuals? Just as the advances in public health over the past 100 years have enabled a shift from infectious to chronic disease, we must now take steps to counter the effects of chronic disease on our aging population. Some public health researchers are working toward better educating our population on living healthier lifestyles, while others are working with the government to develop policies to improve overall health, and some are studying the genetic and molecular biomarkers of diseases in populations.

I believe Heidenreich, et.al, said it quite clearly in terms of the next steps for heart disease and his statement can be extended to many other chronic diseases of the elderly: "In the public-health arena, more evidence-based effective policy, combined with systems and environmental approaches, should be applied in the prevention, early detection, and management of cardiovascular disease risk factors. Through a combination of improved prevention of risk factors and treatment of established risk factors, the dire projection of the health and economic impact of cardiovascular disease can be diminished." (Heidenreich PA, et.al, 2011). The key to lowering the burden of chronic disease is to prevent the chronic diseases from occurring.

In the current age of unequal medical coverage in the US, this will be a difficult task. Many individuals never visit a doctor unless they are already ill. Without getting into a political discussion of health care coverage or public health programs, it is clear that first, we must have people seen regularly starting at young ages by their health professionals. We need to be able to detect individuals at higher risk of developing these costly chronic conditions and intercede early in the process or before the course of the conditions begin.

My current work toward this goal combines the use of statistical genetics and epidemiologic principles to better understand the role genetics plays in the aging process, focusing on quantitative measures for the three chronic conditions highlighted above: cardiovascular disease, obesity, and chronic kidney diesase. If we can recognize individuals at a higher risk of accelerated aging and at a higher risk of age-related diseases very early in life, caregivers and clinicians can better target individuals in need of close monitoring or early and

preventative treatment. As the CARDIA study (Lee DH, et.al, 2003) has shown, early intervention may be the key. We need to take action before clinical symptoms appear. Although the use of genetics in medicine and public health is still in its infancy, it may be the key to identifying these high-risk individuals early.

Another approach to improving chronic conditions is highlighted in recent research at the Mayo Clinic in the use genetics to better understand the pathophysiological process of aging to devise better treatments. This research has implicated a category of cells, known as senescent cells that promote aging in tissues (Baker, et.al, 2011). These cells accumulate in aging tissues, like arthritic knees, cataracts, and arterial plaque and cause damage by secreting agents that stimulate the immune system and cause low-level inflammation. There is hope that cleansing the body of these cells may postpone many of the diseases of aging. A possible mechanism for clearance involves a characteristic marker gene, $p16^{Ink4a}$, which is switched on by the senescent cells, making many hopeful this finding will be clinically useful.

Findings such as that of senescent cells are very exciting, because aging is such a complex process that it can be difficult to study. One challenge of investigating the genetics of aging, in particular, is in having a measure of aging that can be applied to living individuals of varying age. Many studies have focused on specifically studying centenarians (Sebastiani and Perls, 2012; Beekman, et.al, 2013), but these studies are often very small. Another approach has been to use age at death as a measure of aging, but again there are limitations, such as the generalizability of findings to the living (Walter, et.al, 2011).

I believe the key to making early prevention and intervention possible is having good early biomarkers and genetic markers available. Because the physical decline that accompanies aging typically involves multiple systems that often interact, such as the cardiovascular system and the renal system, the first goal of this dissertation (Chapter II) aims to develop a measure of aging, based on three easily measured physiologic traits: blood pressure, waist circumference, and serum creatinine. It was also important for this measure to be applicable to a wide range of ages and among individuals who are still living. Two definitions of aging are explored. Both measures estimated a "biologic" age for each individual. The first measure accomplishes this by calculating a predicted age, while the other uses the individual's risk of mortality to assign them a "risk" age which is equal to the age at which someone in the census population has the same mortality risk. Both measures are then compared to the person's real age. In this way we can classify individuals as either being younger or older biologically than their actual age.

These measures of aging can be used to identify new biomarkers or predictors of aging. If there is a genetic component to the traits, we would expect them to have a significant degree of heritability. Using four family-based cohort studies, with differing pedigree structures, heritability of both traits was explored. After confirming that both traits are moderately heritable in Chapter II, Chapter III contains a genome-wide association study (GWAS), incorporating an additional unrelated cohort of middle-aged individuals, and ultimately will contain genotype and phenotype data on over 30,000 individuals after replication.

Based on evolutionary theories of aging and past research, expected findings from the GWAS should likely include genes that control the levels of activities, such as DNA repair and

antioxidant defense, and thus indirectly regulate longevity (Kirkwood, 2008). This is because aging has not been shown to be programmed, but rather results from accumulation of somatic damage, owing to limited investments in maintenance and repair. People may wonder how these deleterious variants have escaped natural selection. The two main theories are that the variants have late action late-acting deleterious effects, after the age where natural selection is greatest (sexual maturity) or that there is a trade-off between benefit at an early age against harm at older ages, termed antagonistic pleiotropy (Ness and Williams, 1996).

Because kidney function is one of the best predictors of age and because of my personal interest in it from working on the Dialysis Outcomes and Practice Patterns Study (DOPPS) for over 10 years, Chapter IV of this dissertation focuses specifically on chronic kidney disease. As previously mentioned, the rates of CKD and ESRD are increasing all over the world. Mortality rates once a person reaches end-stage are over 20% per year. The key to improving these numbers is to keep patients from reaching ESRD.

Over our lifetime, kidney function decreases for all individuals. Among healthy individuals, it has been reported that kidney function, measured as estimated glomerular filtration rate (eGFR), declines on average 0.4 – 1.2 ml/min/year after age 40 (Vlassara, et.al, 2009). There are known clinical predictors of accelerated decline in kidney function, such as diabetes, hypertension, smoking, and family history. Recent studies, by such consortia as CKDGen (Köttgen, et.al, 2009 and Böger, et.al, 2011), have also determined several genetic loci that are predictive of kidney function. To examine the potential clinical and public health utitility of genetic markers, I have investigated the use of genetic risk score for the decline in

kidney function. The score is based on the reported loci from CKD-Gen significantly associated with eGFR and the prevalence kidney disease. This was accomplished by exploiting the longitudinal nature of the SardiNIA Study on Aging and examining the change in kidney function over time.

Through this work, it is my hope to demonstrate the utility of incorporating the use of genetic information into clinical practice at an early stage may lead to improvements in chronic disease outcomes. While the clinical use of using genetic scores remains to be tested, they could potentiallly reduce the societal burden of care as well as improved survival and quality of life for individuals as they age.

CHAPTER II

Development of Two Measures of Aging and Assessment of Their Heritability in Cohorts with Varied Pedigree Structures

2.1 Introduction

When one thinks of aging, they commonly think of heart disease, cancer, and diseases such as Alzheimer's, but these age-related pathologies are actually a side-effect of the human body's loss or malfunction in cell processes. Also, unlike pathologies, aging occurs in every human given sufficient time (Hayflick, 2004). While gerontology studies age-related disease and degeneration, with the goals of preserving health and prolonging human life, it does not look at the fundamental causes of aging or predisposition for individuals to age at such variable rates. The fundamental cause of aging and its biological variation are rooted in the cellular and molecular mechanisms that are consequences of the interplay of environmental and genetic factors.

Genetic studies have been conducted to better understand aging and assess its heritability. However, these have mainly focused on studying 1) longevity, using very selected cohorts of centenarians (Sebastiani and Perls, 2012; Beekman, et.al, 2013); 2) non-human

populations (Frankowski, et.al., 2012; Osiewacz, et.al., 2013); or 3) specific age-related pathologies, such as cancer (Han, et.al, 2013). While these studies have identified genetic loci and chromosomal regions of interest, we do not know how generalizable animal studies are to human populations. Also, there hasn't been replicated evidence of key genetic elements underlying aging in human populations.

To begin to better understand the genetic mechanisms of aging within population-based studies of human, we first focus on defining a measure of aging among five cohorts representing over 30,000 individuals of ages ranging from 14 to 90 years of age who have not been selected on the basis of any health measures or outcomes. Next, to determine if there is a significant genetic component to this measure of aging, the heritability was also estimated in three different pedigree structures: large pedigrees (up to five generations), small pedigrees (two to three generations), and within twins (one generation).

2.2 Methods

2.2.1 Samples

Aging traits were first developed based on individuals aged 20-89 years of age and of European-descent from five genetic cohorts. The sample included 9,612 individuals from the Atherosclerotic Risk in Communities Study (ARIC), 3,018 individuals from the offspring cohort of the Framingham Heart Study (FHS), 6,135 individuals from the SardiNIA Study on Aging (SardiNIA), 9,998 siblings from the Swedish TwinGene Study (TG), and 4,838 siblings from the

TwinsUK Study (TwinsUK), totally 33,601 individuals. Heritability was assessed among the four cohorts ascertained as relatives (FHS, SardiNIA, TG, and TwinsUK), which included 23,989 individuals.

The ARIC study is a prospective study designed to investigate the etiology of atherosclerosis (The ARIC Investigators, 1989) and included only individuals aged 44-65 years of age at baseline. While there were four recruitment sites (Forsyth County, North Carolina, Jackson, Mississippi, the suburbs of Minneapolis, Minnesota, and Washington County, Maryland) in the study with a total sample of 15,792 individuals, this investigation contains only the participants who were of European descent and had complete data for the variables of interest (n=9,633). Participants of African descent were not included, so the sample would be more homogeneous genetically to the other cohorts. Baseline data collection occurred between 1986 and 1989. Because this cohort does not contain family information, their data was employed in creation of the aging traits, but not used in the heritability analyses.

Although the FHS was started in 1948 as a prospective investigation of coronary heart disease (Dawber, 1950) in Framingham, Massachusetts, this investigation included participants of the offspring cohort, who were the descendants of the initial cohort participants. Initial examinations for this cohort began in 1971 (Kannel, 1979), but only information from the seventh visit contained the variables of interest and was included. Visit seven occurred in the early 2000's. Use of the offspring cohort also allowed us to study the heritability of the traits in small pedigrees. This sample contained a total of 1,749 families, with over two-thirds of the family structures representing sib-ships. Only adult men and women who were at least 20

years-old and have completed an informed consent were eligible for participation. Individuals were excluded from the study if they had an allergy to latex; had active Reynaud's disease, as manifested by daily attacks of Reynaud's currently blue fingers or ischemic finger ulcers; had in the past had a radical mastectomy on right side; or who refused or withdrew from the test.

The SardiNIA Study is a prospective study containing close to 60% of the inhabitants of four regions in southeast Sardinia, Italy. Baseline visits began in late 2001 (Pilia, 2006) and the study is currently conducting the fourth visits. This investigation used information from the first three visits, which were conducted approximately 3-years apart. This cohort contains a rich selection of physiologic measurements on all participants and includes 588 large multi-generation pedigrees. The number of generations represented ranged from one to five, with an average of three generations.

Both of the twin studies included in this investigation are European (TwinGene in Sweden and TwinsUK in the United Kingdom) and contain a combination of monozygotic and dizygotic twins. Each study contains approximately two-thirds dizygotic twins and one-third monozygotic twins. All pedigrees contained only the twins and therefore included only one generation of individuals. The studies differed in their recruitment of participants and age-distributions.

The TG project contains approximately 10,000 participants drawn from the Swedish Twin Registry (Katsika, et.al, 2010). The Swedish Registry was established in the 1960s to study how smoking affects our health and at present contains information on approximately 85,000 twin pairs. Recruitment for participants in the TG project was restricted to twins born before

1958 and contained individuals who completed self-reported questionnaires on health and medication data are collected from self-reported questionnaires. Blood sampling materials were then collected at a local health care center and a health check-up was administered. In the simple health check-up, height, weight, circumference of waist and hip, and blood pressure were measured (http://ki.se/ki/jsp/polopoly.jsp?l=en&d=13903&a=30244).

The TwinsUK study contains approximately 12,000 twin volunteers from the UK Adult Twin Registry, which was started in 1992 (Moayyeri, 2012). Blood draws for collection of DNA for use in genetic studies occurred between April 2004 and May 2007 and were collected on a total of 5,024 individuals who were active and consented for genetic studies. Although the age range of the full cohort spanned 18 to 108 years, individuals participating in the genetic portion of the study tended to be younger, with an age range of 18 to 80 years.

2.2.2 Predicted Age Differential

To quantify the variation in aging between individuals, we first estimated a "biological" age based on physiological traits for each individual. Physiologic trait selection was done originally using the SardiNIA sample and began with forty quantitative traits known or hypothesized to be associated with aging and which has very low rates of missing values (< 0.5%). Traits examined included: four anthropomorphic traits (waist circumference, height, weight, and BMI), immunologic measures (basophils, eosinophils, monocytes, lymphocytes, neutrophils, white blood cell count, IL-6, and C-reactive protein), lipid measures (total cholesterol, HDL, LDL, and triglycerides), liver measures (AST, ALT, bilirubin, and gamma GT), blood count measures (platelet count, red blood cell count, serum iron, hemoglobin,

transferrin, and mean corpuscular volume), adipokines and glycoproteins (adiponectin, leptin, and fibrogen), glucose metabolism (insulin, glucose, and HbA1c), atherosclerosis measures (intima medial thickness, SBP, and DBP), and serum measures of uric acid, sodium, potassium, and creatinine (or estimated glomerular filtration rate, using the CKD-Epi formula). A detailed description of the measurements of all traits has been previously described in the 2006 paper by Pilia, et.al.

Using linear regression with forward selection and assessing the partial R^2 of each trait, it was found that the majority of the variation in age could be explained by three traits that were common between all of the cohorts involved in the study, which included: serum creatinine, waist circumference, and systolic blood pressure. After inclusion of these three traits in the model, additional traits explained less than 1% of the remaining variability. A full-list of the traits investigated along with their univariate estimates can be found in Table S2.1. To account for the use of anti-hypertensive medication use, the standard adjustment used in genetic studies of adding 10 mmHg to the reported SBP for all individuals taking anti-hypertensive medications was implemented.

The first trait estimated, predicted age differential, is the difference between this estimated "biological" age for each person and their real age (i.e., the residual from the predicted age model). More specifically, predicted age differential = Age_i – Predicted Age_i, where Predicted $Age_i = \hat{\alpha} + \sum_{i=1}^{3} \hat{\beta}_i \, x_i$. The trait yields a positive number if the individual is older biologically or a negative number if the individual is younger biologically than their actual calendar age.

Examination of the predicted ages vs. age revealed a floor/ceiling effect or regression to the mean effect, where very young individuals were much more likely to be predicted older than they were and very old individuals were predicted to be younger than they were. Efforts were made to improve the fit, by employing the use of squared and cubic terms, but the extra terms did not improve the fit (Figure S2.1). This was expected, as closer investigation within the SardiNIA study showed a fairly linear association between age and the three physiologic measurements used to predict it (Figure S2.2). To correct for this bias, the differences between predicted age and actual age were calculated for each individual were standardized within age decade, as a post-hoc adjustment.

To validate the predicted age differential was actually indicative of health; associations were examined between the standardized trait and death, as well as, the prevalence of comorbid conditions within the SardiNIA cohort, using logistic regression. These models accounted for family clustering using a generalized estimating equation, with compound symmetry covariance structure. Conditions examined included: gastro-intestinal, blood disorders, bone, cancer, cardiac, cerebrovascular, depression, diabetes, endocrine, hypertension, immunological, kidney/urinary, dyslipidemia, liver, lung, metabolic, neurologic, and skin.

In the final models, predicted ages were obtained separately by each cohort, using linear mixed models, which accounted for family clustering. As a sensitivity analysis, using the SardiNIA cohort, a linear mixed model was also fit, using the calculated kinship matrix as the

random effect. The two models did not differ meaningfully and therefore the simpler and more expedient approach was used for all cohorts. Models were adjusted for sex.

2.2.3 Mortality Risk Differential

A second aging trait was also explored that incorporated an individual's risk of death based on the same three physiologic traits used in the predicted age modeling. This method employed the use of census mortality rates or counts per population (specific to each of the cohort populations in the study). Each person's estimated 1-year survival predicted by the three physiologic traits was compared to the national statistics and individuals were assigned a "risk" age that was equal to the census age at which the mortality risks were equivalent. The difference between this risk age and actual age was then computed. An example person is displayed in Supplementary Figure S2.3.

US mortality rates were determined from the Social Security Administration actuarial life tables (http://www.ssa.gov/oact/STATS/table4c6.html), Sardinian mortality estimates were calculated from Italian government website (http://demo.istat.it/unitav/index.html#), Swedish estimates were taken from Eurostat European Commission data base (http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database), and UK mortality was estimated from the UK Office for National Statistics (http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Mortality+Rates).

Because survival curves and predicted risk of 1-year mortality are expected to be very different at different ages, Cox proportional hazards regression was employed to estimate the

1-year survival for each individual based on their covariate values and the estimated baseline hazard for all cohorts where specific entry time and censoring data was available (ARIC, FHS, TG, and TwinsUK). In the SardiNIA study, where age of death was known for individuals who died during follow-up, a discrete survival analysis was employed using a repeated-measures logistic model, which employed the information available from the three visits for each individual. This model was used to estimate the 3-year odds of survival for each individual. Mortality analyses were run separately for males and females, since some sex-specific differences were observed in the association between physiologic traits and mortality risk. For display in the graphics, the risk age was centered for each cohort. This has no effect on the estimation of heritability or interpretation of results because it is simply a scalar shift.

To allow for comparison between the SardiNIA mortality results and the other cohorts by using a Cox regression, two assumptions were made to create the follow-up time information needed. First, individuals who did not die during the study follow-up were censored at their last visit. Secondly, for individuals who died during the study, their date of death was assumed to be during the middle of the year of their reported age of death.

All analyses excluded individuals greater than 90 years of age, since the association between physiologic measures and age often differs in this age group, compared to younger individuals. All modeling was completed using SAS 9.2 and 9.3 software (SAS, Carry, NC).

2.2.4 Heritability Analysis

Narrow-sense heritability was estimated for each trait by cohort, based solely on their phenotypes, using variance-component based estimation. This was accomplished by fitting a simple model with two variance components (a heritable additive polygenic component and an individual specific environmental component) and two covariates (sex and age). Estimates were calculated for the full-sample within each cohort and also for a restricted sample of individuals aged 45-65. This age range was common across the five cohorts and within the range of data where model fit was best. This age range also contains ages at which humans begin to see age-related changes in health. Heritability was estimated using Merlin software (Abecasis, 2002).

2.3 Results

2.3.1 Traits

Differences were observed in the age distributions between the different cohorts, as displayed in Figure 2.1. The SardiNIA study had the widest age range (14-90 years), while the ARIC study had the smallest age range (44-66 years). FHS and TwinGene studies lacked young individuals and contained individual's whose ages ranged from the mid-forties to 90. The TwinsUK study was the youngest cohort included with ages ranging from 18 to 80, but contained very few individuals over 70 years of age.

The characteristics of interest for this study are shown in Table 2.1 for each cohort. As illustrated by the distributions displayed in figure 2.1, the SardiNIA and TwinsUK studies had the youngest mean ages, in the mid-forties, while FHS and TwinGene samples had mean ages in the

low to mid-sixties. All cohorts were > 50% female, but the TwinsUK was overwhelmingly female with only 8.6% males. Serum creatinine values were highest in the ARIC and FHS studies, with means greater than 1 mg/dl. In all cohorts males had higher creatinine values than females. Average SBP ranged from 120 to 140 mmHg between the cohorts, with the highest means observed in the oldest cohorts. Waist circumference was also highest in the oldest cohorts and larger among the male participants. The percentage of individuals receiving an anti-hypertensive medication differed markedly, with a higher prevalence of use in the oldest cohorts.

Tables 2.2 and 2.3 contain the model fit estimates by cohort. As expected, mainly significant and positive associations were seen between each physiologic measure and age (Table 2.2), although variation is seen between the cohorts. The magnitude of the regression coefficient was largest for the SardiNIA study, which had the largest range of ages, while the smallest effect sizes were seen in the ARIC cohort which had the tightest age range.

Estimates from the Cox mortality models showed more variation between the cohorts (Table 2.3), although almost all were in the expected direction of higher values being associated with a higher hazard ratio. The mortality HR for creatinine and waist circumference was largest in the older cohorts among males (SardiNIA: HR=1.06 for 0.1 higher creatinine and HR=1.16 for 5 cm larger waist circumference; TwinGene: HR=1.08 for 0.1 higher creatinine and HR=1.09 for 5 cm larger waist circumference).

Very similar estimates were seen for the odds of death using the discrete survival model when compared to the Cox model results in SardiNIA, with a 6% higher odds of death for every

0.1 higher creatinine (p=0.0005) and a 9% increase in the risk of death for every 5 cm larger waist circumference (p=0.005). Systolic blood pressure was not highly predictive of mortality, although border-line significant (p=0.05), and was in the opposite direction as expected. The C-statistic for concordance was quite high at 0.86.

Table 2.4 contains the odds ratios for the association between the predicted age difference trait and the odds of mortality, as well as prevalence of other comorbid conditions, among the SardiNIA sample. The trait was significantly associated with death (OR=1.60, p=0.0002). It was also significantly associated with the comorbid conditions typically defined by the three physiologic traits used to create the trait: metabolic syndrome (OR=1.55 per 1 SD higher, p<0.0001), hypertension (OR=1.48 per 1 SD higher, p<0.0001), and kidney/urological disease (OR=1.41 per 1 SD higher, p<0.0001). It was also found to be associated with other comorbid measures such as immunologic disease (OR=1.27 per 1 SD higher, p=0.03), diabetes (OR=1.24 per 1 SD higher, p<0.0001), coronary artery disease (OR=1.20 per 1 SD higher, p<0.0001), and depression (OR=1.11 per 1 SD higher, p=0.03).

Figure 2.2a displays the floor/ceiling effect seen in the predicted age estimates from the linear models before standardization, for each cohort. The floor/ceiling effect is more pronounced for cohorts with a smaller age range, where as the model fits much better in the SardiNIA cohort with the large age-range, except in the very young and very old individuals. Figure 2.2b shows the risk age estimates plotted by age. Although the estimates don't appear to be linear along the full range of ages, less deviation between the expected and observed ages are seen at the extreme ages.

Histograms of the two traits are displayed in Figure 2.3. The predicted age differential is normally distributed, as expected, because it was standardized within age decade. The distribution of the mortality risk differential nears normal, but it slightly bi-modal, with a small second node to the left. This reflects that fact that separate mortality models were estimate per sex and compared to sex-specific census values in each cohort. Heritability analyses were adjusted for sex to assure this did not affect the estimates.

2.3.2 Heritability

Heritability estimates for the four related cohorts are shown in Figure 2.4. Estimates vary markedly when comparing the twin studies with the cohorts that include parental phenotypes. Estimates for the heritability of the predicted age differences differ less markedly and range from 0.25-0.33 in the non-twin studies compared to 0.70 in the twin studies. Estimates of h^2 for the twin studies for the risk age trait are questionably high and reasons for this will be discussed, while estimates for the SardiNIA and FHS studies, using the Cox model were markedly low. The heritability estimated within the SardiNIA study using the discrete survival model, which estimated a 3-year showed higher heritability (h^2=0.30) than the 1-year survival estimates from the Cox model.

When heritability estimation was restricted to the sample of 45-65 year old individuals in each cohort, the heritability estimates were much more similar (h^2=0.30 in FHS, h^2=0.54 in SardiNIA, h^2=0.60 in TwinGene, and h^2=0.67 in TwinsUK) for the predicted age differential.

Estimates were more variable for the risk age differences, likely due to the difference age-ranges in which the models estimates were devised.

2.4 Discussion

Two different measures of aging were developed in this study among five cohorts with varying ranges of ages and pedigree structures. The measure based on predicted age showed more consistent heritability estimates across the different cohorts ranging from 25% in the full SardiNIA sample to 70% in the TwinsUK study. The trait based on an individual's risk of mortality was less consistent across the studies and within the SardiNIA cohort varied depending on the type of mortality model used to estimate 1-year and 3-year survival. Estimates within the twin studies were quite high and possibly not valid, due to the lack of substantial differences seen in the predicted 1-year survival between twin pairs, even when the co-twins had differing physiologic traits.

Both aging traits are highly dependent on obtaining good predictions of both age and mortality risk. Having a larger range of ages in a cohort yielded the best model fit for predicting age. The three physiologic traits chosen for this analysis were based on work in the SardiNIA study, which had the largest age range. These same three measures were then examined in the other cohorts, so that heritability estimates would be comparable and future genome-wide association studies could be conducted. Without that aim in mind, it is highly likely that different physiologic measures may have been better at predicting age in the other cohorts and may have yielded better heritability estimates.

The cox proportional hazards models depend highly on having a sufficient number of individuals of advanced age and deaths for good fit. All three of the physiologic traits (serum creatinine, SBP, and waist circumference) showed a positive association with age, but the association was stronger in cohorts with more individuals of older ages. Therefore cohorts, such as the TwinsUK cohort whose average age is only 47.5 years and contains very few individuals over the age of 70 years, showed weaker associations between the physiologic traits and mortality. The ARIC study, which had the smallest age range, also did not show large associations between the measures and mortality, nor did the SardiNIA study, which has the highest percentage of very young individuals.

Having a way to measure aging among individuals of varying ages could be very useful in furthering our understanding of the genetics of aging. Most work in this field has been limited to studies of longevity, focusing on individuals who have attained a specific age. These studies have shown heritability estimates between 20-30% (Sebastiani and Perls, 2012), even among twins. Other studies among humans have focused on the mitochondrial DNA (De Benedictis, 1999) haplogroups of centenarians and found variants associated with longevity. Animal studies have had more success than studies among humans in detecting variants with large effects on longevity (Paaby and Schmidt, 2009), and while these studies have informed human studies, they aren't often directly applicable.

While estimates of predicted age were fairly consistent across the cohorts in our study, there may be an alternate explanation of the estimates of heritability for the risk age trait in the twin studies. Estimates of survival are highly dependent on age. Because the twin pairs are the exact same age, their predicted 1-year survival estimates are also very similar. In the majority of

pairs, this led to both twins being assigned the same risk age and in-turn the same risk age difference. With near perfect correlation within twin pairs, the h^2 estimates will be biased upward.

Although the results of this work may not be directly clinically applicable, there are genetic components of aging that can be discerned with this technique. Future work focused on discovering the loci and pathways involved may facilitate the ability to detect an individual's risk for accelerated aging. Individual's at high risk for accelerated aging could be followed more closely by their physicians and could potentially lower their risk by avoiding high risk and unhealthy behaviors.

There are some limitations in this study. In particular to allow for comparability between the different cohorts, the same three physiologic measures were used to predict age and the risk of mortality. These measures were selected from SardiNIA may not be the most appropriate for other studies because of differences in individual characteristics and age. Improvements are expected if cohorts developed their own list of physiologic traits to be used in the predictive equations. We also would not expect all predictors of age to be linear across the entire age range, although they were fit this way in the current analysis. Squared and quadratic terms were tested and did not show an improvement in fit. Had data been available on hormone levels or menopause status, this would have been an important factor to examine in relation to the shape of the physiologic and age associations.

Also, the use of 1-year survival estimates vs. a longer follow-up may have limited the variability in the predicted survival estimates. The use of 3-year survival estimates from the Cox

model could improve this limitation, as evidenced from the discrete survival model employed in the SardiNIA sample. Heritability estimates were greatly increased by the use of 3-year probability of dying vs. the 1-year survival estimate from the Cox model.

This study indicates that investigations in the genetics of aging do not need to be limited to special cohorts of very advanced ages. By defining an individual's biological age we can investigate the difference between their actual age and the age of their body. Variation does exist in this difference that is heritable and future work will attempt to link this variability to specific genetic loci.

2.5 Acknowledgements

I owe a special thank you on this work to Dr. Hal Morgenstern for his willingness to talk through the results over casual meetings and in the inspiration and guidance in development of the risk age trait. I also would like to acknowledge Ana Veldez and Sarah Metrustry of King's College London for their collaboration and contribution of the TwinsUK data and to Per Svensson and Johannes Arpegård at Karolinska University in Sweden for their collaboration and contribution of the TwinGene data.

I would also like to acknowledge Yan Zhou, who helped me examine repeated measures linear mixed models, accounting for the pedigree structure in the SardiNIA, while exploring the predicted age models. These models were so computationally intense that regular software

could not be employed, so I appreciate very much Yan's expertise in writing R code to make the investigation of these models possible.

Figure 2.1: Comparison of Age Distributions, by Cohort. A: ARIC, n=9,633. B: FHS, n=3,232. C: SardiNIA, n=6,161. D: TwinGene, n=10,664. E: TwinsUK, n=4,838.

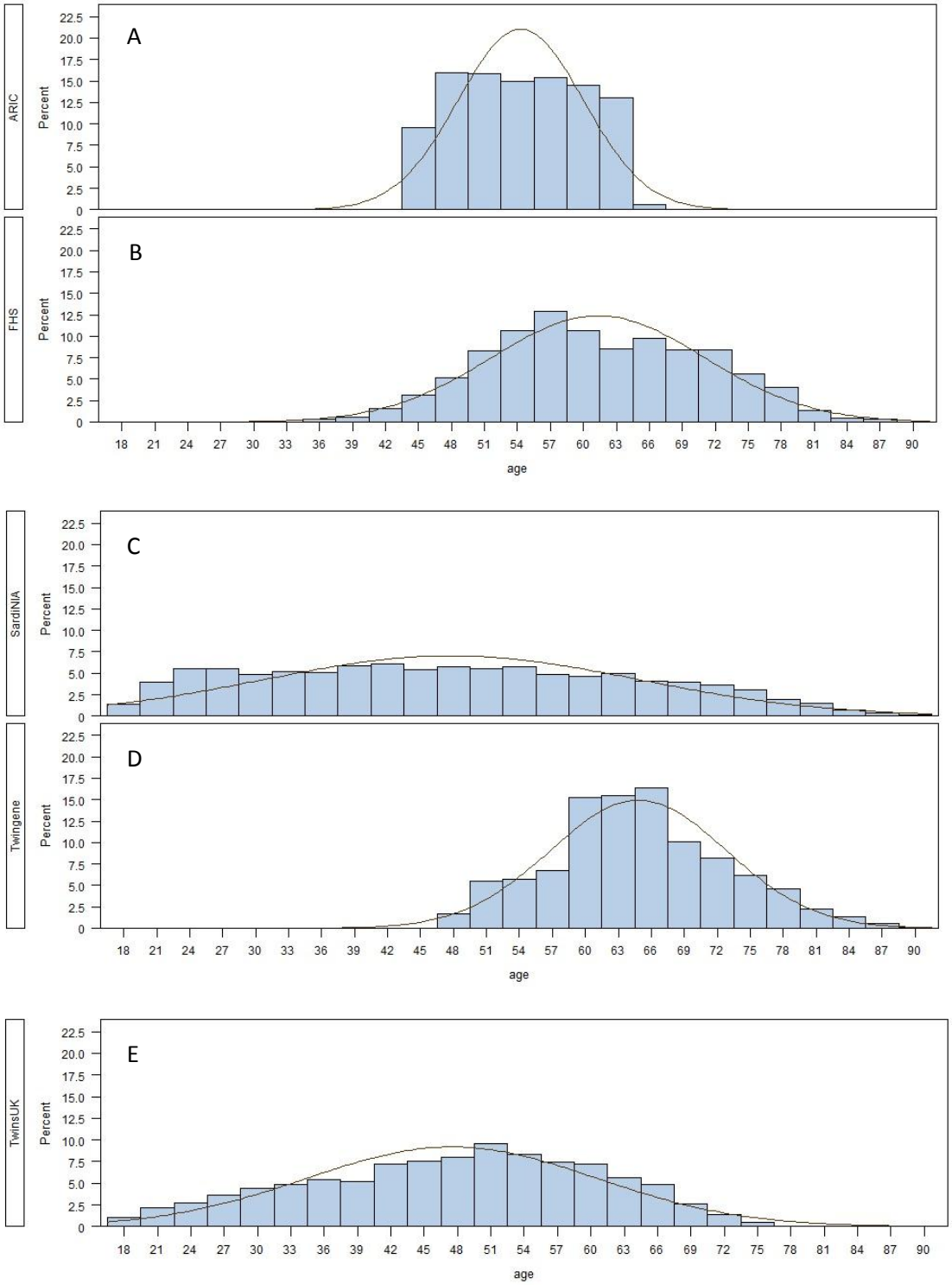


Table 2.1: Characteristics Used in Estimating Measures of Aging, by Cohort.

Measure	Mean (SD) or Percent					ANOVA F-statistic P-value
	ARIC (n=9,633)	FHS (n=3,232)	SardiNIA (n=6,161)	TwinGene (n=10,664)	TwinsUK (n=4,838)	
Demographics:						
Age (years)						
All	54.3 (5.7)	61.3 (9.7)	49.6 (17.6)	64.8 (8.5)	47.5 (13.0)	<0.0001
Males	54.7 (5.7)	61.3 (9.7)	49.7 (18.0)	65.5 (8.0)	46.5 (13.4)	<0.0001
Females	54.0 (5.7)	61.4 (9.7)	49.6 (17.4)	64.3 (8.1)	47.6 (13.0)	<0.0001
Male (%)	47.0	46.1	42.5	47.5	6.9	<0.0001
Physiologic Traits:						
Creatinine (mg/dl)						
All	1.09 (0.18)	1.07 (0.32)	0.76 (0.22)	0.88 (0.32)	0.76 (0.18)	<0.0001
Males	1.20 (0.16)	1.19 (0.35)	0.86 (0.24)	0.99 (0.40)	0.92 (0.17)	<0.0001
Females	0.98 (0.14)	0.97 (0.24)	0.68 (0.18)	0.79 (0.16)	0.75 (0.17)	<0.0001
SBP (mmHg)*						
All	120.4 (18.5)	130.6 (20.9)	127.1 (20.2)	141.3 (20.4)	120.8 (17.6)	<0.0001
Males	122.1 (17.5)	131.8 (19.4)	131.9 (18.9)	142.5 (20.0)	120.3 (17.7)	<0.0001
Females	118.9 (19.2)	129.6 (22.0)	123.6 (20.5)	140.3 (20.7)	127.8 (15.1)	<0.0001
Waist circ. (cm)						
All	96.2 (13.3)	99.8 (14.1)	84.8 (13.1)	91.7 (12.3)	80.8 (11.3)	<0.0001
Males	99.6 (10.4)	103.3 (11.2)	90.2 (11.3)	97.1 (10.5)	94.1 (10.1)	<0.0001
Females	93.1 (14.8)	96.8 (15.5)	80.8 (13.0)	86.8 (11.7)	79.8 (10.7)	<0.0001
Treatment:						
Antihypertensive medication (%)						
All	19.7	34.4	13.8	23.9	8.0	<0.0001
Males	20.1	37.5	14.5	24.8	0.0	<0.0001
Females	19.4	31.8	13.4	23.1	8.6	<0.0001

*Individuals taking anti-hypertensive medication had 10 mmHg added to their reported SBP

Table 2.2: Linear Mixed Model Estimates by Cohort for Predicted Age.*

Cohort	N	R^{2**}	Serum Creatinine (per 0.1 mg/dl)		SBP (per 10 mm/Hg)		Waist (per 5 cm)	
			β	p	β	p	B	p
ARIC	9,612	0.09	0.30	<0.0001	0.9	<0.0001	0.02	0.34
FHS	3,018	0.18	0.53	<0.0001	1.8	<0.0001	-0.02	0.75
SardiNIA	6,135	0.47	0.69	<0.0001	3.6	<0.0001	2.71	<0.0001
TwinGene	9,999	0.13	0.34	<0.0001	1.3	<0.0001	-0.10	0.002
TwinsUK	4,838	0.19	0.69	<0.0001	2.4	<0.0001	1.12	<0.0001

*Adjusted for sex

**R-squares taken from model that did not account for family clustering

Table 2.3: Hazard Ratios for Time to Death Models* for all Cohorts and Odds Ratio for Discrete Survival in SardiNIA.

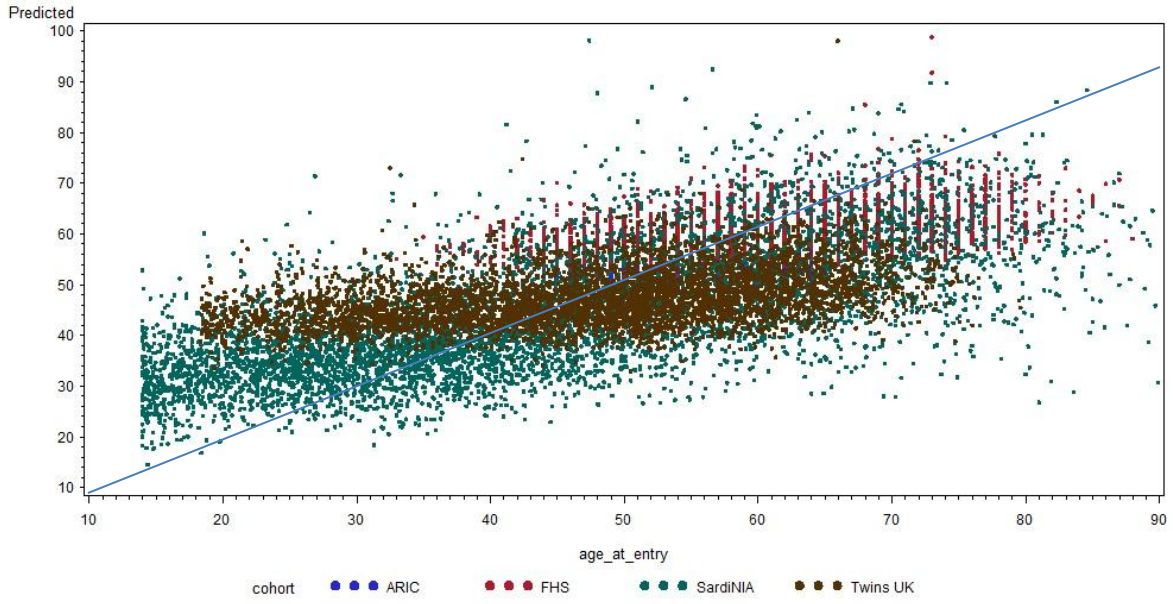
Cohort	N	Deaths	Serum Creatinine (per 0.1 mg/dl)		SBP (per 10 mm/Hg)		Waist (per 5 cm)	
			HR	p	HR	p	HR	p
ARIC								
Females	5,089	572	1.03	0.34	1.10	<0.0001	1.03	0.048
Males	4,521	860	1.01	0.66	1.05	0.02	1.08	<0.0001
FHS								
Females	1,606	123	0.96	0.33	1.07	0.11	1.01	0.75
Males	1,410	202	1.07	<0.0001	1.08	0.046	1.11	0.0014
SardiNIA								
Females	3,392	101	1.03	0.40	0.94	0.18	1.02	0.51
Males	2,505	133	1.06	0.10	1.04	0.36	1.16	0.005
TwinGene								
Females	5,265	191	1.08	<0.0001	0.97	0.30	1.09	0.003
Males	4,734	304	1.03	<0.0001	0.99	0.73	1.03	0.35
TwinsUK								
Females	4,505	235	1.02	0.35	1.04	0.19	1.01	0.71
Males	333	17	1.16	0.25	1.18	0.25	1.33	0.035

Discrete Survival Model (3-year risk of death)	C-statistic	Serum Creatinine (per 0.1 mg/dl)		SBP (per 10 mm/Hg)		Waist (per 5 cm)	
		OR	p	OR	p	OR	p
SardiNIA	0.86	1.06	0.0005	0.94	0.05	1.09	0.005

* Left-truncated to account for survival time before entering the study.

Figure 2.2: Comparison of age traits to age, by cohort

A. Predicted age by age, before standardization



B. Centered risk age by age

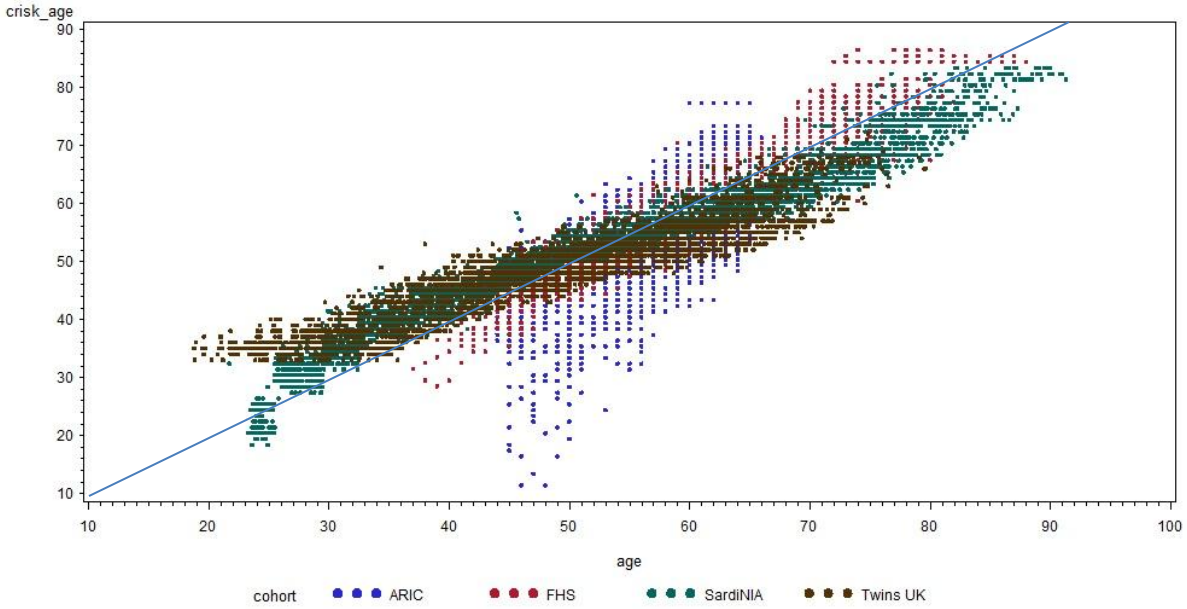
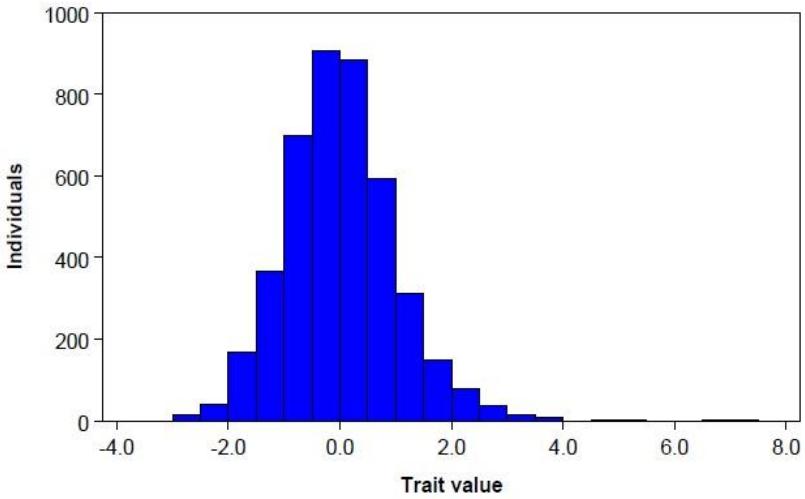


Figure 2.3: Histogram of the Distribution in Predicated Age Differential and Risk Age Differential

A. Predicted Age Differential (Standardized within Age Decade)



B. Risk Age Differential

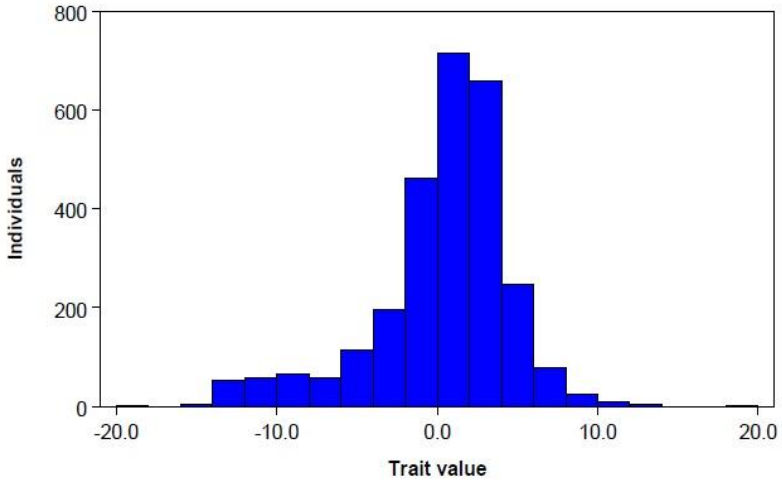
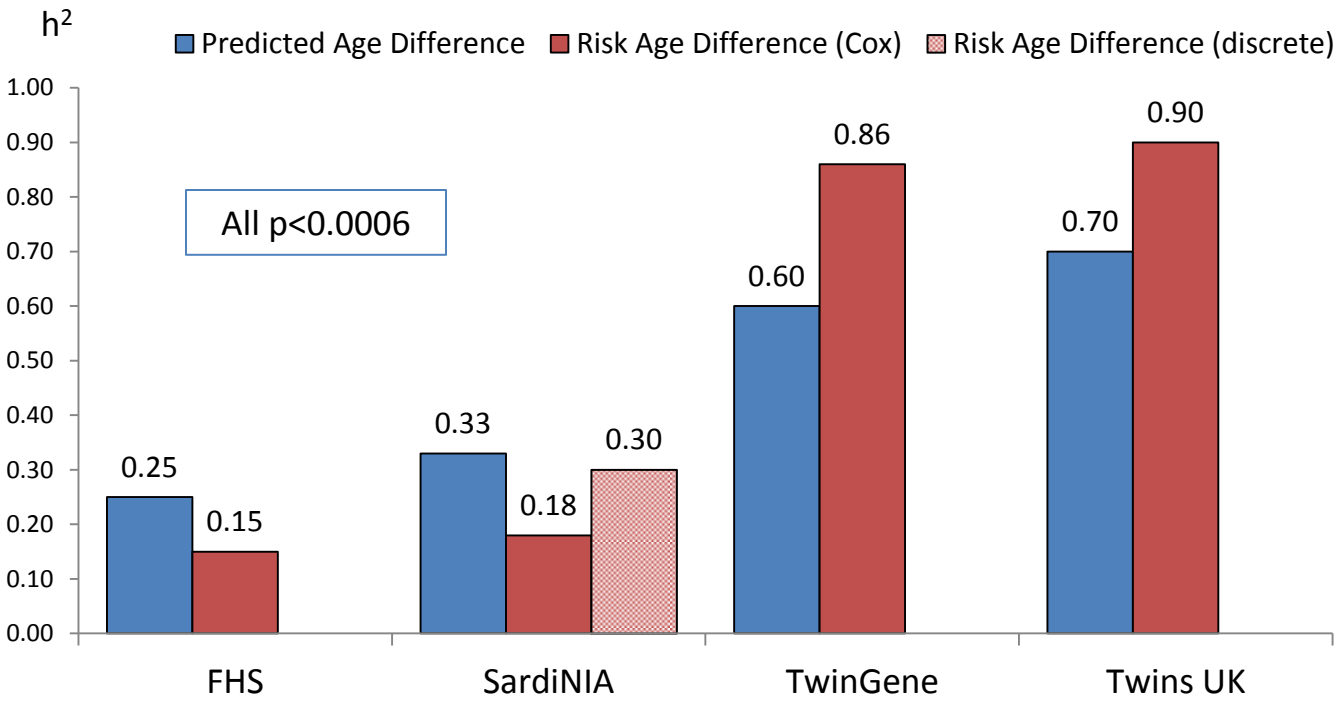


Table 2.4: Odds Ratios of Death or New Disease in SardiNIA Sample by Predicted Age

Measure	**Odds Ratio for 1 SD Higher Standardized Predicted Age Differential**	**p-value**
Death	*1.60*	*0.0002*
Metabolic Disorder	*1.55*	*6.2e-9*
Hypertension	*1.48*	*<1e-30*
Kidney or Urologic Disease	*1.41*	*<1e-30*
Immunological Disorder	*1.27*	*0.03*
Diabetes	*1.24*	*1.4e-7*
Coronary Heart Disease	*1.20*	*3.1e-6*
Cerebrovascular Disease	1.17	0.21
Depression	*1.11*	*0.03*
Gastro-intestinal Disorder	1.07	0.12
Skin Disease	1.05	0.33
Dyslipidemia	1.04	0.30
Lung Disease	1.04	0.49
Liver Disease	1.02	0.75
Endocrine Disorder	1.00	0.96
Bone Disorder	0.99	0.91
Cancer	0.98	0.81
Neurologic Disorder	0.94	0.38
Blood Disorder	0.93	0.10

Figure 2.4: Heritability Estimates by Trait and Cohort

A. Full Age Ranges for Each Cohort



B. Limited Age Ranges for Each Cohort: 45-65 years

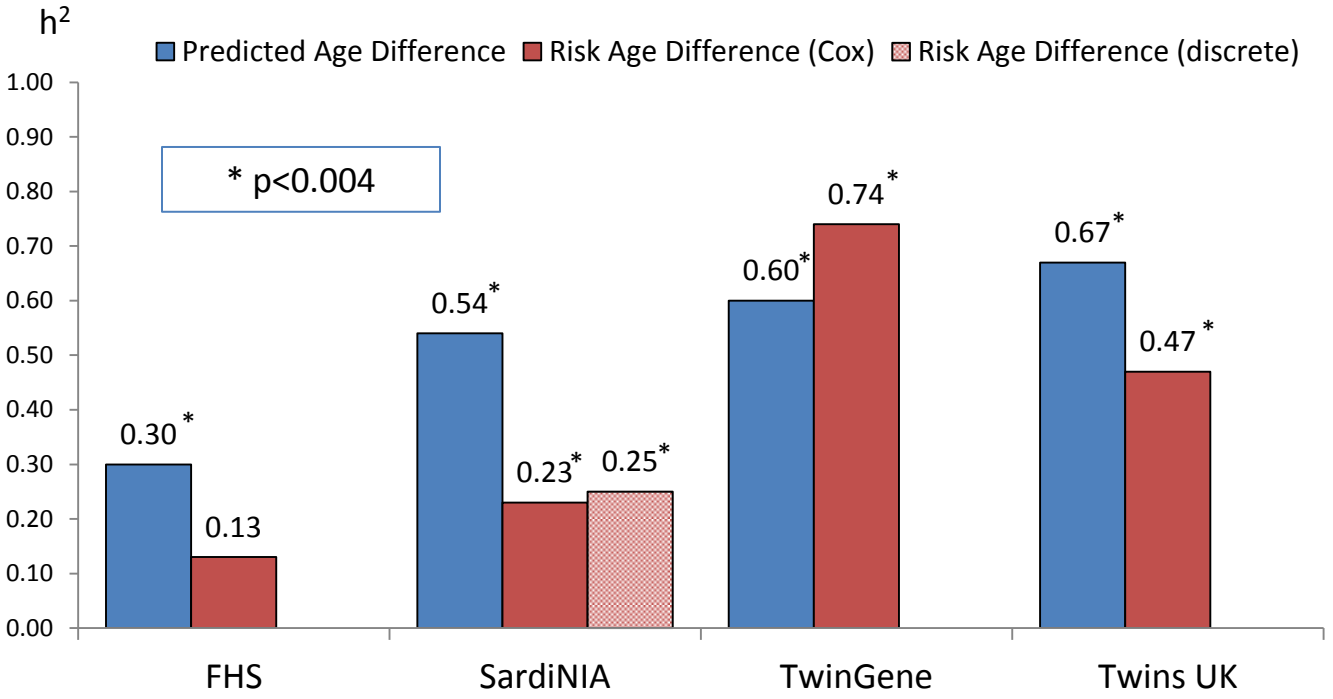
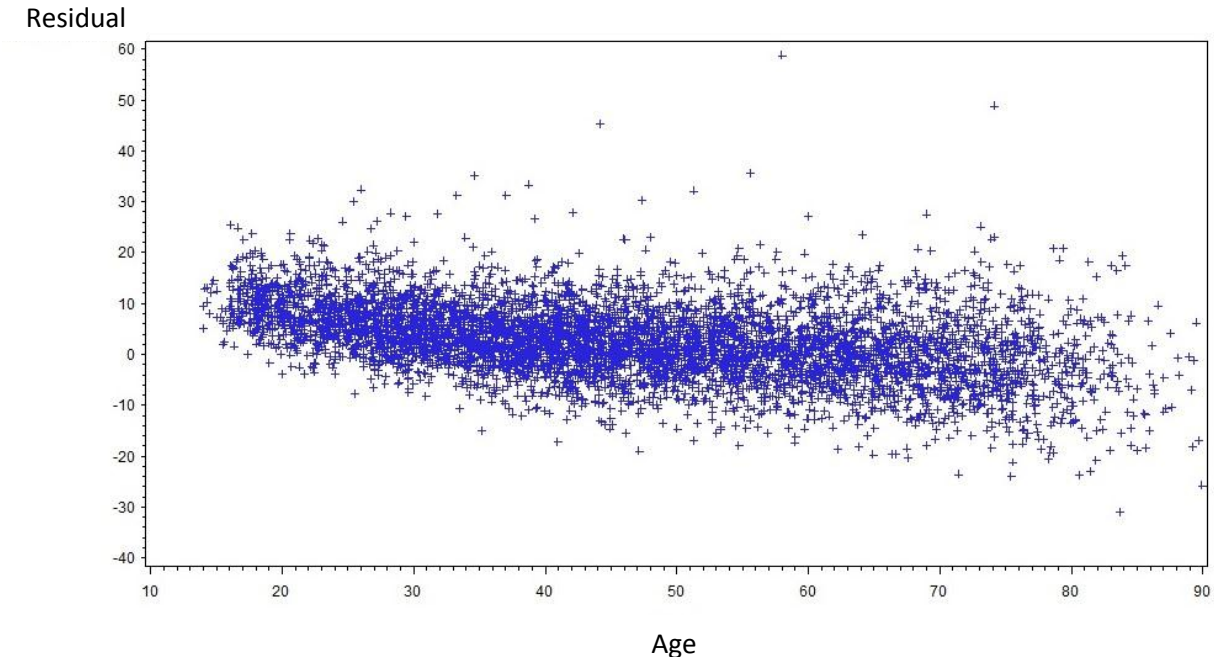


Table S2.1: Univariate Models for Choosing Traits for Predicted Age Model, SardiNIA Study, N=6,114.

Measure	Beta	P-value	Measure	Beta	P-value
S. Creatinine (per 0.1 mg/dl)	1.0	<1e-30	MCV (per 1 fL)	0.3	<1e-30
eGFR (per 10 ml/min)	-6.3	<1e-30	CRP (per 1 mg/dl)	1.5	6.3e-13
SBP (per 10 mm/Hg)	5.2	<1e-30	Platelets (per 10^3/ul)	-0.04	<1e-30
DBP (per 5 mm/Hg)	3.9	<1e-30	Basophils (per 1 %)	0.5	0.3520
Waist (per 5 cm)	3.7	<1e-30	Eosinophils (per 1 %)	0.6	2.7e-7
Weight (per 5 Kg)	1.8	<1e-30	Lymphocytes (per 1 %)	-0.3	<1e-30
Height (per 1 cm)	-0.71	<1e-30	Monocytes (per 1 %)	0.3	0.0022
Ln IMT (per 0.1)	6.4	<1e-30	Neutrophils (per 1 %)	0.2	1.9e-11
BMI (per 1)	1.95	<1e-30	Potassium (per 1 mEq/L)	8.8	<1e-30
Bilirubin (per 0.1 mg/dl)	1.5	4.2e-7	RBC (per 10^6/uL)	-1.9	2.6e-6
ALT (per 5 U/L)	0.4	2.3e-14	Iron (per 5 microU/L	0.1	0.0016
AST (per 5 U/L)	0.9	<1e-30	Sodium (per 1 mEq/L)	0.6	<1e-30
Uric Acid (per 1 mg/dl)	3.4	<1e-30	Transferrin (per 10 mg/dl)	-0.4	<1e-30
Total Chol (per 10 mg/dl	1.7	<1e-30	Triglyc. (per 10 mg/dl	0.5	<1e-30
GammaGT (per 10 U/L)	0.8	<1e-30	Fibrinogen (per 10 mg/dl)	0.7	<1e-30
Glucose (per 10 mg/dl)	2.4	<1e-30	WBC (per 10^3/ul)	-0.9	6.5e-11
HDL (per 5 mg/dl)	1.4	<1e-30	IL6 (per ug/mL)	1.0	<1e-30
HbA1C (per 0.5)	3.8	<1e-30	Leptin (per 100 pg/mL)	0.03	<1e-30
Insulin (per 5 mU/L)	0.8	1.4e-8	Adiponectin (per mg/mL)	0.9	<1e-30
LDL (per 10 mg/dl	1.8	<1e-30	Hemoglobin (per mg/dl)	0.6	7.8e-5

Figure S2.1: Comparison of Residual Plots for Predicted Age using Linear and Squared Physiologic Traits as Predictors

A. Linear Predictors of Age (S. Creatinine, SBP, Waist):



B. Squared Predictors of Age (S. Creatinine^2, SBP^2, Waist^2):

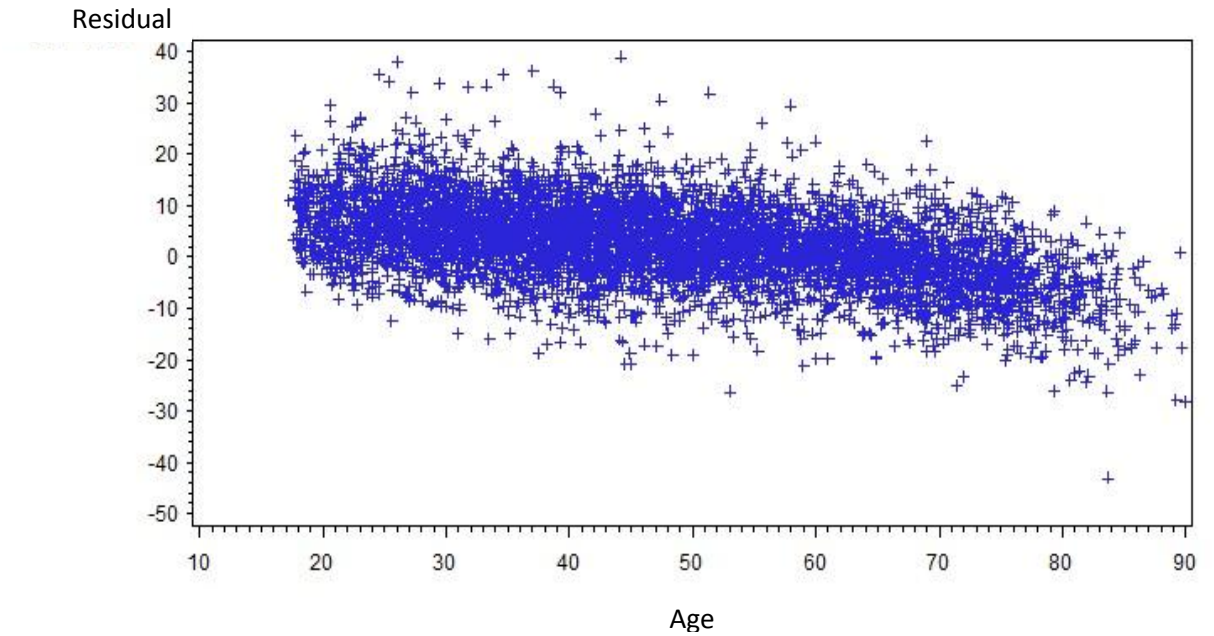
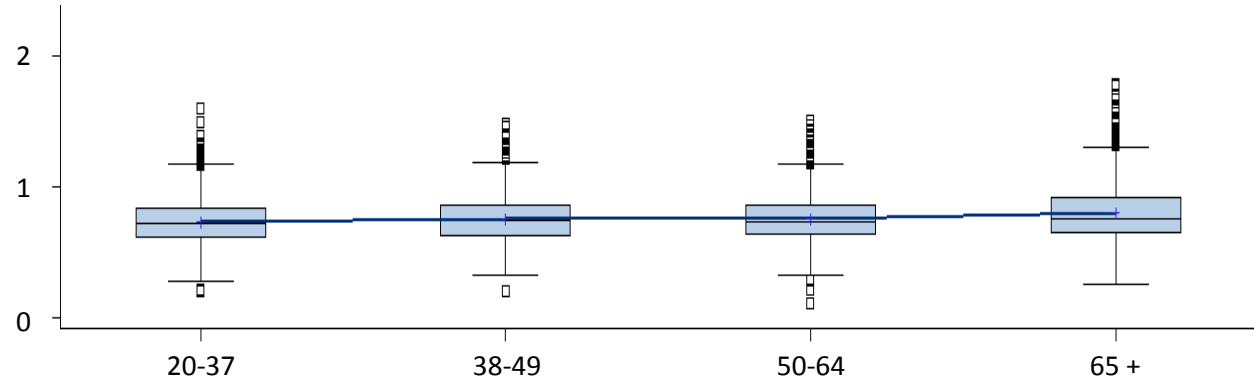
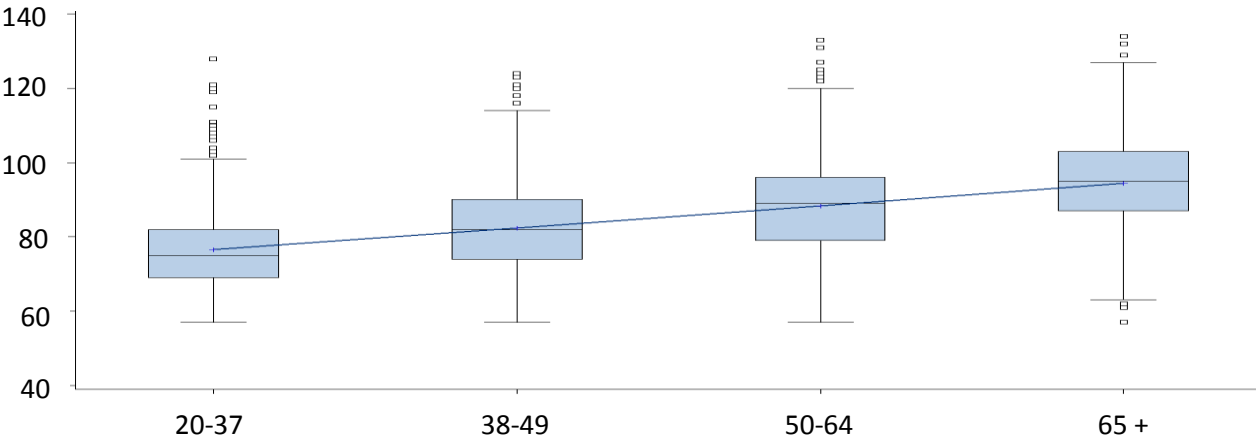


Figure S2.2: Distribution of Three Physiologic Measurements used to Predict Age and Mortality Risk over Age Quartile Range from the SardiNIA Study

A. Creatinine (mg/dl)



B. Waist Circumference (cm)



C. Adjusted Systolic Blood Pressure (mmHg)

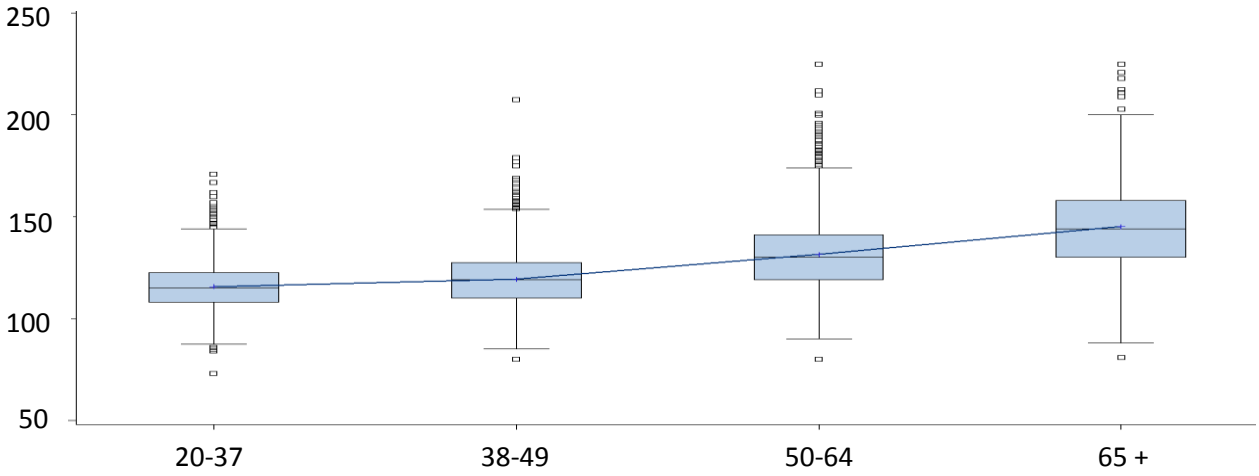
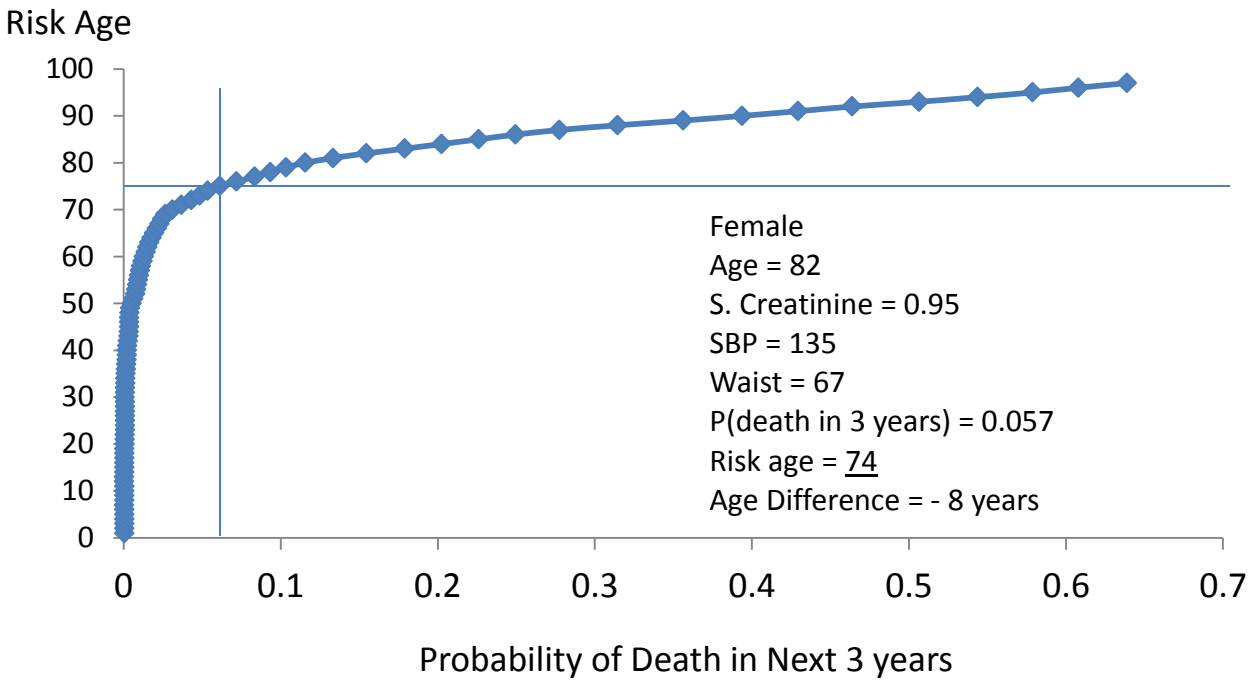


Figure S2.3: Example of Risk Age Trait for an Individual Aged 82, with S. Creatinine of 0.95 mg/dl, SPB of 135 mmHg, and Waist Circumference of 67 cm, from SardiNIA



CHAPTER III

Genome-wide Study of Two Aging Traits in 20,000 Individuals

3.1 Introduction

While genome-wide significant studies (GWAS) have uncovered a host of loci associated with age-related diseases (Jeck, et.al, 2012), to date, no genome-wide association consortia of aging have detected any single nucleotide polymorphisms (SNPs) significant at $p < 5 \times 10^{-8}$. A large meta-analysis of genome-wide association studies from nine studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium was recently conducted for two outcomes: a) all-cause mortality and b) survival free of major disease or death and also failed to produce genome-wide significant findings (Walter, et.al, 2011).

As evidenced in the Walter et.al study, work in this area has typically focused on morality as the measure of aging. We propose a different approach that will be more applicable to general population studies with a wide range of ages. As described in detail in previous work (Chapter III), we have developed two different measures of aging that compare a person's "biological" age to their actual age, in an attempt to quantify the variation in aging. Using family

information from four family-based cohorts, both traits have moderate heritability with estimates ranging from 0.25 in a small pedigree setting to 0.70 among twins for the predicted age differential.

The current work focuses on these traits in five cohorts of varying pedigree structures to determine if genomic regions associated with measures of aging could be identified and replicated.

3.2 Methods

3.2.1 Samples

The same five cohorts that were described in Chapter II were used in this analysis. All five cohorts are longitudinal cohort studies and included the participants of European-descent. A total included: 9,612 individuals from the Atherosclerotic Risk in Communities Study (ARIC, The ARIC Investigators, 1989), 3,018 individuals from the offspring cohort of the Framingham Heart Study (FHS, Dawber, 1950 and Kannel, 1979), 6,135 individuals from the SardiNIA Study on Aging (SardiNIA, Pilia, et.al, 2006), 9,998 twins from the Swedish TwinGene Study (TG, Katsika, et.al, 2010), and 4,838 twins from the TwinsUK Study (TwinsUK, Moayyeri, 2012), totally 33,601 individuals. ARIC, FHS, SardiNIA, and TwinsUK (approximately 20,000 individuals with genotype data) were used in the discovery stage of this analysis and TwinsUK will be used for replication (approximately 10,000 individuals with genotype data).

3.2.2 Phenotypes

Two traits were developed (predicted age differential and mortality risk differential) to measure aging among population samples of individuals of varying ages. A full description can be found in Chapter II of this dissertation, but briefly, the traits are both based on the difference between an individual's estimated "biologic" age and their actual age. Both traits employed three common physiologic measures to estimate biologic age: serum creatinine, systolic blood pressure, and waist circumference. An increase in the value of each trait was generally associated with an increase in estimated biologic age.

The first trait employed mixed linear regression, accounting for family clustering, to predict and individual's biologic age. The second trait was based on the individual's risk of mortality, (1) using Cox proportional hazards models for all cohorts for 1-year and (2) Cox models for all cohorts, except for SardiNIA, for 3-year survival estimates. The SardiNIA study employed discrete survival analysis, using repeated measures logistic regression compared to the mortality risk of the appropriate country's census data. Each individual was then assigned a "risk" age that was equal to the age at which the individual's survival estimate was equivalent to the census survival estimate. Both traits are the difference between the estimated biologic age and actual age. For example, predicted age differential = Age_i – Predicted Age_i, where Predicted $Age_i = \hat{\alpha} + \sum_{i=1}^{3} \hat{\beta}_i \, x_i$. Both traits yield a positive number if the individual is older biologically or a negative number if the individual is younger biologically than their actual calendar age.

3.2.3 Genotype Data and Imputation

Genotype data came from a variety of chip platforms, but were all imputed on the HapMap2 reference panel, release 22, build 36, CEU population (http://hapmap.ncbi.nlm.nih.gov/, The International HapMap Consortium, 2007). Data was downloaded from dbGaP (http://www.ncbi.nlm.nih.gov/gap) for the ARIC and FHS cohorts, while data was contributed through collaboration with the TwinGene and TwinsUK cohorts.

The FHS cohort employed a combination of Affymetrix 500K and MIPS 50K chip (for a total of 534,982 genotyped SNPs) and used MACH (version 1.0.15, Li, et.al, 2010 and Li, et.al, 2009) to impute autosomal SNPs.

Genotype data for the ARIC study came from the Affy 6.0 chip, while the SardiNIA cohort used a combination of the Affy 50K, 500K, and 6.0 chips. The ARIC and SardiNIA cohort data was imputed, in-house. As all ARIC samples were genotyped on the same chip a single imputation was run using a combination of MACH for phasing and Minimac software (Kim, http://genome.sph.umich.edu/wiki/Minimac) for imputation.

As the SardiNIA cohort contained individuals genotyped on three different chips, a two-step imputation was run: one haplotype-based and one family-based (Naitza, 2012). The first was based on imputing separately samples typed with the 500K (N=1412) and with the 6.0 array (N=1097), treating individuals as unrelated. The best guessed genotype (not the dosages) were chosen for all SNP with RSQR >0.3. Using the best guessed genotypes from the two imputations, samples were put back in family pedigrees, and additional imputed SNPs were

discarded if there were an excess of Mendelian errors. Finally, the individuals typed with the 10K array were merged back into the data and family-based imputation was implemented using Merlin, which imputes SNPs based on the estimated haplotype transmission. This imputation was possible because samples typed with 500K and 6.0 were mostly founders and key descendents, so one needed to follow the haplotype inheritance flow estimated using markers in common between the 10K chip and the 500K/6.0 arrays to accurately impute SNPs (Burdick, et.al, 2006).

For the TwinsUK sample, 5710 twins have undergone a genome-wide scan of either 317,000 SNP markers (Illumina HumanHap300 Bead Chip) or 610 000 SNPs (Illumina HumanHap610 Quad Chip). The data have been fully imputed using IMPUTE version 2 software (Howie BN, et.al., 2009) and quality checked (Moayyeri, 2012).

A summary table of genotyping and imputation procedures can be found in Supplementary table S3.1.

3.2.4 Statistical Analysis

Genome-wide association models were run using Merlin software (Abecasis, 2002) for the ARIC, FHS, and SardiNIA Studies and Genabel (Aulchenko, 2007) for the TwinsUK Cohort using an additive model and fitted dosage data for each SNP. The reported pedigree structure was used in the form of the kinship matrix to account for relatedness in the three family-based studies. Each trait was examined unadjusted, adjusted for age, and adjusted for age and sex. It was important to examine the associations adjusted for age, particularly for the predicted age

trait, to account for the standardization of the trait within age decade. On average 2.5M SNPs were available for analysis in each cohort. All SNPs with minor allele frequency (MAF) greater than zero were examined.

Quantile-quantile (QQ) plots and genomic control (GC) values were calculated using R software (R Development Core Team, 2008), to assess population substructure. GC values ranged from 0.98 to 1.09 for the predicated age trait (Figure 3.1 displays plot for each cohort adjusted for age) and 0.92 to 1.09 for the risk age trait (Figure 3.2 displays plot for each cohort adjusted for age). Although the GC values and QQ plot for FHS appear to show an overcorrection for pedigree structure in the age risk trait, no problems could be detected from the pedigree data downloaded from dbGaP. All SNPs with imputation quality < 0.30 were removed before analysis and therefore not the cause.

Meta-analysis was performed, using METAL software (Willer, 2010), to combine the evidence for association from the four individual studies. Two weighting schemes, sample size and standard error, were investigated and showed similar results. Heterogeneity between the studies was also assessed. SAS (SAS, Cary, NC) was employed for creation of all traits and Locus Zoom (Pruim and Welch, 2010) was used to create plots of interesting genomic regions.

3.2.5 Bioinformatic analysis of functional implications of genetic variants

Variants with $p < 5 \times 10^{-6}$ were analyzed by computational methods to infer potential functional relevance. Conservation analysis was performed using the comparative genomics feature in Ensembl (http://useast.ensembl.org/index.html) to compare the human variant with

nine eutherian mammals, including chimp, mouse and pig. To analyze the putative biological effects of each exonic or intronic variation, either the major or minor allele was input into predictive programs flanked on either side by 100 bp of sequence. Potential human transcription factor binding sites were analyzed using default parameters of the transcription element search system from the University of Pennsylvania (http://www.cbil.upenn.edu/cgi-bin/tess/tess). SNPs were interrogated for effects on expression levels, i.e. whether they are expression quantitative trait loci (eQTLs) for that gene using the SNP and CNV Annotation (SCAN) Database (http://www.scandb.org/newinterface/about.html) from the University of Chicago. This database includes information on expression levels in numerous populations, including those from the HapMap.

3.3 Results

3.3.1 Meta-Analysis

One SNP (rs10496861) reached genome-wide significance in this study for the trait based on the difference between predicted age and age, after adjustment for age (table 3.1). The SNP is located within an intronic region of the gene *LRP1B* which produces low density lipoprotein receptor-related protein 1B. The direction of association was the same within all four cohorts and there was no significant heterogeneity found. Minor allele frequency (MAF) for this SNP is 34%. The meta-effect size is +0.28 with a $p = 5.75 \times 10^{-9}$. Because the trait was standardized by age decade this effect size translates into varying age differences, based on an individual's age. Standard deviations by age decade ranged from 7 to 11 years in all cohorts, so an effect size of +0.28 translates into approximately a +2 year to +3 year higher than average

age difference for each age range. The Locus Zoom plot of this SNP (Figure 3.3) shows the typical spike compared to the surrounding regions.

Table 3.1 also displays the SNPs that were found to be marginally significant ($5 \times 10^{-8} < p < 5 \times 10^{-7}$) and SNPs of interest for validation ($5 \times 10^{-7} < p < 5 \times 10^{-6}$). Of particular interest is a set of SNPs on Chromosome 21 located between 20.7 and 20.8 Mb (Figure 3.4). These SNPs are approximately 500 Kb upstream of the NCAM2 gene. Seven SNPs in this region had a meta-analysis p-values of $5 \times 10^{-8} < p < 5 \times 10^{-7}$ and 20 more SNPs in the region had meta-analysis p-values $5 \times 10^{-7} < p < 5 \times 10^{-6}$.

Although one SNP (rs9472826 on Chromosome 6) also reached genome-wide statistical significance for the risk age trait, data was only available for analysis from the SardiNIA and TwinsUK cohorts and there was significant heterogeneity detected between the associations (Table 3.2). The MAF for this SNP was less than 1% and therefore the finding is thought to be a false positive. Similar findings for a marginally significant SNPS (rs17476005) also on Chromosome 6, but in a different region, also point to it potentially being a false positive result. Two regions of modest interest were detected on Chromosome 18. The first region is very small and located around position 41.2 Mb (Figure 3.5), which is upstream of the SLC14A2 gene. Three SNPs were detected in this region with $5 \times 10^{-7} < p < 5 \times 10^{-6}$, with MAF of approximately 2.5%. The second region is wider, spanning positions 53 to 53.1 Mb, and is upstream of two genes: ST8SIA3 and ONECUT2 (Figure 3.6). This region contained nine SNPs with significance levels $5 \times 10^{-7} < p < 5 \times 10^{-6}$.

All SNPs with $p < 5 \times 10^{-6}$ are currently being follow-up in replication work planned among 10,000 individuals in the TwinGene cohort.

3.3.2 Bioinformatic analysis of functional implications of genetic variants

The *LRP1B* gene is very large (1,900,279 bases) and belongs to the low density lipoprotein (LDL) receptor gene family (http://www.genecards.org/cgi-bin/carddisp.pl?gene=LRP1B). These receptors play a wide variety of roles in normal cell function and development due to their interactions with multiple ligands (Liu et al., 2001) by producing cell surface proteins that bind and internalize ligands in the process of receptor-mediated endocytosis. This gene has homologues in six other species: *Mus musculus, Gallus gallus, Pan troglodytes, Canis lupus familiaris, Bos Taurus, and Danio rerio* and has 41 known protein products (http://refgene.com/gene/53353). Using the University of California Santa Cruz genome browser (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=329521287&c=chr2&g=cons46way), a comparison of this SNP was made across 46 different vertebrate species (Figure S3.2). Based on these findings is appears that this position/region is not highly conserved, nor highly divergent.

Our top SNP rs10496861 is located in one of the 93 distinct introns [NM_018557.2] and has been shown to alter expression levels of the LOC407835, MAP2K2, and SLC38A5 genes within the Nigerian population (YRI) at a level p<0.0001, n=283. Expression work for the *LRP1B* gene have shown that it is expressed in the human brain, thyroid gland, skeletal muscle, and to

a lesser amount in testis but absent in other tissues, including heart, kidney, liver, lung, and placenta (Haas, 2011).

3.4 Discussion

Most human genetic studies of aging and longevity have focused on centenarians' survival, or have been candidate gene studies focused on genomic regions suspected to be associated with aging. In a new approach, using two novel measures of aging developed in and applicable to adults of ages 20 to 90 years of age, this study was able to detect a genome-wide significant genetic association for aging. Each measure used an estimated biologic age for each individual and computed the difference between that age and the individual's chronological age.

One SNP (rs10496861) on Chromosome 2 within the *LRP1B* gene reached genome-wide statistical significance with a p-value of 5.78×10^{-9} for the trait comparing predicted age to age in the meta-analysis. Because additive models were used, individuals with one additional copy of the minor allele (T compared to the wild-type G, with MAF=0.34) were on average +0.28 standardized years older than individuals with the wild-type allele. Depending on the individual's actual age, this translates into a 2 to 3 year above average biologic age compared to actual age, for each copy of the T allele. Another region of interest was found for this trait on chromosome 21 between position 20.7 and 20.8 Mb, with the most statistically significant SNP (rs1786357) just above genome-wide significance with a p-value of 7.64×10^{-8}. Seven SNPs with $5 \times 10^{-8} < p < 5 \times 10^{-7}$ were found within this region.

Although no SNPs reached genome-wide significance with the risk age trait, two regions of interest on Chromosome 18 showed promise (around position 41.2 Mb and between 53 to 53.1 Mb). These regions are currently being followed-up in replication work in the TwinGene cohort.

These findings are very exciting because if validated, may provide a way to detect an individual's risk of accelerated aging before clinical symptoms can be detected for comorbid conditions such as hypertension, diabetes, and chronic kidney disease. Work from the CARDIA study (Lee DH, et.al, 2003) has found that early intervention at a young age can have effects long after. With the current shift in the age distribution of the population, having a way to determine an individual's risk for accelerated aging could be invaluable. We need to be able to single out high risk individuals and focus efforts to change behaviors and, if necessary, closely monitor them for early signs of diseases or comorbidities, such as high blood pressure or high glucose levels.

There is support for the current *LRP1B* finding from previous work investigating successful aging (Poduslo, et.al, 2009). The study was very small (n=63) and included individuals aged greater than 85 years with Mini Mental State Examination (Folstein, 1983) scores > 26 and no major illness (cardiovascular problems, diabetes, obesity, major cancer diseases, or dementia in the family) were compared to a cohort of Alzheimer's patients. They found that specific haplotypes (3 SNPs: rs12474609, rs10201482, and rs980286) in intron 18 of the *LRP1B* gene were significantly associated with successful aging after Bonferroni correction. These three SNPs are located 32 Kb downstream from the SNP in the current study, but linkage

disequilibrium between them is very low, with r^2 ranging from 0.02 to 0.033. Since their finding was based on Alzheimer's patients and a large GWAS of Alzheimer's patients was published in 2011 (Hollingworth, et.al), their findings were examined for any hits in the LRP1B gene and non were found.

LRP1B expression appears to be restricted to specific organ. mRNA was been detected in the human brain, thyroid gland, skeletal muscle, and testes (Haas, 2011). It is absent in other tissues, such as heart, kidney, liver, lung, and placenta. This could point to a speciation function of *LRP1B* in certain organs. Many of the *LRP1B* ligands are well known factors in blood coagulation and lipoprotein metabolism, which may suggest a possible role in atherosclerosis, which is directly correlated with aging.

The biological implications described in the mRNA work above corroborate other studies that have linked this gene to phenotypes such as thyroid cancer, where inactivation of *LRP1B* results in changes to the tumor environment, conferring cancer cells an increased growth and invasive capacity (Prazeres, 2010) and insulin-resistance (Burgdorf, 2012). The rs2890652 locus upstream of *LRP1B* was associated with both increased BMI and decreased insulin sensitivity.

Through bioinformatic investigation, our most significant SNP was shown to alter expression levels of the *LOC407835*, *MAP2K2*, and *SLC38A5* genes *MAP2K2* (mitogen-activated protein kinase kinase 2) is located on 19p13.3 and the protein encoded by this gene is a dual specificity protein kinase that belongs to the MAP kinase kinase family. This kinase is known to play a critical role in mitogen growth factor signal transduction and mutations in this gene cause cardiofaciocutaneous syndrome (CFC syndrome), a disease characterized by heart

defects, mental retardation, and distinctive facial features similar to those found in Noonan syndrome. The inhibition or degradation of this kinase is also found to be involved in the pathogenesis of Yersinia and anthrax. A pseudogene (*LOC407835*), which is located on chromosome 7, has been identified for this gene and was also shown to have altered expression by *LRP1B* (http://www.ncbi.nlm.nih.gov/gene/5605). The protein encoded by SLC38A5 (solute carrier family 38, member 5), located on the X-chromosome at p11.23, is a system N sodium-coupled amino acid transporter involved in the transfer of glutamine, asparagine, histidine, serine, alanine, and glycine. The encoded protein does not transport charged amino acids, imino acids, or N-alkylated amino acids (http://www.ncbi.nlm.nih.gov/gene/92745).

A possible mechanism for the increased rate of aging that we observe in association with variation in our most significant SNP (rs10496861) that follows our evolutionary expectations is that the minor allele causes a reduction in LRP1B's endocytosis capability. Endocytosis and autophagy work as partners to take in and break up molecules. Both pathways are highly conserved as is gene *LRP1B*, demonstrated by the presence of homologues in six other organisms. Constitutive autophagy has a housekeeping role and is essential for survival, development and metabolic regulation (Markaki, 2011). Research has also shown that autophagy is also responsive to stress and can degrade damaged proteins and organelles, oxidized lipids and intracellular pathogens and that defects in the autophagic degradation system have been linked to disease pathogenesis and aging. Different signaling pathways converge on autophagy to regulate lifespan in diverse organisms and autophagy is a critical regulator of metabolic homeostasis and molecular mechanisms that promote longevity.

Another study showed that genetic inhibition of autophagy induces degenerative changes in mammalian tissues that resemble those associated with aging, and normal and pathological aging are often associated with a reduced autophagic potential (Rubinsztein, 2011). This study also showed that pharmacological or genetic manipulations that increase life span in model organisms often stimulate autophagy, and its inhibition compromises the longevity-promoting effects of caloric restriction, Sirtuin 1 activation, inhibition of insulin/insulin growth factor signaling, or the administration of rapamycin, resveratrol, or spermidine, which could be insightful for considering future targets.

With GWAS there is always a chance that the findings are false positives, which is very likely the case for rs9472826 and rs17476005 in the risk age difference meta-results. SNP rs9472826 has a MAF of only 0.6% in our sample and this SNP was only present in two of the cohorts (n=7,242 with SardiNIA and TwinsUK combined). This suggests that only 44 copies of the minor allele are present in the entire sample. SNP rs17476005 is even rarer with a MAF of 0.08% (9 copies out of the 11,319 individuals from ARIC and TwinsUK). With such a small number of copies in the sample, the results could be highly affected by error rates in both genotype calling and imputation, which are less accurate for rare alleles.

The findings for the top SNP (rs10496861) which was genome-wide significant and located in the *LRP1B* gene appear to be quite robust. The MAF is 34%, which is well in the range for effects we would expect to indentify using GWAS. Also, we saw similar direction of association in all four cohorts (n=19,860) and there was no significant heterogeneity found in the associations between cohorts. After inspection of other SNPs in the region we see the

characteristic spike on a LocusZoom plot (Figure 3.3), showing that although less significant, our analysis picked up nearby SNPs in linkage disequilibrium with our SNP.

The current study was limited to individuals of European descent to remove potential bias from inadequately accounting for population admixture, but it does limit the generalizability of the findings. Future work should be done investigating this gene and polymorphism in other ancestry groups. Another potential issue with the current work was demonstrated in the QQ plot for the risk age difference in FHS, where GC values were 0.91. This cohort created their imputed data set from samples that were genotyped on two different chips. Because of the direction of bias in the FHS sample, we would expect to find less significant results and therefore this should not be a major limitation in interpreting any significant findings. The SardiNIA cohort also had samples genotyped on different samples, but took great care in imputation, employing the family information, as described by Burdick, et.al.

There are limitations in this work due to how the traits were defined, as well. It is possible that one of the physiologic factors used to predict age or mortality risk may be driving our findings. Since the GWAS results from the CKD-Gen studies are publically available (https://intramural.nhlbi.nih.gov/labs/CF/Pages/CKDGenConsortium.aspx), I was able to verify that genome-wide significant SNP was not significant in their kidney-specific analyses. In particular, the p-value for the odds of CKD for SNP rs10496861 was 0.7728 and for the continuous measure of eGFR the p-value was 0.5380.

Analysis is currently underway to replicate the findings from this study, focusing on associations with $p < 5 \times 10^{-6}$, using 10,000 individuals from the TwinGene study. While

consistent replication can greatly improve the credibility of a genotype-phenotype association, there is still a chance it may not eliminate spurious associations due to biases shared by many studies (Kraft, 2009). If these findings do replicate and the bioinformatics hypotheses are plausible, then a next step would be to begin looking for the causal variant. This may be accomplished by accessing publically available data from the Exome Sequencing Project (https://esp.gs.washington.edu/drupal/), which is available from dbGaP. In instances where this data will not be suitable, a multi-pronged approach will be needed. Due to the fact that our most significant SNP is located in an intron, we would want to sequence the surrounding exons, including 50 bp on each side of the exons to clearly capture the intron/exon boundaries. If nothing if found within the immediate range then targeted sequencing of the entire gene may be in order, being sure to include possible promoter regions in the 2000 bp at the 5' end.

There are times when even such targeting sequencing is unable to generate hypotheses for the biological processes taking place and then one would need to start considering epi-genetic factors, such as methylation or histone modification, and consider the role of gene by environment interactions, or regulatory mechanisms.

The findings of this study have demonstrated that aging can be investigated in cohorts of varying ages and not restricted to the very elderly or need to be based on survival. An association between a polymorphism in the *LRP1B* gene and the difference in predicted age and age reached genome-wide significance and is supported by a previous study investigating successful aging. Bioinformatics investigation of this SNP has shown it to alter expression levels of the *LOC407835*, *MAP2K2*, and *SLC38A5* genes within the Nigerian population (YRI) at a level

$p<0.0001$ Replication work and specific investigations into determining the causal variant are underway. If replication of this SNP is successful, it could be a target for determining an individual's risk of accelerated aging and allow for early interventions to prolong healthy life.

3.5 Acknowledgements

I would like to acknowledge three individuals in particular whose assistance was invaluable in this project. First Serena Sanna for giving me the skills years ago to analyze GWAS data and for creating the SardiNIA genotype files needed. Secondly, to Carlo Sidore for writing a python script that I have used countless times which creates the analysis files needed for Merlin from MACH output files. Lastly, I would like to thank Xiaowei Zhan for his aide in the re-imputation of the ARIC data onto the HapMap2 reference panel.

I'd also like to acknowledge the assistance of Margaret DeAngelis and Margaux M. Morrison for their expertise and assistance in using bioinformatics techniques to infer potential functional consequences of genetic variants.

Table 3.1: Meta-analysis of Difference between Predicted Age and Age in 20,000 Individuals of European Descent, Adjusted for Age.

Locus	Chr	SNPs in Region in Significance Range	Lead SNP	Alleles/MAF	Effect Size	P	Direction of association*
Genome-wide Significant: P < 5 x 10^{-8}							
LRP1B	2	1	rs10496861	T/G/0.34	+0.28	5.78E-09	++++
Marginally Significant: 5 x 10^{-8}< P < 5 x 10^{-7}							
LRP1B	2	1	rs6732847	T/C/0.34	+0.16	3.94E-07	++++
ERC2	3	1	rs1167245	T/C/0.46	+0.20	2.35E-07	++++
PSD3	8	1	rs335222	A/C/0.04	-0.80	2.31E-07	?---
OTOGL	12	1	rs10778728	C/G/0.36	+0.21	1.80E-07	++++
KLHL1	13	1	rs2154199	G/A/0.04	+0.46	4.21E-07	++++
MDGA2	14	2	rs1683210	T/C/0.04	+0.54	1.01E-07	++++
NCAM2	21	7	rs1786357	G/A/0.16	+0.33	7.64E-08	++++
Suggestive for Replication: 5 x 10^{-7} < P < 5 x 10^{-6}							
MTOR	1	3	rs3765897	G/A/0.06	+2.07	5.14E-07	+??+
EDEM1	3	1	rs17043984	A/G/0.02	+0.60	1.12E-06	++++
CPNE4	3	1	rs16837395	C/G/0.42	+0.22	2.07E-06	++++
SI	3	1	rs12630444	G/T/0.01	-0.77	2.82E-06	----
LPP	3	1	rs9820072	T/A/0.43	+0.07	3.12E-06	+?+?
CPEB2	4	1	rs218819	G/A/0.03	+0.79	3.31E-06	+??+
NUDT12	5	4	rs7702688	A/C/0.03	+0.59	2.89E-06	++++
HLA-L	6	1	rs3094078	A/T/0.09	+0.48	1.58E-06	++++
C6orf105	6	1	rs4140558	T/C/0.28	+0.07	1.16E-06	+++?
PHACTR1	6	1	rs2327591	T/C/0.10	+0.38	2.74E-06	++++
EYA4	6	1	rs2184784	A/C/0.13	+0.37	8.89E-07	++++
IMMP2L	7	1	rs6968496	C/T/0.24	+0.26	7.11E-07	++++
FBXO25	8	2	rs17737960	C/G/0.07	+0.38	3.00E-06	++++
NKAIN3	8	1	rs10099792	G/A/0.10	+0.35	1.67E-06	++++
IL33	9	1	rs13284060	C/A/0.07	+0.39	1.89E-06	++++
KDM4C	9	4	rs4742295	A/C/0.03	-0.58	5.27E-07	----
TMEM38B	9	1	rs3010957	G/A/0.03	+0.47	3.27E-06	++++
MUSK	9	1	rs2846447	G/C/0.20	+0.28	7.70E-07	++++
GRID1	10	1	rs17106322	T/C/0.03	+0.57	7.97E-07	++++
ADAM12	10	1	rs4559596	T/C/0.43	+0.21	1.46E-06	++++
ENDOD1	11	2	rs626246	A/C/0.16	+0.33	1.01E-06	++++
KLHL1	13	1	rs9529821	A/T/0.04	+0.35	2.75E-06	++++
MDGA2	14	1	rs1769488	A/G/0.04	+0.44	1.45E-06	++++
WSCD1	17	1	rs12942336	G/C/0.05	+0.40	4.06E-06	++++
NCAM2	21	20	rs1786368	A/G/0.16	+0.30	6.16E-07	++++

* Order of studies: SardiNIA, FHS, ARIC, TwinsUK

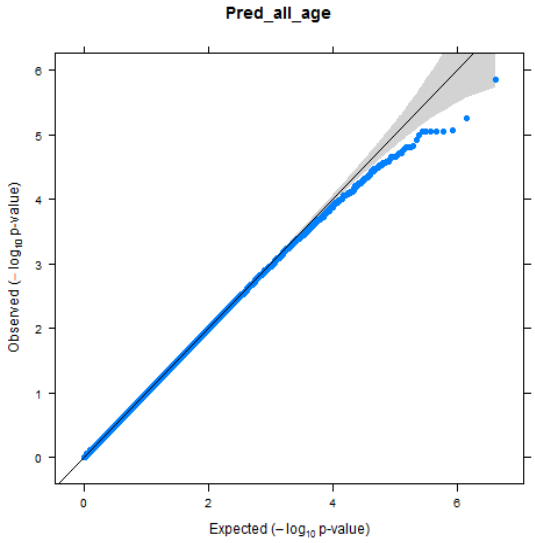
Table 3.2: Meta-analysis of Difference between Risk Age and Age in 20,000 Individuals of European Descent, Adjusted for Age.

Locus	Chr	SNPs in Region in Significance Range	Lead SNP	Alleles/MAF	Effect Size	P	Direction of association*
Genome-wide Significant: P < 5 x 10^{-8}							
PLA2G7	6	1	rs9472826	C/T/***0.006***	+12.0	4.17E-09	+??+
Marginally Significant: 5 x 10^{-8}< P < 5 x 10^{-7}							
BAI3	6	1	rs17476005	G/A/***0.0008***	-10.6	3.26E-07	??--
Suggestive for Replication: 5 x 10^{-7} < P < 5 x 10^{-6}							
ASB3/ CHAC2	2	1	rs6735011	C/T/0.35	+0.31	8.98E-07	++++
GABRA1	5	1	rs1531263	G/A/0.49	-0.31	4.81E-06	---?
ACO1/ MIR873	9	1	rs10738859	G/T/***0.003***	-2.6	7.18E-07	----
LOC286370	9	1	rs4744117	A/G/0.34	+0.32	3.54E-06	?+++
C12orf10	12	1	rs4759054	C/T/***0.008***	+4.5	8.01E-07	+??+
CMKLR1	12	1	rs11113865	A/G/0.07	+.56	4.46E-06	++++
SLC14A2	18	3	rs1462152	C/T/***0.02***	-0.85	5.48E-07	----
ST8SIA3/ ONECUT2	18	9	rs4801093	A/C/0.24	+0.33	1.75E-06	++++

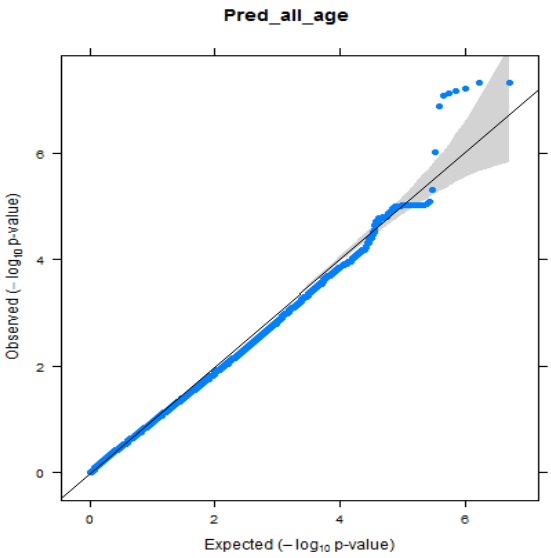
* Order of studies: SardiNIA, FHS, ARIC, TwinsUK

Figure 3.1: QQ plots of Difference between Predicted Age and Age, Adjusted for Age, by Cohort

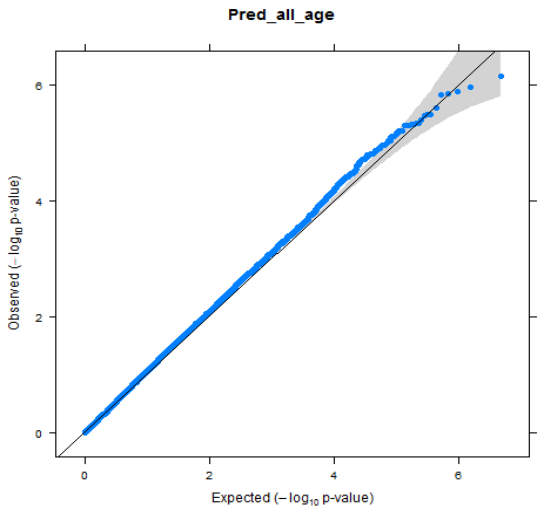
A. ARIC



B. Framingham



C. SardiNIA



D. TwinsUK

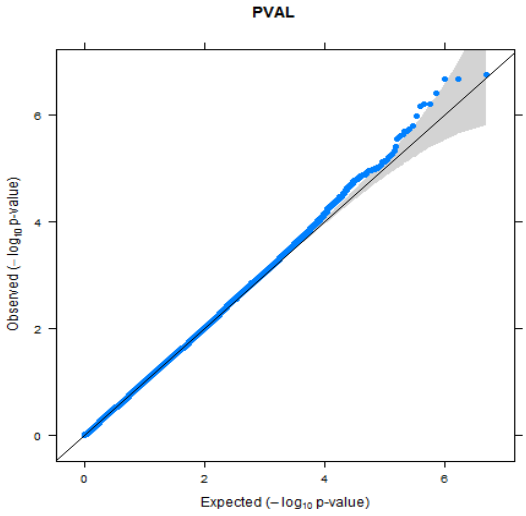
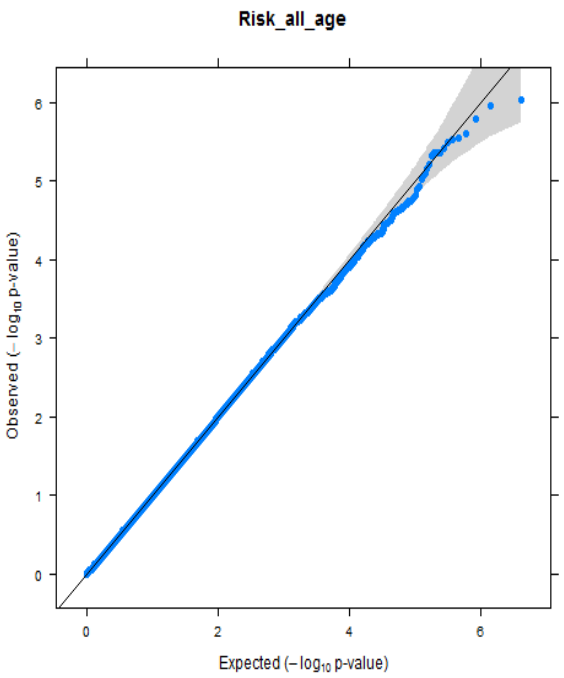
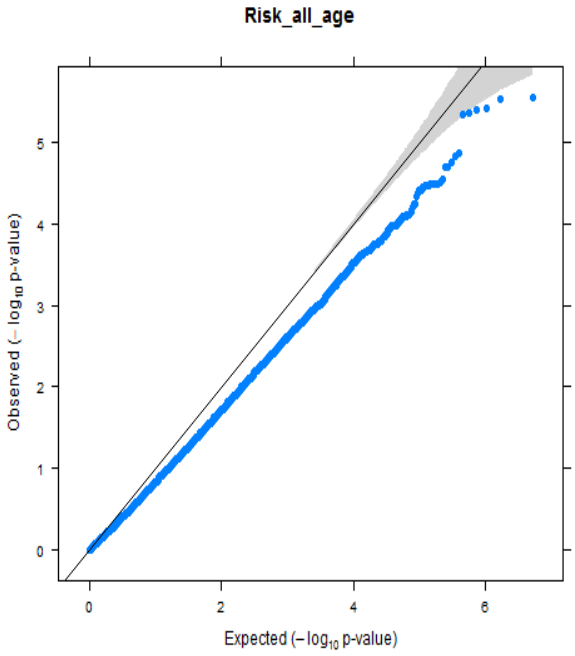


Figure 3.2: QQ plots of Difference between Risk Age and Age, Adjusted for Age, by Cohort

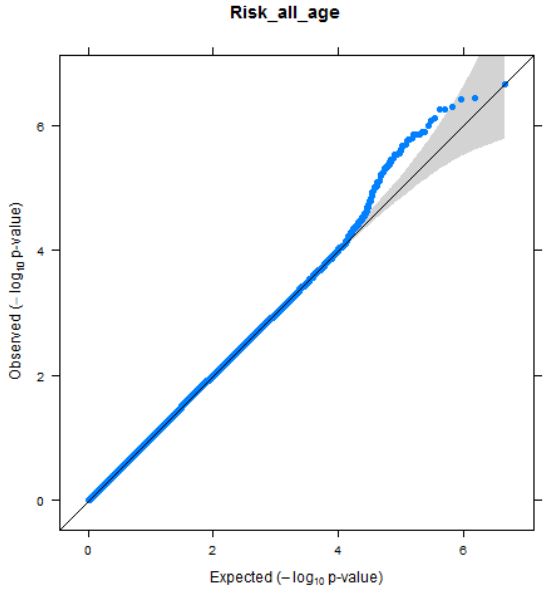
A. ARIC



B. Framingham



C. SardiNIA



D. TwinsUK

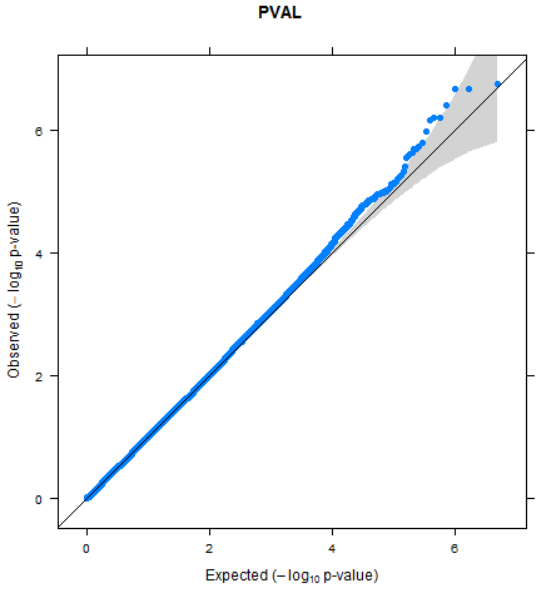


Figure 3.3: Locus Zoom Plot of Gene *LRP1B* on Chromosome 2 for Predicted Age Difference, Adjusted for Age, From Meta-Analysis, N=20,000 Individuals of European Descent

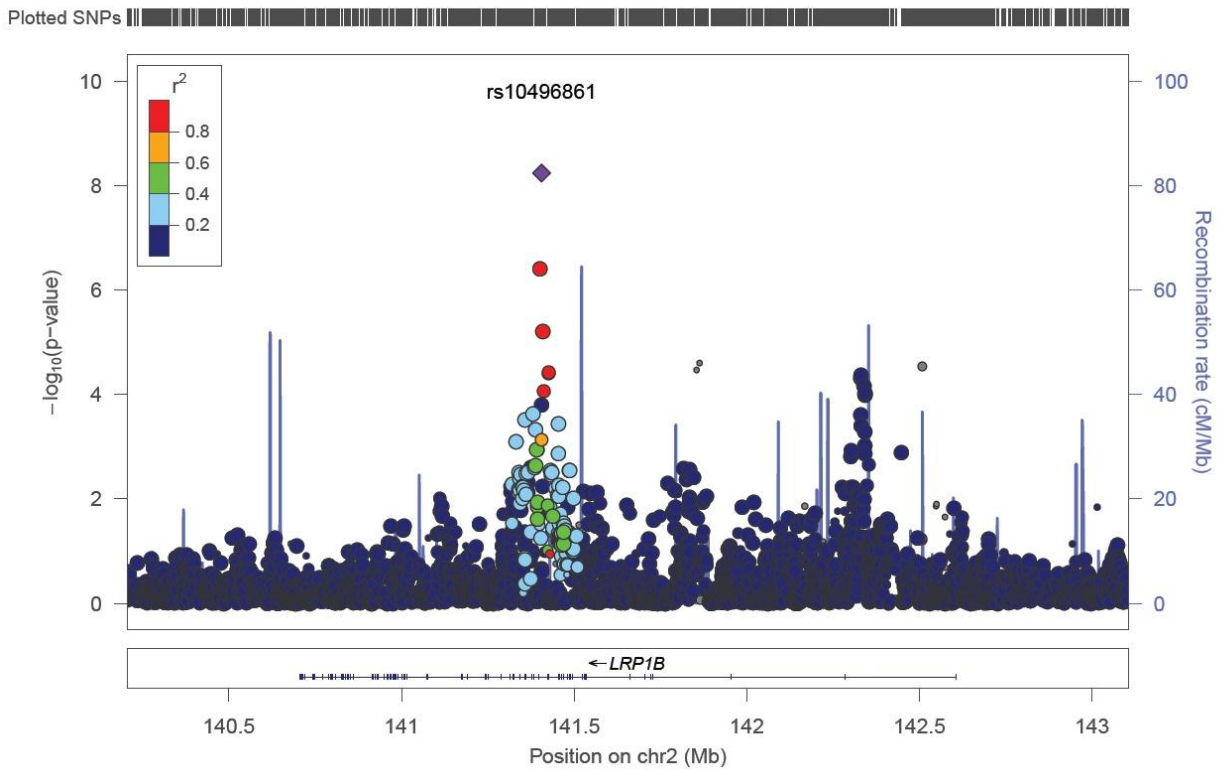


Figure 3.4: Locus Zoom Plot Chromosome 21, Positions 20.7 to 21.3 Mb, for Predicted Age Difference, Adjusted for Age, From Meta-Analysis, N=20,000 Individuals

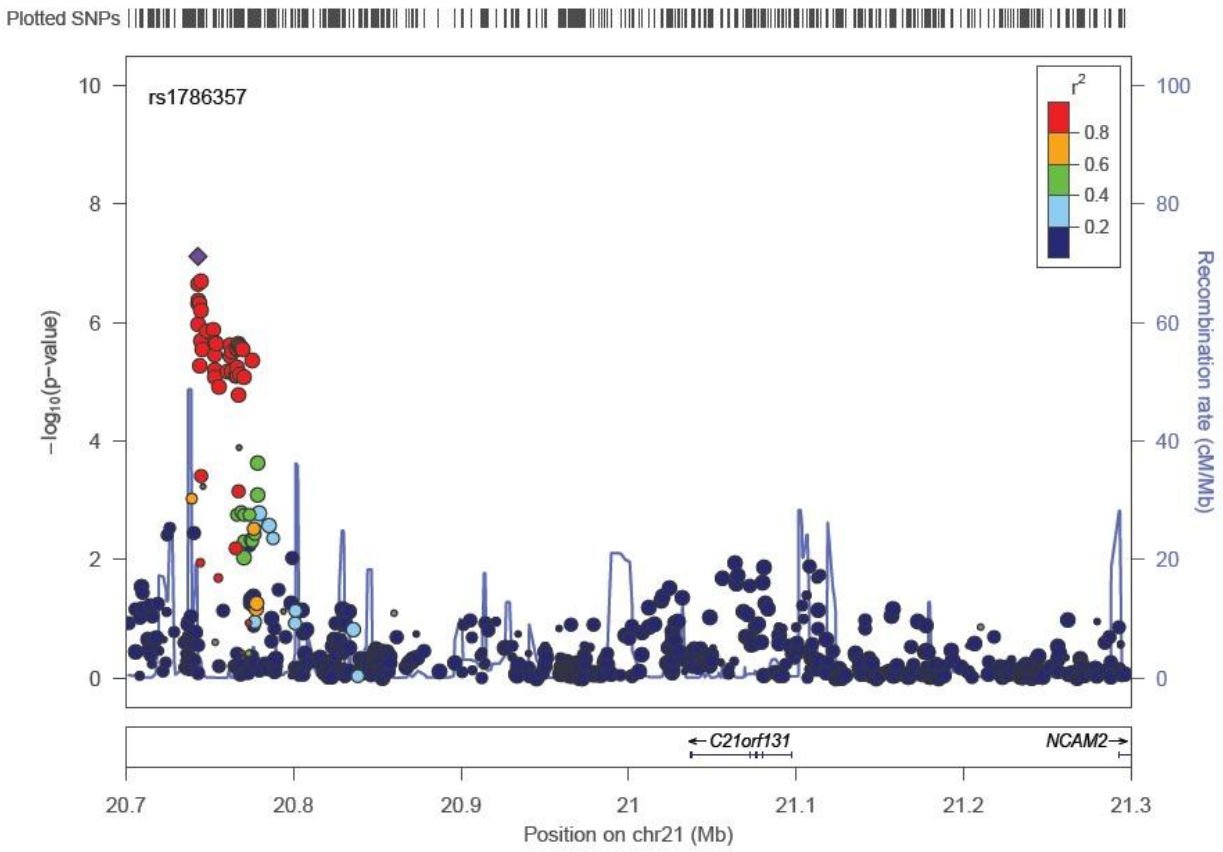


Figure 3.5: Locus Zoom plot of Chromosome 18, position 52.7 to 53.5 Mb for Difference between Risk Age and Age in 17,500 Individuals of European Decent, Adjusted for Age

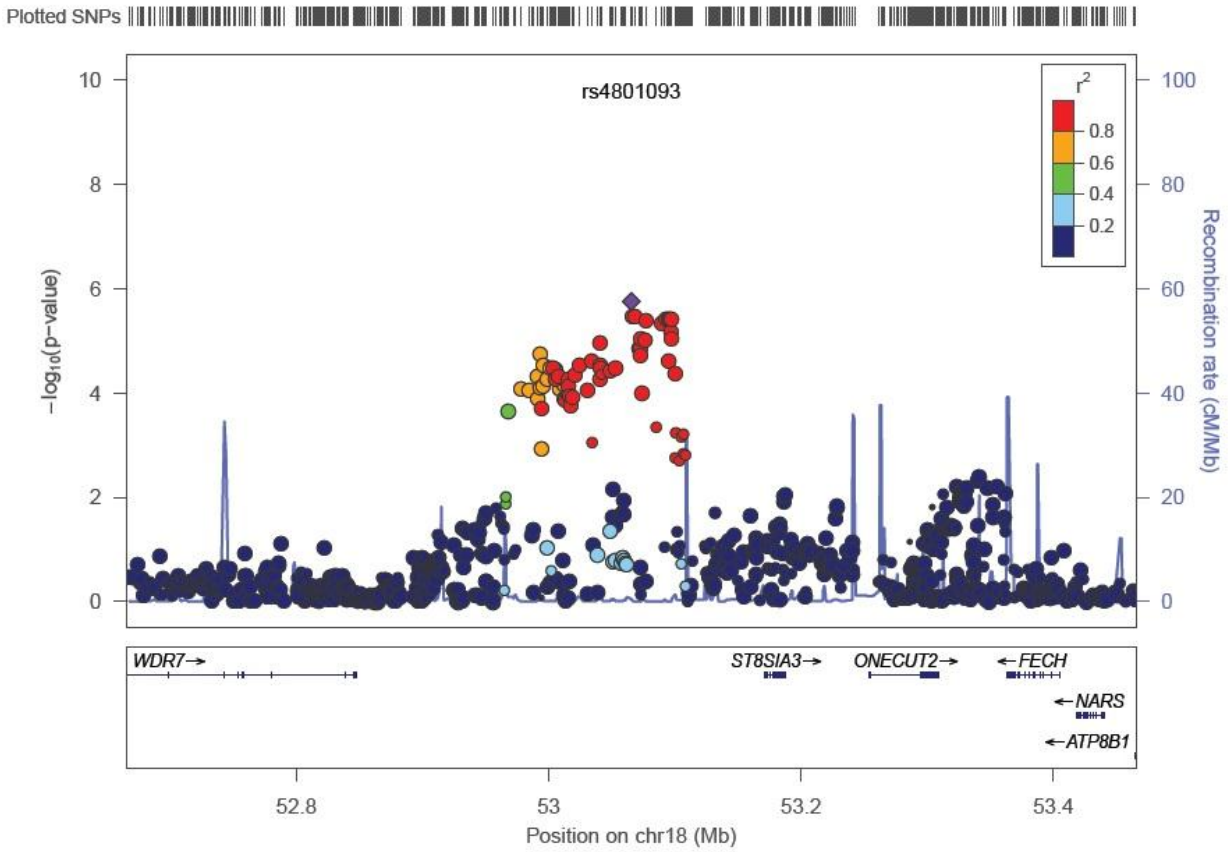


Figure 3.6: Locus Zoom plot of Chromosome 18, position 39.9 to 41.6 Mb for Difference between Risk Age and Age in 17,500 Individuals of European Decent, Adjusted for Age

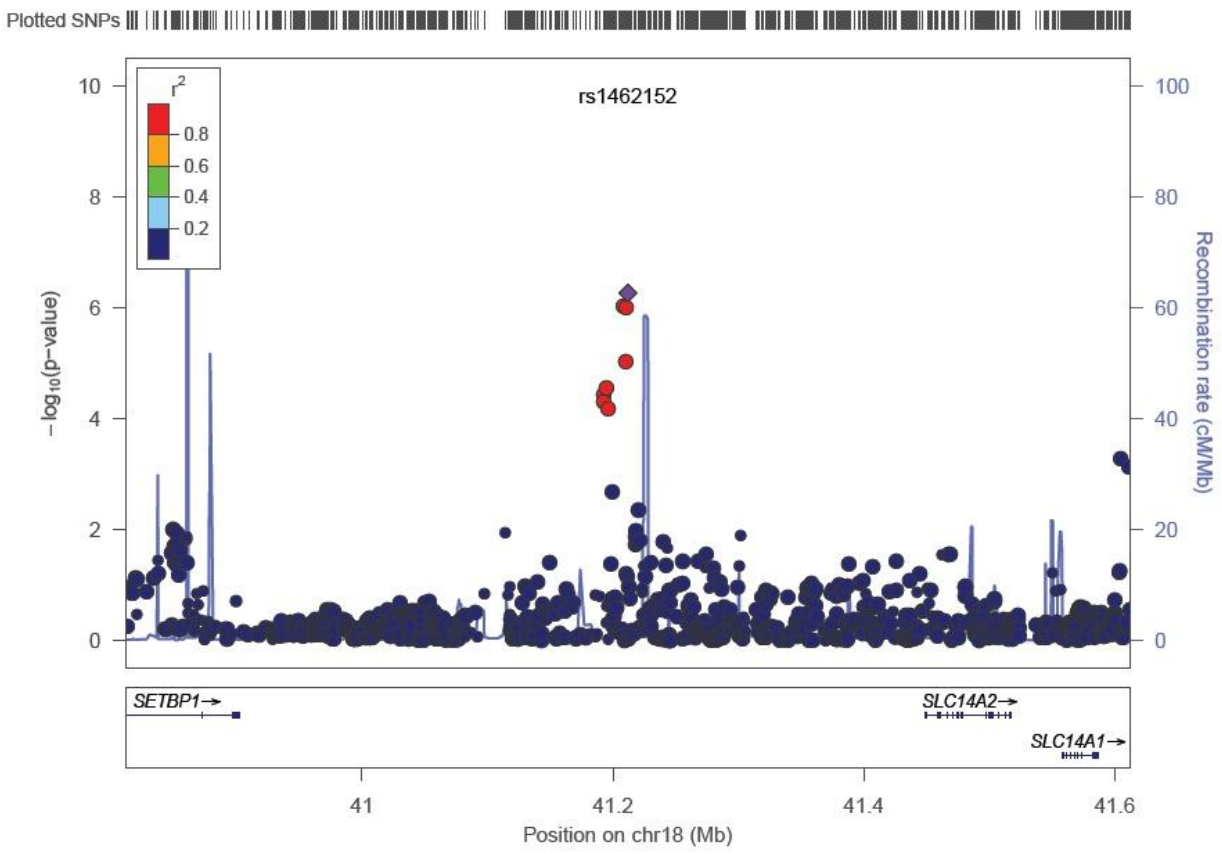


Figure S3.1: Manhattan Plot for Meta-Analysis of the Difference between Predicted Age and Age, Adjusted for Age, N=20,000 Individuals

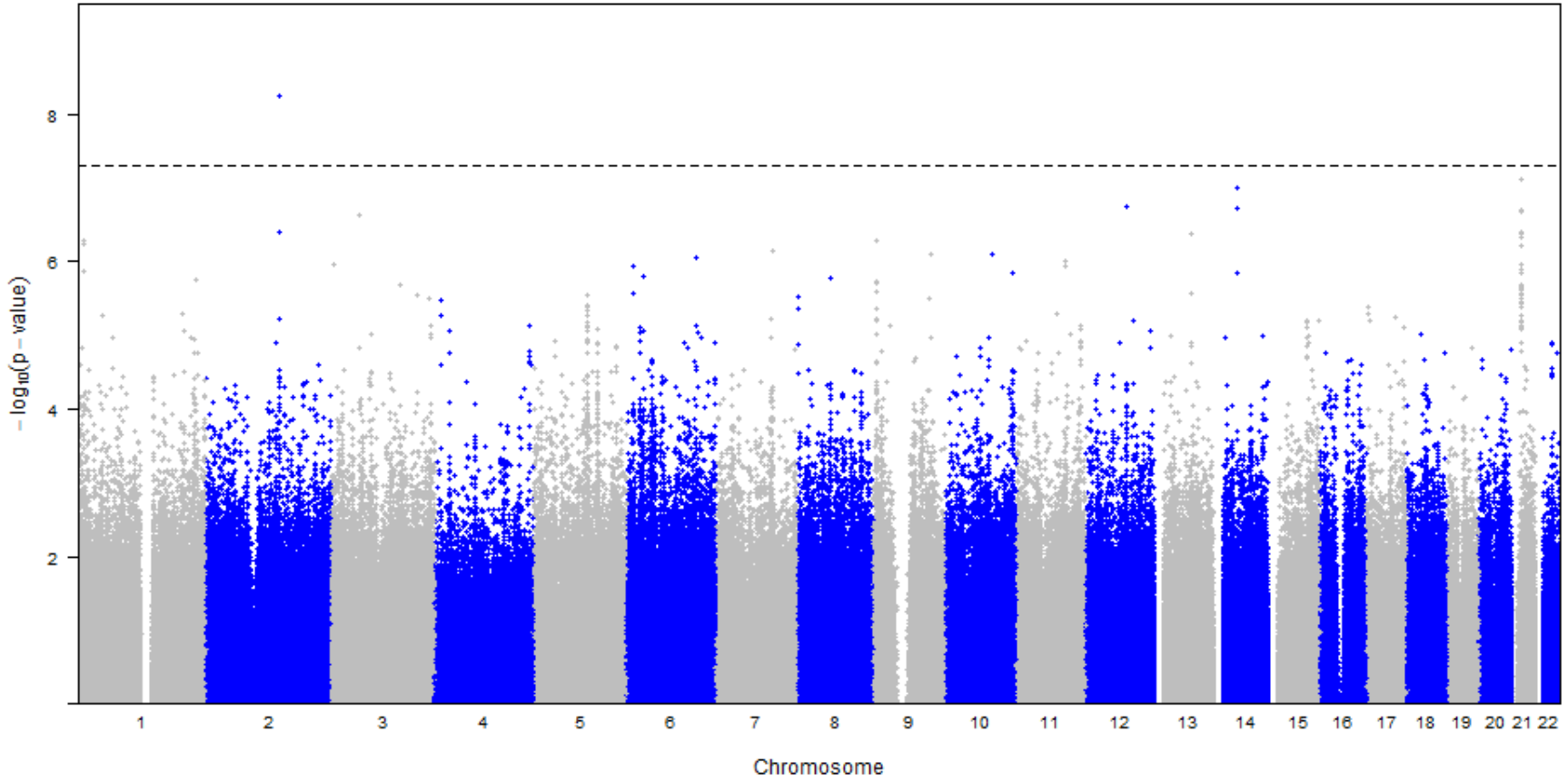


Figure S3.2: Manhattan Plot for Meta-Analysis of the Difference between Risk Age and Age, Adjusted for Age, N=17,500 Individuals

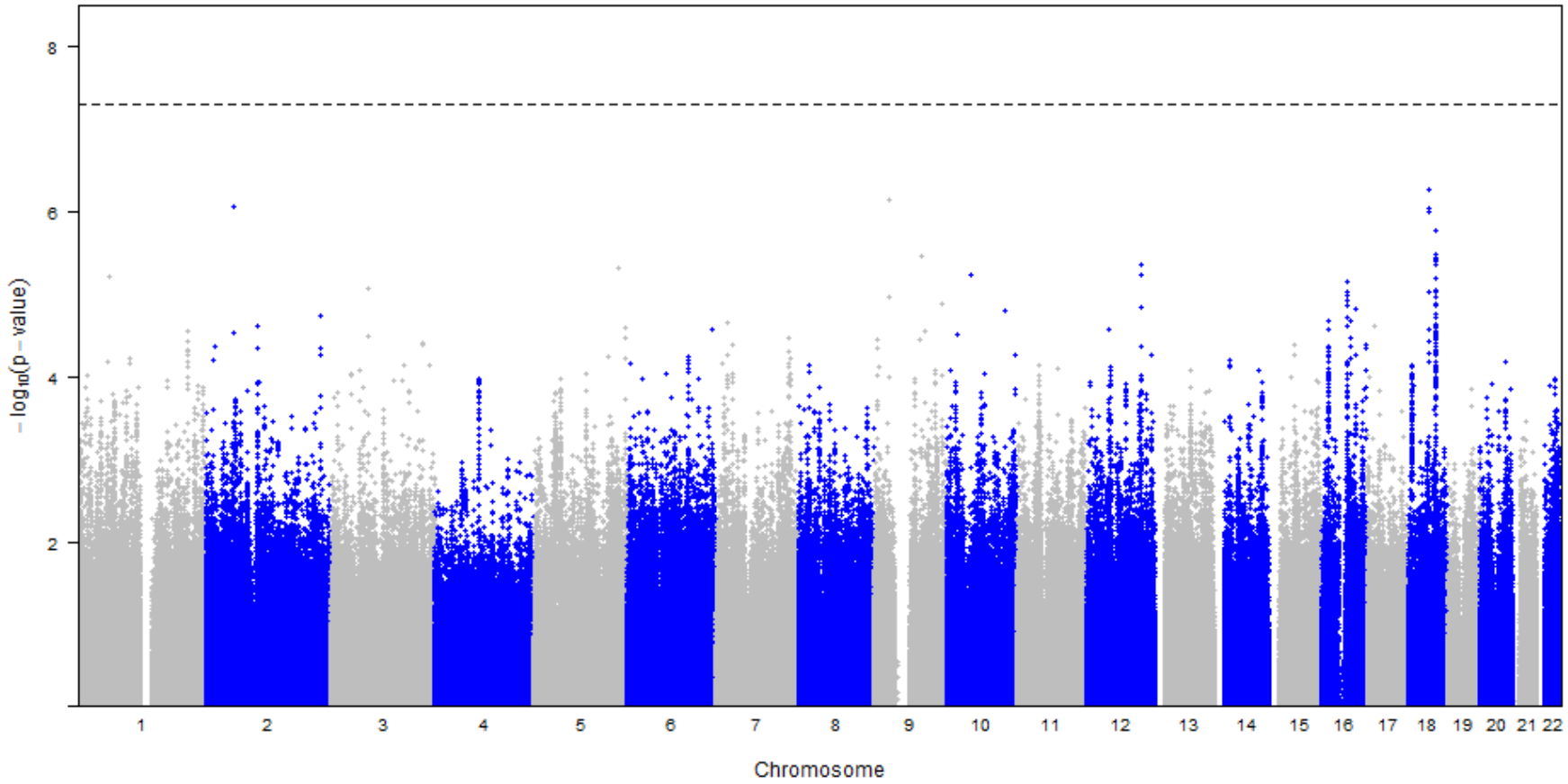


Figure S3.2: Comparison of rs10496861 across 46 Species

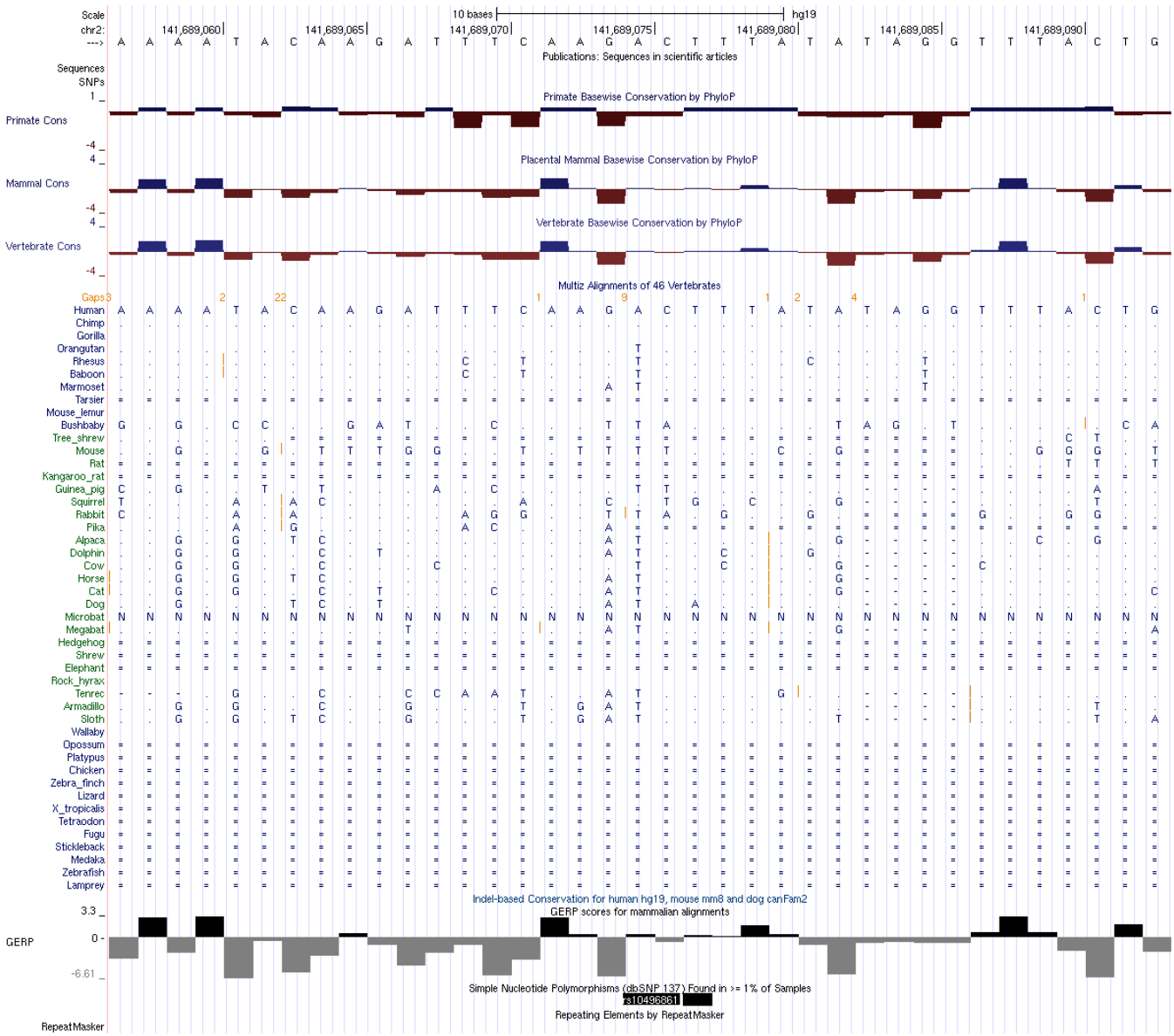


Table S3.1: Summary of Study Design and Sample QC, by Cohort

Study	Study design	Ethnicity	Total sample size for analysis	Sample QC	
				Call rate*	other exclusions
ARIC	Population Cohort	White European	8,718	>95%	Missing phenotypes
Framingham	Family-based	White European	3,018	>95%	Missing phenotypes
SardiNIA	Family-based	White European	4,774	>90%	Missing phenotypes
TwinGene	Twin-based	White European	9,999		Missing phenotypes
TwinsUK	Twin-based	White European	4,838	>98%	Missing phenotypes, heterozygosity across all SNPs ≥2 s.d. from the sample mean, evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations, observed pairwise IBD probabilities suggestive of sample identity errors

Table S3.2: Summary of Genotyping and Imputation Procedures, by Cohort

Study	Genotyping						Imputation			
	Platform	Genotype calling algorithm	Inclusion criteria			SNPs that met QC criteria	Imputation software	Inclusion criteria		Reference Population
			MAF	Call rate*	p for HWE			MAF	Imput quality*	
ARIC	Affymetrix 6.0	Birdseed	>0%	> 95%	> 10^{-6}	566,802	MACH/ Minimac	>5%	>0.30	HapMap2 (release 22, build 36, CEU)
Framingham	Affymetrix 500K and MIPS 50K	Dynamic Modeling	>0%	> 97%	> 10^{-6}	534,982	MACH	>0%	>0.30	HapMap2 (release 22, build 36, CEU)
SardiNIA	Affymetrix10K, Affymetrix 500K, Affymetrix 6.0	BRLMM for 10K and 500K arrays, Birdseed_ v2 for 6.0 array	≥ 0%	>90% for 10K and 500 arrays, > 95% for 6.0 array	> 10^{-6}	759,213	MACH	>0%	>0.30	HapMap2 (release 22, build 36, CEU)
TwinGene										
TwinsUK	Illumina HumanHap300 Bead Chip and Illumina HumanHap610 Quad Chip	Illuminus	≥1%	>97%	> 10^{-6}	~927000	IMPUTE (v2)	≥1%	>0.30	HapMap2, (release 22, combined CEU+YRI+ ASN panels)

CHAPTER IV

Prevalence of chronic kidney disease in the SardiNIA study cohort and its relationship with eGFR-related loci and clinical risk factors (CKD-SardiNIA Study)

4.1 Introduction

Universal concern about chronic kidney disease (CKD) and renal failure (National Kidney Foundation, 2002, Levey et.al, 2005, Levey et.al, 2007) has been expressed over geographical differences in CKD prevalence (Zhang, et.al, 2008, Zoccali, et.al, 2010). However, only a small (though increasing) number of large, adequately powered studies have estimated CKD prevalence in general populations. The "template" for such studies is the National Health and Nutrition Examination Surveys (NHANES) (Coresh, et.al, 2007), which showed an alarming increase of CKD prevalence during the last two decades in the USA. Other surveys in Europe (Nynke, 2011, Hallan et.al, 2011, Gambaro, et.al, 2010, Viktorsdottir, et.al, 2005, Wetzels, et.al, 2007, Cirillo, et.al, 2006) and developing countries (Zhang, et.al, 2008, Zhang, et.al, 2012, Nugent, et.al, 2011) showed a lower but variable prevalence. The differences can be partially

explained by differences in the prevalence of risk factors such as diabetes, hypertension, obesity, and atherosclerosis; but other factors (Levey and Coresh, 2012) can also influence prevalence. In addition to epidemiological and hormonal influences, specific genetic loci were recently demonstrated to be significantly associated with estimated glomerular filtration rate (eGFR) and with end stage renal disease (ESRD) in large genome-wide association studies of individuals of European descent (Böger, et.al, 2011, Köttgen, et.al, 2009).

General decline in eGFR is in fact a long-established feature of aging (Glassock and Winearls, 2008), but some groups of individuals may show a more rapid loss of renal function. How much clinical and genetic conditions influence decline of eGFR has been largely conjectural. Therefore, we investigated the clinical, epidemiological, and genetic factors that may influence both cross-sectional and longitudinal renal function in a well-powered Sardinian founder population cohort (SardiNIA study). In particular, given the recent success in finding and replicating genetic regions for eGFR, we investigation a genetic risk score for CKD in this population.

4.2 Methods

4.2.1 Study Design

The SardiNIA study, which began in 2001, has measured > 300 traits (endophenotypes, quantitative risk-related genetic, and environmental factors) that can be scored on a continuous scale for both epidemiological and genetic analyses. The sample was drawn from the 10,982 residents of 4 mountain villages (Lanusei, Arzana, Elini and Ilbono) in Ogliastra, a

province in eastern Sardinia (Pilia, et.al, 2006). In 2002, letters were sent to all residents inviting them to participate. About 56% (n=6,162) had initial visits. The recruited cohort, ranging from 14 to 102 years of age, is representative of the overall age distribution of the region in 2002, with a somewhat greater number of females. Visits were repeated approximately every three years and data from two additional visits have been collected, with 5,204 individuals completing a visit during the third to sixth year of the project, and 4,842 completing a visit during the seventh to ninth years. A total of 4,074 individuals had three visits, with a total average follow-up of seven years.

4.2.2 Screening and follow-up

Participants were interviewed during the first visit to collect detailed socio-demographic information, medical and family history, lifestyle, health behaviors (smoking, drinking, coffee intake, etc.), and medications taken. Anthropometric measures (height, weight, and waist circumference) and resting blood pressure were determined. Blood samples were collected by venipuncture after an overnight fast of at least 12 h at each visit. Urine specimens were only collected at the third visit, in 95% of the participants. Blood tests included serum creatinine, blood urea nitrogen [BUN], uric acid, glucose, hemoglobin A1c (HbA1c), and lipid levels. At the third visit, urine dipstick proteinuria and microalbuminuria were determined (Clark, et.al, 2011).

4.2.3 Definitions

Diabetics were defined according to the guidelines of the American Diabetes Association as individuals with either HbA1c ≥ 6.5 %, or fasting plasma glucose (no caloric intake for at least 8 h) ≥ 126 mg/dl (7.0 mmol/L), or on anti-diabetic therapy, or when they reported a diagnosis of diabetes. Blood pressure (BP) was measured using a calibrated desktop sphygmomanometer after at least 5 minutes of supine rest. BP was measured three times at intervals of at least 5 minutes, and the reported BP was the average of the last two measurements. Volunteers were classified as hypertensive when BP was ≥ 140 mmHg systolic or ≥ 90 mmHg diastolic, or when they reported taking antihypertensive medication. Obesity was defined as BMI (body mass index) ≥ 30 kg/m2, according to the World Health Organization's definition. Abdominal circumference was considered high when it was >94 cm for men and >80 cm for women. Metabolic syndrome was defined according to the International Diabetes Federation (IDF) guidelines (Alberti, et.al, 2006).

Participants whose albuminuria ranged from 3 to 30 mg/dl and whose proteinuria on a urinary spot test was < 30 mg/dl were classified as microalbuminuric, while individuals with proteinuria > 30 mg/dl at urinary spot test classified were classified as having macroalbuminuria. Cigarette smoking was defined as at least 10 cigarettes a day for a year. Previous cardiovascular (CV) events included coronary heart disease, heart attack, heart failure, or stroke, and were self-reported. Total cholesterol ≥ 200 mg/dl, triglycerides ≥ 130 mg/dl, LDL cholesterol (LDL-cholesterol)≥ 110 mg/dl, and uric acid serum levels ≥ 6 mg/dl (360 µmol/L) for

women and ≥ 7 mg/dl (420 µmol/L) for men were considered high. HDL cholesterol (HDL-C) < 40 mg/dl for men and < 50 mg/dl for women was considered low.

4.2.4 Genotype Data

Genome-wide markers assayed on a combination of Affymetrix platforms (500K and 1.0) (http://www.affymetrix.com) and imputed using HapMap2 samples as a reference, were used to calculate a genetic risk score from a list of 18 published loci found to be associated with CKD (Köttgen, et.al, 2012, Böger, et.al, 2011). Sixteen of the loci were available for analysis and ultimately 13 loci that were found to have the same direction of association with CKD in the SardiNIA sample were included: rs13538, rs347685, rs626277, rs881858, rs1731274, rs4744712, rs11959928, rs17319721, rs1260326, rs10109414, rs13038305, rs2467853, and rs12917707 (Table 3). For each locus, dosages were coded so that a value of 0 indicated the presence of no risk alleles, 1 indicated the presence of one risk allele, and 2 indicated the presence of two risk alleles. Dosages were added for the 13 loci to create a composite measure with a possible range of 0-26. More complicated genetic risk scores were explored, employing weights based on their strength of association with each outcome. These scores yielded smaller p-values in the association models, but for ease of being able to interpret the risk per 1 additional risk allele, the simple score is presented.

4.2.5 CKD classification and measures of kidney function

CKD was defined as eGFR <60 mL/min/1.73 m^2 and/or eGFR >60 ml/min/1.73 m^2 with kidney damage, and was staged according to the KDOQI classification (Levey, et.al, 2005). Kidney damage was quantified by albuminuria (micro or macro), and decreased kidney function was quantified by eGFR assessed by serum creatinine (sCr) concentrations (Levey, et.al, 2005). The Chronic Kidney Disease Epidemiology Collaboration (CKD-Epi) formula was employed in all association models because it is considered to be best to estimate GFR in general population-cohort studies (Wieneke, et.al, 2010). We also did estimation of GFR with MDRD 175 study equation to facilitate comparisons with the other surveys.

4.2.6 Calibration of serum creatinine

Measurements of sCr in the SardiNIA Laboratory (NIALab) were performed with a kinetic alkaline picrate assay at the first and third visits, but using different instruments, a Bayer Express Plus Chemistry Analyzer at first visit and a Biosystem A25 Chemistry analyzer at third visit. Calibration was done (Selvin, et.al, 2007) to estimate correctly the prevalence of kidney disease in the study cohort, to assess rates of change in kidney function, and to compare the data to other surveys. Calibration was carried out by assaying 109 randomly chosen, thawed samples from the first visit at the Central Laboratory of the Brotzu Hospital (CLB), Cagliari, Italy, where sCr measurements were performed with an Olympus analyzer (Olympus Mishima Co., Ltd., Shizuoka, Japan), using Jaffe's kinetic method. The creatinine concentration for the serum calibrator was traceable to the IDMS method.

A total of 63 randomly chosen specimens were also sent from the CLB to the NIA Lab for the measurement of sCr values using the same instrument that had been used for the third visit analysis. Extreme outliers (difference > 3 standard deviations, SDs, from the mean) were excluded because they would not contribute useful information to the calibration. Deming linear regression (Y = CCRL on X = original serum creatinine) was conducted for each survey to correct the regression models for measurement error (Selvin, et.al, 2007).

Two calibration equations were generated from the results and applied accordingly:

1) First/second visit: y (CLB) = -0.107+ 1.066*Creatinine NIALab

2) Third visit: y (CLB) = -0.195+ 1.0977*Creatinine NIALab

After standardization, we compared sCr values in subgroups of individuals in the same age range (20- 40 yr) and found that no statistically significant differences (Figure S4.1).

4.2.7 Statistical Analyses

Quantitative data are presented as the mean ± SD, and categorical data are presented as percentages. Differences between groups were examined using chi-square statistics for categorical variables.

The unadjusted odds ratios (OR) between risk factors and CKD were calculated using univariate logistic regression analysis, whereas the adjusted OR was calculated by multivariable logistic regression analysis, accounting for family membership by using generalized estimating equation methods. P-values less than 0.05 were considered significant. All parameters that

were significant ($p<0.10$) in univariate regression models were entered into a full multivariable model. In instances in which variables were known to be strongly correlated to one another (e.g., glucose and diabetes) only the one with the strongest association was included in the final model. Final multivariable models included predictors that were significant at the $p<0.10$ level. Because it is possible that variable that were not significant in the univariate models could become significant predictors in the multivariable models (Simpson's Paradox), we examined all combinations of multivariable models and this was not found to be the case. We evaluated both the continuous and categorical variables, and since results were very similar, the categorical variables are shown for ease of interpretation.

4.2.8 Definition of Decline in eGFR

Changes in eGFR during the study were assessed in individuals for whom measurements at all three visits were available. Linear regression was employed to determine the slope for each individual. These models imposed a linear rate of decline on each individual, which was a close estimation for most individuals (Figure S4.1) and allowed for a single value estimate for use in later association models. Linear mixed models, accounting for family membership as a repeated variable with compound symmetry covariance, were used to examine the association between known risk factors (including a genetic risk score) and change in eGFR (slope).

Individuals were defined as having a "fast decline" in eGFR if their slope was steeper than -13.18 ml/min over the 6 years of follow-up, as this value indicated that these individuals had a decline greater than one SD below the mean. Clinical and genetic risk factors were

examined by logistic regression to determine their association with classification as “fast decline”. All analyses were performed using SAS 9.3 software (SAS, Cary NC).

4.2.9 Methodological considerations for cross study comparisons

To date, our cohort is one of the largest European cohorts to have been investigated using the CKD-Epi eGFR formula. For the first time, a genetic risk score for CKD was tested in a general population. The SardiNIA study cohort differs in ethnicity from the NHANES, HUNT II, Beijing, and INCIPE cohorts. Like the HUNT II and NHANES studies, we enrolled a large, representative sample of pedigrees from the general population and achieved high participation and completion rates. In order to optimize the comparison of our cohort with the HUNT II and NHANES populations we used the same eGFR formulas and stratified the data by age. Therefore, prevalence rates of CKD can be compared, although the very low prevalence of CKD stage 4 and 5 described here may be biased by an underestimation of sicker individuals, whose participation in the study tended to be curtailed.

Calibration of sCr assays included a correction factor similar to that employed in the NHANES and HUNT II analyses (Selvin, et.al, 2007, Hallan, et.al, 2006). With this correction, the MDRD 175 formula and the CKD-Epi formula can be considered unbiased in the cohort. The precision of the CKD-Epi equation is limited as compared to measured GFR, but the formula corrects the bias of previous formulas in the classification of normal and mildly decreased eGFR groups (Skali, et.al, 2011). Using the MDRD study equation we avoided biases comparing our eGFR results versus the results of the other cohorts or populations.

The SardiNIA project was not designed as an epidemiological study of CKD, and in addition to the lack of determinations of micro-macroalbuminuria and ACR ratio in the first visit (see above), quantitative determination of microalbuminuria was evaluated only qualitatively by dipstick (albustick), and only once. Consequently, the prevalence of micro-macro-albuminuria and stages 1 and 2 CKD may have been overestimated. In order to reduce any overestimation of albuminuria, which is the strongest risk factor for CKD progression and thus for the development of cardiovascular disease(Sarnak, et.al, 2003), subjects were defined as being affected by microalbuminuria when values were greater than 3 mg/dl, in both genders. Encouragingly, the prevalence of persistent microalbuminuria was also higher than macroalbuminuria in CKD stages 1 and 2 in the NHANES study (Coresh, et.al, 2007).

4.3 Results

4.3.1 General Demographic Characteristics of the SardiNIA study cohort

This cohort was representative of the regional population in Ogliastra with regard to mean age (43.7, SD 17.6), gender (58% female), and age groups, as per the National Census data shown in Figure S4.2 (Pilia, et.al, 2006). At the third visit, after a median of seven years of follow up, 4,842 individuals were examined. As shown in Table 4.1, from the first to the third visit, the prevalence of all traditional risk factors increased.

4.3.2 Longitudinal renal function evaluation

As expected from the increasing age of the cohort, mean eGFR was lower at the third visit in the whole population: 98.6 ml/min (SE 19.3) compared to the first visit 104.6 ml/min (SE 20.7, p= 0.001). Therefore individuals in the normal renal function group decreased while individuals in the mild, moderate, and severe groups increased (Figure 4.1). The average decline in renal function evaluated by eGFR during the follow-up period was approximately 1 ml/min/year among the 4,074 individuals that completed all three visits (Figure 4.2).

4.3.3 Prevalence of albuminuria and CKD

We evaluated the prevalence of CKD at the third visit. The proportions of individuals with micro- and macro-albuminuria were 9.5 % and 3.4%, respectively. The overall estimate of CKD stages 1 to 5 was 14.5%. The specific prevalence of CKD for each stage was: 7.4% for CKD stage 1, 4.1% for CKD stage 2, 2.9% for CKD stage 3, 0.07% for CKD stage 4, and 0.05 for CKD stage 5 (Figure 4.1). The overall prevalence of CKD among men was 12.9% whereas it was 15.4% among women (Table S4.1). The general trend toward higher prevalence of CKD stages 1 and 2 over time differed among age categories. Consistent with expected aging trends, the proportion of individuals in stage 1 was higher in the younger age group, i.e., below 30 years of age (> 10%), compared to 7.0% in the 30-39 year-old group and 5% in the 40-49 year-old group (Table S4.1). Total prevalence of CKD increased among age categories from 12.7% in the 50-59 year-old group, to 16.3% in the 60-69 group, to 36.5% in the 70-79 group, to 48.1% in the over eighty group. The highest prevalence of CKD was observed in individuals affected by diabetes (35.3%),

hypertension (27.9%), high BMI (22.8%), hyperuricemia (28.4%), and metabolic syndrome (27.9%). The lowest CKD prevalence, i.e., 8.5%, was observed in individuals who had none of the risk factors listed above (Table S4.1).

4.3.4 Comparison of SardiNIA study results with NHANES (1988-1994 and 1999-2004), HUNT II and Beijing study results

To compare our population to the other populations we used CKD prevalence data obtained by the MDRD 175 formula stratified by age. SardiNIA's CKD prevalence was about the same as NHANES (16.1 vs 16.5) and Beijing (12.4 vs 11.2), higher than in the HUNT II (16.1 vs 11.2), and lower than NHANES 1999-2004 (16.8 vs 20.3), although these were next tested for statistically significant differences (Table 4.2).

Similar results can are seen by taking into account kidney function, grouped as normal, mildly reduced, moderately reduced, and severely reduced, and stratifying by age. The Sardinian cohort included significantly more individuals with normal renal function than Americans (54.4% vs 51.9% in NHANES 1988-1994 and 52.7% vs 40.7% in NHANES 1999-2004) and fewer than in the HUNT II (52.6% vs 56.7%) and Beijing (56.3% vs 64.7%) studies. The opposite trend was evident for mildly reduced and moderately reduced eGFR, and no significant differences were observed for severely reduced eGFR (Table 4.2).

4.3.5 Risk factors associated with CKD

In univariate analysis, we evaluated 15 variables and found that all of them were significantly associated with the presence of CKD, except for smoking, major lipid profile, and cortical thickness. In the final reduced multivariable model, older age (per 10 years, OR= 1.31), female gender (OR= 1.28), diabetes (OR=1.48), and genetic risk score (per 1 risk allele, OR = 1.07) were independently associated with CKD. High uric acid (OR= 1.28, p= 0.06) and abnormal kidney length (OR= 1.26, p= 0.06) showed an association trend with CKD (Table 4.3).

4.3.6 Risk factors associated with change in eGFR

In univariate analysis, only smoking, low HDL-C, and cortical thickness were not significantly associated with decline in eGFR. In the final reduced multivariable model, baseline eGFR (per 10 ml/min, - 0.52 ml/min), age (-3,5 ml/ min), male gender (1.23 ml/ min), diabetes (-3.13 ml/ min), hypertension (-1.69 ml/ min), high uric acid (-1.36 ml/ min), and genetic risk score (per 1 risk allele, - 0.23 ml/min) were associated with a change in eGFR (data are expressed in ml/ min per 10 years) (Table 4.4).

In univariate analysis of the dichotomous outcome of fast decline of eGFR (-2.3 ml/min/year) and the predictors listed above as the independent variables, only smoking, lipid profile and cortical thickness were not significant in the multivariable model, whereas previous cardiac disease only showed marginally significant positive trend (p=0.06). In the final reduced multivariable model, age (per 10 years, OR= 1.67), female gender (OR=1.39), hypertension (OR=1.58), high uric acid (OR=1.97) and genetic risk score (per 1 risk allele, OR= 1.05) were

significantly independently associated with a faster decline of eGFR, whereas diabetes (p=0.11) was not (Table 4.4).

4.4 Discussion

Assessed for the first time in a European founder population, CKD prevalence and geographical variability have been confirmed. It is likely the rapid increase of traditional risk factors has contributed to the high prevalence of early stages of CKD. We also present evidence for additional clinical risk factors, especially hyperuricemia and ultrasound renal length, but not renal cortical thickness. This result counters to what has been reported in a small group of patients with advanced stages of CKD and GFR estimated by Cockcroft-Gault and MDRD 186 formulas (Beland, et.al, 2010). Instead, in our work, these two parameters have been tested for the first time as risk factors in such large sample of pedigrees from a general population versus early stages of CKD and with GFR estimated by CKD-Epi formula. For the first time we have also shown that a genetic renal risk score supplies an independent risk factor for CKD and fast eGFR decline.

The cross-sectional CKD prevalence was high (14.6%) in the SardiNIA cohort, although the population was relatively young (mean age 49.8 ys). Early stages of CKD were the most represented (CKD 1 + 2: 11.3%), which corresponds to the high prevalence of proteinuria (12.8% in the whole cohort). The prevalence estimate might be reduced if we had used ACR to detect proteinuria and did urinalysis more than once; but it seems more likely that the high level of early CKD stages reflected an increasing level of risk factors (i.e., diabetes, obesity) as a

more American lifestyle/diet has been adopted, mainly in the last 10 years (Table 4.1). Furthermore, based on the level of early CKD, the percentage of individuals with eGFR < 60 ml/min is expected to rise in the future (and may already be underestimated because patients affected by chronic disease tend to participate less in repeated follow-up visits). Further understanding of the interactions between obesity, metabolic syndrome and CKD could represent a potential strategy to reduce end stage renal disease (ESRD) in the future.

In addition to using the CKD-Epi formula, as recommended (Skali, et.al, 2011), we also applied the MDRD 175 formula to compare SardiNIA to other large cross-sectional CKD studies. Accepting the first-order accuracy of overall estimates, all the populations compared showed a greater CKD prevalence in females (15.4%) than in males (12.9%). Data stratified by age showed CKD prevalence in the SardiNIA study similar to to the comparably rural population in Beijing – 2008 (12.4% vs 11.2%) and to NHANES 1988-1994 (16.1% vs 16.5%), but lower than NHANES 1999-2004 (16.8% vs 20.3%) and higher than HUNT II (16.1% vs 11.2%) (Table 4.2). These differences likely result from a combination of factors. Looking at general risk factors, for example, the results are consistent with the higher and increasing prevalence of diabetes and metabolic syndrome in the NHANES American population; however, Norwegians have by far the highest prevalence of hypertension, often a major cause of ESRD, but had the lowest prevalence of CKD. Differences across the world may also depend on additional parameters that have not yet been assessed, including environmental and hormonal factors, genetic variation (see below), and even differences in health policy, particularly screening programs that can differentially affect awareness, treatment, and control (Zoccali, et.al, 2010).

Longitudinal analysis of data in the 4,074 individuals who performed all 3 visits showed an overall reduction in mean eGFR. In a mean of 7 years of follow-up, individuals with normal renal function decreased, whilst there was a consistent increase in the prevalence of mildly and moderately reduced eGFR. This was partially expected as a result of aging (Prakash, et.al, 2009) of the cohort with the increased prevalence of clinical risk factors. In particular, diabetes, a major cause of ESRD in developed countries, showed an increase of almost 100%. It was an independent risk factor for CKD and its presence was associated with a significant additional change in eGFR (-3.13 ml/m'/yr). However it was not a predictor of fast eGFR decline, probably because of the high prevalence of CKD stage 1; in early stages of diabetic nephropathy glomerular hyperfiltration is observed, resulting in a misleading "improvement" of renal function.

Also, Obesity had an increasingly high prevalence (18.2%), especially compared to the Italian mainland population (8-10%) (Eurostat, 2011). As suggested by Zoccali et al, obesity can amplify the cost of CKD, hypertension, diabetes, and cardiovascular disease consuming a large fraction of healthcare resources (Zocalli, et.al, 2010). However, its correlation with CKD did not remain significant in multivariable analysis. Hypertension, the other main cause of ESRD in developed countries. Hypertension increased in prevalence weakly and was associated with fast eGFR decline. However hypertension was not an independently significant risk factor for CKD. This is again most likely because CKD stages 1-2, the most prevalent in SardiNIA, are characteristically associated wth proteinuric conditions like diabetes and obesity, whilst hypertensive nephroangiosclerosis is associated with more advanced renal damage. Moreover, in contrast to what was observed in Africans, Caucasians show greater response to ACE

inhibitors (Materson, et.al, 1993), and better control of hypertension might have weakened its correlation to CKD.

Ultrasound renal length, according to other studies and in our experience, is strongly associated with CKD when advanced renal damage is present (Sanusi, et.al, 2009). Using a restrictive cut-off for renal length (< 10 cm) we observed a correlation with CKD and eGFR decline. Hyperuricemia has not been extensively assessed in published surveys, but the prevalence was high in SardiNIA. The correlation of uric acid with CKD and with fast eGFR decline was significant, suggesting it as a risk factor. Experimental and clinical studies have indicated that elevated uric acid can itself lead to kidney disease without deposition of uric acid crystals. Glomerulosclerosis, interstitial fibrosis and arteriolar disease without intrarenal urate crystals are the principal lesions in rats with elevated uric acid levels (Kang, et.al, 2002, Mazzali, et.al, 2001). Uric acid may thus be implicated in renal disease in human, and lowering its blood levels may slow disease progression, especially in patients with hyperuricemia (Mazzali, et.al, 2001, Siu, et.al, 2006, Kanbay, et.al, 2007). It remains to be further documented whether uric acid is a general independent biomarker of early renal damage and possible prognosis of progression (Feig, et.al, 2008).

For the first time in a study on CKD prevalence and on the longitudinal renal function conducted in a general population, we have further tested a risk score to assess the relevance of genetic factors. A multivariable model inclusive of traditional risk factors and other measures associated with CKD showed that the genetic renal risk score, based on 13 published loci associated with CKD, was independently associated with outcomes. An individual with one

additional risk allele had a 7% higher odds of having CKD, a greater decline in eGFR (-0.23, p= 0.004), and 5% increased odds of fast eGFR decline. Although these estimates are relatively modest for 1 additional risk allele, when one takes into account the range of data in the cohort (6-24), risk increases are quite substantial: 337% higher odds of CKD, -4.14 greater decline in eGFR over 10 years, and 240% higher odds of "fast decline". Further work is required to cross-compare the reproducibility of results in Sardinia and in other ethnicities, and to refine predictive models for clinical use based in effect sizes of individual genetic loci.

4.5 Acknowledgements

I would like to acknowledge Dr. Antonello Pani and an equal contributor of this work. His nephrology expertise and knowledge was very important on putting the results into a context for the renal community.

Table 4.1: Demographics and CKD Risk Factors for the SardiNIA Study, by Wave *

Variable	Total Sample				Individuals in All Three Visits (n=4,074)			
	Wave 1 (n=6,162)	Wave 2 (n=5,204)	Wave 3 (n=4,842)	Trend Test or X^2 p-value	Wave 1	Wave 2	Wave 3	Trend Test or X^2 p-value
Categorical Measures:								
Age (years, %):								
< 20	8.8	3.6	0.7		7.04	3.04	0.00	
20 – 39	37.0	34.3	30.6		37.78	34.00	28.62	
40 – 59	33.4	37.3	40.1	< 0.0001	37.33	39.74	41.26	< 0.0001
60 – 69	12.4	13.8	15.4		12.86	14.58	16.47	
70 +	8.4	11.0	13.2		4.98	8.64	13.65	
Male (%)	42.6	41.8	42.3	0.64	41.5	41.5	41.5	-
Smoking (%)	17.0	21.4	23.7	< 0.0001	17.7	21.9	24.3	< 0.0001
Metabolic syndrome (%)	6.3	6.8	13.1	< 0.0001	7.6	9.6	19.3	< 0.0001
Obesity (%)	15.7	16.7	18.1	< 0.0001	14.8	16.3	18.6	< 0.0001
Large waist (%)	39.6	43.5	49.2	< 0.0001	39.3	43.2	50.1	< 0.0001
Diabetes (%)	5.2	7.4	9.1	<0.0001	3.7	6.4	9.2	< 0.0001
High glucose (%)	3.8	4.4	6.2	<0.0001	2.7	3.5	6.3	< 0.0001
Hypertension (%)	29.9	31.8	32.0	<0.0001	28.1	31.1	34.4	< 0.0001
High Blood Pressure (%)	25.5	24.9	22.1	0.22	24.2	24.9	24.5	0.73
Previous cardiac disease (%)	4.2	4.7	5.6	< 0.0001	3.56	4.22	5.94	< 0.0001
High uric acid (%)	13.2	15.0	16.9	<0.0001	12.0	14.5	17.1	< 0.0001
High total cholesterol (%)	56.2	55.7	63.7	< 0.0001	58.1	57.3	65.0	< 0.0001
High LDL (%)	66.2	67.2	76.3	< 0.0001	67.6	68.5	78.0	< 0.0001
Low HDL (%)	7.0	6.1	19.3	< 0.0001	6.0	5.7	20.1	< 0.0001
Continuous Measures, mean (sd):								
Age (years)	43.7 (17.6)	46.9 (16.8)	49.7 (16.3)	< 0.0001	43.3 (15.8)	46.7 (15.8)	50.7 (15.8)	< 0.0001
BMI	25.3 (4.7)	25.6 (4.6)	25.9 (4.7)	< 0.0001	25.4 (4.5)	25.6 (4.5)	26.1 (4.7)	< 0.0001
Waist circumference (cm)	84.8 (13.2)	85.6 (12.5)	87.2 (12.3)	< 0.0001	84.5 (12.8)	85.4 (12.4)	87.3 (12.2)	< 0.0001
Glucose (mg/dl)	90.1 (23.7)	90.8 (23.3)	98.3 (25.0)	< 0.0001	88.7 (20.7)	89.9 (21.7)	98.8 (24.6)	< 0.0001
Sys Blood pressure (mm/Hg)	125.6 (18.6)	124.5 (17.9)	119.4 (26.0)	0.23	124.9 (17.6)	124.6 (17.4)	125.4 (18.9)	0.26
Uric Acid ()	4.3 (1.5)	4.5 (1.5)	4.6 (1.4)	< 0.0001	4.2 (1.4)	4.4 (1.5)	4.6 (1.4)	< 0.0001
Total Cholesterol (mg/dl)	208.5 (42.2)	206.9 (38.9)	215.3 (40.3)	< 0.0001	210.0 41.0	208.0 (38.7)	216.6 (40.0)	< 0.0001
LDL Cholesterol (mg/dl)	126.8 (35.4)	125.7 (32.7)	135.5 (34.6)	< 0.0001	127.8 34.7	126.8 (32.5)	136.9 (34.3)	< 0.0001
HDL Cholesterol (mg/dl)	64.1 (14.9)	63.0 (13.7)	56.8 (14.1)	< 0.0001	64.7 14.8	63.3 (13.5)	56.8 (14.2)	< 0.0001
Triglycerides (mg/dl)	88.1 (68.2)	91.4 (70.7)	115.3 (69.3)	< 0.0001	87.7 69.7	89.9 (66.8)	114.9 (69.8)	< 0.0001

Kidney Measures:					
Reduced Length (< 10 cm)	-	-	18.1	-	
Reduced Cortical Thick (< 10mm)	-	-	2.5	-	
Microalbuminuria (%)	-	-	9.5	-	
Macroalbuminuria (%)	-	-	3.4	-	

*Trend tests account for family clustering

Table 4.2: Prevalence of CKD stages and of kidney function categories, stratified by age in SardiNIA study cohort individuals aged 14 years or older based on third visit (SardiNIA 3) and compared to NHANES 1988- 1994 and 1999- 2004, Beijing, and HUNT II populations (glomerular filtration rate estimated by MDRD 175).

		Prevalence of CKD (95% CI)				
		CKD 1	**CKD 2**	**CKD 3**	**CKD 4**	**TOT**
Not stratified	SardiNIA 3 (4,477)	5.5 (4.9- 6.2)	5.5 (4.9- 6.2)	4.0 (3.5- 4.7)	0.1 (0.0- 0.3)	15.2 (14.2- 16.3)
Stratified	SardiNIA 3	5.6 (4.9- 6.4)	5.7 (5.0-6.5)	4.6 (3.9 – 5.4)	0.1 (0.0- 0.3)	16.1 (14.9 – 17.4)
	NHANES 1988-1994 (14,319)	4.2 (3.9- 4.5)	4.8 (4.4- 5.2)	7.1 (6.7- 7.5)	0.3 (0.2- 0.4)	16.5 (15.8- 17.1)
	SardiNIA 3	5.6 (4.9 – 6.4)	6.0 (5.3 – 6.9)	5.0 (4.3 -5.8)	0.1 (0 – 0.3)	16.8 (15.5 – 18.1)
	NHANES 1999- 2004 (12,216)	4.0 (3.6- 4.3)	5.5 (5.1- 5.9)	10.2 (9.6- 10.8)	0.6 (0.5- 0.8)	20.3 (19.5- 21.1)
	SardiNIA 3	5.5 (4.8- 6.3)	5.9 (5.1 – 6.6)	4.6 (4.0 – 5.4)	0.1 (0.0–0.3)	16.1 (14.9 – 17.4)
	HUNT II (65,181)	3.1 (3.0- 3.2)	3.4 (3.3- 3.5)	4.5 (4.3- 4.7)	0.16 (0.13- 0.19)	11.2 (10.9- 11.5)
	SardiNIA 3	5.5 (4.8- 6.3)	4.5 (3.9- 5.2)	2.4 (2.0- 2.8)	0.1 (0.0- 0.2)	12.4 (11.4- 13.4)
	Beijing(13,925)	5.5 (5.2- 6.0)	3.8 (3.5- 4.2)	1.7 (1.5- 2.0)	0.1 (0.1- 0.2)	11.2 (10.7- 11.8)
		Kidney function (MDRD eGFR), mL/min/1.73 m2				
		Normal (≥ 90) %	**Mildly reduced (60- 90) %**	**Moderately reduced (30- 60) %**	**Severely reduced (15- 30) %**	
Not stratified	SardiNIA 3	53.4 (51.3- 55.5)	42.6 (40.7- 44.5)	4.0 (3.4- 4.6)	0.1 (0.0- 0.2)	
Stratified	SardiNIA 3 (4,731)	54.4 (52.2- 56.7)	41.0 (39.1- 42.9)	4.5 (3.8- 5.2)	0.2 (0.1- 0.4)	
	NHANES 1988-1994 (15,488)	51.9 (50.7 – 53.1)	42.4 (41.3- 43.5)	5.4 (5.0- 5.8)	0.21 (0.14- 0.29)	
	SardiNIA 3	52.7 (50.6- 54.9)	42.3 (40.4- 44.2)	4.9 (4.2- 5.7)	0.2 (0.1- 0.4)	
	NHANES 1999- 2004 (13,233)	40.7 (39.6- 41.8)	51.2 (50.0- 52.5)	7.7 (7.2- 8.2)	0.35 (0.25- 0.46)	
	SardiNIA 3	52.6 (50.6- 54.7)	42.8 (40.9- 44.7)	4.5 (3.8- 5.2)	0.2 (0.1- 0.3)	
	HUNT II (65,181)	56.7 (56.1- 57.3)	38.6 (38.1- 39.1)	4.5 (4.3- 4.7)	0.16 (0.13- 0.19)	
	SardiNIA 3	56.3 (54.6- 59.1)	40.9 (39.0- 42.8)	2.3 (1.9- 2.7)	0.1 (0.0- 0.2)	
	Beijing	64.7 (63.4- 66.1)	33.4 (32.4-34.4)	1.8 (1.5-2.0)	0.1 (0.1- 0.2)	

Figure 4.1: Prevalence of Chronic Kidney Disease in SardiNIA defined by eGFR alone and by NHANES, by wave

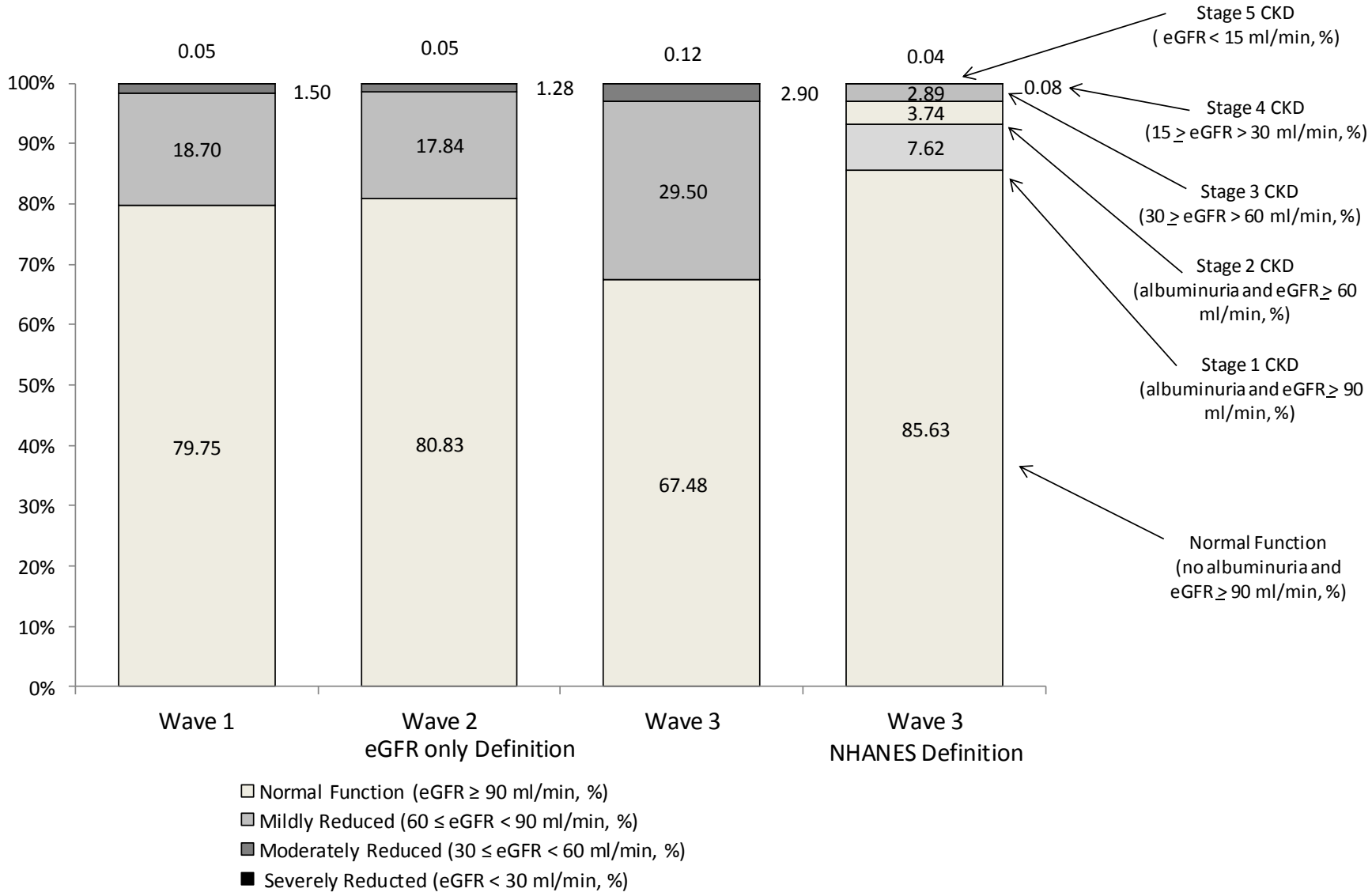


Figure 4.2: Histogram of Change in eGFR

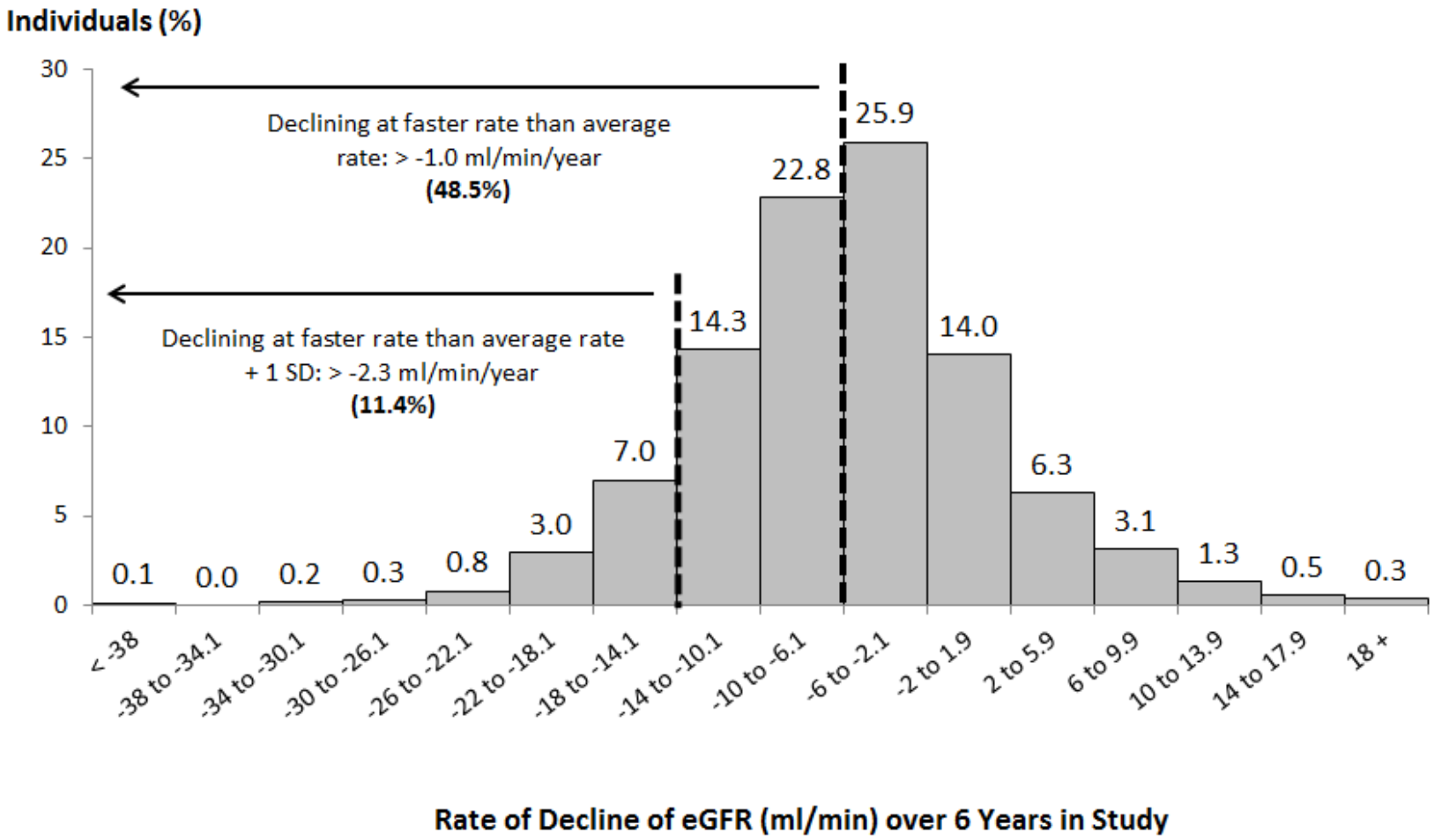


Table 4.3: Previously Identified Genetic Loci used in SardiNIA Risk Score for eGFR, with Univariate Beta Estimate from SardiNIA Data.

Marker	Chr	Position	Risk Allele	Gene	SardiNIA β	CKD-Gen β	Comments
Identified in Köttgen, 2009							
rs12917707	16	20,275,191	G	UMOD	-0.3701	-1.02	Familial juvenile hyperuricemic nephropathy type 1 (FJHN1) & medullary cystic kidney disease type 2 (MCKD2)
rs17319721	4	77,587,871	A	SHROOM3	-0.4081	-1.01	A susceptibility gene for kidney disease in an obese mouse model of type II diabetes
rs2467853	15	43,486,085	G	SPATA5L1-GATM	-1.0190	-1.01	SNPs in region association with renal tumors
rs13038305	20	23,558,262	C	CST3-CST9	-0.0084	+1.07	SNPs in region associated with kidney function and endocrine-related traits
rs1731274	8	23,822,264	A	STC1	-0.2326	-1.02	May play a role in regulation of renal and intestinal calcium
Identified in Böger, 2011							
rs11959928	5	39,375,121	A	DAB2	-0.5816	N/A	May modulate growth factor/Ras pathways
rs626277	13	72,347,446	A	DACH1	-0.4470	N/A	Regulates gene expression and cell fate determination during development. Expression of this gene is lost in some forms of metastatic cancer, and is correlated with poor prognosis.
rs10109414	8	23,750,901	C	STC1	-0.3411	N/A	May play a role in the regulation of renal and intestinal calcium
rs13538	2	73,868,078	A	NAT8	-1.7967	N/A	Specifically expressed in kidney and liver. May affect cell adhesion and gastrulation movements.
rs1260326	2	27,730,690	C	GCKR	-0.1417	N/A	Considered a susceptibility gene candidate for a form of maturity-onset diabetes of the young (MODY)
rs4744712	9	71,434,457	A	PIP5K1B	-0.4639	N/A	May be involved in stable platelet adhesion
rs881858	6	43,806,359	A	VEGFA	-0.2690	N/A	Mutations in this gene have been associated with proliferative and nonproliferative diabetic retinopathy
rs347685	3	141,806,887	A	TFDP2	-0.6510	N/A	Transcriptional activation of cell cycle regulated genes

Table 4.4: Multivariable Odds Ratios for CKD (NHANES definition) at Visit 3* and Decline in eGFR during study N=4,273

Measure (yes/no)	Odds of CKD		Additional change in eGFR (ml/min)		Odds of Fast Decline	
	OR (95% CI)	P-value	Estimate per 10 year	P-value	OR (95% CI)	P-value
Baseline eGFR (per 10 ml/min)	-	-	-0.52	<0.0001	-	-
Age (per 10 years)	1.31 (1.22 – 1.40)	<0.0001	-3.5	<0.0001	1.67 (1.44 – 1.95)	<0.0001
Male (yes/no)	0.72 (0.56 – 0.91)	0.006	1.23	0.0007	0.61 (0.47 – 0.76)	<0.0001
Diabetes (yes/no)	1.48 (1.10 – 2.00)	0.01	-3.13	0.01	1.64 (0.90 – 2.96)	0.11
Hypertension (yes/no)	-	-	-1.69	0.0003	1.58 (1.17 – 2.12)	0.003
High uric acid (yes/no)	1.28 (1.00 – 1.67)	0.06	-1.36	0.03	1.97 (1.39 – 2.80)	0.0001
Abnormal Kidney Length (yes/no)	1.26 (1.00 – 1.59)	0.06	-	-	-	-
Genetic Risk Score (per 1 risk allele)	1.07 (1.03 – 1.12)	0.001	-0.23	0.004	1.05 (1.003 – 1.10)	0.04

* Accounts for family clustering

Figure S4.1: Plots of eGFR Decline with Age, by Individual, by Age Decade

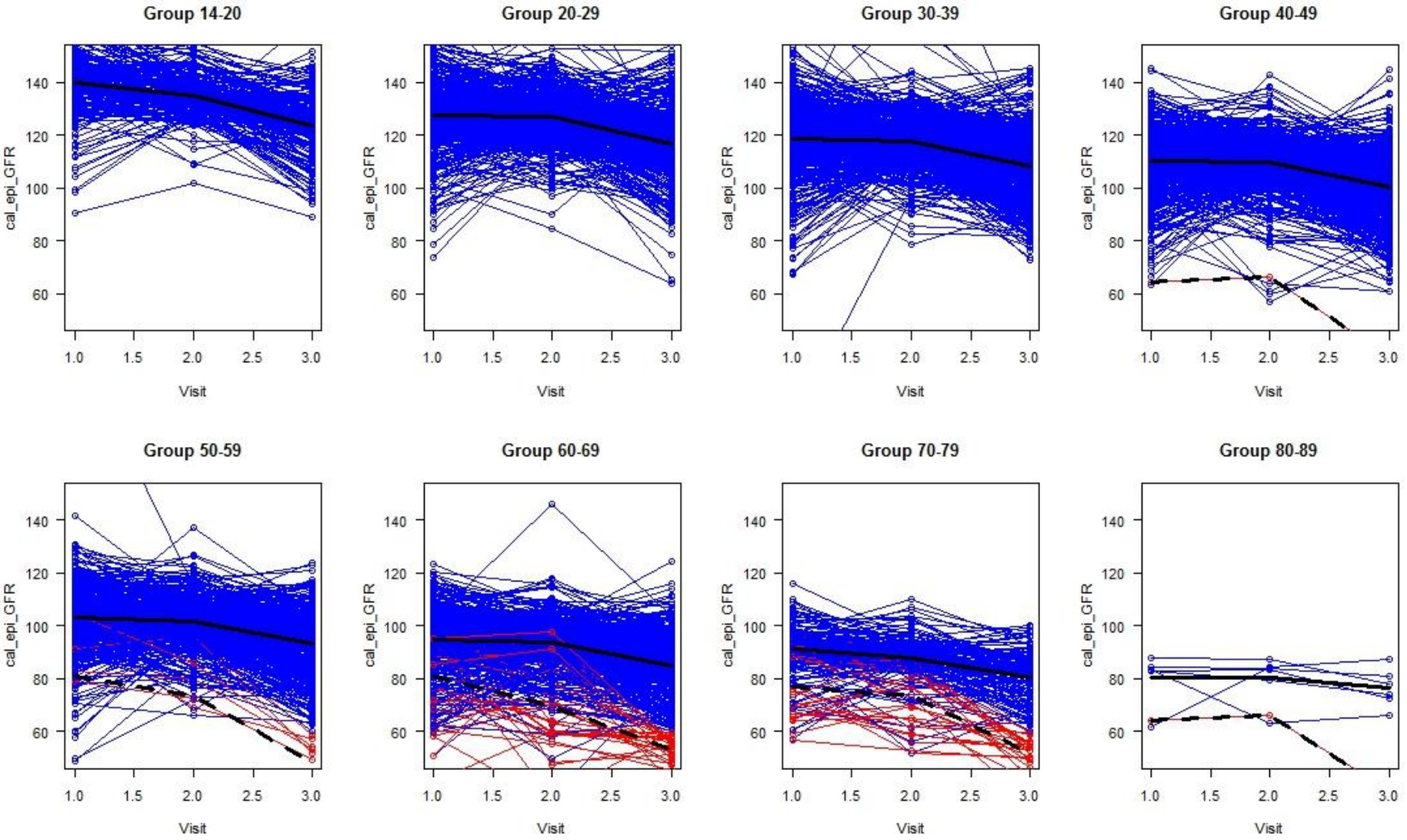


Table S4.1: Comparison of Characteristics by CKD Status and Fast Decline Status

Variable	CKD (n=696)	No CKD (n=4,146)	X^2 or T-test p-value	eGFR Decline > 1 SD* (n=551)	eGFR Decline ≤ 1 SD* (n=3,522)	X^2 or T-test p-value
Categorical Measures:				***Categorical Measures:***		
Age (years, %):						
< 20	0.4	0.8		8.5	6.7	
20 – 39	21.0	32.2		25.5	39.7	
40 – 59	34.9	41.0	<0.0001	37.5	37.4	< 0.0001
60 – 69	17.4	15.0		19.0	12.0	
70 +	26.3	11.0		9.5	4.2	
Male (%)	38.1	43.1	0.01	37.4	42.1	0.04
Smoking (%)	25.6	23.4	0.21	15.9	17.9	0.27
Metabolic syndrome (%)	30.0	20.1	<0.0001	15.4	9.4	<0.0001
Obesity (%)	23.6	17.2	0.0002	17.9	14.3	0.03
Large waist (%)	60.5	47.3	<0.0001	46.0	38.2	0.0003
Diabetes (%)	15.5	8.0	<0.0001	6.6	3.3	0.0002
High glucose (%)	10.6	5.5	<0.0001	5.1	2.3	0.0004
Hypertension (%)	44.5	29.9	<0.0001	40.0	26.2	<0.0001
High Blood Pressure (%)	28.7	21.0	<0.0001	34.8	22.6	<0.0001
Previous cardiac disease (%)	9.5	5.2	0.0003	5.5	3.5	0.03
High uric acid (%)	22.0	16.1	0.0004	16.9	11.2	0.0002
High total cholesterol (%)	66.3	63.3	0.14	62.6	57.4	0.02
High LDL (%)	77.0	76.2	0.66	69.4	67.4	0.35
Low HDL (%)	24.7	18.3	0.0002	5.1	6.2	0.32
Continuous Measures, mean (sd):				***Adjusted mean (se of mean):***		
Age (years)	56.0 (17.2)	48.7 (15.9)	<0.0001	47.5 (0.50)	42.7 (0.19)	<0.0001
BMI	26.9 (4.9)	25.7 (4.7)	<0.0001	25.9 (0.19)	25.3 (0.07)	0.002
Waist circumference (cm)	89.7 (13.2)	86.8 (12.1)	<0.0001	85.9 (0.52)	84.2 (0.20)	0.003
Glucose (mg/dl)	103.7 (33.5)	97.4 (23.2)	<0.0001	92.0 (0.89)	88.3 (0.35)	<0.0001
Sys Blood pressure (mm/Hg)	122.5 (27.0)	118.2 (25.8)	<0.0001	129.5 (0.28)	124.3 (0.72)	<0.0001
Uric Acid ()	4.90 (1.61)	4.58 (1.38)	<0.0001	4.42 (0.06)	4.20 (0.02)	0.001
Total Cholesterol (mg/dl)	216.1 (41.6)	215.2 (40.1)	0.58	212.1 (1.7)	209.7 (0.66)	0.19
LDL Cholesterol (mg/dl)	135.3 (34.8)	135.5 (34.6)	0.88	128.3 (1.5)	127.8 (0.56)	0.74
HDL Cholesterol (mg/dl)	56.7 (15.5)	56.9 (13.8)	0.74	64.5 (0.65)	66.2 (0.25)	0.01
Triglycerides (mg/dl)	121.4 (82.4)	114.3 (66.8)	0.03	87.5 (3.0)	88.2 (1.2)	0.84
Genotype Score	16.51 (2.3)	16.17 (2.3)	0.002	16.4 (0.11)	16.2 (0.04)	0.20

* Adjusted for baseline eGFR and accounts for family clustering

Figure S4.2: Creatinine Calibration Plots for Wave 1 and Wave 3

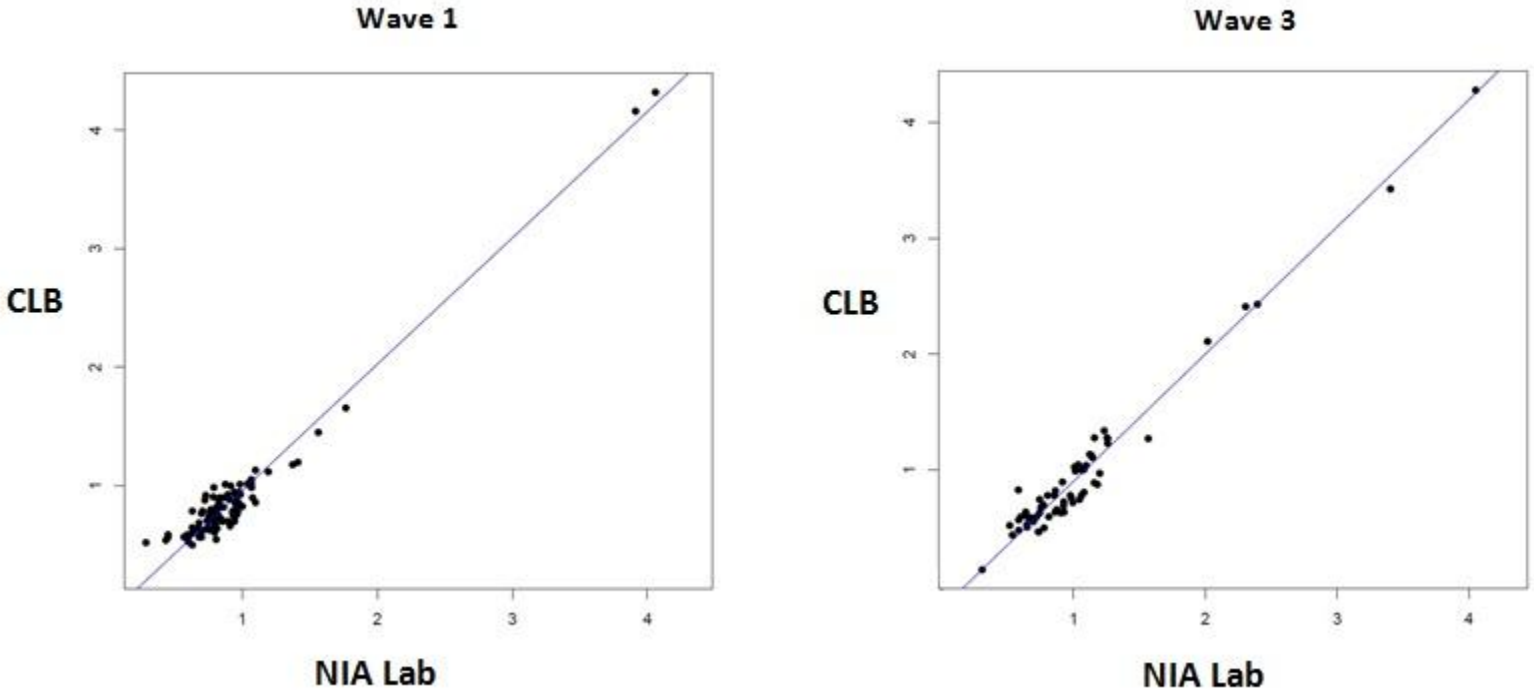
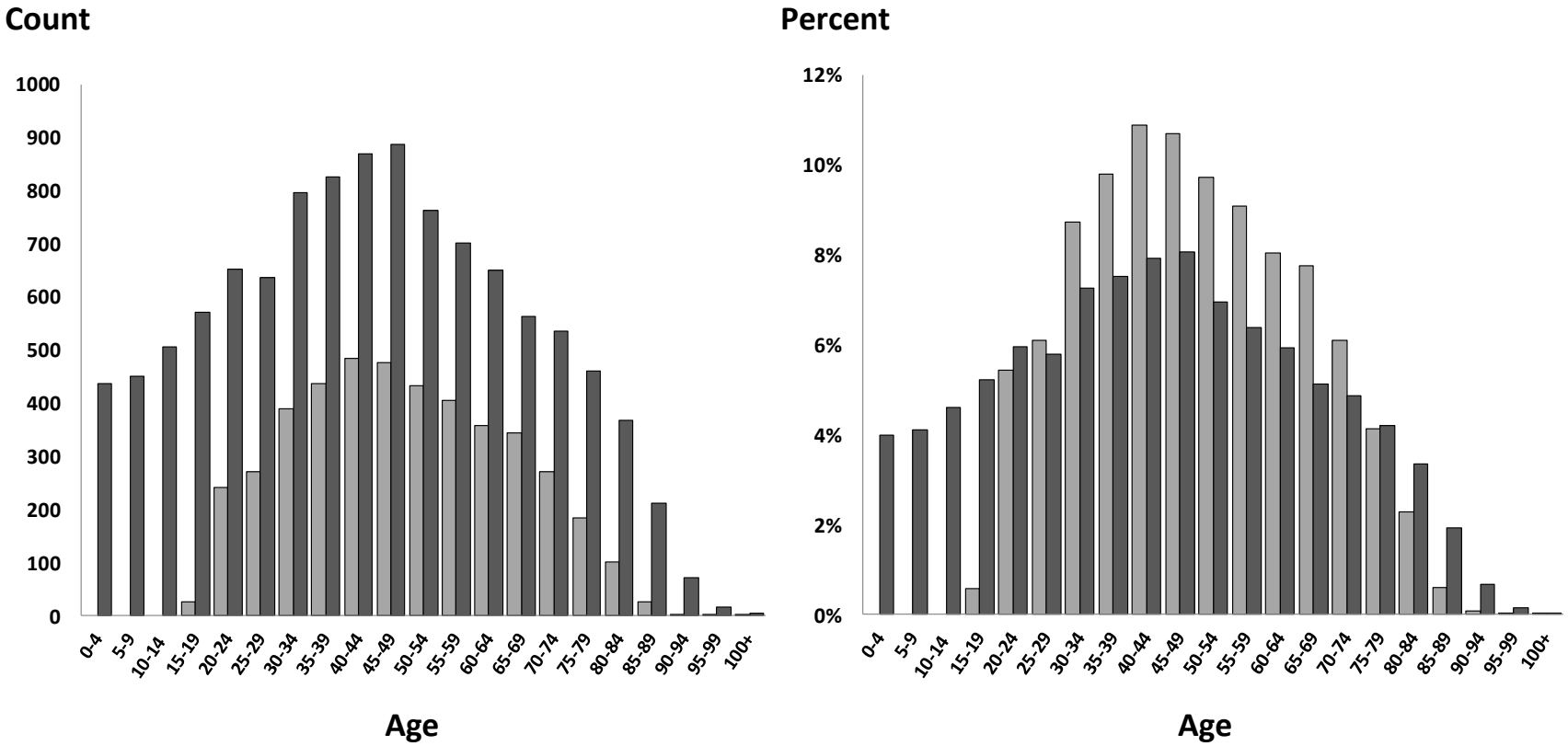


Figure S4.3: Distribution of SardiNIA Sample Compared to Distribution of Population by Age in the Territory of Lanusei



CHAPTER V

CONCLUSION

Due to the huge population age shift that the world is currently experiencing, our emphasis on disease prevention and management of chronic diseases needs to increase. In 2006, almost 500 million people worldwide were 65 and older. By 2030, that total is projected to increase to 1 billion, comprising approximately 1 in every 8 of the earth's inhabitants (Dobriansky, et.al., 2007). The most rapid increases in the 65-and-older population are occurring in developing countries, which will see a jump of 140 percent by 2030 (Dobriansky, et.al., 2007).

The number of deaths due to chronic health conditions have superseded the number of deaths from communicable causes world-wide, but strikingly so in middle to high-income countries, such as the United States (CDC, 2013). In high-income countries, 7 out of the 10 leading causes of death are chronic health conditions and include: heart disease, cancer, chronic lower respiratory diseases, stroke/cerebrovascular diseases, Alzheimer's disease, diabetes, and nephritis/nephrotic syndrome/nephrosis.

In most instances, such as in cardiovascular diseases, diabetes, and kidney disease, damage has occurred to the individual's health before there are clinical symptoms. The CARDIA study (Friedman, et.al, 1988) was designed to investigate the development of heart disease in black and white adults and has helped bring attention to the need for early recognition and intervention for heart disease. They have found that the early intervention can have lasting effects years into the future and are more beneficial than intervention at later ages (Lee DH, et.al, 2003).

In order to develop better early recognition and intervention procedures for these chronic conditions, we must first understand more about the biological process of aging. The goal of this dissertation was to develop a measure of aging, applicable to a wide range of ages. Further, it aimed to determine the genetic basis of this measure, in the hopes that genetic biomarkers could be discovered. With such measures available from the time of conception, we can begin to assess the potential value of incorporating such genetic risk biomarkers into individual-level health care. We are still just beginning to enter the age of personalized genomic medicine, but it's possible that we can begin to incorporate the use of genetic information in the care of individual's with specific diseases, such as chronic kidney disease. Therefore, this work also investigated the utility of adding a genetic risk score, along with the accepted clinical risk factors for CKD, in the prediction of the decline in kidney function.

In the second chapter of this dissertation, I constructed two different measures of aging that showed moderate heritability. My method was quite novel compared to the approaches taken in the past, which have tended to focus on longevity among individuals of very advanced age, such as centenarians. Most of the previous research also used morality as the outcome

measure. Instead of using this approach, I examined adults aged 20 to 90 years of age and developed two measures of biologic age which were based on three easily measured physiologic indicators of health: serum creatinine, systolic blood pressure, and waist circumference. This biologic age was then compared to the individual's actual age and the difference was then assessed for a genetic component. These difference measures can be interpreted as the number of years the individual is physically older or younger than the average person with the same physiologic measurement values.

By using only three common measures, I was able to obtain data from five cohorts with genetic data on individuals of European ancestry. Data from the ARIC and FHS were obtained through dbGaP, while data from SardiNIA, TwinGene, and the TwinsUK came through collaboration channels. Data on over 33,000 individuals was available for phenotype development.

The first estimation of biologic age was created by fitting a linear mixed model for each cohort to determine each individual's predicted age, based on their physiologic health measure. The second estimate of biologic age incorporated the individual's risk of mortality predicted by the same three physiologic measures. This risk of mortality was then compared to the country specific death risk reported from census data. Each individual was assigned a "risk" age that was equivalent to the census age that had the equivalent mortality risk.

Upon investigation of the fitted models by cohort, it became apparent that the age distribution of sampled individuals had a strong influence on the ability to obtain the best model fit for the prediction of biologic age. The predicted age models fit best within the

SardiNIA cohort, which has a very large age distribution (included 20-90 year olds in the current study), while the mortality risk models were best estimated in cohorts with a larger proportion of older individuals that contained more deceased individuals, likely due to the use of 1-year mortality estimates.

In order to assess the traits for a genetic component, heritability analyses were run among the related cohorts. Both traits were found to be moderately heritable with the lowest estimates seen for the risk age trait in FHS (h^2=0.15), where there were few very old individuals, and in the SardiNIA study (h^2=0.18), where there were only 234 deaths among the sample of nearly 6,000 individuals. Heritability estimates were highest for the risk trait among the twin studies (h^2=0.86-0.90), but may be unreliable, due to the fact that the use of age at entry of the study was necessary in the models and twins typically entered at the same age. The predicted age differential showed more consistent heritability across the cohorts, especially when restricted to individuals between 45-65 years of age (the age range that was present in all cohorts), with estimates from 0.25-0.33 in the family pedigrees to 0.60-0.70 in the twin studies.

In both cases, there was good evidence of a genetic component to the traits, so both traits were followed-up using genome-wide association methods, separately for four of the cohorts (ARIC, FHS, SardiNIA, and TwinsUK), and these results were meta-analyzed. GWAS is currently being run for the TwinGene cohort of approximately 10,000 individuals and these results will be used for replication.

Based on past research of aging and evolutionary theory, we hypothesized that significant associations with the aging trait would be found within two broad classes of genes.

First, it would be very plausible to find associations with genes involved in controlling the levels of activities, such as DNA repair and antioxidant defense, thus regulating longevity. Secondly, genomic loci that contained late-acting deleterious genes that had escaped the force of natural selection or that traded benefit at an early age against harm at older ages, termed antagonistic pleiotropy, could be expected to be discovered (Kirkwood, 2008).

One genome-wide significant association was detected in the meta-analysis for the predicted age differential and was located within the *LRP1B* gene on chromosome 2. We found that individuals with the minor allele for SNP rs10496861 (MAF=0.34 in our samples) had a 2-3 year higher difference in predicted age compared to age ($p=5.78 \times 10^{-9}$) than an average individual in the same age decade.

Past research, highlighted in Chapter III, has shown the *LRP1B* gene to be highly pleiotropic. The *LRP1B* gene is very large (1,900,279 bases) and produces 41 different known proteins. Previous studies have found association between this gene and phenotypes such as thyroid cancer, where risk was increased due to inactivation of the gene (Prazeres, 2010), and insulin-resistance (Burgdorf, 2012). Interestingly, another study on aging, that employed a very different study design, also uncovered associations between 3 SNPs in this same gene and successful aging without cognitive impairment (Poduslo, et.al, 2009). The SNPs in this previous study ranged in distance from our top hit by only 32 to 72 Kb. Bioinformatic investigation of our polymorphism showed it to alter expression levels of three other genes, *LOC407835, MAP2K2,* and *SLC38A5*.

The goal of the research in Chapter IV was to demonstrate the next step in taking experimental science findings to application. Working with the longitudinal data from the SardiNIA study, I created a genotype risk score for CKD based on loci found to be significantly associated with kidney function in two large meta-analyses of GWAS studies (Köttgen, 2009 and Böger, 2011). Thirteen loci from the total of eighteen published loci were available in the SardiNIA cohort and showed similar direction of association to that of the published meta-analysis results. For ease of interpretation the thirteen loci were combined in an additive manner, so that association results can be interpreted as the effect of having one 1 additional risk allele.

The genetic risk score was significantly associated with current kidney function and also future kidney function, after taking into account known clinical risk factors like diabetes and hypertension. First, the genetic risk score was predictive of baseline CKD status (OR=1.07 per one additional risk allele, p=0.001). More importantly, though, the risk score was also highly associated with future decline in kidney function over the study follow-up (β=-0.23 per one additional risk allele, over the following 10 years, p=0.004). The risk score was also significantly associated with the odds of being a "fast decliner", defined as having a decline in kidney function that was greater than 1 SD below the mean (or a decline more extreme than 2.3 ml/min/year, OR=1.05 per one additional risk allele, p=0.04).

Although these chapters have been presented as a complete project, there is much more to be done to make the findings useful in combating chronic disease burden. GWAS are simply the first step in identifying regions of interest associated with specific phenotypes. The goal now is to determine which SNP is actually the causal SNP and to elucidate the functional

consequences more fully. Many times the top findings in a GWAS is close to the causal SNP, but may not be the exact polymorphism. To accomplish this, the next step should be an analysis of the rare variants within and around the *LRP1B* gene, either through the use of exome chip data, exome sequencing data, or potentially imputing the current data on the 1000 Genomes reference panel (http://www.1000genomes.org/), which contains rare loci. The Exome Sequencing Project data in dbGaP would be a good first place to explore, if the data is available to create the two aging traits.

If publically accessible data is not available, then it would be worthwhile to begin further work by sequencing the exons the *LRP1B* gene both upstream and downstream of the hit. Because our variant was found within an intron, it is important to sequence 50 bp outside of each exon to be sure that the intron/exon borders are captured. If nothing is found in the immediate region, then targeted sequencing of the whole *LRP1B* gene would be necessary, being sure to include the promoter region (usually 2000 bp at the 5' end). After attempting to find the variant with sequencing, if nothing is found then we should being looking into epigenetic explanation. There are many inexpensive assays available to quantify methylation of the gene. Also, gene-environment interactions and potential regulatory mechanisms should be explored.

When the causal variant is determined, it would be useful to look at the associations between the genotypes and expression level of the gene. Since we know that the *LRP1B* gene is an eQTL loci for 3 other genes, expression levels of those genes should also be examined. If a sound mechanism can be hypothesized that is supported by this multi-pronged approach, then

one could think about moving into animal models for testing the hypotheses. This should not be difficult for this particular gene, since it has homologues in six different well studied species.

If the true mechanism for the association observed between the *LRP1B* gene and aging is actually related to endocytosis and autophagy as proposed in Chapter III, previous studies (Rubinsztein, 2011) have shown that it could be a target for pharmacological or genetic manipulations. In model organisms, these manipulations have demonstrated that stimulating autophagy often increases life span. Conversely inhibition of autophagy compromises the longevity-promoting effects of caloric restriction, Sirtuin 1 activation, inhibition of insulin/insulin growth factor signaling, or the administration of rapamycin, resveratrol, or spermidine.

While much work remains to be done to discern the true mechanism and utility of the general aging findings discussed in Chapters II and III, the kidney risk score explored in Chapter IV could potentially begin to be explored in the clinical setting. Over the past 100 years great strides have been made in renal research to allow individuals with little or no kidney function to continue to live, but mortality rates remain high for individuals receiving dialysis.

Until the 1940's individuals with acute or chronic kidney function had no options and often died within days of their kidneys failing. John Abel is credited with having the first idea of an artificial kidney in 1914, but it took the personal investment and interest of many individuals (Abel, Rountree, Turner, Haas, Hess and McGuigan, Macallum, Lambert and Vogel, Thalheimer, Murray, Macneil and the Anthone twins, Kolff, Skeggs and Leonard, Alwall, Kiil, Rosenak and others less recognized were) to see the potential widespread clinical use, and Gordon

Murray, along with Delorme and Thomas, to construct an artificial kidney in Murray's basement with a personal investment of $8,000 in 1946 (Schreiner GE, 2000).

Dialysis is expensive and in the 1960's communities would often have bake sales and fundraisers to help raise money for friends or loved ones in need of dialysis. State coverage for care was very limited and the cause began to be picked up by the media. "The coverage extended from local weeklies to TV and Life Magazine: "Who shall live and who shall die?" The people and their families and their surrogate, the National Kidney Foundation, recognized the possibility of a "national" solution, and historically, several things converged at a national level." (Schreiner GE, 2000). In 1972 President Nixon signed the bill that allowed patients with end stage renal disease (ESRD) to be classified as disabled and to be eligible for Medicare coverage, regardless of age.

ESRD coverage by Medicare has extended countless lives, but has also created a large burden on society, from loss of individuals in the work force to increased Medicare spending. In the US, costs for the care of ESRD patients in 2009 rose 3.1% to $29 billion (USRDS, 2011). Mortality rates for patients with ESRD have been decreasing slightly due to work by researches, focused on improving patient care, but the rates are still very high (> 20% mortality/year in the US) and patients in the US receiving dialysis for renal replacement therapy have only a 34% survival rate after 5 years on treatment (USRDS, 2011).

Individuals with CKD, who are not yet at end-stage of the disease, have much better survival and quality of life than individuals on hemodialysis, and are much less healthcare

expense. If we can identify these individuals earlier in the course of the disease, there is hope that many could live without ever having to begin dialysis.

I feel the next step in trying to incorporate the genetic risk score into patient care should be a refinement of the measure, incorporating data from more cohorts. As mentioned, the score developed in this project was meant as a proof of concept and it is likely a more predictive measure could be created from the 13 SNPs not in linkage disequilibrium with each other. A possible refinement that should be more thoroughly explored is the use of weighting for each SNP in the score by its effect size. Another option would be to explore the use of principal component analysis to reduce the dimension from 13 separate variables in a meaningful way.

After a suitable measure is created, we would need to test it utility, including sensitivity and specificity characteristics. This would be a very long term project if we started with a population from the general public, as development of CKD does not usually begin until well into adulthood. A more plausible study design may be to begin within a nephrology clinic that sees some patients who are still in the early stages of the disease and not yet requiring dialysis. If they could be genotyped and followed, monitoring their changes in kidney function over time, we may find that we can split the patients into those who we'd expect to see a faster decline and those who we expect a more gradual decline. Then one could begin to customize treatment options specifically to this expectation and either be more or less aggressive in treating the other conditions that contribute to kidney disease.

The sample size for this follow-up project would need to be quite large, as there are a certain set of individuals who may not be as responsive to normal therapies and who may show no benefit. One possible study to approach with this objective is the Chronic Kidney Disease Surveillance Project, which is funded by the CDC and run out of the University of Michigan's Kidney Epidemiology and Cost Center (http://www.sph.umich.edu/kecc/html/ckdsurveillance.html). This study was funded in order to pilot a National Surveillance System for Chronic Kidney Disease in the US. The study uses data from combined sources, such as the University of Michigan Health Center and Blue Cross Blue Shield of Michigan. Successful feasibility and pilot testing of the proposed system will allow the establishment of national CKD surveillance that will likely shape quality improvement, promote research and inform health policy related to CKD.

In summary, work must continue in the area of genetics to help us better understand the etiology of chronic diseases and help us shape possible new treatments, in order to effectively manage the healthcare of our aging population structure. The findings in this dissertation have demonstrated that there are many ways to assess the genetic contributors to aging and specific aging-related diseases. Using multi-pronged methods will allow us to assess the quality of our findings, and hopefully in the near future, put them into practice to help save productive quality years of life for our elderly.

BIBLIOGRAPHY

Abecasis GR, Cherny SS, Cookson WO and Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 30:97-101

Ahima RS (2009) Connecting obesity, aging and diabetes. *Nature Medicine*. 15(9). Pg 996-997.

Aulchenko YS, Ripke S, Isaacs A, and van Duijn CM (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23 (10):1294-1296.

Affymetrix: http://www.affymetrix.com

Alberti KG, Zimmet P, Shaw J. (2006, May) Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med.* 23(5):469-80.

Beland MD et al. (2010) Renal Cortical Thickness Measured at Ultrasound: Is It Better Than Renal Length as an Indicator of Renal Function in Chronic Kidney Disease? *American Journal of Roentgenology*; 195: w146-w149.

Böger CA et al. (2011) CUBN is a gene locus for albuminuria. *JASN* 2011; 22: 555-570.

Böger CA, Gorski M, Li M, Hoffmann MM, Huang C, Yang Q, et al; (2011) CKDGen Consortium. Association of eGFR-Related Loci Identified by GWAS with Incident CKD and ESRD. *PLoS Genet*.; 7(9):e1002292.

Burdick JT, Chen WM, Abecasis GR, Cheung V. (2006) In silico method for inferring genotypes in pedigrees. *Nat Genet*. 2006 September ; 38(9): 1002–1004.

CDC - National Center for Health Statistics – Fastats. (2013) http://www.cdc.gov/nchs/fastats/lcod.htm. January 27.

Centers for Disease Control and Prevention (2012). Summary health statistics for U.S. adults: National Health Interview Survey, 2010. Hyattsville, MD: National Center for Health Statistics. Vital and Health Statistics 10(252); 2012. Available online: http://www.cdc.gov/nchs/data/series/sr_10/sr10_252.pdf.

Cirillo M, Laurenzi M, Mancini M, Zanchetti A, Lombardi C, De Santo NG. (2006) Low glomerular filtration in the population: prevalence, associated disorders, and awareness. Kidney Int. 2006 Aug;70(4):800-6.

Clark WF, Macnab JJ, Sontrop JM, Jain AK, Moist L, Salvadori M, Suri R and Garg AX,. (2011) Dipstick proteinuria as a screening strategy to identify rapidly renal decline. J Am Soc Nephrol 2011;22:1729-1736.

Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, Van Lente F, Levey AS. (2007) Prevalence of Chronic Kidney Disease in the United States. *JAMA*, 298(17):2038-2047

Dawber T, Meadors G, and Moore F. (1950). Epidemiological Approaches to Heart Disease: The Framingham Study. Presented at a joint session of the Epidemiology, Health Officers, Medical Care, and Statistics Sections of the APHA.

De Benedictis, G. Rose, G. Carrieri, M. De Luca, E. Falcone, G. Passarino, M. Bonafe´, D. Monti, G. Baggio, S. Bertolini, D. Mari, R. Mattace, And C. Franceschi. (1999, September). Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. G. *The FASEB Journal*. Vol. 13, pg. 1532-1536.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W. et al. (2007), A genome-wide association study of global gene expression. *Nat. Genet*. Vol. 39, pp. 1202–1207.

Eurostat news release (2011, November 24). European Health Iterview Survey. Between 8% and 25% of adults are obese across member states. No systematic differences between women and men. 172/2011.

Feig DI, Kang DH, Johnson RJ. (2008) Uric Acid and Cardiovascular Risk. *N Engl J Med* 359:1811-21.

Flegal KM, Carroll MD, Kit BK, Ogden CL. (2012) Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *Journal of the American Medical Association*. 307(5):491–97.

Folstein MF, Robins LN, Helzer JE. (1983) The Mini-Mental State Examination. *Arch Gen Psychiatry.* 40(7):812.

Frankowski H, Alavez S, Spilman P, Mark K, Nelson JD, Mollahan P, Rao RV, Chen S, Lithgow G, Ellerby HM. (2010). Dimethly sulfoxide and dimethyl fornamide increase lifespan in *C. elegans* in liquid. *Mechanisms of Ageing and Development*.

Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, Jr, Liu K, Savage PJ. (1988). CARDIA: Study design, recruitment and some characteristics of the examined subjects. *Journal of Clinical Epidemiology*. 41(11):1105-1116.

Gambaro G, Yabarek T, Graziani MS, Gemelli A,, Abaterusso C, Frigo AC, Marchionna N, Citron L, Bonfante L, Grigoletto F, Tata S, Ferraro PM, Legnaro A, Meneghel G, Conz P, Rizzotti P, D'Angelo A, and Lupo A, for the INCIPE Study Group. (2010) Prevalence of CKD in Northeastern Italy: Results of the INCIPE Study and Comparison with NHANES. *Clin J Am Soc Nephrol* 5: 1946–1953

Glassock RJ, Winearls C. (2008) An epidemic of CKD: fact or fiction? *NDT*; 23: 1117-21.

Haas J, Beera AG, Widschwendtera P, et.al. (2011) LRP1b shows restricted expression in human tissues and binds to several extracellular ligands, including fibrinogen and apoE – carrying lipoproteins. *Atherosclerosis* 216:342–347.

Hallan SI, Coresh J, Astor BC, Åsberg A, Powe NR, Romundstad S, Hallan HA, Lydersen S, and Holmen J. (2006) International Comparison of the Relationship of Chronic Kidney Disease Prevalence and ESRD Risk. *J Am Soc Nephrol* 17: 2275–2284.

Hallan S, Astor B, Lydersen S. (2006, June) Estimating glomerular filtration rate in the general population: the second Health Survey of Nord-Trondelag (HUNT II). *Nephrol Dial Transplant.*, 21(6):1525-33.

Han J, Hubbard B, Lee J, Montagna C, Lee HW, Sinclair D, and Suh Y. (2013). Analysis of 41 cancer cell lines reveals excessive allelic loss and novel mutation in the SIRT1 gene. *Cell Cycle*. 12:2, 263-270.

Heidenreich PA, Trogdon JG, Khavjou OA, et al. (2011) Forecasting the future of cardiovascular disease in the United States. *Circulation*. 123:933-944.

Hollingworth P, Harold D, et.al. (2011) Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nature Genetics* 43:5. 429-436.

Howie BN, Donnelly P, Marchini J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529.

Kanbay M, Ozkara A, Selcoki Y, et al. (2007) Effect of treatment of hyperuricemia with allopurinol on blood pressure, creatinine clearence, and proteinuria in patients with normal renal functions. *Int Urol Nephrol*., 39:1227-33.

Kang DH, Nakagawa T, Feng L, et al. (2002) A role for uric acid in the progression of renal disease. *J Am Soc Nephrol*., 13:2888-97.

Köttgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M et al. (2009) Multiple loci associated with indices of renal function and chronic kidney disease. *Nature Genetics*, 2009; 41 (6) 712-717

Köttgen A, Yang Q, Shimmin LC, Tin A, Schaeffer C, Coresh J, et al. (2012) Association of estimated glomerular filtration rate and urinary uromodulin concentrations with rare variants identified by UMOD gene region sequencing. *PLoS One*. 7(5):e38311.

Kraft P, Zeggini E, and Ioannidis JPA. (2009) Replication in Genome-Wide Association Studies. *Statistical Science.* 2009, Vol. 24, No. 4, 561–573.

Lee DH, Jacobs DR Jr., Gross M, Kiefe CI, Roseman J, Lewis CE, and Steffes M. (2003). γ-Glutamyltransferase Is a Predictor of Incident Diabetes and Hypertension: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Clinical Chemistry* 49:8. 1358–1366.

Levey AS, Eckardt KU, Tsukamoto Y, et al. (2005) Definition and classification of chronic kidney disease: a position statement from kidney disease: Improving Global Outcome (KDIGO). *Kidney Int*.; 67(6):2089-2100

Levey AS, Atkins et al. (2007) Chronic Kidney Disease as a global public. *Kidney Int*., 72: 247-249.

Levey, Coresh. (2012) Chronic kidney disease. *Lancet*., 379: 165-80

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.

Li Y, Willer CJ, Sanna S and Abecasis GR (2009) Genotype Imputation. *Annu Rev Genomics Hum Genet* 10:387-406.

Liu CX, Li Y, Obermoeller-McCormick LM, Schwartz AL, and Bu G. (2001). The Putative Tumor Suppressor *LRP1B*, a Novel Member of the Low Density Lipoprotein (LDL) Receptor Family, Exhibits Both Overlapping and Distinct Properties with the LDL Receptor-related Protein. *The Journal of Biological Chemistry.* Vol. 276, No. 31, Issue of August 3, pp. 28889–28896.

Markaki M, Tavernarakis N. (2011) The role of autophagy in genetic pathways influencing ageing. *Biogerontology.* 12:377–386.

Materson BJ, Reda DJ, Cushman WC et al. (1993) Single-drug therapy for hypertension in men – a comparison of six antihypertensive agents with placebo. *N Engl J Med*., 328: 914-21.

Mazzali M, Hughes J, Kim YG, et al. (2001) Elevated uric acid increases blood pressure in the rat by a novel crystal-independent mechanism. *Hypertension*, 38: 1101-6.

Moayyeri A, Hammond C, Valdes A, and Spector T. (2012 January) Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int. J. Epidemiol*. 1–10.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L. et al. (2007), 'Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma'. *Nature* Vol. 448, pp. 470–473.

Morrison MA, Alexandra CS, Huynh N, Jun G, Silvia ES, et.al. (2011 October) Systems biology-based analysis implicates a novel role for vitamin D metabolism in the pathogenesis of age-related macular degeneration. *Human Genomics*. VOL 5. NO 6. 538–568.

Naitza S, Porcu E, S Maristella, Taub DD, et.al. (2012) A Genome-Wide Association Scan on the Levels of Markers of Inflammation in Sardinians Reveals Associations That Underpin Its Complex Regulation. *PLoS Genetics*. Volume 8. Issue 1. Pg 1-12.

National Kidney Foundation. (2002) K/DOQI Clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification. *Am J Kidney Dis*.; 39(2)(suppl 1):S1-S266.

Nesse, R.M., Williams, G.C. (1996). Why We Get Sick. Pp. 107-122. First Vintage Books.

Nugent RA, Fathima SF, Feigl AB, Chyung D. (2011) The burden of chronic kidney disease on developing nations: a 21st century challenge in global health. *Nephron Clin Pract*. 118(3):c269-77.

Nynke Halbesma, Desiree F. Jansen, Ronald P. Stolk, Paul E. De Jong, Ronald T. Gansevoort and for the PREVEND Study group. (2011) Changes in renal risk factors versus renal function outcome during follow-up in a population-based cohort study. *Clin J Am Soc Nephrol* 6: 2121–2131.

Ogden CL, Carroll MD. (2010) Prevalence of overweight, obesity, and extreme obesity among adults: United States, trends 1960–1962 through 2007–2008. NCHS Health E-

Stat. Hyattsville, MD: National Center for Health Statistics. (http://www.cdc.gov/nchs/data/hestat/obesity_adult_07_08/obesity_adult_07_08.pdf)

Ogden CL, Carroll MD, Kit BK, Flegal KM. (2012) Prevalence of obesity and trends in body mass index among US children and adolescents, 1999–2010. Journal of the American Medical Association. 307(5):483–90.

Osiewacz H, Hamann A, and Zintel S. (2013) Assessing Organismal Aging in the Filamentous Fungus *Podospora anserina*. *Cell Senescence: Methods and Protocols, Methods in Molecular Biology*, vol. 965.

Paaby AB. and Paul S. Schmidt. (2009). Dissecting the genetics of longevity in Drosophila melanogaster. *Landes Bioscience.* Pg. 1-10.

Pilia G, Chen WM, Scuteri A, Orru´ M, Albai G, Dei M, et al. (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* 2(8): e132.

Poduslo SE, Huang R, and Spiro III A. (2009) A Genome Screen of Successful Aging Without Cognitive Decline Identifies *LRP1B* by Haplotype Analysis. *American Journal of Medical Genetics*. Part B. Pages 114-119.

Prakash S, O'hare AM. (2009) Interaction of aging and chronic kidney disease. *Semin Nephrol*, 29: 497-503.

Pruim RJ*, Welch RP*, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. (2010) LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*. September 15; 26(18): 2336.2337.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rubinsztein DC, Mariño G, Kroemer G. (2011) Autophagy and Aging. *Cell*. 146, September 2.

Sanusi AA, Arogundade FA, Famurewa OC, et al. (2009) Relationship of ultrasonographically determined kidney volume with measured GFR, calculated creatinine clearance and other parameters in chronic kidney disease (CKD). *NDT*; 24:1690–1694.

Sarnak MJ, Levey AS, Schoolwerth AC et al. (2003) Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. *Circulation*. 108: 2154–2169.

SAS. SAS Institute, Cary NC.

Schreiner GE. (2000) How End-Stage Renal Disease (ESRD)-Medicare Developed. American Journal of Kidney Diseases - Volume 35, Issue 4.

Selvin E, Manzi J, Stevens LA, Van Lente F, Lacher DA, Levey AS, Coresh J. (2007) Calibration of serum creatinine in the National Health and Nutrition Examination Surveys (NHANES) 1988-1994, 1999-2004. *Am J Kidney Dis*. Dec;50(6):918-26.

Siu YP, Leung KT, Tong MK, Kwan TH. (2006) Use of allopurinol in slowing the progression of renal disease through its ability to lower serum uric acid level. *Am J Kidney Dis*; 47:51-9.

Skali H, Uno H, Levey AS, Inker LA, Pfeffer MA, Solomon SD. (2011) Prognostic assessment of estimated glomerular filtration rate by the new Chronic Kidney Disease Epidemiology Collaboration equation in comparison with the Modification of Diet in Renal Disease Study equation. Am Heart J. Sep;162(3):548-54.

US Renal Data System (2011), USRDS 2011 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2012.

Viktorsdottir O, Palsson R, Andresdottir MB, Aspelund T, Gudnason V, Indridason OS. (2005) Prevalence of chronic kidney disease based on estimated glomerular filtration rate and proteinuria in Icelandic adults. *Nephrol Dial Transplant*.; 20(9):1799-807.

Wetzels JF, Kiemeney LA, Swinkels DW, Willems HL, den Heijer M. (2007) Age- and gender-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study. *Kidney Int*. Sep;72(5):632-7.

Weight-control Information Network. (2012) NIH Publication No. 04–4158. Updated October 2012. (http://win.niddk.nih.gov/publications/PDFs/stat904z.pdf).

Wieneke MM, DC Grootendorst, M Verduijn, EG Elliott, FW Dekker, and RT Krediet. (2010) Performance of the Cockcroft-Gault, MDRD, and New CKDEPI Formulas in Relation to GFR, Age, and Body Size. *Clin J Am Soc Nephrol* 5: 1003–1009.

Willer CJ, Y Li, and GR Abecasis (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. Vol. 26 no. 17, pages 2190–2191.

World Health Organization Media Centre (2011) - Fact Sheet # 310. (http://www.who.int/mediacentre/factsheets/fs310/en/index.html).

Zhang QL, Rothenbacher D. (2008) Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health*. 8:117.

Zhang L, Zhang P, Wang F, Zuo L, Zhou Y, Shi Y et al (2008) Prevalence and Factors Associated With CKD: A Population Study From Beijing. *American Journal of Kidney Disease*, 51: 373-384.

Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al. (2012) Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet*, 379(9818):815-22.

Zoccali et. Al. (2010) Epidemiology of Chronic Kidney Disease in Europe: an uncertain scenario. *NDT,* 25: 1731- 1733.

Zoccali C, Kramer A and Jager K. (2010) Chronic kidney disease and end-stage renal disease—a review produced to contribute to the report 'the status of health in the European union: towards a healthier Europe'. *NDT Plus*, 3:213-224.