DISENTANGLING THE EFFECTS OF BODY SIZE AND GENETICS ON INSULIN-LIKE GROWTH FACTOR-1 AND ITS RELATIONSHIP TO MORTALITY

by

Rehab Sherlala

MBBCh, Zawiya University, Libya, 2006

MPH, University of Pittsburgh, 2013

Submitted to the Graduate Faculty of the

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Rehab A. Sherlala

It was defended on

April 8, 2021

and approved by

Candace M. Kammerer, PhD, Associate Professor,
Department of Human Genetics

Eleanor Feingold, PhD, Interim Chair and Professor,
Departments of Human Genetics and Biostatistics

Allison L. Kuipers, PhD, Assistant Professor,
Department of Epidemiology

Dissertation Director: Ryan L. Minster, PhD, MSIS, Assistant Professor,
Department of Human Genetics

DISENTANGLING THE EFFECTS OF BODY SIZE AND GENETICS ON INSULIN-LIKE GROWTH FACTOR-1 AND ITS RELATIONSHIP TO MORTALITY

Rehab A. Sherlala, PhD

University of Pittsburgh, 2021

Abstract

The number of Americans age 65 and older is expected to increase to over 98 million by 2060 and results in an increased demand for health services due to increased rates of aging-related diseases. Studies of traits that are associated with longevity, such as the insulin-like growth factor-1 (IGF-1) levels, may provide insights to mitigate effects of these diseases. The relationship between IGF-1 and body mass index (BMI), another trait associated with morbidity and mortality, is unclear. Furthermore, only seven genetic loci are known to be associated with IGF-1 levels. In this study, I analyzed data from the Long Life Family Study participants, a unique cohort of two-generation families. My analyses indicated that the relationship between IGF-1 and BMI differs across age groups, and this pattern was also seen in non-Hispanic White, non-Hispanic Black, and Mexican American participants in the third National Health and Nutritional Examination Survey. Genetic analysis revealed a novel locus associated with IGF-1 levels on chromosome 11 (LOD = 3.48) as well as a previously known region on chromosome 7p12.3 ($p \leq 0.00023$), although the identity of the specific gene (or genes) involved is unclear. Greater serum IGF-1 levels were associated with lower risk of mortality (HR = 0.61, 95% CI = 0.53–0.70, $p = 1.4\times 10^{-10}$), an association that was attenuated after adjusting for age; however, the genetic variants associated with IGF-1 levels themselves were not associated with the risk of mortality. Additional genetic studies are required to elucidate the role that IGF-1 plays in age-related morbidity and mortality. My study has contributed to our understanding of the interplay between genetic and environmental factors that

influence IGF-1 levels and this knowledge may enable the development of methods to mitigate the development of age-related diseases and improve public health.

Table of Contents

List of Tables

List of Figures

Preface

While this project is the result of hard work, pain, and pressure, it has also allowed me to taste success and joy. I would not have been able to complete this project without the help and support of many people. I would like to express my thanks to all who contributed in one way or another to making this research a reality: the participants as well as the research staff who recruited and surveyed them and gave me the opportunity to be part of this project. In addition, special thanks to the University of Zawiya and the Ministry of Higher Education in Libya through the Libyan North American Scholarship Program (LNASP) for supporting and sponsoring my scholarship. I sincerely express my gratitude to my committee members, Candace M. Kammerer, Eleanor Feingold, Allison L. Kuipers, and Ryan L. Minster for their advice, guidance, and support. My special thanks to Candace M. Kammerer and Ryan L. Minster for accepting to be my mentor and walking me through the whole PhD program with care and passion. I would like to thank everyone who helped and worked with me to shape this research: Emily Russell, Ryan Cvejkus, Hannah Campbell, and my dear friend and study partner Zeynep Erdogan-Yildirim.

Last and not least, I want to thank my husband, Esam Abuharba, and my kids: Ahmed, Rafeef, Reem, and Maryam for their moral support, encouragement, and patience. In addition, my friends and family here and back home, in Libya, thank you for your support, care, and prayers.

1.0 INTRODUCTION

1.1 Public health relevance

The number of Americans age 65 and older is expected to increase from an estimated 46 million in 2016 to over 98 million by 2060, representing an increase in the proportion of the population from 15% to 24% [1, 2]. Globally, the number of older persons, 65 years and older, is expected to double from 962 million in 2017 to nearly 2.1 billion by 2050 according to "World Population Aging: Highlights 2017" [3]. This rise in longevity is expected to have significant health and social impacts. There will be an increased demand for health services due to higher rates of aging-related obesity and chronic diseases including cardiovascular disease, dementia, cancer, and diabetes. This increased demand will result in challenges for public health and social and economic stability [1]. In addition to environmental factors (for example, lifestyle, stress, nutrition, and air and water quality) that play a large role in determining lifespan, genetic factors also affect the process of aging and longevity, both in humans and in other organisms [4–7]. Knowledge of the involved genetic factors may facilitate identification of individuals at risk for aging-related diseases and/or development of prophylaxis.

Although the process of aging in humans is still not fully understood, evidence from model organism studies suggests the involvement of numerous candidate genes in longevity. Several biological pathways—including inflammation, oxidative stress and stress response, DNA damage and repair, cellular senescence, and the insulin-like growth factor (IGF, a table of abbreviations can be found in **Appendix A**) pathway—appear to play a key role in the process underlying longevity [5, 8]. The IGF pathway is an evolutionarily conserved pathway that influences the

process of aging among eukaryotes [9]. The pathway is involved with mammalian growth, development, and metabolism and locally affects cell proliferation, differentiation, cell migration, and survival [10]. Diverse species, including yeast, worms, fruit flies, and mice, have mutations in genes in the IGF pathway that have been associated with lifespan expansion and similar aging-related characteristics [9]. Some of these shared phenotypes include decreased IGF-1 levels, reduced insulin signaling, and increased sensitivity to insulin. The role of genetic variation in the IGF pathway in longevity, however, is inconclusive in humans. Studies have suggested that genetic variation in human loci within the IGF pathway is associated with variation in plasma IGF-1 levels and that such associations were more apparent in long-lived participants (median age, 99 years) than in younger participants (median age, 60 years) [11]. Studies of Ashkenazi Jewish centenarians suggest that reduced activity in the IGF pathway due to variants in genes within the pathway—in particular, *IGF1R*—is associated with longevity [12].

Additional studies of the genetic and environmental factors influencing the IGF pathway may provide insights into the underlying physiology involving lifespan and healthy aging. For this reason, several population-based studies have investigated the association of serum levels of IGF-1 (the main ligand in the IGF pathway) with either mortality or age-related chronic diseases. Many epidemiological studies have reported an association between IGF-1 serum concentration and elevated risks of type 2 diabetes [13], cancer [14], cardiovascular disease [15], and mortality [16, 17]. In addition, because IGF-1 has a wide range of biological effects on different cell types and it functions as a mediator of most of the anabolic effects of growth hormone (GH), IGF-1 has been proposed as therapeutic agent for many conditions, such as diabetes, obesity, GH resistance, and osteoporosis [10]. However, such therapeutic use may increase the risk of IGF-1 side effects, such as asystole and hypotension (from short-term IGF-1 administration) or hypoglycemia, tachycardia,

and facial and hand edema (from multiple subcutaneous IGF-1 injections) [10]. Knowledge of the genetic and environmental factors that influence IGF-1 levels may enable the development of more targeted therapies.

To date, only one linkage analysis of IGF-1 levels has been performed [18], but four genome-wide association studies have been reported [19–22]. Ten SNPs associated with serum IGF-1 levels and its main binding protein, IGF binding protein 3 (IGFBP3), were identified by Teumer et al. [21], of which four had been previously identified in Kaplan et al. [20]. Five of these identified SNPs were in loci known to be associated with longevity, age-related, and metabolic traits [21]. These findings may indicate possible pleiotropic effects on serum IGF-1 levels, metabolic traits, and the aging-related process.

Further exploration of the genetic architecture of serum IGF-1 levels in a unique family study of exceptional longevity, such as the Long Life Family Study (LLFS), might identify additional loci influencing variation in IGF-1 levels. The identification of such loci may provide insights about the involvement of some shared genetic variants in the association between serum IGF-1 variation and metabolic traits, and aging-related process. In addition, because participants in the LLFS are older than those in previous GWASs (the mean age of the proband generation is 90.8 years), such studies may provide insights regarding the genetic relationship between IGF-1 levels and aging.

1.2 Background and significance

1.2.1 Physiological role of IGF-1

Insulin-like growth factor-1 (IGF-1) or somatomedin C, has structural homology to insulin and characteristics of both a circulating hormone and a local tissue growth factor [23]. IGF-1 is a member of the IGF pathway, an axis that includes three ligands (insulin, IGF-1, and IGF-2), three receptors (the insulin receptor [IR], the IGF-1 receptor [IGF-1R], and the mannose-6-phosphate/IGF-2 receptor [M6P/IGF-2R]), and six high-affinity insulin-like growth factor binding proteins (IGFBP1–6) [23]. In addition, IGF-1 circulates as a ternary complex bound to IGFBPs and the IGFBP acid-labile subunit (IGFALS). IGF-1 and -2 are approximately 70% identical. IGF-1 circulates as a single-chain peptide of 70 residues in four domains (A–D); whereas IGF-2 circulates as a single-chain peptide of 67 residues. They also share about 50% identity in domains A and B with human insulin [23, 24]. While IGF-1 starts being expressed just after birth, IGF-2 is mainly expressed prenatally, and its main beneficial functions are fetal development and placental growth [24]. In addition, IGF-2 is not affected by GH stimulation and thus has little effect on somatic growth postnatally. Although IGF-1 and -2 are both synthetized and released by the liver, IGF-2 production is not under the control of growth hormone (GH). In addition, IGF-1 and -2 are the production of two distinctive genes [25]. *IGF1* is located on chromosome 12, and *IGF2* is located on chromosome 11.

Circulating IGF-1 is produced mainly, but not exclusively, by the liver due to GH stimulation [24]. Releasing IGF-1 systemically as an endocrine hormone serves to mediate GH actions in promoting growth, development, and metabolism. GH is a major factor that regulates IGF-1 hepatic biosynthesis. Insulin and nutritional status (as dietary protein and long-term caloric

restrictions) also regulate IGF-1 [26]. Serum concentration of IGF-1 is low at birth, increases during childhood and puberty, and reaches its highest concentration during early adulthood. It then starts to decline in the third decade of human life [10].

Extrahepatically, IGF-1 is produced by most body tissues, for instance fat tissue, to perform its paracrine/autocrine action in promoting cellular growth, differentiation, and apoptosis [24]. Extrahepatic IGF-1 expression is regulated by prostaglandin E_2 (PGE_2), parathyroid hormone (PTH), and thyroid hormone (TSH), as well as GH [23]. In addition to the main actions of IGF-1 on cell progression, proliferation, differentiation, and cell death, it is involved in hormone and neurotransmitter secretion (e.g., negative feed-back on GH secretion), amino acid and glucose uptake, chemotaxis/chemokinesis (e.g., increase the chemotactic migration in T lymphocytes), immune response, and overexpression of IGF-1 receptors [27]. Anabolically, IGF-1 affects glucose metabolism (increased peripheral glucose uptake and decreased hepatic glucose production), fat metabolism (decreased serum ketones, free fatty acid [FFA], and triglycerides), and renal function (increased glomerular filtration rate, and renal plasma flow) [27].

IGF-1 also regulates protein metabolism by increasing protein synthesis, decreasing nitrogen excretion, and increasing total body protein accretion [27]. The anabolic effect of IGF-1 is strongly affected by individual nutrient status. After eating, GH secretion is suppressed, and insulin secretion increases. This promotes glucose translocation into the skeletal tissue and stimulates glyco- and adipogenesis [28–30]. When fasting, insulin secretion decreases, and GH (through IGF-1) promotes lipolysis and thereby increases FFA release and hepatic glucose uptake. In addition, when food is spare, the GH/IGF-1 axis switch from the use of carbohydrate and protein to the use of lipids as fuel [28, 30].

The six IGFBPs, which are found in both the circulating and extracellular fluid, have been reported to have high affinity to IGF-1 and -2 [27]. The IGFBPs act as IGF-1 carriers: playing roles in transporting IGF-1 from the circulation to the peripheral tissue, maintaining the IGF-1 levels in the circulation by prolonging the IGF-1 half-life, mediating the IGF-independent biological effect, and inhibiting or potentiating IGF-1 action [24]. IGFBP3, the largest IGFBP, is the predominant circulating binding protein for IGF-1; about 90% of circulating IGF-1 binds to it [24, 27].

The high affinity binding of IGF-1R to IGF-1 and the widespread expression of it in almost all human tissues and cell types allows for the initiation and promotion of physiological actions of IGF-1 all over the body [24, 27]. In addition to its primary activating response to insulin, insulin receptor can also be activated by IGF-1 to initiate an intracellular response due to this ligand's structural similarity to IGF-1R (both are tyrosine kinases).

1.2.2 IGF-1 and the risk of mortality

Reports of the relationship between IGF-1 and all-cause mortality or cardiovascular mortality are contradictory. A U-shaped relationship has been reported in several studies between IGF-1 levels and all-cause mortality, cancer mortality, and cardiovascular mortality in the general population [15–17, 31–35]. Other investigators have concluded that there is no association between IGF-1 serum levels and all-cause mortality, but suggested a greater association between mortality and levels of IGFBP1 and IGFBP2 [36–39]. These conflicting results may be due to confounding factors, such as the presence of underlying disease, obesity status, age, and the time to follow-up. Other factors that are difficult to measure and adjust for, such as poor nutritional status, immobility,

and muscle weakness, which is generally associated with lower IGF-1, may also confound the association between IGF-1 and mortality.

1.2.3 Serum IGF-1 levels and BMI

Body mass index (BMI) is a strong mediator of the risk for chronic disease development associated with aging. Epidemiological studies have reported associations between BMI and all-cause mortality, type 2 diabetes, cancer, and cardiovascular disease, in addition to the association of BMI with insulin and insulin-like growth factors [40]. The relationship between BMI and IGF-1, however, is neither consistent nor clear in most of the studies' results. Some studies observe a negative correlation between IGF-1 and BMI regardless of age, sex, or race [18, 41–45]. Others, however, suggest a positive association [36, 46]. Yet others reported no association [47]. The data on the relationship between BMI and IGF-1 are obviously inconsistent and not fully understood (**Table B1 Appendix B**). Studies that reported a negative relationship were primarily performed on participants between the ages of 10 y and 60 y. In contrast, studies that observed a positive association between BMI and IGF-1 were performed on older individuals with a range in age from 45 y to 98 y. In addition, several of these studies were small or were comprised of a highly selected group, such as obese/overweight individuals who might be experiencing weight-related disruption in insulin and GH secretion [44]. Furthermore, most of these studies measured the relationship between IGF-1 and BMI by stratifying the samples based on BMI categories, not age, although the study participants' ages varied widely. To date, no study has assessed the influence of age on the association between BMI and IGF-1.

1.2.4 Genetic epidemiology of serum IGF-1 levels

IGF-1 is a product of a single gene (*IGF1*) located on 12q. *IGF1* consists of six exons that can be alternatively spliced to create various transcripts, each encoding different pre–IGF-1 proteins. All transcript isoforms yield the same mature IGF-1 peptide after processing and use the same receptors [32, 48].

The heritability of serum IGF-1 levels is estimated to range from 38% to 63% [18, 26, 49, 50]. A genome-wide association study done by Kaplan et al. [20] was the first study to use the GWAS approach to identify nucleotide polymorphisms (SNPs) associated with serum IGF-1 and IGFBP3 levels. Genotype and phenotype data were available on 10,280 men and women of European origin who were enrolled in four community-based cohort studies from the USA and Germany (CHS, FHS, KORA, and SHIP). The mean age of participants in this GWAS study ranged from 39.9 y to 71.5 y. The main findings are summarized in **Table 1**. A novel and independent locus on chromosome 7p12.3 (rs700752), located 790 kbp upstream of *IGFBP3*, was significantly associated with higher IGF-1 ($p = 4.9 \times 10^{-9}$) levels and higher IGFBP3 ($p = 4.4 \times 10^{-21}$) [20]. They also identified a SNP 50 kbp upstream of *IGFBP3* on 7p12.3 (rs11977526) that was strongly associated with higher IGFBP3 concentration. This SNP was also associated with lower IGF-1 levels after adjusting for IGFBP3 levels. A third SNP, 18 kbp upstream of *IGFBP3* (rs1496499), was strongly associated with IGF-1 levels after adjustment for IGFBP-3. These results might indicate that the binding affinity between IGF-1 and IGFBP3 can affect the detection of associations.

Other SNPs had marginally significant associations with serum IGF-1 levels: (1) rs1245541 on 10q22.1, which is a region that has been associated with late-onset Alzheimer

disease [51], and (2) rs2153960 on 6q21, which is located within *FOXO3* and has been linked with longevity [52].

In addition to results of analyses on IGF-1, two additional loci were significantly associated with IGFBP3 only: (1) rs4234798 on chromosome 4p16.1, inside *SORCS2*, a novel finding; and (2) rs1065656 on chromosome 16p13.3, within *NUBP2* and downstream of *IGFALS*, a confirmatory finding. Some of these genes, such as *IGFBP3* and *IGFALS*, were already known to be involved in the IGF regulatory or signaling pathways [53].

Table 1. Significant findings of previous GWASs of IGF-1.

SNP	Chr	Phenotype	Nearby Gene	Study
rs780093	2	IGF-1	*GCKR*	Teumer et al. 2017
rs934073	2	IGF-1	*ASXL2*	Teumer et al. 2017
rs509035	3	IGF-1	*GHSR*	Teumer et al. 2017
rs4234798	4	IGFBP3	*SORCS2*	Kaplan et al. 2011, Teumer et al. 2017
rs2153960	6	IGF-1	*FOXO3*	Teumer et al.
rs11977526	7	IGFBP3	*IGFBP3*	Kaplan et al. 2011, Teumer et al. 2017
rs700752	7	IGFBP3, IGF-1	*TNS3, IGFBP3*	Kaplan et al. 2011
rs700753	7	IGFBP3, IGF-1	*TNS3, IGFBP3*	Teumer et al. 2017
rs978458	12	IGF-1	*IGF1*	Teumer et al. 2017
rs1065656	16	IGFBP3, IGF-1	*NUBP2, IGFALS*	Kaplan et al. 2011, Teumer et al. 2017

The most recent GWAS meta-analysis was done in 2017 [21] using data on 14,424 men and 16,460 women (total 20,884 individuals) of European ancestry from twenty-one studies, The mean age of participants across these twenty-one studies ranged from 18.9 y to 76 y. Teumer and colleagues identified several new genome-wide significant loci associated with circulating IGF-1 in or near *GCKR*, *IGF1*, *FOXO3*, *ASXL2*, *NUBP*, and *GHSR* in addition to rs700753, 800 kbp upstream of *IGFBP3* and 575 kbp downstream of *TNS*; rs700753 is in perfect linkage disequilibrium with the previously reported SNP, rs700752, in individuals of European ancestries in the 1000 Genomes Project.

The IGF-1-associated SNPs were also examined in association with anthropometric and age-related traits. For rs780083 in *GCKR*, the allele that was associated with higher IGF-1 levels was also associated with higher levels of fasting glucose, fasting insulin, homeostatic model assessment for insulin resistance (HOMA-IR), and elevated risk of type 2 diabetes. It was also associated with greater height, lower waist–hip ratio, and higher spine bone mineral density. rs934073 near *ASL2* was associated with decreased IGF-1 was associated with survival beyond 90 y, higher levels of BMI, and lower lumber spine bone mineral density [21]. These results suggest strong concordance between SNPs that influence serum IGF-1 levels and other age-related (e.g., diabetes, bone mineral density, and survival) traits.

1.2.5 Specific Aims

An increased understanding of the roles that IGF-1 plays in both physiological and pathological conditions could lead to better characterization of the use of IGF-1 as a safe potential therapeutic agent or as a biomarker to categorize individuals at high risk of age-related diseases and mortality. However, the relationship between IGF-1 and all-cause mortality and cardiovascular mortality [31, 34, 38, 39] as well as the relationship between IGF-1 and BMI as a measure of adiposity is neither consistent nor clear across studies [18, 36, 38, 41, 42]. The well-known role that the IGF pathway plays in the underlying process of longevity in animal models [7], as well as its apparently contradictory relationship with measures of aging [54], supports additional studies seeking genetic variants that are associated with variation in serum IGF-1 levels. Genetic studies thus far have identified several novel candidate genes that are associated with IGF-1 serum variation, however, whether this genetic variation affects disease risk and disease endpoints has not been determined.

The two-generation family members who participated in the Long Life Family Study range in age from 24 y to 110 y, enabling studies of age-related effects on a variety of traits. In addition, family members have been followed over time, thus longitudinal analyses can be done to assess changes over time versus genotypes and/or, in the older generation, with mortality. Furthermore, because LLFS is a family study, relationships among traits and genotypes within the proband generation can be compared with results in the offspring generation. This structure will enable the assessment of genetic and environmental factors affecting a variety of age-related traits, as well as the effects on mortality.

To address several of these gaps and contradictions in the literature using data on a unique population of families, I assessed three main research questions (Specific Aims):

(1) Does the relationship between serum IGF-1 levels and BMI vary in an age- or sex-specific manner? To address this question, I will
    a) Compare the relationship in other cohorts
    b) Assess whether the relationship differs by race and ethnicity
    c) Assess the relationship between other adiposity measurements
(2) What is the genetic architecture of serum IGF-1 levels in LLFS? To address this question, I will
    a) Estimate the heritability of serum IGF-1 levels and BMI in LLFS and the genetic correlation between these two traits.
    b) Identify possible quantitative trait loci (QTL) associated with serum IGF-1 level in LLFS using quantitative linkage analyses

c) Identify possible SNPs associated with IGF-1 levels in LLFS using genome-wide association and replicate those SNPs using data from the Framingham Heart Study (FHS)

(3) What is the relationship between serum IGF-1levels and risk of mortality? To address this question, I will

a) Estimate the probability of survival in LLFS

b) Assess whether the survival proportion differs between gender or among serum IGF-1 tertiles

c) Determine whether serum IGF-1 levels or specific genetic variants predict the risk of mortality

Chapter 2 describes methods used for all the aims and sub-aims. The results of Aim1 are described in Chapter 3. In Chapter 4, I describe the results from Aim 2 (heritability, genetic correlations, and linkage analyses). In Chapter 5, results from the GWAS in Aim 2 are presented, as well as attempted replication of interesting findings in the Framingham Heart Study. In Chapter 6, I describe the results of the Aim 3 mortality analyses. Finally, I present my overall discussion and conclusions in Chapter 7.

2.0 ANALYTIC MATERIAL AND METHODS

One of the aims of my project is to assess the relationship between serum IGF-1 level with BMI (as a moderator of risk) and the risk of mortality in a unique cohort of families that were recruited based on exceptional healthy longevity. In addition, I assessed whether genetic factors might affect these associations. To achieve these aims I first assessed the relationship between IGF-1 and BMI within age quartiles using LLFS participants' data and then replicated the analysis using the third National Health and Nutritional Examination Survey (NHANES III) participants' data. Secondly, using LLFS data, in addition to assessing the phenotypic association between IGF-1 and BMI, I estimated the genetic correlation between the two variables both overall and by generation. Then, I determined the heritability of IGF-1 in LLFS and I used linkage analysis to determine whether quantitative trait loci influenced serum IGF-1 levels in these families. Third, I performed a GWAS of serum IGF-1 levels using the genotyped data of LLFS participants; all SNPs with a suggestive level of association were then tested for possible replication using Framingham Heart Study (FHS).

2.1 Study populations

2.1.1 Long Life Family Study

LLFS is an international multi-center family-based cohort study designed to determine the genetic and behavioral/environmental risk factors that promote exceptional survival, longevity, and healthy aging [56]. The families were recruited through elderly probands (in their 90s) from across

the USA and Denmark at four collection sites (New York, Boston, and Pittsburgh in the United States and in Denmark) who self-reported on the survival history of their parents and siblings. Families that showed clustering of exceptional survival and Family Longevity Selection Score (FLoSS) ≥ 7 were recruited [55]. FLoSS, which combined the estimated exceptionality of survival with a bonus for living sibling(s), is a measure of exceptional survival that was used as a tool of inclusion by scoring and selecting families for the study. If a proband's family was FLoSS-eligible then additional enrollment criteria had to be met; the proband, at least one living sibling, and one of their offspring (minimum family size of 3) were able to give informed consent, and were willing to get interviewed and be examined including the collection of a blood sample for serum and DNA extraction [56]. The total number of enrolled family members at baseline was 4,953 (99% European ancestries) in 539 families with age range from 24.0 y to 110.0 y (proband age range = 49 y–110 y, and offspring and their spouses age range = 24 y–88 y) [57]. Participants had their first in-person home visit between 2006 and 2009, then were contacted annually by telephone to update vital status, medical history, and general health. A second in-person visit was completed between 2014 and 2017, which was then followed up annually by telephone. The annual telephone follow-up is still ongoing, and plans for third in-person visit are in progress [57]. The key characteristics of LLFS participants from visit 1, which is the source of data used in this analysis, can be found in **Table 2**.

Table 2. Descriptive statistics for LLFS.

	LLFS (n = 4270)	
	Male (45.3%)	Female (54.7%)
Total Participants	1936	2334
Probands	641	764
Offspring	895	1230
Measurements · mean (s.d.)		
Age at Enrollment (y)	70.6 (15.3)	69.5 (15.9)
IGF-1 (ng/ml)	134.6 (54.0)	123.2 (51.6)
BMI (kg/m^2)	27.5 (4.0)	26.7 (5.3)
Height (cm)	173.6 (7.7)	159.6 (7.8)
Weight (kg)	83.2 (14.9)	68.2 (14.9)
WC* (cm)	99.5 (11.0)	90.3 (13.9)
Smokers	3%	4%
Hypertension	24%	27%
Diabetes	4%	3%
Stroke	3%	4%
Heart Disease	6%	3%

*WC: waist circumference

2.1.2 NHANES III

NHANES III is a sample of ~39,000 participants aged 2 months and older and was designed to be representative of the US population. It was conducted from 1988 to 1994 in two phases. Of the total sample of adults (n = 20,024), we selected a subset of 2,555 non-Hispanic White participants (20 y–90 y) with a complete record of the study variables. In addition, a few analyses were conducted using data from non-Hispanic Black participants (n = 1,639, 20 y–90 y) and Mexican American participants (n = 1,607, 20 y–90 y).

2.1.3 Framingham Heart Study

The Framingham Heart Study (FHS) was first established in 1948 as a community-based study investigating cardiovascular risks of adult from the town of Framingham, Massachusetts. It currently includes data on three generations of participants [58]. To replicate the results from my GWAS of IGF-1 in LLFS, genotype and phenotype data from participants in the FHS offspring exam 7 cohort was analyzed, because IGF-1 levels were measured in this cohort. The 2,833 participants (45.7% men and 54.3% women, mean age = 61.1 y ± 9.5 y) were recruited from 2002 to 2004. Per the agreement between LLFS and FHS investigators, I wrote an analysis plan for replication of all suggestive SNPs from the GWAS analysis performed in LLFS and submitted it to LLFS Data Coordinating Center, and they performed the replication analysis.

2.2 Serum IGF-1 assays

In LLFS, fasting peripheral blood samples were obtained following a standardized venipuncture protocol by staff at in-person visit 1 [55]. Approximately 50 mL of blood was collected in serum tubes and kept at room temperature for 30 min to 45 min before centrifugation. The centrifuged serum tubes and other unprocessed blood tubes were shipped to the Advanced Diagnostics and Research Laboratory (ADRL) at the University of Minnesota. IGF-1 was measured by the ADRL in serum using a solid-phase enzyme-linked chemiluminescent immunoassay on an Immulite 2000 system (Siemens Healthcare Diagnostics, Inc.). The inter-assay coefficient of variability was 8.7%.

In NHANES III, fasting serum samples were collected from 1988 to 1994, and IGF-1 concentrations were quantified by IGF-I enzyme-linked immunosorbent assay (DSL 10-5600)

including an extraction step which separates IGF-1 from its binding protein. The samples were reanalyzed if the coefficient of variation for replicate samples was greater than 15% [59]. In the current study, given the differences in the assays used for serum IGF-1 measurements in LLFS and NHANES III, we compared the relationships among traits within and not across studies.

In FHS, fasting blood sample were collected according to protocol [18]. Samples were centrifuged and aliquoted immediately for storage at −70 °C. Standardized enzyme-linked immunosorbent assay (ELISA) was used to measure serum IGF-1. Quality control measures according to a strict protocol were done, and the intraassay coefficient of variation for serum IGF-1 was 5.3%.

2.3 Covariate selection

Based on published studies and biological plausibility with serum IGF-1 levels, the following covariates were considered for analysis; age (and age^2), sex, BMI, waist circumference (WC), weight, height, medication history, insulin level, smoking status, diabetes status, hypertension history, and measures of nutritional status (diet group: meat-eater, vegetarian, or vegan) and physical activity status (inactive, moderately inactive, moderately active, active). Several of these covariates were not measured in LLFS, i.e., nutritional status and physical status. BMI, WC, weight, height, and insulin levels were measured directly. Age, sex, medication history, and smoking status were self-reported. Variables such as history of hypertension and diabetes status were inferred from each participant's self-reported diagnosis, their list of current medications, and/or direct measurements of blood pressure or fasting glucose, respectively. Next, ignoring the familial dependency of the observations, bi-directional stepwise multiple regression was

performed to identify a subset of covariates to be included in the subsequent, relationship-adjusted, analyses. Interaction terms age × BMI and sex × BMI and LLFS field center (Boston, New York, Pittsburgh, Denmark) were also included as possible covariates.

In the analysis to assess the relationship between IGF-1 and BMI (**Section 3.3.4**), the following covariates were remained in the model and were included: age, age^2, sex, BMI, age × BMI, sex × BMI, history of hypertension, history of diabetes and LLFS field centers. In the relationship-adjusted genetic analyses, the covariates were age, sex, BMI, field center, and age × BMI (**Chapter 5)**.

2.4 IGF-1 and BMI by age (Aim 1)

All analyses for Aim 1 were performed using R version 3.4.0 [60]. Prior to analysis, IGF-1 levels and BMI were log-transformed to approximate normality and for consistency with previous reports. Phenotype data on 4,241 LLFS participants (1,391 probands, 2,119 offspring, and 731 offspring's spouses) were included. To assess the correlation between serum IGF-1 levels and adiposity measures (BMI, weight, WC), in addition to age and height, Pearson correlation coefficients were estimated. To account for non-independence due to relatedness, bootstrap Pearson correlation coefficients were estimated using the R *boot* package.

Linear regression of IGF-1 levels on BMI was performed with the sample non-stratified and stratified by sex. I used in linear mixed-model regression as implemented by the lmekin() function of the R *coxme* package with age, age^2, male sex, BMI, field center, diabetes, hypertension, BMI × age, and BMI × sex as fixed-effects covariates. lmekin()accounts for the

relatedness between the participants by incorporating a kinship matrix as a random-effects covariate.

To assess the relationship between IGF-1 and BMI by age, the overall sample was divided into age quartiles, and linear regression of IGF-1 on BMI was performed within each age quartile. Each age quartile was further stratified into male and female participants, and linear regression was performed to assess if the relationship between IGF-1and BMI differed between sexes.

Parallel analyses were performed with 2,555 non-Hispanic White, 1,639 non-Hispanic Black, and 1,607 Mexican American participants from NHANES III. These analyses will allow me to assess whether the relationship between IGF-1 and BMI differed among racial/ethnic groups.

2.5 Heritability of IGF-1 and the genetic correlation between IGF-1 and BMI (Aim 2A)

To estimate the proportion of variance in IGF-1 levels due to additive genetic effects, heritability analysis was performed using Sequential Oligogenic Linkage Analysis Routines (SOLAR) [61], a pedigree-based maximum-likelihood method, adjusting for age and sex. Heritability was estimated in each of the proband and offspring generations of the LLFS. To determine the genetic correlation between IGF-1 and BMI, bivariate heritability analysis was performed using SOLAR in the total study sample and stratified by generation, adjusting each analysis for age and sex.

2.6 Genotyping, Linkage Markers, and Imputation

The Center for Inherited Disease Research (CIDR) assayed all LLFS subjects using Illumina Human Omni 2.5 v1 genotyping arrays [62]. Quality control (QC) of the genotypes and imputation were carried out by LLFS investigators in the Division of Statistical Genomics, Washington University in Saint Louis. Genotype QC included checking of Mendelian errors and verification of reported pedigree relationship using GRR (graphical representation of relationship errors). A set of 3,647 SNPs with high Mendelian error rates and 83,774 SNPs with a call rate < 98% per marker were dropped. Eighteen subjects with call rates < 97.5% were also dropped. Finally, 153,363 genotypes flagged as Mendelian errors were set to missing. Additional genotypes were imputed based on the cosmopolitan-phased haplotypes of the Haplotype Reference Consortium (HRC) [63]. SNPs with imputation quality scores $r^2 < 0.3$ and those that were not in Hardy–Weinberg equilibrium (exact test p value < 0.0001) were dropped from analyses. The final count of observed and imputed SNPs examined in the GWAS was 9,354,374 on the autosomes and 286,048 on chromosome X.

For linkage markers, the LLFS Coordinating Center generated 6,570 multi-allelic haplotypes from genotype data on 1–5 adjacent assayed SNPs on chromosomes 1–22. The mean centimorgan position of the SNPs in each set was used as the position for the haplotype. The mean distance between haplotypes was 0.5 cM with a maximum of 3.49 cM and a median of 0.5 cM. These haplotypes were used to generate sex-specific multipoint identity-by-descent (IBD) matrices.

In FHS, genotyping was performed on the Affymetrix GeneChip Human Mapping 500K Array Set and 50K Human Gene Focused Panel. The mean call rate was 98%, and the quality

control filtration included removal of SNPs with call rate < 95% or a Hardy–Weinberg equilibrium $p < 10^{-6}$ [20]. Phased 1000 Genomes genotypes were used for the imputation reference panel.

2.7 Quantitative Trait Linkage (QTL) analyses (Aim 2B)

Quantitative trait linkage analyses were performed to identify possible identity-by-decent (IBD) allele shared as quantitative trait loci (QTLs) for serum IGF-1 levels in LLFS participants. Multipoint QTL analyses were done using SOLAR [61], adjusted for age and sex. The result was reported as logarithm of the odds (LOD) score, with a suggestive cutoff of 3 or significant cutoff of 3.3. The corresponded significant interval range was located against the sex-averaged human genetic map (~3,337 cM).

2.8 GWAS and candidate gene analysis (Aim 2C)

To determine whether specific genetic variants, i.e., SNPs, were associated with IGF-1 serum levels, a GWAS was performed over all LLFS individuals using data on both assayed and imputed genotypes with a minor allele frequency (MAF) ≥ 0.004 (9,354,374 autosomal SNPs and 286,048 chromosome X SNPs). Using the R *GENESIS* package [64], I performed mixed model regression to evaluate the association between each genetic variant and serum IGF-1 levels, accounting for both relatedness and population stratification. Covariates used in the model were age at enrollment, sex, BMI, field center, and age × BMI. SNPs were coded using linear additive scoring of 0, 1, or

2 copies of the minor allele of each SNP. A significant threshold of $p < 5 \times 10^{-8}$ was genome-wide significant, and $p < 5 \times 10^{-6}$ was a suggestive level of significance.

I also extended the analysis to test for the association between serum IGF-1 levels and independent SNPs within genes that were either were associated with IGF-1 in previous GWASs [21] (**Table 1**) or have been reported as candidate genes due to their membership in the IGF pathway (*IGF1*, *IGFALS*, *SST*, *SSTR5*) [53]. In total, independent SNPs within nine genes (*IGF1, FOXO3, NUBP2, GHSR, TNS3, SST, SSTR5, IGFALS, IGFBP3)* were assessed for association. A conservative Bonferroni correction was used for each gene separately (0.05 ÷ number of SNPs per gene) to obtain the *p* value cut-off for statistical significance.

To replicate the result from the GWAS and candidate gene analysis in an independent sample, an analysis plan and the list of SNPs (with MAF > 0.004 and discovery $p < 5 \times 10^{-6}$) were submitted to LLFS Data Coordination Center to be tested using the data of the FHS participants (**Section 2.1.3**). A generalized linear model (GLM) was used as the statistical method. Again, Bonferroni correction was used to estimate a significant *p* value for replication.

To assess the involvement of IGF-1 associated variants with gene regulation, the SNPs that were statistically significant in the GWAS and the SNPs that were suggestively significant in the GWAS and significant upon replication were examined for evidence that they are expression quantitative trait loci (eQTLs). The Genotype–Tissue Expression Program (GTEx) is a publicly available data resource with reports on statistical tests of association between genetic variants and gene expression in an array of tissues. Information on gene expression from all available human tissues in GTEX v8 [65] and HaploReg v4.1 [66] will be used to annotate the GWAS results.

2.9 Survival analysis and the risk of mortality (Aim 3)

All analysis were performed using R version 3.4.0 (R Core Team, 2019). First, I performed Kaplan–Meier analyses to estimate the probability of survival among the LLFS proband participants (n = 1,482) over ten years of follow-up. Second, to compare the survival proportion difference between male (n = 668) and female (n = 814) probands and between low, high, and median IGF-1 levels, I performed a log-rank test for differences in survival curves between the sexes and between IGF-1 tertiles. Third, to test for association between serum IGF-1 levels (as a continuous variable) and the risk of mortality, I performed Cox (proportional hazards) regression. I also used Cox regression to test for differences in mortality between IGF-1 tertiles. Lastly, to assess whether the risk of mortality can be predicted by IGF-1–associated SNPs, their individual effects on mortality risk was calculated using Cox regression and the hazard ratios (HRs) for mortality (death status 0 or 1) were estimated. SNPs were coded using linear additive scoring of 0, 1, or 2 copies of the minor allele of each SNP.

3.0 RELATIONSHIP BETWEEN SERUM IGF-1 AND BMI DIFFERS BY AGE

Rehab A. Sherlala, MPH[1], Candace M. Kammerer, PhD[1], Allison L. Kuiper, PhD[2], Mary K. Wojczynski, PhD[3], Svetlana V. Ukraintseva, PhD[4], Mary F. Feitosa, PhD[3], Jonas Mengel-From, MSc, PhD[5], Joseph M. Zmuda, PhD[2], Ryan L. Minster, PhD, MSIS[1]

[1] Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

[2] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

[3] Department of Genetics, Washington University School of Medicine in St. Louis, MO

[4] Social Science Research Institute, Duke University, Durham, NC

[5] Molecular Epidemiology, IST–EBB/Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark

On behalf of the Long Life Family Study

Correspondence:

Rehab A. Sherlala, MPH

Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

E-mail: ras148@pitt.edu

3.1 Introduction

Insulin-like growth factor 1 (IGF-1) is a member of the IGF pathway [23], which appears to play a key role in the processes underlying longevity [67]. Many epidemiological studies report that serum IGF-1 levels are associated with elevated risk of type 2 diabetes [13], cancer [14, 68], cardiovascular disease [16, 69], and mortality [15, 17]. IGF-1 has structural homology to insulin and characteristics of both a circulating hormone that mediates growth hormone (GH) actions in promoting growth, development, and metabolism [23] and a local tissue growth factor that promotes cellular growth, differentiation, and apoptosis [24]. Serum IGF-1 levels are heritable, with estimated heritability ranging from 38% to 63% [18, 26, 49, 50], and are influenced by obesity, age, sex, physical activity, GH level, and nutritional status [26, 70]. Across the lifespan, serum IGF-1 levels are low at birth, increase during childhood and puberty, and reach their highest concentration during early adulthood then start to decline in the third decade of life [10].

Body mass index (BMI) is also strongly associated with risk for chronic disease development associated with aging [40]. Because both IGF-1 and BMI are associated with disease risk and disease endpoints, several studies have assessed the relationship between them. Understanding the relationship between these two predictors could help categorizing those at risk of disease development or event. However, the relationship between BMI and IGF-1 across studies is neither consistent nor clear. Several studies have reported that IGF-1 levels are inversely correlated with BMI [18, 41–45]; whereas others report a positive correlation [36] or no correlation [47]. Most of these studies estimated the relationship between IGF-1 and BMI by stratifying the samples based on BMI categories, and not age, although the study participants' ages varied widely [42, 44, 45, 47].

To date, no study has assessed the influence of age on the association between BMI and IGF-1. We hypothesize that the relationship between IGF-1 and BMI varies by age.

Here we present a cross-sectional study of the relationship between IGF-1 and BMI in a large sample of 4,241 participants from the Long Life Family Study (LLFS) with validation of the relationships in a large sample of 2,555 participants from the third National Health and Nutritional Examination Survey (NHANES III). In particular, we assessed whether the relationship between IGF-1 and BMI varies in an age- and sex-specific manner.

3.2 Method

3.2.1 Study Population

The primary sample for this study is a set of participants from LLFS. LLFS is a multi-center family-based cohort study of 539 families that was designed to determine the genetic and behavioral/environmental risk factors that promote exceptional longevity [56]. The families were recruited between 2006 and 2009 from the USA and Denmark at four enrollment sites (New York, Boston, and Pittsburgh in the United States and nationwide in Denmark). The total number of enrolled participants is 4,953, consisting of long-lived probands and their siblings (n = 1,727), the offspring of this generation and their spouses (n = 3,226). Participants without measurements of serum IGF-1 levels or BMI were excluded, therefore, the total sample size for this analysis is 4,241 participants (aged 24 y–110 y) consisting of 1,391 from the proband generation (49 y–110 y), 2,119 from the offspring generation (32 y–87 y), and 731 offspring spouses (24 y–88 y). All participants self-identified as non-Hispanic White.

The findings in LLFS were replicated using participants from NHANES III ("NHANES III (1988-1994)," n.d.). NHANES III is a sample of ~39,000 participants aged 2 months and older and was designed to be representative of the US population. It was conducted from 1988 to 1994 in two phases. Of the total sample of adults (n = 20,024), we selected a subset of 2,555 non-Hispanic White participants (20 y–90 y) with complete record of the study variables. Parallel analyses were conducted with non-Hispanic Black participants (n = 1639, 20 y–90 y) and Mexican American participants (n = 1607, 20 y–90 y).

3.2.2 Participant characteristics

In both LLFS and NHANES III, standing height, weight, and WC were assessed by trained interviewers with a standardized protocol and skill level. BMI was calculated as weight in kilograms per the square of the height in meters. Age, race, ethnicity, and sex were taken by self-report during the interview. For the analysis in LLFS, presence of diabetes was defined as use of diabetes medications or fasting glucose ≥ 126 mg/dL. Presence of hypertension was defined as SBP ≥ 140 mmHg or DBP ≥ 90 mmHg or self-report confirmed by use of antihypertensive medication. For NHANES III, presence of diabetes was defined as history of diabetes diagnosis, or fasting glucose ≥ 126 mg/dL, or current use of oral hypoglycemics or insulin. Presence of hypertension was defined if a participant reported both ever being told that he or she had high blood pressure and current use of antihypertensive medication, or if the average measured BP was ≥ 140 mmHg systolic, or ≥ 90 mmHg diastolic. We excluded smoking status as a covariate because of the large number of missingness within the NHANES III participants, in addition to the non-significant association between IGF-1 and smoking status in LLFS.

3.2.3 Laboratory Assays

In LLFS, fasting peripheral blood samples were obtained from participants and then shipped to the Advanced Diagnostics and Research Laboratory at the University of Minnesota [72]. IGF-1 was measured in serum using a solid-phase enzyme-linked chemiluminescent immunoassay on an Immulite 2000 system (Siemens Healthcare Diagnostics, Inc.). The inter-assay coefficient of variability was 8.7%.

In NHANES III, fasting serum samples were collected from 1988 to 1994 and IGF-1 concentrations were quantified by IGF-I enzyme-linked immunosorbent assay (DSL 10-5600) including an extraction step which separates IGF-1 from its binding protein. The samples were reanalyzed if the coefficient of variation for replicate samples was greater than 15% [59]. In the current study, given the differences in the assays used for serum IGF-1 measurements in LLFS and NHANES III, we only compared the relationships among traits between studies.

3.2.4 Statistical approach

All data analysis was performed using R version 3.4.0 [60]. To approximate normality, both IGF-1 and BMI were natural log–transformed for the analysis. We calculated the Pearson coefficient of correlation between IGF-1 levels and BMI, height, weight, WC, and age.

We used two sample t-tests to assess the mean age difference and mean IGF-1 difference between the LLFS and NHANES III. Also, we used the lstrends() function to estimate and compare the slopes of fitted lines between male and females in both studies.

We regressed log(IGF-1) on log(BMI) to get an overall assessment of their relationship in each cohort. We performed linear mixed effect model using *coxme* package [73] and adjusting for

covariates. Covariates were chosen based on their known association with serum IGF-1 and included age, age^2, male sex, log(BMI), field center, diabetes, hypertension, log(BMI) × age, and log(BMI) × sex as fixed variables and kinship as a random variable accounting for the relatedness between LLFS participants. To assess whether the relationship differed by sex, for all analyses described below we also stratified by sex, and regressed IGF-1 on BMI as above without the sex and log(BMI) × sex terms.

First, we regressed IGF-1 on BMI in all samples regardless of age but included other covariates. Then, to assess the relationship between IGF-1 and BMI by age, we divided the LLFS sample into age quartiles and performed linear mixed-model regression of IGF-1 with BMI as a fixed variable and kinship as a random variable within each age quartile. In addition, we also conducted all the previous analyses with IGF-1 and WC, as another measure of adiposity.

To validate the results, the same approach was used with the NHANES III sample of 2,555 non-Hispanic White participants, although we did not adjust for kinship, as the participants were assumed to be unrelated. The age-quartile thresholds in LLFS were used as age-group thresholds in NHANES III. We then assessed whether the relationship between IGF-1 and BMI also differed by age quartile in non-Hispanic Black (n = 1,639) and Mexican American (n = 1,607) participants.

3.3 Results

The means, standard deviations, and proportions of key characteristics of the study samples are presented in **Table 3**. The Long Life Family Study participants' mean age was 70 years (range of 24 years–110 years) and the prevalence of diabetes, and hypertension were 7%, 51%, respectively. The overall mean serum IGF-1 level was 128.3 ng/mL and ranged from 26 ng/mL to 745 ng/mL.

The mean BMI was 27 kg/m^2. In NHANES III, the participants' mean age was 53.2 years (range of 20 years–90 years) and the prevalence of diabetes, hypertension were 6%, 28%, respectively. The overall mean IGF-1 was 249.5 ng/mL and ranged from 25.3 ng/mL to 863.8 ng/mL. The mean BMI was 26.48 kg/m^2.

Both studies had a wide age range, but on average, NHANES III participants were 16.8 years younger than LLFS participants ($p < 0.0001$). The age distribution in LLFS is bimodal due to the family-study design with some overlap between the LLFS generations; the age distribution in NHANES III is approximately uniform across its range (**Figure C1 Appendix C**). As expected, log(IGF-1) levels were negatively correlated with age in both LLFS and NHANES III, $r = -0.42$ ($p < 0.001$) and $r = -0.47$ ($p < 0.001$), respectively **(Figure 1a**). In addition, mean serum IGF-1 levels were 121.2 ng/ml lower in LLFS compared to NHANES III ($p < 0.0001$).

Figure 1

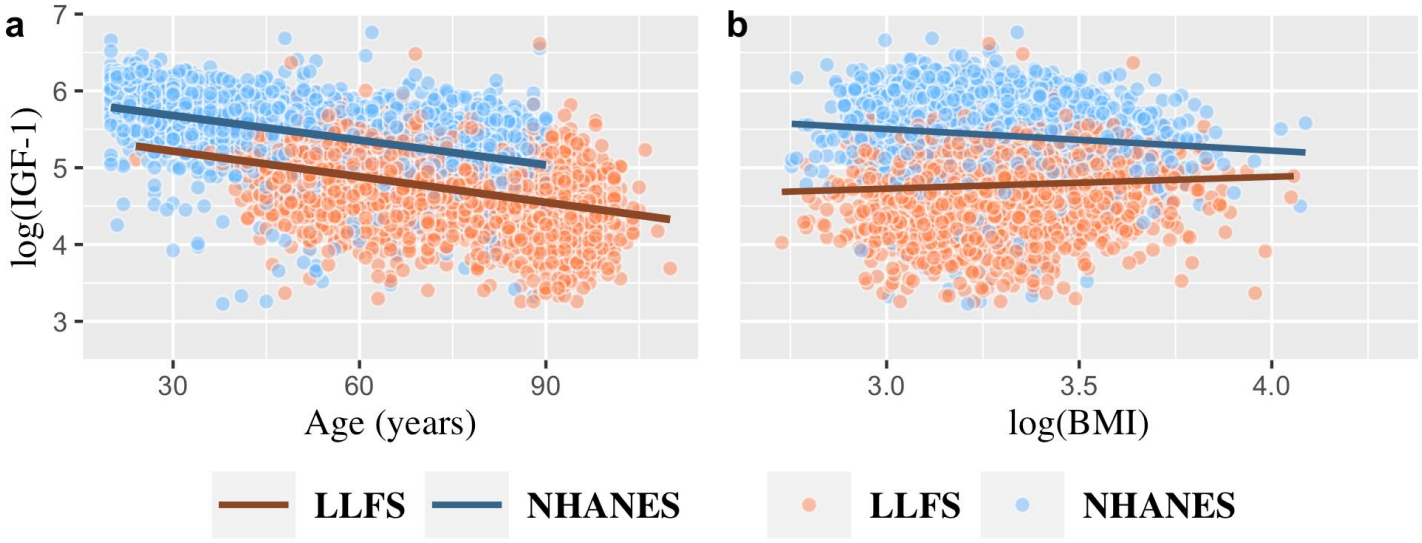


Figure 1. a) scatter plot of log(IGF-1) by age and b) scatter plot of log(IGF-1) by log(BMI) in both LLFS and NHANES III.

In LLFS, across all participants, log(IGF-1) levels were positively correlated with both log(BMI) and WC, $r = 0.06$ ($p < 0.001$) and $r = 0.03$ ($p = 0.06$), respectively (**Table 4**). In contrast, in NHANES III, log(IGF-1) levels were negatively correlated with log(BMI) and WC, $r = -0.12$ ($p < 0.001$) and $r = -0.18$ ($p < 0.001$), respectively. However, log(IGF-1) was positively correlated with height among both LLFS and NHANES III participants ($r = 0.27$, $p < 0.001$ and $r = 0.24$, $p < 0.001$, respectively) **(Table 4**).

Table 3. Descriptive statistics of age, IGF-1, and anthropometric measurements in LLFS and NHANES III.

Measurements	LLFS			NHANES III		
	Male (45.3%)	Female (54.7%)	Total	Male (45.2%)	Female (54.8%)	Total
All Participants (20 y–110 y) (*n*)	1924	2317	4241	1155	1400	2555
1st Age Quartile/Group (20 y–58 y) (*n*)	421	640	1061	643	840	1483
2nd Age Quartile/Group (58 y–66 y) (*n*)	507	553	1060	140	144	284
3rd Age Quartile/Group (66 y–86 y) (*n*)	497	563	1060	346	381	727
4th Age Quartile/Group (87 y–110 y) (*n*)	499	561	1060	26	35	61
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Age at Enrollment (y)	70.6 (15.3)	69.5 (15.9)	70.0 (15.6)	54.0 (19.5)	52.6 (19.7)	53.2 (19.6)
IGF-1 (ng/ml)	134.6 (54.0)	123.2 (51.6)	128.3 (52.9)	264.7 (97.6)	236.9 (103.6)	249.5 (101.9)
BMI (kg/m^2)	27.5 (4.0)	26.7 (5.3)	27.8 (4.8)	26.7 (4.8)	26.3 (5.7)	26.5 (5.3)
Height (cm)	173.6 (7.7)	159.6 (7.8)	166.0 (10.5)	175.5 (7.1)	161.3 (7.0)	167.7 (9.9)
Weight (kg)	83.2 (14.9)	68.2 (14.9)	75.0 (16.5)	82.5 (16.5)	68.4 (15.4)	74.8 (17.4)
WC (cm)	99.5 (11.0)	90.4 (13.9)	94.5 (13.4)	97.7 (12.7)	89.5 (14.5)	93.2 (14.3)
Hypertension	24%	27%	51%	12%	16%	28%
Diabetes	4%	3%	7%	3%	3%	6%

In the regression analysis—adjusting for sex, diabetes, and hypertension, and, in LLFS, field center and kinship, log(IGF-1) was associated positively with log(BMI) ($\beta = 0.20$, $p = 4.4 \times 10{-}^{12}$); whereas in NHANES III, the relationship was significant and negative ($\beta = -0.23$, $p = 1.4 \times 10^{-6}$) (**Table 5, Figure 1b)**. In NHANES III, Hypertension was negatively associated with log(IGF-1) in the overall sample ($\beta = -0.12$, $p = 6.9 \times 10^{-10}$) and in the overall sample stratified by sex **(Data not shown).** However, in LLFS, hypertension was not associated with log(IGF-1) in the overall sample, but it was negatively associated with log(IGF-1) in the stratified overall sample by sex (both $p < 0.0001$) **(Data not shown).**

Table 4. Pearson correlation coefficients between log(IGF-1), age, and anthropometric measures in LLFS.

		LLFS					
		log(IGF-1)	Age	log(BMI)	Height	Weight	WC
NHANES III	log(IGF-1)	1	−0.42***	0.06***	0.27***	0.19***	0.03
	Age	−0.47***	1	−0.12***	−0.38***	−0.31***	0.03*
	log(BMI)	−0.12***	0.10***	1	0.08***	0.82***	0.81***
	Height	0.24***	−0.24***	0.01	1	0.62***	0.80***
	Weight	0.02	−0.05***	0.85***	0.52***	1	0.80***
	WC	−0.18***	0.28***	0.87***	0.23***	0.86***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In LLFS, interaction for both log(BMI) and age, and log(BMI) and sex had significant effects on log(IGF-1) (both $p < 0.0001$) (Data not shown). Whereas in NHANES III, there was a significant interaction effect between log(BMI) and age ($p = 7.3 \times 10^{-5}$) on log(IGF-1), but no significant interaction between log(BMI) and sex ($p = 0.4$) (**Data not shown**).

We further investigated the interaction between age and log(BMI) on log(IGF-1) in LLFS using age quartiles (**Table C1 Appendix C)**. As can be seen in **Figure 2a**, the relationship between log(IGF-1) and log(BMI) differed by age quartile. In the first (youngest) age quartile (20 y–58 y)

the relationship was significant and negative ($\beta = -0.2$, $p = 0.0022$), in the second (59 y–66 y) and third (67 y–86 y) age quartiles the relationship was non-significant, but in the fourth (oldest) quartile (87 y–110 y), the relationship was significant and positive ($\beta = 0.31$, $p = 2.6 \times 10^{-4}$) (**Table 5**). When the NHANES III data was stratified using the LLFS age-quartile thresholds, a similar pattern was observed (**Table 5**, **Figure 2a**). We also stratified LLFS and NHANES III using age thresholds derived from NHANES III and applied them to LLFS; a similar pattern was observed (**Figure C2 Appendix C**).

Table 5. Stratified results of linear mixed-model regression of log(IGF-1) on log(BMI).

	LLFS			NHANES III		
log(IGF-1) on log(BMI) in	β	(SE)	*p* value	β	(SE)	*p* value
All Participants	0.20	(0.04)	4.4×10^{-12}	−0.23	(0.05)	1.4×10^{-6}
Male Participants	0.45	(0.06)	4.5×10^{-12}	−0.23	(0.07)	1.4×10^{-3}
Female Participants	0.05	(0.05)	0.28	−0.22	(0.06)	5.2×10^{-4}
1st age quartile/group (20 y–58 y)	−0.20	(0.06)	2.2×10^{-3}	−0.40	(0.05)	2.6×10^{-12}
2nd age quartile/group (58 y–66 y)	−0.07	(0.06)	0.32	0.01	(0.13)	0.96
3rd age quartile/group (67 y–86 y)	−0.01	(0.07)	0.84	0.03	(0.08)	0.76
4th age quartile/group (87 y–110 y)	0.31	(0.08)	2.6×10^{-4}	0.84	(0.39)	0.03

We next investigated the relationship between log(BMI) and sex on log(IGF-1) levels in LLFS by stratifying each age quartile by sex. Among females, log(IGF-1) was significantly and negatively associated with log(BMI) in the first (20 y–58 y) age quartiles ($\beta = -0.28$, $p = 3 \times 10^{-4}$), but there was no significant association in males (**Figure 2b**). In the second (59 y–66 y) and third age quartile (67 y–86 y), log(IGF-1) was not associated with log(BMI) in either sex. However, in the oldest quartile (87 y–110 y), log(IGF-1) was significantly and positively associated with log(BMI) in both sexes ($\beta = 0.2$, $p = 0.04$ and $\beta = 0.4$, $p = 0.0014$, respectively) (**Figure 2b**). The relationship between log(IGF-1) and log(BMI) did not significantly differ by sex in NHANES III, except in the fourth quartile (87 y–110 y) wherein the association was positive and significant in

males (p = 0.002), but not in females (p = 0.4), though the sample sizes were small. In addition, we observed a significant slope difference by sex in the relationship between log(IGF-1) and log(BMI) in the first age quartile of LLFS only (p = 0.002).

Figure 2

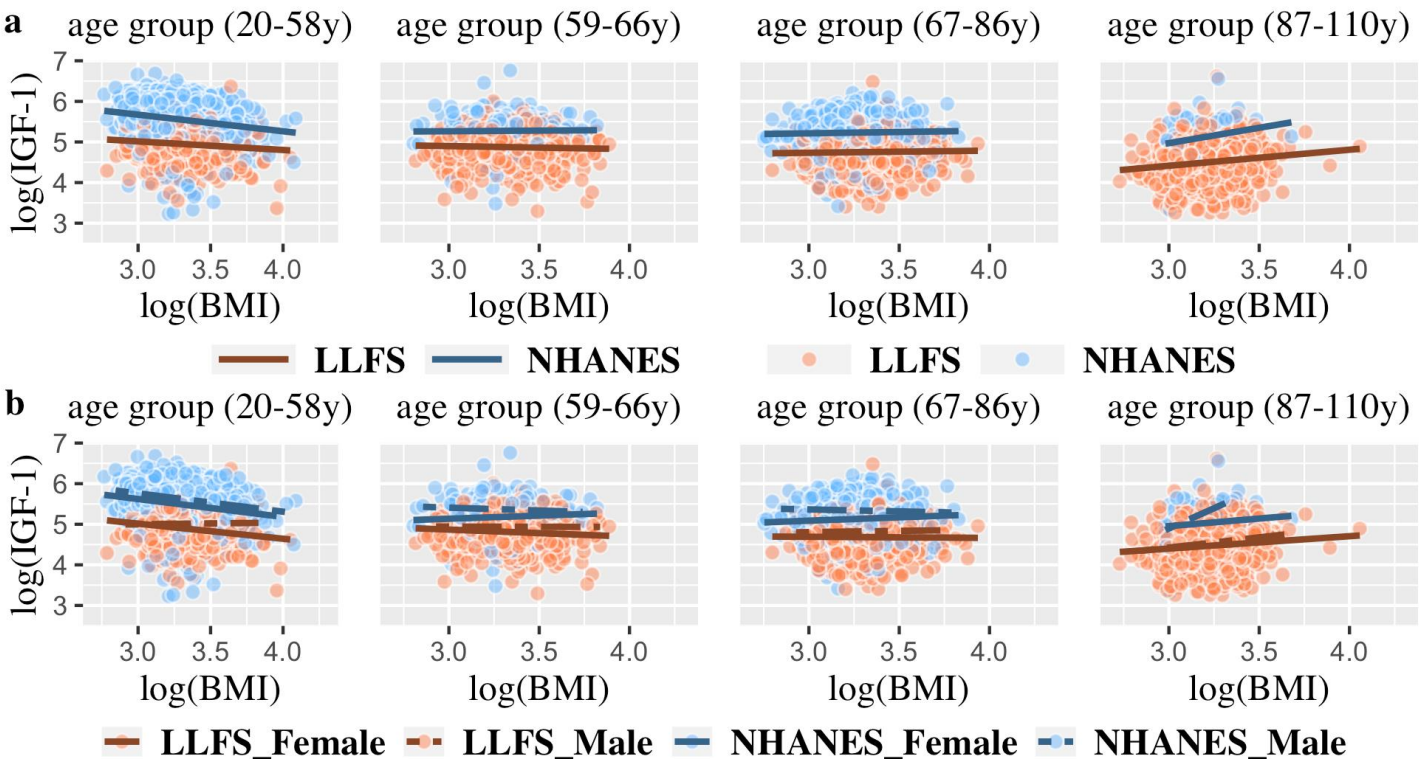


Figure 2. a) scatter plot of log(IGF-1) by log(BMI) stratified by age group, b) scatter plot of log(IGF-1) by log(BMI) per age group and stratified by sex in LLFS and NHANES III. The same age quartile thresholds for LLFS were applied onto NHANES III.

Similar results were observed for the relationship between IGF-1 and WC in both studies (**Table C2 Appendix C**). Also, similar patterns were seen among non-Hispanic Black and Mexican American participants in NHANES III (**Figure C3 Appendix C).**

3.4 Discussion

In this cross-sectional study of LLFS, a unique family-based cohort of exceptional longevity, we examined the age- and sex specific effects of the relationship between serum IGF-1 levels and BMI. Younger participants (24 y–58 y), had a negative relationship between IGF-1 and BMI, while older participants (87 y–110 y) had a positive relationship. There was no statistically significant relationship for the age-groups in between. The same pattern was observed in an independent sample of non-Hispanic White adults of similar age range recruited from the general population in the NHANES III. In addition, we did not observe a consistent sex-specific difference in the relationship between IGF-1 and BMI across the age groups. The discrepancies in the relationship between serum IGF-1 and BMI among studies in the literature [18, 36, 41–45, 47], may be explained by the ages of the cohorts used in the previous studies. Studies reporting the negative relationship between BMI and IGF-1 were primarily conducted in participants with ages ranging from 10 years to 60 years [18, 41–45]. In contrast, studies that reported a positive relationship were often performed with older individuals with ages ranging from 45 y to 90 y [36]. These studies also showed similar pattern between the relationship of IGF-1 and WC [36, 42–44]. In addition, several of these studies were comprised of highly selected groups, such as obese/overweight individuals, who might be experiencing weight-related disruption in insulin and growth hormone secretion [44].

The age-related difference in the relationship between IGF-1 and BMI might be due to height, which is a component of BMI. However, the pattern between IGF-1 and WC (a measure of central adiposity that is independent of height) was like that with BMI. This result indicates that the relationship is driven by adiposity rather than height (**Table C2 Appendix C**).

In non-Hispanic Black and Mexican American participants, other investigators have reported an inverse association between IGF-1 and BMI [41, 42]; whereas others have reported no association [74]. However, in our study, we saw similar patterns by age group in non-Hispanic Black and Mexican American participants within NHANES III as we saw in the non-Hispanic White participants, despite smaller sample sizes. (**Figure C3 Appendix C**). These results suggest that the relationship between IGF-1 and BMI by age group is similar among different racial/ethnic groups.

In this cohort we observed that younger participants had higher mean IGF-1 level compared to older participants, and this is consistent with known IGF-1 biology in adolescent and early-adulthood [10]. Although the interaction between log(BMI) and sex in predicting log(IGF-1) was a statistically significant overall, the slope difference between male and female was not only statistically significant except in the youngest LLFS age quartile. The latter results might reflect the sex differences in development during puberty and early adulthood.

LLFS and NHANES III data were collected 10 years apart, thus, period or cohort effects may exist, in addition to the age-effect we demonstrate. However, despite this potential period effect, the patterns were consistent for both LLFS and NHANES III cohorts, for BMI and WC, and across different racial/ethnic groups. In addition, mean serum IGF-1 levels were 121.2 ng/ml lower in LLFS compared to NHANES III. The most likely reason for this difference is the use of different assay kits in the measurement of serum IGF-1 levels between the two studies. Previous studies have reported significant differences in serum IGF-1 levels when using different assay kits, even though the samples were from the same population [75–77]. Different assay kits have different age- and sex-specific reference ranges, and this might affect the upper and lower limit of each study's serum IGF-1 levels. Furthermore, NHANES III was conducted using a sample from

the general population, whereas LLFS sampled healthy long-lived individuals[56]. Thus, study population and assay type are confounded, and it is impossible to determine if the difference in mean IGF-1 levels between the cohorts is due to ascertainment differences or to assay differences given these data. However, all statistical comparisons in this paper were done within-study, so the mean differences in IGF-1 levels between studies should not affect our conclusions, especially given that the patterns across age groups were similar.

This current study was cross-sectional and not longitudinal; therefore, we could not measure the relationship between IGF-1 and BMI on the same participants throughout their lifespan to determine the patterns of change in this relationship. Instead, we stratified our samples by age group and are extrapolating these cross-sectional results to reflect individual changes related to aging. However, additional longitudinal data are needed to confirm these findings. Also, we relied on BMI and WC to measure adiposity, which capture body size but not body composition. These anthropometric measurements are not as precise as imaging-based measurements, such as fat mass as estimated from dual X-ray absorptiometry or peripheral quantitative computed tomography [78]. Such measures were not available for these studies. Another limitation in our study is the lack of data on potential confounders such as physical activity, diet, and GH level. These factors can have major effects on BMI and IGF-1 level and may be important to consider as potential confounders of this association.

In summary, we identified age-related differences in the relationship between serum IGF-1 levels and BMI, as well as WC, in non-Hispanic White, non-Hispanic Black, and Mexican American participants. This finding clarifies that the apparent contradiction in the previous literature on the relationship between IGF-1 and adiposity is likely due to differences in cohort age ranges. As such, the clinical implication of this is that age should be considered when evaluating

the relationship between adiposity and IGF-1. However, longitudinal studies and further investigation into the underlying biology affecting the relationship between serum IGF-1 and measures of adiposity across the lifespan is needed to understand these observations. Such an understanding might help categorize individuals at risk of disease or inform interventions to delay disease depending on their age-dependent BMI and IGF-1 level.

4.0 LINKAGE ANALYSIS FOR SERUM IGF-1 LEVELS IN LLFS

4.1 Introduction

IGF-1, the product of a single gene (*IGF1*) located on the long arm of chromosome 12 [48], is the mediator of growth hormone (GH) action in human body and GH is a major stimulator of IGF-1 production. Furthermore, IGF-1 levels are age-dependent; its level peaks at puberty then starts declining in adulthood [10]. In Chapter 3, I also reported a negative correlation between serum IGF-1 level with age. However, the relationship between IGF-1 and BMI varied by age, such that the relationship between IGF-1 and BMI in younger participants differed from that in the oldest participants. The genetic and environmental components (such as age, sex, and nutritional status) on individual variation in serum IGF-1 levels has been reported in a few family studies [18, 26, 49, 50, 70]. Previous studies in adults have estimated that the heritability of serum IGF-1 levels ranges from 38% to 63%. Only one study has reported conducting a genome-wide linkage analysis to assess chromosomal regions linked to serum IGF-1 variation [18]. Lam and colleagues identified two suggestive regions of linkage: one on chromosome 12 at 8 cM (LOD score = 2.41) and another on chromosome 1 at 36 cM (LOD score = 2.41). Neither of these regions contain *IGF1*.

In this chapter, I assessed the contribution of genetic factors to variation of serum IGF-1 levels (heritability) and the genetic and environmental correlation between serum IGF-1 levels and BMI. Because the phenotypic relationship between IGF-1 and BMI varied by age, I also performed analysis of genetic correlations stratified into probands and offspring. Finally, I used quantitative

trait loci (QTL) linkage analysis, an unbiased genetic mapping approach, to identify potential regions of the genome influencing serum IGF-1 levels.

4.2 Method

4.2.1 Study population

For the heritability analyses, analyses were performed using the data of 4,400 participants (proband, offspring, and controls) from the Long Life Family Study. For bivariate analyses estimating the genetic correlation between IGF-1 and BMI, data were available for 4,203 participants, because of missing data on BMI. In analysis of genetic correlations stratified by generation, data were available for 1,391 proband generation participants and for 2,090 offspring generation participants (data for 722 married-in spouses of offspring generation participants were not used in the stratified analyses).

4.2.2 Statistical and genetic analysis

As described in **Chapter 2**, the preliminary statistical analyses were performed using R, and the genetic analyses were performed using SOLAR. Because the distributions of both serum IGF-1 and BMI were skewed, natural log transformation was applied to reduce non-normality.

4.2.2.1 Heritability of IGF-1

Heritability (h^2) was estimated by SOLAR using maximum-likelihood methods adjusting for age and sex. Heritability was also estimated separately for both proband and offspring generations. The estimated IGF-1 heritability, which is the fraction of the additive genetic variance (σ_A^2) that can be attributed to the phenotypic variability (σ_P^2) of serum IGF-1 levels, was defined as:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

4.2.2.2 Genetic correlation between serum IGF-1 and BMI

The shared genetic and environmental components of variance between IGF-1 and BMI were estimated by bivariate genetic correlation analysis using SOLAR, adjusting for age and sex. The genetic (ρ_G) and environmental (ρ_E) correlations were estimated in the non–generation-stratified cohort of 4,203 LLFS participants using this model:

$$r = \sqrt{h^2_{\text{IGF-1}}}\sqrt{h^2_{\text{BMI}}} \cdot \rho_G + \sqrt{1 - h^2_{\text{IGF-1}}}\sqrt{1 - h^2_{\text{BMI}}} \cdot \rho_E$$

$h^2_{\text{IGF-1}}$ is the heritability for IGF-1

h^2_{BMI} is the heritability for BMI

ρ_G is the genetic correlation

ρ_E is the environmental correlation

Because the phenotypic relationship between IGF-1 and BMI differed by age, I assessed whether the genetic relationship differed by age by estimating the genetic correlation within each generation, adjusting for age and sex. In addition, a parallel analysis was performed to assess the genetic correlation between IGF-1 and WC. Although the age distributions of the proband and offspring generations overlap, I analyzed data within each generation because inclusion of relatives

is necessary to estimate genetic correlations and stratifying by an age cutoff would have disconnected informative relationships.

4.2.2.3 Quantitative trait locus (QTL) linkage analysis of IGF-1

Multipoint quantitative trait linkage analysis was performed, using SOLAR, to identify the shared genetic regions associated with serum IGF-1 levels in LLFS participants using SOLAR.

QTL linkage analysis is a method of linkage analysis that identifies identity-by-decent (IBD) allele sharing between relative pairs that is linked to the quantitative phenotype [61]. The multipoint IBD methodology extends the analysis of heritability with the added assumption of the effect of the QTL (σ^2_{QTL}) as a component of the additive genetic variance (σ^2_A). Using the variance components method, the QTL effect can be tested by calculating the maximum-likelihood ratio comparing the null hypothesis of no linkage to the alternative hypothesis of linkage. The general form of variance model is:

$$\sigma^2_P = \sigma^2_E + \sigma^2_A + \sigma^2_e$$

where σ^2_P is the total phenotypic variance, σ^2_E is the enviromental component, σ^2_A is the genetic component (which include the QTL component σ^2_{QTL}), and σ^2_e is the error component. The statistical evidence of linkage, modulated by the QTL component (H_1: $\sigma^2_{QTL} > 0$ or $\theta < 0.5$), was reported as logarithm of the odds (LOD) score ($\log_{10}$ of the likelihood ratio), with a suggestive cutoff of LOD = 3 or significant cutoff of LOD = 3.3. All analyses were adjusted for age and sex. The corresponding region of interest for the QTL was defined as the region from the peak LOD score out to the LOD score that is the difference between the peak LOD and 1.8, on both both sides of the peak LOD score. The region of interest (in centimorgans) was then mapped against the physical map in basepairs. To further narrow the linkage peak region of interest so that the involved

gene or genes could be more easily ascertained, QTL linkage analyses were performed with the data from families whose familywise LOD scores were in the top 20%. This subset of families has the strongest evidence of linkage within them, and the exclusion of data from families with weaker evidence of linkage might highlight the narrower region I am interested in.

To test the possibility that different QTLs were segregating in different families (genetic heterogeneity), heterogeneity LOD scores (HLOD) were calculated, using SOLAR. If there is evidence of heterogeneity, this means some families will show evidence of linkage at a QTL while others will not. Heterogeneity of this type will reduce the power to detect linkage at such QTLs. The HLOD function in SOLAR, via homo program, tests heterogeneity using an admixture model [80]. The program performs three tests: linkage assuming homogeneity [H_0 (no linkage) vs H_1 (linkage under homogeneity)], for heterogeneity given linkage [H_1 vs H_2], and for joint linkage and heterogeneity [H_0 vs H_2]. In the admixture model, proportion (α) of families in the sample are linked to the marker of interest (recombination rate $\theta < 0.5$) while the rest of families ($1 - \alpha$) are unlinked ($\theta = 0.5$). This test is considered significant only if $p < 0.0001$ and is used at a single position only after significant evidence of linkage has been obtained (that is, at a position of a statistically significant LOD score peak).

4.2.2.4 Bivariate QTL Linkage analysis of serum IGF-1 levels and BMI

To further evaluate whether any QTLs influence IGF-1 and BMI pleiotropically, bivariate linkage analysis of IGF-1 and BMI, adjusted for age and sex, was performed using SOLAR. These analyses were done in the full cohort (n = 4,203) and by generation proband (n = 1,392) and offspring (n = 2,090).

4.3 Results

4.3.1 Estimated heritability of IGF-1

The estimated heritability of serum IGF-1 (h^2) in the total sample, adjusted for age and sex, was 0.41 ± 0.04 ($p < 0.0001$), and the proportion of variance attributed to the covariates was 0.19. Due to the known phenotypic correlation between age and IGF-1 levels, and to assess possible age-dependent effects on IGF-1 heritability, I also estimated heritability in each generation separately. The heritability, adjusted for age and sex, in the proband generation ($h^2 = 0.48 \pm 0.07$) was similar to that in the offspring generation ($h^2 = 0.49 \pm 0.06$**)**

4.3.2 Estimated genetic correlation between serum IGF-1 and BMI

In the analysis of genetic correlation between IGF-1 and BMI in the total sample, adjusted for age and sex, the heritability of BMI was 0.44 ± 0.04 for BMI ($p < 0.0001$) **(Table 6)**. The environmental correlation (ρ_E) was 0.03 ($p = 0.7$). The genetic correlation (ρ_G) was −0.02 ($p = 0.8$). Stratified by generation, the genetic correlation was statistically significant in the proband generation ($\rho_G = 0.29$, $p = 0.01$) but not in the offspring **(Table 6)**.

In genetic analysis of IGF-1 and WC, the heritability of WC was 0.51 ± 0.04 ($n = 3418$, $p < 0.0001$). The environmental correlation (ρ_E) between the two was 0.003 ($p = 0.9$), and the genetic correlation (ρ_G) was 0.001 ($p = 0.9$). As between IGF-1 and BMI, there was statistically significant genetic correlation between IGF-1 and WC in the proband generation ($n = 1357$, $\rho_G = 0.36$, $p = 0.001$), but not in the offspring.

Table 6. The genetic correlation between log IGF-1 and BMI. Analyses were adjusted for age and sex in the overall and stratified by generation analysis.

	Proband (n = 1391)	Offspring (n = 2090)	Overall (n = 3481)
h^2_{IGF} ± s.e. (p value)	0.48 ± 0.07 (< 0.0001)	0.49 ± 0.06 (< 0.0001)	0.42 ± 0.04 (< 0.0001)
h^2_{BMI} ± s.e. (p value)	0.42 ± 0.06 (< 0.0001)	0.54 ± 0.06 (< 0.0001)	0.44 ± 0.04 (< 0.0001)
ρ_G (p value)	0.29 (0.01)	–0.02 (0.62)	–0.02 (0.8)
ρ_E (p value)	–0.05 (0.64)	–0.14 (0.63)	0.03 (0.7)
r (p value)	0.10 (0.0001)	–0.06 (0.01)	0.008 (0.6)

h^2_{IGF} is the heritability of IGF-1. h^2_{BMI} is the heritability of BMI.
ρ_G is the genetic correlation. ρ_E is the enviromental correlation. r is the phenotypic correlation.

4.3.3 Quantitative trait locus (QTL) linkage analysis of IGF-1

Multipoint linkage analysis in the overall cohort identified a novel locus on chromosome 11 (LOD = 3.48; **Figure 3**). The region of interest for the location of the QTL is between 67 cM and 88 cM, approximately equivalent to chr11:20,676,782–31,034,467 (**Figure 4**).

To refine the linkage region and reduce the number of potential candidate genes under the linkage peak, families with familywise LOD scores among the top 20% of familywise LOD scores were selected for follow-up linkage analysis. The summed LOD score decreased from 3.48 to 2.78 on chromosome 11; however, the width of the region of interest for the QTL was not decreased (**Figure 5**). Because the region of interest was relatively large, 21 cM, there are approximately 32 possible candidate genes under the peak. No further analysis to identify possible causal genes under the peak were performed.

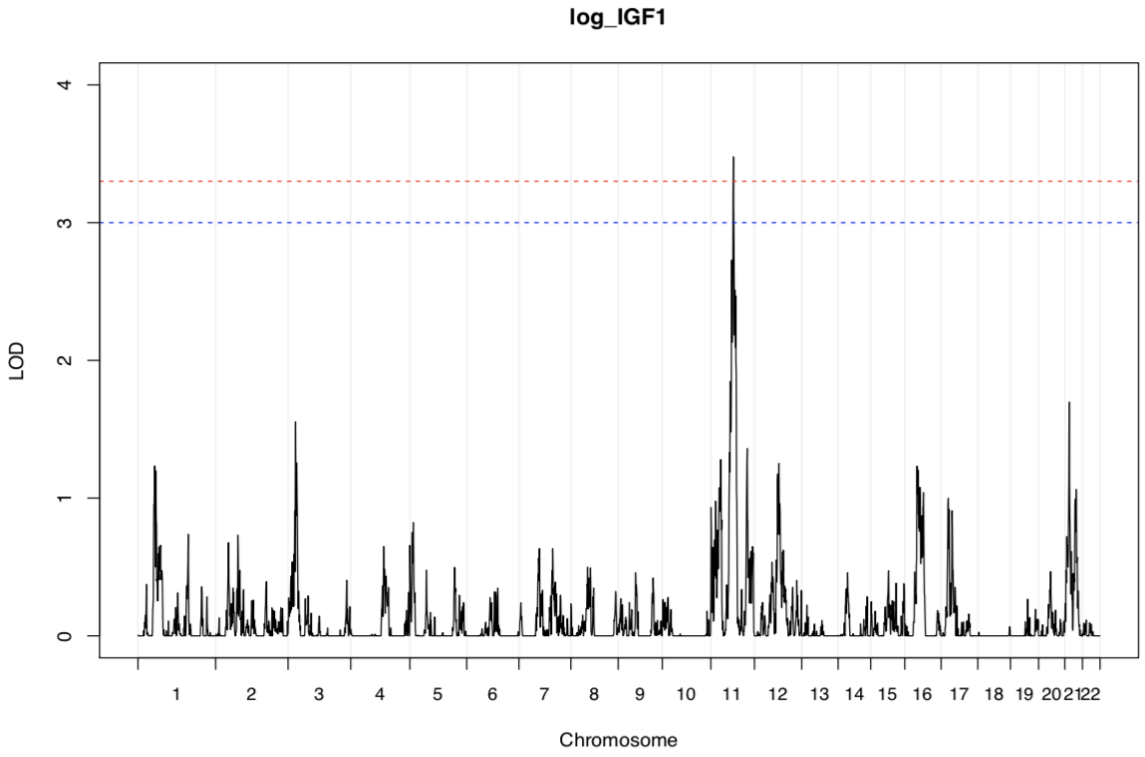


Figure 3. Genome-wide multipoint quantitative linkage analysis of IGF-1 adjusting for age and sex. Significant linkage region associated with serum IGF-1 on chromosome 11 with LOD = 3.48.

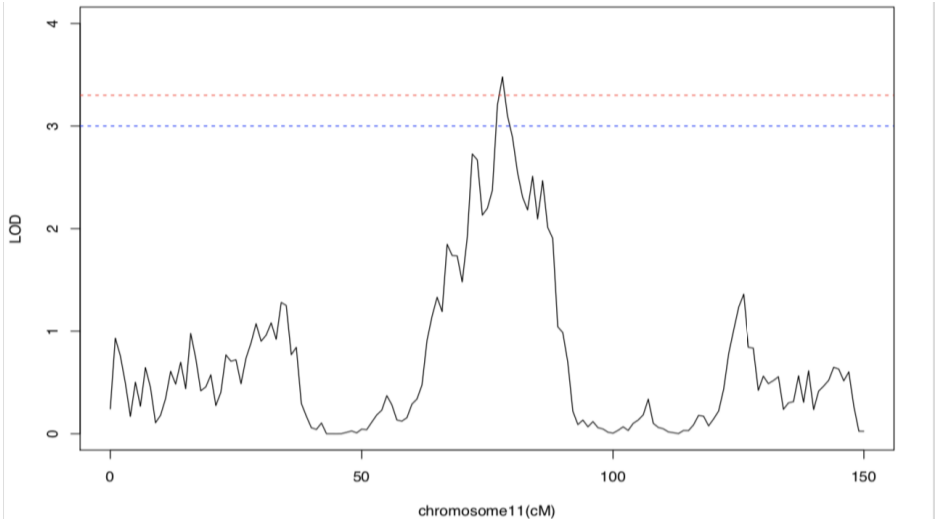


Figure 4. Genome-wide multipoint quantitative linkage analysis of IGF-1, adjusting for age and sex, on chromosome 11 with LOD = 3.48.

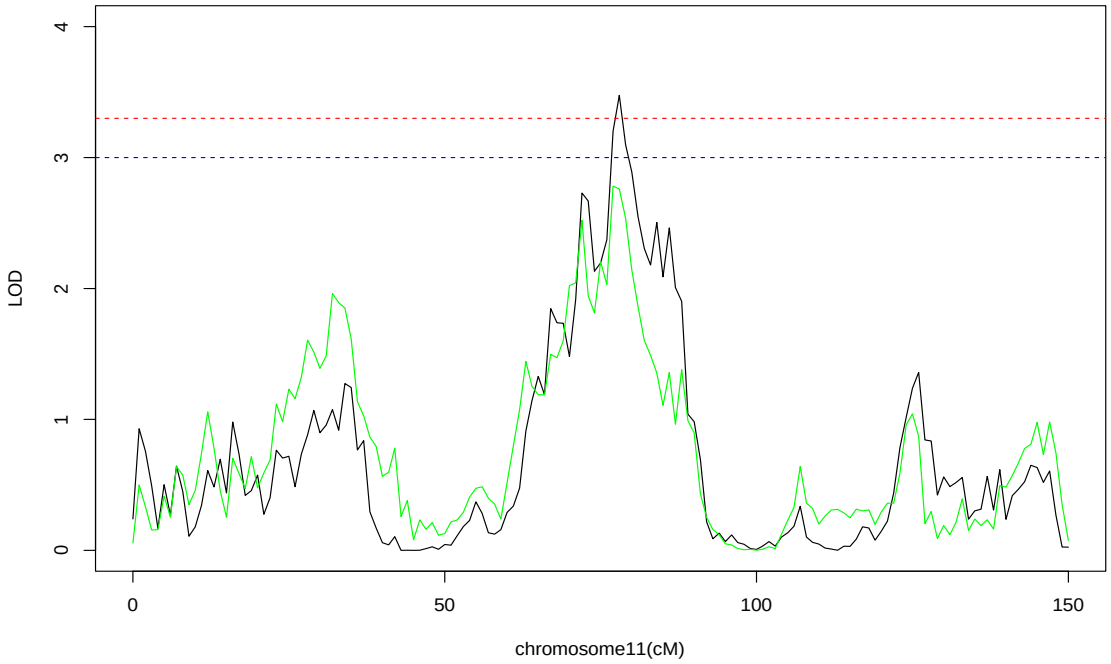


Figure 5. Linkage analysis of families in the top 20% of LOD scores. The linkage analysis LOD scores for all families are the black line; the LOD scores for the families with a LOD score in the top 20% of LOD scores are the green line.

To test for possible heterogeneity in the linkage among the LLFS families, I used the HLOD function in SOLAR. The test for linkage (H_1: $\theta < 0.5$) versus no linkage (H_0: $\theta = 0.5$) was highly significant (LOD = 3.47, $p = 0.00003$), and the test for linkage with heterogeneity given linkage (H_2: $\theta < 0.5$, $\sigma^2_{QTL} > 0$) versus no linkage (H_0: $\theta = 0.5$) was significant ($p = 0.001$). However, the test of linkage under homogeneity (H_1: $\sigma^2_{QTL} > 0$) versus linkage with heterogeneity (H_2: $\theta < 0.5$, $\sigma^2_{QTL} > 0$) was not statistically significant ($p = 0.5$) **(Table 7**). Therefore, there is no evidence of heterogeneity of linkage among families in this sample.

Table 7. Heterogeneity test results in LLFS.

Test	LOD score	χ^2 statistic	*p* value
H_0 vs. H_1 (linkage under homogeneity)	3.47	16.0	0.00003
H_1 vs. H_2 (heterogeneity given linkage)	0.00	0.0	0.5
H_0 vs. H_2 (linkage and heterogeneity)	3.47	16.0	0.0001

H_0 = no linkage, H_1 = linkage, homogeneity; H_2 = linkage, heterogeneity; significance level is $p < 0.0001$.

4.3.4 Bivariate QTL Linkage analysis of serum IGF-1 levels and BMI

Because there was genetic correlation between IGF-1 and BMI, I performed bivariate linkage analysis. In the total sample, there was no significant evidence for bivariate linkage was observed; the highest LOD score was on chromosome 4 at 117 cM (LOD = 2.67) and on chromosome 11 at 78 cM (LOD = 2.50) **(Figure 6).**

Results of bivariate linkage analysis between IGF-1 and BMI also showed no statistically significant evidence of linkage in either the proband generation or the offspring generation participants (**Figure 7 and 8**). There were two suggestive regions within the offspring generation on chromosome 7 at 48 cM (LOD = 3.24) and on chromosome 17 at 25 cM (LOD = 3.22) (**Figure 8**).

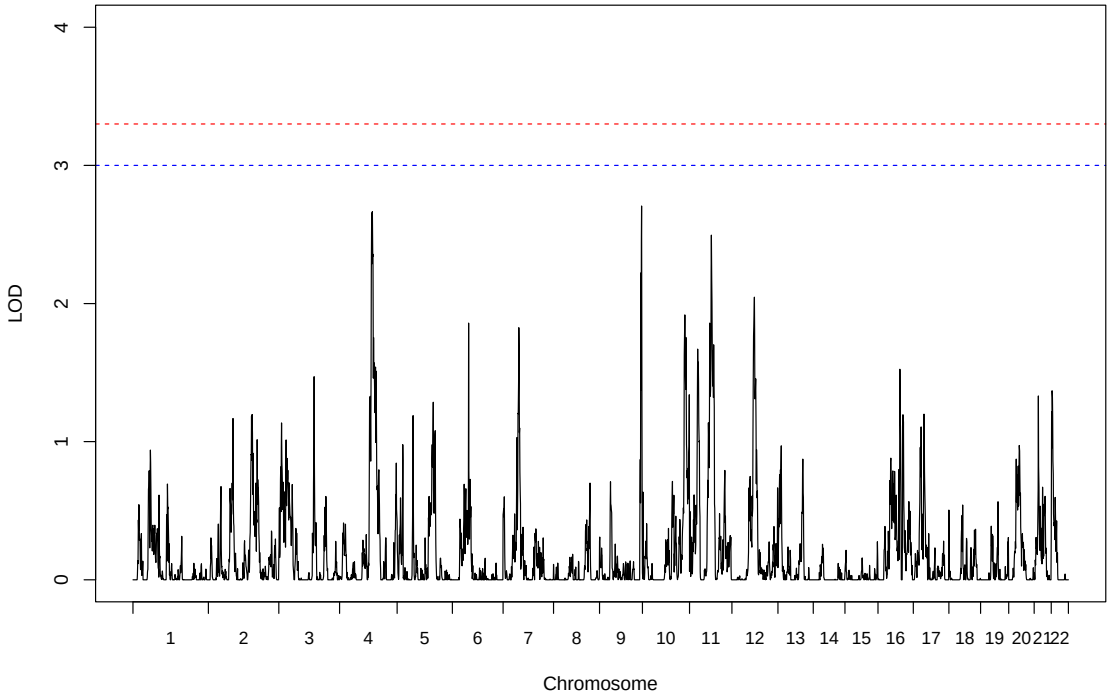


Figure 6. Bivariate linkage analysis of IGF-1 and BMI, adjusted for age and sex, in the total sample. No significant linkage was observed.

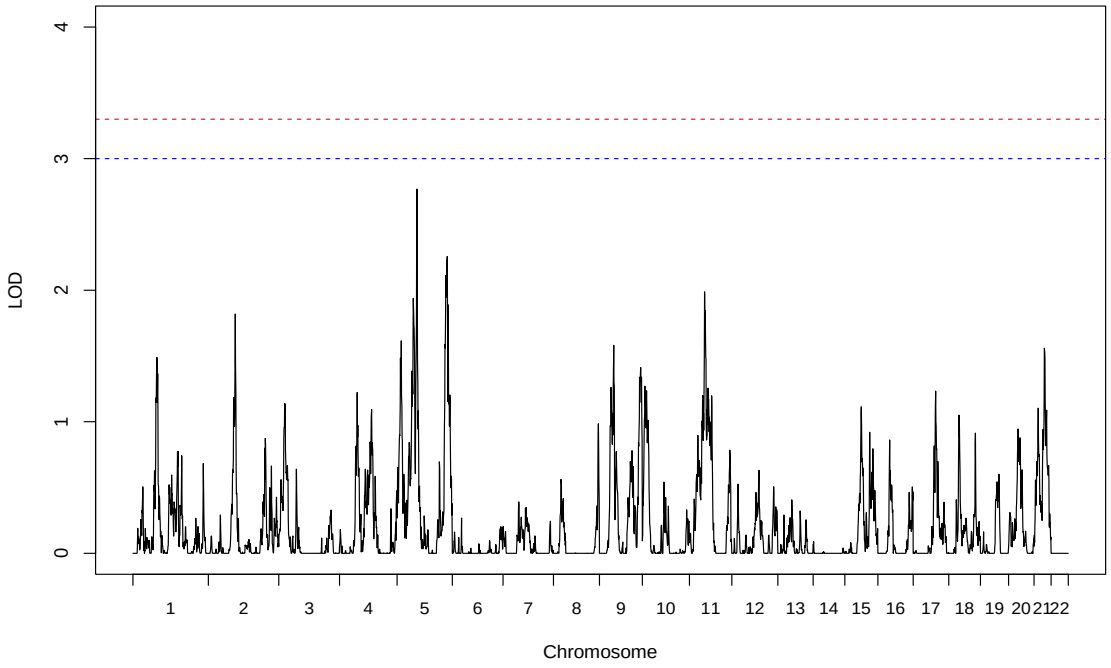


Figure 7. Bivariate linkage analysis of IGF-1 and BMI, adjusted for age and sex, in proband generation. No significant linkage was observed.

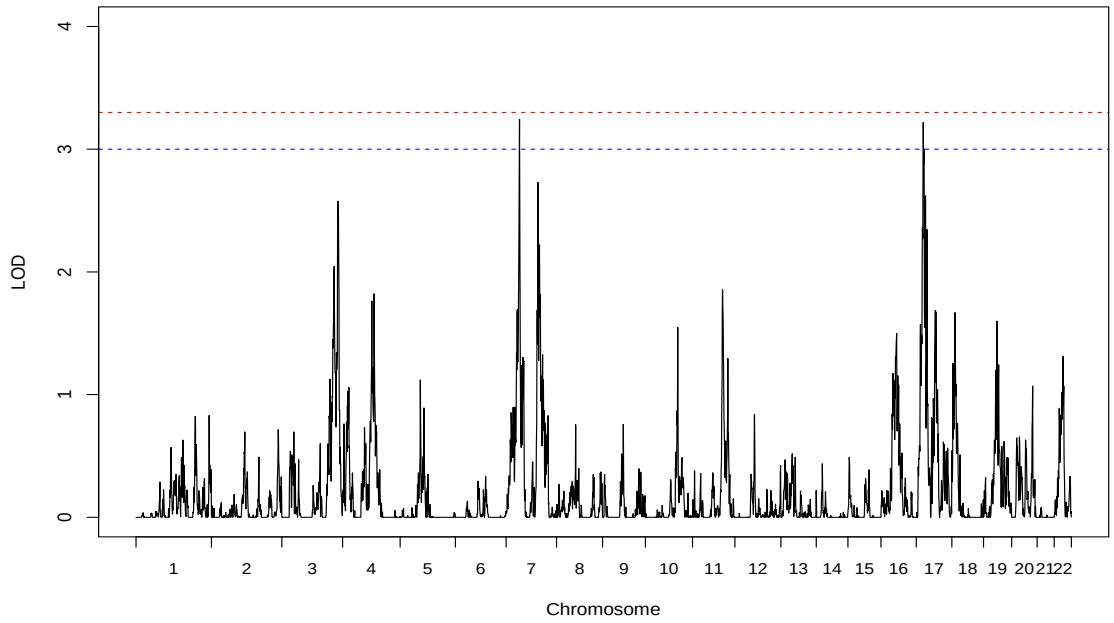


Figure 8. Bivariate linkage analysis of IGF-1 and BMI adjusted for age and sex in offspring generation. Two suggestive regions on at chromosome 7 at 48 cM (LOD = 3.24) and chromosome 17 at 25 cM (LOD = 3.22).

4.4 Discussion

Although many epidemiological studies have assessed the relationships between IGF-1 levels and endogenous and exogenous environmental factors, such as age, sex, nutrition, and physical activity, few have investigated genetic factors that influence IGF-1 levels. In the current study, I estimated the heritability of serum IGF-1 levels, its genetic correlation with BMI, and whether possible QTLs are linked to IGF-1 levels in a unique cohort of families. Because I observed that there was evidence of age-related differences in the correlation between IGF-1 levels and BMI (**Chapter 3**), I also assessed possible genetic factors influencing IGF-1 levels and BMI within each generation—stratifying by age breaks too many relationships to be useful for analysis here.

Within the total LLFS sample, the heritability of serum IGF-1 levels (adjusted for age and sex) was $h^2 = 0.41 \pm 0.004$. This estimate is consistent with previously reported heritability estimates that, ranged from 0.38 (in adult twin pairs) to 0.63 (also in adult twin pairs) [18, 26, 49, 50]. Although the generation-specific heritability estimates were slightly higher than the overall estimate, I observed no difference in the generation specific estimates of the LLFS proband and offspring generations, $h^2 = 0.48 \pm 0.07$ and $h^2 = 0.49 \pm 0.06$, respectively. This result is similar to what Franco and colleagues reported in a previous study assessing age-dependent heritability [81]

Adjusting for age and sex, I detected no genetic correlation between IGF-1 and BMI in the total sample ($\rho_G = -0.02$, $p \geq 0.80$) or in the offspring generation ($\rho_G = -0.02$, $p \geq 0.62$). However, in the proband generation, the genetic correlation between IGF-1 levels and BMI was statistically significant ($\rho_G = 0.29$, $p = 0.01$). These results indicate that part of the phenotypic correlation between IGF-1 and BMI in the proband generation is attributable to variation in a similar set of genes and that this genetic correlation emerges as individuals age. Similar result was observed between IGF-1 and WC, where no genetic correlation in the total sample or the offspring generation, but the genetic correlation was statistically significant in the proband generation ($\rho_G = 0.36$, $p = 0.001$).

To identify possible QTLs and potentially specific genes that influence IGF-1 levels, I next performed QTL linkage analyses and identified a novel locus with a maximum LOD score of 3.48 on chromosome 11 between 67 cM and 88 cM. Approximately 32 loci resided in the QTL region of interest. I then attempted to narrow the region of interest by testing whether a subset of families was showing linkage in the regions. I detected no evidence of heterogeneity among families and was thus unable to narrow the region of interest. I also performed bivariate linkage analyses for IGF-1 levels and BMI, particularly because of the significant genetic correlation between IGF-1

and BMI in the proband generation. However, I detected no significant evidence for the presence of a QTL simultaneously influencing IGF-1 levels and BMI.

In conclusion, the present study is the first to report genetic correlation between IGF-1 levels and BMI in older individuals. In addition, I identified a novel QTL locus on chromosome 11 that is linked to serum IGF-1 levels. This region is not near any of the known genes of the IGF-1 pathway. Additional analyses in other samples are needed to replicate this finding and additional work is necessary to perform follow-up fine mapping to identify putative causal genes.

5.0 GENETIC VARIATIONS ASSOCIATED WITH SERUM IGF-1 LEVELS IN LLFS

5.1 Introduction

Insulin-like growth factor 1 (IGF-1) is a main ligand in the IGF pathway, a conserved pathway that is involved with growth, development, and metabolism starting in the prenatal period and extending throughout adulthood [10, 82]. IGF-1 is synthesized in most cells, but mainly in the liver, and travels in the circulatory system after binding to insulin-like growth factor binding protein 3 (IGFBP3). In addition to being the primary mediator of growth hormone (GH) function [24], IGF-1 and other enzymes and proteins involved in the IGF pathway have been associated with longevity [12, 83], risk of cancer [14, 84], mortality [16, 31], and common age-related disease [13, 15] in many population studies.

Estimated heritability of serum IGF-1 levels ranges from 38% to 63% [18, 26, 49, 50], and is 40% in this study (**Chapter 4, Section 4.3.1**). Several genetic studies of serum IGF-1 levels have been conducted to identify genetic variants associated with IGF-1 levels and to assess whether these variants influence age-related disease risk. Al-Zahrani and colleagues tested for and reported association between serum IGF-1 and genetic variants within *IGF1* or *IGFBP3* [84]. Other researchers have investigated whether single nucleotide polymorphisms (SNPs) within other IGF pathway genes were associated with IGF-1 levels and risk of developing cancer [53, 85–87]. To date, only two genome-wide association studies (GWAS) have been conducted of serum IGF-1 levels, and seven SNPs were identified by them [20, 21].

In this study, I conducted a GWAS using data on serum IGF-1 levels of participants in the Long Life Family Study, a unique family study in which participants were recruited based on a

healthy aging phenotype. Because the IGF pathway is associated with longevity both in human and animal models [4], performing GWASs on the participants from longevous families may provide additional insights about the biology behind the healthy aging. In addition, specific genetic variants identified in this analysis will be tested in subsequent analyses with IGF-1 levels and with the risk of mortality in LLFS (**Chapter 6, Section 6.3.4**).

5.2 Methods

5.2.1 Study population

Long Life Family Study (LLFS) is multicenter international study of families displaying healthy aging. The study sample is comprised of 4,953 men and women of mostly (99%) European ancestry from families with two generations ascertained (proband generation n = 1,727, offspring generation n = 3,226) and is 55% female. Participants missing genotyping data or serum IGF-1 levels were excluded from this GWAS. It was performed using genotype and phenotype data of the 4,070 participants of European ancestries with available data in LLFS.

Replication analysis of significant and suggestively significant variants was performed with participants from the Framingham Heart Study (FHS). FHS is a community-based study of cardiovascular risk. The study was begun in 1948 and has recruited to date three generations. Replication analysis was complete with the phenotype and genotype data of 2,833 participants (mean age ± SD = 61.09 y ± 9.45 y, 54% women) in the FHS Offspring exam 7 cohort (generation 3, recruited 2002–04).

5.2.2 IGF-1 measurement

Serum IGF-1 levels in LLFS were measured using a solid-phase enzyme-linked chemiluminescent immunoassay on an Immulite 2000 system (Siemens Healthcare Diagnostics, Inc.). The inter-assay coefficient of variability was 8.7%. More details are available in **Section 2.2**

In FHS, standardized immunoassay (ELISA) was used to measure serum IGF-1. Quality control measures according to strict protocol were performed, and the intraassay coefficient of variation for serum IGF-1 was 5.3%

5.2.3 Phenotype and covariates

To approximate normality and mitigate test statistic inflation due to a non-normal phenotype distribution, IGF-1 was transformed using indirect rank-based inverse normal transformation as two-stage approach for genetic association analysis [88]. First, serum IGF-1 levels (Y) were regressed on covariates (X: age, age^2, and sex).

$$\text{Stage 1: Regress } Y \sim X, \text{ giving } \varepsilon = Y - \hat{\beta}X$$

An inverse transformation was applied to the residuals (ε), and these values were used in the GWAS, adjusting for the X covariates again.

$$\text{Stage 2: Test genotype } (G) \text{ association based on regression } \varepsilon \sim X + G$$

BMI was transformed using a direct rank-based inverse normal transformation. The variables used as covariates in the GWAS models were age, sex, transformed BMI, age × transformed BMI, and field center.

Parallel transformations and analyses were performed with the FHS replication sample.

5.2.4 Genotyping and Imputation

The Center for Inherited Disease Research (CIDR) assayed all LLFS subjects using the Illumina Human Omni 2.5 v1 chip [62]. Additional imputed genotypes were generated based on the cosmopolitan-phased haplotypes of the Haplotype Reference Consortium (HRC) build 37 [63]. Quality control (QC) was carried out by LLFS investigators in the Division of Statistical Genomics, Washington University in Saint Louis. QC that was performed before imputation included checking of Mendelian errors and verification of reported pedigree relationship using GRR (graphical representation of relationship errors). SNPs with high Mendelian error rates were dropped (n = 3,647); SNPs with a call rate < 98% per marker were also dropped (n = 83,774). Eighteen subjects with a call rate < 97.5% were dropped. Genotypes determined to be Mendelian errors were set to missing (n = 153,363). In the GWAS, I included all assayed and imputed genotypes for SNPs with a minor allele frequency ≥ 0.004. I also removed poorly imputed SNPs (imputation quality score of $r^2 < 0.3$) as well as variants that were not in Hardy–Weinberg equilibrium (exact test p value < 0.0001). The GWAS included 9,354,374 observed and imputed SNPs on the autosomes and 286,048 observed and imputed SNPs on chromosome X.

In FHS, genotyping was performed on the Affymetrix GeneChip Human Mapping 500K Array Set and 50K Human Gene Focused Panel. The mean call rate was 98%, and the quality control filtration included removal of SNPs with call rate < 95% or a Hardy–Weinberg equilibrium $p < 10^{-6}$ [20]. Imputation of genotypes in FHS had been performed using 1000 Genomes as the reference panel.

5.2.5 Genetic analyses

5.2.5.1 GWAS discovery analyses

To identify possible genetic variants associated with serum IGF-1 levels in LLFS, I conducted GWA using the genotype data on all participants of LLFS. To account for population and pedigree structure, mixed-model association analysis was performed using the *GENESIS* package for R [64]. Serum IGF-1 level, transformed as described in **Section 5.2.3**, was the outcome variable. Age at enrollment, sex, transformed BMI, age × transformed BMI, field center, and principal components of ancestry were included as fixed-effects covariates and a kinship matrix was included as a random-effects covariate, all as described in **Section 5.2.3**. The principal components of ancestry and kinship coefficients were generated using PC-AiR and PC-Relate, respectively, within the *GENESIS* package [89]. To assess any potential systemic bias, I calculated the genomic inflation factor (λ_{GC}) was calculated and plotted a quantile–quantile (QQ) plot of the tests' *p* values. $p < 5 \times 10^{-8}$ was used for genome-wide significance, and $p < 5 \times 10^{-6}$ was used as for suggestively significance. Visualization of the region of interest was performed using LocusZoom [90]. The GWAS Catalog [91] and Variant Effect Predictor (VEP) [92] were used to annotate SNPs of interest.

5.2.5.2 Candidate gene analysis

In addition to the agnostic GWAS, I also tested for association between serum IGF-1 levels and SNPs within genes that either were associated with IGF-1 in previous GWAS studies [21] **(Table 1)** or reported as candidate genes because of their membership in the IGF pathway [53]. In total, I assessed SNPs within nine genes (*IGF1*, *FOXO3*, *NUBP2*, *GHSR*, *TNS3*, *SST*, *SSTR5*, *IGFALS*, *IGFBP3*). After determining the boundary of each gene (transcription start to transcription end)

using Locus Zoom [90] and the UCSC Genome Browse [93], independent (uncorrelated) SNPs within each gene were selected for testing by retaining SNPs with LD $r^2 < 0.8$ between them within 50 SNP windows (windows sliding in steps of five SNPs), using PLINK 1.9 [94]. Specific SNPs that had been previously reported to be associated with IGF-1 in prior candidate gene studies were explicitly kept in the list of SNPs for each gene. Each SNP was tested for association using the sample models as in the GWAS. For gene-wise significance levels, a Bonferroni correction was computed for each gene separately (0.05 ÷ number of SNPs in each gene).

5.2.5.3 Replication Study

To replicate the results of the GWAS and the candidate gene study, using an independent sample, those SNPs a discovery $p < 1 \times 10^{-6}$ were submitted to LLFS Data Coordination Center for testing in the FHS sample. A generalized linear model (GLM) was used to test for association. Statistical significance was determined to be $p < 0.00023$, 0.05 ÷ total number of SNPs ($n = 212$) tested for replication.

5.2.5.4 eQTL analysis

The statistically significant SNPs from the GWAS and the suggestively significant SNPs from the GWAS that replicated were examined for evidence of association with expression of genes within particular tissue contexts using the Genotype-Tissue Expression portal (GTEx analysis release V8) [65]. GTEx is a publicly available resource to study tissue-specific gene expression and regulation. Associations could be with any one of up to 18,795 genes within one of up to 49 tissues assessed for eQTL evidence by the GTEx Consortium [65].

5.3 Results

5.3.1 Study characteristics

I analyzed data of 4,070 men and women from LLFS with a mean age of 69.9 y ± 15.5 y, a mean IGF-1 level of 128.6 ng/ml ± 53.2 ng/ml, a mean BMI of 27.1 kg/m^2 ± 4.2 kg/m^2. Women comprised 54.6% of the participants. These values and those of other key characteristics are similar to those presented in **Chapter 2 (Table 2)**.

5.3.2 GWAS Discovery

The distribution of *p* values in the GWAS was not inflated, $\lambda_{GC} = 1.10$. The QQ plot of the *p* values is shown in **Figure 9**. The results of the GWAS are presented in **Figure 10**, and, in this figure, SNPs are plotted on the *x* axis according to their position on each chromosome (1–23, where chromosome 23 is chromosome X) and on the *y* axis is the $-\log_{10} p$ value. The red horizontal line indicates the threshold for the genome wide significance ($p < 5 \times 10^{-8}$), and the blue line indicates the threshold for suggestive significance ($p < 5 \times 10^{-6}$). One SNP, rs72696993 on chromosome 14q21, was associated with IGF-1 levels at genome-wide significance, $p = 4.16 \times 10^{-8}$. The minor allele (MAF = 0.02) was associated with lower IGF-1 levels ($\beta = -0.39$) **(Table 8**).

A LocusZoom plot of the region around the significant SNP is shown in **Figure 11**. The left *y* axis is the $-\log_{10} p$ value of the test statistics between serum IGF-1 levels and each SNP, and the right *y* axis is the recombination rate. SNPs with *p* values > 10^{-2} are displayed in the not figure. The diamond is the "lead SNP" of the locus, rs72696993 ($p = 4.16 \times 10^{-8}$). The colors of the

plotted SNPs correspond to their LD r^2 with the lead SNP. The locations of genes within this region are marked in blue below the plot. *TCL1B* is highlighted with a gray box. rs72696993 and those SNPs in high LD with it (in red, $r^2 \geq 0.8$) are located upstream of *TCL1B*.

Besides this one SNP associated at genome-wide significance with IGF-1, 214 SNPs were associated with IGF-1 at suggestively significant p values $< 1 \times 10^{-6}$.

Table 8. Loci associated with IGF-1 levels at genome-wide significance in LLFS or suggestively associated with IGF-1 levels in LLFS and replicated at Bonferroni-corrected signifiance in FHS.

		Position	Nearest			LLFS			FHS	
SNP	Chr	(bp)	Gene	A1	A2	MAF	*p* value	β	*p* value	β
rs72696993	14	96143822	*TCL6*	T	C	0.020	4.16×10^{-8}	−0.390	0.996	−0.0005
rs700750*	7	46753491	*IGFBP3*	A	C	0.383	2.49×10^{-6}	0.109	6.95×10^{-5}	0.11
rs700752*	7	46753553	*IGFBP3*	G	C	0.357	1.26×10^{-6}	0.114	1.29×10^{-4}	0.11
rs700753*	7	46753684	*IGFBP3*	G	C	0.357	1.26×10^{-6}	0.114	1.86×10^{-4}	0.11
rs856582*	7	46741843	*IGFBP3*	T	C	0.387	4.95×10^{-6}	0.106	2.20×10^{-4}	0.11

* SNPs that were been significantly replicated in FHS
SNP, single nucleotide polymorphism; A1, minor allele; A2, reference allele; MAF, minor allele frequency; Chr, chromosome

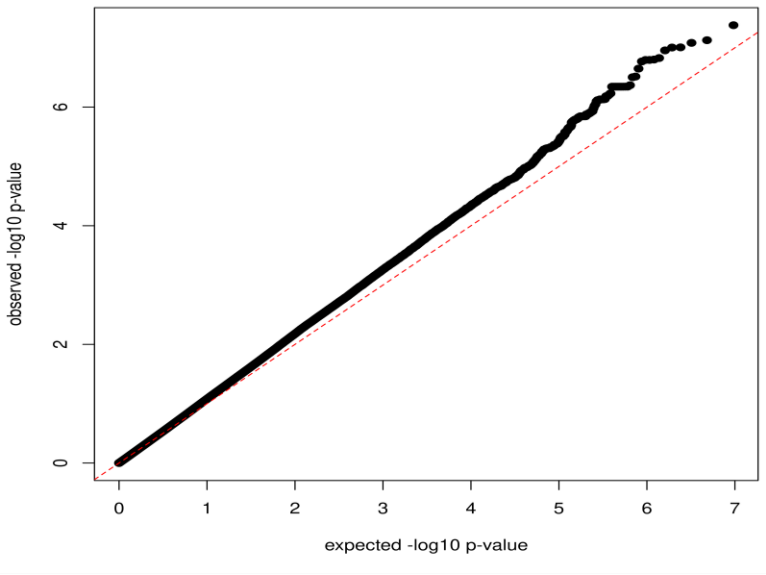


Figure 9. QQ plot of discovery GWAS *p* values for serum IGF-1 levels in LLFS.

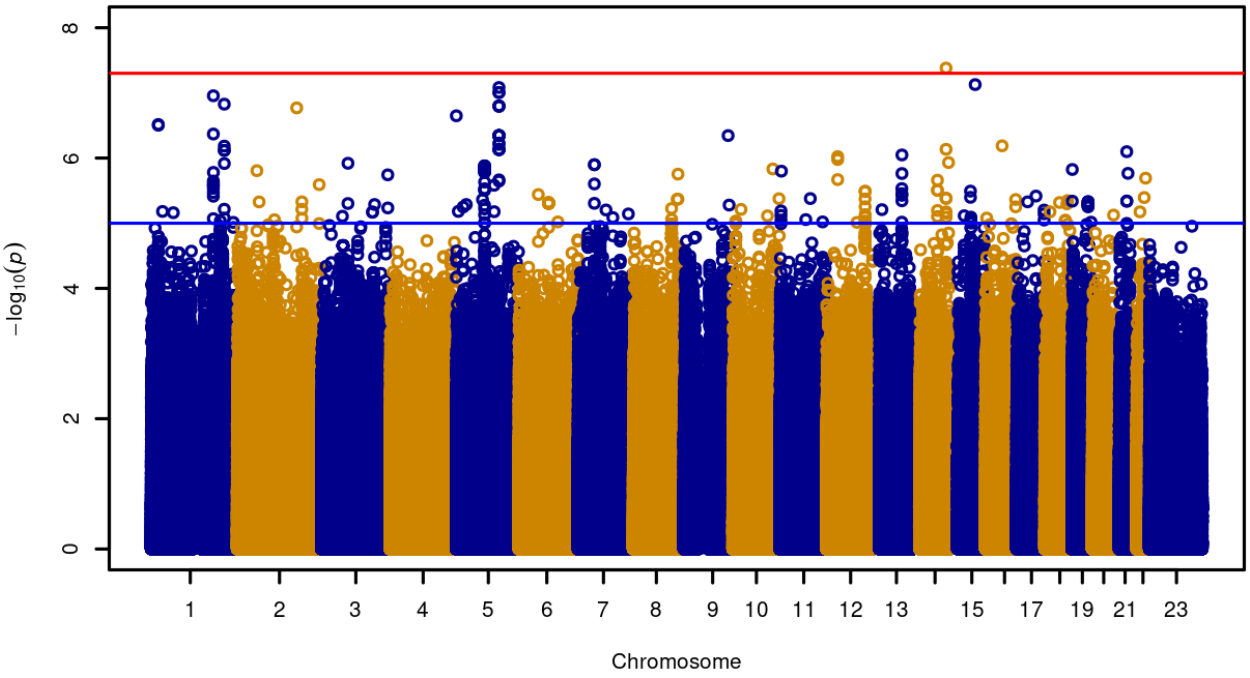


Figure 10. Manhattan plot of discovery GWAS *p* values for serum IGF-1 levels in LLFS.

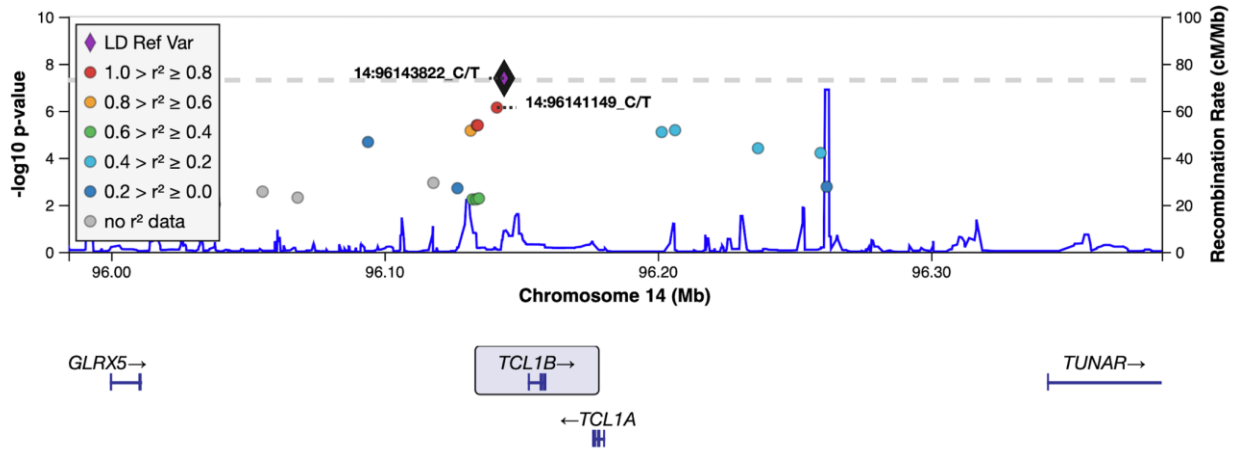


Figure 11. Regional LocusZoom plot around rs72696993 on chromosome 14.

SNPs from nine candidate genes were also examined for association in LLFS (**Table 9**). The number of SNPs tested in each candidate gene ranged from 36 SNPs for *IGFBP3* to 286 SNPs for *FOXO3*. After adjusting for multiple testing within each gene, one SNP, rs12313279 in *IGF1*, was associated with serum IGF-I levels ($p = 2.20 \times 10^{-4}$, $\beta = -0.035$, MAF = 0.30). The minor allele was associated with lower IGF-1 levels.

Table 9. Genes tested in the candidate gene analysis.

Study	Gene	Chr	SNPs Tested (*n*)	Significance Threshold	SNPs significant in LLFS
Teumer et al.	*IGF1*	12	218	2.29×10^{-4}	rs12313279
	FOXO3	6	286	1.75×10^{-4}	—
	NUBP2	16	185	2.70×10^{-4}	—
	GHSR	3	94	5.32×10^{-4}	—
	TNS3	7	784	6.38×10^{-5}	—
	IGFBP3	7	36	1.39×10^{-3}	—
Fangyi et al.	*SSTR5*	12	243	2.06×10^{-4}	—
	IGFALS	16	188	2.66×10^{-4}	—
	SST	3	110	4.55×10^{-4}	—

5.3.3 Replication

The study participants in FHS had a mean age of 61.1 y ± 9.5 y, and 54.3% were female. The mean IGF-1 levels were 113.4 ng/ml ± 36.5 ng/ml, and the mean BMI was 28.2 kg/m^2 ± 5.30 kg/m^2. Of the 216 SNPs prioritized for replication (215 SNPs from the discovery GWAS, and one SNP from candidate gene analyses [rs12313279]), 212 SNPs were available for replication in FHS.

After Bonferroni correction for 216 tests, statistical significance was $p < 0.00023$. Four SNPs were statistically significant: rs700750, rs700752, rs700753, and rs856582 on chromosome 7p12.3, all located within *AC011294.1* and 781 kbp–793 kbp upstream of *IGFBP3* **(Table 8).** LocusZoom was used to examine the region around these significant SNPs using the *p* values from the GWAS **(Figure 12)**. In this figure, the left *y* axis indicates $-\log_{10} p$ value for association with serum IGF-1 levels from the discovery GWAS, and the right *y* axis indicates recombination rate. The diamond indicates rs700752 (the lead SNP) with the most significant discovery *p* value. The SNPs colors indicate LD (r^2) with the lead SNP. The locations of genes within this region are marked in blue below the plot. SNPs with $p > 10^{-2}$ are not plotted. Two of the IGF-1 binding proteins, *IGFBP3* and *IGFBP1,* lie approximately 790 kbp upstream of the four SNPs. The effect sizes of the four SNPs on IGF-1 levels were nearly the same in LLFS and FHS (0.11 ng/ml). The discovery genome-wide significant SNP, rs72696993 on chromosome 14, was not replicated in the FHS sample ($p = 0.996$).

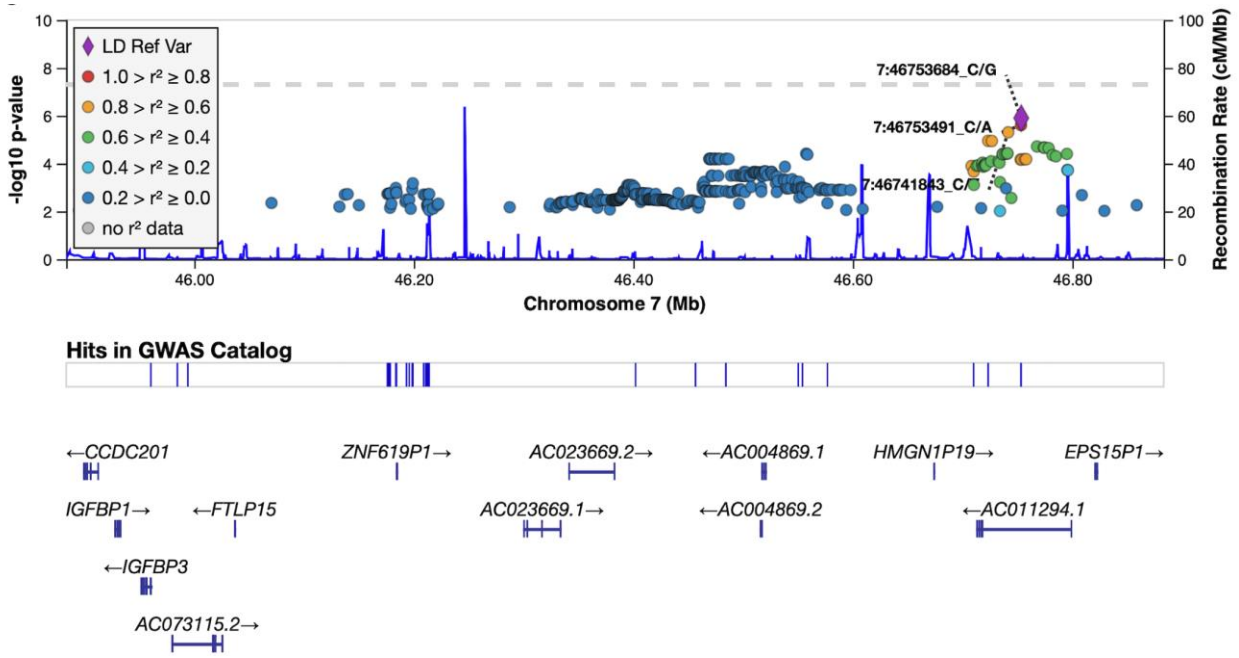


Figure 12. Regional LocusZoom plot of significant replicated SNPs (rs700750 [7: 46753491_C/A], rs700752 [purple diamond], rs700753 [7:46753684_C/G], and rs856582 [7:46741843_C/T]) using LLFS discovery *p* values.

5.3.4 eQTL of the significant replicated SNPs

The GTEx portal was used to check for evidence that significant SNPs (either from the discovery or replication GWAS) were also eQTLs. The genome-wide significant SNP rs72696993 (on chromosome 14 near *TCF6*) from the discovery GWAS was associated with *TCL6* expression in whole blood ($p = 1.9 \times 10^{-5}$). None of the four SNPs on chromosome 7 near *IGFBP3* that were replicated in FHS (rs700750, rs700752, rs700753, and rs856582) were associated with the expression of any gene in any tissue in GTEx.

5.4 Discussion

The GWAS of serum IGF-1 levels in 4,070 participants from LLFS identified a single SNP associated with IGF-1 at genome-wide significance in the discovery analysis (rs72696993 on 14q21, $p = 4.16 \times 10^{-8}$). This SNP had not been previously reported to be associated with serum IGF-1 levels. It is located near *TCL6*, which has been associated with QT interval [95]. Furthermore, there is evidence that this variant affects *TCL6* expression in whole blood ($p = 1.9 \times 10^{-5}$), although how *TCL6* and its expression might be connected to serum IGF-1 levels, beyond the association observed here, is unclear. Additionally, association between this variant and IGF-1 levels was not observed in FHS. Thus, this result may be a false positive.

The discovery analysis also identified 214 SNPs that were associated with serum IGF-1 levels at $p < 1 \times 10^{-6}$. Furthermore, examination of SNPs in nine candidate gene identified an additional SNP (rs12313279 at 12q23.2 in *IGF1*) associated with IGF-1 levels. This SNP has previously been reported to be associated with brain stem volume [96].

To confirm the discovery GWAS results, a replication analysis of 212 SNPs was performed using genotype and phenotype data from FHS. (Of the 215 SNPs to be replicated here, 212 were available for analysis in FHS.) After a Bonferroni correction per each gene was applied, four SNPs were replicated: rs700750, rs700752, rs700753, and rs856582 ($p \leq 0.00023$). The effect size and direction of these SNPs on IGF-1 were similar in both LLFS and FHS. These SNPs are on chromosome 7p12.3 in *AC011294.1* and are 790 kbp upstream of *IGFBP3* and *IGFBP1*. This region has previously been associated with IGF-1 and IGFBP3 levels in GWASs [20, 21]. In addition, this region has also been associated with glomerular filtration rate [97], thyroid stimulating hormone levels [98], urate levels [99], and obesity [19].

The repeated association of this region on 7p12.3 with IGF-1 levels in multiple studies is evidence that this region is truly associated with variation in serum IGF-1 levels potentially by affecting IGF-1 binding protein IGFBP3. However, the precise mechanism by which genetic variants in this region affect IGF-1 levels—even through effects on IGFBP3—remains unknown and is worthy of additional study. In my study, I did not observe any SNPs within *IGFBP*3 itself that were associated with IGF-1 levels. If it is *AC011294.1* and not *IGFBP3* that is affecting IGF-1 levels in this region on chromosome 7p12.3, the mechanism of action for this gene is unclear.

The main limitation of this analysis was the absence of measurements of important factors that affect serum IGF-1 levels, such as IGFBP3, which is the main carrier of IGF-1 in the circulation before it binds to its receptor. Kaplan et al. [20], in their GWAS, observed that the association between serum IGF-1 levels and rs700752 was attenuated when the analysis was conditioned on IGFBP3 levels. In addition, SNPs within the *IGFBP3* gene became significantly associated with IGF-1 after adjusting for IGFBP3 concentration. I was unable to test for any relationship between IGFBP3 levels and this same region around and within *IGFBP3* on chromosome 7p12.3.

A strength of this study was the replication of the results in another independent population-based sample (FHS). FHS has an age range similar to LLFS. Having FHS available to serve as a replication sample heightens the confidence in the findings on chromosome 7, which first observed in LLFS, were replicated in FHS.

Another limitation is the sample size of the genetic analysis in LLFS, which is underpowered given typical common variant effect sizes on quantitative phenotypes. Furthermore, LLFS is a unique population given the focused ascertainment of healthy long-lived participants, so the distribution of IGF-1 and its relationship to genetic variants might not be representative of

the general population. Genetic studies conducted with this unique study (compared to studies with samples of the general population) may reveal additional shared SNPs associated with serum IGF-1 (which is known to be connected to aging) variation between/within these families.

In summary, the GWAS of serum IGF-1 levels in LLFS and replicated in FHS did not identify any new loci that affect IGF-1 levels. The analysis confirmed previously observed associations between serum IGF-1 levels and variants in *AC011294.1* on chromosome 7p12.3, located approximately 800 kb upstream of *IGFBP3* gene, which has been associated with other aging-related traits, such as thyroid stimulating hormone levels, obesity, and glomerular filtration rate of the kidneys. One of the main study limitations is the absence of data on serum IGFBP3 levels and free IGF-1 levels, which are important aspects of the metabolism of IGF-1 [100]. Future studies to assess the association of these significant SNPs with aging-related diseases such as diabetes, cancer, and mortality is recommended. I follow up these associated variants in **Section 6.3.4** in a study of IGF-1 and the risk of mortality. Using both the phenotypic information—such as age, sex, BMI, IGF-1 levels—and the genetic information from associations may inform categorization of individuals at risk of early mortality.

6.0 SURVIVAL ANALYSIS AND THE RISK OF MORTALITY IN LLFS

6.1 Introduction

The increase in age-related morbidity and mortality from greater numbers of persons living to age 65 and older is expected to result in a significant rise in health services demands [1, 2]. Although the determinants of variation in the process of aging in humans are not fully understood, evidence from model organism studies suggests the involvement of numerous candidate genes for longevity [5, 8, 9]. The IGF pathway, which plays an essential role in growth, development, and metabolism, has been recognized as a key regulator of aging and longevity. However, epidemiologic studies investigating the relationship between serum IGF-1 levels, which is the main ligand in the IGF pathway, and age-related diseases and survival have reported inconsistent findings. An inverse relationship between serum IGF-1 levels and mortality in the general population has been suggested by some investigators [15, 34, 36, 101, 102], whereas other investigators reported a positive relationship [17, 103]. Yet other investigators concluded that there was no association between IGF-1 serum levels and mortality, but suggested a greater association between mortality and higher levels of IGFBPs [39, 104].

The contradictory and inconclusive reports may be attributed to multiple factors. First, several studies were conducted in high-risk populations, and the activity of the IGF pathway and IGF-1 are known to be affected by acute and chronic disease [105]. Thus, some of the reported relationships between IGF-1 levels and mortality might be due to reverse causation [17]. Second, poor nutritional status, immobility, and muscle weakness, which are generally associated with lower IGF-1, may also confound the association between IGF-1 and mortality. Furthermore,

several of these studies were conducted with participants from a wide age range, and serum IGF-1 levels are attenuated by age, that is, younger adults have higher serum IGF-1 levels than older adults [35, 104].

To date, no study has examined the relationship between genetic variants that are associated with serum IGF-1 levels and the risk of mortality. Using, as a predictor variable, the genotypes of SNPs associated with variation in serum IGF-1 levels to interrogate their effects on the risk of mortality might provide more insight regarding mechanisms and may enable mortality prediction. The aim of this chapter is to describe effects of sex, IGF-1 tertiles, serum IGF-1 levels, and SNP genotypes on the probability of survival by calculating Kaplan–Meier (KM) estimators and applying Cox mixed-model regression.

6.2 Method

6.2.1 Study population

The analysis for this aim uses the proband generation of LLFS [56]. The average age of the proband generation was 91.8 y, and more individuals in this generation were likely to die over the course of the study than those in the offspring generation. Mortality data for the LLFS proband generation was taken from death reports for the first ten years (2007–17) of follow-up. The total number of proband participants was 1,581. A set of 96 participants were excluded due to absence of serum IGF-1 measurements, and three participants were excluded because there was no information regarding death status. Thus, data was available for 1,482 proband participants.

6.2.2 Phenotype, predictors, and covariates

Death status (status = 1 is for dead, status = 0 is for alive) and the survival time as calculated from the initial visit date to either the death date or the last date of follow-up were the main phenotype. Serum IGF-1 levels were measured as described in **Chapter 2 Section 2.2**. The participants' age, sex, anthropometric and physical health characteristic were obtained from the in-person visit 1 examination. BMI was calculated as weight in kilograms per the square of the height in meters. The genotypes for four SNPs (rs700750, rs700752, rs700753, and rs856582) that had been associated with IGF-1 in LLFS and replicated in FHS, as described in **Chapter 5**, were also to be used as predictors. The SNPs of interest were extracted from the variant call format (VCF) file of all LLFS genotypes information on Unix using bcftools version 1.9-40s, and the genotypes of each SNP were merged with the phenotype data file.

6.2.3 Statistical analyses

To reduce nonnormality, serum IGF-1 levels and BMI were log transformed. A *t* test was used to assess whether the serum IGF-1 level differed with statistical significance between LLFS male and female participants in the proband generation. In addition, ANOVA was used to test for mean differences in age among IGF-1 tertiles. A significant ANOVA test statistic was followed by computing Tukey honest significant difference test statistics for pairwise-comparisons between the group means.

6.2.3.1 Kaplan–Meier estimation of survival proportions

Using the *Survival* package and its survfit() function, nonparametric KM estimators were computed to estimate the probability of survival in the LLFS proband participants over the ten years of follow-up. A KM estimator is a statistic used to measure the fraction of survivors within a specific amount of time (not controlling for covariates) with three assumptions [106]: (1) individuals who are censored at any time are assumed to have the same survival prospects as those who continue to be followed, (2) the survival probability is assumed to be the same for those who were recruited early and late in the study, and (3) the event is assumed to happen at the time specified. The survival probability at any specific time calculated as:

$$\hat{S}(t) = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

For each time t_i, the survival probability $\hat{S}(t)$ is the number of subjects surviving d_i divided by number of individuals at risk n_i. So, at each time point in the study when an individual was lost to follow-up or died, the proportion of individuals who survived to that point (survival) and the standard error and 95% confidence interval (CI) for that proportion is calculated.

KM curves were generated in the total proband generation sample, then stratified by sex in the total sample, by IGF-1 tertile in the total sample, and finally by IGF-1 tertile separately in male and female participants. To test whether there was a statistically significant survival proportion difference between or among the stratifications, the log-rank test was used. This tests the hypothesis that the survival curves (KM curve) of the groups differ against the null hypothesis that they do not.

6.2.3.2 Cox regression

To test for the effects of other variables on survival, accounting for covariates, and to estimate the association between serum IGF-1 level as a continuous variable and the risk of mortality, Cox mixed-model (proportional hazard) regression was used. Hazard is defined as the probability of dying at a given time, and the hazard ratio (HR) is the ratio of the risk of hazard occurring at any given time in one group compared to risk in the other at the same time. An HR > 1 indicates a higher risk of death by a specific condition. An HR < 1, on the other hand indicates a lower risk of death. To perform a mixed-model form of the regression, I used the *Coxme* package and the coxme() function. I adjusted for relatedness as a random effect using the varlist() function and the pedigree-derived kinship matrix. In addition, to account for fixed-effect covariates, the same analysis was performed adjusting for sex and BMI first, then age. Similarly, the hazard risk of mortality was tested with IGF-1 tertiles as a fixed-effect predictor variable (the 2nd tertile was the reference group) and kinship included as a random-effect covariate. Then, Cox regression adjusted for sex and BMI as fixed-effect covariates and kinship as a random-effect covariate was performed, and then finally age was added as a fixed-effect covariate.

I next assessed whether the risk of mortality can be predicted by IGF-1–associated SNPs obtained from the analyses in **Chapter 5**. I calculated the total death count per genotype of each SNP and the mean IGF-1 level between genotypes. Using coxme() as above, I performed Cox regression of survival against each SNP, coded additively where the coded allele was the minor allele, with kinship as a random-effect covariate to estimate the Hazard ratios (HRs). Then, Cox regression adjusted for sex and BMI as fixed-effect covariates and kinship as a random-effect covariate was performed, and then finally age was added as a fixed-effect covariate.

All the analysis were done using R version 3.4.0 [60].

6.3 Results

6.3.1 Baseline characteristics of the study participants

The study included 1,482 LLFS proband participants, whose baseline characteristics are shown in **Table 10.** The age range of participants was 49 y–110 y, with no significant difference in mean age between male and female participants (89.4 y vs 89.4 y respectively, $p = 0.99$). Male participants in the LLFS proband generation had higher mean serum IGF-1 than female participants (106.4 ng/ml vs 100.1 ng/ml, respectively, $p = 0.005$). The mean age of participants in the 1st IGF-1 tertile (lowest serum IGF-1 level) was significantly higher than those in the 3rd IGF-1 tertile (91.4 y vs 87.6 y, respectively, $p < 2.2 \times 10^{-16}$**)**. Participant survival did not differ among different recruitment field centers ($p = 0.16$, **Figure D1 Appendix D**).

6.3.2 Survival proportions

Over the course of ten years of follow-up, approximately 70% of the total proband generation had died; the median survival time was 1,983 days **(Figure 13)**. In the KM curves visualized in **Figures 13–15**, the *y* axis is the survival probability, and the *x* axis is the time in days from the first interview visit to the survival event (death or end of follow-up). At time zero, the survival probability is 1.0 meaning 100% of the proband generation is alive. The median survival time (the time at which 50% of the proband generation had died) was 2,367 days.

Survival curves differed significantly between male and female participants of the proband generation, the median survival time was higher in female (2,187 d) compared to male (1,787 d) participants, and this difference was significant ($p = 0.0004$) **(Figure 14)**.

Table 10. Key characteristics of LLFS proband generation participants.

	Female Participants (n = 816)	Male Participants (n = 668)
Age at Enrollment (y), mean ± SD	89.4 ± 7.7	89.4 ± 5.8
IGF-1 (ng/ml), mean ± SD	100.1 ± 45.0	106.4 ± 42.0
BMI (kg/m^2), mean ± SD	26.0 ± 4.7	26.4 ± 3.6
Smoking, %	2.2%	1.8%
Hypertension, %	72.0%	60.0%
Diabetes, %	8.0%	11.5%

	1st tertile (n = 495)	2nd tertile (n =495)	3rd tertile (n = 494)
Female, %	59%	57%	49%
Age at Enrollment (y), mean ± SD	91.4 ± 5.9	89.3 ± 6.6	87.6 ± 7.5
IGF-1 (ng/ml), mean ± SD	60.8 ± 13.2	96.9 ± 10.1	151.2 ± 36.4
BMI (kg/m^2), mean ± SD	25.4 ± 4.0	26.2 ± 4.0	26.9 ± 4.5

	Status = 1, died (n = 1020)	Status = 0, survived (n = 464)
Female, %	50.6%	64.4%
Age at Enrollment (y), mean ± SD	91.8 ± 5.3	84.2 ± 7.1
IGF-1 (ng/ml), mean ± SD	98.7 ± 42.3	112.2 ± 45.3
BMI (kg/m^2), mean ± SD	25.8 ± 4.1	27.1 ± 4.4

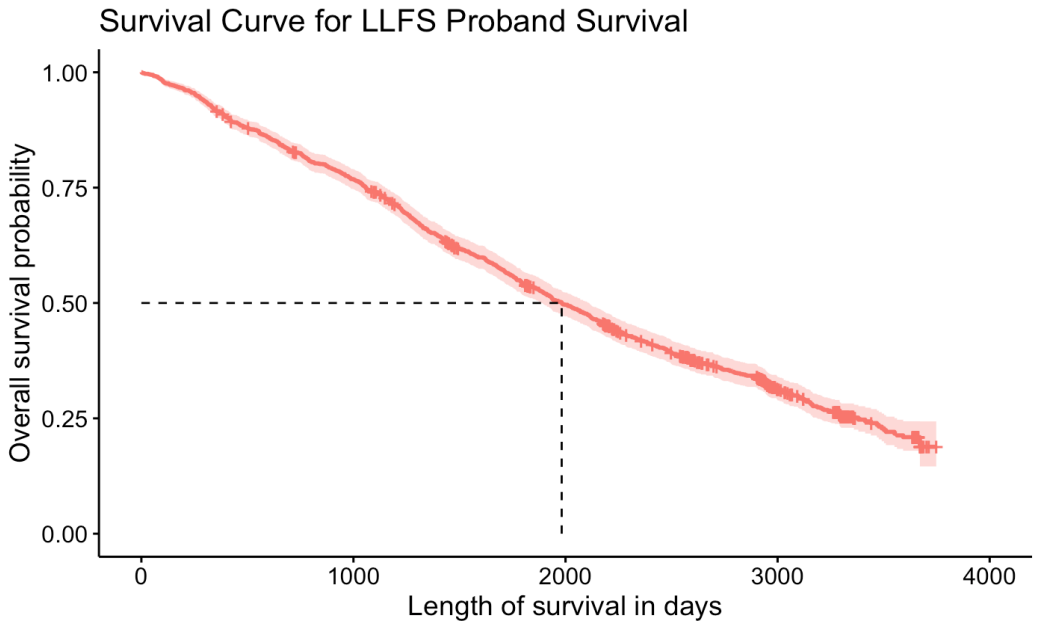


Figure 13. The KM curve of the survival probability proportion in the LLFS proband generation. The black dotted line is the median survival time.

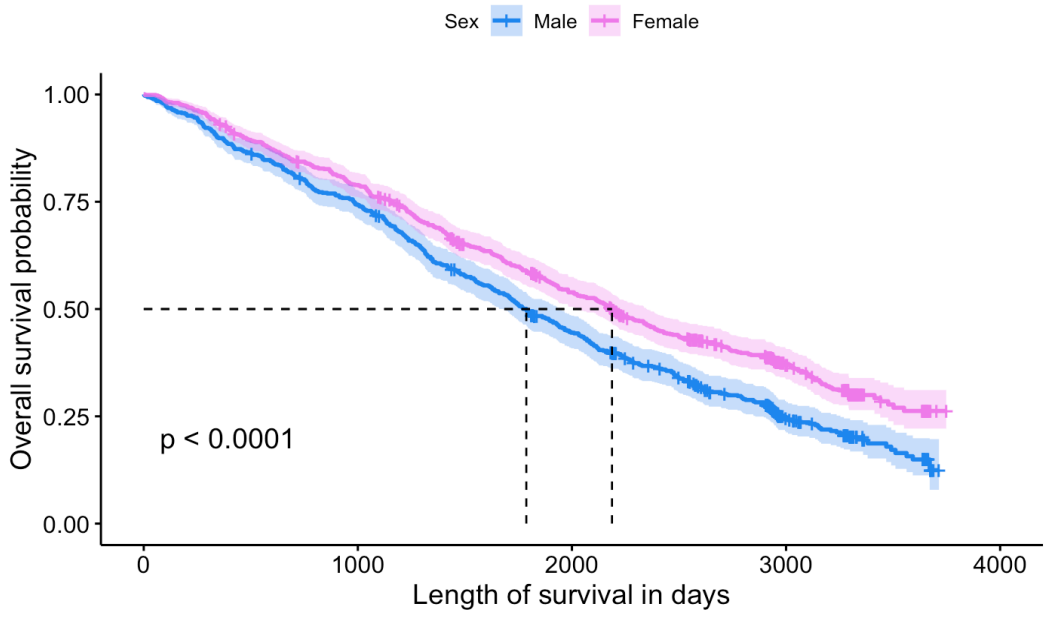


Figure 14. The KM curve of the survival probability between male and female participants in the LLFS proband generation. The black dotted line is the median survival time.

Comparing the KM curves among the IGF-1 tertiles in the total proband generation, participants in the 3rd tertile (highest IGF-1 levels) had a higher survival rate than in the 2nd tertile which were higher than in the 1st tertile (lowest IGF-1 levels) (**Figure 15a**). The difference among the three curves was statistically significant ($p < 0.0001$); a formal analysis of the differences pairwise is given in the next section.

When survival was stratified by sex and IGF-1 tertile, there was a statistically significant survival proportion difference among the tertiles curves, in both male ($p < 0.0001$, **Figure 15b**) and female ($p < 0.0001$, **Figure 15b**) participants. A similar pattern, where 3rd tertile participants had higher survival than 2nd tertile participants which had higher survival and 1st tertile participants. Again, a formal analysis of the differences pairwise is given in the next section.

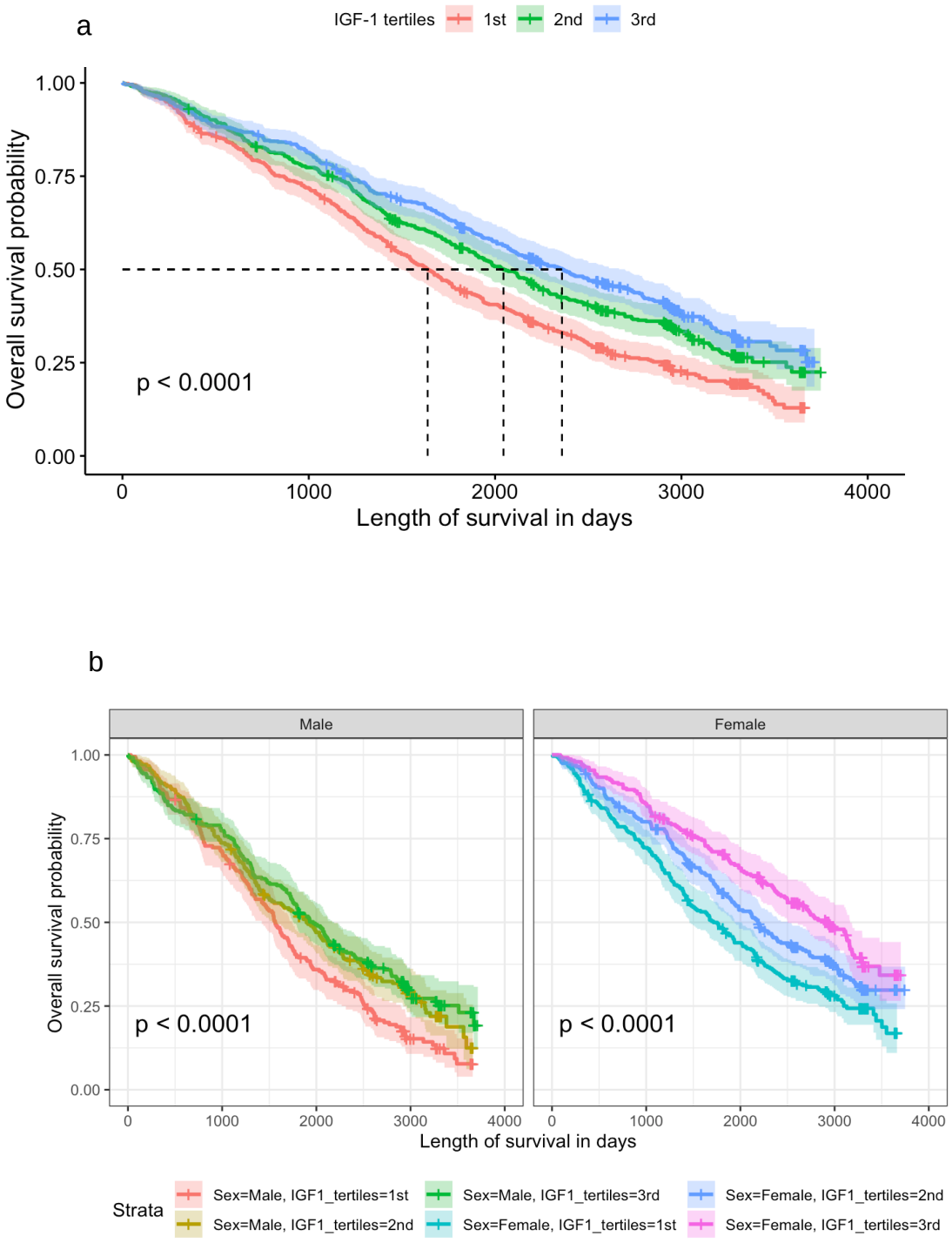
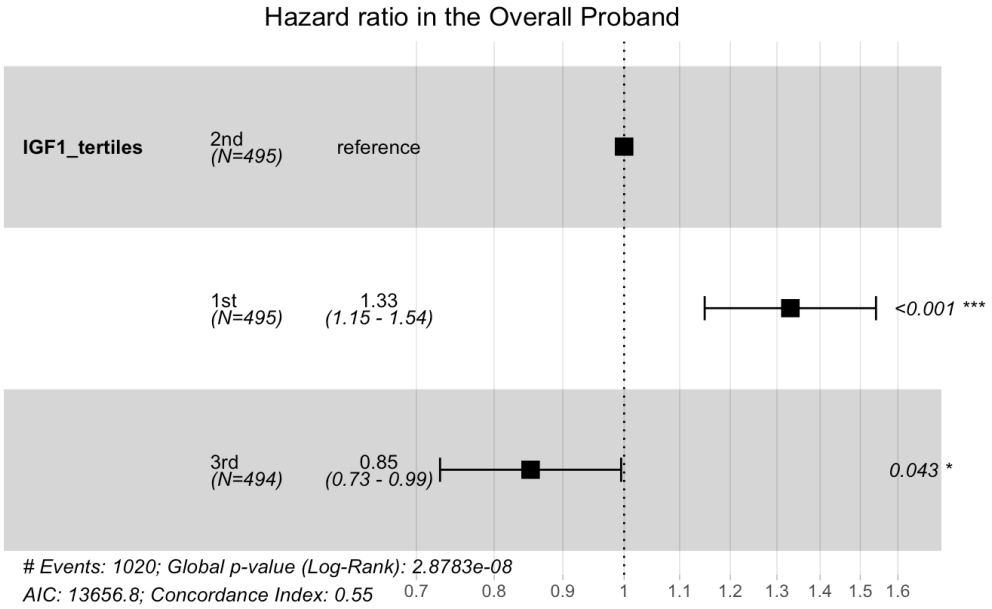


Figure 15. Survival probability between IGF-1 tertiles (a) overall and (b) stratified by sex. The difference in survival between IGF-1 tertile was significant overall and within each sex. IGF-1 is log transformed. The black dotted line in (a) is the median survival time.

6.3.3 Serum IGF-1 levels and the risk of mortality in LLFS

Results of Cox regression of risk of mortality on baseline IGF-1 levels, adjusted only for the relatedness among participants, showed that higher baseline IGF-1 was significantly associated with lower risk of mortality (HR = 0.6, 95% CI = 0.53–0.70, $p = 1.4\times 10^{-10}$).

In the results of regressing risk of mortality on IGF-1 tertiles, participants in the 1st IGF-1 tertile (lowest IGF-1 levels) were at a statistically significant higher risk of death (HR = 1.33, $p = 0.0001$) compared to the 2nd tertile in the total proband generation sample as well as when separately regressed in male (HR = 1.34, $p = 0.009$) and female (HR = 1.34, $p = 0.004$) participants. Furthermore, participants in the 3rd tertile (highest IGF-1 levels) were at a statistically significant lower risk of death (HR = 0.85, $p = 0.04$ compared with the 2nd tertile in the total proband generation sample as well as in female participants (HR = 0.75, $p = 0.01$) (**Figure 16**). There was not a statistically significant lower risk of death in male participants in the 3rd tertile as compared with the 2nd.



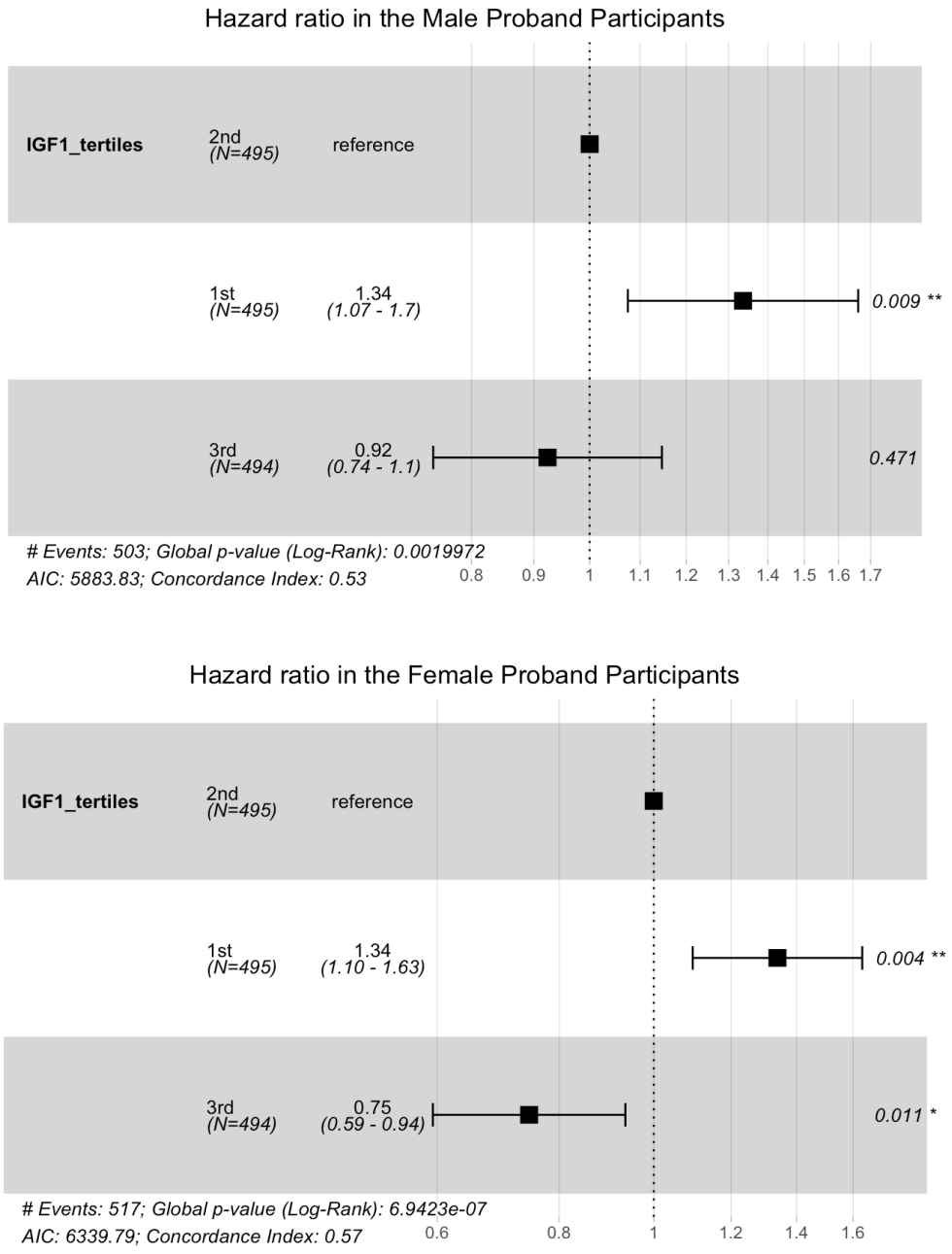


Figure 16. Forest plot of the hazard ratio per IGF-1 tertiles in the overall cohort and statified by sex.

In Cox regression of survival on IGF-1, including kinship, sex, and BMI as covariates, female sex, higher BMI and higher IGF-1 were associated with statistical significance with lower risk of mortality in the LLFS proband generation (**Figure 17**). However, after adjusting for age,

the effect of IGF-1 levels on the risk of death was attenuated (HR = 0.9, 95% CI = 0.7–1.0, p = 0.1), while the protective effect of higher BMI remained statistically significant (**Figure 18**). This attenuation was also observed for IGF-1 tertiles as the predictor of interest (**Figure 19**).

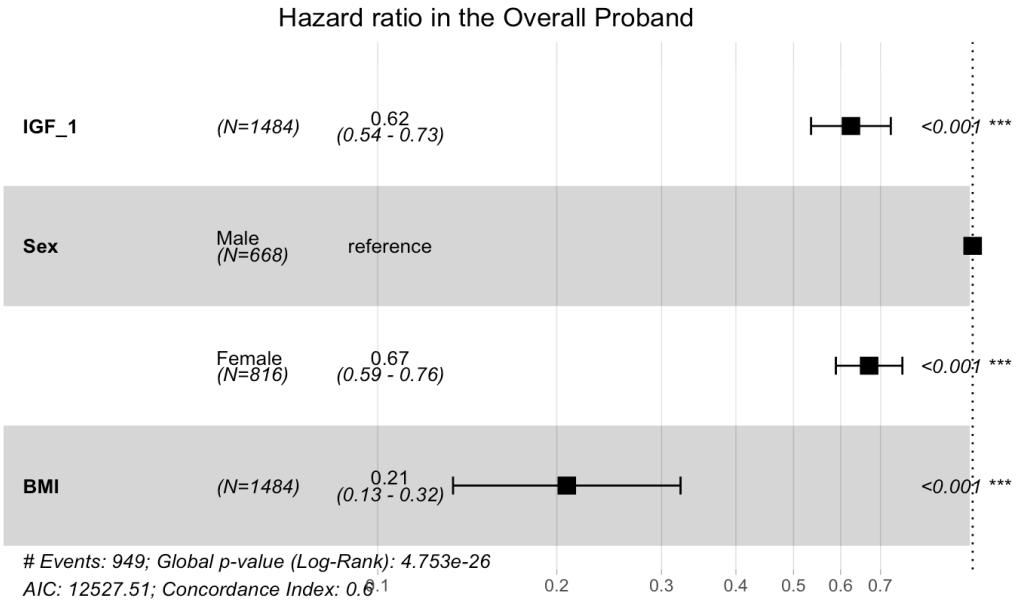


Figure 17. Forest plot of the hazard ratio in multivariable Cox regression model. IGF-1 and BMI are both log transformed.

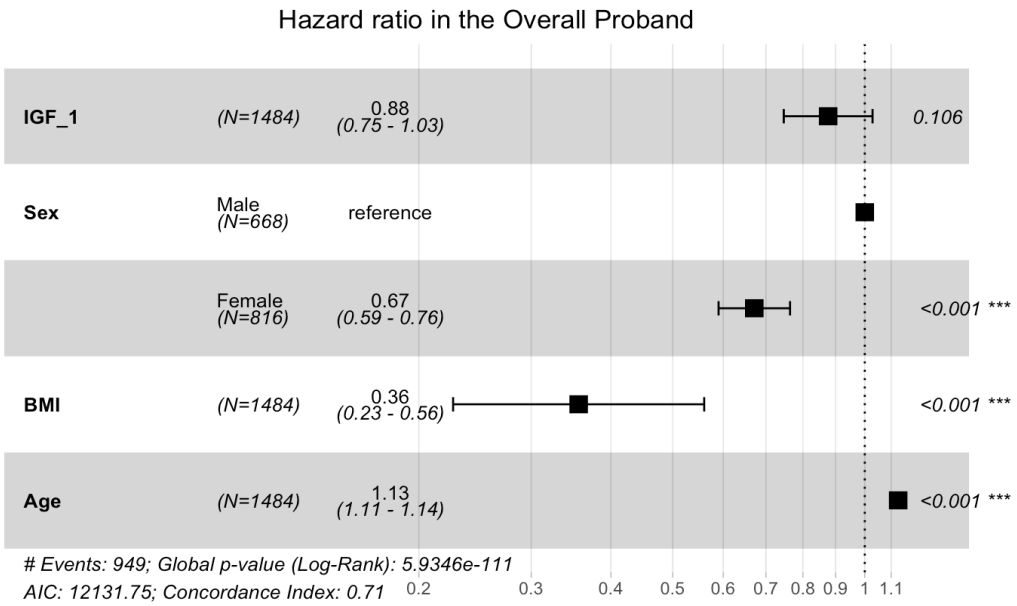
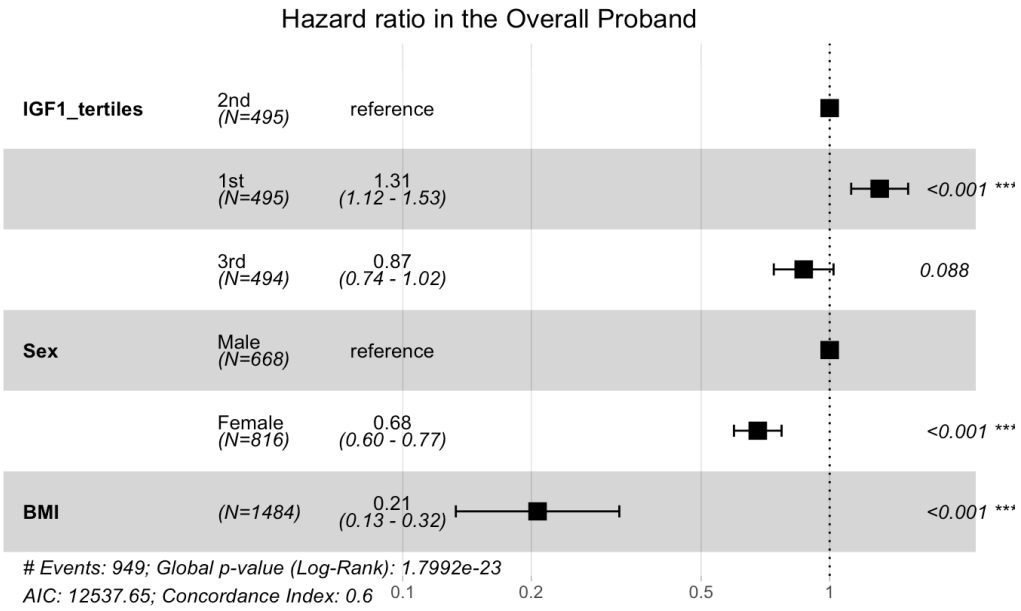


Figure 18. Multivariate cox regression modle after adding the age as covariate with log IGF-1 as continous variable.



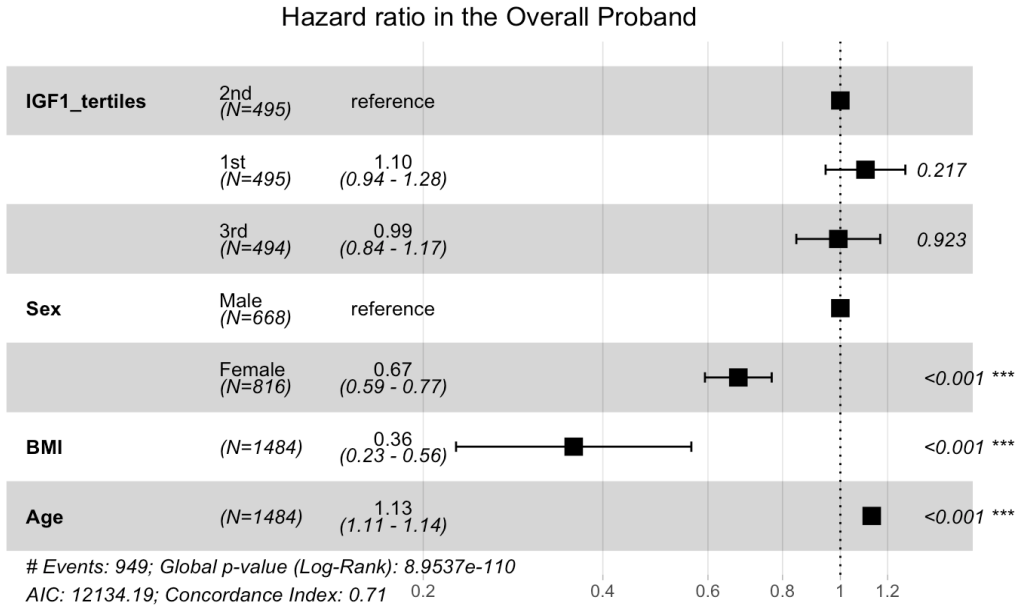


Figure 19. Forest plot of the hazard ratio in multivariate Cox regression model. IGF-1 in modeled as tertiles, without and with age as a covariate.

6.3.4 IGF-1 genetic variation and the risk of mortality

In this analysis to see if SNPs associated with IGF-1 levels affect mortality risk, participants from the LLFS proband generation with missing genotypes were excluded, and the total number of participants available was 1,323. The total number of deaths in this group was 1,020 (77%). The number of deaths (status = 1) by each SNP's genotype is presented in **Figure 20**. Using linear regression, mean log IGF-1 differed by SNP genotype, and the AA genotype (minor allele homozygote) had the lowest level of log IGF-1 for all four SNPs. **(Figure 21).** However, survival probabilities did not differ significantly by genotypes groups for any of the four SNPs (log-rank tests of KM estimators, $p > 0.05$). KM curves for rs700750 genotypes are shown in **Figure 22**. KM curves for the other SNPs are shown in **Figure 2D** in **Appendix D**.

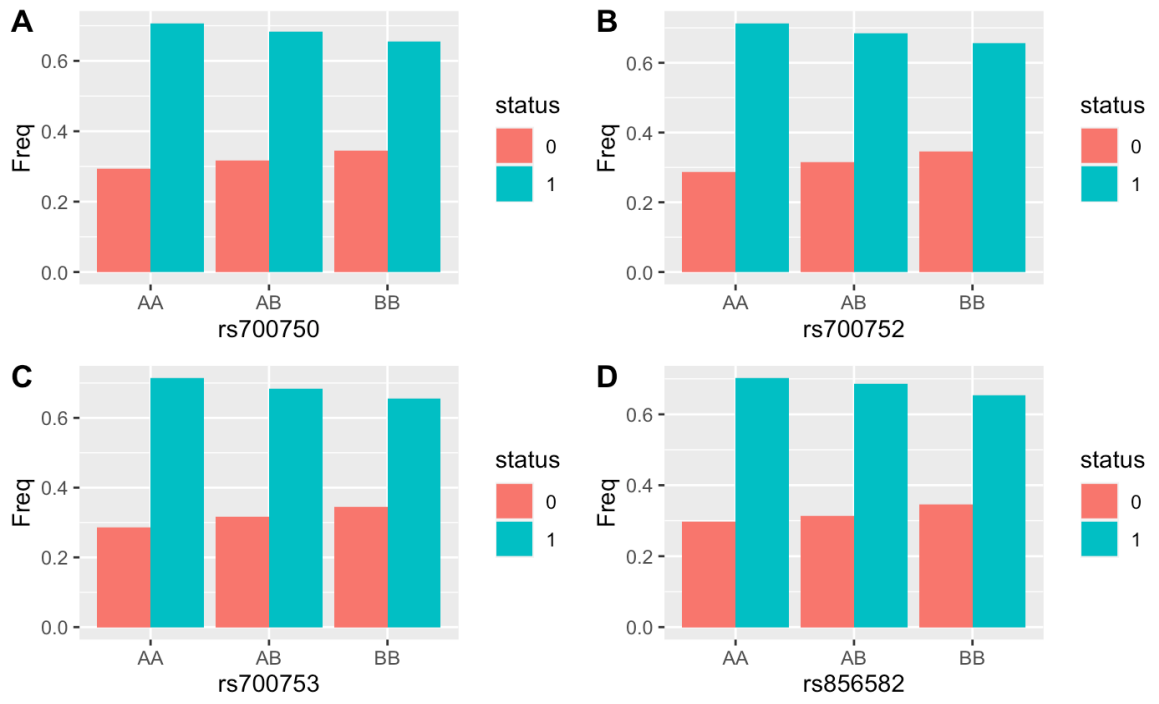


Figure 20. Frequency of participant per status for each SNP genotype group. Status = 0 (survived), status = 1 (died).

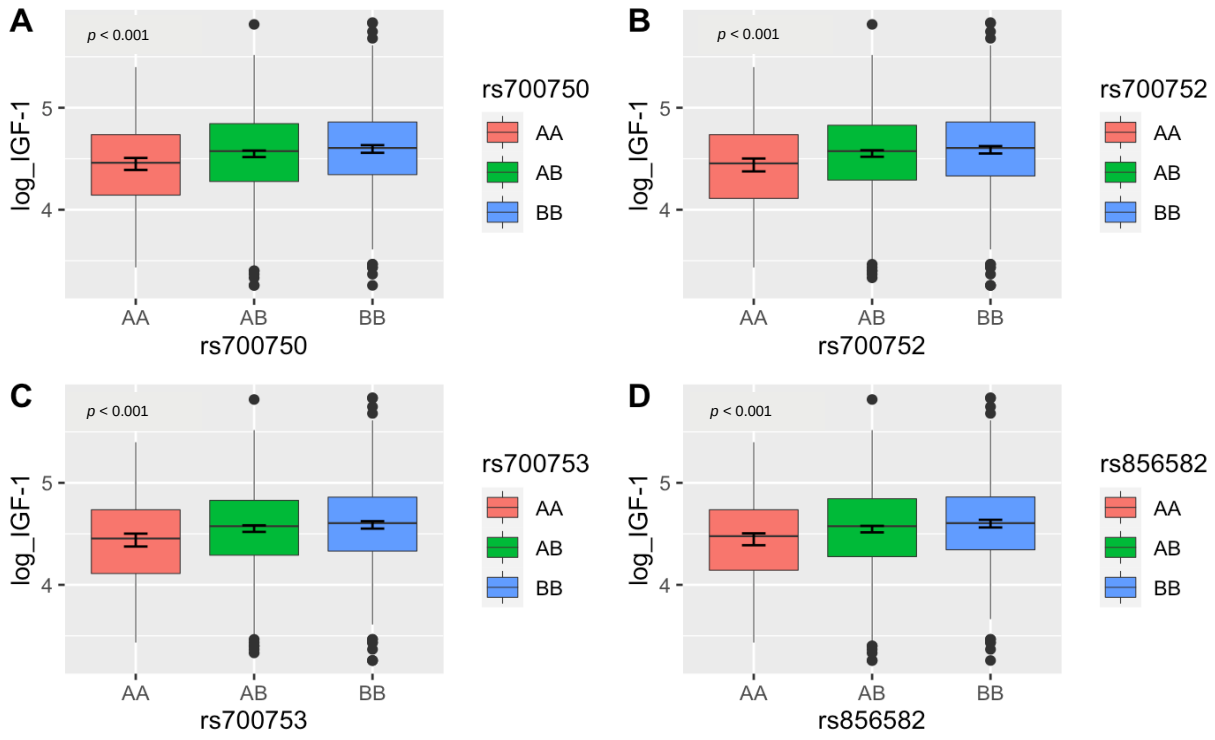


Figure 21. Box plot of log IGF-1 by SNP genotype. Mean IGF-1 difference is statistically significant across SNP genotype groups, $p < 0.001$.

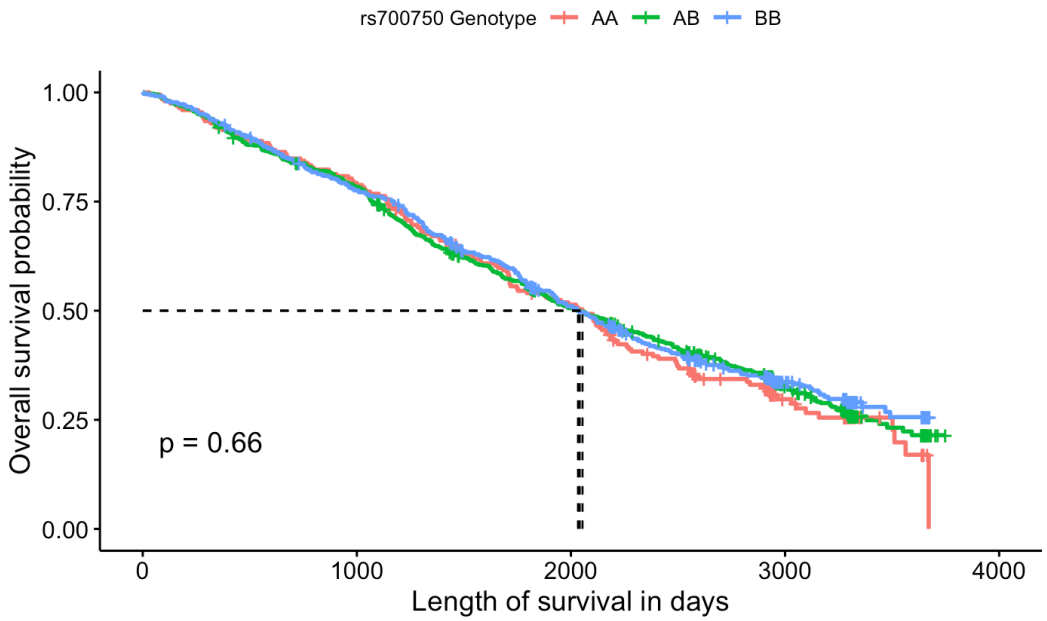


Figure 22. Survival curve of proband within SNP rs700750 genotype groups. No statistical significant difference of surviving within the genotype groups. Black dotted line is the survival median time.

Additionally, Cox regression showed no statistical association between genotype and the risk of mortality in any model (kinship only; kinship, sex, and BMI; and kinship, age, sex, and BMI). In all genetic models, higher age is associated with higher risk of mortality (HR = 1.13, $p < 0.0001$) while female sex and greater BMI are associated with lower risk of mortality (HR = 0.6 and HR = 0.3, respectively, both $p < 0.0001$) **(Figure D3 Appendix D).**

6.4 Discussion

The present study shows significant differences in survival proportions between male and female participants over the course of the ten years of follow-up in the LLFS. It is well established that women live longer than men in most contexts and that this difference is consistent across the lifespan.

Analyses of differences in survival by IGF-1 tertiles and quantitative IGF-1 levels showed consistently showed that higher IGF-1 levels were associated with higher survival and lower risk of mortality, an effect that was attenuated after adjusting for age. This contradicts results observed in mice, associating lower IGF-1 levels and higher survival [9]. This inverse correlation has also been reported in several studies of human participants [15, 34, 102, 103]. However, other studies in humans having findings consistent with what I observed in this study here [15, 33, 35, 104, 107]. The attenuation of the IGF-1 effect after adjusting for age indicates the strong age effect on IGF-1.

The analysis was extended to assess the relationship between IGF-1–associated genetic variants and risk of mortality. Although these SNPs influenced serum IGF-1 variation, they were not associated with mortality status or with differences in survival probability.

The inconsistency of results of association between IGF-1 and mortality in previous studies might be because of the inclusion of wide variety of age ranges in the study samples, and the assumption that the relationship is similar between younger participants, in whom survival to baseline has been high, and older participants, whose characteristics might be skewed due to survival bias [35, 104]. For this reason, the present analysis was performed using data across participants within a narrower age range (the proband generation, range 49 y–110 y old) to minimize the variability of serum IGF-1 levels due to the age effect. (Younger participants have higher IGF-1 levels comparing to older participants.) In addition, the phenotype (mortality) is already known for most of the members of this generation, which is one of the strengths of the study. Finally, participants in this analysis are healthier than participants in other studies [56] thus, minimizing the effect of acute or chronic disease status on serum IGF-1 levels [105]. However, this feature—healthily long-livedness—might also restrict the generalizability of the analysis results of such a unique family-based study to general population. Another study strength is the large sample size and the length of follow-up to allow observation of the end-point (mortality) for most participants, as compared with other studies [34, 39, 102, 103]. Finally, this study is the first to assessing the risk of mortality based on the associated genetic variant with serum IGF-1levels.

One of the study limitations is that the cause of death is unknown, which is necessary to determine whether the association of IGF-1with mortality differs based on the underlying cause of death. In addition, I did not adjust for common survival covariates such as smoking, diabetes, and other chronic disease in these analyses, although since the statistically nonsignificant effect of IGF-1 on mortality is unlikely to strengthen with the inclusion of such covariates. Another limitation is the absence of traits that are known to affect serum IGF-1 function and bioavailability, such as the binding proteins IGFBP-1–6 and levels of free IGF-1, which is more available to attach

with cell receptor. Several investigators have reported an association between IGFBPs and/or free IGF-1 with mortality, but not between serum IGF-1, which is a measurement of bound IGF-1, and mortality [17, 35, 38].

In summary, I observed an association between IGF-1 levels and risk of mortality in the total sample and in the sample stratified by sex. Participants in lower IGF-1 tertile have higher mortality risk than participants in higher IGF-1 tertiles. A similar inverse relationship between IGF-1 levels and risk of mortality was observed in male and female participants, separately. These associations were statistically significant after adjusting for baseline BMI and sex but were not statistically significant after age was included in the models. Future studies of the other components of the IGF-1 pathway, such as the IGFBPs, might clarify possible mechanisms by which the IGF pathway influences the process of aging and longevity.

7.0 CONCLUSION

Several biological pathways, including the IGF pathway, appear to play a role in aging and the development of aging-related diseases, but the mechanisms by which they act are not fully understood in humans or animal models [5, 8, 9]. IGF-1, the main ligand in IGF pathway, has been the focus of researchers due to its role in growth, development and metabolism by regulating the actions of growth hormone [10]. Multiple investigators have reported an association between serum IGF-1 levels and age-related disease and mortality [13, 14, 16]. However, the relationship between IGF-1 and other risk factors for aging-related diseases, such as BMI, is unclear. In addition, only seven genetic loci have been identified that influence IGF-1 levels, and many more might exist. The identification of additional genetic variants that affect IGF-1 levels should provide insights into the possible mechanisms involved in IGF-1 metabolism. Studies of IGF-1 and other traits in long-lived families may provide additional insights into healthy aging.

In this study, I investigated three main questions: (1) does the relationship between IGF-1 and measures of body size, such as BMI, differ by age, (2) what is the genetic architecture of serum IGF-1 levels, and (3) what is the relationship between serum IGF-1 levels and genetic variants that affect it, and the risk of mortality?

7.1 The study result summary

7.1.1 Does the relationship between serum IGF-1 levels and BMI vary in an age- and sex-specific manner*?*

As described in the background (**Section 1.2.3**), previous studies of the relationship between IGF-1 levels and BMI have reported contradictory results. I hypothesized that the apparent contradictory results may have been due to differing ages and sexes of the participants in the different studies. To assess the age- and sex specific effects of the relationship between IGF-1 and BMI, the 4,241 study participants of LLFS (age 24 y–110 y) were divided into age quartiles and linear regression of IGF-1 as the dependent variable and BMI as an independent variable was performed in each age quartile. This showed that younger participants (24 y–58 y) had a negative relationship between IGF-1 and BMI, while older participants (87 y–110 y) had a positive relationship **(Figure 2a)**. There was no statistically significant relationship in the intermediate age-groups. In addition, I observed no sex-specific difference in the relationship between IGF-1 and BMI in any age quartile **(Figure 2b)**. A similar relationship pattern between IGF-1 and BMI was observed in another sample of non-Hispanic White participants (NHANES III) **(Figure 2a, b)** and in non-Hispanic Black participants and Mexican American participants **(Figure C3 Appendix C).** This relationship was also observed with another measure of obesity: waist circumference (**Table C2 Appendix C**).

7.1.2 What is the genetic architecture of serum IGF-1 levels in LLFS?

To explore the genetic architecture of serum IGF-1 levels, the phenotype and genotype data of LLFS and FHS participants were used. I estimated the heritability of serum IGF-1 levels in LLFS, which is 40% of serum IGF-1 variation. Genetic correlation between IGF-1 and BMI was explored, in LLFS, as an extension of the phenotypic correlation that was observed in **Chapter 3.** No statistically significant genetic correlation between IGF-1 and BMI was observed in either all study participants or the offspring generation participants. However, statistically significant genetic correlation was observed in the proband generation participants (**Table 6**).

Genome-wide linkage analysis revealed a novel QTL linked to IGF-1 on chromosome 11 with a LOD score of 3.48. There was no evidence of heterogeneity between the study families, and HLOD analysis did not successfully refine the linkage region to highlight any hypothetical causal genes.

To identify specific genetic variants associated with serum IGF-1 levels, candidate gene studies and a GWAS were performed. Candidate gene analysis identified single SNP in *IGF1* associated with IGF-1 in LLFS; however, the association was not statistically significant in FHS. In the GWAS one SNP was genome-wide significant, located near *TCL6* (rs72696993, on 14q21, $p = 4.16 \times 10^{-8}$) (**Figure 9**). This SNP has previously been reported to be associated with QT interval and has also been observed to be associated with *TCL6* gene expression in whole blood. Analysis of all SNPs with $p = 5 \times 10^{-6}$ was performed with data from FHS, and four SNPs in high LD with each other in *AC011294.1* on chromosome 7p12.3. were statistically significantly associated with serum IGF-1 levels. This region has been previously reported to be associated with IGFBP-3 levels as well, and the *IGFBP3* gene is located ~800 kb upstream of it.

7.1.3 What is the relationship between serum IGF-1 levels, its genetic variants, and risk of mortality?

Higher baseline IGF-1 levels were associated with lower mortality risk in regression analyses adjusted for kinship and in analyses adjusted for kinship, sex, and BMI (**Figure 17**). Specifically, participants with lower IGF-1 levels (in the 1st tertile) had a 13% greater risk of mortality, while participants with higher IGF-1 levels (in the 3rd tertile) had a 15% lower risk of mortality, both compared to participants in the 2nd tertile. Upon adjustment for age at baseline, the predictive effect of IGF-1 was attenuated (**Figure 18**). The analysis was extended to assess the association of the IGF-1–associated genetic variants with the risk of mortality. There were no statistically significant associations between any variant genotypes and the risk of mortality in LLFS in any examined regression model.

7.2 Study strengths and limitations

The wide age range of the participants in this study project allowed for the age-group stratification as appropriate in assessing age-related differences in the relationship between serum IGF-1 levels and BMI. In addition, the study sample size was large and well-powered for this analysis compared to other studies. This study was able to replicate some of the analyses (**chapter 3, and 5**) using data from NHANES III and FHS, demonstrating generalizability and serving to validate the study results. Another strength is the use of a family study design with multiple generations, which robustly increases the power for the genetic analysis of variance components. Segregation of allelic sharing through IBD in large pedigree will therefore be detected more easily where it exists.

The main study limitation of these analyses is that all analyses are cross-sectional, which is not useful to assess IGF-1 trajectory across the lifespan or to infer temporality of associations. Furthermore, potential confounders that affect IGF-1 variation and function were not measured, including physical activity, diet, and the IGFBPs.

7.3 Future directions

The underlying effect of genetic variation in the IGF pathway and serum IGF-1 levels in humans on the process of longevity is still not well understood. Studying the association of serum IGF-1 levels with age-related mediators and examining the impact on morbidity and mortality might give more insight to the underlying biological processes of aging and the effect that IGF-1 might play in it. Some of the analysis that would be useful to understand the genetic architecture and genes affecting this complex trait:

- Fine mapping the QTL region on chromosome 11 to narrow the linkage peak and identify the underlying genes is a priority. Adding the genetic data of the third generation (currently under collection) and performing linkage analysis with them might have higher power and identify families with large segregating effects.
- Alternative models for estimating the genetic correlation between IGF-1 and BMI might provide more insight, mainly because the age effect on the phenotypic relationship between IGF-1 and BMI confounds their relationship. LD Score regression is an alternative method for estimation of the genetic correlation using GWAS summary statistics [108]. Another possibility would be an assessment of the effect of BMI–associated SNPs on serum IGF-1 variation in a sample stratified by age groups.

- To follow-up of the GWAS, it would be interesting to explore the biological effect of the variants in *AC011294.1* (and near *IGFBP3*) on 7p12.3 with functional analyses in animal models or cell lines (fibroblast, whole blood cell, or adipocytes). Knockout of this region for examination might help elucidate its effect on IGF-1 levels and the illuminate any subsequent consequences on body fat, development, and mortality.

7.4 Public health significance

The unprecedented increase in longevity worldwide is associated with public health challenges such as increasing demand of health services. In this study, we measured the association between serum IGF-1 (main ligand of IGF pathway, which is involved in the longevity process) and BMI as a chronic disease moderator by age. In addition, I assessed the relationship between serum IGF-1 and associated genetic variations on the risk of mortality. Such information might help categorize elderly individuals into risk groups based on their BMI, age, and IGF-1 information. Since serum IGF-1 is used clinically as a biomarker in the diagnosis of GH-associated illness, understanding the association of IGF-1 with aging may also reveal its usefulness as a biomarker for age-related disease and mortality.

Appendix A Abbreviation

ASL Argininosuccinate Lyase
ASXL2 ASXL Transcriptional Regulator 2
BMI Body mass index
BU Boston
cM centimorgan
CIDR Center for Inherited Disease Research
DNA Deoxyribonucleic acid
DK Denmark
eQTL expression quantitative trait loci
ELISA Enzyme-linked Immunosorbent Assay
FOXO Forkhead Box O
FHS Framingham Heart Study
FFA free fatty acid
FLoSS Family Longevity Selection Score
fc field centers of recruiting
GCKR Glucokinase Regulator
GHSR Growth Hormone Stimulating Receptor gene
GH Growth Hormone
GWAS Genome-wide Association Study
GTEx Genotype-Tissue Expression
HOMA-IR Homeostatic Model Assessment Insulin Resistance
HRC Haplotype Reference Consortium
HR Hazard ratio
HLOD Heterogeneity LOD score
IGF1 Insulin Like Growth Factor1
IGF-1R Insulin-like Growth Factor-1 Receptor
IGF-1 Insulin Like Growth Factor-1
IGF-2 Insulin Like Growth Factor-2
IGFBP3 Insulin Like Growth Factor Binding Protein
IGFBP1–6 Insulin-like Growth Factor Binding Protein 1–6
IGFALS Insulin Like Growth Factor Binding Protein Acid Labile Subunit
IGFALS Insulin-Like growth factor acid-labile subunit
IR Insulin Receptor
IBD identity-by-descent
LLFS Long Life Family Study
LOD logarithm of the odds

LD Linkage disequilibrium
MAF Minor Allele Frequency
M6P/IGF2R the mannose-6-phosphate/IGF-2 receptor
NHANES III third National Health and Nutritional Examination Survey
NUBP Nucleotide Binding Protein
NY New York
PIT Pittsburgh
PTH Parathyroid hormone
PGE2 Prostaglandin E_2
QTL quantitative trait loci
QC Quality control
Q11st quartile
Q2 2nd quartile
Q3 3rd quartile
Q4 4th quartile
s.d. standard deviation
s.e. standard error
SST Somatostatin
SSTR5 Somatostatin Receptor 5
SORCS2 Sortilin Related VPS10 Domain Containing Receptor 2
SNP Single Nucleotide polymorphism
SOLAR Sequential Oligogenic Linkage Analysis Routines
TNS3 Tensin 3
TCL6 T Cell Leukemia/Lymphoma 6
TSH Thyroid hormone
WC Waist Circumference
WT Weight

Appendix B Chapter 1 Supplement

Table B 1. Literature review table of the relationship between serum IGF-1 levels and BMI.

Author (Date)	**Topic/focus question**	**Study population/ Ethnicity/ Mean Age/**	**Study method and IGF-1 assay type.**	**Main finding**	**Comments comparing to LLFS samples**
Alderete et al. (2012)	To examine interrelationships (IGF-1), (IGFBPs) and adiposity.	126 participants 49 AA, 77 Latino. obese adolescents. Mean age 15.3y.	Randomized control trials. immunoradiometric assay kits	IGF-1 and IGFBP-1 were inversely correlated with BMI, total fat mass, visceral adipose tissue (VAT), and hepatic fat fraction (HFF) ($r = -0.20$ to -0.33, $P < 0.05$).	Small sample size. Not Europeans. Obese. Adolescents. Inverse correlation.
Lam et al. (2010)	Hypothesized that lower IGF-1 and higher IGFBP-3, would be related to greater risk of metabolic risk.	Framingham Heart Study. Recruited from 2002-2004, total n = 3977, mean age 40±9	Cross-sectional study. IGF-1 was measured by standard immunoassay.	IGF-1 concentrations were negatively associated with age, diabetes, total cholesterol, BMI, alcohol consumption, and renal function.	Younger than the overall LLFS population. Mean age close to the mean age of the offspring generation. Inverse correlation.
Fauple et al. (2010)	Aimed to examine the associations between IGF-1, (IGFBP-3), and the IGF-1/IGFBP-3 molar ratio with anthropometric measures by race/ethnicity and gender.	NHANES III. n = 5803, mean Age 40. 3,168 women and 2,635 men. 44% non-Hispanic White, 28.2% non-Hispanic Black, and 27.7% Mexican American	Population-based study 1988–1994 IGF-I measured using ELISA.	BMI was inversely associated with IGF-1 levels in all. WHR and WC were inversely associated with IGF-1 levels in all groups except non-Hispanic Black men and Mexican-American females. Height was positively associated with IGF-1 levels only in Mexican- Americans	Larger sample size. Multi-ethnic group. Inverse association. positive association with Height in one group. Age close to the offspring generation
Schernhammer et al. (2007)	Aimed to evaluate associations between IGF and body size throughout life	Nurses' Health Study II (NHS II). 592 healthy premenopausal women. median age 43.5 years (1996-1999).	Cross- sectional study. IGF-1 assayed by ELISA.	Inverse trend between early somatotypes and IGF-I levels, WHR and WC and IGFBP-1 levels. Positive association between birth weight and IGFBP-1 levels. Weight at birth was weakly positively associated with adult IGF-I and IGFBP-1 levels. Inverse associations between body shape at various stages in life and IGF levels	Small sample size. Only women. Only premenopausal. Age different from the mean age of LLFS overall generation. Inverse association.
Gram et al. (2006)	Aimed to examine the relationship between BMI and WHR with serum	Case–control study European 2139 women. 1992-1998.	Cross-sectional study. IGF1 measured by ELISA	Mean IGF-I were inversely related to age. Mean serum IGF-I showed nonlinear inverse association with BMI. IGFBP-3 was linearly positively related to WC and WHR.	Only women Wide age rang Inverse, non-linear European

Author (Date)	Topic/focus question	Study population/ Ethnicity/ Mean Age/	Study method and IGF-1 assay type.	Main finding	Comments comparing to LLFS samples
	levels of IGF-I and IGFBP-3	mean age 55y(32-77y).		IGF-I/IGFBP-3 ratio had a non-linear relation with BMI and a linear inverse relationship with WHR.	
Laughlin A et al. (2004)	the prospective association of serum IGF-I and IGFBP-1 with all cause, non-IHD CVD, and IHD mortality among community-dwelling older men and postmenopausal women.	Rancho Bernardo Study cohort. 633 men and 552 women Caucasian. aged 51–98yr. 1988 –1992 followed through 2001	Prospective cohort study. IGF-I was determined by RIA	Age negatively correlated to IGF-I levels Serum IGF-I levels were positively related to BMI, WHR, and WC. IGFBP-1 were inversely related to adiposity The risk of IHD death increased 38% for each sd (40 ng/ml) decrease in IGF-I	Small sample size Caucasians. Age rang close to the Proband LLFS generation. IGF-1positivly correlated with BMI, and WC.
Fowke et al. (2010)	investigate the role of obesity upon serum IGF-1, IGFBP-3, and IGF-2 levels. white and African American women	study population included 1637 participants (816 white and 821 AA women. Age 40-79y	Southern Community Cohort Study. ELISA kits.	AA premenopausal and postmenopausal women had higher IGF-1 lower IGF-2 levels, compared to white women. Inverse associations between IGF-1 and BMI at age 21 in white women	Inverse correlation Multi-ethnicity Only women
Jernström et al. (2001)	Aimed to evaluate interindividual variability in circulating IGF-I and IGFBP-3 levels in relation to specific genetic and nongenetic factors	women between 17 and 35 years of age. white women		mean plasma IGF-I level and the mean plasma IGFBP-3 level declined with age. Women who used OCs had significantly lower IGF-I levels, higher IGFBP-3 levels. no significant association between birth weight, height, current weight, or BMI and the plasma level of IGF-I.	Young Only women No association between IGF-1 and BMI, WT, and height.

Appendix C Chapter 3 Supplement

Table C 1. The key characteristic of the age group in LLFS and NHANES III.

	Age group			
	20 y–58 y	59 y–66 y	67 y–86 y	87 y–110 y
Participants, *n*				
LLFS	1061	1060	1060	1060
NHANES III	1483	284	726	61
Female, *n* (%)				
LLFS	640 (60%)	553 (52%)	563 (53%)	561 (53%)
NHANES III	840 (56%)	144 (50%)	381 (53%)	35 (57%)
IGF-1 (ng/mL), mean (s.d.)				
LLFS	150.9 (54.1)	139.0 (47.9)	124.1 (48.6)	99.3 (46.6)
NHANES III	283.8 (104.0)	209.9 (82.6)	200.3 (72.5)	187.6 (97.6)
BMI (kg/m^2), mean (s.d.)				
LLFS	27.1 (5.2)	27.9 (4.8)	27.7 (4.7)	25.7 (4.0)
NHANES III	26.3 (5.7)	27.4 (4.9)	26.6 (4.6)	24.9 (3.9)

Table C 2. Linear regression of between log(IGF-1) on WC in both the whole LLFS and NHANES III samples and per age quartile.

	LLFS Age Quartiles			NHANES III Age Groups		
	β (s.e.)	*p*	*n*	β (s.e.)	*p*	*n*
Total sample	0.0004 (0.0 0)	0.4	4170	−0.007 (0.00)	< 0.001	2492
1st age quartile/group (20 y–58 y)	−0.003 (0.00)	0.002	1046	−0.007 (0.00)	< 0.001	1452
2nd age quartile/group (59 y–66 y)	−0.0006 (0.00)	0.5	1048	0.001 (0.00)	0.5	280
3rd age quartile/group (67 y–88 y)	−0.0002 (0.00)	0.7	1040	0.0007 (0.00)	0.5	696
4th age quartile/group (89 y–110 y)	0.004 (0.00)	0.001	1036	0.004 (0.00)	0.5	55

Note, the same age-quartile thresholds for LLFS were applied onto NHANES III.
s.e. = Standard Error, β = regression coefficients of log(BMI) on WC, adjusted for diabetes, hypertension, and sex in both studies, as well as field center and the kinship in LLFS.

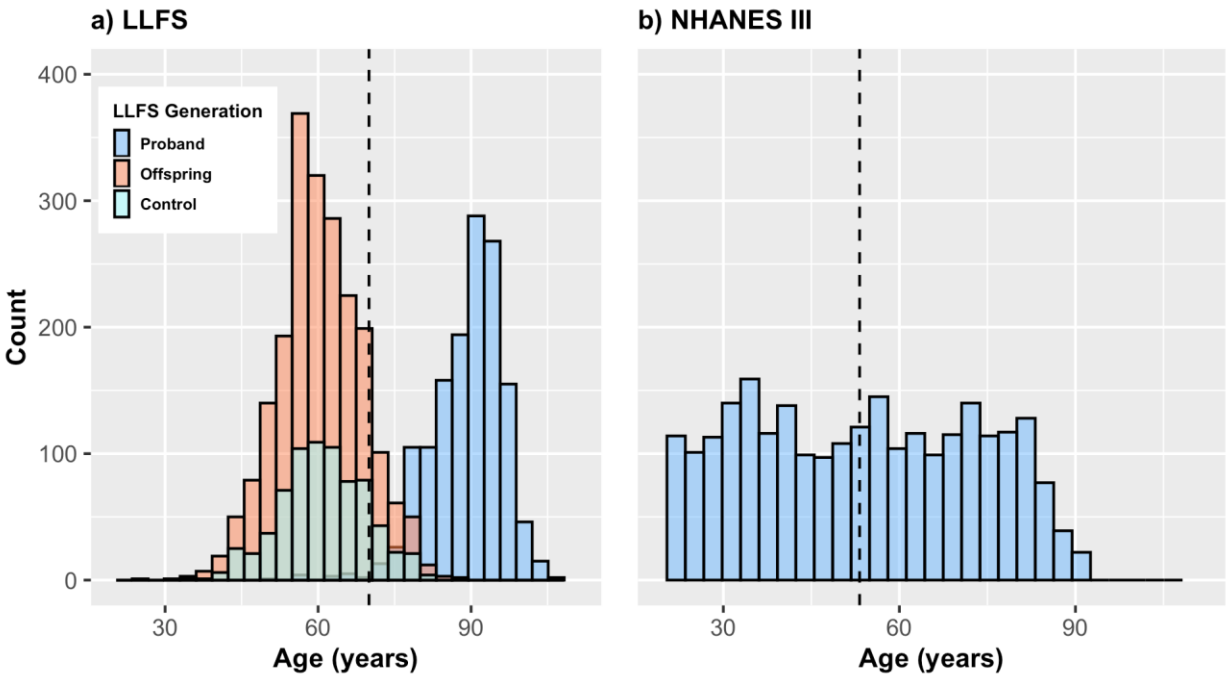


Figure C 1. (a) Age distribution in LLFS (proband generation, offsprings generation and offspring generation's spouses, overlaying and not stacked. (b) Age distribution of participants in NHANES III.

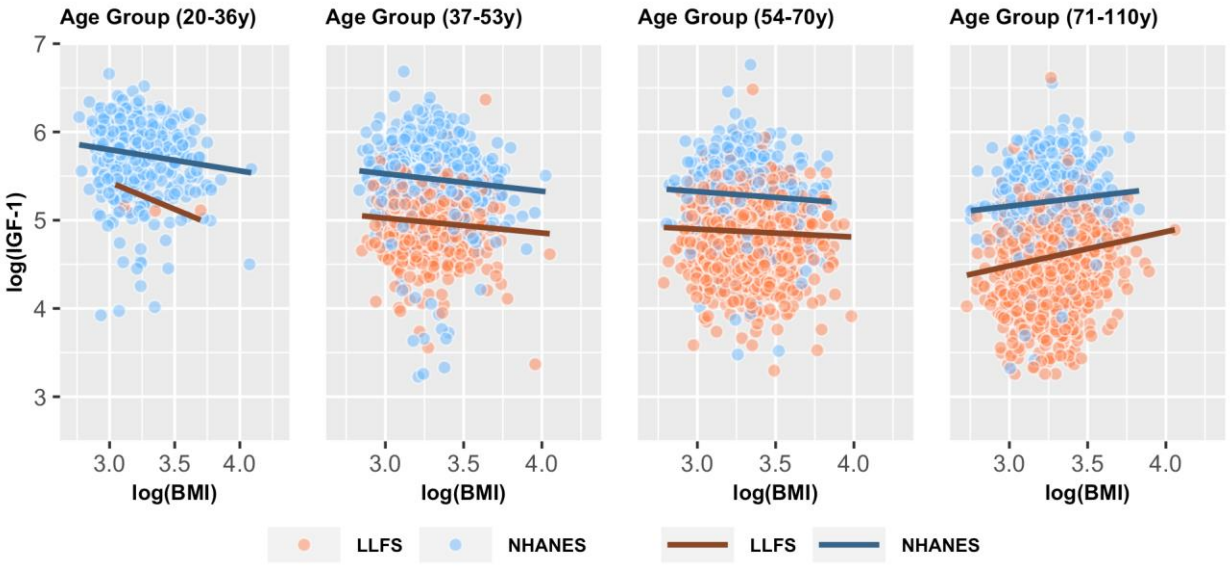


Figure C 2. Scatter plot of log(IGF-1) by log(BMI) stratified by age in LLFS and NHANES III (applying the NHANES III age quartile threshold to LLFS.

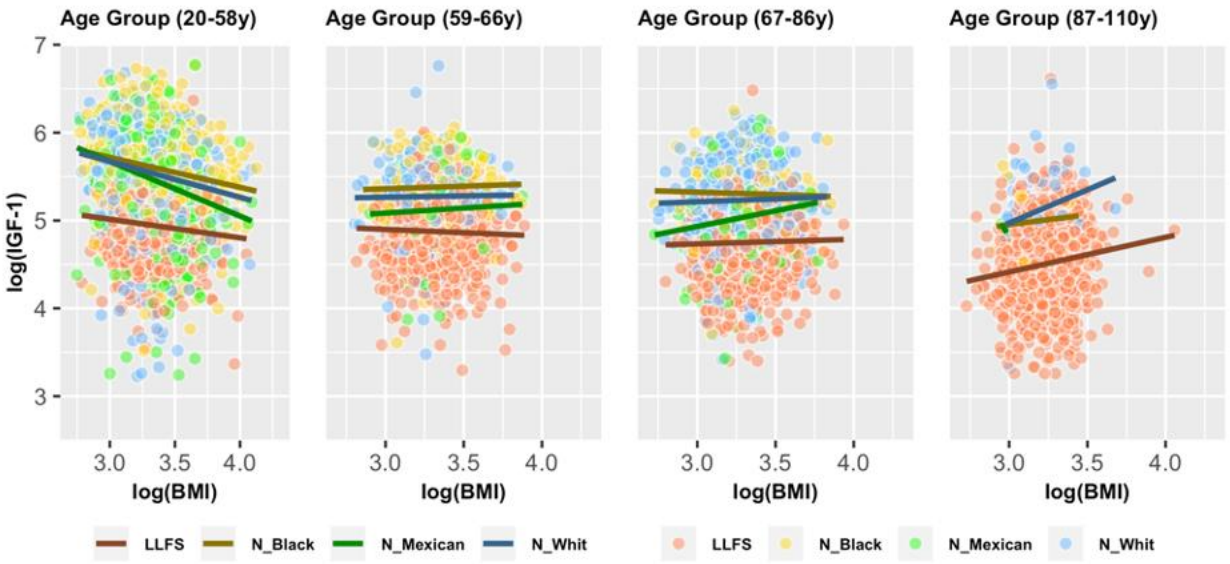


Figure C 3. Scatter plot of log(IGF-1) by BMI stratified by age groups in White participants in LLFS and in non-Hispanic White, non-Hispanic Black, and American Mexican participants in NHANES III.

Appendix D Chapter 6 Supplement

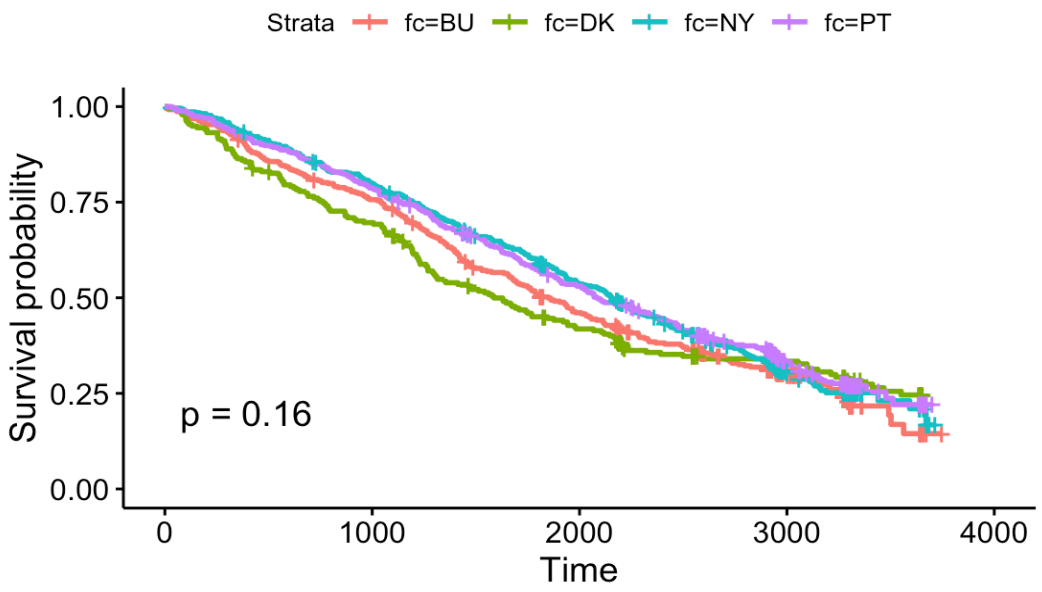
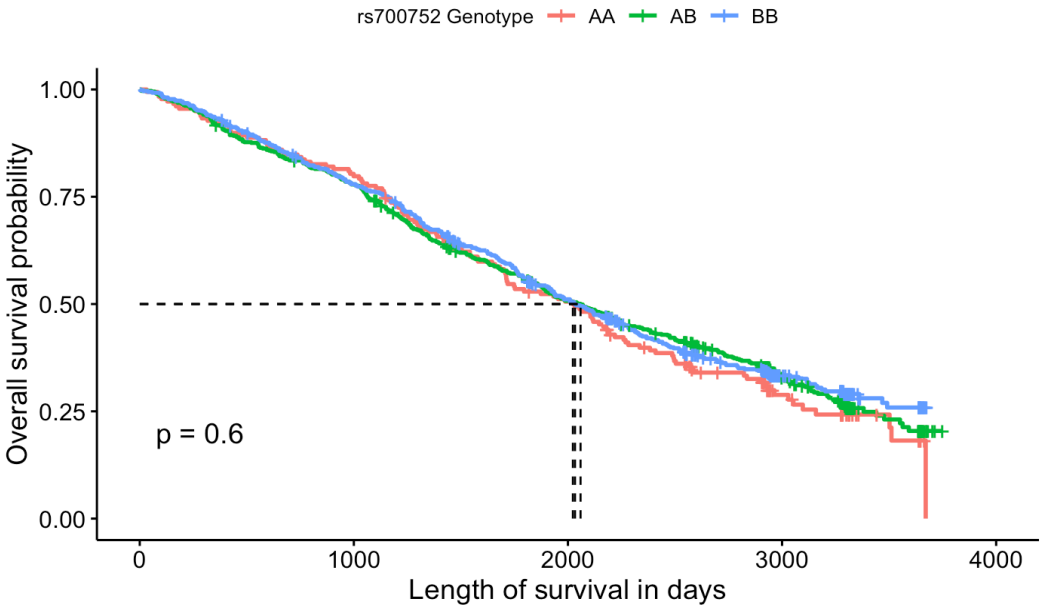


Figure D 1. Survival probability curve for each field center. fc = field center, BU = Boston, DK = Denmark, NY = New York, PT = Pittsburgh.



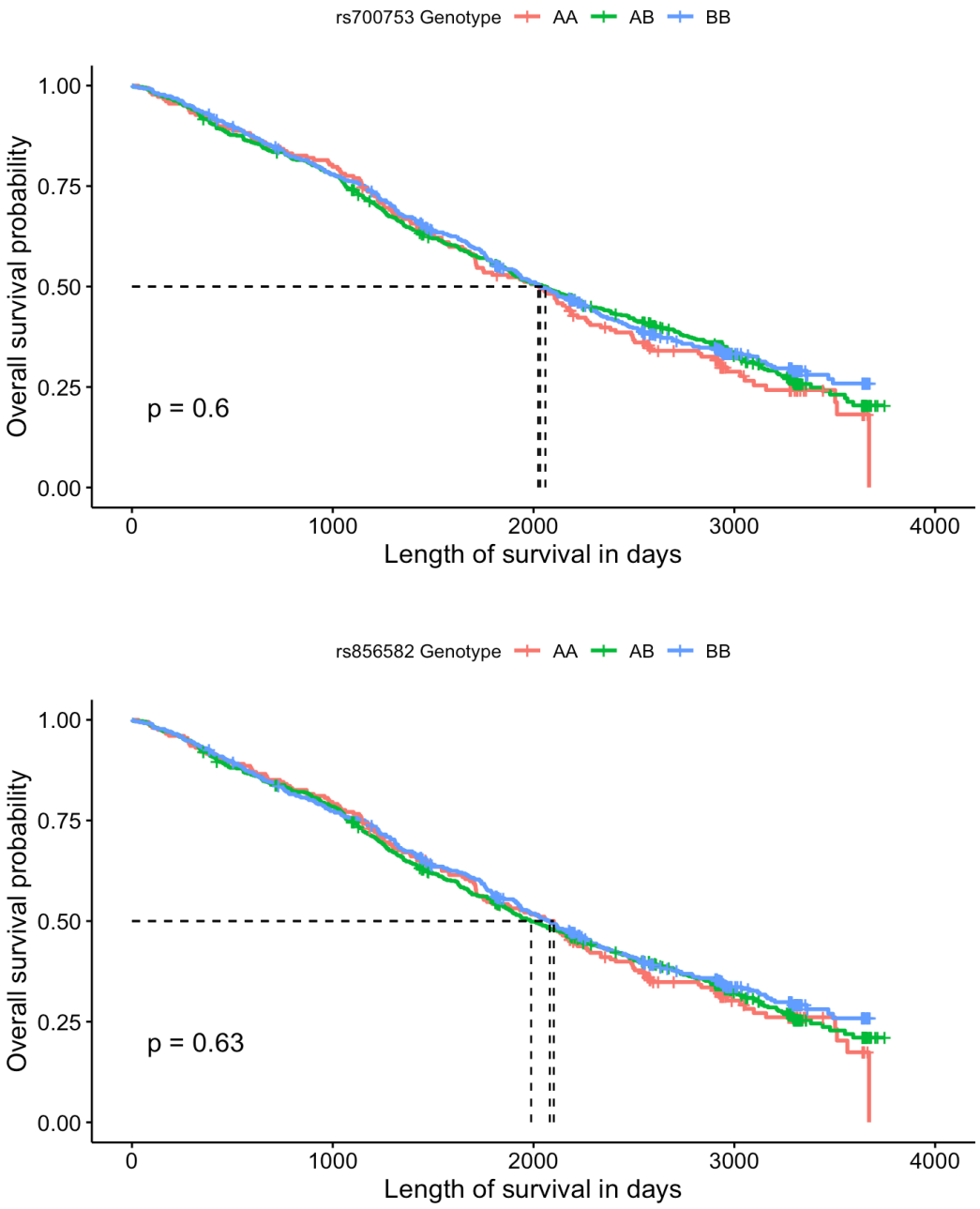
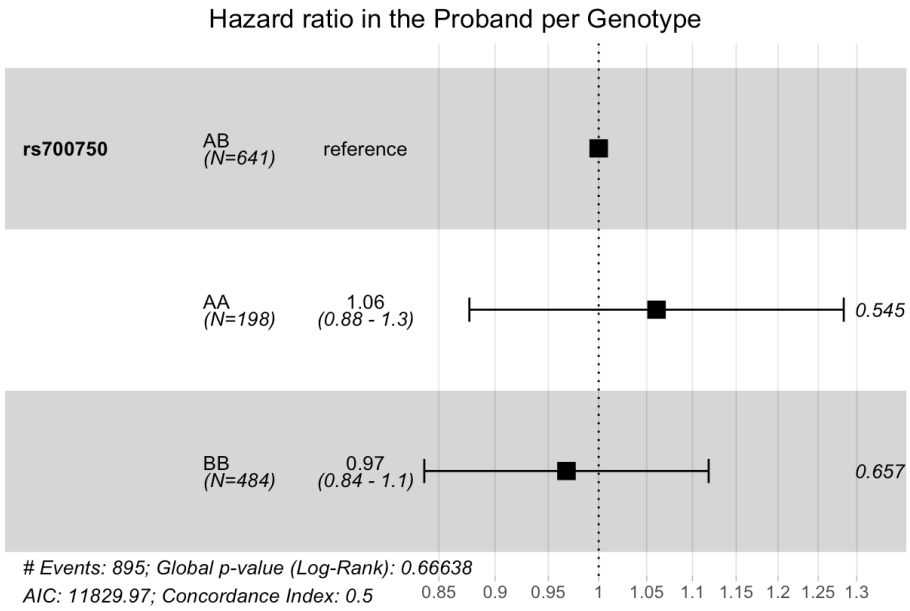
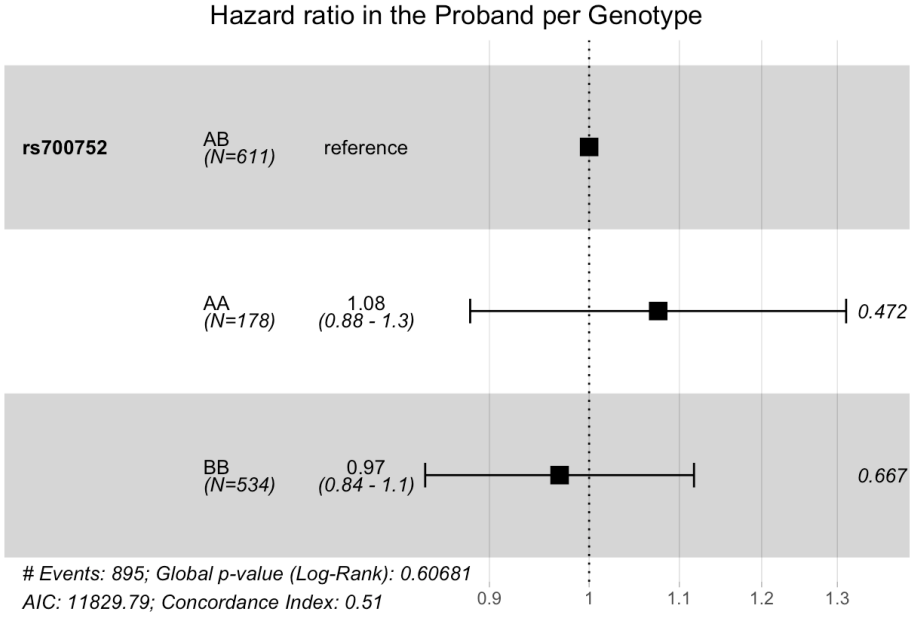
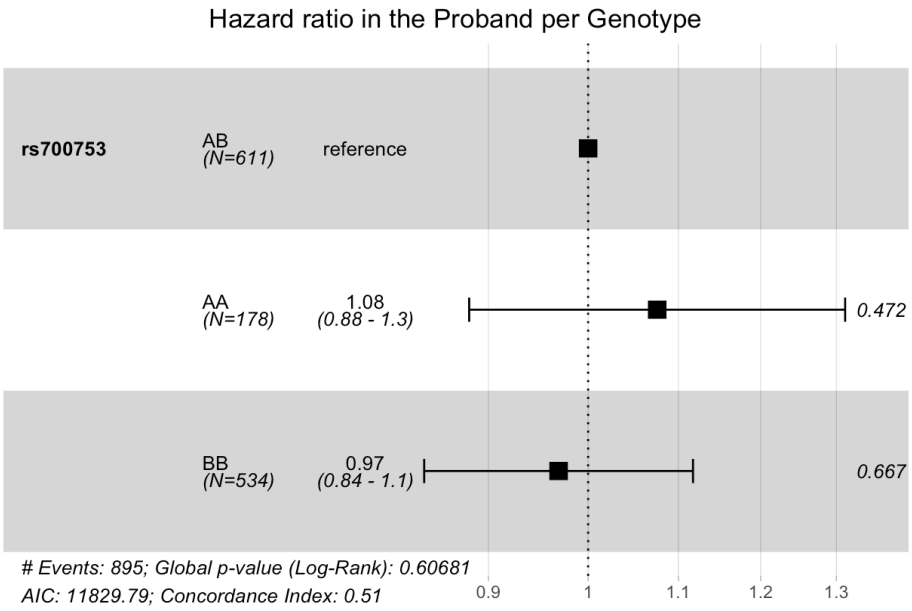


Figure D 2. Survival probability curves for each SNP's Genotype.


Hazard ratio in the Proband per Genotype
rs700750
AB
(N=641)
reference
AA
(N=198)
1.06
(0.88 - 1.3)
0.545
BB
(N=484)
0.97
(0.84 - 1.1)
0.657
Events: 895; Global p-value (Log-Rank): 0.66638
AIC: 11829.97; Concordance Index: 0.5
0.85
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3


Hazard ratio in the Proband per Genotype
rs700752
AB
(N=611)
reference
AA
(N=178)
1.08
(0.88 - 1.3)
0.472
BB
(N=534)
0.97
(0.84 - 1.1)
0.667
Events: 895; Global p-value (Log-Rank): 0.60681
AIC: 11829.79; Concordance Index: 0.51
0.9
1
1.1
1.2
1.3



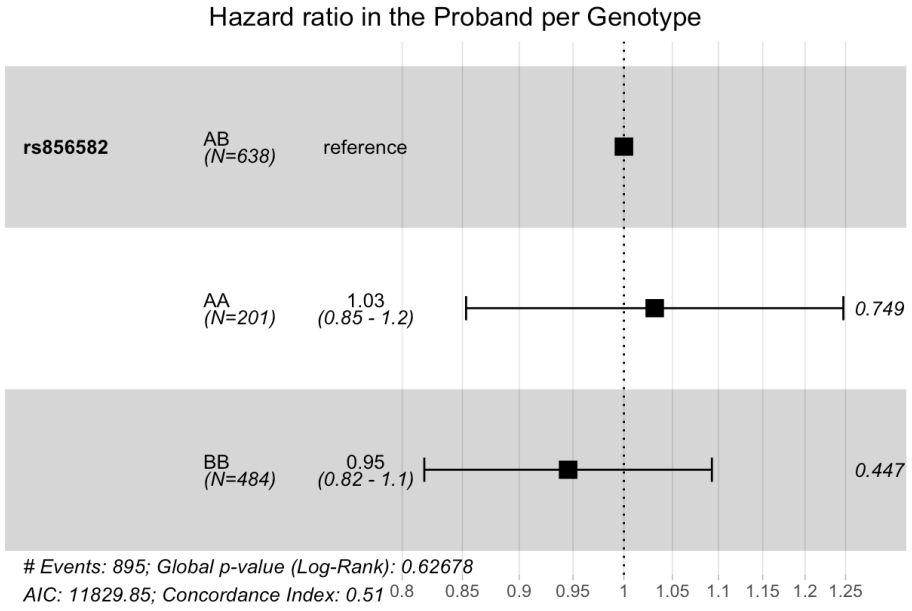


Figure D 3. Hazard risk of mortality per SNP's genotypes, not adjusted for age**.**

Bibliography

1. Smith S, Schoenecker E, Bachrach CA, et al (2015) www.prb.org POPULATION BULLETIN 70. www.prb.org. Accessed 27 Sep 2018
2. Nikolich-Žugich J, Goldman DP, Cohen PR, et al (2016) Preparing for an aging world: Engaging biogerontologists, geriatricians, and the society. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 71:435–444
3. United Nations (2015) World population ageing (highlights)
4. Butler RN, Austad SN, Barzilai N, et al (2003) Longevity genes: from primitive organisms to humans. J Gerontol A Biol Sci Med Sci 58(7):581–584
5. Newman AB, Walter S, Lunetta KL, et al (2010) A Meta-analysis of Four Genome-Wide Association Studies of Survival to Age 90 Years or Older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. Journals Gerontol Ser A 65A(5):478–487. https://doi.org/10.1093/gerona/glq028
6. Lunetta KL, D'Agostino RB, Karasik D, et al (2007) Genetic correlates of longevity and selected age-related phenotypes: A genome-wide association study in the framingham study. BMC Med Genet. https://doi.org/10.1186/1471-2350-8-S1-S13
7. Ziv E, Hu D (2011) Genetic variation in insulin/IGF-1 signaling pathways and longevity. Ageing Res. Rev. 10:201–204
8. Sonntag WE, Csiszar A, De Cabo R, Ferrucci L, Ungvari Z (2012) Diverse roles of growth hormone and insulin-like growth factor-1 in mammalian aging: Progress and controversies. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 67 A:587–598
9. Barbieri M, Bonafè M, Franceschi C, Paolisso G (2003) Insulin/IGF-I-signaling pathway:

An evolutionarily conserved mechanism of longevity from yeast to humans. Am J Physiol - Endocrinol Metab 285(5 48-5). https://doi.org/10.1152/ajpendo.00296.2003

10. Le Roith D (1997) Seminars in medicine of the Beth Israel Deaconess Medical Center. Insulin-like growth factors. N Engl J Med 336:336–640. https://doi.org/10.1159/000109323
11. Bonafè M, Barbieri M, Marchegiani F, et al (2003) Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: Cues for an evolutionarily conserved mechanism of life span control. J Clin Endocrinol Metab 88(7):3299–3304. https://doi.org/10.1210/jc.2002-021810
12. Suh Y, Atzmon G, Cho M-O, et al (2008) Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proc Natl Acad Sci 105(9):3438–3442. https://doi.org/10.1073/pnas.0705467105
13. Rajpathak SN, He M, Sun Q, et al (2012) Insulin-like growth factor axis and risk of type 2 diabetes in women. Diabetes 61(9):2248–2254. https://doi.org/10.2337/db11-1488
14. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M (2004) Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: Systematic review and meta-regression analysis. Lancet 363:1346–1353
15. Schutte AE, Conti E, Mels CM, et al (2016) Attenuated IGF-1 predicts all-cause and cardiovascular mortality in a Black population: A five-year prospective study. Eur J Prev Cardiol 23(16):1690–1699. https://doi.org/10.1177/2047487316661436
16. Andreassen M, Raymond I, Kistorp C, Hildebrandt P, Eaber J, Kristensen LØ (2009) IGF1 as predictor of all cause mortality and cardiovascular disease in an elderly population. Eur J Endocrinol 160(1):25–31. https://doi.org/10.1530/EJE-08-0452
17. Zhang WB, Aleksic S, Gao T, et al (2020) Insulin-like Growth Factor-1 and IGF Binding

Proteins Predict All-Cause Mortality and Morbidity in Older Adults. Cells 9(6):1–18. https://doi.org/10.3390/cells9061368

18. Lam CSP, Chen MH, Lacey SM, et al (2010) Circulating insulin-like growth factor-1 and its binding protein-3: Metabolic and genetic correlates in the community. Arterioscler Thromb Vasc Biol 30(7):1479–1484. https://doi.org/10.1161/ATVBAHA.110.203943

19. Comuzzie AG, Cole SA, Laston SL, et al (2012) Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. PLoS One 7(12):51954. https://doi.org/10.1371/journal.pone.0051954

20. Kaplan RC, Petersen A-K, Chen M-H, et al (2011) A genome-wide association study identifies novel loci associated with circulating IGF-I and IGFBP-3. Hum Mol Genet 20(6):1241–1251. https://doi.org/10.1093/hmg/ddq560

21. Teumer A, Qi Q, Nethander M, et al (2017) Genomewide meta-analysis identifies loci associated with IGF-I and IGFBP-3 levels with impact on age-related traits. Aging Cell 16(4):898–898. https://doi.org/10.1111/acel.12490

22. Prins BP, Kuchenbaecker KB, Bao Y, et al (2017) Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. Sci Rep 7(1). https://doi.org/10.1038/s41598-017-10812-1

23. Le Roith D (2003) The Insulin-like Growth Factor System. Exp. Diab. Res 4:205–212

24. Stewart CE, Rotwein P (1996) Growth, differentiation, and survival: multiple physiological functions for insulin-like growth factors. Physiol Rev 76(4):1005–26. https://doi.org/10.1152/physrev.1996.76.4.1005

25. Bentov I, Werner H (2013) Insulin-like Growth Factor-1. Handb Biol Act Pept 1627–1632. https://doi.org/10.1016/B978-0-12-385095-9.00222-0

26. Hong Y, Pedersen NL, Brismar K, Hall K, De Faire U (1996) Quantitative Genetic Analyses of Insulin-Like Growth Factor I (IGF-I), IGF-Binding Protein-1, and Insulin Levels in Middle-Aged and Elderly Twins. J Clin Endocrinol Metab 81(5):1791–1797. https://doi.org/10.1210/jcem.81.5.8626837

27. Jones, JI and Clemmons D (1995) Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev 16(1):3:34

28. Berryman DE, Glad CAM, List EO, Johannsson G (2013) The GH/IGF-1 axis in obesity: pathophysiology and therapeutic considerations. Nat Rev Endocrinol 9(6):346–356. https://doi.org/10.1038/nrendo.2013.64

29. RABINOWITZ D, KLASSEN GA, ZIERLER KL (1965) EFFECT OF HUMAN GROWTH HORMONE ON MUSCLE AND ADIPOSE TISSUE METABOLISM. J Clin Invest 44(1):51–61. https://doi.org/10.1172/JCI105126

30. Møller N, Jørgensen JOL (2009) Effects of Growth Hormone on Glucose, Lipid, and Protein Metabolism in Human Subjects. Endocr Rev 30(2):152–177. https://doi.org/10.1210/er.2008-0027

31. Burgers AMG, Biermasz NR, Schoones JW, et al (2011) Meta-analysis and dose-response metaregression: Circulating Insulin-Like Growth Factor I (IGF-I) and mortality. J Clin Endocrinol Metab 96(9):2912–2920. https://doi.org/10.1210/jc.2011-1377

32. HUMBEL RE (1990) Insulin-like growth factors I and II. Eur J Biochem 190(3):445–462. https://doi.org/10.1111/j.1432-1033.1990.tb15595.x

33. Van Bunderen CC, Caroline Van Nieuwpoort I, Van Schoor NM, Deeg DJH, Lips P, Drent ML (2010) The Association of Serum Insulin-Like Growth Factor-I with Mortality, Cardiovascular Disease, and Cancer in the Elderly: A Population-Based Study.

https://doi.org/10.1210/jc.2010-0940

34. Roubenoff R, Parise H, Payette HA, et al (2003) Cytokines, insulin-like growth factor 1, sarcopenia, and mortality in very old community-dwelling men and women: The Framingham Heart Study. Am J Med 115(6):429–435. https://doi.org/10.1016/j.amjmed.2003.05.001

35. Friedrich N, Haring R, Nauck M, et al (2009) Mortality and serum insulin-like growth factor (IGF)-I and IGF binding protein 3 concentrations. J Clin Endocrinol Metab 94(5):1732–1739. https://doi.org/10.1210/jc.2008-2138

36. Laughlin GA, Barrett-Connor E, Criqui MH, Kritz-Silverstein D (2004) The Prospective Association of Serum Insulin-Like Growth Factor I (IGF-I) and IGF-Binding Protein-1 Levels with All Cause and Cardiovascular Disease Mortality in Older Adults: The Rancho Bernardo Study. J Clin Endocrinol Metab 89:114–120. https://doi.org/10.1210/jc.2003-030967

37. Nolte AA, Movin M, Lundin H, Salminen H (2015) IGFBP-1 predicts all-cause mortality in elderly women independently of IGF-I. Growth Horm IGF Res 25(6):281–285. https://doi.org/10.1016/j.ghir.2015.09.001

38. Hu D, Pawlikowska L, Kanaya A, et al (2009) Serum insulin-like growth factor-1 binding proteins 1 and 2 and mortality in older adults: The health, aging, and body composition study: Brief reports. J Am Geriatr Soc 57(7):1213–1218. https://doi.org/10.1111/j.1532-5415.2009.02318.x

39. Kaplan RC, Bùžková P, Cappola AR, et al (2012) Decline in circulating insulin-like growth factors and mortality in older adults: Cardiovascular health study all-stars study. J Clin Endocrinol Metab 97(6):1970–1976. https://doi.org/10.1210/jc.2011-2967

40. Gallagher EJ, LeRoith D (2015) Obesity and Diabetes: The Increased Risk of Cancer and Cancer-Related Mortality. Physiol Rev 95(3):727–748. https://doi.org/10.1152/physrev.00030.2014

41. Alderete TL, Byrd-Williams CE, Toledo-Corral CM, Conti D V., Weigensberg MJ, Goran MI (2011) Relationships between IGF-1 and IGFBP-1 and adiposity in obese African-American and latino adolescents. Obesity 19(5):933–938. https://doi.org/10.1038/oby.2010.211

42. Faupel-Badger JM, Berrigan D, Ballard-Barbash R, Potischman N (2009) Anthropometric Correlates of Insulin-Like Growth Factor 1 (IGF-1) and IGF Binding Protein-3 (IGFBP-3) Levels by Race/Ethnicity and Gender. Ann Epidemiol 19(12):841–849. https://doi.org/10.1016/j.annepidem.2009.08.005

43. Schernhammer ES, Tworoger SS, Eliassen AH, et al (2007) Body shape throughout life and correlations with IGFs and GH. Endocr Relat Cancer 14(3):721–732. https://doi.org/10.1677/ERC-06-0080

44. Gram IT, Norat T, Rinaldi S, et al (2006) Body mass index, waist circumference and waist-hip ratio and serum levels of IGF-I and IGFBP-3 in European women. Int J Obes 30(11):1623–1631. https://doi.org/10.1038/sj.ijo.0803324

45. Fowke JH, Matthews CE, Yu H, et al (2010) Racial differences in the association between body mass index and serum IGF1, IGF2, and IGFBP3. Endocr Relat Cancer 17(1):51–60. https://doi.org/10.1677/ERC-09-0023

46. Sugimoto T, Nakaoka D, Nasu M, Kanzawa M, Sugishita T, Chihara K (1998) Age-dependent changes in body composition in postmenopausal Japanese women: relationship to growth hormone secretion as well as serum levels of insulin-like growth factor (IGF)-I

and IGF-binding protein-3. Eur J Endocrinol 138(6):633–639

47. Jernström H, Deal C, Wilkin F, et al (2001) Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev 10(4):377–384

48. Oberbauer AM (2013) The regulation of IGF-1 gene transcription and splicing during development and aging. Front. Endocrinol. (Lausanne). 4:39

49. Harrela M, Koistinen H, Kaprio J, et al (1996) Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. J Clin Invest 98(11):2612–2615. https://doi.org/10.1172/JCI119081

50. Pantsulaia I, Trofimov S, Kobyliansky E, Livshits G (2005) Genetic regulation of the variation of circulating insulin-like growth factors and leptin in human pedigrees. Metabolism 54(7):975–981. https://doi.org/10.1016/j.metabol.2005.02.014

51. Myers A, De-Vrieze FW, Holmans P, et al (2002) Full genome screen for Alzheimer disease: Stage II analysis. Am J Med Genet - Neuropsychiatr Genet 114(2):235–244. https://doi.org/10.1002/ajmg.10183

52. Willcox BJ, Donlon TA, He Q, et al (2008) FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A 105(37):13987–13992. https://doi.org/10.1073/pnas.0801030105

53. Gu F, Schumacher FR, Canzian F, et al (2010) Eighteen insulin-like growth factor pathway genes, circulating levels of IGF-I and its binding protein, and risk of prostate and breast cancer. Cancer Epidemiol Biomarkers Prev 19(11):2877–2887. https://doi.org/10.1158/1055-9965.EPI-10-0507

54. Puche JE, Castilla-Cortázar I (2012) Human conditions of insulin-like growth factor-I (IGF-I) deficiency. J. Transl. Med. 10:224

55. Sebastiani P, Hadley EC, Province M, et al (2009) A family longevity selection score: Ranking sibships by their longevity, size, and availability for study. Am J Epidemiol 170(12):1555–1562. https://doi.org/10.1093/aje/kwp309

56. Newman AB, Glynn NW, Taylor CA, et al (2011) Health and function of participants in the Long Life Family Study: A comparison with other cohorts. Aging (Albany NY) 3(1):63–76. https://doi.org/10.18632/aging.100242

57. Wojczynski MK, Province MA (2019) The Long Life Family Study (LLFS). Encycl Gerontol Popul Aging (July):1–7. https://doi.org/10.1007/978-3-319-69892-2_1101-1

58. Tsao CW, Vasan RS (2015) Cohort Profile: The Framingham Heart Study (FHS): Overview of milestones in cardiovascular epidemiology. Int J Epidemiol 44(6):1800–1813. https://doi.org/10.1093/ije/dyv337

59. Berrigan D, Potischman N, Dodd KW, et al (2009) Race/ethnic variation in serum levels of IGF-I and IGFBP-3 in US adults. Growth Horm IGF Res 19(2):146–155. https://doi.org/10.1016/j.ghir.2008.08.005

60. R Core Team (2019) R: A Language and Environment for Statistical Computing

61. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62(5):1198–1211. https://doi.org/10.1086/301844

62. Lee JH, Cheng R, Honig LS, et al (2013) Genome wide association and linkage analyses identified three loci-4q25, 17q23.2, and 10q11.21-associated with variation in leukocyte telomere length: The long life family study. Front Genet 4(JAN):310. https://doi.org/10.3389/fgene.2013.00310

63. McCarthy S, Das S, Kretzschmar W, et al (2016) A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48(10):1279–1283. https://doi.org/10.1038/ng.3643

64. Gogarten SM, Sofer T, Chen H, et al (2019) Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics 35(24):5346–5348. https://doi.org/10.1093/bioinformatics/btz567

65. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (80-) 369(6509):1318–1330. https://doi.org/10.1126/science.aaz1776

66. Ward LD, Kellis M (2016) HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 44(D1):D877–D881. https://doi.org/10.1093/nar/gkv1340

67. Vitale G, Pellegrino G, Vollery M, Hofland LJ (2019) ROLE of IGF-1 system in the modulation of longevity: Controversies and new insights from a centenarians' perspective. Front Endocrinol (Lausanne) 10(FEB):1–11. https://doi.org/10.3389/fendo.2019.00027

68. Murphy N, Carreras-Torres R, Song M, et al (2020) Circulating Levels of Insulin-like Growth Factor 1 and Insulin-like Growth Factor Binding Protein 3 Associate With Risk of Colorectal Cancer Based on Serologic and Mendelian Randomization Analyses. Gastroenterology 158(5):1300-1312.e20. https://doi.org/10.1053/j.gastro.2019.12.020

69. Carlzon D, Svensson J, Petzold M, et al (2014) Both low and high serum IGF-1 levels associate with increased risk of cardiovascular events in elderly men. J Clin Endocrinol Metab 99(11):E2308–E2316. https://doi.org/10.1210/jc.2014-1575

70. Ketelslegers J, Maiter D, Maes M, Underwood LE, Thissen J (1995) Nutritional Regulation of Insulin-Like Growth Factor-I. metabolism 4(10):50–57

71. CDC (2020) NHANES III (1988-1994). https://wwwn.cdc.gov/nchs/nhanes/nhanes3/default.aspx. Accessed 6 Apr 2019

72. Sebastiani P, Thyagarajan B, Sun F, et al (2016) Age and Sex Distributions of Age-Related Biomarker Values in Healthy Older Adults from the Long Life Family Study. J Am Geriatr Soc 64(11):e189–e194. https://doi.org/10.1111/jgs.14522

73. Therneau TM coxme: Mixed Effects Cox Models

74. Henderson KDL, Goran MI, Kolonel LN, Henderson BE, Le Marchand L (2006) Ethnic disparity in the relationship between obesity and plasma insulin-like growth factors: The multiethnic cohort. Cancer Epidemiol Biomarkers Prev 15(11):2298–2302. https://doi.org/10.1158/1055-9965.EPI-06-0344

75. Zhu H, Xu Y, Gong F, et al (2017) Reference ranges for serum insulin-like growth factor I (IGF-I) in healthy Chinese adults. PLoS One 12(10):1–15. https://doi.org/10.1371/journal.pone.0185561

76. Pokrajac A, Wark G, Ellis AR, Wear J, Wieringa GE, Trainer PJ (2007) Variation in GH and IGF-I assays limits the applicability of international consensus criteria to local practice. Clin Endocrinol (Oxf) 67(1):65–70. https://doi.org/10.1111/j.1365-2265.2007.02836.x

77. Krebs A, Wallaschofski H, Spilcke-Liss E, et al (2008) Five commercially available insulin-like growth factor I (IGF-I) assays in comparison to the former Nichols Advantage IGF-I in a growth hormone treated population. Clin Chem Lab Med 46(12):1776–1783. https://doi.org/10.1515/CCLM.2008.349

78. Kuriyan R (2018) Body composition techniques. Indian J. Med. Res. 148:648–658

79. Oberbauer AM (2013) The regulation of IGF-1 gene transcription and splicing during development and aging. Front Endocrinol (Lausanne) 4(MAR):1–9.

https://doi.org/10.3389/fendo.2013.00039

80. Smith CAB (1961) Homogeneity test for linkage data. Proc Sec Int Congr Hum Genet 1:212–213

81. Franco L, Williams FMK, Trofimov S, et al (2014) Assessment of age-related changes in heritability and IGF-1 gene effect on circulating IGF-1 levels. Age (Omaha) 36(3):1443–1452. https://doi.org/10.1007/s11357-014-9622-7

82. Le Roith D, Scavo L, Butler A (2001) What is the role of circulating IGF-I? Trends Endocrinol Metab 12(2):48–52. https://doi.org/10.1016/s1043-2760(00)00349-0

83. Deelen J, Uh HW, Monajemi R, et al (2013) Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Omaha) 35(1):235–249. https://doi.org/10.1007/s11357-011-9340-3

84. Al-Zahrani A, Sandhu MS, Luben RN, et al (2006) IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet 15(1):1–10. https://doi.org/10.1093/hmg/ddi398

85. Patel A V., Cheng I, Canzian F, et al (2008) IGF-1, IGFBP-1 and IGFBP-3 polymorphisms predict circulating IGF levels but not breast cancer risk: Findings from the breast and prostate cancer cohort consortium (BPC3). PLoS One 3(7). https://doi.org/10.1371/journal.pone.0002578

86. Johansson M, McKay JD, Wiklund F, et al (2007) Implications for prostate cancer of insulin-like growth factor-I (IGF-I) genetic variation and circulating IGF-I levels. J Clin Endocrinol Metab 92(12):4820–4826. https://doi.org/10.1210/jc.2007-0887

87. Schumacher FR, Cheng I, Freedman ML, et al (2010) A comprehensive analysis of common IGF1, IGFBP1 and IGFBP3 genetic variation with prospective IGF-I and IGFBP-3 blood

levels and prostate cancer risk among Caucasians. Hum Mol Genet 19(15):3089–3101. https://doi.org/10.1093/hmg/ddq210

88. Sofer T, Zheng X, Gogarten SM, et al (2019) A fully adjusted two-stage procedure for rank-normalization in genetic association studies. Genet Epidemiol 43(3):263–275. https://doi.org/10.1002/gepi.22188

89. Conomos MP (2017) Population Structure and Relatedness Inference using the GENESIS Package Data Reading in Genotype Data. 39(4):1–11

90. Pruim RJ, Welch RP, Sanna S, et al (2011) LocusZoom: Regional visualization of genome-wide association scan results. In: Bioinformatics. Oxford University Press, pp 2336–2337

91. Buniello A, Macarthur JAL, Cerezo M, et al (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47(D1):D1005–D1012. https://doi.org/10.1093/nar/gky1120

92. Hunt SE, McLaren W, Gil L, et al (2018) Ensembl variation resources. Database (Oxford) 2018:1–12. https://doi.org/10.1093/database/bay119

93. Kent WJ, Sugnet CW, Furey TS, et al (2002) The Human Genome Browser at UCSC. Genome Res 12(6):996–1006. https://doi.org/10.1101/gr.229102

94. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW DM& SP (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. In: Am. J. Hum. Genet. https://zzz.bwh.harvard.edu/plink/contact.shtml. Accessed 3 Mar 2021

95. Marroni F, Pfeufer A, Aulchenko YS, et al (2009) A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: The EUROSPAN project. Circ Cardiovasc Genet 2(4):322–328.

https://doi.org/10.1161/CIRCGENETICS.108.833806

96. Satizabal CL, Adams HHH, Hibar DP, et al (2019) Genetic architecture of subcortical brain structures in 38,851 individuals. Nat Genet 51(11):1624–1636. https://doi.org/10.1038/s41588-019-0511-y

97. Hellwege JN, Velez Edwards DR, Giri A, et al (2019) Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. Nat Commun 10(1):1–11. https://doi.org/10.1038/s41467-019-11704-w

98. Zhou W, Brumpton B, Kabil O, et al (2020) GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. Nat Commun 11(1). https://doi.org/10.1038/s41467-020-17718-z

99. Gill D, Cameron AC, Burgess S, et al (2021) Urate, Blood Pressure, and Cardiovascular Disease. Hypertension 77(2):383–392. https://doi.org/10.1161/hypertensionaha.120.16547

100. Rajaram S, Baylink DJ, Mohan S (1997) Insulin-like growth factor-binding proteins in serum and other biological fluids: regulation and functions. Endocr Rev 18(6):801–31. https://doi.org/10.1210/edrv.18.6.0321

101. van der Spoel E, Rozing MP, Houwing-Duistermaat JJ, et al (2015) Association analysis of insulin-like growth factor-1 axis parameters with survival and functional status in nonagenarians of the Leiden Longevity Study. Aging (Albany NY) 7(11):956–963. https://doi.org/10.18632/aging.100841

102. Bourron O, Le Bouc Y, Berard L, et al (2015) Impact of age-adjusted insulin-like growth factor 1 on major cardiovascular events after acute myocardial infarction: Results from the Fast-MI registry. J Clin Endocrinol Metab 100(5):1879–1886. https://doi.org/10.1210/jc.2014-3968

103. Milman S, Atzmon G, Huffman DM, et al (2014) Low insulin-like growth factor-1 level predicts survival in humans with exceptional longevity. Aging Cell 13(4):769–771. https://doi.org/10.1111/acel.12213

104. Saydah S, Graubard B, Ballard-Barbash R, Berrigan D (2007) Insulin-like growth factors and subsequent risk of mortality in the United States. Am J Epidemiol 166(5):518–526. https://doi.org/10.1093/aje/kwm124

105. Ross R, Miell J, Freeman E, et al (1991) Critically ill patients have high basal growth hormone levels with attenuated oscillatory activity associated with low levels of insulin–like growth factor-I. Clin Endocrinol (Oxf) 35(1):47–54. https://doi.org/10.1111/j.1365-2265.1991.tb03495.x

106. Kishore J, Goel M, Khanna P (2010) Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 1(4):274. https://doi.org/10.4103/0974-7788.76794

107. Sanders JL, Guo W, O'Meara ES, et al (2018) Trajectories of IGF-I Predict Mortality in Older Adults: The Cardiovascular Health Study. Journals Gerontol - Ser A Biol Sci Med Sci 73(7):953–959. https://doi.org/10.1093/gerona/glx143

108. Lee JJ, McGue M, Iacono WG, Chow CC (2018) The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. Genet Epidemiol 42(8):783–795. https://doi.org/10.1002/gepi.22161