Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online)
Vol.8, No.4, 2018


# Association of Measures of Muscle Density and SNPs in Homologs of Muscle Function Genes in C.elegans

Sevinç Akçay*
*Department of Molecular Biology and Genetics, Ahi Evran University, Kirsehir, TURKEY
Email Address: sevinc.akcay@ahievran.edu.tr

Candace Kammerer
University of Pittsburgh, Department of Human Genetics
Email Address: cmk3@pitt.edu

**Abstract**
Sarcopenia, the loss of muscle mass and strength that occurs with aging, has a crucial role in development of frailty. Although sarcopenia and other muscle-related traits are heritable, the underlying genetic variation contributing to the development of sarcopenia is unclear. Our aim was to assess whether there is a relationship between >700 single nucleotide polymorphisms (SNPs) total in 27 genes (14 genes identified from *C. elegans* and 13 genes in the mevalonate and ubiquinone pathways) and six traits related to sarcopenia using data from the HealthABC cohort. No SNPs were statistically significant at the experiment-wide p-value ($p< 0.00001$). SNPs in the locus *MAGOHB* were associated with THIMF, THMUSD, and TOTPF in European Americans, whereas SNPs in *CYP3A5* were associated with these traits in African Americans ($p<0.01$ for all). Furthermore, SNPs in *GOLGA4* ($P<0.01$) and *RALGABP* ($P<0.001$) were associated with TOTPF in African Americans, and SNPs in RALGABP were also associated with THIMF and TOTLEAN in African Americans. The results indicate that variation in some genes related to development of muscle-wasting in an animal model, *C.elegans*, may have pleiotropic effects on traits related to sarcopenia in older men and women.
**Keywords:** Sarcopenia, C.elegans, association, Health ABC

## 1. Introduction

Sarcopenia is characterized by a progressive loss of muscle quantity and quality and muscle mass with aging (Evans, 1995). The prevalence of sarcopenia ranges from 8 to 40% of people over 60 years (Van, 2009). The consequences of sarcopenia are decreased energy expenditure, reduced insulin sensitivity, reduction in muscle mass and strength, and increased physical disability, falls, and mortality.

Sarcopenia has a complex and multifactorial etiology that changes with age. Between ages 20 and 40 years, as a result of a decline in physical activity and a type II muscle fiber size, the sprinting capacity becomes lower. Between ages 40 and 60 years, as a result of the rapid decline in the motor units, decline in physical activity and increase in body fat, there is a decline in aerobic capacity and muscle protein synthesis. At the ages of 60-70 years, the physical activity is deficient, and body fat is very high. Thus the declines in muscle protein synthesis are huge. After the age of 70 years, there is a further reduction in physical activity that causes increased muscle protein breakdown. There is some evidence about the possible mechanisms of sarcopenia including, a decrease in muscle self-repair ability (Collins et al. 2005), inactivity, decreased protein intake, loss of α-motor neurons, hormonal changes, inflammatory effects (increased IL-6), altered caloric intake, and changes in muscle cell physiology. However, the primary mechanism of sarcopenia is still unclear (Volpi, 2004). This lack of knowledge regarding mechanisms also limits the effective treatment of sarcopenia. Thus, there is a need to develop new model systems to understand the mechanism of sarcopenia with subsequent follow-up studies in humans. Results from these studies could lead to the development of new pharmacological agents for the effective treatment of sarcopenia.

The non-parasitic soil nematode C.elegans has become a robust model system in various areas of biology including molecular biology, genetics, developmental biology and neurobiology in recent years (Baumeister, 1997). C. elegans has a variety of experimental advantages including (1) it can be quickly and cheaply cultured in the lab, (2) it has a short lifespan, (3) it has a completely sequenced genome, and (4) the ability to use RNA interference ( Timmons,1998). Under laboratory conditions, worms develop from egg to adult in 3 days and reproduce at 4 or 5 days. The adult C.elegans has 12 to 18 days lifespan. C.elegans has also been used commonly for genetic studies of longevity (Glenn, 2004). Using genetic studies, investigators determined that longevity in C. elegans is regulated by genes involved in the insulin and insulin-like growth factor (IGF) signaling pathway (Kimura, 1997). The Insulin-like pathway includes the daf-2 gene that encodes a protein similar to vertebrate insulin and IGF-I receptors (Kimura et al., 1997). Insulin-like pathways are found to be evolutionarily conserved in C.elegans (Tatar et al., 2001).

As in humans, C.elegans also develops sarcopenia that is characterized by muscle loss that starts in midlife, increases with advancing age, and results in decreased mobility (Herndon et al., 2002, Glenn et al. 2004). The

sarcopenia observed during aging has significant effects on worms. Because of the structural changes in the muscles, there is decreased mobility in the affected worms (Herndon et al.2002). Worms at young ages have increased physical activity that includes sinusoidal swimming on the agar plate. With the aging of worms, there is an apparent physical decline in the physical activity (Herndon et al., 2002). The reductions in the mobility of worms were irreversible (Herndon et al., 2002). The reduced movement is due to disorganized myofibrils and decreased myosin filaments (Herndon, 2002). Sarcopenia does not only affect the mobility of worms, but it also a significant predictor of mortality during aging (Herndon et al., 2002).

Herdon and colleagues reported that aging in C.elegans is associated with sarcopenia (Herndon et al., 2002). Other investigators reported that animals with specific mutations in daf-2 pathway are resistant to the development of sarcopenia (Duhon and Johnson, 1995). Using RNA interference in muscle, Kashyap and colleagues subsequently identified 18 genes which are required for the delay in the onset of sarcopenia in daf-2 mutations (Kashyap et al., 2010). The identified genes included splicing factors, vacuolar sorting proteins, transcription factors, and metabolic enzymes.

Two most commonly studied sarcopenia-related phenotypes are muscle mass and muscle strength. Muscle mass and muscle strength have multiple factors that differ among individuals. Although genetic factors primarily determine the variation among individuals, the specific genes that control these phenotypes are mostly unknown (Tan et al., 2010).

Results of segregation analyses have identified several loci that influence variation in whole body lean mass (Rice et al., 1992). Segregation analyses in Chinese (Liu et al.2004a) and other populations (Borecki et al., 1994, Karasik et al., 2008) have confirmed that there are many genes with moderate effects that determine variations in whole body lean mass.

There is significant evidence that many traits related to muscle strength have high heritability (h2). The heritability of hand grip strength (GS), quadriceps strength (QS) and lower limb muscle strength (LLMS) were estimated as 30-52%(Arden and Spector 1997; Fredericksen et al., 2002), 31-78%( Tiainen et al.,2008) and 42-64% (Fredericksen et al.,2002) based on twin studies.

There is only one linkage study performed on older people. Investigators performed linkage analyses using 397 microsatellites genotypes in 217 twin pairs from the Finnish Twin Study who were 66-75 years of age. They detected a significant linkage between knee extensor isometric strength and a QTL on chromosome 8 (Tiainen et al.2008). Candidate gene studies of possible associations with sarcopenia have included a limited number of genes and have only analyzed a few SNPs within these genes. The results of the candidate gene studies have reported inconsistent results, probably because of population stratification, ethnic variability, sample size differences and gender differences (Tan et al.,2012).

The number of linkage studies for sarcopenia-related phenotypes is deficient. Based on the results of two studies, only 7q21, 7q32, 7p15.3, 15p13, and 20q13 were significant (Tan et al.,2012)

There are a few genes identified that are related to variation in skeletal strength or skeletal mass. Based on the results of two or more studies, the ACE, ACTN3, MSTN, CNTF and VDR genes have been found to be associated with skeletal strength and skeletal mass (Tan et al.,2012). The substantial genetic contribution of MSTN gene to variation in skeletal muscle phenotypes was confirmed by association studies, linkage studies, and expression studies. The IGF1 and IL-6 are also potentially significant genes that were supported by linkage and association studies. Identification of genes that influence sarcopenia-related phenotypes could lead to the development of new pharmacological interventions (Tan et al.,2012).

Our aim of this study was to assess whether SNPs in a total of 27 genes (including 14 that are human homologs of C.elegans muscle function genes and 13 that are mevalonate and ubiquinone pathway genes) influence muscle-related traits in the Health, Age and Body Composition (HealthABC) cohort. We also analyzed data available on the following muscle-related traits include isokinetic leg muscle maximum torque (KCTMAX), thigh intermuscular fat area (THIMF), thigh intermuscular fat area (THMUS), thigh total muscle density (THMUSD), total lean mass (TOTLEAN,) and total percent fat (TOTPF).

## 2. Subjects and methods

### 2.1. Subjects

We used data from Health, Aging and Body Composition Study (Health ABC). Briefly, the HealthABC study population consists of 3,075 well-functioning people from both sexes, who were 70-79 years old at baseline. Data on similar numbers of each sex by ancestry group were available, although slightly more women (N=1584) than men (N=1491) were recruited (Table 1). Health ABC subjects were recruited from Pittsburgh, PA (N=1527) and Memphis, TA (N=1548) in March 1997- July 1998. The University of Pittsburgh and University of Tennessee institutional review board approved the protocol and all participants gave informed consent. Height was measured by a stadiometer and weight was measured by a calibrated balance-beam scale. Subjects' height and weight were measured while they were wearing a hospital gown and no shoes. Body mass index was also one of the body composition measures. Intermuscular fat area of the left thigh and right thigh were measured by

CT.

DXA was used to measure total lean mass ( Goodpaster et al.,2006) and computed tomography was used to measure thigh muscle cross-sectional area (Goodpaster et al., 2006).

### 2.2. Phenotypes examined in this study

The summary of phenotypes is found in Table 2. Genotype data were available on a total of 1737 SNPs across 14 genes, which are human homologs of *C. elegans* muscle function genes and 13 mevalonate and ubiquinone pathway genes.

Phenotypes examined in this study are;

1. KCTMAX: isokinetic leg muscle maximum torque (NM): 80-40deg
2. THIMF: mean of right and left thigh intermuscular fat area (cm sq)
3. THMUS: mean of right and left thigh total muscle area (cm sq)
4. THMUSD: mean of right and left thigh total muscle density (HU)
5. TOTLEAN: total lean (gm)
6. TOTPF: total percent fat

### 2.3. Genotyping and Imputation

Whole genome genotyping of the HealthABC individuals was conducted by the Center for Inherited Disease Research (CIDR) by the Illumina Human 1M-Duo BeadChip system. Three quality control measures were applied by the Wake Forest team:

data from individual samples were eliminated from the data, if sample genotyping the was not successful at all (<97% SNPs genotyped), if the first-degree relatedness was identified by the single-nucleotide polymorphism (SNP) data or if the chromosome sex did not pair with the reported sex. Specific SNPs retained if the SNP had a minor allele frequency (MAF) >1, was called with >97% and had a Hardy-Weinberg equilibrium (HWE) test p value $>10^{-6}$ (Minster, 2011). Genotyping of 1,151,215 autosomal SNPs was performed in 1.663 individuals of European American ancestry and were promoted to imputation (Minster, 2011).

Imputation was conducted by MACH 1.0.16 (Li et al. 2010b) and the HapMap II phased haplotypes (Frazer et al.2007) as the reference. The number of available genotypes were 914,263 SNPs based on the HapMap Centre d'Étude du Polymorphisme Humain (CEPH) reference panel (rel. 22, b36). 2,543,887 genotyped and imputed autosomal SNPs were available for analysis of the "HapMap SNP" set (Minster, 2011).

African Americans and European Americans' quality control protocols were similar to each other; samples were eliminated from the data, if the genotyping was not successful, if the first-degree relatedness was identified by the SNP data or if the chromosome sex did not pair with the reported sex. 1,151,215 SNPs were successfully genotyped in 1,139 African American individuals (Minster 2011). SNPs with an MAF > 1%, a call rate > 97% and a HWE test $p$ value > $10^{-6}$ were used for imputation. 3,021,329 SNPs were available for analysis in the African Americans and were carried forward to imputation. Imputation was conducted by MACH 1.0.16 (Li et al. 2010b) and the HapMap II phased haplotypes (Frazer et al.2007) as the reference (Minster, 2011).

For the single SNP-single trait association analysis whole genome genotyping data were available on 1794 European American and 1281 African Americans. We included all SNPs (both assayed and imputed) that were located within the 27 genes of interest, plus all SNPs that were within ±5kb of the 5' and 3' untranslated regions of the 27 genes. The number of assayed plus imputed SNPs available (for the 27 genes of interest) was 736 for European Americans and 999 for African Americans. For the association analyses in the current study, SNPs with the minor allele frequencies less than 0.05 were excluded. After MAF less than 0.05 removals, there were 627 SNPs in European American and 778 SNPs in African Americans. The number of SNPs within each gene after removal of MAF and the results are given in Supplementary Table 1.

### 2.4. Statistical Methods

All statistical analyses were done using the R 2.11.1 [R Development Core Team 2010]. First, the distribution of the six muscle-related traits was analyzed to assess the outliers and deviations from normality. If necessary, the traits were transformed to reduce non-normality and values that were ±4sd were removed prior to subsequent analyses. Next, linear regression analyses were performed to identify the effects of significant covariates, such as age, and sex. Because unknown admixture among individuals can bias results of population-based association studies, the measures of admixture using principle components of admixture derived by the Wake Forest team were also included. Based on the results of these linear regression analyses, the effects of significant covariates, including admixture were removed. All subsequent analyses were performed using these residual, transformed traits. Correlation among traits was estimated using the residuals. Finally, because some of the genotypes contained imputed data, the genotype data was represented as "probable allele dosage". In other words, the genotype score ranged from 0 – 2. Single SNP-single trait analysis were done using PLINK (Whole genome association analysis toolset) 1.07 (Purcell et al., 2007). To control for multiple testing, results were not considered to be statistically significant unless the p-value was approximately $\leq 5\times10^{-5}$. Multiple testing is one of

the most important concerns in association tests (Yongchao et al.2003). Single SNP-single trait association tests require many tests. Bonferroni method correction is one of the methods to decrease the concerns about multiple testing (Simes 1986). In our study, we used 6 traits and approximately 1400 SNPs and we are performing 8400 tests. Thus, by chance, at *p* value 0.05, we expect 410 false positive results. If we adjusted for multiple testing in our association study, we would need a p value= 0.05/8400 which is approximately $5\times10^{-5}$. With Bonferroni method, we reduced the number of false positives in our study. Effects of possibly significant SNPs on the traits were plotted using box and whiskers plots. Finally, for genes with multiple, potentially significant SNPs, LocusZoom 1.1 was used (Pruim et al., 2010) to assess the location of the significant SNPs and linkage disequilibrium among the SNPs.

## 3. Results

### 3.1. Phenotype characterization

All six phenotypes underwent quality control procedures including transformations to normality and outlier removal. KCTMAX, THMUS, THMUSD and TOTLEAN were transformed by square roots. TOTPF was not transformed. THIMF was transformed by natural logarithms. Observations were called as outliers if they are more than four standard deviations and they were removed from the study. Values that were ±4sd were removed prior to subsequent analyses. In the HealthABC cohort, only one observation was removed from THIMF, THMUSD and KCTMAX. No other observation was removed from THMUS, TOTLEAN, and TOTPF.

We analyzed effect of covariates (age, sex, and measures of admixture) and they were included in the regression models (Supplementary Table 2). To be conservative, including these results as well as information from the literature, we removed the effects of age, sex, and admixture in European and African Americans separately. The characteristics of residuals are given in Supplementary Table 3 and 4. These residuals were used in all subsequent analyses.

### 3.2. Correlation among traits

Phenotypic correlations were computed in order to assess the co-variation among the muscle-related traits. Traits with high correlations may show similar results from the subsequent association analyses. All analyses were done overall and within European Americans and African Americans. The highest phenotypic correlations were between KCTMAX, TOTLEAN, and THMUS (Table 3). The correlations between KCTMAX and TOTLEAN, KCTMAX and THMUS, and TOTLEAN and THMUS were 0.72, 0.75 and 0.90 (all p-value < 00001), respectively. THIMF and THMUSD were significantly, but negatively correlated (r=-0.61, p< 0.0001). Although significant, correlations between the remaining phenotypes were moderate and low. The phenotypic correlations among muscle-related phenotypes for European Americans and African Americans were similar (Table 4 and Table 5), however, overall correlations were higher for European Americans. In African Americans, correlations between THIMF and THMUSD, THMUS-KCTMAX, THMUS-TOTLEAN, and KCTMAX-TOTLEAN were 0.61, 0.70, 0.87, and 0.634, respectively (all p-value <0.0001). Similarly, in European Americans, correlations between THIMF-THMUS, THMUS-KCTMAX, THMUS-TOTLEAN, and KCTMAX-TOTLEAN were 0.60, 0.79, 0.92, and 0.77, respectively (all p-value <0.0001).

### 3.3. Single SNP-single trait association results

We performed the single SNP-single trait association analyses using PLINK in the European American and African American cohort separately. Although a total of 1735 SNPs from 27 genes were initially available, 221 SNPs were excluded in African Americans and 109 SNPs were in European Americans due to low minor allele frequencies (MAF<0.05). A summary of the most significant results are given in Table 5 and 6. As described above, based on the number of tests, a significant result required $p \leq 5\times10^{-5}$. In Figure 2, the $-\log_{10}$p-value for the SNP in each gene is plotted for each trait. The SNPs are plotted by chromosome and base-pair position, and the horizontal line marks p= 0.01. None of the SNPs achieved experiment-wide significance with any trait in either European or African Americans. The most significant association (P <0.00015) was between SNP (rs805554) in *RALGABP* and TOTPF in African Americans (Figure 1 and Table 7).

Although no SNPs achieved experiment wide significance, several genes had multiple SNPs that were associated with a muscle-related trait at p<0.01. To investigate these results further, we tallied the number of SNPs within each gene had p-values $10^{-2}$ for each trait in European Americans and African Americans and also listed the most significant p-value for each gene (Table 7 and 8). Although several genes contained SNPs that were associated (at p< 0.01) with several traits, there was no overlap between European American and African American results. That is, the same SNPs did not affect the same traits in the two groups.

### 3.4. Association Results in European Americans

In European Americans, there was minimal clustering of associations within genes or across traits (Table 6). However, three SNPs (#rs11053877, #rs10845181, and #rs1026462) within MAGOHB locus on chromosome 12 are associated with THIMF (p= $3.062\times10^{-3}$, $8.433\times10^{-3}$, and $2.282\times10^{-3}$, respectively). In addition, 2 of these SNPs (#rs1026442 and #11053877, p= $2.237\times10^{-3}$ and $2.582\times10^{-3}$, respectively) were associated with variation in thigh muscle density (THMUSD) and 1 SNP (rs12811247) was associated with total percent fat (TOTPF). These three traits are moderately correlated with each other (see table 4 and table 5: Phenotypic Correlations). These results may indicate that variation in *MAGOHB* has pleiotropic effects on multiple traits. There was no apparent clustering among the remaining results with p-values <0.01.

### 3.5. Association Results in African Americans

In African Americans, there were several clusters of p-values within genes and across traits (Table 7). As mentioned above, a SNP (rs805554) located in *RALGAPB (KIAA1219)* is significantly associated with total percent fat (TOTPF) in African Americans (p value= $1.517\times10^{-4}$). *RALGAPB* (is on chromosome 20 and codes for the Ral GTPase activating protein, beta subunit (non-catalytic). Five additional SNPs (rs1115600, rs1295992, rs12625459, rs805559, and rs805552) in *RALGAPB* are also associated with TOTPF (p= $8.404\times10^{-3}$, $5.133\times10^{-3}$, $3.33\times10^{-3}$, $5.305\times10^{-3}$, and $4.658\times10^{-3}$, respectively). In addition, eight SNPs (rs2185594, rs805554, rs130577, rs1315305, rs1303571, rs11699660, rs41812025, and rs2206783) in the *RALGAPB* locus are associated with thigh intermuscular fat, of which rs805554 is the same as the associated with total percent fat. Finally, two SNPs (rs9941745 and rs9941761) in the *RALGAPB* locus are associated with total lean mass.

Variation in total percent fat (TOTPF) in African Americans was also associated with SNPs in other genes including: (1) four SNPs (rs1652344, rs1631685, rs1652348, and rs1788781) within the NPC1 locus on chromosome 18 (p=$8.776\times10^{-3}$, $7.433\times10^{-3}$, $6.498\times10^{-3}$, and $4.256\times10^{-3}$, respectively); (2) six SNPs (rs6800842 , rs11129756, rs6550470, rs7639447, rs4635655, and rs17266090) in the GOLGA4 locus on chromosome 3 (p= $5.149\times10^{-3}$, $2.606\times10^{-3}$, $2.438\times10^{-3}$, $2.528\times10^{-3}$, $2.488\times10^{-3}$, and $4.71\times10^{-3}$, respectively); and (3) three SNPS (#rs10242455, # rs4646457, and # rs776746) in the CYP3A5 locus on chromosome 7 (p= $1.325\times10^{-3}$, $5.09\times10^{-3}$, and $5.395\times10^{-3}$, respectively). Of particular interest, genotypes at SNP rs10242455 in CYP3A5 were also associated with both thigh intermuscular fat (p= $4.768\times10^{-3}$) and thigh muscle density (p=$6.564\times10^{-3}$). Thus, this latter SNP may have pleiotropic effects on multiple traits.

SNPs in additional loci were associated (P< 0.01) with other muscle-related traits, but there were no apparent clusters among the remaining results.

### 3.6. Linkage disequilibrium and mean effects of SNPs in genes of interest

As we described previously, single SNP-single trait analysis revealed several loci associated with specific phenotypes within each ancestry cohort, although none of these results achieved experiment-wide significance. However, to further assess the strength of some of the associations, we plotted regional significant association results from single gene-single trait association results using the LocusZoom 1.1 tool (Pruim et al., 2010).

In European Americans, multiple SNPs within *MAGOHB* were associated with thigh intermuscular fat, thigh muscle density, and total percent fat. Figure 2 is a plot of the physical position of all tested SNPs within this chromosomal region by the $-\log_{10}$p values for association with thigh intermuscular fat (THIMF). Recombination hotspots (blue solid line) are at 10.63 Mb and 10.68 Mb. The purple diamond represents the SNP with the lowest *p* value (rs1026442). The colors of the other SNPs represent their correlation with the most significant SNP. Yellow circle shows the $0.6 < r^2 < 0.8$ and green circle shows the $0.4 < r^2 < 0.6$. Blue circle shows the $0.2 < r^2 < 0.4$. As can be seen, the two SNPs with the most significant p-values, rs1026442 (purple diamond) and rs11053877 (red circle) are in high LD with each other. Thus, although these SNPs are 10.37kb apart, the significant associations with THIMF and THMUSD likely represent the same "true" signal. In addition, other SNPs within MAGOHB and KLRA1 show no evidence for association and are not correlated with the two most significant SNPs.

In African Americans, multiple SNPs within *RALGPB* were associated with thigh intermuscular fat, total lean, and total percent fat. Figure 3 is a plot of the physical position of all tested SNPs within this chromosomal region by the $-\log_{10}$p values for association with total percent fat (TOTPF) using SNP rs805554 and rs805552 as the index SNPs. As can be seen, multiple SNPs across the *RALGAPB* locus are associated with the TOTPF. However, when rs805554 is used as the index SNP, several SNPs are not in high LD with this SNP – which could indicate that variation in more than one region of *RALGAPB* may influence the muscle-related traits. Although the "true" causal variant or region is not clear, when rs805552 is used as the index SNP, 4 SNPs across the gene are in high LD with the index SNP (These two SNPs are 4.234 kb apart from each other). These results may indicate that variation in possibly two regions within *RALGAPB* influence muscle-related traits. Figure 4 is a plot of the physical position of all tested SNPs within this chromosomal region by the $-\log_{10}$p values for

association with total lean (TOTLEAN) using SNP rs9941761 as the index SNP. When rs9941761 (which is near the 3' end of *RALGABP*) is used as the index SNP, two SNPs across the gene are in high LD with the index SNP and one of these SNPs is near the middle of *RALGABP*. From all of these results it is not clear whether there is more than one 'true' signal for RALGABP, nor is it clear where the signal might be located.

In African Americans, multiple SNPs within *CYP3A5*were associated with thigh intermuscular fat (THIMF), thigh total muscle density (THMUSD), and total percent fat (TOTPF). Figure 4 is a plot of the physical position of all tested SNPs within this chromosomal region by the $-\log_{10}$p values for association with total percent fat (TOTPF) using SNP rs10242455 as the index SNP. When rs10242455 (which is in the 3' UTR) is used as the index SNP, 2 SNPs across the gene are in high LD with the index SNP, one of which is near the 5' end. These results may indicate that variation in possibly this region within *CYP3A5* influences muscle-related traits, although the location of a 'true' causal variant is unclear.

In African Americans, multiple SNPs within *GOLGA4* were associated with thigh total percent fat (TOTPF). Figure 5 is a plot of the physical position of all tested SNPs within this chromosomal region by the $-\log_{10}$p values for association with total percent fat (TOTPF) using SNP rs6550470 as the index SNP. When rs6550470 is used as the index SNP, 2 SNPs across the gene are in high LD with the index SNP, but others are not. These results may indicate that variation in possible multiple regions within GOLGA4 influences muscle-related traits.

## 4. Discussion

Advancing age results in major changes in body composition, including decreases in skeletal and muscle mass. Sarcopenia, the loss of muscle mass and strength that occurs with aging, plays a crucial role in frailty and results in increased morbidity and mortality among older individuals. The loss of muscle mass and function varies greatly among older individuals, however the underlying causes and mechanisms of this variation are unclear. Because sarcopenia and measures of muscle mass and strength are heritable, many genetic studies, including large genome-wide association consortium studies, have been done to identify genes that influence traits related to sarcopenia, such as total lean mass and obesity (e.g., BMI). These studies have identified several genes that influence lean mass and BMI, but these genes do not account for all of the known heritability of the trait. In addition, few studies have been performed in older individuals or using traits more strongly associated with sarcopenia, such as muscle function and muscle density. In order to better understand the genetic basis of sarcopenia, I performed the single SNP-single trait association analysis using genotype data on 778 SNPs in 27 candidate genes and phenotype data on 6 measures of muscle function and density (KCTMAX, TOTLEAN, TOTPF, THIMF, THMUS and THMUSD) from 3075 men and women aged 70-79 years in the HealthABC cohort. The muscle density measures of THIMF, THMUS, and THMUSD were calculated as the average of the measurements from left and right thigh. Analyses were performed separately in the European American (n=1794) and African American (n=1281) cohorts separately.

Overall, none of the associations in either ancestral group achieved the experiment-wide significance level ($p<10^{-5}$). The most significant result was an association in African Americans between total percent fat and SNPs in the *RALGAPB* locus (minimum p-value = $1.5\times10^{-4}$). I further investigated all tests that achieved a more liberal significance level (p-value < $10^{-2}$). Interestingly, even at this more liberal significance level, there was minimal overlap in results from analyses in the European Americans versus African American cohorts; only three genes showed an association $<10^{-2}$ in both cohorts (*MAGOHB*, *MVK*, and *NPC1*) but the associations were between different SNPs and different traits (Tables 6 and 7). This result indicates that at least some of the variation in these muscle-related traits is influenced by different genes between the two groups. Such an observation is not surprising because of known phenotypic and genetic differences in body composition traits between individuals of predominantly African versus European ancestry (Shaffer et al., 2007).

One of most promising results was the association in African Americans between multiple SNPs within *RALGAPB* locus in African Americans and variation in total percent fat, thigh intermuscular fat, and total lean mass. The genomic location of *RALGAPB* in humans is on 20q11.23. The LD between two SNPs, rs805554 and rs805552, with the most significant association results is high and equals 0.797. These SNPs are also in moderate to high LD ($0.6 < r^2 < 0.9$) with other associated SNPs in *RALGAPB* (Figure 3). Because the LD among all of the SNPs is not uniformly high (e.g, LD>0.80), it is not clear whether one or more variants potentially influence variation in total percent fat, thigh intermuscular fat, or total lean mass. Furthermore, it is not clear where in *RALGAPB* the potentially 'causal' variants are located because associated SNPs are located across the gene. Both SNP rs805554 (T>A) and SNP rs805552 (C>G) are located in the intron 1 region of *RALGAPB*, and are unlikely to be causal. Variants in introns can have causal effects because introns have many short sequences which are significant in effective splicing. Acceptor and donor sites at both ends of the intron are important in proper splicing. It was found that introns affect the expression level of host gene (Chorev and Carmel, 2012). Introns also enhanced the gene expression by intron-mediated enhancement (Jonsson et al.,1992). However, we do not know whether these two SNPs are causal or whether they are in high LD with another, unassayed but

causal variant.

The two most significant SNPs in *RALGAPB* were associated with variation in total percent fat and thigh intermuscular fat. Total percent fat and total intermuscular fat have a phenotypic correlation (r) equal to 0.47 ($r^2$ = 0.22). Total lean mass has a positive, but low correlation with thigh intermuscular fat (r = 0.28), and a negative, moderate correlation with total percent fat (r = 0.42). The mean effects of genotypes at the rs805554 locus on THIMF and TOTPF are consistent with the phenotypic correlation among the traits. Thus, the association results may indicate that polymorphisms within *RALGAPB* have pleiotropic effects on the two muscle-related traits.

*RALGAPB*, the homolog of KIAA1219, is a Ral GTPase activating protein, beta subunit (non-catalytic) and is a non-catalytic subunit of the heterodimeric RalGAP1 and RalGAP2 complexes that act as GTPase activators for the Ras-like small GTPases RALA and RALB. *RALGAPB* gene is the homolog of a putative gene D.2085.5 that is located on chromosome II of *C.elegans*. How variation in RALGAPB may influence muscle-related traits (and sarcopenia) is unclear. However, a RalGAP complex is involved in insulin–stimulated glucose trafficking in muscle and adipose tissue (Chen et al., 2011). As stated earlier, insulin-signalling pathways are associated with longevity, thus the involvement of Ral GAP and other members of this pathway in insulin signaling makes it a reasonable candidate for future study. Furthermore, another member of this pathway, the Ras-like small G protein, RalB, may be involved in the activation of the autophagosome in the presence of nutrient deprivation (Bodemann, et al. 2011). Thus, further study of this pathway may provide novel insights into the development of sarcopenia.

Another potentially interesting result is the association with SNPs within *MAGOHB* locus and thigh intermuscular fat, thigh muscle density, and total percent fat in European Americans. The chromosomal location of *MAGOHB* in humans is 12p13.2. The two SNPs, rs1026442 and rs11053877, with the strongest association results are in high LD with each other (LD= 0.97), thus the association between SNPs and MAGOHB and the muscle-related traits is likely due to a single, 'causal' variant. However, the location of the potentially causal variant is unclear because rs1026442 is located within the first intron whereas rs11053877 is located in the 3' untranslated region. As stated above, variants in introns may affect protein function and expression. Furthermore, 5' UTR and 3' UTR have roles in mRNA stability, translational efficiency and mRNA localization. 3' UTR influences regulation of transcript stability (Laroia et al., 1999). In addition, SNPs on 3' UTR influence microRNA function (Hu and Bruno, 2011). To determine whether these associated SNPs may be causal or whether they are in high LD with an unassayed variant will require additional analyses.

Both SNPs in *MAGOHB* were associated with variation in two correlated traits: THIMF and THMUSD density. Total percent fat and thigh intermuscular fat have a phenotypic correlation (r) equal to 0.47 ($r^2$ = 0.22), whereas thigh intermuscular fat is negatively correlated (r=-0.61) with thigh muscle density. The genotypic means of the three traits for SNP rs1026442, rs11053877, and rs12811247 in *MAGOHB* (Table 6) reflect the phenotypic relationship among the traits. Thus, variation in *MAGOHB* likely has pleiotropic effects on the correlated traits. *MAGOHB* (mago-nashi homolog B) is involved in mRNA splicing and in the nonsense-mediated decay (NMD) pathway (Roignant and Treisman, 2010) and has not been previously identified as potential candidate genes for development of sarcopenia, although control of RNA splicing could definitely have effects on muscle-related traits. Recent reports also indicate that Magoh, as a component of the exon junction complex, regulates division of neural stem cells (Silver et al., 2010). Given the connection between peripheral nerve function and muscle strength (Strotmeyer et al, 2009), members of this pathway also are reasonable candidates for further study of the development of sarcopenia.

## 5. Conclusion

To identify genes that may influence development of sarcopenia, I performed association analyses between several muscle-related traits and SNPs in 27 candidate genes that were homologs of genes that influence the onset of sarcopenia in *C. elegans*. None of these candidate genes have been associated with lean mass or obesity (measured as BMI) in (mostly) middle-aged adults in studies from large consortia (http://www.genome.gov/26525384). However, none of the consortia studies have identified all of the genes that influence lean mass and BMI. Furthermore, lean mass and BMI in middle-aged adults are only crudely correlated with sarcopenia in older individuals and no large studies of sarcopenia or measures of muscle mass and function have been done, especially in older adults.

Although none of the results of this study were statistically significant at an experiment-wide level, several of the results were suggestive. In particular, variants in *MAGOHB* may influence fat infiltration into muscle among individuals with European ancestry, whereas variants in *RALGABP* may influence fat infiltration among individuals with African ancestry. Neither of these genes, or biological pathways, has been previously identified as playing a role in sarcopenia, but further studies may potentially lead to novel insights regarding development of sarcopenia. Replication in other populations is necessary, as well as additional bioinformatic and mechanistic studies, but these are beyond the scope of this thesis. However, these results indicate the potential usefulness of

animal models to generate hypotheses for further study in men and women.

**Acknowledgement**
I would also like to thank all the participants of Health, Aging and Body Composition Study.

**References**

Arden N.K. and Spector T.D. (1997). Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. *Journal of Bone and Mineral Research*, vol. 12, no. 12, pp. 2076–2081.

Baumeister R, Leimer U, Zweckbronner I, Jakubek C, Grunberg J, Haass C. (1997). Human presenilin-1, but not familial Alzheimer's disease (FAD) mutants, facilitate *Caenorhabditis elegans* Notch signalling independently of proteolytic processing. *Genes Funct.*, 1:149–159.

Bodemann BO, Orvedahl A, Cheng T, Ram RR, Ou YH, Formstecher E, Maiti M, Hazelett CC, Wauson EM, Balakireva M, et al. (2011). RalB and the exocyst mediate the cellular starvation response by direct activation of autophagosome assembly. Cell, 144:253–267.

Borecki IB, Rice T, Perusse L, Bouchard C, Rao DC (1994) An exploratory investigation of genetic linkage with body composition and fatness phenotypes: the Quebec family study. Obes Res 2:213–219

Chen X-W., Leto D., Xiong T., Yu G., Cheng A., Decker S., Saltiel A.R. (2011) A Ral GAP complex links PI 3-kinase/Akt signaling to RalA activation in insulin action. *Mol Biol Cell*, 22(1): 141-152.

Chorev M. and Carmel L (2012). The function of introns. Front Genet. , 3: p. 55 doi: 10.3389/fgene.2012.00055.

Collins CA, Olsen I, Zammit PS, Heslop L, Petrie A, Partridge TA, Morgan JE. (2005). Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell*, 122:289–301

Duhon SA, Johnson TE. (1995). Movement as an index of vitality: comparing wild type and the *age-1* mutant of *Caenorhabditis elegans*. J Gerontol A Biol Sci Med Sci ,50:B254–261.

Evans, W. J. (1995). What is sarcopenia? *The Journals of Gerontology Series*, 50A:5-8.

Frazer K. A, Ballinger D., Cox D., Hinds D., Stuve L., Gibbs R., Belmont J., Boudreau A., Hardenbol P., Leal S. et al., (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature ,449, 851–86110.1038/nature06258.

Ge Y., Dudoit S., Speed T. (2003). Resampling-based multiple testing for microarray data analysis. Technical Report *#633*, UC Berkeley Dept. of Statistics

Glenn, C. F., Chow, D. K., Gami, M. S., Iser, W. B., Hanselman, K. B., Wolkow, C. A., David, L., Goldberg, I. G., and Cooke, C. A. (2004). Behavioral deficits during early stages of aging in *Caenorhabditis elegans* result from locomotory deficits possibly linked to muscle frailty. *Gerontol A Biol Sci Med Sci* ,59: 1251–1260.

Fredericksen H., Gaist D., Petersen H. C., Hjelmborg J., McGue M., Vaupel J. W., et al. (2002). Hand grip strength: a phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. *Genet. Epidemiol.*, 23 110–122. 10.1002/gepi.1127

Goodpaster B.H., Park S.W., Harris T.B., Kritchevsky S.B., Nevitt M., Schwartz A.V., Simonsick E.M., FTylavsky F.A., Visser M., and Newman A.B. (2006). The loss of skeletal muscle strength, mass, and quality in older adults: the Health, Aging and Body Composition Study. *J. Gerontol. A Biol. Sci. Med. Sci*., 61:1059–1064.

Herndon LA, Schmeissner PJ, Dudaronek JM, Brown PA, Listner KM, Sakano Y, Paupard MC, Hall DH, Driscoll M. (2002). Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans. *Nature,* 419:808–814.

Jonsson JJ, Foresman MD, Wilson N, McIvor RS (1992). Intron requirement for expression of the human purine nucleoside phosphorylase gene. Nucleic Acids Res, 20:3191–3198.

Karasik D, Dupuis J, Cupples LA, Beck TJ, Mahaney MC, Havill LM, Kiel DP, Demissie S. (2007). Bivariate linkage study of proximal hip geometry and body size indices: the Framingham study. *Calcif Tissue Int.*, 81:162–73.

Kasyap L, Perera S., Fisher AL. (2012). İdentification of novel genes involved in sarcopenia through RNAi screening in Caenorhabditis elegans. *J Gerontol A Biol Sci Med*,67(1):56-65.

Kim J, Wang Z, Heymsfield SB, Baumgartner RN, Gallagher D (2002) Total-body skeletal muscle mass: estimation by a new dual-energy X-ray absorptiometry method. Am J Clin Nutr 76:378–383

Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science*, 277:942–946.

Laroia G, Cuesta R, Brewer G, Schneider RJ (1999). Control of mRNA decay by heat shock-ubiquitin–proteasome pathway. Science, 284: 499–502.

Li Y., Willer C. J., Ding J., Scheet P., Abecasis G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol*., 34, 816–83410.1002/gepi.20488.

Liu PY, Qin YJ, Recker RR, Deng HW. (2004). Evidence for a major gene underlying bone size variation in the

Chinese. *Am J Hum Biol*., 16:68–77.
Minster R. (2011). Pleiotropic relationships among measures of muscle density, bone geometry, lean muscle mass and fat mass. PhD, University of Pittsburgh, Pittsburgh, USA.
Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., and Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26:2336–2337.
Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., Bakker P., Daly M.J.*et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses *Am J Hum Genet* ,81 pp. 559-575.
Rice T, Borecki IB, Bouchard C, Rao DC. (1992). Commingling analysis of regional fat distribution measures: the Québec family study. *Int J Obes Relat Metab Disord.*, Oct;16(10):831–844.
Roignant JY, Treisman JE (2010). Exon junction complex sub-units are required to splice*Drosophila MAP kinase*, a large heterochromatic gene. Cell, 143:238–250.
Shaffer J.R., Kammerer C.M., Reich D., McDonald G., Patterson N., Goodpaster B., Bauer D.C.,. Li J., Newman A.B., Cauley J.A., Harris T.B., Tylavsky F., Ferrell R.E.,. Zmuda J.M., and Health ABC Study (2007). Genetic markers for ancestry are correlated with body composition traits in older African Americans. *Osteoporos. Int*., 18:733–741.
Silver DL, Watkins-Chow D., Schreck K., Pierfelice T., Larson D., Burnetti A., Liaw H., Myung K., Walsh C., Gaiano N. et al. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. Nat. Neurosci,. 13:551–558
Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika,* 73, 751—4.
Strotmeyer ES, de Rekeneire N, Schwartz AV, Resnick HE, Goodpaster BH, Faulkner KA, Shorr RI, Vinik AI, Harris TB, Newman AB (2009). Health ABC Study. Sensory and motor peripheral nerve function and lower-extremity quadriceps strength: the health, aging and body composition study.J *Am Geriatr Soc,* 57(11):2004-10.
Tan, L.-J., Liu, S.-L., Lei, S. -F., Papasian, C. J., and Deng, H.-W. (2012). Molecular genetic studies of gene identification for sarcopenia. *Hum Genet,* 131:1–31, 2012.
Tatar M, Kopelman A, Epstein D, Tu MP, Yin CM, et al. (2001). A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science*, 292:107–110.
Tiainen KM, Perola M, Kovanen VM, Sipila S, Tuononen KA,
Rikalainen K, Kauppinen MA, Widen EI, Kaprio J, Rantanen T, Kujala UM (2008) Genetics of maximal walking speed and skeletal muscle characteristics in older women. Twin Res Hum Genet 11:321–334
Tiainen KM, Perola M, Kovanen VM, Sipilä S, Tuononen KA, Rikalainen K, Kauppinen MA, Widen EI, Kaprio J, Rantanen T, Kujala UM (2008). Genetics of maximal walking speed and skeletal muscle characteristics in older women. *Twin Res Hum Genet*,11:321–334
Timmons L, Fire A. (1998). Specific interference by ingested dsRNA. *Nature*, 395:854.
Van G.A. (2009). Epidemiology and consequences of sarcopenia.. *The Journal of Nutrition, Health and Aging* , 13: 708-712.
Volpi, E., Nazemi, R., and Fujita, S. Muscle tissue changes with aging. *Curr Opin Clin Nutr Metab Care*, 7:405–410, 2004.

Table 1. HealthABC Characteristics

| | | GENDER | | |
|---|---|---|---|---|
| | | **Male** | **Female** | **Total** |
| **RACE** | **White** | 939 | 855 | 1794 |
| | **Black** | 552 | 729 | 1281 |
| | **Total** | 1491 | 1584 | 3075 |

Table 2. Health ABC summary of phenotypes across both sexes and ancestral groups

| **Phenotypes** | **Min** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max** | **N** |
|---|---|---|---|---|---|---|---|
| **KCTMAX (nm)** | 12.30 | 78.91 | 100.60 | 106.30 | 132.40 | 461.70 | 2646 |
| **THIMF (cm sq)** | 0.100 | 6.048 | 8.955 | 10.310 | 13.030 | 91.250 | 3011 |
| **THMUS (cm sq)** | 47.77 | 89.93 | 109.00 | 111.40 | 130.50 | 226.20 | 3011 |
| **THMUSD (HU)** | 2.425 | 31.110 | 36.600 | 35.380 | 40.270 | 70.060 | 3011 |
| **TOTLEAN (gm)** | 247.50 | 406.60 | 483.90 | 89.50 | 565.40 | 839.80 | 3075 |
| **TOTPF (%)** | 13.10 | 28.90 | 34.60 | 34.99 | 41.20 | 56.00 | 3075 |
| **Weight (kg)** | 33.50 | 65.50 | 75.20 | 75.81 | 85.00 | 141.00 | |
| **Height (cm)** | 1370 | 1590 | 1658 | 1662 | 1732 | 2007 | |

Table 3. Phenotypic correlations (r) among muscle-related traits (transformed and adjusted for significant covariates)

| | | logTHIMF | sqrtTHMUS | sqrtTHMUSD | sqrtKCTMAX | sqrtTOTLEAN | numericTOTPF |
|---|---|---|---|---|---|---|---|
| logTHIMF | | 1 | .214** | -.614** | .039* | .275** | .465** |
| | Sig. (2-tailed) | | .000 | .000 | .046 | .000 | .000 |
| sqrtTHMUS | | . | 1 | .061** | **.746**** | **.901**** | -.418** |
| | Sig. (2-tailed) | | | .001 | .000 | .000 | .000 |
| sqrtTHMUSD | | | | 1 | .205** | .015 | -.505** |
| | Sig. (2-tailed) | | | | .000 | .422 | .000 |
| sqrtKCTMAX | | | | | 1 | **.716**** | -.464** |
| | Sig. (2-tailed) | | | | | .000 | .000 |
| sqrtTOTLEAN | | | | | | 1 | -.415** |
| | Sig. (2-tailed) | | | | | | .000 |
| numericTOTPF | | | | | | | 1 |
| | Sig. (2-tailed) | | | | | | |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4. Phenotypic correlations (r) among muscle-related traits (transformed and adjusted for significant covariates) in European Americans

| | | logTHIMF | sqrtTHMUS | sqrtTHMUSD | sqrtKCTMAX | sqrtTOTLEAN | numericTOTPF |
|---|---|---|---|---|---|---|---|
| logTHIMF | | 1 | .215** | **-.603**** | .102** | .287** | .421** |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .000 | .000 |
| sqrtTHMUS | | | 1 | .044 | **.791**** | **.922**** | -.484** |
| | Sig. (2-tailed) | | | .064 | .000 | .000 | .000 |
| sqrtTHMUSD | | | | 1 | .155** | .001 | -.488** |
| | Sig. (2-tailed) | | | | .000 | **.**981 | .000 |
| sqrtKCTMAX | | | | | 1 | **.773**** | -.493** |
| | Sig. (2-tailed) | | | | | .000 | .000 |
| sqrtTOTLEAN | | | | | | 1 | -.479** |
| | Sig. (2-tailed) | | | | | | .000 |
| numericTOTPF | | | | | | | 1 |
| | Sig. (2-tailed) | . | | | | | |

**. Correlation is significant at the 0.01 level (2-tailed).

a. RACE = European

Table 5. Phenotypic correlations (r) among muscle- related traits (transformed and adjusted for significant covariates) in African Americans

| | | logTHIMF | sqrtTHMUS | sqrtTHMUSD | sqrtKCTMAX | sqrtTOTLEAN | numericTOTPF |
|---|---|---|---|---|---|---|---|
| logTHIMF | Sig. (2-tailed) | 1 | .138** | **-.610**** | -.048 | .235** | .522** |
| | | | .000 | .000 | .121 | .000 | .000 |
| sqrtTHMUS | Sig. (2-tailed) | | 1 | .139** | **.706**** | **.876**** | -.375** |
| | | | | .000 | .000 | .000 | .000 |
| sqrtTHMUSD | Sig. (2-tailed) | | | 1 | .280** | .059* | -.523** |
| | | | | | .000 | .038 | .000 |
| sqrtKCTMAX | Sig. (2-tailed) | | | | 1 | **.634**** | -.435** |
| | | | | | | .000 | .000 |
| sqrtTOTLEAN | Sig. (2-tailed) | | | | | 1 | -.355** |
| | | | | | | | .000 |
| numericTOTPF | Sig. (2-tailed) | | | | | | 1 |
| | | . | | | | | |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. RACE = African

Table 6. Genes with SNP associations $p < 10^{-2}$ by trait in European Americans

| | TRAITS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **GENE** | **KCTMAX** | **Total SNPs (#sig)** | **THIMF** | **Total SNPs (#sig)** | **THMUS** | **THMUSD** | **Total SNPs (#sig)** | **TOTLEAN** | **TOTPF** | **Total SNPs (#sig)** |
| *MAGOHB* | - | | $2.8\times10^{-3}$ | 46(3) | - | $2.2\times10^{-3}$ | 46(2) | - | $9.7\times10^{-3}$ | 46(1) |
| *MVK* | $4.8\times10^{-3}$ | 32(1) | - | | - | - | | - | - | |
| *NPC1* | - | | $3.1\times10^{-3}$ | 40(2) | - | - | | - | - | |
| *SLC27A4* | - | | - | | - | $3.3\times10^{-3}$ | 9(2) | - | - | |

Table 7. Genes with SNP associations $p < 10^{-2}$ by trait in African Americans

| | TRAITS | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **GENE** | **KCTMAX** | **Total SNPs (#sig)** | **THIMF** | **Total SNPs (#sig)** | **THMUS** | **Total SNPs (#sig)** | **THMUSD** | **Total SNPs (#sig)** | **TOTLEAN** | **Total SNPs (#sig)** | **TOTPF** | **Total SNPs (#sig)** |
| *ADCK3* | $9.9\times10^{-3}$ | | - | | - | | - | | | | - | |
| *COQ2* | - | | - | | $5.4\times10^{-3}$ | 15(1) | - | | | | - | |
| *CYP3A5* | - | | $4.7\times10^{-3}$ | 25(1) | - | | $6.5\times10^{-3}$ | 24(1) | | | $1.3\times10^{-3}$ | 27(3) |
| *FDPS* | $3.5\times10^{-3}$ | 7(2) | - | | - | | - | | | | - | |
| *GOLGA4* | - | | - | | - | | - | | | | $2.4\times10^{-3}$ | 68(6) |
| *JUP* | - | | - | | - | | - | | $8.0\times10^{-3}$ | 23(1) | - | |
| *MAGOHB* | - | | - | | $2.5\times10^{-3}$ | 62(1) | - | | $3.4\times10^{-3}$ | 59(1) | - | |
| *MVK* | - | | - | | $1.8\times10^{-3}$ | 26(1) | - | | | | - | |
| *NPC1* | - | | - | | - | | - | | | | $4.2\times10^{-3}$ | 59(4) |
| *PDSS1* | - | | $8.7\times10^{-3}$ | 63(3) | - | | - | | | | $8.2\times10^{-3}$ | 70(2) |
| *RALGABP* | - | | $3.1\times10^{-3}$ | 81(6) | - | | - | | $4.1\times10^{-3}$ | 91(2) | $1.5\times10^{-4}$ | 84(6) |
| *SAC1ML* | $8.5\times10^{-3}$ | 75(1) | - | | - | | - | | | | - | |

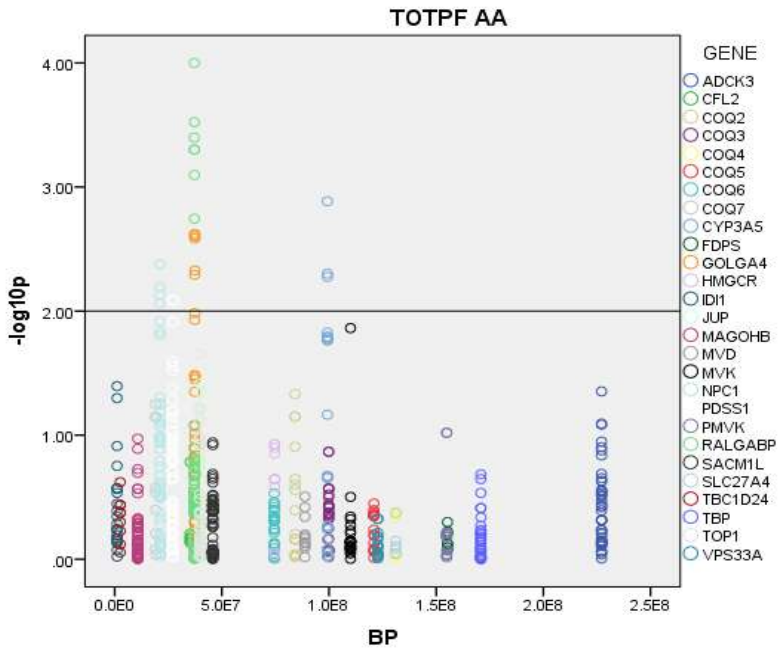



Figure 1. Plot of –($\log_{10}$p-value) and location (in bp) for TOTPF in AA

The - $\log_{10}$p-value for the SNP (rs805554) in RALGABP is plotted for TOTPF.rs805554 SBP is plotted by chromosome and base-pair position.

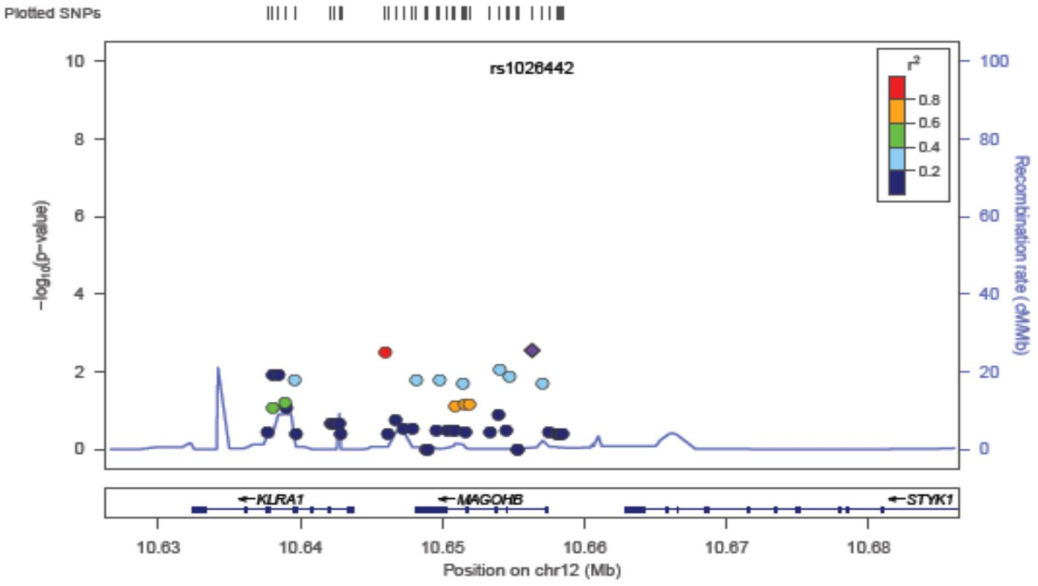



Figure 2. Association between THIMF and SNPs around *MAGOHB* (rs1026442) on 12p13.2 in European Americans

Figure 2 is a plot of the physical position of all tested SNPs within this chromosomal region by the –log10p values for association with thigh intermuscular fat (THIMF). Recombination hotspots (blue solid line) are at 10.63 Mb and 10.68 Mb. The purple diamond represents the SNP with the lowest p value (rs1026442). The colors of the other SNPs represent their correlation with the most significant SNP. Yellow circle shows the 0.6 < r2 < 0.8 and green circle shows the 0.4 < r2 < 0.6. Blue circle shows the 0.2 < r2 < 0.4.

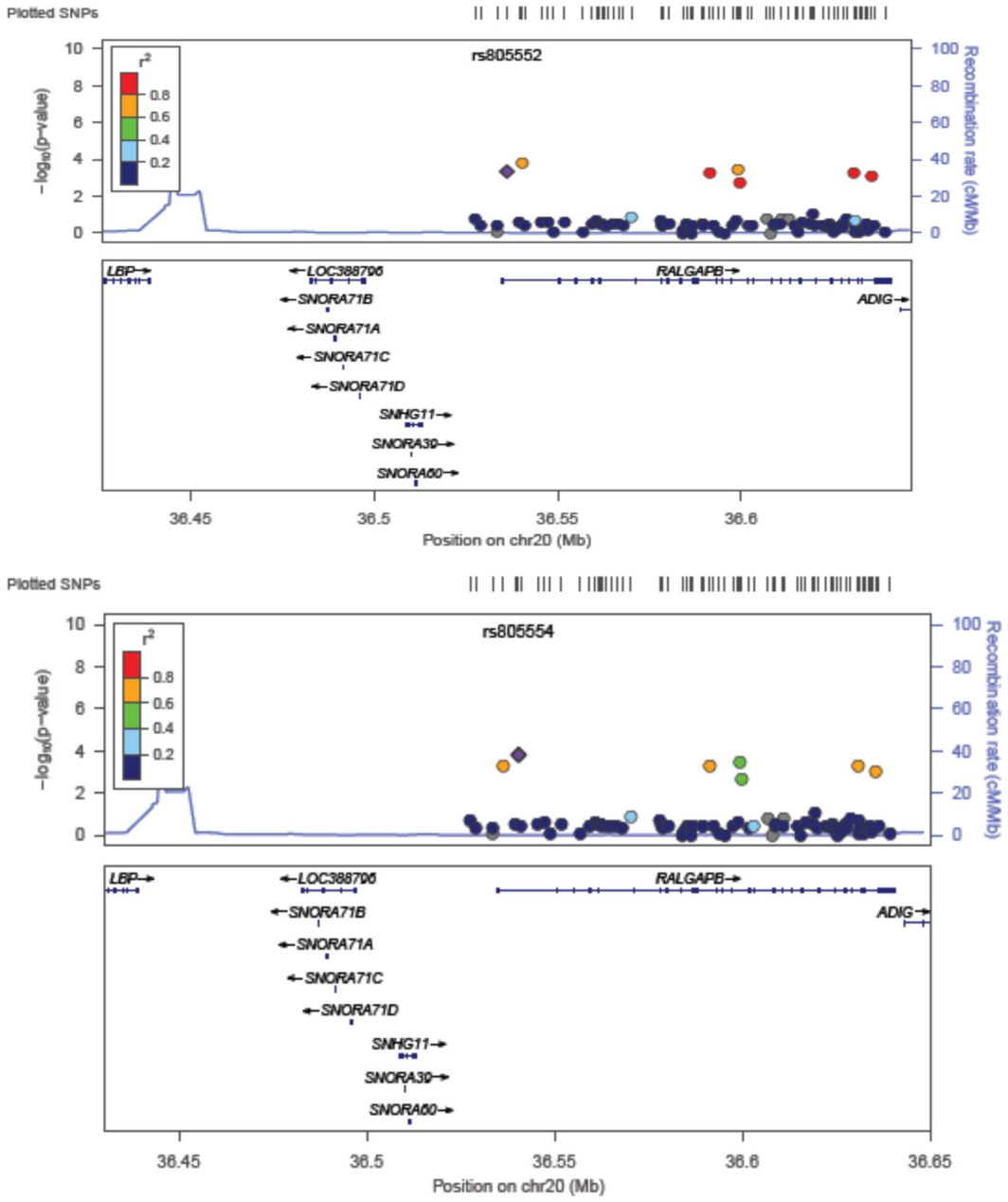



Figure 3. Association between TOTPF and SNPs near RALGBP in African Americans using SNPs rs805554 and rs805552 as the index SNP.

Figure 3 is a plot of the physical position of all tested SNPs within this chromosomal region by the –log10p values for association with TOTPF. Recombination hotspots (blue solid line) are at 36.45 Mb and 36.6 Mb. The purple diamond represents the SNP with the lowest p value (rs805554 and rs805552). The colors of the other SNPs represent their correlation with the most significant SNP. Yellow circle shows the $0.6 < r2 < 0.8$ and green circle shows the $0.4 < r2 < 0.6$. Blue circle shows the $0.2 < r2 < 0.4$.

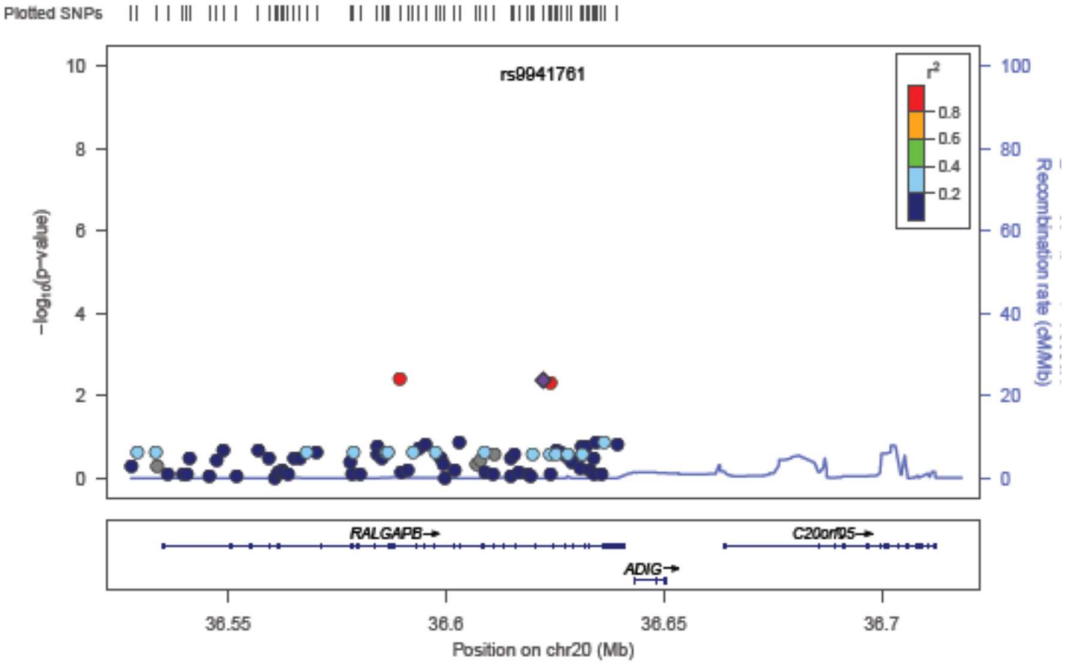



Figure 4. Association between TOTLEAN and SNPs near RALGBP in African Americans using SNP rs9941761 as the index SNP.

Figure 4 is a plot of the physical position of all tested SNPs within this chromosomal region by the –log10p values for association with TOTLEAN. Recombination hotspots (blue solid line) are at 36.55 Mb and 36.7 Mb. The purple diamond represents the SNP with the lowest p value (rs9941761). The colors of the other SNPs represent their correlation with the most significant SNP. Blue circle shows the 0.2 < r2 < 0.4.

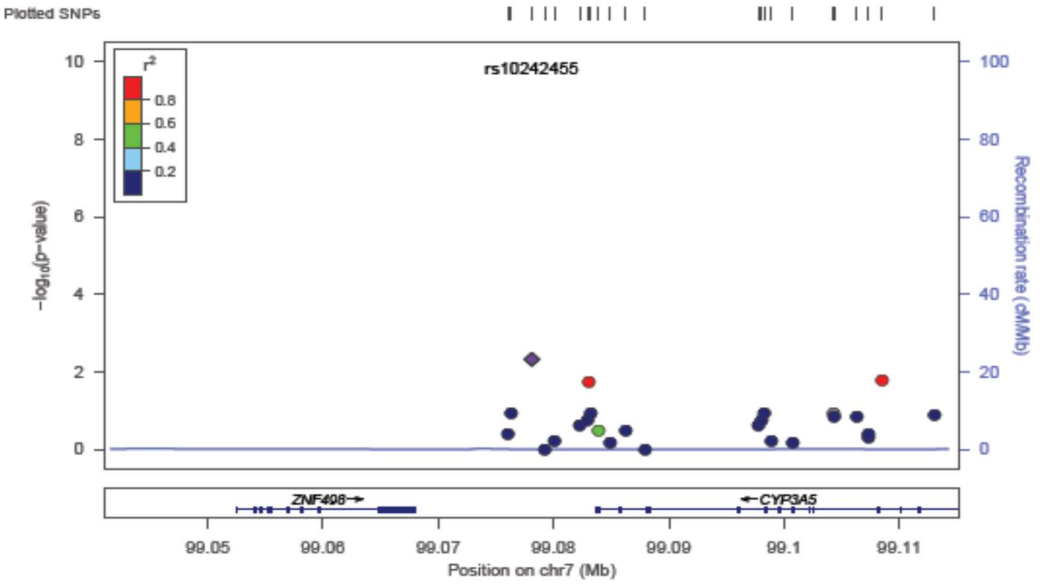



Figure 5. Association between THIMF and SNPs near *CYP3A5* in African Americans using SNP rs10242455 as the index SNP.

Figure 5 is a plot of the physical position of all tested SNPs within this chromosomal region by the –log10p values for association with THIMF. Recombination hotspots (blue solid line) are at 99.05 Mb and 99.11Mb. The purple diamond represents the SNP with the lowest p value (rs10242455). The colors of the other SNPs represent their correlation with the most significant SNP. Blue circle shows the 0.2 < r2 < 0.4.

**Appendix**

Supplementary Table 1. List of Candidate Genes (and number of SNPs for African Americans and European Americans) used in the current study

| Genes | African Americans Total number of SNPs= 778 | European Americans Total number of SNPs =627 |
|---|---|---|
| *ADCK3* | 45 | 41 |
| *CFL2* | 6 | 5 |
| *COQ2* | 15 | 12 |
| *COQ3* | 24 | 23 |
| *COQ4* | 7 | 4 |
| *COQ5* | 17 | 16 |
| *COQ6* | 26 | 24 |
| *COQ7* | 11 | 9 |
| *CYP3A5* | 24 | 8 |
| *FDPS* | 7 | 7 |
| *GOLGA4* | 64 | 35 |
| *HMGCR* | 12 | 17 |
| *IDI1* | 23 | 23 |
| *JUP* | 22 | 18 |
| *MAGOHB* | 62 | 46 |
| *MVD* | 10 | 13 |
| *MVK* | 27 | 32 |
| *NPC1* | 47 | 40 |
| *PDSS1* | 62 | 53 |
| *PMVK* | 14 | 13 |
| *RALGAPB* | 81 | 30 |
| *SACM1L* | 74 | 73 |
| *SLC27A4* | 6 | 5 |
| *TBC1D24* | 9 | 9 |
| *TBP* | 34 | 34 |
| *TOP1* | 42 | 36 |
| *VPS33A* | 13 | 12 |

Supplemantary Table 2. Effects of covariates (beta-values) on the muscle-related traits in European and African Americans

| | KCTMAX sqrt (nm) | TOTLEAN sqrt (gm) | THIMF log (cm sq) | THMUS Log (cm sq) | THMUSD sqrt (HU) | TOTPF (%) |
|---|---|---|---|---|---|---|
| **European Americans** | | | | | | |
| **Sex** | -0.704** | -0.823** | -0.034(NS) | -0.789** | -0.197** | 0.705** |
| **Age** | -0.130** | -0.056(NS) | -0.002(NS) | -0.082** | -0.089** | -0.051(NS) |
| **PC1** | 0.083** | 0.064(NS) | 0.159** | -0.005(NS) | -0.136** | 0.051(NS) |
| **PC2** | -0.037** | -0.001(NS) | 0.010** | -0.005(NS) | -0.003** | -0.017(NS) |
| **African Americans** | | | | | | |
| **Sex** | -0.613** | -0.704** | 0.160** | -0.666** | -0.314** | 0.746** |
| **Age** | -0.113** | -0.100(NS) | -0.039(NS) | -0.134** | -0.111** | -0.062(NS) |
| **PC1** | -0.057 (NS) | -0.102** | -0.043(NS) | -0.101** | -0.040(NS) | 0.093** |
| **PC2** | -0.029(NS) | -0.055** | -0.025(NS) | -0.035** | -0. 019(NS) | 0.035** |

** p< 0.00

Supplementary Table 3. Summary of residuals of transformed traits in European Americans

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max |
|---|---|---|---|---|---|---|
| **SqrtKCTMAX(nm)** | -5.186 | -0.739 | 0.095 | 0.00 | 0.802 | 9.600 |
| **SqrtTOTLEAN(gm)** | -39.69 | -9.47 | -0.58 | 0.00 | 8.74 | 56.58 |
| **LogTHIMF(cm sq)** | -2.67 | -0.33 | 0.03 | 0.00 | 0.36 | 2.37 |
| **LogTHMUS(cm sq)** | -2.59 | -0.53 | 0.01 | 0.00 | 0.53 | 2.85 |
| **SqrtTHMUSD(HU)** | -3.22 | -0.31 | 0.04 | 0.00 | 0.39 | 2.57 |
| **TOTPF (%)** | -1374 | -228.3 | -14.61 | 0.00 | 234.5 | 1477 |