ABSTRACT

Identification of risk factors for sarcopenia and its roles in Koreans through Genome Wide Association Study

Ye Seul Yang, M.D.

(Directed by Professor Kyong Soo Park, M.D., Ph. D.)

Department of Molecular Medicine and Biopharmaceutical Sciences,

Graduate School of Convergence Science and Technology,

and College of Medicine,

Seoul National University

Sarcopenia, the progressive decrease of skeletal muscle mass and function, is related to functional impairment in the elderly. Muscle mass, mainly composed of lean body mass, is known to be a heritable trait. Sarcopenia has been a subject of increasing genetic researches. However, genetic risk factors in Koreans were not thoroughly evaluated. The

aim of this study is to identify genetic variants associated with lean body mass in Koreans.

We performed a genome-wide association analysis of the two cohorts (Stage I), then did a joint analysis to discover common SNPs (Single Nucleotide Polymorphisms) in both cohorts (Stage II). A total of 13,105 participants (5,817 from Ansung-Ansan cohort and 7,288 form Gangnam center cohort) were analyzed. Joint analysis was conducted using METAL. We performed linear regression analysis including sex, age, height and body fat mass as covariates. Arbitrarily defined P value threshold of less than 0.00001 was used for suggestive evidence of association.

In joint analysis, none of the SNP markers passed genome-wide significance threshold of $P < 5 \times 10^{-8}$. There are nine loci (in/near *FTO, GALNT16, ERH, PLCE1, STK39, PRF1* and *ADAMTS14* genes) which were associated with lean body mass with suggestive genome-wide significance.

Among them, five variants including the most significant loci (rs3751812) were located in *FTO*. The most recent study showed that *FTO* expression increased during myoblasts differentiation, while the silence of *FTO* inhibited the differentiation. So, we planned to evaluate the correlation between *FTO* expression and muscle protein degradation, which is another muscle atrophy phenotype in sarcopenia model. After 2 days of FTO siRNA transfection in our study, there were no significant difference of MuRF-1 and

Atrogin-1 expression, the two muscle-specific ubiquitin ligases that are important regulators of ubiquitin-mediated protein degradation in skeletal muscle.

In this first genome-wide association joint analysis study for lean body mass in Korean participants including the representative cohorts of Korea, we identified nine loci (in/near *FTO, GALNT16, ERH, PLCE1, STK39, PRF1* and *ADAMTS14* genes) for lean body mass which are crucial in sarcopenia diagnosis.

The actual function of the variants and the underlying mechanisms of the most suggestive gene, as well as *FTO*' s involvement in skeletal muscle biology still need to be further elucidated by *in vitro* and animal experiments.

.....

Keywords: aging, lean body mass, sarcopenia, genome wide association study, FTO

Student Number: 2015-26008

Contents

List of Tables

List of Figures

Introduction

With aging, there is a progressive loss of muscle mass, and a concurrent increase in fatty infiltration and fibrosis of muscle. A progressive loss of muscle mass occurs approximately from the age of 40. This loss has been estimated at about 8 % per decade until the age of 70, after which the loss increases to 15 % per decade [1].

In 1988, Irwin Rosenberg proposed that sarcopenia is an age-related decrease of skeletal muscle mass and its function [2]. The prevalence of sarcopenia reported varies widely depending on the definition: Appendicular skeletal muscle mass (ASM) divided by height squared (ASM/ht^2) or by weight (ASM/wt), and methods of assessment. It ranges from 8 % to 40 % of people aged over 60 [3, 4]. Sarcopenia is usually accompanied by physical inactivity, decreased mobility, slow gait, and poor physical endurance. It may reach a critical point at which time functional impairment and disability occurs [5] and may increase mortality in the elderly [6].

In fact, the annual direct healthcare costs that are attributable to sarcopenia in the United States were estimated to be in excess of 18 billion dollars in 2000. It is estimated that a 10.5 % reduction of the prevalence of sarcopenia could lead to a reduction of healthcare costs by

1.1 billion dollars per year in the United States [7]. Korea is rapidly becoming an aging society. According to the Korea National Statistical Office, 13.1 % of the Korean population was aged 65 or older (i.e., elderly) in 2015. The figure is expected to rise to 40.1 % in 2060 [8]. Therefore, research on sarcopenia is essential for the development of public health programs for the increasing elderly Korean population.

Muscle mass and muscle strength are two widely-studied sarcopenia-related phenotypes used for genetic studies. Among them, lean body mass primarily consists of skeletal muscle, and is highly heritable phenotype with heritability estimates around 0.57-0.66, as estimated from family and twin studies [9, 10]. Since Genome-Wide Association Study (GWAS) became a promising and effective approach for discovery of genes related to complex disease or quantitative traits, many studies have been performed to reveal the genetic background of body mass index (BMI) and fat mass. However, few studies have been done in order to search for genes associated with lean body mass.

The first GWAS related to lean body mass tested 379,319 eligible SNPs in 1,000 unrelated US whites and found two SNPs, rs16892496 (P = 7.55 x 10^{-8}) and rs7832552 (P = 7.58 x 10^{-8}), within the thyrotropin releasing hormone receptor (*TRHR*) gene [11]. A copy number variation (CNV) located in the *GREM1* gene was reported to be associated with lean body mass in a GWAS

of 1,627 Chinese [12]. Guo et al. identified a locus in/near *GLYAT* genes at 11q12.1 in a bivariate GWAS for bone size phenotypes and appendicular lean mass in 1,627 Chinese and replicated in 2,286 European ancestry individuals [13]. Most recently in a study of Japanese women, the *PRDM16* gene was suggested to be associated with lean mass [14].

However, these previous results did not prove to be consistent and no SNPs have been found to be associated with lean body mass with a genome-wide significance level ($P < 5 \times 10^{-8}$), to date. In addition, genetic risk factors in Korean have not been evaluated yet. Therefore, we performed GWAS joint analysis in over 10,000 Korean participants of two cohorts in order to identify genetic variants associated with lean body mass in Koreans.

Materials and Methods

We focused on lean body mass. In the previous studies, lean body mass measured by dual energy X-ray absorptiometry (DXA) is widely used as an index for muscle mass. On the other hands, the bioelectrical impedance analysis (BIA) estimates the volume of fat and lean body mass and is a noninvasive, portable, quick, and inexpensive method for measuring body composition. Thanks to these advantages, BIA is widely used to measure body composition in many health care centers and hospitals. BIA results have been found to correlate well with MRI predictions [15] and the results measured by DEXA [16].

We performed a genome-wide analysis of the two cohorts (Stage I), then joint-analyzed to discover common SNPs in both cohorts (Stage II). Since lean body mass is correlated with fat mass and height, analyses were adjusted for these potential confounders in addition to sex and age to focus our search for genes contributing to lean body mass independent of those of body height and fat mass.

Study Population

In Stage I Discovery, participants for the GWAS were recruited from two cohorts.

One is the combined cohort of rural Ansung and urban Ansan cohorts (Ansung-Ansan cohort). Initiated in 2001, as part of the Korean Genome Epidemiology Study (KoGES), the initial samples included 5,018 and 5,020 participants from the age of 40 to 69. The sampling base for the cohort is in Gyeonggi Province. This cohort was designed to allow longitudinal prospective study and participants were examined every two years since baseline, with the seventh follow-up study. More than 260 traits were extensively examined through epidemiological surveys, physical examinations, and laboratory tests applied to Ansung-Ansan cohort members. Data from 2007 to 2008 were included for analyses in this study.

The other one is Seoul National University Hospital Gangnam Healthcare Center cohort (Gangnam center cohort). It is a cohort composed of participants who regularly receive a screening program and designed to identify environmental and genetic factors for major chronic diseases in Koreans.

All Cohorts were approved by their institutional ethics review committees (IRB No. AJIRB-CRO-07-012 for Ansung-Ansan cohort and IRB No. H-1103-127-357 for Gangnam center cohort) and all participants provided written informed consent. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Lean Body Mass Measurements

Lean body mass was measured in both cohorts using BIA. It relies on the geometrical relationship between impedance (Z), length (L) and volume (V) of an electrical conductor. Adapted to the human body, V corresponds to the volume of fat-free mass (FFM) and L to the height of the subject. Z is composed of the pure resistance (R) of the conductor, the FFM, and the reactance (Xc), which produced by the capacitance of cellular membranes, tissue interfaces and non-ionic tissues: $Z^2 = R^2 + Xc^2$.

BIA was performed using InBody 3.0 and InBody 720 (Biospace Co., Ltd, Seoul, Korea) in Ansung-Ansan cohort and Gangnam center cohort, respectively. The subject stood on the footplate in are feet and held both the hand electrodes. The screen automatically displays measurements of lean body mass (kg), skeletal muscle mass (kg), body fat mass (kg), and body fat percentage (%).

Anthropometric Measurements

Anthropometric parameters and blood pressure were measured by standard methods. Fasting plasma glucose (FPG), total cholesterol, triglycerides, high-density lipoprotein (HDL), low-density lipoprotein (LDL) and HbA1c were measured in a central laboratory after a 12 hours fast.

Stage I: Genome-wide association analyses in each cohorts

Genotyping and Imputation

We extracted genomic DNA from the participants' peripheral leucocytes. The genome scan was done with an Affymetrix Genome-Wide Human Single Nucleotide Polymorphism (SNP) Array 5.0 for Ansung-Ansan cohort and Affymetrix Axiom Customized Biobank Genotyping Arrays for Gangnam center cohort. Genotypes were called using the Birdseed version 2.0 algorithm. Briefly, only unrelated participants with genotype missingness of less than 5% were included in the analysis.

5,817 participants in Ansung-Ansan cohort and 7,288 participants in Gangnam center cohort were available for genetic association analysis after quality-control filtering. Markers with significant deviations from the Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-6}$), a genotype call rate of less than 0.95, and minor allele frequency of less than 0.01 were excluded. Genotype imputation was conducted using IMPUTE version 2.0 with 1000G as reference.

Stage II: Joint analysis of the two cohorts

Joint analysis of Ansugn-Ansan cohort and Gangnam center cohort was performbed with METAL (http://www.sph.umich.

edu/csg/abecasis/metal/). Prior to joint analysis, we filtered out SNPs with low minor allele frequency, MAF (< 1 %) and poor imputation quality and applied genomic control correction where the genomic control parameter lambda (λ_{GC}) was > 1.0.

Animals

Animal studies were performed in accordance with the Institutional Animal Care and Use Committee of Seoul National University Hospital. C57BL/6 mice were sacrificed at 20 weeks of age. For tissue analysis, mice were anesthetized with Avertin (500mg/kg). All efforts were made to minimize the pain during procedure. Tissues were isolated, snap-frozen in liquid nitrogen, and stored at -80 ℃ until use.

Cell culture, differentiation and transfection

Mouse skeletal muscle cell line, C2C12 myoblasts were maintained in High glucose (25mM) Dulbecco's Modified Eagle Medium (DMEM) (Hyclone laboratories, Logan, UT) supplemented with 10 % FBS (invitrogen, Carlesbad, CA) at 37 ℃ under a 5 % CO_2. Differentiation was induced using DMEM with 2 % horse serum (Invitrogen, MA, USA), and the media were changed every two days. Seven days after 2

% horse serum addition, C2C12 myotubes were transfected with 50 nM FTO siRNA (Dharmacon, Cat. No. L-062238-01, Chicago, II) using Lipofectamine RNAiMAX (Invitrogen, MA, USA) for 48 hours.

RNA preparation and Real-time PCR

Total RNAs from differentiated C2C12 cells were isolated using TRIzol reagent (Invitrogen, MA, USA) according to the manufacturer's instruction. To synthesize cDNA, 1㎍ of total RNA, 5㎕ of reaction buffer, 2.5㎕ of 100mM DTT, 1.25㎕ of 10mM dNTP, 0.5㎕ of Oligo dT, 0.25㎕ of RNase inhibitor and 1㎕ of RTase (Invitrogen, MA, USA) were mixed and RNA free water added up to 25㎕. The mixture was incubated at 37 ℃ for 1 hour and at 72 ℃ for 10 min using PCR system. Expression levels of genes were determined by using quantitative real-time PCR with SYBR-MASTER MIX (Takara, Otsu, Shiga, Japan) and ABI 7500 Real-time PCR system (Applied Biosystem, CA, USA). The primer sequences for qPCR were listed on Table 1.

Western blot analysis

Cells were harvested in RIPA buffer (0.5M Tris-HCl, pH 7.4, 1.5M NaCl, 2.5 % deoxycholic acid, 10 % NP-40, 10mM EDTA) (Merck Millipore, Temecula, CA) containing

protease inhibitors (10 μg/μl aprotnin, 10 μg/μl leupeptin and 1mM PMSF). Cell debris were removed by centrifugation (13,000 rpm) for 20 min at 4 ℃.

Frozen mouse tissues were powdered in liquid nitrogen and incubated in RIPA buffer on the rotator for 2 hours at 4 ℃. Concentration of protein was determined by bicinchoninic-acid (BCA) assay. About 15~20 μg of proteins were separated on SDS-PAGE and transferred to a nitrocellulose membrane (Whatman, Dassel, Germany). Membranes were incubated in blocking solution (5 % skim milk in Tris-buffered saline with Tween 20) for 1 hour at room temperature and then incubated with primary antibody overnight at 4 ℃ or 3 hour at room temperature. Antibodies aginst FTO (ABcam, Cambrige, UK) and GAPDH were used.

Table 1. Primer sequences used for quantitative PCR

Target		Primer sequences (5' → 3')
FTO	Forward	5' TTC ATG CTG GAT GAC CTC AAT G 3'
	Reverse	5' GCC AAC TGA CAG CGT TCT AAG 3'
MuRF-1	Forward	5' GCT GGT GGA AAA CAT CAT TGA CAT 3'
	Reverse	5' CAT CGG GTG GCT GCC TTT 3'
Atrogin-1	Forward	5' CAC ATT CTC TCC TGG AAG GGC 3'
	Reverse	5' TTG ATA AAG TCT TGA GGG GAA 3'
GAPDH	Forward	5' AGG TCG GTG TGA ACG GAT TTG 3'
	Reverse	5' TGT AGA CCA TGT AGT TGA GGT CA 3'

Statistical analysis

Most of the genome wide association testing was performed using PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell /plink/). In each cohort, a linear regression model with additive genetic effect was applied to test. Covariates adjusted in the model included sex, age, height and fat mass measured by BIA.

The QQ plot, which shows the distribution of the observed P values of the linear regression analysis against the expected distribution under the null hypothesis, was generated by using the R statistical package (http://www.r-project.org). The Manhattan plot, which depicts the negative log10 of P values derived from the linear regression analysis, was plotted against the chromosomal position by using the R statistical package. The dense regional association result of joint analysis was plotted using LocusZoom software (http://csg.sph.umich.edu/locuszoom/).

A threshold of $P < 5 \times 10^{-8}$ was pre-specified as being genome-wide significant, while a threshold of $P < 1.0 \times 10^{-5}$ was used to select SNPs for suggestive genome-wide significant.

Statistical analysis of experimental data was used for GraphPad Prism version 5.0 (GraphPad Software, Inc., La

Jolla, CA). Results are expressed as the mean ± SEM. Statistical significance was calculated using the Mann-Whitney U-test. A p-value less than 0.05 was denoted as statistically significant difference.

Results

Clinical characteristics of the study cohort

A total of 13,105 participants, 5,817 participants from Ansung-Ansan cohort and 7,288 participants from Gangnam center cohort, were included for the analysis. The mean age of Ansung-Ansan cohort was older than that of Gangnam center cohort (57.9 ± 8.7 years than 50.4 ± 10.2 years). Female particicpants had lower lean body mass than male in both two cohorts. The other descriptions and clinical characteristics of the study populations are shown in table 2.

Table 2. Clinical characteristics of Study cohort

Characteristics	Ansung-Ansan cohort			Gangnam center cohort		
	Female	Male	Total	Female	Male	Total
N (%)	3,060 (52.6)	2,757 (47.4)	5,817	3,063 (42.0)	4,225 (58.0)	7,288
Age, years	58.3±8.8	57.4±8.5	57.9±8.7	49.1±10.2	51.3±10.1	50.4±10.2
Height, cm^2	153.6±5.8	166.9±5.8	159.9±8.8	159.8±5.2	171.6±5.7	166.6±8.0
Weight, kg	58.3±8.3	67.7±9.6	62.8±10.1	55.0±7.3	71.5±9.2	64.6±11.8
Waist Circumference, cm	82.9±10.7	85.9±8.1	84.3±9.6	77.1±7.6	86.7±7.1	82.6±8.7
SBP, mmHg	117.2±17.0	118.2±14.7	117.7±16.0	111.2±13.4	118.1±12.5	115.2±13.3
DBP, mmHg	73.9±9.0	76.9±9.2	75.3±9.2	71.2±9.5	79.1±9.4	75.7±10.2
HbA1c, %	5.8±0.9	5.8±0.9	5.8±0.9	5.6±0.5	5.7±0.6	5.7±0.5
Fasting glucose, mg/dL	96.7±27.5	104.2±33.1	100.3±30.5	93.8±13.9	102.6±17.8	98.9±16.9
Total cholesterol, mg/dL	199.9±34.9	191.1±34.6	195.7±35.0	195.3±33.8	190.5±33.5	192.5±33.7
Triglycerides, mg/dL	130.2±78.4	153.7±103.7	141.4±92.0	84.5±47.4	125.7±82.0	108.4±72.5
HDL-cholesterol, mg/dL	45.5±10.5	42.7±10.5	44.2±10.6	59.0±12.0	49.9±10.4	53.8±12.0
LDL-cholesterol, mg/dL	129.7±32.1	121.3±34.4	125.7±33.5	117.1±30.2	121.0±29.5	119.4±29.8
Lean body mass, kg	40.0±4.5	53.1±6.4	46.2±8.6	38.4±3.8	54.7±6.0	47.8±9.6
Fat mass, kg	18.5±5.2	14.7±4.9	16.7±5.4	16.6±5.1	16.9±5.1	16.8±5.1

Data are means±SD

Stage I: Genome-wide association analyses in each cohorts

In the stage 1 genome-wide association analysis, linear regression analysis using an additive genetic model was used to test for the association between the genotypes and lean body mass. QQ plot showed the distribution of the observed P values from the linear regression analysis for the stage 1 genome-wide association analyses against the expected distribution under the null hypothesis (Figure 1A and 2A). The negative log10 of the *P* values from the association test was plotted against their genomic position in Figure 1B (Ansung-Ansan cohort) and 2B (Gangnam center cohort).

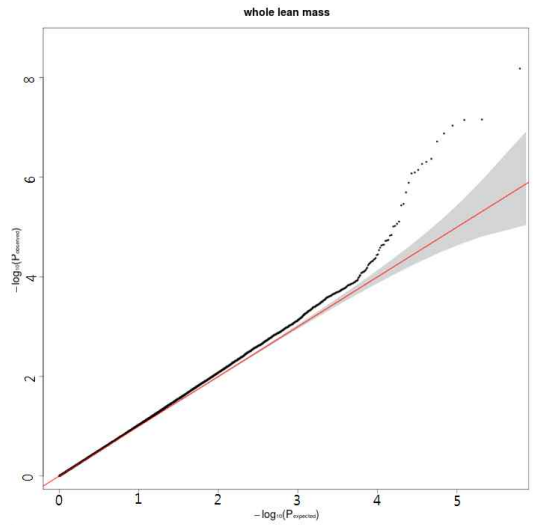
Ansung-Ansan cohort

Top SNPs for Ansung-Ansan cohort are listed in Table 3. A total of twenty independent SNPs were suggestive of an association according to our predefined arbitrary threshold of $P < 1.0 \times 10^{-5}$. Among these, a variant in *SCAPER* (rs2629032) showed a genome-wide significant association with lean body mass ($P = 6.57 \times 10^{-9}$).

Also, three variants (rs12122759, rs6681959, rs7526822) which showed suggestive genome-wide significant association with lean body mass are commonly located in *CDC42BPA*.

Figure 1. (A) QQ plot and (B) Manhattan plot of Ansung-Ansan cohort

(A)



(B)

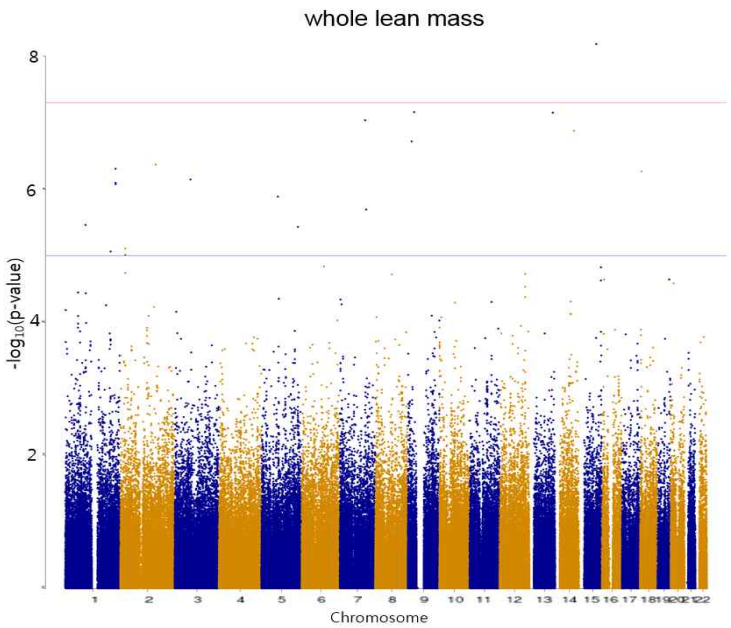
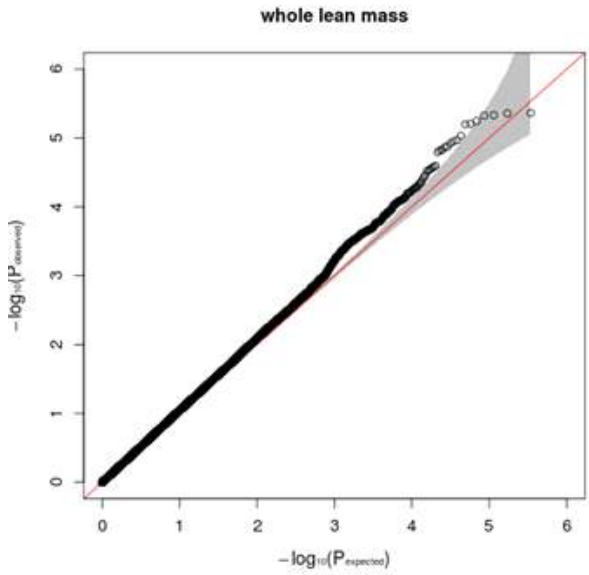


Figure 2. (A) QQ plot and (B) Manhattan plot of Gangnam center cohort

(A)



(B)

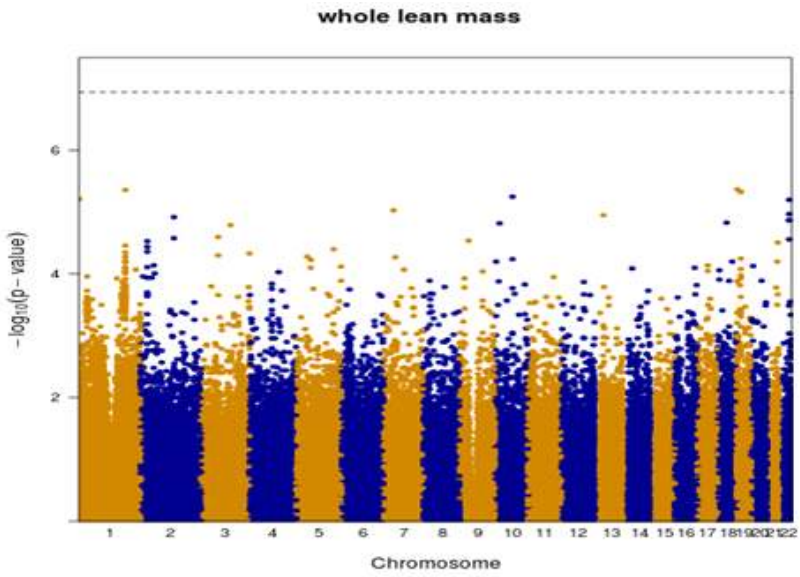


Table 3. Top SNPs for Ansung-Ansan cohort

SNP ID	Chromosome	Position	Minor Allele	Beta	SE	*P*-value	Related Gene
rs2629032	15	77,035,637	T	0.2022	0.03479	6.57E-09	*SCAPER*
rs16912073	9	27,960,365	G	0.179	0.03315	6.92E-08	*LINGO2*
rs2669460	13	104,643,963	A	0.1795	0.03326	7.06E-08	*DAOA-AS1*[*]
rs2968848	7	113,435,362	T	0.2007	0.03751	9.17E-08	*PPP1R3A*[*]
rs10484108	14	84,272,455	C	0.1538	0.02912	1.32E-07	*Intergenic*[**]
rs4961725	9	16,599,255	T	0.1514	0.02904	1.91E-07	*BNC2*
rs9973369	2	157,791,556	G	0.1161	0.02292	4.26E-07	*GPD2*[*]
rs12122759	1	227,217,340	C	-0.05375	0.01067	4.91E-07	*CDC42BPA*
rs16950424	18	6,646,790	C	0.1672	0.03332	5.40E-07	*ARHGAP28*[*]
rs2699494	3	70,863,811	C	0.1204	0.02426	7.12E-07	*FOXP1*[*]
rs6681959	1	227,241,676	G	-0.05289	0.01071	8.02E-07	*CDC42BPA*
rs7526822	1	227,263,491	T	-0.05264	0.01067	8.33E-07	*CDC42BPA*

[*] Intergenic

[**] Not near any genes within 50 kb

Analyses stratified by sex

We performed sex-specific analyses with the Ansung-Ansan cohort. In women, variants near *PPP1R3A* (rs2968848) had Genome-wide significant *P* values ($P = 2.02 \times 10^{-8}$). Another twelve SNPs which showed suggestive genome-wide significant association with lean body mass were listed in table 4.

Top SNPs associated with lean body mass in men are shown in table 5. A variant, rs16998308, near DMD showed the strongest association with lean body mass in men ($P = 7.70 \times 10^{-21}$).

Gangnam center cohort

None of the association reached genome side significance in Gangnam center cohort. Eight variants which showed suggestive genome-wide association were listed in table 6.

Compared to the result of the Ansung-Ansan cohort, there were no common SNPs showing significant association with lean body mass in both cohorts considering the results of each cohort.

Table 4. Top SNPs for Ansung-Ansan cohort according to sex (Females)

SNP ID	Chromosome	Position	Minor Allele	Beta	SE	*P*-value	Related Gene
rs2968848	7	113,435,362	T	0.2728	0.04849	2.02E-08	*PPP1R3A*[*]
rs2629032	15	77,035,637	T	0.2322	0.04422	1.61E-07	*SCAPER*
rs16912073	9	27,960,365	G	0.212	0.04168	3.85E-07	*LINGO2*
rs10503642	8	18,831,664	G	0.08688	0.01781	1.13E-06	*PSD3*
rs10484108	14	84,272,455	C	0.1834	0.03768	1.19E-06	*Intergenic*[**]
rs745732	1	92,012,189	G	-0.04835	0.009942	1.21E-06	*CDC7*[*], *TGFBR3*[*]
rs2669460	13	104,643,963	A	0.1992	0.04135	1.53E-06	*DAOA-AS1*[*]
rs1456806	23	65,485,431	G	0.1201	0.02526	2.06E-06	*Intergenic*[**]
rs16950424	18	6,646,790	C	0.1991	0.0425	2.91E-06	*ARHGAP28*[*]
rs17160628	5	107,159,013	T	0.07647	0.01671	4.90E-06	*EFNA5*[*], *FBXL17*[*]
rs7844906	8	18,830,330	T	0.08255	0.01811	5.37E-06	*PSD3*
rs10849279	12	5,612,994	G	-0.05157	0.01133	5.56E-06	*NTF3*[*], *ANO2*[*]

[*] Intergenic

[**] Not near any genes within 50 kb

Table 5. Top SNPs for Ansung-Ansan cohort according to sex (Males)

SNP ID	Chromosome	Position	Minor Allele	Beta	SE	*P*-value	Related Gene
rs16998308	23	32,442,346	C	0.1623	0.01719	7.70E-21	*DMD*[*]
rs11033031	11	35,244,643	A	-0.1023	0.02023	4.49E-07	*CD44*
rs1203605	23	135,566,495	C	0.08686	0.01931	7.10E-06	*ADGRG4*[*], *BRS3*[*]
rs1073260	7	131,464,021	G	0.05415	0.01232	1.14E-05	*PODXL*[*]
rs7016751	8	74,021,818	G	0.2094	0.04881	1.85E-05	*SBSPON*[*], *C8orf89*[*]
rs5024581	4	26,780,582	G	-0.04608	0.01083	2.18E-05	*TBC1D19*[*], *STIM2*[*]
rs11177202	12	68,791,570	C	0.1807	0.0426	2.29E-05	*MDM1*[*]
rs12651043	4	10,762,637	T	0.06457	0.01526	2.39E-05	*CLNK*[*], *MIR572*[*]
rs6696501	1	204,680,577	T	0.04672	0.01106	2.50E-05	*LRRN2*[*], *NFASC*[*]
rs997250	4	15,739,498	G	-0.078	0.01858	2.78E-05	*BST1*[*], *CD38*
rs604302	8	37,004,569	T	0.04974	0.01203	3.65E-05	*MIR1268A*
rs10207503	2	60,586,661	A	-0.05616	0.0137	4.26E-05	*MIR4432HG*

[*] Intergenic

Table 6. Suggestive SNPs for Gangnam center cohort

SNP ID	Chromosome	Position	Minor Allele	Beta	SE	*P*-value	Related Gene
rs7259757	19	1,996,352	C	0.03298	0.007169	4.30E-06	*BTBD2*
rs3010648	1	188,209,895	G	-0.0336	0.007309	4.38E-06	*RLA2G4A*[*]
rs62104727	19	15,869,508	A	-0.05493	0.01198	4.64E-06	*OR10H3*[*]
rs10410357	19	15,800,331	G	-0.05349	0.01168	4.71E-06	*CYP4F12*
rs10762395	10	72,353,700	G	0.03608	0.007943	5.65E-06	*PALD1*[*], *PRF1*[*]
rs2494641	1	2,362,827	T	-0.0338	0.007474	6.21E-06	*PLCH2*, *PEX10*[*]
rs17002034	22	40,996,367	T	-0.03501	0.007747	6.31E-06	*MKL1*
rs4723738	7	38,240,908	A	0.03195	0.0072	9.23E-06	*STARD3NL*

[*] Intergenic

Stage II: Joint analysis of the two cohorts

We perfomed a joint analysis with the Ansugn-Ansan cohort and the Gangnam center cohort using METAL package. The results are shown using a QQ plot and a Manhattan plot (Figure 3). Five SNPs in *FTO* (rs3751812, rs9939609, rs7202116, rs17817449, rs8050136) were associated with lean body mass in our joint analysis with suggestive genome-wide significance. The most significant of the SNPs showed a *P* values of 1.95×10^{-6}. Four other variants in/near *GALNT16, ERH, PLCE1, STK39, PRF1, ADAMTS14* were detected in the joint analysis. These common SNPs and their *P* values in each cohort are listed in Table 7.

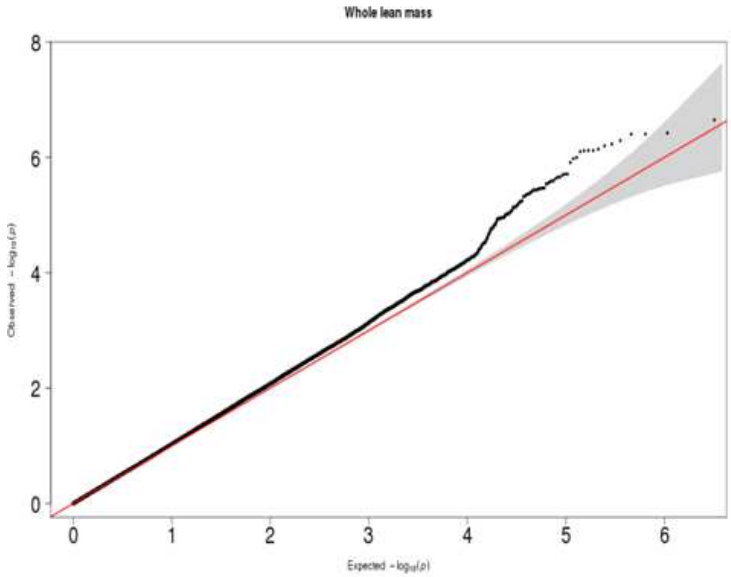
The dense regional association plot of SNPs near *FTO* imputed with 1000G database is depicted in Figure 4. It shows that these suggestive loci in *FTO* were in strong linkage disequilibrium (LD) with each other.

Analyses stratified by sex

When the joint analysis was performed according to sex, there was no suggestive genome-wide significant loci in women (Table 8). In men, eight SNPs showed suggestive genome-wide significant association with lean body mass (Table 9). A variant, rs12251307, near *IL2RA* and *RBM17* showed the strongest association with lean body mass in men ($P = 3.95 \times 10^{-7}$).

Figure 3. (A) QQ plot and (B) Manhattan plot of joint analysis

(A)



(B)

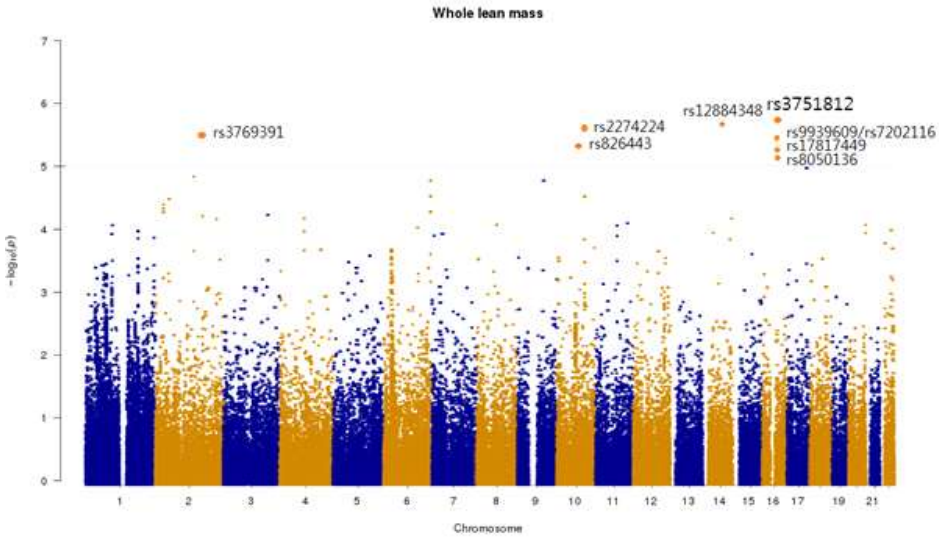
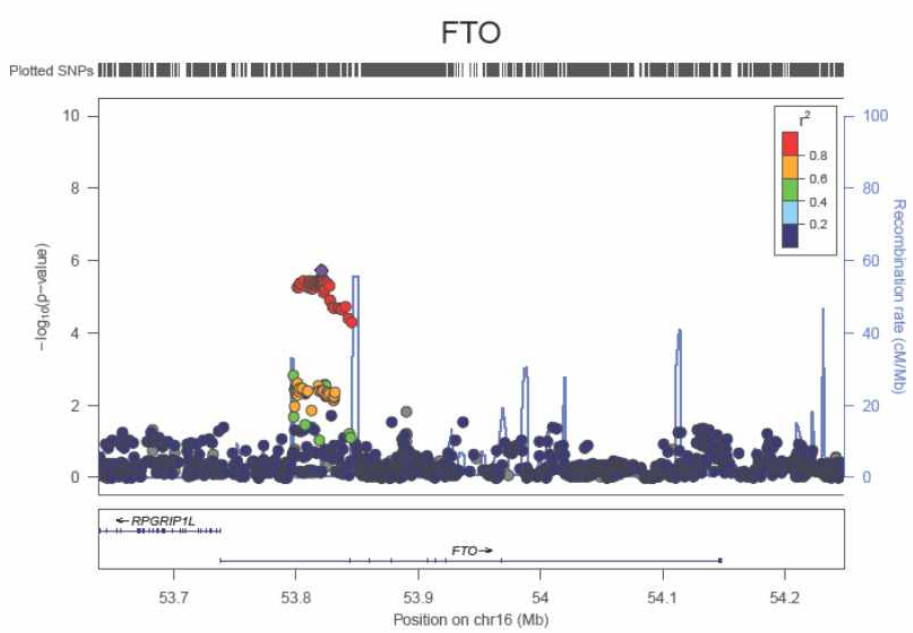


Table 7. Common SNPs for joint analysis

SNP ID	Chr	Position	Related Gene	Minor Allele	*P*-value	Ansung-Ansan cohort			Gangnam center cohort		
						Beta	SE	*P*-value	Beta	SE	*P*-value
rs3751812	16	53,818,460	*FTO*	T	1.95E-06	0.03558	0.01369	1.46E-03	0.03855	0.01089	4.05E-04
rs12884348	14	69,826,100	*GALNT16** *ERH**	T	2.20E-06	0.03596	0.01821	7.23E-05	0.02470	0.00881	5.08E-03
rs2274224	10	96,039,597	*PLCE1*	C	2.57E-06	-0.03410	-0.04860	4.17E-06	-0.01564	0.00709	2.75E-02
rs3769391	2	169,013,688	*STK39*	T	3.40E-06	0.02706	0.00951	2.52E-03	0.03093	0.00876	4.16E-04
rs9939609	16	53,820,527	*FTO*	A	3.50E-06	0.03505	0.01314	1.72E-03	0.03685	0.01077	6.25E-04
rs7202116	16	53,821,615	*FTO*	A	3.67E-06	0.03547	0.01339	1.65E-03	0.03680	0.01083	6.83E-04
rs826443	10	72,375,542	*PRF1*[*] *ADAMTS14*[*]	T	4.77E-06	0.02765	0.00475	1.8E-02	0.04632	0.01151	5.77E-05
rs17817449	16	53,813,367	*FTO*	T	5.79E-06	0.03539	0.01351	1.54E-03	0.03516	0.01081	1.16E-03
rs8050136	16	53,816,275	*FTO*	A	7.38E-06	0.03503	0.01322	1.65E-03	0.03465	0.01082	1.38E-03

[*] Intergenic

Figure 4. Dense regional association plot near *FTO*



The hash marks above the panel represent the positions of each SNP that was genotyped or imputed. The negative log10 of *P* values from linear regression are shown in the panel. The purple diamond indicates the SNP with the most significant association in the joint analysis of imputed genome scan. The estimated recombination rates are plotted to reflect recombination hot spots. The SNPs in LD with the most significant SNP are color coded to represent their strength of LD. The locations of genes, exons and introns are shown in the lower panel (taken from the Human Genome hg18 build).

Table 8. Common SNPs for joint analysis according to sex (Female)

SNP ID	Chr	Position	Related Gene	Minor Allele	*P*-value	Ansung-Ansan cohort			Gangnam center cohort		
						Beta	SE	*P*-value	Beta	SE	*P*-value
rs9384679	6	108,864,419	*LACE1** *FOXO3**	T	5.31E-05	-0.03909	-0.06110	5.08E-04	-0.02836	0.01266	2.52E-02
rs2844529	6	31,353,593	*HLA-B** *MICA**	A	5.62E-05	0.02283	0.00323	2.25E-02	0.03787	0.01109	6.46E-04
rs2428486	6	31,354,104	*HLA-B** *MICA**	T	6.22E-05	0.02254	0.00295	2.42E-02	0.03784	0.01110	6.62E-04
rs13172457	5	83,929,767	*EDIL3* *NBPF22P*	T	6.49E-05	0.03043	0.00646	1.29E-02	0.04308	0.01361	1.56E-03
rs9375218	6	123,406,692	*GSG1L*	A	6.64E-05	-0.01747	-0.03995	1.28E-01	-0.05041	0.01220	3.94E-05
rs11074881	16	27,823,714	*FTO*	T	7.16E-05	0.04646	0.01116	9.93E-03	0.06291	0.02069	2.38E-03
rs6037255	20	2,639,570	*IDH3B*	T	7.82E-05	-0.02916	-0.05048	7.38E-03	-0.03466	0.01191	3.65E-03
rs5755921	22	36,104,730	*APOL6** *APOL5**	T	7.85E-05	0.02962	0.00759	8.44E-03	0.03630	0.01229	2.85E-03
rs6500983	16	7,605,888	*RBFOX1*	A	8.50E-05	0.03332	0.00796	1.01E-02	0.04297	0.01439	4.90E-03
rs7180119	15	94,193,200	*RGMA*	T	9.51E-05	-0.02930	-0.05050	6.82E-03	-0.03321	0.0118	1.18E-02

* Intergenic

Table 9. Common SNPs for joint analysis according to sex (Male)

SNP ID	Chr	Position	Related Gene	Minor Allele	*P*-value	Ansung-Ansan cohort			Gangnam center cohort		
						Beta	SE	*P*-value	Beta	SE	*P*-value
rs12251307	10	6,123,495	*IL2RA** *RBM17**	T	3.95E-07	0.04024	0.01509	1.73E-03	0.04299	0.01077	6.66E-05
rs3769391	2	169,013,688	*STK39*	T	4.28E-07	0.03991	0.0139	2.66E-03	0.04641	0.01138	4.66E-05
rs12217047	1	188,203,764	*BRINP3**	A	1.18E-06	0.01702	-0.006376	1.54E-01	0.05098	0.01000	3.54E-07
rs2046606	1	188,157,788	*BRINP3**	T	1.62E-06	0.01786	-0.005517	1.34E-01	0.0495	0.00997	7.22E-07
rs10846172	12	15,564,674	*PTPRO*	A	1.86E-06	-0.04427	-0.07409	3.65E-03	-0.05164	0.01364	1.56E-04
rs7387249	8	71,436,375	*TRAM1*	A	4.33E-06	-0.1053	-0.1561	5.11E-05	-0.05586	0.02116	8.34E-03
rs11589480	1	209,028,716	*PLXNA2** *MIR205HG**	A	5.07E-06	-0.05314	-0.09803	2.04E-02	-0.07524	0.01883	6.57E-05
rs2274224	10	96,039,597	*PLCE1*	C	8.59E-06	-0.04248	-0.06397	1.10E-04	-0.02397	0.00920	9.20E-03
rs17034377	4	105,145,350	*TACR3** *CXXC4**	T	1.02E-05	0.05872	0.02162	1.94E-03	0.0487	0.01534	1.51E-03
rs4438452	2	169,071,348	*STK39*	T	1.66E-05	0.0322	0.006885	1.27E-02	0.03912	0.01109	4.24E-04

* Intergenic

Functional study of the candidate gene from joint analysis

Expression of FTO in mouse white adipose tissue, liver and skeletal muscle

The C57BL/6 mice were housed for 20 weeks and their white adipose tissues, liver and gastrocnemius muscle were subjected to protein and mRNA extraction. FTO protein level was relatively decreased in gastrocnemius muscle (GM) compared to the white adipose tissue (Figure 5A). FTO mRNA levels in gastrocnemius muscle were lower than FTO mRNA levels in white adipose tissue and liver (Figure 5B).

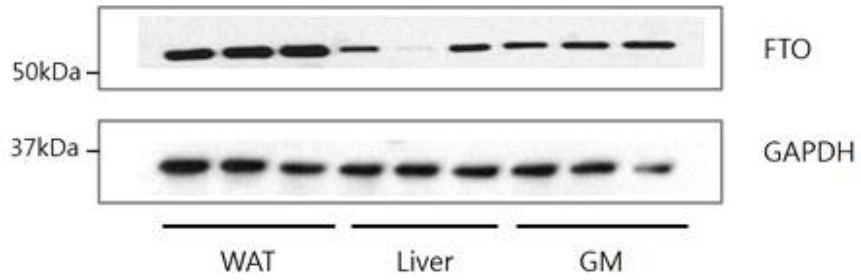
Silencing FTO and C2C12 degradation by ubiquitin-proteasome pathway

Then we silenced FTO to see if the alteration of FTO can influence the muscle degradation by ubiquitin-proteasome pathway. C212 cells were transfected with siNS or siFTO on day 7 of differentiation. FTO knockdown was successfully performed after 48 hours of transfection (Figure 6).

The relative mRNA expressions of two muscle-specific ubiquitin ligases, MuRF-1 and Atrogin-1, which are important regulators of ubiquitin-mediated protein

degradation in skeletal muscle were measured by RT-PCR. There was no significant difference between siNS and siFTO in each mRNA expression of MuRF-1 and Atrogin-1 (Figure 7).

(A)



(B)

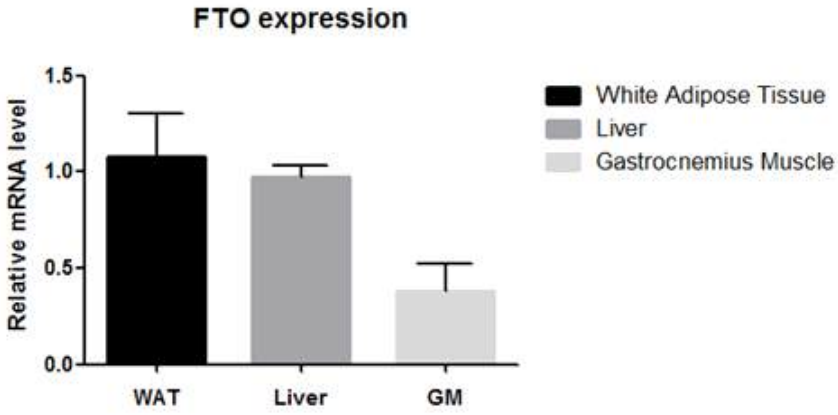


Figure 5. FTO expression in various mouse tissues

(A) Protein extracts were prepared from epididymal fat, liver and gastrocnemius muscle of chow diet mice.

(B) Total RNA was isolated from epididymal fat, liver and gastrocnemius muscle. The RNA (1 μg) was subjected to real-time PCR with primers specific to FTO (Table 1). WAT (n=6), Liver (n=4), GM (n=6).

WAT, white adipose tissue; GM, gastrocnemius muscle

(A)

(B)

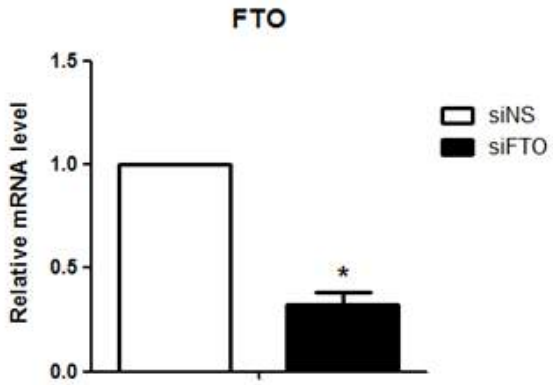


Figure 6. Silencing FTO in C2C12

(A) Western blot of FTO. Protein level of FTO in C2C12 decreased by 50nM siFTO for 48 hours. n=3

(B) Relative mRNA expression of FTO in C2C12 myotubes after siFTO transfection for 48 hours showing 62 % reduction compare to siNS. n=3 (* = $p < 0.001$ vs siNS)

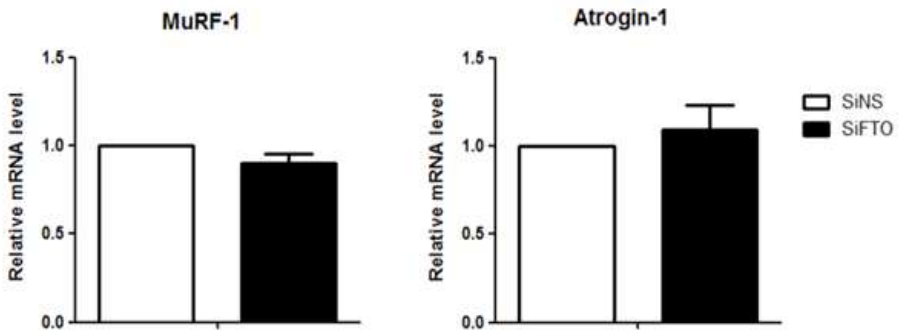


Figure 7. mRNA expression of C2C12 degradation by ubiquitin-proteasome pathway

The relative mRNA expressions of MuRF-1 and Atrogin-1, which are two muscle-specific ubiquitin ligases, in C2C12 myotubes transfected with siNS or siFTO on day 7 for 48 hours. n=3

Discussion

In this first genome-wide association join analysis study of lean body mass in Korean participants including the representative cohorts of Korea, we identified nine GWS loci (in/near *FTO, GALNT16, ERH, PLCE1, STK39, PRF1* and *ADAMTS14* genes) for lean body mass which are crucial in sarcopenia diagnosis.

Among them, five variants including the most significant loci (rs3751812) were located in *FTO*. As for the *FTO* locus, the *FTO* gene has been found to be associated with adiposity/obesity [17] and its related traits, such as BMI [18] and metabolic syndrome/type 2 diabetes [19]. Apart from its role in adiposity, SNPs in *FTO* were found to be associated with both DXA-derived fat and lean mass in two recent candidate gene studies. In a previous study [20], the association for lean mass was only slightly attenuated after fat mass adjustment. In the study [20], they suggest that the *FTO* association with body size variation is mediated via both fat body mass and lean body mass, but not fat mass specifically. Similar to the result, we detected that variants in *FTO* were associated with lean body mass adjusted by fat mass.

SNPs in *FTO* from the result were not suggested to be the causal variants affecting *FTO* expression, when

examining the association through eQTL analysis. However, since these SNPs were in strong LD, while one of them, rs9939609, was recently reported to be associated with lean body mass [21], we can expect the SNPs to play their roles in lean body mass.

In another study, the expression of FTO protein was detected in all of the major mouse tissues examined, with higher levels of expression in the brain and lower in the skeletal muscle [22]. Consistent with the previous results, the expression of FTO protein and gene in gastrocnemius muscle were relatively lower than those of white adipose tissue and liver in our study (Figure 5).

The role of *FTO* in skeletal muscle is poorly understood, although a couple of recent studies mentioned that the cells lacking *FTO* showed a decrease in activity of the mTORC1 pathway [23] and the FTO-deficient mice showed a reduction in lean mass [24].

There are two phenotypes in sarcopenia model; myoblasts differentiation and muscle protein degradation. Interestingly, *FTO* expression has already been shown to increase during myoblasts differentiation, while the silence of *FTO* inhibits the differentiation. In addition, skeletal muscle development was impaired in skeletal muscle FTO-deficient mice [25].

Therefore, we focused on the correlation between *FTO* expression and muscle protein degradation. Muscle

atrophy is a tightly-controlled process involving many signaling pathways. In particular, the ubiquitin-proteasome pathway plays an important role in muscle protein degradation. MuRF-1 and Atrogin-1 are two muscle-specific ubiquitin ligases that are important regulators of ubiquitin-mediated protein degradation in skeletal muscle [26]. To investigate whether the ubiquitin-proteasome system was activated in C2C12 cells when transfected with FTO, mRNA expression levels of MuRF-1 and Atrogin-1 were determined using RT-PCR. However, there wes no significant difference of MuRF-1 and Atrogin-1 expression after 2 days of FTO siRNA transfection in our study.

There was no significant effect of *FTO* in muscle degradation. However, based on previous studies of myoblast differentiation and in vivo study, *FTO* gene may be involved in the regulation of lean mass. So the actual function of the variants and the underlying mechanisms of *FTO*' s involvement in skeletal muscle biology still need to be further elucidated by *in vitro* and animal experiments.

During the study, we also found that rs12122759 in *CDC42BPA* showed different gene expression according to risk alleles. The protein encoded by *CDC42BPA* is a member of the Serine/Threonine protein kinase family, which is an important downstream effector of *CDC42* and plays a role in the regulation of cytoskeleton reorganization and cell migration. Although there is no literature about *CDC42PBA*

and its association with sarcopenia to date, it can be a novel candidate marker for sarcopenia.

Among the Genome-wide significant results in this study, *DMD,* which showed an association with lean body mass in males of Ansung-Ansan cohort, is a protein coding gene. A protein encoded by *DMD* forms a component of the dystrophin-glycoprotein complex (DGC), which bridges the inner cytoskeleton and the extracellular matrix. It is well known that deletions, duplications, and point mutations at this gene locus may cause Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), or cardiomyopathy. Sarcopenia and muscular dystrophy possess several similar characteristics as pointed out in more recent review by Rudolf et al [27]. So, the variants in *DMD* may play significant roles in mechanisms of sarcopenia. Following *in vitro* and animal experiments need to confirm the expected roles.

In conclusion, through genome-wide association join analysis study of lean body mass in Korean participants in the representative cohorts of Korea, we identified nine GWS loci (in/near *FTO, GALNT16, ERH, PLCE1, STK39, PRF1* and *ADAMTS14* genes) for lean body mass which are crucial in sarcopenia diagnosis. The actual function of the variants and the underlying mechanisms of their roles in skeletal muscle biology are yet to be clearly elucidated by following *in vitro* and animal experiments.

References

1. Grimby, G. and B. Saltin, *The ageing muscle.* Clin Physiol, 1983. **3**(3): p. 209-18.

2. Rosenber, I.H., *Epidemiologic and methodologic problems in determining nutritional status of older persons. .* Am J Clin Nutr, 1989. **50**(5 Suppl): p. 1121-235.

3. Abellan van Kan, G., *Epidemiology and consequences of sarcopenia.* J Nutr Health Aging, 2009. **13**(8): p. 708-12.

4. Kim KM, L.S., Choi KM, Kim JH, Yu SH, Kim TN, Song W, Lim JY, Won CW, Yoo HJ, Jang HC, Sarcopenia Study Group of Korean Geriatrics Society., *Sarcopenia in Korea: Prevalence and Clinical Aspects.* J Korean Geriatr Soc 2015;19:1-8., 2015.

5. Janssen, I., et al., *Skeletal muscle cutpoints associated with elevated physical disability risk in older men and women.* Am J Epidemiol, 2004. **159**(4): p. 413-21.

6. Kim, J.H., et al., *Sarcopenia: an independent predictor of mortality in community-dwelling older Korean men.* J Gerontol A Biol Sci Med Sci, 2014. **69**(10): p. 1244-52.

7. Janssen, I., et al., *The healthcare costs of sarcopenia in the United States.* J Am Geriatr Soc, 2004. **52**(1): p. 80-5.

8. Korea National Statistical Office, *Elderly person statistics.* 2015.

9. Deng, H.W., et al., *Characterization of genetic and lifestyle factors for determining variation in body mass index, fat mass, percentage of fat mass, and lean mass.* J Clin Densitom, 2001. **4**(4): p. 353-61.

10. Hsu, F.C., et al., *Heritability of body composition measured by DXA in the diabetes heart study.* Obes Res, 2005. **13**(2): p. 312-9.

11. Liu, X.G., et al., *Genome-wide association and replication studies identified TRHR as an important gene for lean body mass.* Am J Hum Genet, 2009. **84**(3): p. 418-23.

12. Hai, R., et al., *Genome-wide association study of copy number variation identified gremlin1 as a candidate gene for lean body mass.* J Hum Genet, 2012. **57**(1): p. 33-7.

13. Guo, Y.F., et al., *Suggestion of GLYAT gene underlying variation of bone size and body lean mass as revealed by a bivariate genome-wide association study.* Hum Genet, 2013. **132**(2): p. 189-99.

14. Urano, T., et al., *Large-scale analysis reveals a functional single-nucleotide polymorphism in the 5'-flanking region of PRDM16 gene associated with lean body mass.* Aging Cell, 2014. **13**(4): p. 739-43.

15. Chien, M.Y., T.Y. Huang, and Y.T. Wu, *Prevalence of sarcopenia estimated using a bioelectrical impedance analysis prediction equation in community-dwelling elderly people in Taiwan.* J Am Geriatr Soc, 2008. **56**(9): p. 1710-5.

16. M Chen, J.S., H Bai, Y Wang, D Xu, X Zhu, J Chen, Z Bao,, *Muscle mass reference standard for sarcopenia using bioelectrical impedance analysis.* Asian J Gerontol Geriatr, 2015. **10**: p. 16-21.

17. Hinney, A., et al., *Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants.* PLoS One, 2007. **2**(12): p. e1361.

18. Yang, J., et al., *FTO genotype is associated with phenotypic variability of body mass index.* Nature, 2012. **490**(7419): p. 267-72.

19. Yang, Y., et al., *FTO Genotype and Type 2 Diabetes Mellitus: Spatial Analysis and Meta-Analysis of 62 Case-Control Studies from Different Regions.* Genes (Basel), 2017. **8**(2).

20. Livshits, G., et al., *Association of FTO gene variants with body composition in UK twins.* Ann Hum Genet, 2012. **76**(5): p. 333-41.

21. Heffernan, S.M., et al., *Fat mass and obesity associated (FTO) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position.* BMC Genet, 2017. **18**(1): p. 4.

22. Gao, X., et al., *The fat mass and obesity associated gene FTO functions in the brain to regulate postnatal growth in mice.* PLoS One, 2010. **5**(11): p. e14005.

23. Gulati, P., et al., *Role for the obesity-related FTO gene in the cellular sensing of amino acids.* Proc Natl Acad Sci U S A, 2013. **110**(7): p. 2557-62.

24. Fischer, J., et al., *Inactivation of the Fto gene protects from obesity.* Nature, 2009. **458**(7240): p. 894-8.

25. Wang, X., et al., *FTO is required for myogenesis by positively regulating mTOR-PGC-1alpha pathway-mediated mitochondria biogenesis.* Cell Death Dis, 2017. **8**(3): p. e2702.

26. Gumucio, J.P. and C.L. Mendias, *Atrogin-1, MuRF-1, and sarcopenia.* Endocrine, 2013. **43**(1): p. 12-21.

27. Rudolf, R., et al., *Degeneration of neuromuscular junction in age and dystrophy.* Front Aging Neurosci, 2014. **6**: p. 99.

국문초록

근감소증은 노화에 수반되는 현상 중 하나로, 점진적인 근육의 양적, 질적 저하를 뜻하며 노인에서 신체 수행도를 감소시킨다. 그 중 근육량은 주로 제지방으로 구성되며 유전력이 높은 것으로 알려져 있다.

전장유전체연관성분석이 다양한 복합 질환의 유전적 위험 인자를 밝히는데 널리 활용되면서, 근감소증에 대한 유전 연구도 증가하고 있다. 하지만, 근감소증과 관련된 한국인의 유전적 위험 인자에 대해서는 평가된 바가 없다. 이 연구의 목적은 한국인에서 제지방량과 관련된 유전변이형을 확인하는 것이다.

본 연구에서는 한국의 대표적인 두 코호트에 대하여 각각 전장유전체연관성분석을 수행한 후, 각각의 분석에서는 발견되지 않았으나 두 코호트에서 공통적으로 발견되는 유전변이형을 찾기 위하여 공동 분석을 시행하였다. 안성-안산 코호트의 5,817명의 대상자와 강남 센터 코호트의 7,288명의 대상자를 포함한 총 13,105명 대상자의 유전변이형을 분석하였다. METAL 도구를 사용하여 공동분석을 수행하였고 성별, 연령, 신장, 체지방량을 보정하여 선형 회귀 분석을 수행하였다. p 값 0.00001 미만을 유전변이형과 표현형질과의 연관성을 시사할 수 있는 임의의 기준으로 정의하였다.

두 코호트의 공동 분석 결과, 5×10^{-8} 미만의 p 값을 만족하는 유전변이형은 없었다. 하지만 FTO 유전자의 5가지 유전변이형은 제지방량과 연관이 있는 것으로 확인되었다 ($p = 1.95 \times 10^{-6}$). 최근 다른 연구에서 근원세포가 분화하는 동안 FTO 유전자의 표현이 증가하고 FTO의 유전형을 억제하면 근원세포의 분화가 억제되

는 것이 밝혀졌다. 따라서 본 연구에서는 근감소증 모델의 또 다른 근감소 표현형인 근육 단백질 분해와 FTO와의 상관 관계를 확인하고자 하였다. 근원세포의 분화과정에서 FTO siRNA를 처리한 후, 골격근에서 단백질 분해의 중요한 조절 인자로서 역할을 하는 근육 특이성 ubiquitin ligase인 MuRF-1과 Atrogin-1 mRNA 발현을 확인하였으나 유의한 차이는 없었다.

본 연구에서는 국내의 대표적인 두 코호트의 공동 분석을 통해 한국인의 제지방량과 관련있는 9 가지의 유전변이형과 관련 유전자 (*FTO, GALNT16, ERH, PLCE1, STK39, PRF1* 그리고 *ADAMTS14*)를 확인하였다. 이러한 유전변이형들과 후보 유전자 FTO가 실제 근육량과 근감소증에 미치는 영향 및 그 기전에 대해서는 추가적인 연구와 분석이 필요하다.

.....

주요어 : 노화, 근육, 근감소증, 전장유전체연관성분석, FTO 유전자
학 번 : 2015-26008