

SCIENTIFIC REPORTS



OPEN

Gene-based genome-wide association study identified 19p13.3 for lean body mass

Received: 14 October 2016
Accepted: 17 February 2017
Published: 21 March 2017

Shu Ran^{1,*}, Lei Zhang^{2,3,*}, Lu Liu^{2,3}, An-Ping Feng^{2,3}, Yu-Fang Pei^{3,4}, Lei Zhang¹, Ying-Ying Han¹, Yong Lin¹, Xiao Li^{2,3}, Wei-Wen Kong^{2,3}, Xin-Yi You^{2,3}, Wen Zhao^{2,3}, Qing Tian⁵, Hui Shen⁵, Yong-Hong Zhang^{3,4} & Hong-Wen Deng^{1,5}

Lean body mass (LBM) is a complex trait for human health. To identify genomic loci underlying LBM, we performed a gene-based genome-wide association study of lean mass index (LMI) in 1000 unrelated Caucasian subjects, and replicated in 2283 unrelated Caucasians subjects. Gene-based association analyses highlighted the significant associations of three genes *UQCR*, *TCF3* and *MBD3* in one single locus 19p13.3 (discovery $p = 6.10 \times 10^{-5}$, 1.65×10^{-4} and 1.10×10^{-4} ; replication $p = 2.21 \times 10^{-3}$, 1.84×10^{-3} and 6.95×10^{-3} ; combined $p = 2.26 \times 10^{-6}$, 4.86×10^{-6} and 1.15×10^{-5} , respectively). These results, together with the known functional relevance of the three genes to LMI, suggested that the 19p13.3 region containing *UQCR*, *TCF3* and *MBD3* genes was a novel locus underlying lean mass variation.

The muscular tissue, as characterized by lean body mass (LBM), is related to human health. Low LBM may be related to a series of health problems, such as sarcopenia, obesity, and increased mortality^{1,2}. LBM is under genetic control, with heritability over 50%^{3,4}. Previous GWA studies have found novel single nucleotide polymorphisms (SNPs) and genes associated with LBM⁵⁻⁷. However, the vast majority of LBM candidate genes remain to be revealed. LBM can be measured accurately by dual energy X-ray absorptiometry (DXA). Body lean mass index (LMI) is frequently used to predict sarcopenia^{8,9}.

Genomic regions may present allelic heterogeneity to the phenotype, i.e., multiple variants in a region affect the phenotype jointly. In the presence of allelic heterogeneity, gene-based association test can improve statistical power and robustness of genetic association analysis by integrating multiple SNP signals into a single statistic^{10,11}. A variety of statistical methods were developed for gene-based test, such as Versatile Gene-based Association Study (VEGAS)¹². VEGAS combines p -values of multiple SNPs within a gene region into a gene-based score while accounting for linkage disequilibrium (LD) by simulating genotype data from a multivariate normal distribution in permutation-based test. VEGAS is computationally efficient because the number of permutation simulations is adaptive. Its performance is similar to other statistical methods¹², but is superior to the others in certain conditions due to its usage of large population reference panels when performing permutation. Therefore, it uses summary statistics only instead of raw genotype and phenotype data, making it suitable for the summary results from large-scale meta-analyses.

In this study, we reported a gene-based GWAS for LMI to identify genetic loci underlying variation of LBM. The discovery sample included 1000 unrelated Caucasian subjects genotyped with the Affymetrix 500 k genotyping array. The replication sample included 2283 unrelated subjects of Caucasian subjects, genotyped with the Affymetrix SNP6.0 genotyping array. Genotypes in both samples were imputed with the 1000 genomes project sequencing reference panel.

¹Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai, PR China.

²Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University, Jiangsu, PR China.

³Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Jiangsu, PR China. ⁴Department of Epidemiology and Statistics, School of Public Health, Soochow University, Jiangsu, PR China. ⁵Department of Biostatistics, Tulane University, New Orleans, Louisiana, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.-W.D. (email: hdeng2@tulane.edu)

Materials and Methods

Ethics Statement. Study participants were recruited from the cities of Omaha and Kansas city and their neighboring areas. The study was approved by institutional review boards of the Creighton University and the University of Missouri-Kansas city. All participants provided written informed consent documents before entering the study. The methods carried out in accordance with the approved study protocol.

Subjects. *Discovery sample.* The discovery sample consisted of 1,000 unrelated Caucasian subjects of European ancestry, of whom 501 were males and 499 were females. The sample was randomly selected from a large-scale cohort containing over 6000 subjects. The inclusion and exclusion criteria for cases were described in our previous publications¹³.

Replication sample. The replication sample consisted of 2283 unrelated subjects of European ancestry. There were 556 male subjects and 1727 female subjects. All subjects were healthy individuals recruited from the Midwestern United States. There was no overlap between the subjects of the discovery and the replication cohorts.

Phenotyping. All subjects completed a structured questionnaire including lifestyle, medical history, family information, anthropometric variables, etc. Lean body mass and fat body mass (FBM) were measured with a Hologic QDR 4500 W DXA scanner (Hologic Inc., Bedford, MA, USA). Weight was measured in light clothing, on a calibrated balance beam scale. Height was obtained using a calibrated stadiometer. Lean mass index (LMI, kg/m²) was calculated as the ratio of lean mass to square of height¹⁴.

Genotyping and quality control. Genomic DNA was extracted from peripheral blood leukocytes using a commercial isolation kit (Gentra Systems, Minneapolis, MN, USA). Genotyping was performed as described in our previous publication¹⁵. Briefly, the discovery cohort was genotyped with the Affymetrix Mapping 250 K Nsp and Affymetrix Mapping 250 K Sty arrays at the Vanderbilt Microarray Shared Resource at Vanderbilt University Medical Center (Nashville, TN, USA) using the standard protocol recommended by the manufacturer (Affymetrix, Inc., Santa Clara, CA, USA). The Caucasian replication cohort was genotyped using the Affymetrix SNP 6.0 arrays by the standard protocol of the manufacturer.

We followed strict quality control (QC) procedure. Samples that had a minimum call rate of 95% were included. We discarded SNPs that deviated from Hardy-Weinberg equilibrium ($p < 0.0001$) and those containing a minor allele frequency (MAF) less than 0.01. After QC, 379,319 SNPs remained in the discovery sample.

Genotype imputation. Genotype imputation was applied to both the discovery and replication samples, with the 1000 Genomes projects sequence variants as reference panel (as of August 2010). Reference sample included 283 individuals of European ancestry.

The details of genotype imputation process had been described earlier¹⁶. Briefly, strand orientations between reference panel and test sample were checked before imputation, and inconsistencies were resolved by changing the test sample to reverse strand or removing the SNP from the test sample. Imputation was performed with MINIMAC¹⁷. Quality control was applied to impute SNPs with the following criteria: imputation $r^2 > 0.5$ and MAF > 0.01 . SNPs failing the QC criteria were excluded from subsequent association analyses.

SNP-based association analyses. We used VEGAS for gene based association analyses, which takes individual SNP association p -values as input. Therefore, we performed SNP-based association analyses first to generate association signals. We used the principal component based approach for the correction of population stratification problem¹⁸. In both samples, covariates including gender, age, age², fat body mass (FBM) and the first five principal components^{19,20} derived from genome-wide genotype data were screened for significance with the step-wise linear regression model implemented in the R function stepAIC. Raw BMI values were adjusted by significant covariates (age, gender and FBM), and the residuals were normalized by inverse quantiles of standard normal distribution.

Genetic associations were examined between genotyped and/or imputed SNPs and normalized phenotypes under an additive mode of inheritance with MACH2QTL^{21,22}, which fitted phenotype by allele dosage with a linear regression model.

Linkage disequilibrium (LD) measure r^2 was calculated with Haploview²³. To examine potential confounding effect caused by population stratification, we estimated the genomic control inflation factor (λ)²⁴.

Gene-based test using VEGAS. Gene-based association test was examined to identify genes that were associated with the phenotype. We used the VEGAS approach for such analysis, which requires LD structure information and individual SNP p -values as input¹². The SNP inclusion criteria was the same as that for individual SNP tests. Specifically, we required that SNP imputation accuracy $r^2 > 0.5$ and MAF > 0.01 .

Meta-analysis. Significant genes identified in the discovery sample were further replicated in the replication sample. The two gene-based association signals were then jointly analyzed with the Fisher's method. Specifically, the Fisher's statistic was calculated as

$$\chi_{(4)}^2 = -2(\ln(p_1) + \ln(p_2))$$

where p_1 and p_2 were the two gene-level p -values. Under the null hypothesis of no association, this statistic approximately follows the chi-square distribution with 4 degrees of freedom. Note that the Fisher's method is always valid regardless of whether the directions of two effect sizes are consistent.

	Discovery Sample (Caucasian)		Replication Sample (Caucasian)	
	Male	Female	Male	Female
No. of subjects	501	499	556	1727
Age	50.65 (18.60)	50.00 (17.90)	50.71 (16.05)	51.59 (12.92)
Height (m)	1.77 (0.07)	1.64 (0.06)	1.71 (0.07)	1.62 (0.03)
Weight (kg)	89.20 (14.90)	71.39 (16.10)	87.08 (16.70)	71.46 (16.00)
Fat body mass (kg)	23.46 (8.88)	26.92 (10.33)	20.67 (9.30)	20.92 (13.20)
Lean body mass (kg)	63.67 (8.22)	43.49 (6.69)	66.60 (9.59)	46.85 (7.04)
LMI	2.01 (0.22)	1.62 (0.23)	2.15 (1.04)	1.75 (1.31)

Table 1. Basic characteristics of the study subjects.

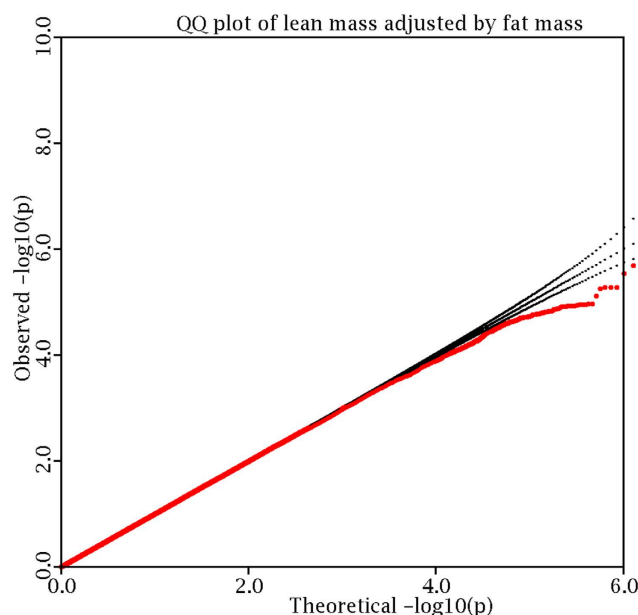


Figure 1. Logarithmic quantile–quantile (QQ) plot of individual SNP-based association.

Gene-level effect direction evaluation. To evaluate the consistency of two gene-level effect directions, we studied the effects of all individual SNPs within the gene and proposed an overall gene-level effect direction measure. We then compared the directions of the two measures. To accomplish this, we set all the SNP effects in the discovery sample to be positive so that a gene-level z-score is estimated by $z_{discovery} = \frac{\sum_{i=1}^m z_i^{(1)}}{m}$, where $z_i^{(1)}$ is the positive z-score of the i th SNP in the discovery sample. If the reported SNP z-score is negative, then the reference allele and the alternative allele will exchange so that the z-score changes to be positive. For example, if the reference allele and the alternative allele are A and G, and the reported z-score of allele A is -1.0 , then the allele G, whose z-score is 1.0 , will change to the reference allele.

Once all reference alleles are determined by referring to the discovery sample, we will calculate a gene-level z-score for the replication sample as $z_{replication} = \frac{\sum_{i=1}^m z_i^{(2)}}{m}$, where $z_i^{(2)}$ is the z-score of the reference allele at the i th SNP. If the $z_{replication}$ is positive, then it has the same direction as $z_{discovery}$, and we declare that the two genes are consistent in effect direction; otherwise, they are opposite in effect direction.

Results

The basic characteristics of the subjects used in both discovery and replication samples are summarized in Table 1.

Discovery sample. Using the genotyped and imputed genotypes in our sample of 1000 Caucasian subjects with GWAS data, we tested SNPs for association with LMI. Overall genomic control inflation factor was 1.05. To correct for potential population stratification, we adjusted individual p-values by the inflation factor. The logarithmic quantile–quantile (QQ) plot of SNP-based association results is displayed in Fig. 1. No evidence of population stratification is observed after adjustment by the genomic control approach. Manhattan plot of association results across the genome is displayed in Fig. 2. There is no genome-wide significant association signal.

Through individual SNP-based tests, we performed gene-based tests using VEGAS software. A total of 17,511 genes are included in the gene-based association test in the discovery sample. The Bonferroni correction is used to declare the genome-wide significance level (GWS, $0.05/17,511 = 2.86 \times 10^{-6}$). The QQ plot of gene-based

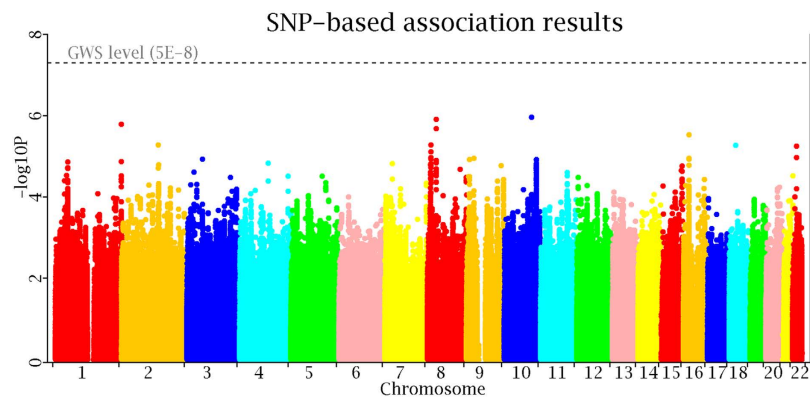


Figure 2. Manhattan plot of individual SNP-based association tests.

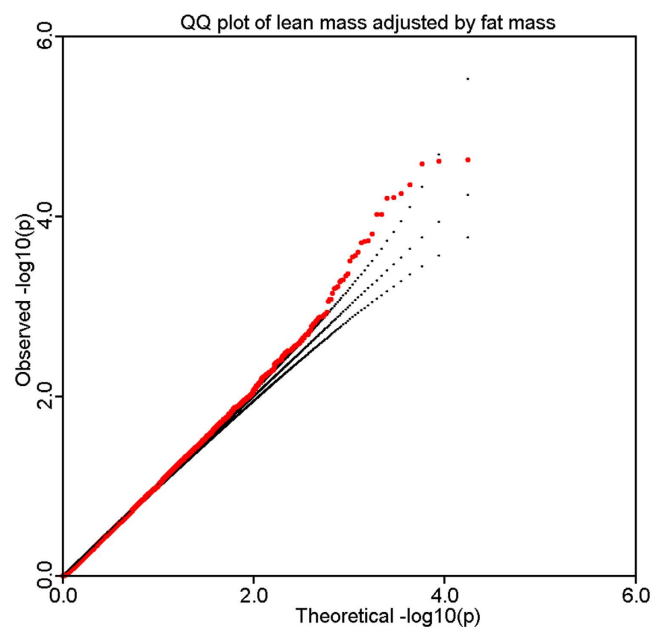


Figure 3. QQ plot of gene-based association tests.

association results is displayed in Fig. 3. A marked deviation is observed in the tail of the distribution, implying true association signals. The Manhattan plot of gene-based association results is displayed in Fig. 4.

None of the genes is significant at the GWS level. At a less stringent borderline level 2×10^{-4} , there are 11 genes that are significant. The most significant gene is *GATA4* at 8p23.1 ($p = 2.40 \times 10^{-5}$). These 11 genes are located into 6 distinct genomic regions.

Replication sample. The 11 suggestively associated genes in the discovery sample are subjected to replication in the replication sample. SNP-based and gene-based association analyses in the replication sample are same to those in the discovery sample. Only three genes from the same locus 19p13.3 (*UQCR* $p = 2.21 \times 10^{-3}$, *TCF3* $p = 1.84 \times 10^{-3}$, *MBD3* $p = 6.95 \times 10^{-3}$), while the other 8 genes are not significant.

In the meta-analysis of both the discovery and the replication samples by the Fisher's method, the signal of *UQCR* achieves the GWS level (discovery $p = 6.10 \times 10^{-5}$, replication $p = 2.21 \times 10^{-3}$, combined $p = 2.26 \times 10^{-6}$), while the signals of *TCF3* and *MBD3* are close to the GWS level (*TCF3* combined $p = 4.86 \times 10^{-6}$, *MBD3* combined $p = 1.15 \times 10^{-5}$).

Combining the evidence from both the discovery and the replication samples, three genes (*UQCR*, *TCF3* and *MBD3*) from one single locus 19p13.3 are convincingly associated with LMI after adjustment by fat mass. The main results are listed in Table 2.

These three genes are apart 170 kb at most. Their tested regions overlap with each other. SNPs in this locus are divided into several haplotype blocks. The regional LD structure and association signals are displayed in Fig. 5.

Gene-level effect direction evaluation. We evaluated the consistency of gene-level effect directions between the discovery and the replication samples. We proposed a gene-level effect direction measure in each study and then compared the two measures. The results are listed in Table 2. The z-scores in the discovery sample

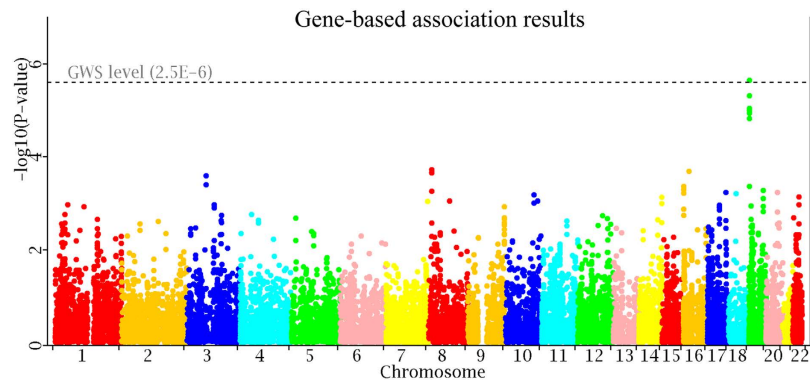


Figure 4. Manhattan plot of gene-based association tests.

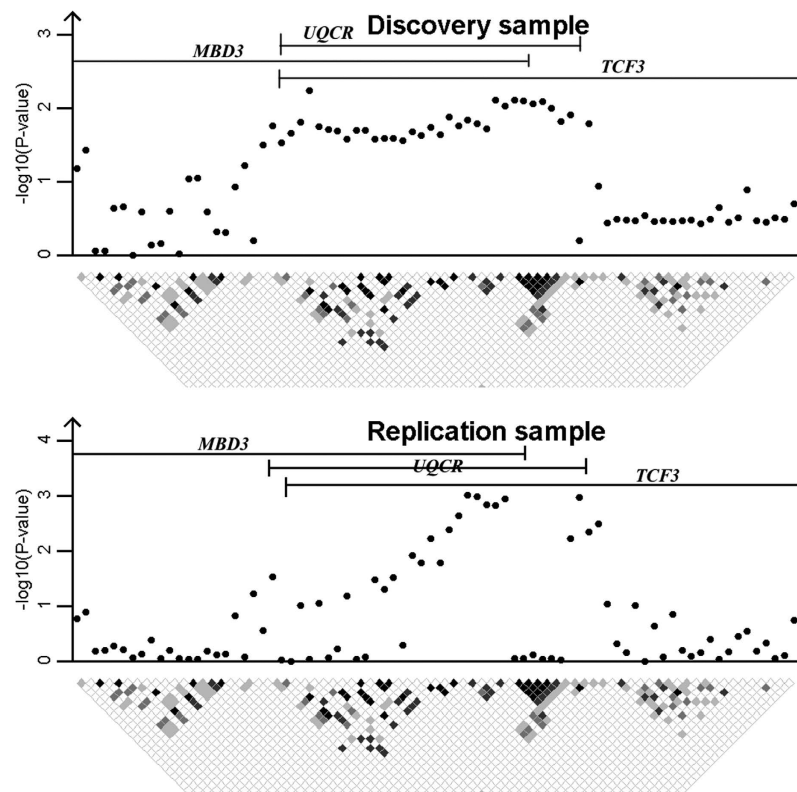


Figure 5. Regional plot of SNP associations around 19p13.3.

Gene	Chr	Start	Stop	No. SNPs	Discovery		Replication		Combined
					z	p	z	p	p
<i>UQCR</i>	19	1548170	1556431	30	2.36	6.10×10^{-5}	0.30	2.21×10^{-3}	2.26×10^{-6}
<i>TCF3</i>	19	1560294	1601277	51	1.84	1.65×10^{-4}	0.41	1.84×10^{-3}	4.86×10^{-6}
<i>MBD3</i>	19	1527677	1543652	47	1.82	1.10×10^{-4}	0.18	6.95×10^{-3}	1.15×10^{-5}

Table 2. Genes identified for LMI by gene-based association test. Notes: Z-score was the average z-scores over all included SNPs. z-score at each SNP was referred to the increasing-allele in the discovery sample whose z-score was positive by definition. Genome-wide significance (GWS) level significant result was marked in bold.

are always positive by definition. It is clear from the table that the z-scores in the replication sample are also positive in all cases, implying that the two studies are consistent in effect direction for all the three evaluated genes.

Discussion

In this study, we have performed a genome-wide association study in 1000 unrelated Caucasian subjects and replicated in 2283 unrelated Caucasian subjects. Combining evidence from both the discovery sample and the replication sample, we have identified three genes *UQCR*, *TCF3* and *MBD3* located at 19p13.3 that were associated with LMI.

Gene density at chromosome 19 is the highest in all the human chromosomes, more than double of the genome-wide average level. The densely located gene families of repetitive DNA sequences implies it being the chromosome of biological and evolutionary significance^{25,26}.

The product of the *UQCR* gene is the ubiquinol-cytochrome c reductase complex, also called mitochondrial complex III. It functions to form a part of the mitochondrial respiratory chain. It may also act as a binding factor for the iron-sulfur protein. Mitochondrial Complex III is composed of one mitochondrial-encoded subunit (MT-CYB) and ten nuclear-encoded subunits. The complex is located within the mitochondrial inner membrane and plays an important role in biochemical synthesis of ATP. It functions to catalyze electrons to transfer from succinate and nicotinamide adenine dinucleotide linked dehydrogenases to mitochondrially encoded cytochrome b. It also functions to utilize the energy to translocate protons across the membrane²⁷. Deficiency of isolated complex III has been detected in patients of neuromuscular and nonneuromuscular disorders in both children and adults²⁸.

TCF3 (Transcription factor 3) gene encodes two basic helix-loop-helix (HLH) transcription factors E12 (synonyms–Immunoglobulin Transcription Factor-1 (ITF1)) and E47 (Transcription Factor-3 (TCF-3))²⁹. These two factors are believed to function in the regulation of cell growth, differentiation of muscle cells, and commitment. The mechanism is that they act as dimerisation partners for lineage-specific HLH proteins such as the myogenic factors myoblast determination (*Myod*) and myogenin (*Myog*)³⁰.

The processes of myogenic specification and differentiation are regulated by a family of myogenic basic HLH transcription factors that include *Myod*, *Myog*, myogenic factor 5 (*Myf5*), and myogenic factor 6 (*Myf6*) in the human³¹. These heterodimers yield a sophisticated pathway (Supplemental Fig. 1) that leads to the expression of critical genes that are specific to muscle. They also promote precursor cells (myoblasts) to develop into multinucleated and differentiated muscle cells (myotubes)³².

The *MBD3* (methyl-binding 3) gene encoded protein is a subunit of the NuRD, a multi-subunit complex comprised of *Hdac1* (histone deacetylase1), *Hdac2*, *Rbbp7* and *Rbbp4* containing nucleosome remodeling and histone deacetylase activities³³. The class I histone deacetylase *Hdac1* is directed associated with *MyoD*. It is capable of deacetylating *MyoD* and blocking the function of the latter from initiating the myogenic program under differentiation conditions³⁴. In differentiating muscle cells, the reduced level of *HDAC1* is associated with increased level of *MyoD* acetylation, which stimulates transcription³⁵.

Despite their potential relevance to lean mass development, it is unlikely that all the three identified genes are causal. This is because these genes are identified through statistical association instead of biological validation. Statistical association is affected by LD structure, and is usually called indirect gene mapping. These three genes have overlapping test regions, and they present strong levels of LD.

To further explore their potentially causal role, we used a recently developed method PrediXcan to analyze the three identified genes³⁶. In brief, the PrediXcan method imputes the unobserved gene expression levels of the GWAS samples, and examines the association between imputed gene expressions and the trait. Unfortunately, none of the three identified genes is significant by the PrediXcan method ($p > 0.05$). We recognized that the PrediXcan method relies on the eQTL effect of SNPs to the analyzed gene. To evaluate whether the eQTL effect is present, we searched the GTEx datasets through Haploreg³⁷, and found that the eQTL effects are very limited for analyzed SNPs. For example, of the 34 SNPs used to calculate the *UQCR* gene signal, none has eQTL activity to its expression. The determination of causal genes are still awaiting further functional studies.

In summary, we identified the 19p13.3 region containing *UQCR*, *TCF3* and *MBD3* genes, which were significantly associated with LMI in the Caucasian population. These results, together with the known functions of the three genes related to LBM, support that the 19p13.3 region containing *UQCR*, *TCF3* and *MBD3* genes is a novel locus underlying human LBM.

References

1. Sipila, S. *et al.* Endogenous hormones, muscle strength, and risk of fall-related fractures in older women. *J Gerontol a-Biol* **61**, 92–96 (2006).
2. Karakelides, H. & Nair, K. S. Sarcopenia of aging and its metabolic impact. *Curr Top Dev Biol* **68**, 123–148 (2005).
3. Hsu, F. C. *et al.* Heritability of body composition measured by DXA in the diabetes heart study. *Obesity Research* **13**, 312–319 (2005).
4. Nguyen, T. V., Howard, G. M., Kelly, P. J. & Eisman, J. A. Bone mass, lean mass, and fat mass: Same genes or same environments? *Am J Epidemiol* **147**, 3–16 (1998).
5. Korostishevsky, M. *et al.* Genomics and metabolomics of muscular mass in a community-based sample of UK females. *Eur J Hum Genet* **24**, 277–283 (2016).
6. Urano, T., Shiraki, M., Sasaki, N., Ouchi, Y. & Inoue, S. Large-scale analysis reveals a functional single-nucleotide polymorphism in the 5'-flanking region of PRDM16 gene associated with lean body mass. *Aging Cell* **13**, 739–743 (2014).
7. Patel, H. P. *et al.* Lean Mass, Muscle Strength and Gene Expression in Community Dwelling Older Men: Findings from the Hertfordshire Sarcopenia Study (HSS). *Calcified Tissue Int* **95**, 308–316 (2014).
8. Baumgartner, R. N. *et al.* Epidemiology of sarcopenia among the elderly in New Mexico. *Am J Epidemiol* **147**, 755–763 (1998).
9. Cruz-Jentoft, A. J. *et al.* Sarcopenia: European consensus on definition and diagnosis. *Age Ageing* **39**, 412–423 (2010).
10. Li, M. X., Gui, H. S., Kwan, J. S. H. & Sham, P. C. GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *Am J Hum Genet* **88**, 283–293 (2011).
11. Huang, H. L., Chanda, P., Alonso, A., Bader, J. S. & Arking, D. E. Gene-Based Tests of Association. *Plos Genet* **7** (2011).

12. Liu, J. Z. *et al.* A Versatile Gene-Based Test for Genome-wide Association Studies. *Am J Hum Genet* **87**, 139–145 (2010).
13. Deng, H. W. *et al.* A genomewide linkage scan for quantitative-trait loci for obesity phenotypes. *Am J Hum Genet* **70**, 1138–1151 (2002).
14. Wang, J. *et al.* Independent associations of body-size adjusted fat mass and fat-free mass with the metabolic syndrome in Chinese. *Ann Hum Biol* **36**, 110–121 (2009).
15. Liu, Y. J. *et al.* Genome-wide association scans identified CTNBL1 as a novel gene for obesity. *Hum Mol Genet* **17**, 1803–1813 (2008).
16. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906–913 (2007).
17. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955–+ (2012).
18. Zhu, X., Li, S., Cooper, R. S. & Elston, R. C. A unified association analysis approach for family and unrelated samples correcting for stratification. *American journal of human genetics* **82**, 352–365 (2008).
19. Liang, J. J. *et al.* Comparison of Heritability Estimation and Linkage Analysis for Multiple Traits Using Principal Component Analyses. *Genet Epidemiol* **40**, 222–232 (2016).
20. Chen, X. *et al.* Pathway-Based Analysis for Genome-Wide Association Studies Using Supervised Principal Components. *Genet Epidemiol* **34**, 716–724 (2010).
21. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. *Annu Rev Genom Hum G* **10**, 387–406 (2009).
22. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: Using Sequence and Genotype Data to Estimate Haplotypes and Unobserved Genotypes. *Genet Epidemiol* **34**, 816–834 (2010).
23. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
24. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155–166 (2001).
25. Venter, J. C. The sequence of the human genome (vol 292, pg 1304, 2001). *Science* **292**, 1838–1838 (2001).
26. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
27. Fernandez-Vizarrá, E. *et al.* Impaired complex III assembly associated with BCS1L gene mutations in isolated mitochondrial encephalopathy. *Hum Mol Genet* **16**, 1241–1252 (2007).
28. Visapaa, I. *et al.* GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am J Hum Genet* **71**, 863–876 (2002).
29. Slattery, C., Ryan, M. P. & McMorro, T. E2A proteins: Regulators of cell phenotype in normal physiology and disease. *Int J Biochem Cell B* **40**, 1431–1436 (2008).
30. Massari, M. E. & Murre, C. Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**, 429–440 (2000).
31. Krauss, R. S. *et al.* Close encounters: regulation of vertebrate skeletal myogenesis by cell-cell contact. *J Cell Sci* **118**, 2355–2362 (2005).
32. Arnold, H. H. & Winter, B. Muscle differentiation: more complexity to the network of myogenic regulators. *Curr Opin Genet Dev* **8**, 539–544 (1998).
33. Kaji, K., Nichols, J. & Hendrich, B. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134**, 1123–1132 (2007).
34. Mal, A., Sturniolo, M., Schiltz, R. L., Ghosh, M. K. & Harter, M. L. A role for histone deacetylase HDAC1 in modulating the transcriptional activity of MyoD: inhibition of the myogenic program. *Embo J* **20**, 1739–1753 (2001).
35. Jo, C., Jang, B. G. & Jo, S. A. MEK1 plays contrary stage-specific roles in skeletal myogenic differentiation. *Cell Signal* **21**, 1910–1917 (2009).
36. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091–1098 (2015).
37. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930–934 (2012).

Acknowledgements

The study was partially supported by startup fund from Shanghai University of Science and Technology and Shanghai Leading Academic Discipline Project (S30501). The investigators of this work were partially supported by grants from NIH (R01AG026564, RC2DE020756, R01AR057049, R01AR050496 and R03TW008221), a SCOR (Specialized Center of Research) grant (P50AR055081) supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and the Office of Research on Women's Health (ORWH), the Edward G. Schlieder Endowment and the Franklin D. Dickson/Missouri Endowment, the National Natural Science Foundation of China (31571291 to L.Z., 31501026 to Y.F.P., 31301092 to Y.L.), the Natural Science Foundation of Jiangsu Province (BK20150323 to Y.F.P.), the undergraduate innovation program of Jiangsu province (201610285040Z to A.P.F.) and a project of the priority academic program development of Jiangsu higher education institutions.

Author Contributions

Conceived and designed the experiments: H.W.D. Performed the experiments: S.R., L.Z., Y.F.P. and Y.Y.H. Analyzed the data: Y.F.P., L.Z., L.Z., L.L., X.L., A.P.F., W.W.K., X.Y.Y., W.Z., Y.L., Q.T., H.S. and Y.H.Z. Literature search: L.L., X.L., A.P.F., W.W.K., X.Y.Y. and W.Z. Wrote the paper: S.R., L.Z. and H.W.D. All authors reviewed and approved the manuscript.

Additional Information

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Ran, S. *et al.* Gene-based genome-wide association study identified 19p13.3 for lean body mass. *Sci. Rep.* **7**, 45025; doi: 10.1038/srep45025 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017