

## ORIGINAL ARTICLE

# Genome-wide association study of copy number variation identified *gremlin1* as a candidate gene for lean body mass

Rong Hai<sup>1,2</sup>, Yu-Fang Pei<sup>1</sup>, Hui Shen<sup>3</sup>, Lei Zhang<sup>1</sup>, Xiao-Gang Liu<sup>4</sup>, Yong Lin<sup>1</sup>, Shu Ran<sup>1</sup>, Feng Pan<sup>4</sup>, Li-Jun Tan<sup>5</sup>, Shu-Feng Lei<sup>5</sup>, Tie-Lin Yang<sup>4</sup>, Yan Zhang<sup>1</sup>, Xue-Zhen Zhu<sup>1</sup>, Lan-Juan Zhao<sup>3</sup> and Hong-Wen Deng<sup>1,3,5</sup>

Lean body mass (LBM) is a heritable trait predicting a series of health problems, such as osteoporotic fracture and sarcopenia. We aim to identify sequence variants associated with LBM by a genome-wide association study (GWAS) of copy number variants (CNVs). We genotyped genome-wide CNVs of 1627 individuals of the Chinese population with Affymetrix SNP6.0 genotyping platform, which comprised of 940 000 copy number probes. We then performed a GWAS of CNVs with lean mass at seven sites: left and right arms, left and right legs, total of limb, trunk and whole body. We identified a CNV that is associated with LBM variation at the genome-wide significance level (CNV2073, Bonferroni corrected *P*-value 0.002 at right arm). CNV2073 locates at chromosome 15q13.3, which has been implicated as a candidate region for LBM by our previous linkage studies. The nearest gene, *gremlin1*, has a key role in the regulation of skeletal muscle formation and repair. Our results suggest that the *gremlin1* gene is a potentially important gene for LBM variation. Our findings also show the utility and efficacy of CNV as genetic markers in association studies.

*Journal of Human Genetics* (2012) 57, 33–37; doi:10.1038/jhg.2011.125; published online 3 November 2011

**Keywords:** association; copy number variation; *gremlin1* gene; lean body mass; 15q13.3

## INTRODUCTION

Loss and function impairment of skeletal muscle is a common skeletal disorder affecting millions of people worldwide, especially in the elderly. Its most severe outcome is to predispose people to sarcopenia. It is also related to a series of other diseases and health problems, such as osteoporosis (MIM 166710), fracture, impaired protein balance, dyslipidemia (MIM 151660), obesity (MIM 601665), insulin resistance, overall frailty and increased mortality.<sup>1,2</sup> Skeletal muscle is characterized by the measurement of lean body mass (LBM), which is the single best predictor for sarcopenia. LBM is highly inheritable, with estimated heritability ranging from 52 to 84%.<sup>3–5</sup> However, only a few genes for LBM have so far emerged,<sup>6</sup> leaving the majority of the genetic background of LBM still unknown.

Traditional association analysis has focused largely on single nucleotide polymorphisms (SNPs). This assessment of SNP variation has proven fruitful; hundreds of common variants have been found to be associated with diseases such as obesity, osteoporosis, type 2 diabetes and immunological disease.<sup>7</sup> However, recent studies have shown that another type of genomic variation, copy number variations (CNVs), has a significant role in influencing common

diseases as well, and are likely to be at reasonably high frequencies in the population. Recent data imply that CNVs account for up to 4 Mb of genetic differences, whereas that for SNP variation is only 2.5 Mb.<sup>8</sup> The widespread distribution of CNVs across the genome has made it an important type of genetic variation for identifying disease-associated genetic loci. Many diseases are found to be associated with copy number (CN) changes, including osteoporosis,<sup>9</sup> lupus glomerulonephritis,<sup>10</sup> autism,<sup>11</sup> and HIV infection and progression.<sup>12</sup> Therefore, investigation of CNVs would contribute to unravel the genetic basis of complex diseases and phenotypes. Nonetheless, to the best of our knowledge, there is no CNV-aimed association study on LBM reported. It is largely unknown whether CNV underlies the variation of LBM. In this study, we report a CNV-based genome-wide association study (GWAS) to identify genetic loci influencing LBM variation.

## MATERIALS AND METHODS

### Study subjects

The study sample consisted of 1627 (802 males and 825 females) unrelated Chinese-Han subjects living in the cities of Xi'an/Changsha and their neighboring areas. The study was approved by the local institutional review

<sup>1</sup>Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai, PR China; <sup>2</sup>The Affiliated hospital of Inner Mongolia medical College, Hohhot, PR China; <sup>3</sup>Department of Biostatistics, Tulane University, New Orleans, LA, USA; <sup>4</sup>The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, PR China and <sup>5</sup>Laboratory of Molecular and Statistical Genetics and Key Laboratory of Protein Chemistry and Developmental Biology of Ministry of Education, College of Life Sciences, Hunan Normal University, Changsha, PR China

Correspondence: Professor H-W Deng, Center of System Biomedical Sciences, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai, Shanghai 200093, PR China.

E-mail: hdeng2@tulane.edu

Received 7 June 2011; revised 21 September 2011; accepted 9 October 2011; published online 3 November 2011

board. After signing an informed consent, all subjects received assistance in completing a structured questionnaire including questions about anthropometric variables, lifestyle, diet, family information and medical history, and so on.

### Phenotyping

The cohort was recruited for studies aimed in searching for genes underlying body compositions (bone mass, fat mass and lean mass). Body composition was measured using a dual-energy X-ray absorptiometry scanner Hologic QDR 4500W (Hologic Inc., Bedford, MA, USA), following the manufacturer's protocol. A dual-energy X-ray absorptiometry scan can accurately measure total body and regional bone mass, fat mass and fat-free mass. Lean mass is calculated by taking bone mass away from fat-free mass.<sup>13,14</sup> After removal of all metals, a subject laid on a bed and was scanned from head to toe. Whole body composition, body compositions at sub-regions, such as head, trunk and limb, were measured by the dual-energy X-ray absorptiometry scanner.

To ensure the quality of collected data, all scans were conducted, reviewed and analyzed by a clinical expert. Body weight, height and age were obtained on the same visit. In this study, lean mass at four limbs, trunk and whole body were analyzed as main phenotypes.

### Genome-wide genotyping and quality controls (QC)

Genomic DNA was extracted from peripheral blood leukocytes using standard protocols. Genome-Wide Human SNP Array 6.0 (Affymetrix Inc., Santa Clara, CA, USA), which includes 906 600 SNPs and 940 000 CN probes, was used to genotype each subject, according to the Affymetrix protocol. Briefly, ~250 ng of genomic DNA was digested with restriction enzyme NspI and StyI. Digested DNA was adaptor ligated and PCR amplified for each sample. Fragment PCR products were then labeled with biotin, denatured and hybridized to the arrays. Arrays were then washed and stained using phycoerythrin on Affymetrix Fluidics Station, and scanned using the GeneChip Scanner 30007G to quantify fluorescence intensities (Affymetrix Inc.). Data management and analyses were conducted using the Affymetrix Genotyping Command Console. The Affymetrix contrast QC threshold was set at the default value of greater than 0.4 for sample QC. The final average contrast QC across the entire sample reached a high level of 2.62.

### Assessment of genetic background

The method of genomic control implemented in the STRUCTURE2.2 program<sup>15</sup> was used to detect possible population stratification of the study sample. For structure analysis, 2000 SNPs were randomly selected at the genome for clustering of all the subjects. The program uses a Markov chain Monte Carlo algorithm to cluster individuals into different cryptic subpopulations based on multilocus genotype data. Potential substructure was estimated under *a priori* assumption of  $K=2$  discrete subpopulations. To cross-validate the results, we also conducted principal component analysis on selected genotypes using EIGENSTRAT.<sup>16</sup> The calculated principal components are informative to correct for potential population stratification in subsequent association analyses.

### CNV determination

CNVs were identified using the CANARY algorithm implemented in the Birdsuite software (Affymetrix Inc.),<sup>17</sup> which utilized a previously defined CNV map based on HapMap samples.<sup>17</sup> In order to generate results with high confidence, we conducted QC filtering both at the sample level and the CNV level, according to the previously reported methods.<sup>17</sup>

First, for the sample level QC, we used three quality metrics reported by the Birdseye method to evaluate the initial 1627 subjects for quality in CN genotyping. The following procedures were adopted: (1) we removed any sample that was greater or less than three s.d. values from the average estimate of CN, which was approximate two copies at genome-wide level; (2) we calculated the variability in CN and SNP probe intensities with each standardized per chromosome. We removed any sample with three s.d. values more than these estimates on the average genome-wide level; (3) we removed any sample in which more than two chromosomes failed any of these three metrics, that is, more than three values in estimated CN or excessive CNV or SNP variability for the chromosome.

Second, we conducted QC filtering at the CNV level. Out of the initial 1280 CNVs, we discarded (1) any CNVs in which more than 5% of the copy calls were uncertain (confidence score > 0.1) or missing, and (2) any CNVs with the frequency of major variant greater than 99%. The filtering procedure resulted in 603 CNVs available for subsequent association analyses.

### Statistical analyses

Lean mass at the following seven sites was analyzed: left and right arms, left and right legs, subtotal of limbs, trunk and whole body. Each phenotype was adjusted by age, gender, and the first two principal components calculated from the 2000 selected SNPs. Residual phenotypes were normalized by inverse quantile of the standard normal distribution, which imposes a standard normal distribution on the phenotype to be analyzed. Covariate adjustment and phenotype normalization were performed with Minitab (Minitab Inc., State College, PA, USA).

Association of lean mass and CNV was performed by a linear regression model using PLINK.<sup>18</sup> In brief, CNVs were treated as predictors for lean mass. The PLINK input genotype file set includes three files: a family file, a map file and a gvar file. The family file and map file describe individuals and variants, and gvar describes the paternal and maternal origins of the derived CNs. Each row in a family file represents one subject with the following fields separated by a tab: *family id, subject id, father id, mother id, sex and phenotype*. Each row in a map file represents one variant with the following fields: *chromosome, CNV id, genetic distance and start physical position*. Each row in a gvar file has seven fields: *family id, subject id, CNV id, first CN, first dosage, second CN and second dosage*. Here, *first CN* and *first dosage* are the allele inherited from the first parent and its dosage, and the same for the second ones. The command implemented the association test is

```
plink - gfile filesetName - pheno phenoFile - pheno - name phenoName
      - linear - out outFile
```

The output file lists *P*-values for all CNVs.

We adopted the strict Bonferroni correction to account for multiple testing comparisons. Raw *P*-values were adjusted by the product of the number of CNVs (603) and the number of phenotypes (7). Significant results were declared at nominal level 0.05 after correction, corresponding to the genome-wide significance level 1.18E-5.

## RESULTS

Basic characteristics of the sample are summarized in Table 1. The STRUCTURE<sup>8</sup> program clustered all subjects into one single homogeneous population (see Supplementary Figure S1). The estimated inflation factor ( $\lambda$ ) from association analyses is 1.02, below the level of typical deviation for population stratification. All these results indicate that population stratification is not likely to present in the studied sample.

There is one CNV, CNV2073, which hits the genome-wide significance level with raw *P*-value 6.22E-7 (Bonferroni corrected

**Table 1 Basic characteristics of the study sample**

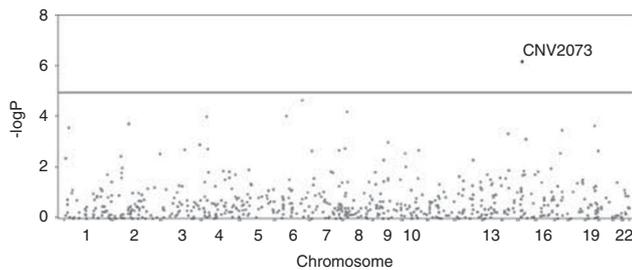
Trait	Total (N=1627)	Male (N=802)	Female (N=825)
Age (year)	34.49 (13.24)	31.43 (11.93)	37.46 (13.77)
Height (cm)	164.25 (8.16)	170.27 (5.96)	158.38 (5.22)
Weight (kg)	65.72 (9.61)	65.74 (9.64)	54.63 (8.09)
W-body-lean (kg)	45.29 (9.26)	52.83 (5.95)	37.95 (4.13)
Trunk-lean (kg)	21.33 (3.98)	24.44 (2.89)	18.31 (2.14)
L-leg-lean (kg)	7.39 (1.72)	8.80 (1.19)	6.01 (0.79)
R-leg-lean (kg)	7.49 (1.76)	8.94 (1.21)	6.09 (0.81)
L-arm-lean (kg)	2.35 (0.75)	2.98 (0.51)	1.74 (0.31)
R-arm-lean (kg)	2.55 (0.81)	3.22 (0.55)	1.91 (0.37)
Subtotal-lean (kg)	41.12 (8.70)	48.38 (5.77)	34.06 (3.97)

Abbreviations: L, left; R, right; W, whole.

Note: data presented are unadjusted means (s.d.).

$P$ -value=0.002) for lean mass at the right arm (R-arm). Figure 1 displays the Manhattan plot of genome-wide scan for this phenotype. The association of this CNV and lean mass is also nominally significant at most other sites, though none of them achieves a genome-wide significance level (Table 2). Table 2 also lists the association results of the top 10 CNVs ranked according to  $P$ -values for lean mass at R-arm.

CNV2073 locates from 283 77 089 to 285 36 721 bp (NCBI build 36.3) at chromosome 15q13.3. Three types of CN exist in the sample: CN=2, 3, and 4, with frequencies 0.06, 0.13 and 0.81, respectively. Compared with subjects with two CNs (normal diploid), subjects with three copies had 6.9% lower mean lean mass at R-arm, and subjects with four copies had 11.2% lower lean mass at R-arm (Figure 2). Linear regression analysis showed that CNV15q13.3 contributed to 1.0% of the total lean mass variance at R-arm.

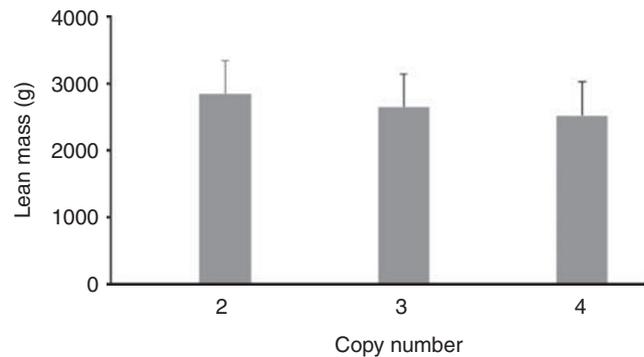


**Figure 1** Manhattan plot of lean mass at R-arm. The y axis represents  $-\log_{10}P$ , and x axis represents the start physical position of CNV along chromosomes. The plot displays  $P$ -values of lean mass at right arm for all 603 CNVs. The line shows the threshold for genome-wide significance level. The figure shows that CNV2073 is significant at genome-wide level  $1.18E-5$ .

Two genes, *gremlin1* and *chrfa7a*, locate in the region 15q13.3, which is covered by CNV2073. Of them, *gremlin1* is of particular interest. It is a candidate gene for LBM reported by both molecular function study<sup>19-21</sup> and previous genetic linkage studies.<sup>22,23</sup> However, none of previous association studies has linked this gene to LBM variation yet.

## DISCUSSION

Lean mass has a considerable heritability. Although previous SNP-aimed association studies have identified several candidate genes,<sup>6</sup> vast majority of genetic mechanisms of lean mass remains unclear, which may reside in CNVs. To the best of our knowledge, this is the first GWAS between lean mass and CNVs in the Chinese population.



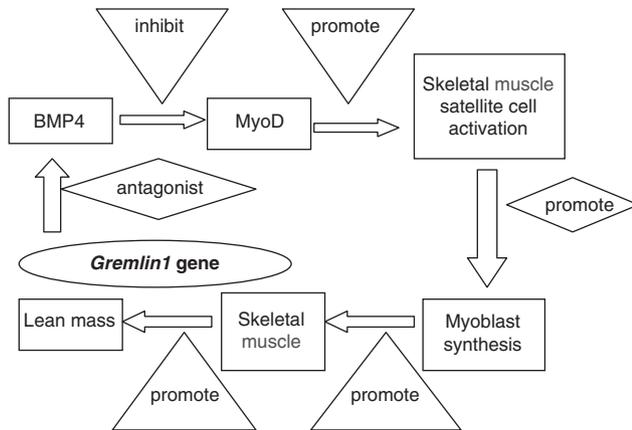
**Figure 2** Mean R-arm lean mass values at different copy numbers for CNV2073. Copy numbers of 2, 3 and 4 exist for CNV2073. The y axis represents the mean lean mass at the right arm for a particular copy number. Error bars denote standard error.

**Table 2**  $P$ -values of the top 10 CNVs

CNV	CHR	Start	End	Freq	$P$ -value						
					R-arm	L-arm	R-leg	L-leg	Subtotal	Trunk	W-body
CNV 2073	15	28377089	28536721	0.19	6.22E-07	6.53E-06	3.74E-04	3.83E-04	1.19E-03	0.12	3.85E-03
					<b>2.62E-03</b>	<b>0.03</b>	1	1	1	1	1
CNV 1010	6	134592448	134598896	0.12	2.00E-05	4.31E-05	0.01	0.01	4.79E-03	0.03	7.82E-03
					0.08	0.18	1	1	1	1	1
CNV 1253	8	12260380	12286526	0.4	5.81E-05	7.78E-04	0.01	0.1	0.07	0.32	0.08
					0.25	1	1	1	1	1	1
CNV 944	6	49039688	49043159	0.19	8.72E-05	3.64E-05	3.97E-04	0.01	0.03	0.36	0.04
					0.37	0.15	1	1	1	1	1
CNV 10674	4	18697657	18733331	0.06	9.24E-05	2.86E-05	0.08	0.29	0.06	0.34	0.11
					0.39	0.12	1	1	1	1	1
CNV 243	2	79184286	79195607	0.03	1.71E-04	5.37E-05	0.3	0.36	0.02	0.86	0.02
					0.72	0.22	1	1	1	1	1
CNV 2406	19	40541333	40553688	0.44	1.96E-04	1.17E-04	0.01	0.01	9.25E-03	0.03	7.73E-03
					0.82	0.49	1	1	1	1	1
CNV 22	1	17067742	17134834	0.36	2.33E-04	2.91E-04	0.02	0.01	5.02E-03	0.09	0.02
					0.98	1	1	1	1	1	1
CNV 2267	17	41006741	41015665	0.46	2.91E-04	3.26E-03	0.03	0.03	8.68E-03	0.09	0.03
					1	1	1	1	1	1	1
CNV 12158	14	40541848	40555277	0.01	4.24E-04	1.66E-05	1.51E-03	3.45E-04	3.03E-03	0.04	3.86E-03
					1	0.07	1	1	1	1	1

Abbreviations: CHR, chromosome; CNV, copy number variant; Freq, frequency of the least frequent copy number; L, left; R, right; w, whole.

$P$ -value was estimated by regression analysis. For each CNV,  $P$ -values at the first row are raw  $P$ -values and that at the second row are Bonferroni corrected ones; significant results after correction are marked in bold.



**Figure 3** Hypothesized functional mechanism of *gremlin1* to lean mass. In this plot, *gremlin1* antagonizes the activity of bone morphogenetic protein 4 (BMP4), which inhibits the expression of MyoD. MyoD is a protein with a key role in upregulating muscle differentiation by stimulating the activity of skeletal muscle satellite cells. As a result, the synthesis of myoblast gets activated, and the levels of skeletal muscle and lean mass increase.

We identified a candidate genomic region 15q13.3 and an associated gene *gremlin1* at the genome-wide significance level. Notably, this region was also indicated to be important for lean mass variation in our two previous linkage studies.<sup>22,23</sup> The first study performed a large-scale whole genome linkage scan for lean mass involving 4498 individuals from 451 Caucasian families. The most pronounced linkage signal was found at 15q13.3 with the LOD score 4.86.<sup>22</sup> The second genome-wide linkage scan of 434 Caucasian pedigrees gave a suggestive linkage signal at this region with the LOD score 2.72.<sup>23</sup> The current study first implied that this region is also associated with lean mass in the Chinese population.

The associated gene *gremlin1* and CNV2073 are about 2 MB apart. The gene is a member of the bone morphogenetic protein (BMP) antagonistic family. It was first cloned from a Xenopus ovarian library for its axial patterning activities. The human *gremlin1* gene encodes for a glycosylated homodimeric peptide of 28 kDa with 184 amino acids.<sup>24</sup> As an antagonist to BMP, the regulation of *gremlin1* is essential for mesoderm induction, establishment of dorsoventral polarity, ectodermal differentiation, somite formation and myogenesis induction.<sup>24</sup>

It is well known that the reduction of lean mass with aging is caused by the atrophy of type II myofiber.<sup>25,26</sup> The capacity of generating new myonuclei for myofiber repair, growth, or replacement is dependent upon the persistence of skeletal satellite cells.<sup>27,28</sup> The proliferation and differentiation of skeletal satellite cell is activated by the expression of the myogenin regulatory factor MyoD, which, in turn, is down-regulated by BMP4. *Gremlin1*, as an antagonist to BMP4, therefore promotes the expression of MyoD, and the generation and repair of lean mass.<sup>29,30</sup> Figure 3 illustrate the regulation pathways of the effect of *gremlin1* on lean mass.

With the same sample pool and analytical approach, we performed another CNV-based GWAS previously.<sup>9</sup> The identified CNVs were successfully validated by real-time PCR in that study. Therefore, though the identified CNV was not further validated by real-time PCR in this study, it may still be highly reliable.

In conclusion, we have conducted a GWAS between CNVs and lean mass in the Chinese population, and identified a significant gene *gremlin1* in the region 15q13.3. Our study strengthens our

understanding of the genetic determinants underlying sarcopenia-related phenotypes, and contributes to further functional studies.

## ACKNOWLEDGEMENTS

The study was partially supported by Shanghai Leading Academic Discipline Project (S30501) and startup fund from Shanghai University of Science and Technology. The investigators of this work were partially supported by grants from NIH (P50AR055081, R01AG026564, R01AR050496, RC2DE020756, R01AR057049, and R03TW008221) and the Franklin D. Dickson/Missouri Endowment and the Edward G. Schlieder Endowment. Lei Zhang was supported by National Natural Science Foundation of China project (31100902). This work was partially supported by Shanghai Pujiang Program (10PJ1407700) for Yan Zhang.

- Sipila, S., Heikkinen, E., Cheng, S., Suominen, H., Saari, P., Kovanen, V. *et al.* Endogenous hormones, muscle strength, and risk of fall-related fractures in older women. *J. Gerontol. A Biol. Sci. Med. Sci.* **61**, 92–96 (2006).
- Karakelides, H. & Nair, K. S. Sarcopenia of aging and its metabolic impact. *Curr. Top Dev. Biol.* **68**, 123–148 (2005).
- Hsu, F. C., Lenchik, L., Nicklas, B. J., Lohman, K., Register, T. C., Mychaleckyj, J. *et al.* Heritability of body composition measured by DXA in the diabetes heart study. *Obes. Res.* **13**, 312–319 (2005).
- Keen-Kim, D., Mathews, C. A., Reus, V. I., Lowe, T. L., Herrera, L. D., Budman, C. L. *et al.* Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. *Hum. Mol. Genet.* **15**, 3324–3328 (2006).
- 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).
- Liu, X. G., Tan, L. J., Lei, S. F., Liu, Y. J., Shen, H., Wang, L. *et al.* Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. *Am. J. Hum. Genet.* **84**, 418–423 (2009).
- Hindorf, L. A., Junkens, H. A., Mehta, J. P. & TA, M. A catalog of published genome-wide association studies. *National Human Genome Research Institute* (online), (<http://www.genome.gov/26525384>) (2009).
- Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
- Yang, T. L., Chen, X. D., Guo, Y., Lei, S. F., Wang, J. T., Zhou, Q. *et al.* Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am. J. Hum. Genet.* **83**, 663–674 (2008).
- Aitman, T. J., Dong, R., Vyse, T. J., Norsworthy, P. J., Johnson, M. D., Smith, J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
- Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–573 (2009).
- Degenhardt, J. D., de Candia, P., Chabot, A., Schwartz, S., Henderson, L., Ling, B. *et al.* Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). *PLoS Genet.* **5**, e1000346 (2009).
- Hansen, R. D., Raja, C., Aslani, A., Smith, R. C. & Allen, B. J. Determination of skeletal muscle and fat-free mass by nuclear and dual-energy x-ray absorptiometry methods in men and women aged 51–84 y (1–3). *Am. J. Clin. Nutr.* **70**, 228–233 (1999).
- Payette, H., Hanusaik, N., Boutier, V., Morais, J. A. & Gray-Donald, K. Muscle strength and functional mobility in relation to lean body mass in free-living frail elderly women. *Eur. J. Clin. Nutr.* **52**, 45–53 (1998).
- Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissono, D., Mannucci, P. M. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Petrovsky, N., Schmechtig, A., Flomen, R. H., Kumari, V., Collier, D., Makoff, A. *et al.* CHRFAM7A copy number and 2-bp deletion polymorphisms and antisaccade performance. *Int. J. Neuropsychopharmacol.* **12**, 267–273 (2009).
- Severance, E. G., Dickerson, F. B., Stallings, C. R., Origoni, A. E., Sullens, A., Monson, E. T. *et al.* Differentiating nicotine- versus schizophrenia-associated decreases of the alpha7 nicotinic acetylcholine receptor transcript, CHRFAM7A, in peripheral blood lymphocytes. *J. Neural. Transm.* **116**, 213–220 (2009).
- Sinkus, M. L., Lee, M. J., Gault, J., Logel, J., Short, M., Freedman, R. *et al.* A 2-base pair deletion polymorphism in the partial duplication of the alpha7 nicotinic acetylcholine gene (CHRFAM7A) on chromosome 15q14 is associated with schizophrenia. *Brain Res.* **1291**, 1–11 (2009).

- 22 Wang, X. L., Deng, F. Y., Tan, L. J., Deng, H. Y., Liu, Y. Z., Papanian, C. J. *et al*. Bivariate whole genome linkage analyses for total body lean mass and BMD. *J. Bone Miner. Res.* **23**, 447–452 (2008).
- 23 Zhao, L. J., Xiao, P., Liu, Y. J., Xiong, D. H., Shen, H., Recker, R. R. *et al*. A genome-wide linkage scan for quantitative trait loci underlying obesity related phenotypes in 434 Caucasian families. *Hum. Genet.* **121**, 145–148 (2007).
- 24 Gazzero, E. & Canalis, E. Bone morphogenetic proteins and their antagonists. *Rev. Endocr. Metab. Disord.* **7**, 51–65 (2006).
- 25 Janssen, I., Heymsfield, S. B., Wang, Z. M. & Ross, R. Skeletal muscle mass and distribution in 468 men and women aged 18–88yr. *J. Appl. Physiol.* **89**, 81–88 (2000).
- 26 Lexell, J. Human aging, muscle mass, and fiber type composition. *J. Gerontol. A Biol. Sci. Med. Sci.* **50** (Spec No), 11–16 (1995).
- 27 Snijders, T., Verdijk, L. B. & van Loon, L. J. The impact of sarcopenia and exercise training on skeletal muscle satellite cells. *Ageing Res. Rev.* **8**, 328–338 (2009).
- 28 Thornell, L. E., Lindstrom, M., Renault, V., Mouly, V. & Butler-Browne, G. S. Satellite cells and training in the elderly. *Scand. J. Med. Sci. Sports.* **13**, 48–55 (2003).
- 29 Reshef, R., Maroto, M. & Lassar, A. B. Regulation of dorsal somitic cell fates: BMPs and Noggin control the timing and pattern of myogenic regulator expression. *Genes Dev.* **12**, 290–303 (1998).
- 30 Frank, N. Y., Kho, A. T., Schatton, T., Murphy, G. F., Molloy, M. J., Zhan, Q. *et al*. Regulation of myogenic progenitor proliferation in human fetal skeletal muscle by BMP4 and its antagonist Gremlin. *J. Cell Biol.* **175**, 99–110 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)