

## ARTICLE

# Rare variant associations with waist-to-hip ratio in European-American and African-American women from the NHLBI-Exome Sequencing Project

Mengyuan Kan<sup>1,2,23</sup>, Paul L Auer<sup>3,23,24</sup>, Gao T Wang<sup>1,23</sup>, Kristine L Bucasar<sup>1</sup>, Stanley Hooker<sup>1</sup>, Alejandra Rodriguez<sup>1,25</sup>, Biao Li<sup>1</sup>, Jaclyn Ellis<sup>4</sup>, L Adrienne Cupples<sup>5,6</sup>, Yii-Der Ida Chen<sup>7,8,9</sup>, Josée Dupuis<sup>5,6</sup>, Caroline S Fox<sup>6</sup>, Myron D Gross<sup>10</sup>, Joshua D Smith<sup>11</sup>, Nancy Heard-Costa<sup>6,12</sup>, James B Meigs<sup>13</sup>, James S Pankow<sup>10</sup>, Jerome I Rotter<sup>7,8</sup>, David Siscovick<sup>14,15</sup>, James G Wilson<sup>16</sup>, Jay Shendure<sup>11</sup>, Rebecca Jackson<sup>17</sup>, Ulrike Peters<sup>3</sup>, Hua Zhong<sup>18</sup>, Danyu Lin<sup>19</sup>, Li Hsu<sup>3</sup>, Nora Franceschini<sup>20</sup>, Chris Carlson<sup>3</sup>, Goncalo Abecasis<sup>21</sup>, Stacey Gabriel<sup>22</sup>, Michael J Bamshad<sup>11</sup>, David Altshuler<sup>22</sup>, Deborah A Nickerson<sup>10</sup>, Kari E North<sup>20</sup>, Leslie A Lange<sup>4</sup>, Alexander P Reiner<sup>3,15</sup>, NHLBI-Exome Sequencing Project and Suzanne M Leal<sup>\*1</sup>

**Waist-to-hip ratio (WHR), a relative comparison of waist and hip circumferences, is an easily accessible measurement of body fat distribution, in particular central abdominal fat. A high WHR indicates more intra-abdominal fat deposition and is an established risk factor for cardiovascular disease and type 2 diabetes. Recent genome-wide association studies have identified numerous common genetic loci influencing WHR, but the contributions of rare variants have not been previously reported. We investigated rare variant associations with WHR in 1510 European-American and 1186 African-American women from the National Heart, Lung, and Blood Institute-Exome Sequencing Project. Association analysis was performed on the gene level using several rare variant association methods. The strongest association was observed for rare variants in *IKBKB* ( $P = 4.0 \times 10^{-8}$ ) in European-Americans, where rare variants in this gene are predicted to decrease WHRs. The activation of the *IKBKB* gene is involved in inflammatory processes and insulin resistance, which may affect normal food intake and body weight and shape. Meanwhile, aggregation of rare variants in *COBLL1*, previously found to harbor common variants associated with WHR and fasting insulin, were nominally associated ( $P = 2.23 \times 10^{-4}$ ) with higher WHR in European-Americans. However, these significant results are not shared between African-Americans and European-Americans that may be due to differences in the allelic architecture of the two populations and the small sample sizes. Our study indicates that the combined effect of rare variants contribute to the inter-individual variation in fat distribution through the regulation of insulin response.**

*European Journal of Human Genetics* (2016) **24**, 1181–1187; doi:10.1038/ejhg.2015.272; published online 13 January 2016

## INTRODUCTION

Waist-to-hip ratio (WHR) is a common anthropometric measurement of body fat distribution, in particular central abdominal fat. A larger WHR indicates more intra-abdominal fat deposition and is an established risk factor for type 2 diabetes<sup>1,2</sup> and cardiovascular

disease.<sup>3,4</sup> Moreover, it has been demonstrated that body shape, rather than weight, is a better predictor of cardiovascular risk.<sup>5,6</sup> WHR is a heritable trait,<sup>7</sup> and many studies have investigated the genetic influence on body fat distribution.<sup>8,9</sup> Recent genome-wide association studies (GWAS) of common variants conducted in different ethnic

<sup>1</sup>Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; <sup>2</sup>Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Graduate School of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China; <sup>3</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; <sup>4</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; <sup>5</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; <sup>6</sup>National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA; <sup>7</sup>Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, at Harbor-UCLA Medical Center, Torrance, CA, USA; <sup>8</sup>Department of Pediatrics, at Harbor-UCLA Medical Center, Torrance, CA, USA; <sup>9</sup>Department of Medicine, at Harbor-UCLA Medical Center, Torrance, CA, USA; <sup>10</sup>Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA; <sup>11</sup>Department of Genome Sciences, University of Washington, Seattle, WA, USA; <sup>12</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA; <sup>13</sup>General Medicine Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; <sup>14</sup>Department of Medicine, University of Washington, Seattle, WA, USA; <sup>15</sup>Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA; <sup>16</sup>Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; <sup>17</sup>Center for Clinical and Translational Science, Ohio State University, Columbus, OH, USA; <sup>18</sup>Division of Biostatistics, Departments of Environmental Medicine and Population Health, NYU Langone Medical Center, New York, NY, USA; <sup>19</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA; <sup>20</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA; <sup>21</sup>Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; <sup>22</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

\*Correspondence: Professor SM Leal, Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, 700D, Houston, TX 77030, USA. Tel: +1 713 798 4011; Fax: +1 713 798 4373; E-mail: sleal@bcm.edu

<sup>23</sup>These authors contributed equally to this work.

<sup>24</sup>Current address: School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA.

<sup>25</sup>Current address: Gene Mapping Group, Universidad de Antioquia, Medellín 050010, Colombia.

Received 6 April 2015; revised 10 November 2015; accepted 26 November 2015; published online 13 January 2016

populations have reported several genetic loci associated with WHR.<sup>10–14</sup> However, these findings explain only a modest percentage of the genetic variance of WHR. The involvement of rare (minor allele frequency (MAF)  $\leq 0.01$ ) variants has not previously been well studied owing to their poor representation on commercial genotyping arrays.

Our study aims to detect the association between rare variants and WHR to better understand the genetic etiology underlying central adiposity. Although rare variants potentially have larger effect sizes than common variants, there is little statistical power to detect association signals when analyzing individual variants.<sup>15</sup> However, taking into account the cumulative effects of multiple rare variants in specific genes or genetic regions can strengthen association signals, thereby increasing the power to detect rare variant associations. We analyzed exome sequence data using several rare variant association methods: Combined Multivariate Collapsing (CMC),<sup>16</sup> Burden of Rare Variants (BRV) method<sup>16–18</sup> Weighted Sum Statistic (WSS),<sup>19,20</sup> Variable Threshold (VT)<sup>21</sup> and Sequence Kernel Association Test (SKAT).<sup>22</sup>

Exome sequencing targets the protein-coding variants in the human genome. It is a proven approach to detect causal variants for Mendelian disorders.<sup>23</sup> There is also great interest in using exome sequencing to elucidate the involvement of rare variants in the genetic etiology of complex traits. With that goal in mind, the National Heart, Lung, and Blood (NHLBI)-Exome Sequencing Project (ESP) was realized. This project sequenced samples from  $\sim 7000$  individuals who were selected for 12 primary traits and also had data on 59 secondary phenotypes.<sup>24–26</sup> One of the secondary phenotypes, WHR, was available on 1534 European-American (EA) and 1216 African-American (AA) women and 784 EA and 296 AA men. For this study, we limited our analysis to women, as it has previously been demonstrated that there are significant differences in the distribution of WHR measurements and in the underlying genetic etiology of WHR between men and women.<sup>10,13</sup> However, owing to the small sample size for men, they were not analyzed.

## SUBJECTS AND METHODS

### Study sample ESP and quality control

A total of 6823 participants in NHLBI-ESP underwent exome sequencing in approximately equal numbers at either the Broad Institute of MIT/Harvard or the University of Washington. In brief, paired-end sequencing ( $2 \times 76$  bp) was performed on either the Illumina Genome Analyzer II or HiSeq2000 sequencers (Illumina, San Diego, CA, USA) to an average depth of  $\sim 90\times$ . The Broad Institute used Agilent SureSelect Human All Exon 50 Mb capture target (Agilent, Santa Clara, CA, USA) while the University of Washington used Roche NimbleGen SeqCap EZ (Roche, Basel, Switzerland). Single-nucleotide variants (SNVs) were called using glfMultiples of the UMAKE pipeline at the University of Michigan that implements a maximum likelihood method to perform multi-variant calling. Reads were mapped to human reference (hg19) with Burrows-Wheeler Aligner (BWA)<sup>27</sup> and summarized in Binary Align/Map (BAM) files as joint calling input and further refined by duplicate removal, recalibration and indel re-alignment using the Genome Analysis ToolKit (GATK).<sup>28</sup> Low-quality reads with phred-scaled mapping quality  $< 20$  were excluded. A support vector machine (SVM) classifier, which is also part of the UMAKE pipeline, was used to separate likely true-positive and false-positive variant sites, and those variant sites that were likely to be false positives were excluded. A total of 1 908 614 SNVs passed the SVM filter. Subsequent data quality control was performed using Variant Association Tools (VAT),<sup>29</sup> unless otherwise denoted. We then removed variant calls with a read depth  $\leq 10\times$ , 284 variant sites with an average read depth across all samples  $> 500\times$  as these regions are likely to contain copy number variants that can induce incorrect variant calls and variant sites that deviated from Hardy–Weinberg equilibrium in either EA ( $N=2592$ ) or AA ( $N=2779$ ) with  $P$ -values  $< 5 \times 10^{-8}$  based on an exact test.<sup>30</sup> To alleviate bias from different capture targets, we removed variant

sites missing  $> 10\%$  of their genotypes and samples with a genotype missing rate  $> 10\%$  per gene in association analysis step. Only those genes with  $\geq 3$  variant sites were analyzed.

Although self-reported ethnicity was available, we designated EA and AA using Multidimensional Scaling (MDS)<sup>31</sup> and removed 30 samples owing to indeterminate race and 27 samples where there was a discrepancy between self-reported and MDS ethnicity. ESP sequenced a number of duplicate samples for quality control. Additionally, several related individuals were included in the study. In order to detect cryptic duplicate and related samples, kinship analysis was performed.<sup>29,31</sup> For duplicate sample pairs, the one with the best sequence quality was retained in the analysis. For related individuals, only one individual per related group was analyzed; selection was based upon availability of phenotype data as well as quality of sequence data. To avoid type I error owing to inclusion of related individuals, only one individual from each family was retained in the analysis. A total of 13 EA samples and 25 AA samples were removed owing to relatedness. It was also evaluated if reported sex differed from chromosomal sex.<sup>29</sup> Fifteen individuals for whom reported sex differed from chromosomal sex or who had Turner or Klinefelter syndrome were removed from the analysis.

Phenotype quality control was performed using PhenoMan.<sup>32</sup> We excluded females who were aged either  $< 18$  or  $> 90$  years, had a height  $< 140$  cm, a body mass index (BMI)  $< 15$  kg/m<sup>2</sup> or had unrealistic WHR values, presumably owing to data entry errors. In total 24 EA and 30 AA females with WHR phenotype data were excluded from the analysis, and 1510 EA and 1186 AA were available for analysis.

### Association analyses

Association analysis was performed in EA and AA separately using linear regression models, which included a number of covariates that could be confounders: age, age squared, BMI, and current smoking status. Covariate selection was performed using PhenoMan under backward selection. In order to control for population stratification and substructure, we estimated MDS components separately for EAs and AAs and included the first two components in the regression model. Additionally, a dummy variable was included in the regression model to control for sampling procedure, cohort membership and capture target.

All the association tests were performed using the VAT software.<sup>29</sup> Before initiation of association testing, all the variants were first annotated using the SeattleSeq Variation Annotation 134 and gene regions were defined by the Reference Sequence (RefSeq) database. We applied five rare variant association tests using a linear regression framework analyzing EAs and AAs separately. Only potentially protein function-altering variant sites, that is, missense, nonsense and splice sites were analyzed. Four fixed-effect methods, CMC, BRV, WSS and VT, were used in the analysis, which involve different allele coding, weighting and maximization procedures. The CMC uses an indicator variable to code whether a rare variant is present or absent within a gene region while the BRV regresses the number of rare variants within a gene region for each individual. The WSS implements BRV coding but weights each variant by its overall sample MAF, thereby up-weighting rare variants. The VT also employs BRV coding but maximizes the test over allele frequencies, adjusting for multiple testing. The random effects test SKAT, which implements a variance component score statistics, was also used to analyze the data. For SKAT, we used Quantile–Quantile (QQ) normalized WHRs, as it was computationally infeasible to obtain permutation-based empirical  $P$ -values while for the fixed-effect tests unadjusted phenotype values were used. For all tests with the exception of VT, we analyzed rare variants with  $MAF \leq 0.01$  based on the population-specific MAF for EAs and AAs estimated from the entire ESP sample. A total of 15 405 genes with 200 526 SNVs and 15 599 genes with 215 400 SNVs were analyzed for EAs and AAs, respectively. For the VT method, low-frequency variants were also analyzed using  $MAF \leq 0.05$ ; a total of 15 602 genes with 210 718 SNVs and 15 904 genes with 238 574 SNVs were analyzed for EAs and AAs, respectively.

Although in the data quality-control steps variant sites missing  $> 10\%$  of variant calls were removed, even lower levels of missing data can still increase type I error. We therefore replaced missing genotypes with an imputed genotype based on the population-specific ESP allele frequencies.<sup>17</sup> For all rare

variant association methods, *P*-values were obtained empirically using adaptive permutation with the exception of SKAT for which analytical *P*-values were acquired owing to the computational intensity of this method. Finally, meta-analysis was used to combine test results from EA and AA, using a sample-size based method<sup>33</sup> for the CMC, BRV, WSS and VT methods and the MetaSKAT package<sup>22</sup> for SKAT.

### Functional evaluation of variant sites

To estimate the evolutionary conservation of the nucleotide and the amino-acid residue changes, PhyloP and GERP scores were used. PhyloP indicates nucleotide conservation based on multiple alignments of 100 vertebrate species under a null hypothesis of neutral evolution. GERP provides position-specific estimates of evolutionary constraint using maximum likelihood evolutionary rate estimation. To assess potential functional consequences, PolyPhen2, PROVEAN, SIFT, CADD, MutationTaster and LRT were included. PolyPhen-2 implements a naive Bayes classifier to predict possible impact of an amino-acid substitution from sequence alignments and protein structural properties. SIFT and PROVEAN compute a combined score based on the degree of conservation of amino-acid residues in the sequence alignments; and PROVEAN can also measure the potential impact of indels. CADD objectively integrates multiple annotations into one measure (*C* score) for each variant. MutationTaster employs a Bayes classifier to calculate probabilities for whether the alteration to be harmful or not. The LRT method utilizes the log likelihood ratio of the conserved relative to neutral model to measure the deleteriousness of a mutation. Using these bioinformatics tools listed above, we annotated rare variants analyzed in selected genes.

## RESULTS

We performed whole-exome sequencing data analysis on 1510 EA and 1186 AA women to study the association between the quantitative trait WHR and rare variants. These analyzed samples were ascertained from six population-based cohorts and classified into different phenotypic cohorts according to their associated primary phenotypes and were sequenced using one of four in-solution capture targets (Supplementary Table S1). EA women had a mean WHR of  $0.84 \pm 0.087$  (refers to mean  $\pm$  SD hereafter) and AA women had a mean WHR of  $0.85 \pm 0.092$ . Additional phenotypic information on the study participants can be found in Table 1. The *P*-values for covariates that were included in the regression analysis are listed in Supplementary Tables S2 and S3. The distributions of variants categorized by type, for example, missense, and frequency, that is,  $\leq 0.01$  and  $> 0.01$  for each population are shown in Supplementary Table S4. More rare variant sites are observed in AAs compared with EAs (ie,  $P < 2.2 \times 10^{-16}$  for both missense and synonymous variants, proportion test), while EAs have more rare nonsense and splice sites than AAs ( $P < 2.2 \times 10^{-16}$  and  $P = 6.63 \times 10^{-13}$ , respectively). For all types of coding variants with a MAF  $> 1\%$ , more sites are observed in AAs compared with EAs, but only missense and synonymous variant sites showed a statistical difference ( $P = 7.95 \times 10^{-4}$  and  $P = 7.29 \times 10^{-4}$ , respectively). Although rare coding variants are predominant within the sample, only a small proportion of rare

coding variants are shared in EA and AA (8.7, 9.6, 6.4 and 5.9% for missense, synonymous, nonsense and splice sites, respectively). Although variant sites with a MAF  $> 0.01$  occur less frequently than rare variants, a larger proportion are shared in both populations (35.9, 40.1, 32.0 and 27.6% for missense, synonymous, nonsense and splice sites, respectively). Notably, synonymous variants have the largest proportion of variants shared between populations regardless of MAF.

Population specific QQ plots and Manhattan plots for each gene-based test are shown in Supplementary Figures S1–S5. The QQ plots demonstrate that for each rare variant association test type I error is well controlled. The most significant genes associated with WHR ( $P < 0.0005$  in any of the five gene-based tests) in EA, AA or meta-analysis are listed in Supplementary Tables S5–S7. Most of the genes with suggestive association with WHR have consistent results for fixed-effect rare variant association methods (CMC, BRV, WSS and VT) while other genes show suggestive associations for the random-effects test SKAT.

An exception was for *IKBKB* (inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta; MIM 603258), which yielded *P*-values less than exome-wide significance ( $P = 2.5 \times 10^{-6}$  Bonferroni corrected *P*-value for testing 20 000 genes) and was detected to be significantly associated with WHR (eg, CMC results  $\beta = -0.131$ ;  $P = 4.00 \times 10^{-8}$ , Table 2) in EA females for all the fixed-effect rare variant association methods, while the *P*-value of  $4.03 \times 10^{-6}$  for SKAT was slightly greater than exome-wide significance (Table 2).

Among the EA women in our data set, eight rare *IKBKB* missense variants were analyzed (Figure 1a, Supplementary Table S8). Three variant sites are novel to ESP. Each of the eight *IKBKB* missense variants was observed in only a single heterozygous EA carrier in our data set. Based on consensus prediction across multiple annotation methods, four variant sites likely have functional consequences, while for two others it is unclear whether they have an impact on protein function and two variant sites are most likely benign. The aggregated effect of *IKBKB* rare variants decreases WHRs in EA ( $\beta = -0.131$ ); and each of the eight EA rare variant carriers have WHRs lower than the mean (0.84), with the exception of one individual who is the carrier of the rare variant at rs150441824 (WHR = 0.88), which is predicted to be benign. The carrier of the rare variant at rs202226005, which is also predicted to be benign, has a WHR of 0.78. Although her WHR is lower than the mean, it is still greater than other females carrying *IKBKB* variants, which are most likely to be functional. *IKBKB* does not show an association with WHR in AA (CMC result  $\beta = -0.00042$ ;  $P = 0.99$ ; meta-analysis with EA:  $P = 3.86 \times 10^{-5}$ , Table 2). For AA, 11 rare missense variants within *IKBKB* were discovered by exome sequencing and none overlap with those observed in EA (Figure 1a, Supplementary Table S8). Additionally, AA rare variant carriers have mean WHR ( $0.85 \pm 0.084$ ) and six AA rare variant carriers have WHR that are greater than the mean WHR ( $0.85 \pm 0.092$ ) for AA women

**Table 1** Phenotypic information for women analyzed for waist-to-hip ratios

Phenotype	EA	AA
Number of study participants	1510	1186
Age (years) <sup>a</sup>	59 $\pm$ 12 (19–84)	59 $\pm$ 10 (18–88)
BMI (kg/m <sup>2</sup> ) <sup>a</sup>	27.69 $\pm$ 5.80 (15.71–49.89)	34.02 $\pm$ 9.94 (17.16–59.90)
WHR <sup>a</sup>	0.84 $\pm$ 0.087 (0.63–1.20)	0.85 $\pm$ 0.092 (0.59–1.15)
Number of nonsmokers/smokers	1216/294	988/198

<sup>a</sup>Data shown are the mean  $\pm$  SD and range (minimum to maximum value).

**Table 2 Association analysis results for *IKBKB***

Population	Sample size	Number of variant sites/alleles			CMC			BRV			WSS			VT			SKAT		
		MAF ≤ 0.01	MAF ≤ 0.05	MAF ≤ 0.05	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
EA	1508	8/8	8/8	8/8	-0.131	0.027	<b>4.00 × 10<sup>-8</sup></b>	-0.131	0.027	<b>1.20 × 10<sup>-7</sup></b>	-0.131	0.027	<b>1.80 × 10<sup>-7</sup></b>	-0.137	0.029	<b>2.00 × 10<sup>-7</sup></b>	-0.137	0.029	<b>4.03 × 10<sup>-6</sup></b>
AA	1186	11/14	13/90	13/90	-0.00042	0.022	0.99	-0.00042	0.021	0.98	0.0040	0.024	0.82	-0.015	0.017	0.78	-0.015	0.017	0.75
Meta	2694	19/22	21/98	21/98	-0.053	0.017	3.86 × 10 <sup>-5</sup>	-0.051	0.017	7.00 × 10 <sup>-5</sup>	-0.054	0.018	1.71 × 10 <sup>-4</sup>	-0.045	0.014	4.55 × 10 <sup>-5</sup>	-0.045	0.014	4.66 × 10 <sup>-4</sup>

P-values that reached exome-wide significant threshold ( $2.5 \times 10^{-6}$ ) are indicated in bold.

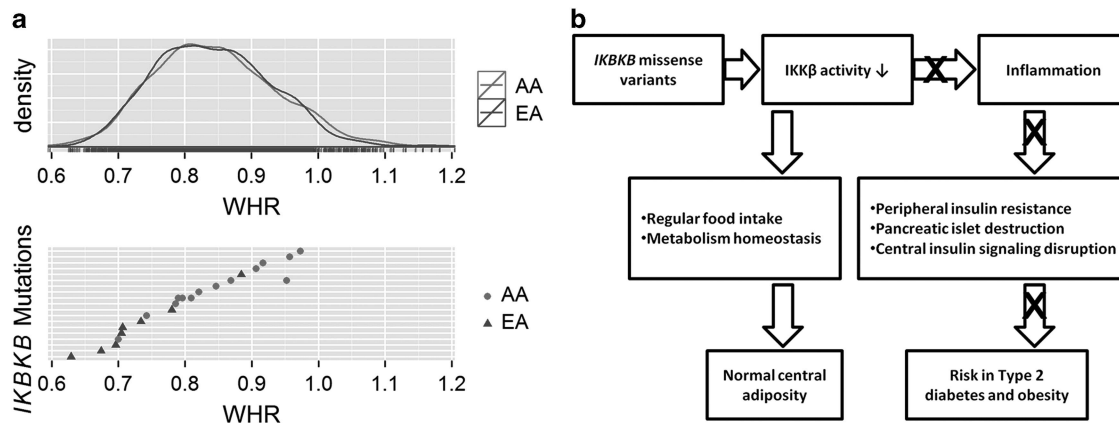
(Figure 1a). Thus the allelic architecture of central adiposity may differ for women of African *versus* European ancestry.

Several additional genes were nominally associated with WHR ( $P < 5 \times 10^{-4}$ , Supplementary Tables S5–S7). Most of them are novel and not within regions previously identified to be associated in GWAS of common variants, with the exception of one gene, *COBLL1* (COBL-like protein 1; MIM 610318; CMC result  $\beta = 0.04$ ;  $P = 2.33 \times 10^{-4}$ , Supplementary Table S5). A few studies<sup>10,12–14</sup> reported 3' UTR and intergenic *COBLL1* variants to be significantly associated ( $P < 5 \times 10^{-8}$ ) with increased WHR; however, these SNPs were not analyzed because they were not exonic and hence were not included in the capture array. To test whether common variant associations have potential effect on rare variant association signal, we obtained genotype data from HapMap3 within a 642-kb region surrounding *COBLL1* (Supplementary Figure S6). A common intronic variant rs6712203 in WHR data set were found in a linkage disequilibrium (LD) block with the GWAS index SNPs ( $r^2 = 0.79$  for rs10195252 and  $r^2 = 0.85$  for rs6717858, Supplementary Figure S6). Conditional regression was performed adjusting for rs6712203 to eliminate potential effects of those index SNPs (Supplementary Table S10). The *COBLL1* rare variant association signal still remained nominally significant after adjustment (CMC result  $P = 3.57 \times 10^{-4}$ , Supplementary Table S10), indicating that the association signal from rare variants is likely to be independent of common variant associations. Additionally, some of the nominally associated genes that were not previously shown to be associated with WHR have data from functional studies that suggest their involvement in adipocyte metabolism or insulin signaling, which includes *CTSB* (MIM 116810) (adipocyte metabolism), *FOXO1* (MIM 136533) (adipocyte metabolism), *ITIH5* (MIM 609783) (adipocyte metabolism) and *PAQR3* (MIM 614577) (insulin signaling) (see Supplementary Tables S5–S7 for additional information on  $\beta$  and P-values for these genes).

Additionally, we performed rare variant association analyses for two significantly associated genes (*IKBKB* and *COBLL1*) for the traits waist circumference and hip circumference, respectively (Supplementary Table S11). Interestingly, rare variants in *IKBKB* were nominally associated with waist circumference ( $P = 9.21 \times 10^{-6}$ ), and in *COBLL1* were weakly associated ( $P = 0.0080$ ). Neither of the genes were associated with hip circumference ( $P > 0.1$ ).

To control for multiple testing due to performing association analysis using five gene-based methods, we used Bonferroni correction. After adjusting for the five methods (CMC, BRV, WSS, VT and SKAT), rare variants in *IKBKB* still showed a significant association with WHR (CMC result  $P = 2.00 \times 10^{-7}$ ) and *COBLL1* showed a suggestive association (CMC result  $P = 0.0012$ , Supplementary Table S12), indicating that multiple testing did not substantially deflate P-values.

We approximated the strength of genetic contributions to WHR by comparing effect size between WHR loci from GWAS<sup>10,13,14</sup> and rare variants in our study. Rare variants in *IKBKB* have an aggregated effect size of 0.13 of decreased WHR, which is much greater than the largest effect size from GWAS, 0.06 of the *LYPLAL1* locus associated with WHR in women.<sup>10,13,14</sup> For the *COBLL1* locus, carriers with the T allele of common variant rs10195252 have an increased WHR of 0.053, and the increase is consistent among different GWAS.<sup>10,13,14</sup> However, rare variants in *COBLL1* have an aggregated effect size of 0.04 to increased WHR, which is about the same scale as the effect size of the common variant. Therefore, there is no clear evidence from this study that rare variants have larger effects than common variants for the WHR phenotype.



**Figure 1** *IKBKB* rare variants associated with decreased WHR in EA. (a) WHR distributions in 1510 EA (blue) and 1186 AA (red) are shown in the upper panel. WHR distributions in *IKBKB* rare missense variant carriers in EA (blue solid triangle) and AA (red solid dot) are shown in the bottom panel. y Axis represents each *IKBKB* rare variant analyzed in gene-based test ordered by increasing WHRs. (b) Potential functional consequences for *IKBKB* rare variants in insulin mechanism. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

## DISCUSSION

We analyzed exome sequence data from 1510 EA and 1186 AA females and using different methods to detect rare variant associations with the quantitative trait WHR. Rare variants in *IKBKB* are strongly associated with WHR in female EA. In addition, we found nominally significant association in *COBLL1*, a gene that was previously shown to be associated with WHR in GWAS of common variants. Several other nominal rare variant associations with WHR are with genes previously shown to have a role in adipocyte and insulin metabolisms.

*IKBKB* encodes the protein IKK $\beta$ , which phosphorylates inhibitor of NF- $\kappa$ B (I $\kappa$ B) to disassociate the inhibitor/NF- $\kappa$ B complex and activate NF- $\kappa$ B in inflammation. IKK $\beta$  is a key mediator in inflammation pathways by several mechanisms.<sup>34</sup> Obesity- or nutritional overload-induced IKK $\beta$ /NF- $\kappa$ B activation initiates the inflammatory process and ultimately results in insulin resistance in hepatocytes and adipocytes.<sup>35,36</sup> In addition to impairment of peripheral insulin sensitivity, the IKK $\beta$ /NF- $\kappa$ B pathway affects glucose metabolism in pancreatic islets by activation of inflammation, causing islet  $\beta$ -cell failure in type 2 diabetes.<sup>37</sup> Furthermore, overnutrition evokes activation of the IKK $\beta$ /NF- $\kappa$ B pathway in the hypothalamus leading to inflammation and the disruption of central insulin and leptin signaling, thus resulting in impaired central nervous system's control of food intake and promoting body weight gain.<sup>38,39</sup> Consequently, inhibition of IKK $\beta$  activation was also identified as a potential therapeutic target to reverse inflammation in obesity-associated type 2 diabetes.<sup>40</sup> Meanwhile, some rare homozygous null mutations in *IKBKB* have been reported in patients with severe combined immunodeficiency, which result in IKK $\beta$  loss of expression and therefore impairment of immune activation.<sup>41</sup> Additionally, a reported GWAS signal near *IKBKB* downstream for the phenotype tissue plasminogen level is in partial LD with a coding variant in the tissue plasminogen activator (*PLAT*) gene (MIM, 173370) nearby,<sup>42</sup> which suggests that central adiposity with insulin resistance may be correlated with plasminogen activator through *IKBKB*. Interestingly, our results showed *IKBKB* rare variants were associated with reduced WHRs which suggested that they are protective against abdominal obesity (Figure 1a). *IKBKB* amino-acid alterations could lead to IKK $\beta$  protein substructure change, and further cause IKK $\beta$  function inactive in inflammation pathways. The malfunction of IKK $\beta$  protein may reverse the insulin resistance and promote normal regulation of food intake and metabolic homeostasis (Figure 1b).

Rare variants within *COBLL1* displayed a nominal association with increased WHR. Previously, GWAS also reported an association between common variants in *COBLL1* (rs10195252 and rs6717858) and increased WHR.<sup>10,12–14</sup> Conditional analysis adjusting for ESP common SNP suggested that the *COBLL1* rare variant association signal is unlikely influenced by associated GWAS common SNPs. Moreover, other common variants in this locus were also found to be associated with elevated fasting insulin,<sup>43</sup> increased high-density lipoprotein cholesterol<sup>44</sup> and risk of developing type 2 diabetes.<sup>45</sup> These prior association findings provide evidence that genetic variations in this region may contribute to multiple biological traits, which could potentially influence WHR.

Among novel genes nominally associated with WHR, some have been found to have functional impact on adipocyte and insulin metabolism. Cathepsin B, encoded by the *CTSB* gene, contributes to adipocyte cell death and macrophage infiltration into adipose tissue associated with adipocyte hypertrophy.<sup>46</sup> FOXO1 (Forkhead box protein O1), encoded by *FOXO1* gene, is a transcription factor that is essential to the decision for a preadipocyte to commit to adipogenesis.<sup>47</sup> ITIH5 (inter-alpha-trypsin inhibitor heavy chain family, member 5), encoded by *ITIH5*, is highly expressed in adipose tissue and is increased in obesity while being reduced after diet-induced weight loss.<sup>48</sup> PAQR3 (progestin and adipoQ receptor family member III), encoded by *PAQR3*, was found to modulate insulin signaling by phosphoinositide 3-kinase pathway.<sup>49</sup> Although these genes did not reach exome-wide significance, these related studies provide additional evidence that rare variants in these genes may have a role in WHR.

In addition to the association with WHR, we identified that rare variants in *IKBKB* showed evidence of association with waist circumference but showed no association with hip circumference. Therefore, rare variants in these WHR-associated genes might exert their impacts primarily through waist circumference instead of hip circumference, indicating that waist circumference might be a potential driver behind WHR rare variant association.

Interestingly, most of our significant findings are confined to either EAs or AAs, with the majority of associations being found for EAs. This can be due to differences in allelic architecture between populations and the effect sizes of causal variants. Most variant sites are EA or AA specific or are private variants, for example, *IKBKB*. Additionally, only a small proportion, 8.7%, of the analyzed rare

variants was shared by EAs and AAs. AAs have a greater number of rare variants compared with EAs, which should increase the power to detect WHR associations unless a greater proportion of AA-specific variants are non-causal. Additionally, the sample size for AAs is approximately three-fourths the size for EAs, reducing the power to detect associations in AAs. The failure to replicate associations between EAs and AAs may indicate that some of these findings are false positives. Our rare variant association study with current sample size ( $N = 1510$  for EAs and  $N = 1186$  for AAs) is underpowered, which can increase false-positive findings. Simulation has been used to evaluate the necessary sample sizes for rare variant association studies. These simulation studies suggest that for the most part large sample sizes, for example, 50 000 individuals, are necessary for a sufficient power.<sup>50</sup>

Replication of our findings using the currently available HumanOmni5Exome BeadChip is not possible, owing to the large number of variants that are absent from the exome chip (Supplementary Tables S6–S7). For example, of the 8 variants observed in EA for the *IKBKB* gene only 2 are found on the exome chip, while for *COBLL1*, of the 26 variants observed for EA only 7 are present on the exome chip. Therefore, deep targeted sequencing or whole exome or genome data sets with information on WHR are necessary to replicate our findings.

When association results for the gene-based tests were compared, it was observed that the results for most genes are correlated, for example, genes with suggestive significant results shared similar results for each burden tests, despite that some genes had a result that was unique to one method (Supplementary Tables S5–S7). Moreover, although performing a variety of gene-based tests come at a cost of multiple testing that can reduce power, there is no single uniquely most powerful test and performance of tests varies depending on the underlying genetic model, which is unknown.<sup>51</sup> For example, fixed-effect test BRV is powerful when the majority of variants have an effect that is unidirectional while the variance component test SKAT is powerful when either a small proportion of variants are causal or the causal variants have bidirectional effects. Even after performing a Bonferroni correction for five gene-based methods, the association between rare variants in *IKBKB* and WHR remained exome-wide significant (Supplementary Table S12).

In summary, we performed the first study to detect the association between rare variants and complex trait WHR using a variety of rare variant association methods using exome sequence data. Our study provides a preliminary understanding of the role of rare variants in WHR and potentially insulin response pathways, which may also contribute to obesity and type 2 diabetes. Although many of our findings are intriguing and limitations are lack of replication and small sample size, replication and functional studies are needed to confirm the results and evaluate whether these same genes also have a role in obesity and type 2 diabetes.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the support of NHLBI and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO) and RC2 HL-102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO).

#### WEB RESOURCES

1000 Genomes: <http://www.1000genomes.org/>  
 Combined Annotation Dependent Depletion (CADD): <http://cadd.gs.washington.edu/>  
 DistilLD: <http://distild.jensenlab.org/>  
 Database of Genotypes and Phenotypes (dbGaP): <http://www.ncbi.nlm.nih.gov/gap>  
 dbSNP: <http://www.ncbi.nlm.nih.gov/SNP/>  
 dbSNP138: <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/snp138.txt.gz>  
 Exome Aggregation Consortium (ExAC): <http://exac.broadinstitute.org/>  
 Exome Variant Server (EVS): <http://evs.gs.washington.edu/EVS/>  
 GERP: <http://mendel.stanford.edu/SidowLab/downloads/gerp/>  
 HapMap3: [ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/2009-01\\_phaseIII/](ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/2009-01_phaseIII/)  
 LRT: [http://www.genetics.wustl.edu/jflab/lrt\\_query.html](http://www.genetics.wustl.edu/jflab/lrt_query.html)  
 MetaSKAT: <http://www.hsph.harvard.edu/skat/metaskat/>  
 MutationTaster: <http://www.mutationtaster.org/>  
 PhenoMan: <https://code.google.com/p/phenoman/>  
 PhyloP: <http://compugen.bscb.cornell.edu/phast/>  
 PolyPhen2: <http://genetics.bwh.harvard.edu/pph2/>  
 PROVEAN: <http://provean.jcvi.org/index.php>  
 RefSeq database: [ftp://ftp.ncbi.nlm.nih.gov/refseq/H\\_sapiens/](ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/)  
 RefSeqGene/  
 SIFT: <http://sift.jcvi.org/>  
 Variant Association Tools (VAT): <http://varianttools.sourceforge.net/>

#### DATA SUBMISSION

The data analyzed are from NHLBI-ESP. The data, that is, BAM and VCF files, and phenotypes are available in the database of Genotypes and Phenotypes (dbGaP) (<http://www.ncbi.nlm.nih.gov/gap>). Additionally, variant-level data are available from dbSNP, the Exome Variant Server and ExAC databases.

- Carey VJ, Walters EE, Colditz GA *et al*: Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. *Am J Epidemiol* 1997; **145**: 614–619.
- Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB: Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* 2005; **81**: 555–563.
- Canoy D: Distribution of body fat and risk of coronary heart disease in men and women. *Curr Opin Cardiol* 2008; **23**: 591–598.
- Rexrode KM, Carey VJ, Hennekens CH *et al*: Abdominal adiposity and coronary heart disease in women. *JAMA* 1998; **280**: 1843–1848.
- Schneider HJ, Friedrich N, Klotsche J *et al*: The predictive value of different measures of obesity for incident cardiovascular events and mortality. *J Clin Endocrinol Metab* 2010; **95**: 1777–1785.
- Yusuf S, Hawken S, Ounpuu S *et al*: Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* 2005; **366**: 1640–1649.
- Souren NY, Paulussen AD, Loos RJ *et al*: Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities. *Diabetologia* 2007; **50**: 2107–2116.
- Rose KM, Newman B, Mayer-Davis EJ, Selby JV: Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins. *Obes Res* 1998; **6**: 383–392.
- Selby JV, Newman B, Quesenberry CP Jr *et al*: Genetic and behavioral influences on body fat distribution. *Int J Obes* 1990; **14**: 593–602.
- Heid IM, Jackson AU, Randall JC *et al*: Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 2010; **42**: 949–960.
- Lindgren CM, Heid IM, Randall JC *et al*: Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 2009; **5**: e1000508.
- Berndt SI, Gustafsson S, Magi R *et al*: Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013; **45**: 501–512.

- 13 Randall JC, Winkler TW, Kutalik Z *et al*: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* 2013; **9**: e1003500.
- 14 Shungin D, Winkler TW, Croteau-Chonka DC *et al*: New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015; **518**: 187–196.
- 15 Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI: Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 2008; **82**: 100–112.
- 16 Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**: 311–321.
- 17 Auer PL, Wang G, Leal SM: Testing for rare variant associations in the presence of missing data. *Genet Epidemiol* 2013; **37**: 529–538.
- 18 Morris AP, Zeggini E: An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol* 2009; **34**: 188–193.
- 19 Lin DY, Tang ZZ: A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 2011; **89**: 354–367.
- 20 Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 2009; **5**: e1000384.
- 21 Price AL, Kryukov GV, de Bakker PI *et al*: Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 2010; **86**: 832–838.
- 22 Lee S, Teslovich TM, Boehnke M, Lin X: General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* 2013; **93**: 42–53.
- 23 Ku CS, Naidoo N, Pawitan Y: Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011; **129**: 351–370.
- 24 Tennesen JA, Bigham AW, O'Connor TD *et al*: Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012; **337**: 64–69.
- 25 Auer PL, Johnsen JM, Johnson AD *et al*: Imputation of exome sequence variants into population-based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *Am J Hum Genet* 2012; **91**: 794–808.
- 26 Lange LA, Hu Y, Zhang H *et al*: Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 2014; **94**: 233–245.
- 27 Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
- 28 McKenna A, Hanna M, Banks E *et al*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
- 29 Wang GT, Peng B, Leal SM: Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. *Am J Hum Genet* 2014; **94**: 770–783.
- 30 Wigginton JE, Cutler DJ, Abecasis GR: A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005; **76**: 887–893.
- 31 Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM: Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; **26**: 2867–2873.
- 32 Li B, Wang G, Leal SM: PhenoMan: phenotypic data exploration, selection, management and quality control for association studies of rare and common variants. *Bioinformatics* 2014; **30**: 442–444.
- 33 Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
- 34 Hayden MS, Ghosh S: Shared principles in NF-kappaB signaling. *Cell* 2008; **132**: 344–362.
- 35 Arkan MC, Hevener AL, Greten FR *et al*: IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med* 2005; **11**: 191–198.
- 36 Cai D, Yuan M, Frantz DF *et al*: Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med* 2005; **11**: 183–190.
- 37 Tornatore L, Thotakura AK, Bennett J, Moretti M, Franzoso G: The nuclear factor kappa B signaling pathway: integrating metabolism with inflammation. *Trends Cell Biol* 2012; **22**: 557–566.
- 38 Purkayastha S, Zhang G, Cai D: Uncoupling the mechanisms of obesity and hypertension by targeting hypothalamic IKK-beta and NF-kappaB. *Nat Med* 2011; **17**: 883–887.
- 39 Zhang X, Zhang G, Zhang H, Karin M, Bai H, Cai D: Hypothalamic IKKbeta/NF-kappaB and ER stress link overnutrition to energy imbalance and obesity. *Cell* 2008; **135**: 61–73.
- 40 Yuan M, Konstantopoulos N, Lee J *et al*: Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of Ikkbeta. *Science* 2001; **293**: 1673–1677.
- 41 Mousallem T, Yang J, Urban TJ *et al*: A nonsense mutation in IKBKB causes combined immunodeficiency. *Blood* 2014; **124**: 2046–2050.
- 42 Huang J, Huffman JE, Yamakuchi M *et al*: Genome-wide association study for circulating tissue plasminogen activator levels and functional follow-up implicates endothelial STXBP5 and STX2. *Arterioscler Thromb Vasc Biol* 2014; **34**: 1093–1101.
- 43 Manning AK, Hivert MF, Scott RA *et al*: A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**: 659–669.
- 44 Teslovich TM, Musunuru K, Smith AV *et al*: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–713.
- 45 Albrechtsen A, Grarup N, Li Y *et al*: Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 2013; **56**: 298–310.
- 46 Gornicka A, Fetting J, Eguchi A *et al*: Adipocyte hypertrophy is associated with lysosomal permeability both in vivo and in vitro: role in adipose tissue inflammation. *Am J Physiol Endocrinol Metab* 2012; **303**: E597–E606.
- 47 Nakae J, Kitamura T, Kitamura Y, Biggs 3rd WH, Arden KC, Accili D: The forkhead transcription factor Foxo1 regulates adipocyte differentiation. *Dev Cell* 2003; **4**: 119–129.
- 48 Anveden A, Sjöholm K, Jacobson P *et al*: ITH-5 expression in human adipose tissue is increased in obesity. *Obesity (Silver Spring)* 2012; **20**: 708–714.
- 49 Wang X, Wang L, Zhu L *et al*: PAQR3 modulates insulin signaling by shunting phosphoinositide 3-kinase p110alpha to the Golgi apparatus. *Diabetes* 2013; **62**: 444–456.
- 50 Zuk O, Schaffner SF, Samocha K *et al*: Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci USA* 2014; **111**: E455–E464.
- 51 Lee S, Abecasis GR, Boehnke M, Lin X: Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; **95**: 5–23.

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