nature genetics

A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry

Genome-wide association studies (GWAS) have identified 36 loci associated with body mass index (BMI), predominantly in populations of European ancestry. We conducted a meta-analysis to examine the association of >3.2 million SNPs with BMI in 39,144 men and women of African ancestry and followed up the most significant associations in an additional 32,268 individuals of African ancestry. We identified one new locus at 5q33 (*GALNT10*, rs7708584, $P = 3.4 \times 10^{-11}$) and another at 7p15 when we included data from the GIANT consortium (*MIR148A-NFE2L3*, rs10261878, $P = 1.2 \times 10^{-10}$). We also found suggestive evidence of an association at a third locus at 6q16 in the African-ancestry sample (*KLHL32*, rs974417, $P = 6.9 \times 10^{-8}$). Thirty-two of the 36 previously established BMI variants showed directionally consistent effect estimates in our GWAS (binomial $P = 9.7 \times 10^{-7}$), five of which reached genome-wide significance. These findings provide strong support for shared BMI loci across populations, as well as for the utility of studying ancestrally diverse populations.

There are notable racial and ethnic disparities in the prevalence of obesity in the United States; nearly 50% of African-American adults are classified as obese compared to 35% of non-Hispanic whites[1]. GWAS have identified 36 BMI loci at statistically significant levels ($P < 5 \times 10^{-8}$)[2–13], 32 of which were identified in individuals of European ancestry[3–8] and 4 of which were identified in east Asian populations[9,10]. Large GWAS of BMI in populations of African ancestry are lacking and will be important for identifying genetic variants that are unique and/or of greater importance to this population[14–17]. In this study we conducted a large GWAS meta-analysis of BMI in men and women of African ancestry to search for new loci and tested associations with common variation at the 36 known loci to better understand their relevance in populations of African ancestry.

We included 36 GWAS, totaling 39,144 men and women of African ancestry, in the stage 1 meta-analysis of as many as 3,283,202 (minor allele frequency (MAF) >1%) genotyped and imputed SNPs (Online Methods, **Supplementary Tables 1–3** and **Supplementary Note**). After applying both study-specific and overall stage 1 genomic-control corrections (**Supplementary Table 2**), 11 SNPs at five loci achieved genome-wide significance ($P < 5 \times 10^{-8}$) (**Table 1**, **Fig. 1** and **Supplementary Fig. 1**). Four of these loci are known BMI loci (1q25, *SEC16B*; 4p12, *GNPDA2*; 16q12, *FTO*; and 18q21, *MC4R*). The fifth locus, at 5q33 (rs7708584, approximately 27 kb upstream of *GALNT10*, $P = 8.02 \times 10^{-9}$), has not been previously associated with BMI at genome-wide significant levels in any population.

We subsequently selected the 1,500 most significantly associated SNPs from stage 1 ($P < 1.19 \times 10^{-3}$) and examined associations with BMI in an independent sample of 6,817 men and women of African ancestry from seven additional studies (stage 2) (Online Methods, **Supplementary Tables 1–3** and **Supplementary Note**). Of these 1,500 SNPs, 179 replicated at nominal significance ($P < 0.05$) and had effects that were directionally consistent with those in stage 1 (**Supplementary Table 4**). A meta-analysis of stages 1 and 2 revealed a second new locus, 6q16 (rs974417, located in an intronic region of *KLHL32*; stage 2 $P = 3.5 \times 10^{-3}$, combined stages 1 and 2 $P = 2.2 \times 10^{-8}$), and confirmed our finding at rs7708584 on 5q33 near *GALNT10* (stage 2 $P = 9.4 \times 10^{-3}$, combined stages 1 and 2 $P = 2.2 \times 10^{-10}$). We further examined the associations of these two variants in a third stage composed of 25,451 individuals of African ancestry from an additional 12 studies. We found support for an association with both variants, although the strength of the association was greater for rs7708584 (*GALNT10*, $P = 7.1 \times 10^{-3}$) than for rs974417 (*KLHL32*, $P = 0.09$). In combining results across all three stages ($n = 71{,}412$), rs7708584 (*GALNT10*) was significantly associated with BMI ($P = 3.4 \times 10^{-11}$), whereas rs974417 (*KLHL32*) was nearly genome-wide significant ($P = 6.9 \times 10^{-8}$) (**Table 1** and **Fig. 2a**,**b**).

To identify additional new loci that may be of importance across populations, we examined the 1,500 most significant SNPs from stage 1 in publicly available data from the GIANT consortium of ~124,000 individuals European ancestry[7] (Online Methods). rs7708584 (*GALNT10*) was significantly associated with BMI in European-ancestry populations (effect allele frequency (EAF) = 0.42, $P = 1.2 \times 10^{-5}$) but rs974417 (*KLHL32*) was not (EAF = 0.85, $P = 0.45$), although it was directionally consistent. Through a meta-analysis of individuals of European and African ancestry, we identified an additional new variant at 7p15 (rs10261878) that was also associated with BMI in European-ancestry populations (GIANT EAF = 0.94, $P = 2.2 \times 10^{-5}$). rs10261878 on 7p15 is located in an intergenic region 39 kb upstream of *MIR148A* (encoding miRNA-148a) and approximately 241 kb upstream of *NFE2L3*. This variant was positively associated with BMI in stages 1 ($P = 1.7 \times 10^{-4}$) and 3 ($P = 1.0 \times 10^{-3}$) in the African-ancestry GWAS, with a directionally consistent but nonsignificant association in the smaller stage 2 ($P = 0.33$) (**Fig. 2c** and **Supplementary Table 5**). In combining results across studies of African (stages 1, 2 and 3) and European ancestry (combined $n = 194{,}247$), both rs7708584 (*GALNT10*, $P = 5.1 \times 10^{-14}$) and

A full list of authors and affiliations appears at the end of the paper.

Received 5 September 2012; accepted 18 March 2013; published online 14 April 2013; doi:10.1038/ng.2608

Table 1 Summary of the eight independent SNPs that were associated with BMI at genome-wide significant ($P < 5.0 \times 10^{-8}$) levels in men and women of African ancestry

	Previously identified BMI loci				Newly identified BMI loci			
	rs543874	rs7586879	rs348495	rs17817964	rs6567160	rs7708584	rs974417	rs10261878[d]
Nearest gene	*SEC16B*	*ADCY3*	*GNPDA2*	*FTO*	*MC4R*	*GALNT10*	*KLHL32*	*MIR148A-NFE2L3*
Chr.	1	2	4	16	18	5	6	7
Position (build 37)	177889480	25116977	45184442	53828066	57829135	153543466	97419598	25917070
Alleles[a]	G/A	T/C	G/A	T/C	C/T	A/G	C/T	C/A
EAF[b]	0.25	0.77	0.34	0.12	0.21	0.32	0.66	0.44
Stage 1								
n	38,899	38,948	39,097	39,080	39,103	38,219	39,120	39,101
β (s.e.)	0.057 (0.009)	0.042 (0.010)	0.048 (0.009)	0.074 (0.012)	0.062 (0.010)	0.050 (0.009)	0.040 (0.008)	0.030 (0.008)
P	1.80×10^{-10}	1.05×10^{-5}	2.70×10^{-8}	2.27×10^{-9}	2.41×10^{-10}	8.02×10^{-9}	1.49×10^{-6}	1.66×10^{-4}
Stage 2								
n	6,805	6,817	6,817	6,769	6,817	6,817	6,816	6,817
β (s.e.)	0.074 (0.020)	0.073 (0.020)	0.067 (0.021)	0.068 (0.027)	0.045 (0.021)	0.047 (0.018)	0.053 (0.018)	0.017 (0.017)
P	1.49×10^{-4}	3.12×10^{-4}	1.19×10^{-3}	0.012	0.032	9.35×10^{-3}	3.47×10^{-3}	0.330
Stage 3								
n						25,337	25,451	25,308
β (s.e.)	N/A	N/A	N/A	N/A	N/A	0.026 (0.010)	0.015 (0.009)	0.029 (0.009)
P						7.08×10^{-3}	0.091	1.01×10^{-3}
Combined								
n	45,704	45,765	45,914	45,849	45,920	70,373	71,387	194,931
β (s.e.)	0.060 (0.008)	0.047 (0.009)	0.051 (0.008)	0.073 (0.011)	0.059 (0.009)	0.040 (0.006)	0.031 (0.006)	0.032 (0.005)
P	2.00×10^{-13}	3.60×10^{-8}	1.60×10^{-10}	1.05×10^{-10}	2.96×10^{-11}	3.37×10^{-11}	6.88×10^{-8}	1.23×10^{-10}
Explained variance[c] (%)	0.21	0.19	0.20	0.10	0.07	0.04	0.02	0.03

[a]The effect allele is listed first. [b]The frequencies shown are from the stage 1 sample. [c]Calculated using the results from stage 2 for previously identified BMI loci and results from stage 2 and stage 3 for newly identified BMI loci; the total fraction of variance explained was calculated using the formula $[2f(1-f) \times a^2] \times 100$, where f is the frequency of the variant, and a is the additive effect of the variant[3]. [d]Shown are the combined results from African-ancestry stages 1, 2, 3 and GIANT (where the GIANT data are $n = 123{,}706$, β (s.e.) = 0.045 (0.011) and $P = 2.21 \times 10^{-5}$). Chr., chromosome; EAF, effect allele frequency; β (beta estimate) reported in inverse normally transformed units; s.e., standard error. *P* values for between-study heterogeneity were all >0.1.

rs10261878 (*MIR148A-NFE2L3*, $P = 1.2 \times 10^{-10}$) were significantly associated with BMI; rs974417 (*KLHL32*) did not meet the genome-wide significance threshold ($P = 5.7 \times 10^{-6}$). In individuals of east Asian descent from the AGEN[10] and RIKEN[9] consortia ($n = 27{,}715$ and $n = 26{,}620$, respectively) (**Fig. 3**, **Supplementary Table 6** and Online Methods), rs7708584 (*GALNT10*, $P = 0.002$) and rs974417 (*KLHL32*, $P = 0.023$) were directionally consistent and significantly associated with BMI, whereas rs10261878 (*MIR148A-NFE2L3*) was neither directionally consistent nor statistically significantly associated with BMI ($P = 0.053$). We then examined the associations with BMI in children of African ancestry ($n = 3{,}751$) (Online Methods) and found that for all three SNPs, the associations were directionally consistent but did not reach statistical significance ($P > 0.05$) (**Supplementary Table 7**).

To further understand differences by ancestral background as well as characterize the functional and genetic epidemiologic architecture of the two new BMI loci (5q33, *GALNT10*; and 7p15, *MIR148A-NFE2L3*) and the suggestive locus at 6q16 (*KLHL32*), we performed several additional analyses. Local ancestry adjustment (in 69% of the stage 1 sample; Online Methods) resulted in numerically similar effect estimates (**Supplementary Table 8**), and we did not detect evidence of significant effect heterogeneity in analyses stratified by local ancestry (**Supplementary Table 9**). We found that the three BMI loci were associated with waist circumference (among $n \approx 20{,}000$ individuals, many of which overlap those studied here) but not with BMI-adjusted waist circumference, waist-to-hip ratio or height[18] (**Supplementary Table 10**); however, SNPs in this region have been associated with waist-to-hip ratio in Europeans, although at SNPs that are not in linkage disequilibrium (LD) with our index SNP[19]. We found no evidence of pleiotropy with adiposity-related metabolic traits using GWAS data provided by trait-specific consortia in men and women predominantly of European ancestry[20–24] (**Supplementary Table 11**).

We examined associations with BMI in our African-ancestry stage 1 sample of the index SNPs reported for the 36 previously established BMI loci in the European and Asian populations[7,9,10] (**Fig. 3** and **Supplementary Table 12**). The associations were directionally consistent with the effects reported in the original papers for 32 of the 36 established BMI loci (binomial test of direction $P = 9.7 \times 10^{-7}$), 16 of which associated with BMI at $P < 0.01$ (binomial test $P < 1.0 \times 10^{-15}$) (**Supplementary Table 12**).

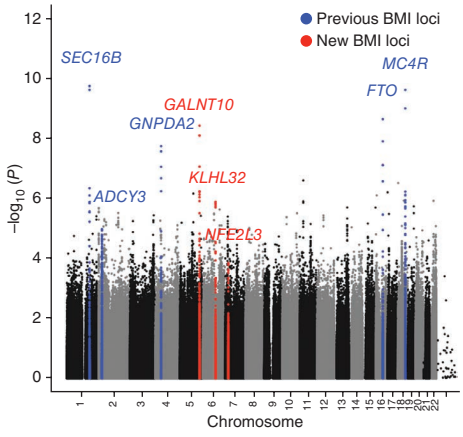


Figure 1 Manhattan plot showing results of the BMI association meta-analysis in the stage 1 studies. Colored genomic loci indicate new associations (red) and those detected previously (blue).

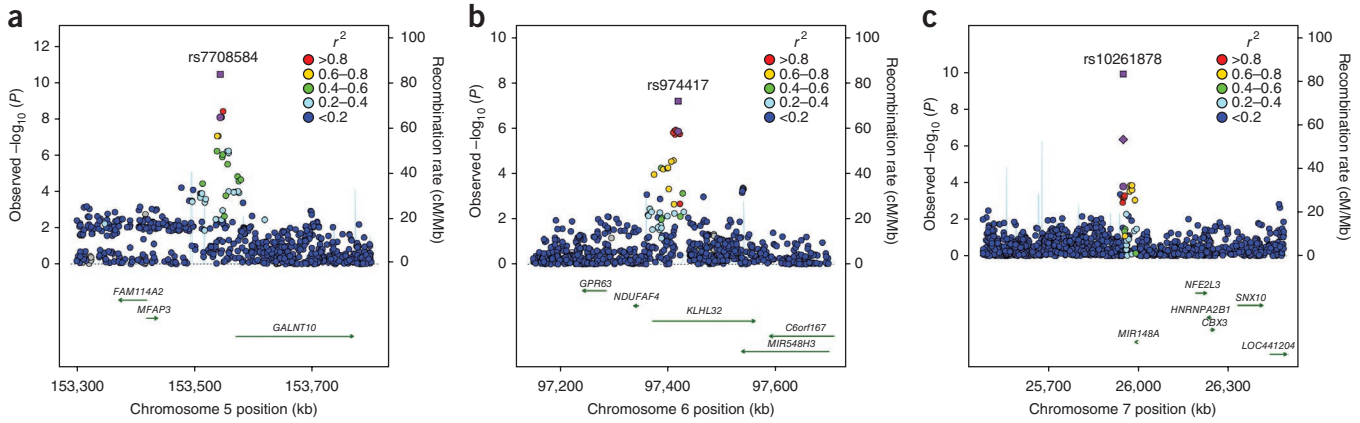


Figure 2 Regional plots of three new genome-wide significant loci identified in men and women of African ancestry. (**a**–**c**) rs7708584 (in the *GALNT10* region; **a**), rs974417 (in the *KLHL32* region; **b**) and rs10261878 (in the *MIR148A-NFE2L3* region; **c**). In **a** and **b**, the stage 1 *P* value is represented by a purple circle, and the combined stages 1, 2 and 3 *P* value is represented by a purple square; in **c**, the stage 1 *P* value is represented by a purple circle, the African-ancestry combined stages 1, 2 and 3 *P* value is represented by a purple diamond, and the combined African ancestry and GIANT consortium *P* value is represented by a purple square. SNPs are plotted by their position within 500 kb on either side of the index SNP on the chromosome against their association ($-\log_{10} P$) with BMI using the stage 1 data. SNPs surrounding the top SNPs are colored to indicate the local LD structure using pairwise r^2 data from the May 2012 AFR panel of 1000 Genomes.

Using the results from the stage 1 meta-analysis, we searched for common variants within the established loci that better captured the association of the index SNP reported in the European and Asian populations. Seven regions (*PTBP2*, *TMEM18*, *DNAJC27* (previously known as *RBJ*), *NUDT3*, *BDNF*, *FTO* and *MC4R*) harbored at least one variant that was correlated with the index SNP in the referent population ($r^2 \geq 0.4$) and was associated with BMI in the African-ancestry GWAS at a significance level that was at least one order of magnitude greater than that observed for the index SNP (Online Methods, **Supplementary Table 13** and **Supplementary Fig. 2a–g**). These variants were also associated with BMI in the GIANT consortium (**Supplementary Table 13**) and are probably better markers of the biologically functional allele, at least in populations of African ancestry. We also interrogated the evidence for possible independent secondary signals by visual inspection of all *P* values of SNP-BMI associations for SNPs with $r^2 < 0.2$ within the 1-Mb region of the index SNP. We did not detect evidence of independent secondary signals at any of the known BMI loci at $P < 6.7 \times 10^{-6}$ (Online Methods). For most loci, the genetic data from African-ancestry populations may assist in refining the location of the risk variant, as there are fewer markers correlated with the strongest signals and/or a more narrowed region in which any proxies reside in this population (**Supplementary Fig. 3**).

To direct us to positional candidate genes, we examined the *cis* associations between the index SNP and expression of gene transcripts within the flanking 1-Mb region (500 kb on each side) in human brain, subcutaneous and omental adipose tissue and liver[25–28] (Online Methods and **Supplementary Table 14**). rs7708584 near *GALNT10* showed nominally significant ($P < 0.05$) associations with GALNT10 expression (for two of the three transcripts available) in liver, omental fat and subcutaneous fat ($P = 0.048$, $P = 0.00010$ and $P = 0.00017$, respectively). We also found suggestive *cis* associations for rs10261878 near *NFE2L3* with NFE2L3 expression in the same three tissues ($P = 0.039$, $P = 0.015$ and $P = 0.036$ for liver, omental fat and subcutaneous fat, respectively). However, despite the consistent associations observed for our lead SNPs in the *GALNT10* and *NFE2L3* loci, other nearby SNPs showed stronger association with the expression levels of the respective transcripts (**Supplementary Fig. 4**). Subsequent conditional analyses adjusting for the most significant expression quantitative trait locus (eQTL) SNP in the region abolished the *cis* associations between the BMI-associated SNPs and the respective transcript expression levels (**Supplementary Table 15**). Taken together, these eQTL analyses could not confirm that the identified BMI-associated SNPs affect GALNT10 and NFE2L3 expression directly.

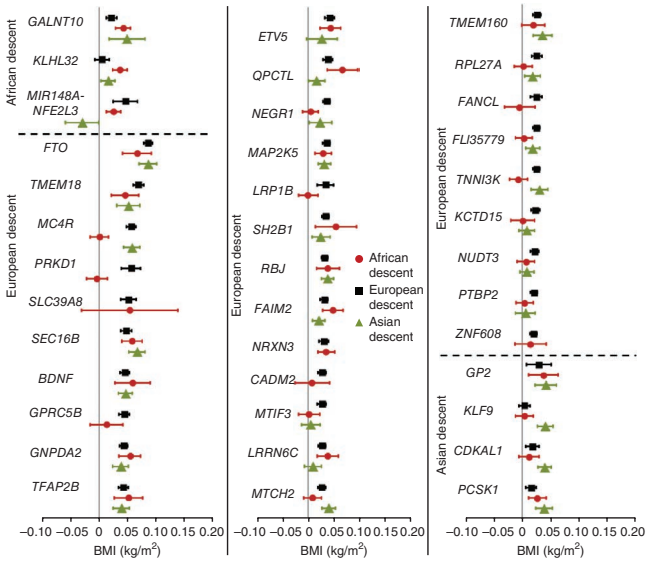


Figure 3 Effect estimates (95% confidence intervals) per BMI-increasing allele for the three new loci discovered in individuals of African ancestry (shown in the first section in descending order of the effect size in this population), the 32 loci discovered in individuals of European ancestry (shown in the second section in descending order of their effect size) and the 4 loci discovered in individuals of Asian ancestry (shown in the third section in descending order of their effect size). Results for individuals of African ancestry are depicted as red dots (combined stages 1, 2 and 3 for new loci and stage 1 for previously discovered loci); results for individuals of European ancestry are depicted as black squares[7], and results for individuals of Asian ancestry are depicted as green triangles[9,10].

We did not find nonsynonymous SNPs in *GALNT10*, *NFE2L3* or *KLHL32* that were correlated ($r^2 > 0.2$) with the most significant SNPs in the 1000 Genomes Project African-ancestry populations (AFR). However, we did detect a number of correlated SNPs ($r^2 > 0.5$) in regulatory sequences determined on the basis of overlapping chromatin marks in multiple cell types, including brain and adipose tissue (Online Methods). Many of these SNPs (or good proxies in the 1000 Genomes Project AFR, with an r^2 range of 0.59–1.0), which are located in putative enhancer and promoter regions, had only marginally weaker associations in stage 1 than the most significant SNPs reported in these regions (**Supplementary Tables 16–19** and **Supplementary Fig. 5a–c**). Together these data suggest that the biologically relevant variants in all three regions may be regulatory in function.

The variant rs7708584 at chromosome 5q33 is located upstream of *GALNT10* (encoding UDP-N-acetyl-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase 10), which catalyzes the first step in the synthesis of mucin-type oligosaccharides (**Supplementary Note**). GALNT10 is highly expressed in the small intestine and at intermediate levels in the stomach, pancreas, ovary, thyroid gland and spleen[29]. Suggestive associations between BMI and *GALNT10* have been observed in a smaller sample of African Americans[14] that was included in the present stage 1 meta-analysis, although the lead SNP differed (rs2033195) and showed only moderate LD ($r^2 = 0.27$) with the lead SNP we discovered here. The variant at 7p15, rs10261878, is intergenic and located 39 kb from a microRNA-encoding gene (*MIR148A*), which has been found to be significantly upregulated during adipogenesis[30], as well as in human adipocytes[31]. In addition, human miR-148a has been shown to regulate *CCKBR* (encoding cholecystokinin B receptor), which has been reported to have a regulatory role in the control of food intake[32]. The next closest gene (241 kb from rs10261878) is *NFE2L3* (encoding the nuclear factor (erythroid-derived 2)-like 3), a transcription factor that binds to antioxidant response elements of target genes and seems to have a role in differentiation, inflammation and carcinogenesis[33].

The most significant SNP at chromosome 6q16 (rs974417) is intronic in *KLHL32* (encoding kelch-like 32). Kelch-like genes have propeller domains that bind substrate proteins, promoting substrate ubiquitination, which modulates protein function. We also detected evidence of recent positive selection in and downstream of *KLHL32* (**Supplementary Figs. 6–9** and **Supplementary Note**).

In the largest GWAS meta-analysis of African-ancestry populations so far, we identified two new loci and one highly suggestive locus influencing BMI. The most informative SNPs in each of these three loci explain 0.10% of the variance in BMI in African-ancestry populations compared to 0.05% in Europeans and 0.03% in Asians (**Table 1** and **Supplementary Table 6**). Using the most significant ancestry-specific markers from each locus, the 36 known BMI loci explain 1.30% of the variance in BMI in men and women of African ancestry compared with 1.67% and 1.25% in European and Asian-ancestry populations, respectively (**Supplementary Tables 12** and **13**). We provide evidence for a shared genetic influence on BMI across populations, as we found directionally consistent associations with the majority of known BMI risk variants. This observation suggests that the biologically functional alleles are ancient and probably arose before migrations out of Africa. In addition, we were able to refine the window of association of some of the previously established BMI loci, which may eventually help identify the biologically functional variant(s). In this study we did not identify common variants for BMI that are likely to contribute to population differences in the prevalence of obesity. The ability to map new loci and replicate signals at established loci found in other populations reflects differences in allele frequency and effect size, which are influenced by population differences in recent demographic history and LD with the functional variant, as well as genetic and environmental modifying factors. Further studies will be needed to test the biologically functional alleles at the known loci, as well as the contribution of less common variation that have not yet been adequately surveyed by genome-wide SNP arrays. Taken together these findings demonstrate the importance of conducting genetic studies in diverse populations to identify new susceptibility loci for common traits.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

A full listing of acknowledgments is detailed in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Design and/or management of the individual studies: A.A., A.A.A., C.B.A., C.I.A., D.K.A., L.A., M.C.A., M.A.A., S.A., C.H.B., D.M.B., D.W.B., E.P.B., E.V.B., G.B., I.B.B., J.P.B., L.B., S.I.B., W.J.B., C.S.C., G. Casey, G.K.C., J.C., L.C., M.C., N.E.C., Q.C., R.S.C., S.J.C., J.D., P.D., R.W.D., S.L.D.-H., M.K.E., T.L.E., C.F., J.K.F., E.M.G., P.J.G., S.F.A.G., A.H., A.J.M.H., B.E.H., B.V.H., C.A.H., C.C.H., D.H., H.H., K.J.H., J.J.H., J.N.H., V.J.H., S.A.I., E.M.J., J.M.J., C.K., D.L.K., E.A.K., E.K.K., L.N.K., L.H.K., R.A.K., S.J.K., S. Kolb, L.L.M., A.M.L., R.J.F.L., S.L., Yongmei Liu, A.B.M., B.M., K.R.M., R.C.M., T.H.M., J.H.M., J.C.M., I.M.-B., K.E.N., M.C.Y.N., S.N., C.N.-D., U.N., B.N., K.L.N., T.O.O., O.O., O.I.O., B.P., U.P., B.M.P., C.A.P., G.J.P., J.R.P., M.F.P., P.A.P., S.R.P., E.A.R.-N., B.A.R., C.N.R., S.R., J.L.R.-G., A.B.S., A.G.S., J.L.S., L.B.S., P.J.S., S.B.S., S.W.-S., M.R.S., M.M.S., I.J.S., S.S.S., H.T., M.J.T., M.A.T., M.V., J.S.W., X.W., J.K.W., S.M.W., L.K.W., M.W., J.J.Y., N.A.Z., R.G.Z., W. Zheng, A.B.Z., K.A.Z., Y.Z. and X.Z.

Genotyping: A.B., U.B., S.J.C., Y.-D.I.C., D.D., T.L.E., S.F.A.G., X.G., D.G.H., J.N.H., T.D.H., T. Haritunians, K.C.J., Yongmei Liu, Y. Lu, W.M., R.N., J.R.P., N.D.P., S.B.S. and D.J.V.D.B.

Phenotyping: A.A.A., D.K.A., M.A.A., E.P.B., R.S.C., E.D., B.I.F., O.G., S.F.A.G., J.N.H., T. Haritunians, K.C.J., A.K., C.K., E.K.K., S.L., J.E.M., M.N., R.N., A.O., H.O.-B., B.M.P., J.R.P., S.R.P., C.N.R., E.R., S.R., B.S., D.S., L.S., B.O.T. and T.R.Y.

Statistical methods and data analysis: A.A.A., D.K.A., L.F.B., C.W.K.C., G.K.C., G. Casey, N.E.C., W.-M.C., G.A.C., Y.-D.I.C., J.D., P.D., T.L.E., C.F., M.F.F., J.P.B., E.M.G., M.G., O.G., X.G., C.A.H., M.R.I., A.K., B.J.K., C.K., E.K.K., S.J.K., C.D.L., G. Lettre, G. Li, H.L., K.L., L.A.L., R.J.F.L., V.L., Yongmei Liu, Youfang Liu, Y. Lu, B.M., K.L.M., Y.A.M., A.N., K.E.N., M.A.N., M.C.Y.N., C.P., J.R.P., E.A.R.-N., S.K.R., B.P., A.P.R., L.J.R.-T., D.A.S., E.K.S., E.E.S., Y.V.S., B.O.T., K.C.T., D.R.V.E., M.K.W., Z.W., L.K.W., T.W.W., L.R.Y., J.Z., J.H.Z., N.A.Z., J.M.Z., W. Zhao, the NABEC Consortium, the UKBEC Consortium, the BioBank Japan Project and the AGEN Consortium.

Writing group: G.K.C., T.L.E., M.G., B.E.H., J.N.H., R.J.F.L., C.A.H., L.A.L., K.L.M., K.E.N., M.C.Y.N., C.P., G.J.P., A.P.R. and K.C.T.

Critical review of the manuscript: A.A.A., A.A., C.B.A., C.I.A., D.K.A., L.A., M.A.A., M.C.A., S.A., A.B., C.H.B., D.M.B., D.W.B., E.P.B., E.V.B., G.B., I.B.B., J.P.B., L.F.B., L.B., S.I.B., U.B., W.J.B., C.W.K.C., C.S.C., F.C., G.A.C., G.K.C., G. Casey, G. Chen, J.C., L.C., M.C., N.E.C., Q.C., R.S.C., S.J.C., W.-M.C., Y.-D.I.C., D.D., E.D., J.D., P.D., R.W.D., S.L.D.-H., M.K.E., T.L.E., C.F., J.K.F., M.F.F., B.I.F., Y.F., E.M.G., G.S.G., M.G., O.G., P.J.G., S.F.A.G., W.T.G., X.G., A.J.M.H., A.H., B.E.H., B.V.H., C.A.H., C.C.H., D.G.H., D.H., H.H., J.J.H., J.N.H., K.J.H., T.D.H., T. Haritunians, T. Harris, V.J.H., J.H.M., M.R.I., S.A.I., E.M.J., J.M.J., K.C.J., A.K., B.J.K., C.K., D.L.K., E.A.K., E.K.K., L.N.K., L.H.K., R.A.K., S.J.K., S.L.R.K., S. Kolb, S. Ketkar, A.M.L., C.D.L., G. Lettre, G. Li, H.L., K.L., L.A.L., L.L., M.C.L., R.J.F.L., S.L., V.L., Yongmei Liu, Youfang Liu, Y. Lu, A.B.M., B.M., J.C.M., J.E.M., L.H.M., K.L.M., K.R.M., R.C.M., T.H.M., W.M., Y.A.M., I.M.-B., A.N., B.N., K.E.N., K.L.N., M.A.N., M.C.Y.N., M.N., R.N., S.N., U.N., C.N.-D., O.I.O., O.O., T.O.O., A.O., H.O.-B., B.M.P., B.P., C.A.P., C.P., G.J.P., J.R.P., M.F.P., N.D.P., P.A.P., S.R.P., U.P., A.P.R., B.A.R., C.N.R., E.R., S.K.R., S.R., J.L.R.-G., E.A.R.-N., L.J.R.-T., A.B.S., A.G.S.,

B.S., D.A.S., D.S., E.K.S., E.E.S., I.J.S., J.L.S., L.B.S., L.S., M.M.S., M.R.S., P.J.S., S.B.S., S.S.S., Y.V.S., B.O.T., K.C.T., M.A.T., H.T., M.J.T., M.V., D.J.V.D.B., D.R.V.E., J.K.W., M.W., S.M.W., S.-Y.W., S.W.-S., T.W.W., X.W., Z.W., L.K.W., J.S.W., M.K.W., J.J.Y., L.R.Y., T.R.Y., A.B.Z., J.Z., J.M.Z., J.H.Z., K.A.Z., N.A.Z., R.G.Z., W. Zhao, W Zheng, X.Z. and Y.Z.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Flegal, K.M., Carroll, M.D., Kit, B.K. & Ogden, C.L. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *J. Am. Med. Assoc.* **307**, 491–497 (2012).
2. Bradfield, J.P. *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).
3. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
4. Frayling, T.M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
5. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
6. Loos, R.J.F. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* **40**, 768–775 (2008).
7. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
8. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
9. Okada, Y. *et al.* Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat. Genet.* **44**, 302–306 (2012).
10. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* **44**, 307–311 (2012).
11. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat. Genet.* **40**, 716–718 (2008).
12. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* **41**, 157–159 (2009).
13. Scherag, A. *et al.* Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genet.* **6**, e1000916 (2010).
14. Ng, M.C. *et al.* Genome-wide association of BMI in African Americans. *Obesity* (*Silver Spring*) **20**, 622–627 (2012).
15. Kang, S.J. *et al.* Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum. Mol. Genet.* **19**, 2725–2738 (2010).
16. Shifman, S., Kuypers, J., Kokoris, M., Yakir, B. & Darvasi, A. Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* **12**, 771–776 (2003).
17. Campbell, M.C. & Tishkoff, S.A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
18. N'Diaye, A. *et al.* Identification, replication, and fine-mapping of loci associated with adult height in individuals of african ancestry. *PLoS Genet.* **7**, e1002298 (2011).
19. Heid, I.M. *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.* **42**, 949–960 (2010).
20. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
21. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
22. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
23. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
24. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
25. Nalls, M.A. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
26. Hernandez, D.G. *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.* **20**, 1164–1172 (2011).
27. Zhong, H., Yang, X., Kaplan, L.M., Molony, C. & Schadt, E.E. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am. J. Hum. Genet.* **86**, 581–591 (2010).
28. Myers, A.J. *et al.* A survey of genetic human cortical gene expression. *Nat. Genet.* **39**, 1494–1499 (2007).
29. Cheng, L. *et al.* Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T10. *FEBS Lett.* **531**, 115–121 (2002).
30. Xie, H., Lim, B. & Lodish, H.F. MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. *Diabetes* **58**, 1050–1057 (2009).
31. Ortega, F.J. *et al.* MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. *PLoS ONE* **5**, e9022 (2010).
32. Clerc, P. *et al.* Involvement of cholecystokinin 2 receptor in food intake regulation: hyperphagia and increased fat deposition in cholecystokinin 2 receptor-deficient mice. *Endocrinology* **148**, 1039–1049 (2007).
33. Chevillard, G. & Blank, V. NFE2L3 (NRF3): the Cinderella of the Cap'n'Collar transcription factors. *Cell Mol. Life Sci.* **68**, 3337–3348 (2011).

Keri L Monda[1,2,156], Gary K Chen[3,156], Kira C Taylor[2,4,156], Cameron Palmer[5–7,156], Todd L Edwards[8,156], Leslie A Lange[9], Maggie C Y Ng[10,11], Adebowale A Adeyemo[12], Matthew A Allison[13], Lawrence F Bielak[14], Guanjie Chen[12], Mariaelisa Graff[2], Marguerite R Irvin[15], Suhn K Rhie[3,16], Guo Li[17], Yongmei Liu[18], Youfang Liu[19], Yingchang Lu[20], Michael A Nalls[21], Yan V Sun[22], Mary K Wojczynski[23], Lisa R Yanek[24], Melinda C Aldrich[25–27], Adeyinka Ademola[28], Christopher I Amos[29,30], Elisa V Bandera[31], Cathryn H Bock[32], Angela Britton[21], Ulrich Broeckel[33], Quiyin Cai[25,26], Neil E Caporaso[34], Chris S Carlson[35], John Carpten[36], Graham Casey[3,16], Wei-Min Chen[37], Fang Chen[37], Yii-Der I Chen[38], Charleston W K Chiang[39], Gerhard A Coetzee[3,16,40], Ellen Demerath[41], Sandra L Deming-Halverson[25,26], Ryan W Driver[42], Patricia Dubbert[43–45], Mary F Feitosa[23], Ye Feng[3], Barry I Freedman[46], Elizabeth M Gillanders[47], Omri Gottesman[20], Xiuqing Guo[38], Talin Haritunians[38], Tamara Harris[48], Curtis C Harris[49], Anselm J M Hennis[50–53], Dena G Hernandez[21], Lorna H McNeill[54,55], Timothy D Howard[11], Barbara V Howard[56,57], Virginia J Howard[15], Karen C Johnson[58], Sun J Kang[59], Brendan J Keating[60], Suzanne Kolb[35], Lewis H Kuller[61], Abdullah Kutlar[62], Carl D Langefeld[63], Guillaume Lettre[64], Kurt Lohman[63], Vaneet Lotay[20], Helen Lyon[65,66], JoAnn E Manson[67], William Maixner[68], Yan A Meng[69], Kristine R Monroe[3], Imran Morhason-Bello[70], Adam B Murphy[71], Josyf C Mychaleckyj[37], Rajiv Nadukuru[20], Katherine L Nathanson[72], Uma Nayak[37], Amidou N'Diaye[65], Barbara Nemesure[50], Suh-Yuh Wu[50], M Cristina Leske[50], Christine Neslund-Dudas[73], Marian Neuhouser[74], Sarah Nyante[75,76], Heather Ochs-Balcom[77], Adesola Ogunniyi[78], Temidayo O Ogundiran[28], Oladosu Ojengbede[70], Olufunmilayo I Olopade[79], Julie R Palmer[80], Edward A Ruiz-Narvaez[80], Nicholette D Palmer[10,11], Michael F Press[81], Evandine Rampersaud[82], Laura J Rasmussen-Torvik[83], Jorge L Rodriguez-Gil[84,85], Babatunde Salako[78], Eric E Schadt[86,87], Ann G Schwartz[32], Daniel A Shriner[10], David Siscovick[17,88],

Shad B Smith[68], Sylvia Wassertheil-Smoller[89], Elizabeth K Speliotes[90–93], Margaret R Spitz[94], Lara Sucheston[95], Herman Taylor[96], Bamidele O Tayo[97], Margaret A Tucker[34], David J Van Den Berg[3,16], Digna R Velez Edwards[98,99], Zhaoming Wang[100], John K Wiencke[101,102], Thomas W Winkler[20,103], John S Witte[104], Margaret Wrensch[101,102], Xifeng Wu[29], James J Yang[73], Albert M Levin[73], Taylor R Young[93], Neil A Zakai[105], Mary Cushman[105], Krista A Zanetti[47,49], Jing Hua Zhao[106], Wei Zhao[107], Yonglan Zheng[79], Jie Zhou[10], Regina G Ziegler[34], Joseph M Zmuda[61], Jyotika K Fernandes[108], Gary S Gilkeson[108], Diane L Kamen[108], Kelly J Hunt[108], Ida J Spruill[109], Christine B Ambrosone[95], Stefan Ambs[49], Donna K Arnett[15], Larry Atwood[110,155], Diane M Becker[24], Sonja I Berndt[34], Leslie Bernstein[111], William J Blot[25,26,112], Ingrid B Borecki[23], Erwin P Bottinger[20], Donald W Bowden[10,11,46], Gregory Burke[113], Stephen J Chanock[34], Richard S Cooper[97], Jingzhong Ding[114], David Duggan[36], Michele K Evans[115], Caroline Fox[116–118], W Timothy Garvey[119], Jonathan P Bradfield[120,121], Hakon Hakonarson[120–123], Struan F A Grant[120–123], Ann Hsing[124–126], Lisa Chu[126], Jennifer J Hu[84,85], Dezheng Huo[127], Sue A Ingles[3], Esther M John[124–126], Joanne M Jordan[19], Edmond K Kabagambe[25], Sharon L R Kardia[107], Rick A Kittles[128], Phyllis J Goodman[129], Eric A Klein[130], Laurence N Kolonel[131], Loic Le Marchand[131], Simin Liu[132], Barbara McKnight[17,133], Robert C Millikan[75,76,155], Thomas H Mosley[96], Badri Padhukasahasram[134], L Keoki Williams[134,135], Sanjay R Patel[136], Ulrike Peters[35], Curtis A Pettaway[137], Patricia A Peyser[107], Bruce M Psaty[17,138,139], Susan Redline[140], Charles N Rotimi[12], Benjamin A Rybicki[73], Michèle M Sale[141–143], Pamela J Schreiner[41], Lisa B Signorello[144], Andrew B Singleton[21], Janet L Stanford[35], Sara S Strom[29], Michael J Thun[42], Mara Vitolins[113], Wei Zheng[25,26], Jason H Moore[145–147], Scott M Williams[145,146], Shamika Ketkar[148], Xiaofeng Zhu[149], Alan B Zonderman[150], NABEC Consortium[151], UKBEC Consortium[151], BioBank Japan Project[151], AGEN Consortium[151], Charles Kooperberg[35], George J Papanicolaou[152], Brian E Henderson[3,16], Alex P Reiner[35], Joel N Hirschhorn[5–7,39], Ruth J F Loos[20,106,153,154], Kari E North[2,157] & Christopher A Haiman[3,16,157]

[1]The Center for Observational Research, Amgen, Inc. Thousand Oaks, California, USA. [2]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [3]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [4]Department of Epidemiology and Population Health, University of Louisville, Louisville, Kentucky, USA. [5]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [6]Division of Genetics, Children's Hospital, Boston, Massachusetts, USA. [7]Division of Endocrinology, Children's Hospital, Boston, Massachusetts, USA. [8]Center for Human Genetics Research, Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, Tennessee, USA. [9]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA. [10]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [11]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [12]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health (NIH), Bethesda, Maryland, USA. [13]Department of Family and Preventive Medicine, University of California San Diego, La Jolla, California, USA. [14]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. [15]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA. [16]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA. [17]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [18]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [19]Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [20]The Charles Bronfman Institute of Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [21]Laboratory of Neurogenetics, National Institute on Aging (NIA), NIH, Bethesda, Maryland, USA. [22]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA. [23]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [24]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [25]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [26]The Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [27]Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [28]Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria. [29]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [30]Department of Family and Community Medicine, Geissel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA. [31]The Cancer Institute of New Jersey, New Brunswick, New Jersey, USA. [32]Karmanos Cancer Institute, Wayne State University, Detroit, Michigan, USA. [33]Department of Pediatrics, Section of Genomic Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. [34]Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), NIH, Bethesda, Maryland, USA. [35]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [36]The Translational Genomics Research Institute, Phoenix, Arizona, USA. [37]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA. [38]Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, California, USA. [39]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [40]Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [41]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, Minnesota, USA. [42]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. [43]South Central VA Mental Illness, Research and Clinical Center, Little Rock, Arkansas, USA. [44]VA Geriatric Research, Education and Clinical Center, Little Rock, Arkansas, USA. [45]Division of Health Services Research, Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA. [46]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [47]Division of Cancer Control and Population Sciences, NCI, NIH, Bethesda, Maryland, USA. [48]Laboratory of Epidemiology and Population Science, NIA, NIH, Bethesda, Maryland, USA. [49]Laboratory of Human Carcinogenesis, Center for Cancer Research, NCI, NIH, Bethesda, Maryland, USA. [50]Department of Preventive Medicine, Stony Brook University, Stony Brook, New York, USA. [51]Chronic Disease Research Centre, University of the West Indies, Bridgetown, Barbados. [52]Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados. [53]Ministry of Health, Bridgetown, Barbados. [54]Department of Health Disparities Research, Division of the Office of the Vice President (OVP), Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [55]Center for Community-Engaged Translational Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [56]MedStar Health Research Institute, Washington DC, USA. [57]Georgetown-Howard Universities Center for Clinical and Translational Sciences, Washington, DC, USA. [58]Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA. [59]Henri Begleiter Neurodynamics Laboratory, Department of

Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center, Brooklyn, New York, USA. [60]Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [61]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [62]Sickle Cell Center, Department of Medicine, Georgia Health Sciences University, Augusta, Georgia, USA. [63]Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [64]Montreal Heart Institute, Université de Montréal, Montréal, Québec, Canada. [65]Program in Genomics and Endocrinology, Division of Genetics, Children's Hospital, Boston, Massachusetts, USA. [66]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA. [67]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [68]Regional Center for Neurosensory Disorders, School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [69]Broad Institute of Harvard and MIT, Metabolic Disease Initiative, Cambridge, Massachusetts, USA. [70]Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria. [71]Department of Urology, Northwestern University, Chicago, Illinois, USA. [72]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [73]Department of Public Health Sciences, Henry Ford Hospital, Detroit, Michigan, USA. [74]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [75]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA. [76]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. [77]Department of Social and Preventive Medicine, University at Buffalo, Buffalo, New York, USA. [78]Department of Medicine, University of Ibadan, Ibadan, Nigeria. [79]Department of Medicine, University of Chicago, Chicago, Illinois, USA. [80]Slone Epidemiology Center at Boston University, Boston, Massachusetts, USA. [81]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [82]Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, Florida, USA. [83]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. [84]Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida, USA. [85]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida, USA. [86]Department of Genetics and Genomic Sciences, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [87]Icahn Institute for Genomics and Multiscale Biology, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [88]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [89]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA. [90]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA. [91]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [92]Division of Gastroenterology. Massachusetts General Hospital, Boston, Massachusetts, USA. [93]The Broad Institute, Cambridge, Massachusetts, USA. [94]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas, USA. [95]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA. [96]Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA. [97]Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, USA. [98]Center for Human Genetics Research, Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, Tennessee, USA. [99]Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, Tennessee, USA. [100]Core Genotype Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland, USA. [101]Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA. [102]Department of Neurological Surgery, University of California San Francisco, San Francisco, California, USA. [103]Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany. [104]Department of Epidemiology and Biostatistics, Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA. [105]Department of Medicine, University of Vermont College of Medicine, Burlington, Vermont, USA. [106]Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [107]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. [108]Department of Medicine, Medical University of South Carolina, Charleston, South Carolina, USA. [109]College of Nursing, Medical University of South Carolina, Charleston, South Carolina, USA. [110]Framingham Heart Study, Boston University School of Medicine, Boston, Massachusetts, USA. [111]Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, Duarte. California, USA. [112]International Epidemiology Institute, Rockville, Maryland, USA. [113]Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA. [114]Section on Gerontology and Geriatric Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [115]Health Disparities Research Section, Laboratory of Population Science, NIA, NIH, Baltimore, Maryland, USA. [116]National Heart, Lung and Blood Institute (NHLBI) Framingham Heart Study, Framingham, Massachusetts, USA. [117]NHLBI Center for Population Studies, Framingham, Massachusetts, USA. [118]Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [119]Department of Nutrition Sciences, University of Alabama at Birmingham and the Birmingham VA Medical Center, Birmingham, Alabama, USA. [120]The Center for Applied Genomics, The Children's Hospital of Philadelphia Research Institute, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [121]Division of Human Genetics, The Children's Hospital of Philadelphia Research Institute, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [122]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [123]Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [124]Cancer Prevention Institute of California, Fremont, California, USA. [125]Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA. [126]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA. [127]Department of Health Studies, University of Chicago, Chicago, Illinois, USA. [128]Department of Medicine, University of Illinois at Chicago, Chicago, Illinois, USA. [129]Southwest Oncology Group (SWOG) Statistical Center, Seattle, Washington, USA. [130]Glickman Urologic and Kidney Institute, Cleveland Clinic, Cleveland, Ohio, USA. [131]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA. [132]Department of Epidemiology, School of Public Health, University of California at Los Angeles, Los Angeles, California, USA. [133]Department of Biostatistics, University of Washington, Seattle, Washington, USA. [134]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA. [135]Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan, USA. [136]Division of Sleep Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [137]Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [138]Department of Epidemiology and Health Services, University of Washington, Seattle, Washington, USA. [139]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [140]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [141]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA. [142]Department of Medicine, University of Virginia, Charlottesville, Virginia, USA. [143]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, USA. [144]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [145]Institute for Quantitative Biomedical Sciences, The Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA. [146]Department of Genetics, The Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA. [147]Department of Community and Family Medicine, The Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA. [148]Division of Biology and Biomedical Sciences, Human and Statistical Genetics Program, Washington University, St. Louis, Missouri, USA. [149]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA. [150]Behavioral Epidemiology Section, Laboratory of Population Science, NIA, NIH, Baltimore, Maryland, USA. [151]A list of contributing members appears in the **Supplementary Note**. [152]Division of Prevention and Population Sciences, NHLBI, NIH, Bethesda, Maryland, USA. [153]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [154]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [155]Deceased. [156]These authors contributed equally to this work. [157]These authors jointly directed this work. Correspondence should be addressed to K.E.N. (kari_north@unc.edu) or C.A.H. (haiman@usc.edu).

ONLINE METHODS

Study design. We used a three-stage design consisting of a GWAS meta-analysis (stage 1), a follow-up of 1,500 SNPs (stage 2) and a focused follow-up of the three new loci (stage 3). Stage 1 included results from 36 GWAS of 39,144 men and women of African ancestry (37,956 African American and 1,188 African; **Supplementary Table 1**). We took forward the 1,500 most significantly associated SNPs ($P < 0.0003$) for examination in 6,817 additional men and women of African ancestry from seven GWAS (stage 2, all African American). The three SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) after the meta-analysis of the results from stages 1 and 2 were taken forward for further confirmation in 25,451 additional African-ancestry subjects from 12 studies. All participants in these studies provided written informed consent for the research, and approval for the study was obtained from the ethics review boards at all participating institutions. A description of each participating study as well as details regarding the measurement and collection of height and weight data are provided in the **Supplementary Note**.

Genotyping and quality control. Genotyping in each study was conducted using Illumina or Affymetrix genome-wide SNP arrays. The size of each study ranged from 50 to 8,421 individuals. The details of the array, genotyping quality-control procedures and sample exclusions for each study that contributed data are listed in **Supplementary Tables 1** and **2**.

Statistical analyses. In all GWAS, imputation to phased haplotype data from the founders of the CEU and YRI HapMap phase 2 samples (build 21) was performed using MACH[34], IMPUTE2 (ref. 35) or BEAGLE[36]. SNPs with lower imputation quality scores ($r^2 < 0.3$) (**Supplementary Table 2**) as well as SNPs with a small number of allele counts after stratifying by sex and case-control status were excluded from the analyses. Local ancestry, defined as the number of European chromosomes (continuous between 0 and 2), was estimated for the majority of the stage 1 African-ancestry studies (**Supplementary Table 8**) using HAPMIX[37]. To evaluate the effect of admixture on the allele distribution between the African and European segments, we stratified the analysis of each variant by local ancestry at each locus (**Supplementary Table 9**).

Stage 1. Genome-wide association analyses were performed by each of the participating studies. BMI was regressed on age, age squared and study site (if needed) to obtain residuals, separately by sex and case-control status, if needed. Residuals were inverse-normally transformed to obtain a standard normal distribution with a mean of 0 and an s.d. of 1. For studies with unrelated subjects, each SNP was tested for additive association with BMI by regressing the transformed residuals on the number of copies of the SNP effect allele adjusting for population structure as measured by the first ten eigenvectors calculated for each study. Analyses were stratified by sex and case-control status (if needed). For studies that included related individuals, family based association tests were conducted that took into consideration the genetic relationships among the individuals. Study-specific λ values ranged from 0.95 to 1.08 (**Supplementary Table 2**). We applied genomic control in the stage 1 analysis (that is, we divided by the median of all χ^2 statistics for each study) to eliminate any remaining overdispersion before combining the GWAS in the meta-analysis. In stage 1, we conducted a fixed-effect meta-analysis using the inverse variance–weighted method implemented in the program METAL[38]. We performed a second genomic control correction of the stage 1 meta-analysis results ($\lambda = 1.136$) before selecting SNPs for follow-up.

Stages 2 and 3. The 1,500 most significant SNPs from stage 1 were examined in an additional 6,817 individuals, with each SNP being analyzed as described for stage 1 and meta-analyzed using the inverse-variance method using METAL. As in stage 1, each SNP was tested for association with BMI by regressing the transformed residuals on the number of copies of the SNP effect allele adjusting for population structure as measured by the first ten eigenvectors calculated for each study. Further testing of the three new variants was conducted in an additional 25,451 individuals (stage 3). Results from all stages were meta-analyzed using the inverse-variance method in METAL.

Examination in individuals of European ancestry. We also examined the 1,500 most statistically significant SNPs from stage 1 in the GIANT consortium ($n = 123{,}706$ individuals of European ancestry)[7]. Of these, 1,390 were genotyped or imputed in GIANT, and 1,328 had data for $n > 50{,}000$ individuals and MAF > 1%. We conducted a meta-analysis of stages 1, 2 and 3 plus GIANT in the same manner as described above. The three new variants were also examined in the AGEN and RIKEN consortia[9,10] and the Pediatric Research Consortium (PeRC) (**Supplementary Note**).

Estimation of variance explained. The total fraction of variance explained was calculated using the formula $2f(1 - f) \times a^2$, where f is the frequency of the variant, and a is the additive effect of the variant[3]. When calculating percentage variance explained in the African-ancestry sample for the previously discovered BMI variants that were not genome-wide significant in stage 1, we used data from the stage 1 sample; for those that were genome-wide significant in stage 1, we used data from the stage 2 sample; and for the new BMI variants, we used data from the combined stage 2 and 3 samples to avoid inflating the estimates as a result of winner's curse. When summing percentage variance explained for the 36 previously discovered BMI variants (**Supplementary Table 12**), we used the more informative SNP discovered through fine mapping at the seven loci (**Supplementary Table 13**). However, for these seven variants, the stage 1 results were used, and estimates may be biased; stage 2 and 3 studies only participated in the look-up of the top SNPs from the preceding stages.

Bioinformatic analysis of the new BMI loci. In an attempt to identify functionality in noncoding regions at the three loci, we used FunciSNP version 0.99 (ref. 39), which systematically integrates the 1,000 Genomes SNP data (1KGP, April 2012) with chromatin features of interest. To capture regulatory elements, we used 73 different chromatin features generated by next-generation sequencing technologies in brain and adipose tissues from the NIH Epigenomics Roadmap[40], as well as known DNaseI hypersensitive locations, formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) peaks and CTCF binding sites from more than 100 different cell types, which were collected from the ENCODE data[41]. All SNPs with $r^2 > 0.5$ with each index SNP in the 1KGP AFR populations in a 1-Mb window around each index variant were cataloged. We used the UCSC Genome Browser (http://genome.ucsc.edu/) to illustrate the correlated SNPs that overlap chromatin features from these tissues, as well as chromatin features from seven cell lines used in the ENCODE Project (**Supplementary Fig. 5a–c**). The results from these analyses are provided in **Supplementary Tables 16–19**.

eQTL analyses. *Liver, subcutaneous fat and omental fat tissue.* The determination of eQTLs in liver, subcutaneous fat and omental fat tissue have been described in detail previously[27]. In brief, liver, subcutaneous fat and omental fat tissue were obtained from patients of European ancestry who underwent bariatric surgery. Expression of a total of 39,280 oligonucleotide probes targeting transcripts representing 34,266 known and predicted genes was assessed. All patients were genotyped on a genome-wide SNP array, and association between SNPs and gene expression data was adjusted for age, race, gender and surgery year using linear regression. Results are presented in **Supplementary Table 14** and **Supplementary Figure 4**.

Brain cortical tissue. We examined the *cis* associations (defined as genes within 1 Mb) between each of the BMI-associated SNPs and expression of nearby genes in brain (cortical tissue)[28]. The eQTL analyses have been described in detail previously (Gene Expression Omnibus (GEO) database GSE8919)[28]. In brief, DNA and RNA of neuropathologically normal cortical brain samples of 193 individuals (average age (range), 81 (65–100) years) of European ancestry were isolated and genotyped for a genome-wide SNP array, and HapMap genotypes were imputed. RNA expression was assessed for 24,357 transcripts, of which 14,078 transcripts met the quality-control criteria. Association analyses between SNPs and expression data assumed an additive model and were adjusted for sex and age at death. Results are presented in **Supplementary Table 14** and **Supplementary Figure 4**.

Association testing of previously established BMI loci. To characterize alleles that might better represent the biologically functional variant at the 36 previously discovered BMI loci, we searched for LD proxies among individuals

doi:10.1038/ng.2608

of African ancestry. Using HapMap data (CEU or JPT/CHB) to estimate LD, we identified all SNPs that were correlated ($r^2 \geq 0.4$) with the index SNP (within 250 kb or larger to include a nearby gene). Next we tested these SNPs for association with BMI in the stage 1 African-ancestry sample. We applied a locus-specific significance criterion, α, which accounts for multiple testing (the number of tag SNPs in the HapMap YRI population that capture (at $r^2 \geq 0.8$) all common SNPs (with MAF ≥ 0.05) correlated with the index signal in the HapMap CEU or JPT/CHB populations). This α level does not account for the number of regions evaluated and reflects a balance between the need to correct for multiple comparisons and the prior knowledge that each region harbors a risk variant for BMI. We also looked for new independent associations, focusing on the genotyped and imputed SNPs that were uncorrelated with the index signal in the initial GWAS populations ($r^2 < 0.2$). We applied a Bonferroni correction for defining new associations as significant in each region as 0.05 divided by the total number of tags needed to capture (at $r^2 \geq 0.8$) all common risk alleles across all risk regions in the YRI population ($\alpha = 6.7 \times 10^{-6}$).

Detection of recent positive selection in Africans and Europeans at a new BMI locus. We evaluated the evidence for recent positive selection at our new loci using several statistical techniques, the BioVU African-American GWAS data and data from the International HapMap Project and the Human Genome Diversity Project (HGDP). We compared adjusted allele frequencies among BioVU and HapMap phase 3 participants from the west African Yoruban (YRI) and east African Luhya (LWK) populations using Treeselect[42]. The LWK sample is differentiated from the YRI sample and samples of African Americans[43]. Allele frequencies in the African-American sample were adjusted by subtracting the expected contribution of European alleles, where p_{AA} is the allele frequency in African Americans obtained from experimental data, p_{EA} is the allele frequency in Europeans obtained from HapMap, p_{AF} is the estimated allele frequency in African founders and α is the average proportion of ancestry from Europeans, or 0.2. The adjustment is then performed by solving the following expression for p_{AF}:

$$p_{AF} = \frac{p_{AA} - \alpha p_{EA}}{(1-\alpha)}$$

We also evaluated the HapMap phase 2 and HGDP data with the integrated haplotype score (iHS)[44] and Haplotter and the crosspopulation-extended haplotype homozygosity (XP-EHH) statistic using the HGDP selection browser[45,46]. We also evaluated the BioVU data using 5,000 random autosomal SNPs with STRUCTURE v2.3.3, and on average, the participants were 20.7% European and 79.3% of African ancestry[47,48].

We observed evidence for recent selection near *KLHL32* within the YRI HapMap data using iHS (**Supplementary Fig. 4**) and in the HGPD African participants (**Supplementary Fig. 5a–d**). Nominal evidence of selection was observed within the YRI and African-American populations using the Treeselect statistic with the transcription factor binding–site SNP rs1206131 ($P = 0.003$ in the African Americans and $P = 0.005$ in YRI) and at the SNP rs9387284 ($P = 0.004$ in the YRI and $P = 0.026$ in the African Americans) (**Supplementary Fig. 6a**,**b**). The Treeselect method also demonstrated a significant allele frequency differentiation between African and African-ancestry populations (F_{st} ~0.01) at the transcription factor binding–site SNP rs1206131. rs1206131 is the most significant SNP for this test in the region ±400 kb. The test from the African-American branch of the tree was slightly less significant at rs1206131, and the most significant SNP was downstream, which is also under the iHS and XP-EHH peaks from Africans in the HGDP and HapMap data. The graph of HGDP allele frequencies at this SNP shows that the ancestral T allele has increased frequencies throughout Africa relative to other major global populations (**Supplementary Fig. 7**). The average (s.d., maximum) F_{st} value in this region between the YRI and African-American populations was 0.001 (0.001, 0.015), between the YRI and CEU populations was 0.040 (0.045, 0.304) and between the African-American and CEU populations was 0.011 (0.013, 0.082).

34. 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).
35. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
36. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
37. Price, A.L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
38. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
39. Coetzee, S.G., Rhie, S.K., Berman, B.P., Coetzee, G.A. & Noushmehr, H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res.* **40**, e139 (2012).
40. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
41. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
42. Bhatia, G. *et al.* Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *Am. J. Hum. Genet.* **89**, 368–381 (2011).
43. Murray, T. *et al.* African and non-African admixture components in African Americans and an African Caribbean population. *Genet. Epidemiol.* **34**, 561–568 (2010).
44. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
45. Pickrell, J.K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
46. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
47. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
48. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).

A Meta-Analysis Identifies New Loci Associated with Body Mass Index in Individuals of African Ancestry

Supplementary Tables and Figures and Supplementary Note

Keri L. Monda[1,2]*, Gary K. Chen[3]*, Kira C. Taylor[2,4]*, Cameron Palmer[5,6,7]*, Todd L. Edwards[8]*, Leslie A. Lange[9], Maggie C.Y. Ng[10,11], Adebowale A. Adeyemo[12], Matthew A. Allison[13], Lawrence F. Bielak[14], Guanji Chen[12], Mariaelisa Graff[2], Marguerite R. Irvin[15], Suhn K. Rhie[3,16], Guo Li[17], Yongmei Liu[18], Youfang Liu[19], Yingchang Lu[20], Michael A. Nalls[21], Yan V. Sun[22], Mary K. Wojczynski[23], Lisa R. Yanek[24], Melinda C. Aldrich[25,26,27], Adeyinka Ademola[28], Christopher I. Amos[29,30], Elisa V. Bandera[31], Cathryn H. Bock[32], Angela Britton[21], Ulrich Broeckel[33], Quiyin Cai[25,26], Neil E. Caporaso[34], Chris S. Carlson[35], John Carpten[36], Graham Casey[3,16], Wei-Min Chen[37], Fang Chen[37], Yii-Der I. Chen[38], Charleston W.K. Chiang[39], Gerhard A. Coetzee[3,16,40], Ellen Demerath[41], Sandra L. Deming-Halverson[25,26], Ryan W. Driver[42], Patricia Dubbert[43,44,45], Mary F. Feitosa[23], Ye Feng[3], Barry I. Freedman[46], Elizabeth M. Gillanders[47], Omri Gottesman[20], Xiuqing Guo[38], Talin Haritunians[38], Tamara Harris[48], Curtis C. Harris[49], Anselm JM Hennis[50,51,52,53], Dena G. Hernandez[21], Lorna H. McNeill[54,55], Timothy D. Howard[56], Barbara V. Howard[57,58], Virginia J. Howard[15], Karen C. Johnson[59], Sun J. Kang[60], Brendan J. Keating[61], Suzanne Kolb[35], Lewis H. Kuller[62], Abdullah Kutlar[63], Carl D. Langefeld[64], Guillaume Lettre[65], Kurt Lohman[64], Vaneet Lotay[20], Helen Lyon[66,67], JoAnn E. Manson[68], William Maixner[69], Yan A. Meng[70], Kristine R. Monroe[3], Imran Morhason-Bello[71], Adam B. Murphy[72], Josyf C.Mychaleckyj[37], Rajiv Nadukuru[20], Katherine L. Nathanson[73], Uma Nayak[37], Amidou N'Diaye[65], Barbara Nemesure[50], Suh-Yuh Wu[50], M. Cristina Leske[50], Christine Neslund-Dudas[74], Marian Neuhouser[75], Sarah Nyante[76,77], Heather Ochs-Balcom[78], Adesola Ogunniyi[79], Temidayo O. Ogundiran[28], Oladosu Ojengbede[71], Olufunmilayo I.Olopade[80], Julie R. Palmer[81], Edward A. Ruiz-Narvaez[81], Nicholette D. Palmer[10,11], Michael F. Press[82], Evandine Rampersaud[83], Laura J. Rasmussen-Torvik[84], Jorge L. Rodriguez-Gil[85,86], Babatunde Salako[79], Eric E Schadt[87,88], Ann G. Schwartz[32], Daniel A. Shriner[10], David Siscovick[17,89], Shad B. Smith[69], Sylvia Wassertheil-Smoller[90], Elizabeth K. Speliotes[91,92,93,94], Margaret R. Spitz[95], Lara Sucheston[96], Herman Taylor[97], Bamidele O. Tayo[98], Margaret A. Tucker[34], David J. Van Den Berg[3,16], Digna R. Velez Edwards[99,100], Zhaoming Wang[101], John K. Wiencke[102,103], Thomas W. Winkler[20,104], John S. Witte[105], Margaret Wrensch[102,103], Xifeng Wu[29], James J. Yang[74], Albert M. Levin[74], Taylor R. Young[94], Neil A. Zakai[106], Mary Cushman[106], Krista A. Zanetti[47,49], Jing Hua Zhao[107], Wei Zhao[108], Yonglan Zheng[80], Jie Zhou[10], Regina G. Ziegler[34], Joseph M. Zmuda[62], Jyotika K. Fernandes[109], Gary S. Gilkeson[109], Diane L. Kamen[109], Kelly J. Hunt[109], Ida J. Spruill[110], Christine B. Ambrosone[96], Stefan Ambs[49], Donna K. Arnett[15], Larry Atwood[111,**], Diane M. Becker[112], Sonja I. Berndt[34], Leslie Bernstein[113], William J. Blot[25,26,114], Ingrid B. Borecki[23], Erwin P. Bottinger[20], Donald W. Bowden[10,11,115], Gregory Burke[116], Stephen J. Chanock[34], Richard S. Cooper[98], Jingzhong Ding[117], David Duggan[36], Michele K. Evans[118], Caroline Fox[119,120,121], W. Timothy Garvey[122], Jonathan P. Bradfield[123,124], Hakon Hakonarson[123,124,125,126], Struan F.A. Grant[123,124,125,126], Ann Hsing[127,128,129], Lisa Chu[127], Jennifer J. Hu[85,86], Dezheng Huo[130], Sue A. Ingles[3], Esther M. John[127,128,129], Joanne M. Jordan[19], Edmond K. Kabagambe[25], Sharon L.R. Kardia[108], Rick A. Kittles[131], Phyllis J. Goodman[132], Eric A. Klein[133], Laurence N. Kolonel[134], Loic Le Marchand[134], Simin Liu[135], Barbara McKnight[17,136], Robert C. Millikan[76,77**], Thomas H. Mosley[97], Badri Padhukasahasram[137], L. Keoki Williams[137,138], Sanjay R. Patel[139], Ulrike Peters[35], Curtis A. Pettaway[140], Patricia A. Peyser[108], Bruce M. Psaty[17,141, 142], Susan Redline[143], Charles N. Rotimi[10], Benjamin A. Rybicki[74], Michèle M. Sale[144,145,146], Pamela J. Schreiner[41], Lisa B. Signorello[147], Andrew B. Singleton[21], Janet L. Stanford[35], Sara S. Strom[29], Michael J. Thun[42], Mara Vitolins[116], Wei Zheng[25,26], Jason H. Moore[148,149,150], Scott M. Williams[148,149], Shamika Ketkar[151], Xiaofeng Zhu[152], Alan B. Zonderman[153], NABEC Consortium[154], UKBEC Consortium[154], BioBank Japan Project[154], AGEN Consortium[154], Charles

Nature Genetics: doi:10.1038/ng.2608

Kooperberg[35], George J. Papanicolaou[155], Brian E. Henderson[3,16], Alex P. Reiner[35], Joel N. Hirschhorn[5,6,7,39], Ruth JF Loos[20,107,156,157], Kari E. North[2,*†], Christopher A. Haiman[3,16*†]

*The authors wish it to be known that, in their opinion, the first five authors should be regarded as joint First Authors.

†These Authors directed the study and contributed equally to the study and are to whom correspondence should be addressed:

Christopher Haiman at Department of Preventive Medicine, USC/Norris Comprehensive Cancer Center, Harlyne Norris Research Tower, 1450 Biggy Street, Room 1504, Los Angeles, CA 90033, USA. Tel: +1 3234427755, Fax: +1 3234427749, E-mail: haiman@usc.edu.

Kari North at Department of Epidemiology, UNC Gillings School of Global Public Health, and Carolina Center for Genome Sciences, 137 E Franklin St., Suite 306, Chapel Hill, NC 27514, USA. Tel: +1 9199662148, Fax: +1 9199669800, E-mail: kari_north@unc.edu.

AFFILIATIONS

[1]The Center for Observational Research, Amgen, Inc. Thousand Oaks, CA, USA

[2]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[4]Department of Epidemiology and Population Health, University of Louisville, Louisville, KY, USA

[5]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[6]Division of Genetics, Children's Hospital, Boston, MA, USA

[7]Division of Endocrinology, Children's Hospital, Boston, MA, USA

[8]Center for Human Genetics Research, Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN, USA

[9]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

[10]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

[11]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

[12]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[13]Department of Family and Preventive Medicine, University of California San Diego, La Jolla, CA, USA

[14]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

Nature Genetics: doi:10.1038/ng.2608

[15]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

[16]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA

[17]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

[18]Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA

[19]Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[20]The Charles Bronfman Institute of Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, USA

[21]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

[22]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

[23]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

[24]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[25]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA

[26]The Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA

[27]Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, TN, USA

[28]Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria

[29]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[30]Department of Family and Community Medicine, Geissel School of Medicine, Dartmouth, Hanover, NH

[31]The Cancer Institute of New Jersey, New Brunswick, NJ, USA

[32]Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA

[33]Department of Pediatrics, Section of Genomic Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA

[34]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Nature Genetics: doi:10.1038/ng.2608

[35]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[36]The Translational Genomics Research Institute, Phoenix, AZ, USA

[37]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

[38]Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA

[39]Department of Genetics, Harvard Medical School, Boston, MA, USA

[40]Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[41]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA

[42]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

[43]South Central VA Mental Illness, Research, and Clinical Center, Little Rock, AR, USA

[44]VA Geriatric Research, Education, and Clinical Center, Little Rock, AR, USA

[45]Division of Health Services Research, Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[46]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[47]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health Bethesda, MD, USA

[48]Laboratory of Epidemiology and Population Science, National Institutes on Aging, National Institutes of Health, Bethesda, MD, USA

[49]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

[50]Department of Preventive Medicine, Stony Brook University, Stony Brook, NY, USA

[51]Chronic Disease Research Centre, University of the West Indies, Bridgetown, Barbados

[52]Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados

[53]Ministry of Health, Bridgetown, Barbados

[54]Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[55]Center for Community-Engaged Translational Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Nature Genetics: doi:10.1038/ng.2608

[56]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

[57]MedStar Health Research Institute, Washington, DC, USA

[58]Georgetown-Howard Universities Center for Clinical and Translational Sciences, Washington, DC, USA

[59]Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA

[60]Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA

[61]Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA, USA

[62]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, USA

[63]Sickle Cell Center, Department of Medicine, Georgia Health Sciences University, Augusta, GA, USA

[64]Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA

[65]Montreal Heart Institute, Université de Montréal, Montréal, Québec, Canada

[66]Program in Genomics and Endocrinology, Division of Genetics, Children's Hospital, Boston, MA, USA

[67]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

[68]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[69]Regional Center for Neurosensory Disorders, School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[70]Broad Institute of Harvard and MIT, Metabolic Disease Initiative, Cambridge, MA, USA

[71]Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria

[72]Department of Urology, Northwestern University, Chicago, IL, USA

[73]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[74]Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI, USA

[75]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Nature Genetics: doi:10.1038/ng.2608

[76]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[77]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

[78]Department of Social and Preventive Medicine, University at Buffalo, Buffalo NY, USA

[79]Department of Medicine, University of Ibadan, Ibadan, Nigeria

[80]Department of Medicine, University of Chicago, Chicago, IL, USA

[81]Slone Epidemiology Center at Boston University, Boston MA, USA

[82]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[83]Department of Human Genetics, Miller School of Medicine, University of Miami, FL, USA

[84]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[85]Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, FL, USA

[86]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, USA

[87]Department of Genetics and Genomic Sciences, The Icahn School of Medicine at Mount Sinai, New York, NY, USA

[88]Icahn Institute for Genomics and Multiscale Biology, The Icahn School of Medicine at Mount Sinai, New York, NY, USA

[89]Department of Epidemiology, University of Washington, Seattle, WA, USA

[90]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

[91]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA

[92]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

[93]Division of Gastroenterology. Massachusetts General Hospital, Boston, MA, USA

[94]The Broad Institute, Cambridge, MA, USA

[95]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA

[96]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA

Nature Genetics: doi:10.1038/ng.2608

[97]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

[98]Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA

[99]Center for Human Genetics Research, Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN, USA

[100]Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN, USA

[101]Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD, USA

[102]Institute for Human Genetics, University of California San Francisco

[103]Department of Neurological Surgery, University of California San Francisco

[104]Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

[105]Institute for Human Genetics, Departments of Epidemiology and Biostatistics and Urology, University of California, San Francisco, San Francisco, CA, USA

[106]Department of Medicine, University of Vermont College of Medicine, Burlington, VT, USA

[107]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

[108]Department of Epidemiology, School of Public Health, University of Michigan, MI, USA

[109]Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

[110]College of Nursing, Medical University of South Carolina, Charleston, SC, USA

[111]Framingham Heart Study, Boston University School of Medicine, Boston, MA, USA

[112]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[113]Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, Duarte. CA, USA

[114]International Epidemiology Institute, Rockville, MD, USA

[115]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[116]Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

[117]Section on Gerontology and Geriatric Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

Nature Genetics: doi:10.1038/ng.2608

[118]Health Disparities Research Section, Laboratory of Population Science, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

[119]National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA

[120]National Heart, Lung and Blood Institute's Center for Population Studies, Framingham, MA, USA

[121]Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[122]Department Nutrition Sciences, University of Alabama at Birmingham and the Birmingham VA Medical Center, Birmingham, AL, USA

[123]The Center for Applied Genomics, The Children's Hospital of Philadelphia Research Institute, The Children's Hospital of Philadelphia, PA, USA

[124]Division of Human Genetics, The Children's Hospital of Philadelphia Research Institute, The Children's Hospital of Philadelphia, PA, USA

[125]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[126]Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Philadelphia PA, USA

[127]Cancer Prevention Institute of California, Fremont, CA, USA

[128]Division of Epidemiology, Department of Health Research & Policy, Stanford University School of Medicine, Stanford, CA, USA

[129]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

[130]Department of Health Studies, University of Chicago, IL, USA

[131]Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA

[132]SWOG Statistical Center, Seattle, WA, USA

[133]Glickman Urologic and Kidney Institute, Cleveland Clinic, Cleveland, OH

[134]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA

[135]Department of Epidemiology, School of Public Health, University of California at Los Angeles, Los Angeles, CA, USA

[136]Department of Biostatistics, University of Washington, Seattle, WA, USA

[137]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI

[138]Department of Internal Medicine, Henry Ford Health System, Detroit, MI

Nature Genetics: doi:10.1038/ng.2608

[139]Division of Sleep Medicine, Brigham and Women's Hospital, Boston MA, USA

[140]Department of Urology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

[141]Department of Epidemiology and Health Services, University of Washington, Seattle, WA, USA

[142]Group Health Research Institute, Group Health Cooperative, Seattle, WA.

[143]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[144]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

[145]Department of Medicine, University of Virginia, Charlottesville, VA, USA

[146]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA

[147]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[148]Institute for Quantitative Biomedical Sciences, The Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA

[149]Department of Genetics, The Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA

[150]Department of Community and Family Medicine, The Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA

[151]Division of Biology and Biomedical Sciences, Human and Statistical Genetics Program, Washington University, St. Louis, MO, USA

[152]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA

[153]Behavioral Epidemiology Section, Laboratory of Population Science, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

[154]A list of contributing members appears in the **Supplementary Note**.

[155]Division of Prevention and Population Sciences, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

[156]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY, USA

[157]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, USA

**In Memoriam

Nature Genetics: doi:10.1038/ng.2608

Contents

Nature Genetics: doi:10.1038/ng.2608

Nature Genetics: doi:10.1038/ng.2608

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1. Study design, number of individuals, and sample quality control for genome-wide association studies contributing to Stage 1 or follow-up.

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	Full name				Call rate	other exclusions			
AABC	African American Breast Cancer Consortium	Case-control studies	5761	1	>95%	1) ancestry outliers 2) relateds 3) missing BMI 4) gender mismatch	5371	self-reported	PMID:21852243 [1]
AAPC1	African Ancestry Prostate Cancer Consortium	Case-control studies	6715	1	>95%	1) ancestry outliers 2) relateds 3) missing BMI 4) gender mismatch	5459	self-reported	PMID:21602798 [2]
AAPC2	African Ancestry Prostate Cancer Consortium	Case-control studies	2835	2	>95%	1) ancestry outliers 2) relateds 3) missing BMI 4) gender mismatch	2350	self-reported	PMID:21602798 [2]
ARIC	Atherosclerosis Risk in Communities Study	Population-based	2989	1	> 95%	1) samples with <18 fingerprinting assays working 2) samples with >3 discordant fingerprinting assays 3) duplicates 4) samples creating lg number of haploid heterozygous calls 5) extreme heterozygosity value 6) low-level IBD/IBS sharing w/ large number of samples 7) nearest neighbor analysis outlier 8) outliers from clustering based on missingness	2775	measured	PMID: 2646917 [3]
BioVu (Stage 1)	BioVU, the Vanderbilt DNA Databank project	DNA biobank linked to electronic medical records via an opt-out deidentified database	1227	1	>=95%	1) cryptic relatedness 2) sex check 3) missing BMI	1136	measured	PMID: 18500243 [4]

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	**Full name**				**Call rate**	**other exclusions**			
BioVu (Stage 3)	BioVU, the Vanderbilt DNA Databank project	DNA biobank linked to electronic medical records via an opt-out deidentified database	913	3	>=95%	1) cryptic relatedness 2) sex check 3) missing BMI	668	measured	PMID: 18500243 [4]
CARDIA	Coronary Artery Risk Development in Young Adults	Population-based	955	1	> 95%	1) severe PC outliers prior to QC 2) samples with <18 fingerprinting assays working 3) samples with >3 discordant fingerprinting assays 4) duplicates 5) samples creating large number of haploid heterozygous calls 6) extreme heterozygosity value 7) low-level IBD/IBS sharing w/ large number of samples 8) nearest neighbor analysis outlier 9) outliers from clustering based on missingness	819	measured	PMID: 3204420 [5]

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	Full name				Call rate	other exclusions			
CFS	Cleveland Family Study	Family-based	632	1	> 95%	1) samples with <18 working fingerprinting assays 2) samples with >3 discordant fingerprinting assays 3) duplicates 4) samples creating large number of haploid hz calls 5) extreme heterozygosity value 6) low-level IBD/IBS sharing w/ large number of samples 7) nearest neighbor analysis outlier 8) outliers from clustering based on missingness 9) samples with high Mendel error rate	468	measured	PMID:12734134 [6]
CHS	Cardiovascular Health Study	Prospective, Population-based	844	1	>=95%	if subject's genotype was discordant with known sex or prior genotyping. If missing weight or height measures	808	measured	PMID: 1669507 [7] PMID:19557197 [8]
FamHS	Family Heart Study	Family-based	624	1	≥ 98%	1) technical errors 2) discrepancies between reported sex and sex-diagnostic markers	624	measured	PMID: 8651220 [9]
GeneSTAR	Genetic Study of Atherosclerosis Risk	Family-based	1220	2	99%	1) sex/gender mismatch; 2) extreme outlier (>10SD) from Eigenstrat; 3)missing height or weight	1137	measured	PMID 20377806 [10] PMID: 15837938 [11]

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	Full name				Call rate	other exclusions			
GENOA	Genetic Epidemiology Network of Arteriopathy	Cohort of sibships enriched for hypertension	1624	1	≥ 95%	1) GENOA participants who also participated in the ARIC Study; 2) sex/gender mismatch; extreme outlier (>6 SD) based on first 10 principal components	996	measured	PMID: 15121494 [12]
Health ABC	Health, Aging, and Body Compostion Study	Population-based	1139	1	> 97%	1) sex discordance 2) discordance between self-reported and genetic ancestry estimates, MDS clustering with HapMap3 ASW as primary reference 3) presence of cryptic relatedness to any other samples at a level of proportional sharing of genotypes > 15% 4) missing data	1139	measured	
HANDLS	Healthy Aging in Neighborhoods of Diversity across the Life Span Study	Population-based	1024 in genetics components, 3722 in total	1	> 95%	1) sex discordance 2) discordance between self-reported and genetic ancestry estimates, MDS clustering with HapMap3 ASW as primary reference 3) presence of cryptic relatedness to any other samples at a level of proportional sharing of genotypes > 15% 4) missing data	994	measured	PMID: 20828101 [13]
HUFS	Howard University Family Study	Population-based	1976	1	> 95%	1) gender missmatch. 2) pedigree information. 3) outlier of IBS. 4) missing data	927	measured	PMID: 19609347 [14]

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	Full name				Call rate	other exclusions			
HyperGen	Hypertension Genetic Epidemiology Network	Family-based	1258	1	>95%	1) pedigree errors 2) blood sample mix up 3) missing BMI or covariate data 4) extreme outlier from Eigenstrat	1181	measured	PMID: 21212386 [15]
IPM	Charles R. Bronfman Institute for Personalized Medicine (IPM) BioBank Genome Wide Association Study of Cardiovascular, Renal and Metabolic Phenotypes	Case-Control	887	2	≥ 95%	1) sex/gender mismatch 2) First-degree relative 3) Outlier based on average IBS (>4SD) 4) missing height and/or weight or other covariate	887	self-reported	PMID: 21573225 [16]
IPM	The Charles Bronfman Institute for Personalized Medicine (IPM) BioBank Genome Wide Association Study of Cardiovascular, Renal and Metabolic Phenotypes	Hospital-based Biobank	3983	3	≥ 95%	1) sex/gender mismatch; 2) individuals with > 5% missing genotype or outlying heteozygosity (Z-value>|6|); 3) duplicates or related samples (PI-HAT > 0.6) 4) missing BMI or other covariates	3470	self-reported	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	Full name				Call rate	other exclusions			
JHS	Jackson Heart Study	Population-based	2145	1	> 95%	1) samples with <18 fingerprinting assays working 2) samples with >3 discordant fingerprinting assays 3) duplicates 4) samples creating lg number of haploid heterozygous calls 5) extreme heterozygosity value 6) low-level IBD/IBS sharing w/ large number of samples 7) nearest neighbor analysis outlier 8) outliers from clustering based on missingness 9) samples with high Mendel error rate	2131	measured	PMID: 10100686 [17]
JoCo	Johnston County Osteoarthritis Project	Population-based	2583	3	98%	1) sex/gender mismatch 2) first-degree relative 3) disagreement between reported and genotypic race 4) missing height and/or weight or other covariate	592	measured	PMID: 17216685 [18]
Maywood	Genetics of Hypertension in Blacks	Population-based	743	1	≥ 95%	1) sex/gender mismatch 2) first-degree relative 3) outlier based on average IBS (>4SD)	743	measured	PMID: 20400458 [19]
MESA	Multi-Ethnic Study of Atherosclerosis	Population-based	6814	1	≥ 95%	1) duplicates 2) gender mismatch 3) non-African	1381	measured	PMID: 12397006 [20]
MESA	MESA Air Pollution and	Population-based	2128	1	≥ 95%	1)duplicates,	947	measured	PMID: 12397006 [20]

Nature Genetics: doi:10.1038/ng.2608

Air/Family	MESA Family Studies	and Family-based				2)gender mismatch 3)non-African			PMID: 19673252 [21]

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	**Full name**				**Call rate**	**other exclusions**			
NCI Lung	NCI Lung Cancer in African Americans GWAS Consortium	Case-control studies	2472	2	≥ 94%	1)sample heterozygosity, 2) sex/gender mismatch, 3) duplicate concordance and imputed ancestry, 4)incomplete covariates and/or phenotype	2444	measured/self-reported	
Nigeria	Genetics of Hypertension in Blacks	Population-based	1188	1	≥ 95%	1) sex/gender mismatch 2) first-degree relative 3) outlier based on average IBS (>4SD)	1188	measured	PMID: 20400458 [19]
ROOT	Genome-Wide Association Study of Breast Cancer in the African Diaspora	Case-control studies	3,774	3	>95%	1) ancestry outliers 2) relateds 3) missing BMI 4) chromosome anomalies	3161	measured/self-reported	PMID: 22357627 [22]
SAPPHIRE	Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity	Case-Control	1536	3	≥ 95%	1) cryptic relatedness 2) sex/gender mismatch 3) call rate filter 4) Adjustment for cryptic relatedness 5) Missing BMI	1012	measured	PMID: 21804549 [23]
SIGNET-REGARDS	Sea Islands Genetics Network	Population-based	2398	3	≥ 95%	1) cryptic relatedness 2) sex/gender mismatch 3) call rate filter	2381	measured	PMID: 15990444 [24] PMID: 18835935 [25] PMID: 20507373 [26] PMID: 19783527 [27]
SIGNET-Sea Islands	Sea Islands Genetics Network	Family-based	1495	3	≥ 95%	1) cryptic relatedness 2) sex/gender mismatch 3) call rate filter 4) Mendelian inconsistency error checking	1268	measured	PMID: 15990444 [24] PMID: 18835935 [25] PMID: 20507373 [26] PMID: 19783527 [27]

Nature Genetics: doi:10.1038/ng.2608

WFSM	Wake Forest School of Medicine Study	Case-control study	1994	1	≥ 95%	1) sex/gender mismatch 2) first-degree relative 3) outlier based on average IBS (>4SD) 4) missing height and/or weight or other covariate	1714	measured	PMID:21701570 [28] _ENREF_20

Supplementary Table 1, continued

Study		Study design	Total sample size (N)	Stage	Sample QC		Samples in analyses (N)	Anthropometric assessment method	References
Short name	**Full name**				**Call rate**	**other exclusions**			
WHI-GARNET	Women's Health Initiative-Genomics and Randomized Trials Network	Nested case-control for CVD	168	3	>99%	1) duplicates 2) had Y-chromosome 3) blind duplicate concordance 4) relatedness	168	measured	PMID: 9492970 [29]
WHI-SHARe	Women's Health Initiative SHARe sample	Population-based	8421	1	99.8%	1) duplicates 2) had Y-chromosome 3) blind duplicate concordance 4) relatedness	8094	measured	PMID: 9492970 [29]

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 2. Information on genotyping methods, quality control of SNPs, imputation, and statistical analysis of genome-wide association studies contributing to Stage 1 or follow-up.

Genotyping							Imputation			Meta-Analysis				
			Inclusion criteria					Inclusion criteria				λ_{GC}		
Study	Platform	Genotype calling algorithm	MAF	Call rate*	P for HWE	SNPs that met QC criteria	Imputation software	MAF	Imputation quality**	SNPs in meta-analysis	Stage	Men	Women	Analysis software
AABC	Illumina 1M Duo	Beadstudio	>1%	>99%	none	1043036	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	3110134	1	--	1.03	C++ code
AAPC1	Illumina 1M Duo	Beadstudio	>1%	>99%	none	1047986	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	3044338	1	1.02	--	C++ code
AAPC2	Illumina 1M Duo	Beadstudio	>1%	>99%	none	1053764	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	1500	2	N/A	N/A	C++ code
ARIC	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v1.33	> 1%	≥95%	none	796384	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	2648165 (men) 2650518 (women)	1	1.02	1.02	PLINK
BioVU	Illumina Human 1M-DuoV3	Beadstudio	> 1%	> 98%	$>10^{-6}$	1024850	IMPUTE v2.1.2	none	none	3120932	1	1.01	1.01	SNPTEST v2.2.0
BioVU(Stage 3)	Illumina Human 1M-DuoV3 Illumina Omni 5M Illumina Omni 1M	Beadstudio	> 1%	> 98%	$>10^{-6}$	1137860	IMPUTE v2.1.2	none	none	3	3	N/A	N/A	SNPTEST v2.2.0
CARDIA	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v1.33	> 1%	≥95%	none	839912	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	2646097 (men) 2650029 (women)	1	1.00	1.02	PLINK
CFS	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v1.33	> 1%	≥95%	none	867495	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	2339282 (men) 2341788 (women)	1	1.00	1.00	R
CHS	HumanOmni1-Quad_v1	Illumina GenomeStudio	none	≥97%	$>10^{-5}$	963248	BEAGLE v3.2.1	none	none	2620194(men) 2621248(women)	1	1.06	1.04	R
FamHS	Illumina Human 1M-DuoV3	Beadstudio-GENCALL v3.0	> 1%	> 98%	$>10^{-6}$	754504	MACH v1.0.16	> 1%	r2_hat ≥ 0.50	2183149 (men) 2183167 (women)	1	1.00	1.01	SAS, R
GeneSTAR	Illumina Human 1Mv1_C	GenomeStudio	≥ 1%	≥95%	$>10^{-6}$	973045	MACH v1.0.16	none	r2_hat ≥ 0.50	1500	2	N/A	N/A	R, GWAF
GENOA	Affymetrix Genome-Wide Human SNP Array 6.0 and Illumina Human 1M-Duo	Birdseed and Illumina GenomeStudio	> 1%	≥95%	None	761839	MACH v1.0.16	≥ 1%	none	2340219	1	1.04	1.02	R
Health ABC	Illumina Human1M-Duo	BeadStudio v3.3.7	≥ 1%	≥97%	$>10^{-6}$	941595	MACH v1.0.16	none	none	3129973	1	0.99	0.99	R
HANDLS	Illumina 1M	BeadStudio	> 1%	> 95%	$>10^{-7}$	907763	MACH v1.0.16	none	r2_hat ≥ 0.30	2939993	1	1.00	0.99	MACH2QTL v1.08
HUFS	Affymetrix 6.0	Birdseed v2	≥ 1%	> 95%	$>10^{-6}$	842152	MACH v1.0.16	≥ 1%	r2_hat ≥ 0.30	2074885	1	1.00	1.01	PLINK, SAS
HyperGen	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	> 1%	> 95%	$>10^{-6}$	846813	MACH v1.0.16	≥ 1%	r2_hat ≥ 0.30	2945497	1	0.95	0.95	Plink
IPM (Stage 2)	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	>1%	> 95%	$>10^{-4}$	665889	MACH v1.0.16	none	none	1500	2	N/A	N/A	mach2qtl
IPM (Stage 3)	HumanOmniExpress-12v1	Illumina GenomeStudio	>1%	> 95%	$>10^{-3}$	635812	IMPUTE v2.1.2		Proper_info> 0.97	3	3	N/A	N/A	Plink and SNPTEST

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 2, continued

Genotyping							Imputation			Meta-Analysis				
			Inclusion criteria					Inclusion criteria				λ_{GC}		
Study	Platform	Genotype calling algorithm	MAF	Call rate*	P for HWE	SNPs that met QC criteria	Imputation software	MAF	Imputation quality**	SNPs in meta-analysis	Stage	Men	Women	Analysis software
JHS	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v1.33	> 1%	≥95%	none	868969	MACH 1.0.16	≥ 1%	r2_hat ≥ 0.30	2650197 (men) 2651305 (women)	1	1.04	1.06	PLINK
JoCo	Illumina 1M-Duo	BeadStudio	>0.5%	>98%	$>10^{-4}$	1065734	MACH	none	none	3	3	N/A	N/A	R, ProbABEL
Maywood	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	> 1%	> 95%	$>10^{-6}$	859332	MACH v1.0.16	≥ 1%	r2_hat ≥ 0.30	2834063	1	1.01	1.02	mach2qtl
MESA	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	> 1%	≥ 95%	none	897979	IMPUTE v2.1.0	none	none	3120116	1	1.02	1.08	R, pedigremm
MESA Air/Family	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	> 1%	≥ 95%	none	897979	IMPUTE v2.1.0	none	none	3120116	1	1.02	1.08	R, pedigremm
NCI Lung	Illumina 1M Duo	Illumina GenomeStudio	none	≥90%	none	1128840	MACH v1.0.16	none	none	1500	2	N/A	N/A	mach2qtl, glu
Nigeria	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	> 1%	≥ 95%	$>10^{-6}$	792857	MACH v1.0.16	≥ 1%	r2_hat ≥ 0.30	2908510	1	1.06	1.07	mach2qtl
ROOT	Illumina HumanOmni2.5-8v1_A	Beadstudio	>0%	>98%	$>10^{-4}$	2,116,675	IMPUTE 2.0	none	info ≥ 0.30	3	3	N/A	N/A	SAS
SAPPHIRE	Affymetrix Axiom AFR array	BRLMM-p in Affymetrix Power Tools	N/A	≥95%	N/A	N/A	IMPUTEv2.2.2	N/A	N/A	3	3	N/A	N/A	EMMAX
SIGNET-REGARDS	Affymetrix Genome-Wide Human SNP Array 6.0	seed v1.33	N/A	N/A	N/A	N/A	MACH v1.0.16		r2_hat ≥ 0.30	3	3	N/A	N/A	R
SIGNET-Sea Islands	Affymetrix Genome-Wide Human SNP Array 6.0	seed v1.33	N/A	N/A	N/A	N/A	MACH v1.0.16		r2_hat ≥ 0.30	3	3	N/A	N/A	GDT
WFSM	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	≥ 1%	≥95%	$>10^{-6}$	836105	MACH v1.0.16	≥ 1%	r2_hat ≥ 0.30	2969127	1	0.99	1.00	PLINK, MACH2QTL
WHI-Garnet	Illumina HumanOmni1-Quad v1-0 B SNP array	BeadStudio version 3.1.3.0 with genotyping module version 3.2.32	≥ 1%	≥97%	$>10^{-4}$	942,224	BEAGLE	≥ 1%	r2_hat ≥ 0.30	3	3	--	N/A	R
WHI-SHARe	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	≥ 1%	≥90%	$>10^{-6}$	829370	MACH	≥ 1%	r2_hat ≥ 0.30	2435278	1	--	1.03	ProbABEL

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3. Study-specific descriptive statistics for all participating studies. Shown are the strata included in the meta-analysis.

			Men						Women					
Study	**Stage**	**Trait**	**n**	**Mean**	**SD**	**Median**	**Min**	**Max**	**n**	**Mean**	**SD**	**Median**	**Min**	**Max**
AABC: CARE Cases	1	Age (yrs)							351	49.05	8.02	49.00	35.00	64.00
		BMI (kg/m²)							351	27.32	5.51	26.40	17.40	43.90
		Weight (kg)							351	74.70	16.04	72.58	45.36	127.01
		Height (m)							351	1.65	0.07	1.65	1.50	1.88
AABC: CARE Controls	1	Age (yrs)							208	47.91	7.95	48.00	35.00	64.00
		BMI (kg/m²)							208	26.52	4.76	25.70	17.50	42.90
		Weight (kg)							208	71.95	13.42	70.31	45.81	117.94
		Height (m)							208	1.65	0.07	1.65	1.50	1.91
AABC: CBCS Cases	1	Age (yrs)							570	51.41	11.85	50.00	23.00	74.00
		BMI (kg/m²)							570	30.80	5.75	30.76	16.94	43.98
		Weight (kg)							570	81.49	16.32	80.00	45.00	146.00
		Height (m)							570	1.63	0.06	1.63	1.42	1.88
AABC: CBCS Controls	1	Age (yrs)							528	51.94	11.45	50.00	26.00	74.00
		BMI (kg/m²)							528	31.02	5.68	30.49	16.60	43.93
		Weight (kg)							528	82.02	16.27	81.00	42.50	139.00
		Height (m)							528	1.63	0.07	1.63	1.40	1.82
AABC: MEC Cases	1	Age (yrs)							651	66.43	9.03	67.00	45.00	86.00
		BMI (kg/m²)							651	28.95	5.34	28.40	16.50	43.95
		Weight (kg)							651	78.70	15.40	77.11	44.45	135.17
		Height (m)							651	1.65	0.07	1.65	1.42	1.91
AABC: MEC Controls	1	Age (yrs)							927	67.13	9.35	68.00	46.00	86.00
		BMI (kg/m²)							927	28.55	5.09	27.86	17.60	43.93
		Weight (kg)							927	76.43	14.70	74.84	43.55	124.29
		Height (m)							927	1.64	0.06	1.63	1.42	1.91
AABC: NBHS Cases	1	Age (yrs)							293	54.43	11.20	54.00	26.00	75.00
		BMI (kg/m²)							293	30.06	5.37	29.46	18.18	43.55
		Weight (kg)							293	81.31	15.20	81.00	44.55	129.15
		Height (m)							293	1.64	0.07	1.65	1.47	1.85
AABC: NBHS Controls	1	Age (yrs)							174	51.97	10.19	52.00	29.00	74.00
		BMI (kg/m²)							174	29.92	5.88	28.95	16.81	43.55
		Weight (kg)							174	81.56	16.94	77.40	47.25	128.25
		Height (m)							174	1.65	0.06	1.65	1.47	1.83

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
AABC: PLCO Cases	1	Age (yrs)							53	68.21	6.74	68.00	59.00	87.00
		BMI (kg/m²)							53	29.88	4.56	29.33	21.30	40.86
		Weight (kg)							53	81.33	13.85	81.82	54.55	116.36
		Height (m)							53	1.65	0.06	1.65	1.50	1.75
AABC: PLCO Controls	1	Age (yrs)							110	67.88	5.99	67.00	57.00	80.00
		BMI (kg/m²)							110	29.80	5.10	29.18	20.64	43.95
		Weight (kg)							110	80.44	14.87	79.55	54.55	127.27
		Height (m)							110	1.64	0.06	1.64	1.50	1.80
AABC: SFBCS/NC-BCFR Cases	1	Age (yrs)							546	51.77	10.59	52.00	22.00	79.00
		BMI (kg/m²)							546	28.77	5.87	28.00	16.80	43.97
		Weight (kg)							546	77.68	16.82	74.84	41.00	136.00
		Height (m)							546	1.64	0.07	1.64	1.45	1.93
AABC: SFBCS/NC-BCFR Controls	1	Age (yrs)							254	54.25	11.60	53.00	31.00	80.00
		BMI (kg/m²)							254	29.41	5.83	29.03	17.73	43.43
		Weight (kg)							254	78.48	16.11	79.38	46.00	122.47
		Height (m)							254	1.63	0.07	1.63	1.46	1.78
AABC: WCHS Cases	1	Age (yrs)							237	49.70	9.75	51.00	22.00	73.00
		BMI (kg/m²)							237	30.85	5.85	30.60	17.80	43.70
		Weight (kg)							237	82.23	16.59	80.30	50.10	146.20
		Height (m)							237	1.63	0.07	1.63	1.47	1.92
AABC: WCHS Controls	1	Age (yrs)							223	49.83	9.36	51.00	23.00	65.00
		BMI (kg/m²)							223	30.28	5.54	30.00	19.70	43.30
		Weight (kg)							223	81.05	16.49	79.30	48.20	146.20
		Height (m)							223	1.63	0.06	1.64	1.46	1.87
AABC: WFBC Cases	1	Age (yrs)							108	55.47	12.01	54.00	30.00	85.00
		BMI (kg/m²)							108	30.14	5.23	29.01	17.81	41.20
		Weight (kg)							108	80.62	15.35	79.38	36.29	127.01
		Height (m)							108	1.63	0.08	1.63	1.42	1.96
AABC: WFBC Controls	1	Age (yrs)							131	55.64	10.38	55.00	37.00	86.00
		BMI (kg/m²)							131	30.59	5.54	30.21	16.55	42.94
		Weight (kg)							131	81.22	15.66	79.83	49.38	127.46
		Height (m)							131	1.63	0.06	1.63	1.47	1.82
AAPC1: CaP Genes Cases	1	Age (yrs)	64	67.06	8.46	67.50	48.00	89.00						
		BMI (kg/m²)	64	27.39	4.48	27.53	19.53	36.94						
		Weight (kg)	64	86.51	16.74	84.37	54.43	127.01						

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	64	1.77	0.07	1.78	1.63	2.01	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
AAPC1: CaP Genes Controls	1	Age (yrs)	78	66.91	8.51	67.00	48.00	84.00						
		BMI (kg/m²)	78	28.88	4.42	28.96	19.26	37.30						
		Weight (kg)	78	92.05	16.38	95.03	55.79	145.15						
		Height (m)	78	1.78	0.09	1.78	1.57	2.03						
AAPC1: CPS-II Cases	1	Age (yrs)	61	70.43	6.34	71.00	56.00	87.00						
		BMI (kg/m²)	61	27.13	3.33	27.05	19.94	35.24						
		Weight (kg)	61	85.50	11.98	85.73	61.24	113.40						
		Height (m)	61	1.77	0.07	1.80	1.57	1.91						
AAPC1: CPS-II Controls	1	Age (yrs)	110	70.85	5.63	71.00	61.00	88.00						
		BMI (kg/m²)	110	26.73	3.17	26.60	20.95	37.66						
		Weight (kg)	110	84.12	11.59	81.65	62.60	122.47						
		Height (m)	110	1.77	0.06	1.78	1.57	1.93						
AAPC1: DCPD Cases	1	Age (yrs)	128	66.82	8.73	67.00	48.00	88.00						
		BMI (kg/m²)	128	26.97	3.96	26.87	18.88	36.64						
		Weight (kg)	128	83.08	13.78	81.65	49.90	125.65						
		Height (m)	128	1.75	0.08	1.75	1.52	1.93						
AAPC1: DCPD Controls	1	Age (yrs)	96	59.19	10.66	60.00	35.00	92.00						
		BMI (kg/m²)	96	27.62	4.23	27.26	19.90	37.66						
		Weight (kg)	96	85.41	16.12	82.10	54.43	129.28						
		Height (m)	96	1.76	0.09	1.75	1.35	1.93						
AAPC1: GECAP Cases	1	Age (yrs)	212	61.66	7.36	62.00	42.00	75.00						
		BMI (kg/m²)	212	28.00	4.11	27.37	18.75	37.73						
		Weight (kg)	212	88.68	14.75	86.64	54.43	133.81						
		Height (m)	212	1.78	0.07	1.78	1.60	1.96						
AAPC1: GECAP Controls	1	Age (yrs)	86	61.63	7.44	62.50	41.00	75.00						
		BMI (kg/m²)	86	28.14	3.67	27.43	20.54	36.28						
		Weight (kg)	86	89.60	13.81	87.54	63.50	127.01						
		Height (m)	86	1.78	0.08	1.78	1.43	1.96						
AAPC1: KCPCS Cases	1	Age (yrs)	134	59.13	6.99	59.00	43.00	74.00						
		BMI (kg/m²)	134	27.63	3.85	27.34	18.88	37.59						
		Weight (kg)	134	89.02	14.98	86.18	49.90	129.28						
		Height (m)	134	1.79	0.08	1.80	1.57	2.01						
AAPC1: KCPCS Controls	1	Age (yrs)	70	54.74	6.48	53.00	45.00	71.00						
		BMI (kg/m²)	70	28.33	4.09	28.42	20.92	37.12						
		Weight (kg)	70	93.32	17.07	91.85	56.70	134.72						

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	70	1.81	0.08	1.83	1.57	2.06	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
AAPC1: LAAPC Cases	1	Age (yrs)	270	63.66	9.27	64.00	42.00	88.00						
		BMI (kg/m²)	270	27.87	4.04	27.57	19.96	37.50						
		Weight (kg)	270	87.94	15.23	86.18	57.15	136.08						
		Height (m)	270	1.77	0.08	1.77	1.56	2.00						
AAPC1: LAAPC Controls	1	Age (yrs)	273	63.89	8.59	64.00	39.00	85.00						
		BMI (kg/m²)	273	27.84	3.96	27.13	18.40	37.67						
		Weight (kg)	273	88.19	15.20	86.18	54.89	154.22						
		Height (m)	273	1.78	0.08	1.78	1.57	2.03						
AAPC1: MDA Cases	1	Age (yrs)	462	60.15	8.51	59.58	38.97	86.82						
		BMI (kg/m²)	462	27.72	4.07	27.13	18.52	38.30						
		Weight (kg)	462	88.03	14.96	86.18	52.00	138.80						
		Height (m)	462	1.78	0.08	1.78	1.57	2.01						
AAPC1: MDA Controls	1	Age (yrs)	409	57.33	9.32	57.87	31.28	81.44						
		BMI (kg/m²)	409	28.34	3.81	28.12	19.12	37.81						
		Weight (kg)	409	90.76	14.30	89.81	58.51	138.35						
		Height (m)	409	1.79	0.08	1.80	1.47	2.03						
AAPC1: MEC Cases	1	Age (yrs)	1001	69.19	7.38	69.00	46.00	88.00						
		BMI (kg/m²)	1001	27.09	3.56	26.63	18.48	37.75						
		Weight (kg)	1001	85.92	13.32	83.92	52.62	132.45						
		Height (m)	1001	1.78	0.07	1.78	1.52	2.03						
AAPC1: MEC Controls	1	Age (yrs)	1000	69.59	7.64	70.00	46.00	87.00						
		BMI (kg/m²)	1000	27.07	3.62	26.63	18.26	37.79						
		Weight (kg)	1000	85.79	13.32	83.92	52.16	133.81						
		Height (m)	1000	1.78	0.07	1.78	1.42	2.11						
AAPC1: PLCO PC Cases	1	Age (yrs)	217	67.76	5.83	68.00	55.00	80.00						
		BMI (kg/m²)	217	27.59	3.74	27.38	19.09	37.31						
		Weight (kg)	217	86.90	13.97	85.91	52.27	131.82						
		Height (m)	217	1.77	0.07	1.78	1.60	1.93						
AAPC1: PLCO PC Controls	1	Age (yrs)	228	63.43	5.36	63.00	55.00	74.00						
		BMI (kg/m²)	228	27.81	4.17	27.74	18.83	37.31						
		Weight (kg)	228	88.42	15.21	86.37	52.73	145.46						
		Height (m)	228	1.78	0.07	1.78	1.60	2.06						
AAPC1: SCCS Cases	1	Age (yrs)	185	61.28	7.57	61.00	40.00	80.00						
		BMI (kg/m²)	185	26.91	4.42	26.64	18.47	37.59						
		Weight (kg)	185	85.23	14.79	83.92	54.43	120.20						

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	185	1.78	0.07	1.78	1.57	2.01	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
AAPC1: SCCS Controls	1	Age (yrs)	372	59.11	7.80	59.00	40.00	79.00						
		BMI (kg/m²)	372	27.47	4.47	27.29	18.21	37.52						
		Weight (kg)	372	87.09	15.89	84.82	53.52	136.08						
		Height (m)	372	1.78	0.08	1.78	1.52	2.01						
AAPC2: MEC Cases	2	Age (yrs)	705	63.42	8.23	64.00	39.00	93.00						
		BMI (kg/m²)	705	28.32	4.85	27.71	14.00	53.70						
		Weight (kg)	705	90.45	17.33	88.45	44.45	181.44						
		Height (m)	705	1.79	0.08	1.78	1.37	2.08						
AAPC2: MEC Controls	2	Age (yrs)	605	66.58	7.67	67.00	48.00	87.00						
		BMI (kg/m²)	605	27.87	4.33	27.34	16.27	50.21						
		Weight (kg)	605	88.69	15.63	86.18	48.53	167.83						
		Height (m)	605	1.78	0.07	1.78	1.52	2.13						
AAPC2: PCBP Cases	2	Age (yrs)	232	66.21	8.85	66.00	47.00	87.00						
		BMI (kg/m²)	232	25.86	4.59	25.47	16.25	47.85						
		Weight (kg)	232	76.16	14.92	74.84	45.36	143.34						
		Height (m)	232	1.71	0.08	1.71	1.36	2.00						
AAPC2: PCBP Controls	2	Age (yrs)	222	66.12	8.77	66.50	49.00	87.00						
		BMI (kg/m²)	222	25.39	4.16	25.28	15.41	40.46						
		Weight (kg)	222	75.12	13.98	74.39	43.09	110.68						
		Height (m)	222	1.72	0.07	1.72	1.50	1.88						
AAPC2: SCCS Cases	2	Age (yrs)	48	62.65	7.32	62.00	48.00	80.00						
		BMI (kg/m²)	48	29.74	6.07	29.29	20.36	53.09						
		Weight (kg)	48	91.65	19.91	88.22	58.97	167.83						
		Height (m)	48	1.75	0.07	1.75	1.60	1.91						
AAPC2: SCCS Controls	2	Age (yrs)	100	59.07	7.51	58.00	45.00	78.00						
		BMI (kg/m²)	100	27.31	6.10	26.58	17.33	49.81						
		Weight (kg)	100	86.32	19.24	81.65	56.25	146.06						
		Height (m)	100	1.78	0.08	1.78	1.57	1.96						
AAPC2: SELECT Cases	2	Age (yrs)	212	64.67	7.02	64.00	51.00	86.00						
		BMI (kg/m²)	212	29.98	5.52	29.36	18.58	56.70						
		Weight (kg)	212	94.20	18.20	92.08	50.80	164.20						
		Height (m)	212	1.77	0.07	1.78	1.60	1.96						
AAPC2: SELECT Controls	2	Age (yrs)	207	64.52	7.08	64.00	51.00	87.00						
		BMI (kg/m²)	207	29.00	4.61	28.62	19.85	44.68						
		Weight (kg)	207	90.84	17.17	88.00	55.79	157.85						

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	207	1.77	0.08	1.75	1.55	1.98	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
ARIC	1	Age (yrs)	1021	53.51	5.99	53.00	44.00	66.00	1754	53.17	5.68	53.00	44.00	65.00
		BMI (kg/m²)	1021	27.98	4.81	27.47	15.46	54.41	1754	30.63	6.43	29.68	14.20	60.62
		Weight (kg)	1017	86.76	16.29	85.28	46.72	165.56	1746	81.34	17.57	78.93	37.19	158.76
		Height (m)	1017	1.76	0.07	1.76	1.56	1.97	1746	1.63	0.06	1.63	1.25	1.88
BioVU Cases	1	Age (yrs)	136	55.32	12.55	55.50	27.00	84.00	313	53.47	13.97	53.00	24.00	92.00
		BMI (kg/m²)	136	32.53	8.34	31.16	14.75	80.87	313	37.02	10.23	35.93	5.84	94.16
		Weight (kg)	136	101.68	25.40	96.89	47.53	217.84	313	97.81	26.42	93.44	47.31	210.00
		Height (m)	136	1.78	0.08	1.77	1.54	2.03	313	1.63	0.08	1.63	1.17	1.83
BioVU Controls	1	Age (yrs)	191	47.89	13.82	47.50	21.00	85.00	496	44.73	14.85	44.00	21.00	87.00
		BMI (kg/m²)	191	29.82	7.77	28.88	16.87	92.45	496	31.56	8.51	30.40	14.88	65.08
		Weight (kg)	191	92.49	20.69	90.13	43.00	168.28	496	85.35	22.97	81.96	40.00	190.51
		Height (m)	191	1.77	0.08	1.75	1.49	1.98	496	1.65	10.06	1.65	1.42	1.88
BioVU Stage 3 participants	3	Age (yrs)	261	49.11	14.37	50.00	21.00	91.00	407	50.22	17.27	50.00	21.00	96.00
		BMI (kg/m²)	261	29.07	6.78	27.76	15.56	64.53	407	32.18	8.65	30.78	14.04	66.60
		Weight (kg)	261	93.52	22.85	90	41	195	407	86.34	23.89	83.00	38.00	236.00
		Height (m)	261	1.78	0.08	1.78	1.61	1.98	407	1.64	0.07	1.65	1.35	1.83
BWHS Cases (incident)	3	Age (yrs)							838	45.96	10.03	45.00	21.00	69.00
		BMI (kg/m²)							838	28.26	6.05	27.29	16.64	50.89
		Weight (kg)							838	77.14	16.92	74.46	44.49	146.19
		Height (m)							838	1.65	0.07	1.65	1.32	1.98
BWHS Controls	3	Age (yrs)							1872	46.64	10.08	46.00	21.00	70.00
		BMI (kg/m²)							1872	28.23	5.75	27.29	15.83	51.54
		Weight (kg)							1872	76.78	16.70	72.64	40.86	153.00
		Height (m)							1872	1.65	0.07	1.65	1.27	2.10
CARDIA	1	Age (yrs)	369	24.41	3.73	24.00	17.00	34.00	585	24.42	3.92	25.00	17.00	34.00
		BMI (kg/m²)	369	24.46	4.19	23.68	16.59	48.03	585	26.14	6.34	24.81	14.88	53.24
		Weight (kg)	369	76.76	14.90	73.66	48.53	158.76	585	69.90	17.83	66.86	36.83	158.76
		Height (m)	369	1.77	0.07	1.77	1.55	1.95	585	1.64	0.07	1.64	1.21	1.86
CFS	1	Age (yrs)	223	33.40	19.10	36.56	2.07	76.11	332	35.11	19.99	35.94	2.30	85.59
		BMI (kg/m²)	223	29.18	9.15	28.69	11.11	57.71	332	31.07	10.43	30.04	14.36	61.68
		Weight (kg)	223	86.85	36.32	90.45	9.55	181.82	332	80.06	32.47	77.80	11.59	169.09
		Height (m)	223	1.68	0.22	1.75	0.86	2.03	332	1.58	0.16	1.62	0.86	1.77
CHS	1	Age (yrs)	302	72.71	5.74	71.50	65.00	92.00	506	72.95	5.64	72.00	64.00	93.00
		BMI (kg/m²)	302	26.68	4.25	26.37	16.12	38.16	506	29.56	5.96	29.01	16.34	58.79
		Weight (kg)	302	80.32	14.38	79.09	47.73	139.55	506	75.44	15.86	73.07	43.64	134.32

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	302	1.73	0.07	1.73	1.54	1.96	506	1.60	0.06	1.59	1.45	1.87

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
FamHS	1	Age (yrs)	214	52.35	10.43	52.00	30.00	81.00	409	53.79	10.96	54.00	31.00	83.00
		BMI (kg/m²)	214	30.26	6.14	29.60	16.92	51.09	409	33.95	7.63	32.85	15.69	60.23
		Weight (kg)	214	94.32	20.56	93.21	48.53	152.86	409	90.02	21.35	85.73	43.99	157.85
		Height (m)	214	1.77	0.07	1.77	1.40	1.95	409	1.63	0.06	1.63	1.45	2.16
GeneSTAR	2	Age (yrs)	432	42.81	10.74	43.00	20.00	71.00	705	43.099	10.599	44.00	20.00	75.00
		BMI (kg/m²)	432	29.24	6.19	28.46	16.89	53.48	705	32.66	8.13	31.35	16.41	81.19
		Weight (kg)	432	92.00	20.70	88.90	47.20	167.60	705	88.20	23.50	85.30	44.70	229.10
		Height (m)	432	1.78	0.07	1.78	1.38	2.01	705	1.64	0.07	1.64	1.40	1.88
GENOA	1	Age (yrs)	295	57.53	10.40	57.30	32.56	86.59	701	55.91	11.42	55.39	20.50	91.17
		BMI (kg/m²)	295	28.17	4.84	27.45	15.24	50.32	701	32.29	7.18	31.76	17.51	61.37
		Weight (kg)	295	89.13	16.66	87.90	50.20	174.10	701	87.41	19.79	85.40	45.00	164.50
		Height (m)	295	1.78	0.07	1.78	1.50	1.96	701	1.65	0.06	1.65	1.48	1.87
Ghana Hypertension Study	3	Age (yrs)	1452	43.65	14.43	43.00	19.00	98.00	1967	42.53	12.92	42.00	19.00	99.00
		BMI (kg/m²)	1452	23.59	3.99	22.88	14.45	50.59	1967	25.51	5.51	24.80	12.95	78.48
		Weight (kg)	1452	68.19	12.54	66.60	38.40	137.00	1967	66.30	14.81	64.60	34.40	135.00
		Height (m)	1452	1.71	0.07	1.71	1.20	1.95	1967	1.61	0.06	1.61	1.26	1.87
HANDLS	1	Age (yrs)	442	48.65	8.73	49.00	30.00	64.00	552	48.43	9.21	49.00	30.00	64.00
		BMI (kg/m²)	442	27.47	5.80	26.66	16.05	52.81	552	31.85	8.99	30.46	15.21	61.89
		Weight (kg)	442	85.78	19.22	84.00	44.00	173.00	552	85.96	24.82	81.00	40.00	179.00
		Height (m)	442	1.77	0.07	1.77	1.42	1.96	552	1.64	0.07	1.65	1.46	1.87
Health ABC	1	Age (yrs)	488	73.52	2.79	73.00	69.00	80.00	651	73.36	2.96	73.00	68.00	80.00
		BMI (kg/m²)	488	27.20	4.37	26.95	14.87	47.97	651	29.60	5.80	29.09	14.59	51.99
		Weight (kg)	488	81.53	14.59	80.45	43.80	141.00	651	75.39	15.46	74.80	33.50	138.80
		Height (m)	488	1.73	0.07	1.73	1.57	2.01	651	1.60	0.06	1.59	1.37	1.81
HUFS	1	Age (yrs)	389	40.05	12.56	46.05	20.00	83.00	536	46.91	13.45	47.00	20.00	88.00
		BMI (kg/m²)	389	28.34	7.18	27.36	15.90	70.18	536	31.59	8.66	30.63	16.16	70.32
		Weight (kg)	389	87.60	23.12	83.63	49.25	232.47	536	84.32	23.95	81.27	40.14	225.31
		Height (m)	389	1.76	0.07	1.76	1.54	1.99	536	1.63	0.07	1.63	1.30	2.04
HyperGen	1	Age (yrs)	383	44.66	13.09	45.00	20.00	85.00	798	46.00	12.90	46.00	20.00	77.00
		BMI (kg/m²)	383	29.80	6.32	28.90	16.50	52.44	798	33.90	8.33	32.62	16.17	73.68
		Weight (kg)	383	92.70	21.96	88.90	49.44	204.11	798	90.30	23.87	87.08	42.63	215.45
		Height (m)	383	1.78	0.08	1.78	1.22	2.05	798	1.64	0.07	1.65	1.27	1.88
IPM	3	Age (yrs)	1198	51.00	14.22	51.00	20.00	93.00	2272	51.93	15.30	52.00	20.00	95.00
		BMI (kg/m²)	1198	28.43	6.54	27.37	15.34	68.11	2272	31.68	8.58	30.18	14.59	79.24
		Weight (kg)	1198	90.07	22.58	86.18	48.53	224.07	2272	85.41	24.35	81.65	36.74	229.52

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	1198	1.78	0.08	1.78	1.05	2.13	2272	1.64	0.07	1.63	1.24	1.88

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
IPM Cases	2	Age (yrs)	141	58.62	12.39	59.00	27.00	87.00	206	64.87	12.08	66.00	30.00	93.00
		BMI (kg/m²)	140	27.79	6.09	26.50	17.80	46.80	206	32.15	8.29	31.20	12.10	58.60
		Weight (kg)	140	89.22	21.08	84.32	54.55	156.82	206	84.32	21.96	81.82	34.09	160.45
		Height (m)	141	1.79	0.08	1.80	1.60	1.97	206	1.62	0.07	1.63	1.32	1.80
IPM Controls	2	Age (yrs)	229	55.86	11.07	55.00	26.00	88.00	311	57.98	12.77	58.00	22.00	94.00
		BMI (kg/m²)	229	28.79	6.52	27.90	14.30	60.50	311	32.39	8.27	31.00	17.20	59.30
		Weight (kg)	229	91.25	22.31	86.36	45.45	231.82	311	86.86	23.48	81.82	44.55	177.27
		Height (m)	229	1.78	0.08	1.78	1.52	1.98	311	1.64	0.08	1.63	1.36	1.96
JHS	1	Age (yrs)	837	49.41	11.98	49.00	21.00	93.00	1296	50.32	12.12	49.00	21.00	91.00
		BMI (kg/m²)	837	30.26	6.51	29.11	17.19	65.05	1296	33.64	8.19	32.19	17.31	91.80
		Weight (kg)	837	96.16	22.37	92.00	49.10	232.40	1296	90.98	22.73	87.50	44.50	197.00
		Height (m)	837	1.78	0.07	1.78	1.58	2.00	1296	1.64	0.07	1.64	1.42	1.88
JoCo	3	Age (yrs)	206	61.31	9.92	60.00	45.00	89.00	386	61.98	10.90	60.50	45.00	93.00
		BMI (kg/m²)	206	30.09	6.41	29.44	16.14	52.29	386	33.64	8.37	32.38	17.09	70.88
		Weight (kg)	206	91.38	21.69	88.86	45.45	169.50	386	87.31	22.99	83.86	44.55	218.20
		Height (m)	206	1.74	0.07	1.73	1.56	1.92	386	1.61	0.07	1.61	1.42	1.84
Maywood	1	Age (yrs)	465	43.13	7.60	43.70	20.00	70.00	278	40.88	7.81	41.07	24.00	71.00
		BMI (kg/m²)	465	25.43	5.96	23.53	15.97	50.11	278	28.83	9.52	24.48	15.23	60.09
		Weight (kg)	465	79.14	19.56	72.95	44.55	159.09	278	77.74	25.45	68.18	40.91	170.00
		Height (m)	465	1.76	0.07	1.77	1.52	1.94	278	1.64	0.07	1.64	1.50	1.94
MEC	3	Age (yrs)	784	61.77	8.50	63.00	45.00	76.00	3543	58.68	8.82	58.00	45.00	77.00
		BMI (kg/m²)	784	27.32	4.13	26.93	15.44	45.71	3543	29.45	5.97	28.40	16.28	54.80
		Weight (kg)	784	86.39	14.50	83.91	47.63	151.05	3543	78.93	16.66	76.66	42.64	156.03
		Height (m)	784	1.78	0.07	1.78	1.57	2.03	3543	1.64	0.07	1.63	1.35	2.06
MESA	1	Age (yrs)	745	62.52	10.28	63.00	45.00	84.00	901	61.99	10.00	62.00	45.00	84.00
		BMI (kg/m²)	745	28.74	4.76	28.44	15.92	46.90	901	31.55	6.46	30.47	15.87	61.86
		Weight (kg)	745	89.00	16.42	87.63	45.50	142.61	901	82.49	17.61	79.92	39.46	158.76
		Height (m)	745	1.76	0.07	1.76	1.53	1.97	901	1.62	0.07	1.62	1.37	1.88
MESA Air/Family	1	Age (yrs)	373	58.41	8.07	58.00	45.00	82.00	574	61.49	7.5626	62.00	47.00	91.00
		BMI (kg/m²)	373	27.34	5.06	26.17	19.28	47.60	574	31.55	5.42	29.90	21.302	49.48
		Weight (kg)	373	93.1	16.57	92.1	53.1	137.20	574	84.9	19.8	80.65	40.28	186.88
		Height (m)	373	1.80	0.06	1.80	1.64	1.93	574	1.64	0.07	1.63	1.46	1.77
NCI Lung: MD Anderson Lung Cancer Epidemiology Study	2	Age (yrs)	93	59.39	9.23	60.00	35.00	80.00	127	56.21	10.30	56.00	27.00	87.00
		BMI (kg/m²)	93	29.02	4.97	28.65	18.17	45.87	127	31.89	7.16	30.73	18.05	63.20
		Weight (kg)	93	93.54	17.36	90.72	58.51	147.42	127	85.63	19.84	81.65	48.08	171.91

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	93	1.80	0.06	1.80	1.63	1.91	127	1.64	0.06	1.65	1.50	1.78

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
NCI Lung: NCI-MD	2	Age (yrs)	173	65.92	8.41	66.00	29.00	84.00	176	63.53	9.63	64.00	24.00	83.00
		BMI (kg/m²)	173	29.07	5.70	28.13	16.21	47.37	176	30.88	7.34	30.11	4.65	52.25
		Weight (kg)	173	92.12	20.39	87.09	51.26	171.91	176	82.03	19.35	80.06	44.45	142.43
		Height (m)	173	1.78	0.08	1.78	1.52	1.98	176	1.64	0.17	1.63	1.47	3.76
NCI Lung: Northern California Lung Cancer Study	2	Age (yrs)	294	63.00	11.71	65.00	29.00	93.00	320	63.58	11.10	63.50	26.00	89.00
		BMI (kg/m²)	294	28.04	5.06	27.43	15.90	46.26	320	29.93	6.45	29.11	18.79	59.86
		Weight (kg)	294	88.85	17.46	86.18	49.90	150.14	320	79.53	17.70	77.11	45.36	157.85
		Height (m)	294	1.78	0.07	1.78	1.60	1.98	320	1.63	0.07	1.63	1.42	1.91
NCI Lung: Project CHURCH	2	Age (yrs)	199	45.14	13.21	47.00	18.00	86.00	559	44.84	12.18	46.00	18.00	77.00
		BMI (kg/m²)	199	31.13	6.04	30.30	18.78	55.25	559	32.35	8.05	31.40	16.69	70.45
		Weight (kg)	199	98.53	21.07	97.02	57.88	185.97	559	86.61	22.64	83.19	43.00	182.12
		Height (m)	199	1.79	0.07	1.79	1.57	1.96	559	1.65	0.06	1.65	1.47	1.84
NCI Lung: SCCS	2	Age (yrs)	302	55.48	8.32	54.00	40.00	81.00	201	56.20	9.80	55.00	40.00	84.00
		BMI (kg/m²)	302	27.20	5.45	26.32	15.81	49.61	201	31.20	8.05	30.47	16.60	54.97
		Weight (kg)	302	85.52	18.12	82.10	47.17	145.15	201	83.60	22.51	81.65	45.36	159.21
		Height (m)	302	1.77	0.09	1.78	1.52	2.03	201	1.64	0.09	1.63	1.22	1.93
Nigeria	1	Age (yrs)	511	46.69	16.30	44.00	16.00	89.00	677	48.60	14.75	48.00	19.00	95.00
		BMI (kg/m²)	511	21.75	3.82	20.99	13.18	41.56	677	24.51	5.54	23.50	14.15	58.23
		Weight (kg)	511	63.01	12.00	60.86	37.02	113.38	677	62.44	14.59	60.33	34.29	139.00
		Height (m)	511	1.70	0.08	1.70	1.02	1.93	677	1.60	0.07	1.60	1.31	1.84
ROOT: NBCS Cases	3	Age (yrs)							692	47.68	11.95	46.00	22.00	90.00
		BMI (kg/m²)							692	25.77	5.73	25.03	15.60	49.22
		Weight (kg)							692	66.31	14.93	64.00	36.00	134.00
		Height (m)							692	1.61	0.08	1.61	1.30	1.89
ROOT: NBCS Controls	3	Age (yrs)							621	45.04	11.96	45.00	18.00	81.00
		BMI (kg/m²)							621	27.02	5.48	26.67	14.34	48.51
		Weight (kg)							621	68.57	14.63	67.00	38.00	121.00
		Height (m)							621	1.59	0.06	1.60	1.42	1.81
ROOT: BNCS Cases	3	Age (yrs)							90	56.47	14.77	54.00	30.00	88.00
		BMI (kg/m²)							90	27.68	6.46	27.05	14.02	47.05
		Weight (kg)							90	72.70	17.09	71.14	35.45	120.50
		Height (m)							90	1.62	0.06	1.62	1.48	1.76
ROOT: BNCS Controls	3	Age (yrs)							228	54.82	13.05	53.00	29.00	85.00
		BMI (kg/m²)							228	29.70	6.03	28.92	15.14	51.35
		Weight (kg)							228	76.82	16.65	74.55	35.45	134.10

Nature Genetics: doi:10.1038/ng.2608

		Height (m)		228	1.61	0.07	1.61	1.42	1.81

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
ROOT: BBCS Cases	3	Age (yrs)							92	53.38	14.13	50.00	28.00	91.00
		BMI (kg/m²)							92	30.77	6.48	30.50	16.95	48.09
		Weight (kg)							92	80.05	17.57	80.29	46.72	131.10
		Height (m)							92	1.61	0.08	1.60	1.27	1.80
ROOT: BBCS Controls	3	Age (yrs)							98	52.69	13.43	51.50	30.00	84.00
		BMI (kg/m²)							98	32.16	8.29	31.73	18.07	53.32
		Weight (kg)							98	84.82	20.61	83.23	46.27	135.60
		Height (m)							98	1.63	0.09	1.63	1.25	1.78
ROOT: RVGBC Cases	3	Age (yrs)							137	46.36	11.11	46.00	23.00	82.00
		BMI (kg/m²)							137	29.69	6.72	28.25	17.49	51.76
		Weight (kg)							137	79.86	17.98	77.11	46.27	135.20
		Height (m)							137	1.64	0.07	1.65	1.22	1.78
ROOT: RVGBC Controls	3	Age (yrs)							254	40.76	12.07	39.00	18.00	92.00
		BMI (kg/m²)							254	29.82	7.23	28.35	17.94	58.88
		Weight (kg)							254	81.03	20.17	77.11	43.09	157.40
		Height (m)							254	1.65	0.07	1.65	1.45	1.80
ROOT: CCPS Cases	3	Age (yrs)							233	46.54	10.28	46.00	23.00	82.00
		BMI (kg/m²)							233	30.38	7.29	29.62	16.94	70.53
		Weight (kg)							233	82.53	21.86	79.00	42.00	216.00
		Height (m)							233	1.65	0.08	1.65	1.37	1.88
ROOT: CCPS Controls	3	Age (yrs)							77	38.61	10.01	38.00	23.00	68.00
		BMI (kg/m²)							77	30.51	8.56	29.05	17.72	64.74
		Weight (kg)							77	84.65	24.19	81.00	49.00	172.00
		Height (m)							77	1.67	0.07	1.68	1.52	1.83
ROOT: SCCS Cases	3	Age (yrs)							215	56.76	9.12	56.95	41.96	81.73
		BMI (kg/m²)							215	32.92	7.61	32.27	16.45	57.75
		Weight (kg)							215	89.63	21.59	88.00	46.27	156.50
		Height (m)							215	1.65	0.07	1.65	1.50	1.88
ROOT: SCCS Controls	3	Age (yrs)							424	56.65	9.01	56.95	41.96	81.73
		BMI (kg/m²)							424	32.68	7.36	32.03	18.66	71.03
		Weight (kg)							424	87.77	20.64	85.73	46.27	181.90
		Height (m)							424	1.64	0.07	1.63	1.50	1.98
SAPPHIRE Cases	3	Age (yrs)	216	35.76	12.17	34.62	20.00	56.88	512	41.23	10.82	42.87	20.13	56.98
		BMI (kg/m²)	216	31.11	8.43	28.68	14.78	68.06	512	34.90	8.96	33.83	17.92	76.74
		Weight (kg)	216	98.19	28.80	92.53	49.44	211.80	512	93.53	25.44	90.04	44.45	222.30

Nature Genetics: doi:10.1038/ng.2608

		Height (m)	216	1.77	0.08	1.77	1.60	2.08	512	1.63	0.07	1.62	1.44	1.88

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
SAPPHIRE Controls	3	Age (yrs)	95	45.52	8.86	46.76	20.13	57.11	189	45.53	9.33	48.08	20.10	56.85
		BMI (kg/m²)	95	30.36	5.83	28.98	18.99	45.26	189	34.25	7.61	33.51	18.42	61.73
		Weight (kg)	95	96.22	19.71	94.35	62.60	141.50	189	90.63	21.11	89.58	47.17	184.20
		Height (m)	95	1.77	0.07	1.77	1.62	1.93	189	1.62	0.07	1.62	1.21	1.77
SIGNET-REGARDS Cases	3	Age (yrs)	414	63.66	8.77	63.00	45.00	90.00	723	64.14	8.65	64.00	45.00	92.00
		BMI (kg/m²)	414	30.92	5.75	30.20	17.10	68.10	723	34.32	7.57	33.10	19.00	73.70
		Weight (kg)	414	97.45	19.83	95.00	50.00	197.00	723	91.93	21.39	88.00	47.00	220.00
		Height (m)	414	1.77	0.07	1.78	1.42	1.96	723	1.64	0.07	1.63	1.37	2.01
SIGNET-REGARDS Controls	3	Age (yrs)	450	62.69	8.32	62.00	45.00	88.00	794	63.34	8.55	63.00	45.00	91.00
		BMI (kg/m²)	450	27.99	4.98	27.75	16.50	43.90	794	31.27	6.83	30.40	16.30	63.20
		Weight (kg)	450	87.92	16.53	87.00	52.00	147.00	794	83.67	18.79	81.00	38.00	167.00
		Height (m)	450	1.77	0.07	1.78	1.45	2.08	794	1.64	0.07	1.63	1.42	1.93
SIGNET-Sea Islands Cases	3	Age (yrs)	236	54.11	13.16	54.00	20.00	84.00	842	54.76	12.85	55.00	20.00	97.00
		BMI (kg/m²)	236	31.62	6.82	30.70	19.10	56.90	842	34.73	8.08	33.80	14.20	73.00
		Weight (kg)	236	97.58	21.90	97.00	56.00	168.00	842	92.16	21.97	90.00	35.00	181.00
		Height (m)	236	1.76	0.08	1.75	1.47	1.96	842	1.63	0.08	1.63	1.22	1.90
SIGNET-Sea Islands Controls	3	Age (yrs)	47	42.72	15.07	44.00	20.00	82.00	143	43.76	14.58	42.00	20.00	83.00
		BMI (kg/m²)	47	29.07	6.87	27.40	19.10	53.00	143	32.96	8.66	32.40	15.10	64.60
		Weight (kg)	47	90.26	18.87	87.00	62.00	158.00	143	88.92	23.90	86.00	40.00	158.00
		Height (m)	47	1.77	0.07	1.78	1.63	1.90	143	1.64	0.08	1.63	1.37	1.85
WFSM Cases	1	Age (yrs)	354	60.20	10.07	60.00	35.00	87.00	545	62.36	10.62	62.00	30.00	94.00
		BMI (kg/m²)	354	28.36	6.16	28.00	15.19	55.99	545	30.61	7.50	29.64	16.12	69.76
		Weight (kg)	354	89.55	20.02	88.53	45.40	182.51	545	81.83	20.92	79.90	39.50	202.48
		Height (m)	354	1.77	0.09	1.78	1.35	2.03	545	1.63	0.07	1.63	1.32	1.85
WFSM Controls	1	Age (yrs)	360	49.89	10.86	49.00	20.00	86.00	455	48.36	12.54	48.00	20.00	91.00
		BMI (kg/m²)	360	27.74	5.71	27.07	15.39	56.42	455	31.87	7.50	31.14	14.49	60.74
		Weight (kg)	360	88.12	18.83	86.26	49.94	158.90	455	85.13	20.84	83.08	40.86	181.60
		Height (m)	360	1.78	0.08	1.78	1.55	2.06	455	1.63	0.07	1.63	1.45	1.85
WHI-GARNET Cases	3	Age (yrs)							83	61.71	7.38	62.00	50.00	79.00
		BMI (kg/m²)							83	33.58	6.92	32.99	21.16	53.45
		Weight (kg)							83	88.68	19.04	84.90	55.40	135.50
		Height (m)							83	1.62	0.07	1.63	1.49	1.77

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 3, continued

Study	Stage	Trait	Men						Women					
			N	Mean	SD	Median	Min	Max	N	Mean	SD	Median	Min	Max
WHI-GARNET Controls	3	Age (yrs)							85	60.84	7.24	60.00	50.00	78.00
		BMI (kg/m²)							85	31.08	6.56	31.05	11.53	45.46
		Weight (kg)							85	81.33	16.95	82.00	32.00	116.00
		Height (m)							85	1.62	0.06	1.62	1.49	1.78
WHI-SHARe	1	Age (yrs)							8094	61.60	7.03	61.00	50.00	79.00
		BMI (kg/m²)							8094	31.02	6.37	30.06	16.32	60.91
		Weight (kg)							8094	81.91	17.36	79.50	45.80	130.20
		Height (m)							8094	1.62	0.06	1.62	1.46	1.77
WHI-PAGE	3	Age (yrs)							3262	61.30	7.36	61.00	50.00	79.00
		BMI (kg/m²)							3262	31.50	6.81	30.59	11.53	68.27
		Weight (kg)							3262	83.17	18.42	81.00	32.00	187.00
		Height (m)							3262	1.63	0.06	1.63	1.00	1.92

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4. SNPs from the top 1500 in Stage 1 that replicated at p<0.05 in Stage 2 and were directionally consistent, sorted by chromosome and position

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs618759	1	176152111	\N	a	g	0.6351	-0.0375	0.008	2.72E-06	0.6373	-0.0461	0.0177	0.009329	--
rs630372	1	176152385	\N	a	g	0.3369	0.0414	0.0082	4.55E-07	0.3297	0.0572	0.0181	0.001583	++
rs10913461	1	176153740	\N	c	g	0.5237	0.032	0.0078	0.0000391	0.5237	0.0416	0.0186	0.02519	++
rs543874	1	176156103	\N	a	g	0.7507	-0.0571	0.009	1.80E-10	0.7523	-0.0739	0.0195	0.0001487	--
rs506589	1	176160910	\N	t	c	0.7513	-0.0567	0.009	2.41E-10	0.7506	-0.052	0.0205	0.01124	--
rs545608	1	176165744	SEC16B	c	g	0.236	0.0531	0.011	1.32E-06	0.242	0.1758	0.0572	0.002119	++
rs575908	1	176166721	SEC16B	t	c	0.7124	-0.0411	0.0085	1.44E-06	0.7195	-0.0529	0.0189	0.005144	--
rs16837085	1	192754916	\N	a	g	0.9191	0.065	0.016	0.0000479	0.9151	0.0967	0.0291	0.000903	++
rs1320331	2	612161	\N	c	g	0.8831	0.0454	0.012	0.0001637	0.8858	0.0679	0.0272	0.01266	++
rs1320330	2	612225	\N	t	g	0.1019	-0.0612	0.0129	2.08E-06	0.1007	-0.0703	0.0289	0.01496	--
rs939582	2	612723	\N	a	g	0.1007	-0.0616	0.013	2.17E-06	0.0991	-0.072	0.0292	0.01378	--
rs2867125	2	612827	\N	t	c	0.1171	-0.045	0.0119	0.0001636	0.1214	-0.0747	0.0271	0.005823	--
rs11127483	2	613691	\N	a	g	0.8974	0.0609	0.013	2.82E-06	0.8987	0.0835	0.0296	0.004763	++
rs11127484	2	613798	\N	t	c	0.1176	-0.0447	0.0119	0.0001809	0.1147	-0.0733	0.0273	0.00728	--
rs6719518	2	613935	\N	t	c	0.8825	0.045	0.0119	0.0001636	0.8853	0.0785	0.0271	0.003737	++
rs6728726	2	613976	\N	t	c	0.1165	-0.0465	0.012	0.0001131	0.1129	-0.0792	0.0273	0.003748	--
rs6711012	2	614034	\N	c	g	0.8826	0.0466	0.0119	0.0000948	0.8853	0.0781	0.0272	0.004065	++
rs2867123	2	614524	\N	c	g	0.8824	0.0468	0.012	0.0001021	0.8854	0.0773	0.0271	0.004344	++
rs2867122	2	614581	\N	a	c	0.1174	-0.0455	0.0119	0.0001382	0.1144	-0.0784	0.0272	0.003946	--
rs2903492	2	614678	\N	a	g	0.8824	0.0462	0.0119	0.0001088	0.8857	0.0785	0.0272	0.003868	++
rs7576635	2	615057	\N	t	c	0.8826	0.0463	0.0119	0.0001052	0.8857	0.0794	0.0272	0.003524	++
rs6732471	2	619914	\N	a	g	0.8494	0.0417	0.011	0.0001457	0.8469	0.0745	0.0246	0.002407	++
rs5017300	2	621099	\N	c	g	0.1179	-0.0513	0.0119	0.0000173	0.1149	-0.0771	0.027	0.004227	--
rs7585056	2	621528	\N	a	g	0.1555	-0.0414	0.011	0.0001626	0.1578	-0.0714	0.0244	0.003392	--
rs11127485	2	622028	\N	t	c	0.8822	0.0517	0.0119	0.0000149	0.8851	0.0764	0.0268	0.004395	++
rs12623218	2	622146	\N	a	t	0.8821	0.0508	0.0119	0.0000209	0.8851	0.056	0.0249	0.0246	++
rs13012571	2	622550	\N	t	c	0.8821	0.0515	0.0119	0.000016	0.8849	0.0797	0.0267	0.002836	++

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs6548238	2	624905	\N	t	c	0.1285	-0.0501	0.0118	0.0000229	0.1287	-0.0613	0.0266	0.02126	--
rs6755502	2	625721	\N	t	c	0.1189	-0.0503	0.0119	0.0000252	0.1169	-0.0642	0.0263	0.01457	--
rs13388043	2	627597	\N	t	c	0.8818	0.0516	0.0119	0.0000154	0.8845	0.0752	0.0271	0.005507	++
rs13393304	2	627830	\N	a	g	0.118	-0.0504	0.0119	0.0000242	0.1153	-0.0779	0.0272	0.004138	--
rs7608976	2	24928785	ADCY3	a	g	0.8187	0.041	0.0104	0.0000867	0.8129	0.0616	0.0219	0.004969	++
rs13387729	2	24929630	ADCY3	a	g	0.1806	-0.041	0.0104	0.0000867	0.1769	-0.0689	0.0225	0.002207	--
rs916485	2	24935777	ADCY3	t	c	0.1754	-0.0398	0.0104	0.0001389	0.1813	-0.0665	0.0221	0.002591	--
rs7567997	2	24950456	ADCY3	t	c	0.1819	-0.0377	0.0103	0.000266	0.1878	-0.0653	0.0221	0.003076	--
rs7580081	2	24950576	ADCY3	c	g	0.824	0.0384	0.0104	0.0002368	0.8168	0.066	0.0223	0.003066	++
rs13410999	2	24951443	ADCY3	t	c	0.1807	-0.0383	0.0103	0.0002119	0.1868	-0.0661	0.0221	0.002776	--
rs6545776	2	24952861	ADCY3	a	c	0.215	-0.0375	0.0097	0.0001106	0.2206	-0.0581	0.0223	0.009233	--
rs2384058	2	24953832	ADCY3	a	g	0.1746	-0.0398	0.0106	0.0001622	0.1809	-0.064	0.0223	0.004138	--
rs2033654	2	24956612	ADCY3	a	c	0.8184	0.0405	0.0103	0.0000896	0.8198	0.0572	0.0222	0.01013	++
rs2033653	2	24956950	ADCY3	t	c	0.1832	-0.041	0.0103	0.0000732	0.1889	-0.0576	0.022	0.008807	--
rs6545790	2	24962806	ADCY3	a	g	0.2122	-0.039	0.0097	0.000058	0.2155	-0.0479	0.0206	0.02039	--
rs6749170	2	24964466	ADCY3	a	g	0.2145	-0.0389	0.0097	0.0000606	0.2182	-0.0469	0.0206	0.02311	--
rs1865687	2	24966389	ADCY3	a	g	0.7727	0.0409	0.0095	0.0000162	0.7654	0.0731	0.0204	0.0003512	++
rs7608185	2	24968936	ADCY3	c	g	0.7718	0.0394	0.0096	0.0000401	0.7671	0.0728	0.0203	0.0003274	++
rs6736711	2	24969960	ADCY3	t	c	0.7727	0.0408	0.0095	0.000017	0.7656	0.0735	0.0204	0.0003217	++
rs7586879	2	24970481	ADCY3	t	c	0.7729	0.0418	0.0095	0.0000105	0.7672	0.073	0.0203	0.0003118	++
rs6545800	2	24972389	ADCY3	t	c	0.7927	0.0383	0.0098	0.000094	0.7878	0.0687	0.0208	0.0009592	++
rs6752483	2	24973590	ADCY3	t	c	0.7805	0.0363	0.0097	0.0001823	0.7719	0.0675	0.0223	0.002428	++
rs11683212	2	24973700	ADCY3	t	c	0.2276	-0.0398	0.0095	0.0000272	0.2333	-0.0688	0.0204	0.0007485	--
rs6723803	2	24974217	ADCY3	c	g	0.7921	0.0377	0.0097	0.0001016	0.7877	0.0691	0.0208	0.0008956	++
rs6713978	2	24974355	ADCY3	t	c	0.2079	-0.0377	0.0097	0.0001016	0.2123	-0.0676	0.0209	0.001233	--
rs6756609	2	24974629	ADCY3	a	c	0.7921	0.0377	0.0097	0.0001016	0.787	0.0594	0.0224	0.008043	++
rs11688665	2	24974672	ADCY3	t	g	0.7787	0.0407	0.0097	0.0000272	0.7719	0.0765	0.0207	0.000211	++
rs4077678	2	24976344	ADCY3	c	g	0.1644	-0.0458	0.0109	0.0000252	0.1734	-0.0661	0.0218	0.002382	--

Nature Genetics: doi:10.1038/ng.2608

rs4077679	2	24976356	ADCY3	t	c	0.2303	-0.0403	0.0095	0.0000216	0.2362	-0.0779	0.0207	0.000162	--

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs4077680	2	24976446	ADCY3	a	g	0.2304	-0.0406	0.0095	0.0000187	0.2365	-0.0736	0.0203	0.0002802	--
rs3903070	2	24976967	ADCY3	c	g	0.2079	-0.0382	0.0098	0.000098	0.2123	-0.0688	0.0208	0.000963	--
rs6545807	2	24977862	ADCY3	t	c	0.769	0.0399	0.0096	0.0000319	0.764	0.0649	0.0216	0.002688	++
rs6545808	2	24977892	ADCY3	a	c	0.7694	0.0402	0.0095	0.0000226	0.7633	0.0733	0.0203	0.0002965	++
rs6712981	2	24979734	ADCY3	a	g	0.2103	-0.038	0.0097	0.0000894	0.2143	-0.0612	0.0201	0.002269	--
rs6726199	2	24979832	ADCY3	c	g	0.7897	0.038	0.0097	0.0000894	0.7853	0.0682	0.0208	0.001068	++
rs6545809	2	24980219	ADCY3	t	c	0.7896	0.038	0.0097	0.0000894	0.7856	0.0519	0.0203	0.01065	++
rs6706316	2	24981855	ADCY3	c	g	0.208	-0.038	0.0097	0.0000894	0.2123	-0.0505	0.0204	0.01344	--
rs10200566	2	24983966	ADCY3	t	g	0.1868	-0.0417	0.0102	0.000046	0.1956	-0.0727	0.0216	0.0007582	--
rs11900505	2	24985490	ADCY3	a	c	0.2225	-0.035	0.0095	0.0002247	0.2267	-0.0488	0.0199	0.01435	--
rs11676272	2	24995042	ADCY3	a	g	0.1656	-0.0429	0.0109	0.0000795	0.1742	-0.0663	0.0286	0.02055	--
rs6752378	2	25003620	\N	a	c	0.8336	0.0438	0.0108	0.0000473	0.8237	0.0621	0.0231	0.0072	++
rs10182181	2	25003800	\N	a	g	0.1673	-0.0427	0.0108	0.000073	0.176	-0.0635	0.0231	0.005949	--
rs2384054	2	25010277	\N	t	c	0.1562	-0.0413	0.0111	0.0001948	0.1584	-0.0608	0.0228	0.007601	--
rs713587	2	25011785	\N	t	c	0.84	0.0417	0.011	0.0001457	0.8356	0.0581	0.0234	0.01295	++
rs1172294	2	25022704	RBJ	a	g	0.1408	-0.0515	0.0117	0.0000112	0.1399	-0.0566	0.0254	0.02601	--
rs1982200	2	25058931	LOC729723	t	c	0.1397	-0.0501	0.0118	0.0000229	0.1397	-0.0566	0.0254	0.02564	--
rs12466350	2	25093473	LOC729723	t	c	0.1381	-0.0498	0.0118	0.0000256	0.1392	-0.0547	0.0252	0.03033	--
rs1077492	2	25139939	EFR3B	t	c	0.1384	-0.0491	0.0118	0.0000333	0.1421	-0.0506	0.0253	0.04583	--
rs483428	2	25158640	EFR3B	t	c	0.1164	-0.0503	0.0126	0.0000636	0.115	-0.0538	0.0265	0.04218	--
rs1530016	2	25159260	EFR3B	t	c	0.1179	-0.0471	0.0125	0.0001589	0.12	-0.062	0.0273	0.02286	--
rs2172169	2	43611369	THADA	c	g	0.4495	-0.0297	0.0079	0.0001663	0.4508	-0.0392	0.0186	0.03552	--
rs6717153	2	43617033	THADA	a	g	0.5716	0.0313	0.008	0.0000903	0.5749	0.037	0.0172	0.03155	++
rs10190234	2	43864824	DYNC2LI1	t	c	0.1646	0.0405	0.0107	0.0001449	0.1659	0.0462	0.0225	0.03976	++
rs1835815	2	43884153	DYNC2LI1	t	c	0.6591	0.0335	0.0082	0.0000447	0.66	0.0431	0.0181	0.01715	++
rs17031599	2	43884679	DYNC2LI1	a	g	0.6514	0.0359	0.0083	0.0000157	0.6515	0.0401	0.0182	0.02721	++
rs3792012	2	43887400	DYNC2LI1	a	g	0.6608	0.033	0.0082	0.000058	0.6614	0.0453	0.0179	0.01145	++
rs2278356	2	43893379	ABCG5	a	c	0.3398	-0.0313	0.0082	0.0001369	0.3397	-0.0377	0.0181	0.03746	--

Nature Genetics: doi:10.1038/ng.2608

rs1864815	2	43896956	ABCG5	a	t	0.4132	-0.0334	0.0082	0.0000471	0.4101	-0.0573	0.0233	0.01417	--

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs7618262	3	14550447	GRIP2	t	c	0.7551	0.0385	0.0098	0.0000864	0.7416	0.0694	0.0201	0.0005479	++
rs4688562	3	65367862	MAGI1	a	t	0.4728	-0.0323	0.0087	0.0002195	0.4696	-0.0453	0.0201	0.0239	--
rs7630511	3	180755175	\N	a	g	0.1742	0.0379	0.0102	0.0002124	0.1774	0.0692	0.0228	0.002393	++
rs12492220	3	180761462	ACTL6A	a	g	0.1719	0.0391	0.0103	0.0001558	0.1759	0.0702	0.0229	0.002151	++
rs6795642	3	180785574	ACTL6A	t	g	0.8161	-0.0389	0.0103	0.0001683	0.8172	-0.0653	0.0238	0.006177	--
rs7635877	3	180796981	MRPL47	a	g	0.1798	0.0394	0.0103	0.0001386	0.1791	0.0653	0.0238	0.006136	++
rs9882051	3	180818399	NDUFB5	a	t	0.187	0.0404	0.0106	0.0001289	0.1915	0.0442	0.0225	0.04946	++
rs1401337	3	181920181	\N	t	g	0.8186	-0.0441	0.0102	0.0000163	0.7996	-0.0468	0.0221	0.03418	--
rs13079982	3	181931334	\N	a	g	0.1193	0.0515	0.0132	0.0000976	0.1281	0.0875	0.039	0.02491	++
rs13078931	3	184261635	MCCC1	a	g	0.1728	0.0425	0.011	0.0001083	0.1725	0.0513	0.0243	0.03461	++
rs2348070	4	44177369	\N	a	g	0.4856	0.0332	0.0078	0.0000198	0.4845	0.0375	0.0175	0.03159	++
rs13130484	4	44870448	\N	t	c	0.2461	0.0479	0.009	8.80E-08	0.2422	0.0649	0.0199	0.001099	++
rs12641981	4	44874640	\N	t	c	0.2406	0.047	0.0091	2.13E-07	0.2366	0.0637	0.0199	0.001378	++
rs1581095	4	44874954	\N	a	c	0.42	0.0421	0.0084	5.74E-07	0.4293	0.0625	0.0202	0.001929	++
rs10938397	4	44877284	\N	a	g	0.751	-0.0534	0.0095	1.81E-08	0.7502	-0.0593	0.0238	0.01264	--
rs348495	4	44879199	\N	a	g	0.6596	-0.048	0.0086	2.70E-08	0.6582	-0.0674	0.0208	0.001192	--
rs348500	4	44881589	\N	t	c	0.422	0.034	0.0078	0.0000125	0.4184	0.0611	0.0172	0.0003811	++
rs11100011	4	156939254	GUCY1B3	a	t	0.9625	-0.1151	0.0304	0.0001513	0.9498	-0.179	0.0667	0.007236	--
rs922133	4	156947894	GUCY1B3	t	g	0.9642	-0.0979	0.0259	0.000157	0.958	-0.136	0.0452	0.002615	--
rs1459854	4	156948690	\N	t	c	0.9647	-0.1068	0.0266	0.0000613	0.9497	-0.1893	0.0662	0.00423	--
rs10011335	4	156948956	\N	t	c	0.0352	0.1069	0.0266	0.0000603	0.0502	0.188	0.0662	0.00455	++
rs10011344	4	156948992	\N	a	c	0.0327	0.0911	0.024	0.0001455	0.0499	0.189	0.0671	0.004878	++
rs10014214	4	156949423	\N	a	g	0.035	0.1084	0.0268	0.0000508	0.0493	0.1913	0.0681	0.004955	++
rs355168	4	166910942	\N	t	c	0.8613	0.0524	0.0142	0.0002188	0.8608	0.0655	0.0261	0.01209	++
rs6858482	4	175185614	\N	a	g	0.4983	-0.0296	0.0078	0.0001423	0.4892	-0.043	0.0174	0.01331	--
rs1373976	5	13994774	DNAH5	a	g	0.818	0.0452	0.01	6.45E-06	0.8291	0.0497	0.0226	0.02743	++
rs261973	5	95895183	\N	t	c	0.7721	-0.0364	0.0095	0.0001245	0.7771	-0.0444	0.0203	0.02902	--
rs750601	5	116825513	\N	t	c	0.1228	0.0439	0.012	0.0002676	0.1319	0.0697	0.0325	0.03199	++

Nature Genetics: doi:10.1038/ng.2608

rs12054772	5	153489550	\N	a	g	0.4043	-0.0302	0.0079	0.0001288	0.4184	-0.0548	0.0172	0.001473	--

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Stage 1				Stage 2				Dir
						Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	
rs2033195	5	153489789	\N	t	c	0.4555	-0.0292	0.0079	0.0002139	0.4503	-0.0468	0.0171	0.006088	--
rs6890277	5	153492593	\N	a	g	0.4535	-0.0298	0.0078	0.0001282	0.4492	-0.044	0.018	0.01437	--
rs815611	5	153498959	\N	a	g	0.4409	-0.0284	0.0078	0.0002624	0.4437	-0.057	0.017	0.0008173	--
rs7715256	5	153518086	\N	t	g	0.6147	-0.041	0.0082	5.87E-07	0.6086	-0.0552	0.0175	0.001626	--
rs7708584	5	153523659	\N	a	g	0.3162	0.0498	0.0086	8.02E-09	0.3172	0.047	0.0181	0.009354	++
rs1366219	5	153525891	\N	a	t	0.3161	0.0498	0.0086	8.02E-09	0.3175	0.0456	0.018	0.01141	++
rs10052189	5	153535357	\N	t	c	0.5511	-0.0404	0.0081	6.13E-07	0.552	-0.0471	0.017	0.0056	--
rs2351228	5	153537783	\N	a	g	0.5513	-0.0401	0.0081	7.42E-07	0.5374	-0.0472	0.017	0.005467	--
rs10038664	5	153550122	GALNT10	a	g	0.4306	-0.0302	0.0078	0.0001039	0.4303	-0.0371	0.017	0.02936	--
rs6868044	5	153553772	GALNT10	a	c	0.4243	-0.0303	0.0078	0.0000986	0.4362	-0.043	0.017	0.01141	--
rs6580057	5	153555538	GALNT10	t	c	0.4241	-0.0304	0.0079	0.0001161	0.4363	-0.0436	0.017	0.01034	--
rs2206277	6	50906485	TFAP2B	t	c	0.1417	0.0454	0.0113	0.0000586	0.1379	0.0677	0.0255	0.007825	++
rs6937736	6	97506493	KLHL32	c	g	0.2727	-0.0346	0.0086	0.0000614	0.2807	-0.045	0.0187	0.01603	--
rs6937950	6	97506648	KLHL32	a	g	0.2727	-0.0346	0.0086	0.0000614	0.2809	-0.0442	0.0189	0.01956	--
rs7769945	6	97507560	KLHL32	a	g	0.7263	0.0347	0.0086	0.0000584	0.7098	0.0441	0.019	0.02042	++
rs13211612	6	97513568	KLHL32	a	g	0.2677	-0.0364	0.0087	0.0000312	0.2793	-0.0459	0.0197	0.01992	--
rs17057164	6	97517257	KLHL32	t	c	0.6576	0.0393	0.0082	1.68E-06	0.6529	0.0554	0.0189	0.003445	++
rs11153285	6	97519872	KLHL32	t	c	0.3688	-0.0386	0.008	1.38E-06	0.3732	-0.0426	0.0176	0.01535	--
rs10485381	6	97520548	KLHL32	a	g	0.6312	0.038	0.008	2.00E-06	0.6189	0.0441	0.0177	0.01267	++
rs7759572	6	97521856	KLHL32	c	g	0.631	0.0387	0.008	1.29E-06	0.6262	0.0425	0.0176	0.0157	++
rs974417	6	97526319	KLHL32	t	c	0.3431	-0.0395	0.0082	1.49E-06	0.357	-0.0526	0.018	0.003466	--
rs2143389	6	97529402	KLHL32	t	c	0.3435	-0.0391	0.0082	1.90E-06	0.3572	-0.0524	0.018	0.003658	--
rs6903798	6	139622485	TXLNB	t	c	0.882	0.0539	0.0139	0.0001003	0.8756	0.0572	0.0289	0.04763	++
rs11753530	6	170226223	\N	t	g	0.1346	-0.0425	0.0114	0.0001942	0.1382	-0.051	0.0245	0.03758	--
rs17144816	7	21643483	DNAH11	a	t	0.9231	0.0645	0.0175	0.0002245	0.9177	0.1147	0.0495	0.02063	++
rs2035425	7	41544492	\N	a	g	0.6356	-0.0307	0.0082	0.0001836	0.649	-0.0432	0.018	0.01614	--
rs7077844	10	90302333	C10orf59	t	c	0.8446	0.0552	0.0147	0.0001749	0.8422	0.0822	0.0293	0.005041	++
rs11819808	11	27637964	BDNF	t	c	0.241	0.0338	0.0093	0.0002675	0.2274	0.0673	0.0202	0.0008532	++

Nature Genetics: doi:10.1038/ng.2608

rs7116340	11	84829155	\N	a	t	0.593	0.0295	0.0079	0.000184	0.5928	0.0406	0.0191	0.03336	++

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs7108137	11	84925621	\N	a	g	0.2916	-0.0319	0.0085	0.0001833	0.2863	-0.0413	0.0194	0.03316	--
rs7973811	12	106536792	BTBD11	c	g	0.9569	0.0969	0.0238	0.0000457	0.9621	0.1646	0.0833	0.04799	++
rs1788062	13	95425793	UGCGL2	t	g	0.8024	-0.0362	0.0097	0.0001899	0.7978	-0.0438	0.022	0.04673	--
rs10150332	14	79006717	NRXN3	t	c	0.6072	-0.0344	0.0083	0.0000351	0.6077	-0.0446	0.0184	0.01522	--
rs7156625	14	79012400	NRXN3	a	g	0.3598	0.0342	0.0085	0.0000605	0.3743	0.0359	0.0179	0.04504	++
rs17763953	14	79038947	NRXN3	t	c	0.256	-0.0335	0.0091	0.0002177	0.2611	-0.0409	0.0195	0.03635	--
rs12593784	15	44667342	\N	t	c	0.066	0.0637	0.0163	0.0000938	0.0691	0.061	0.0308	0.04788	++
rs4506844	15	70296192	PKM2	t	c	0.4656	0.0296	0.0078	0.0001423	0.4592	0.0454	0.0217	0.03609	++
rs1421085	16	52358455	FTO	t	c	0.8865	-0.0602	0.0153	0.0000878	0.9036	-0.1046	0.0405	0.00973	--
rs1558902	16	52361075	FTO	a	t	0.1134	0.0658	0.0128	2.69E-07	0.1122	0.0805	0.0266	0.002503	++
rs3751812	16	52375961	FTO	t	g	0.1118	0.0676	0.0126	7.67E-08	0.0996	0.0618	0.0292	0.03428	++
rs9941349	16	52382989	FTO	t	c	0.1866	0.044	0.01	0.0000113	0.1798	0.0488	0.0223	0.0288	++
rs9931494	16	52384680	FTO	c	g	0.8163	-0.047	0.0101	3.46E-06	0.8168	-0.0504	0.0223	0.02402	--
rs17817964	16	52385567	FTO	t	c	0.1168	0.0739	0.0124	2.27E-09	0.1123	0.0675	0.027	0.01249	++
rs12149832	16	52400409	FTO	a	g	0.1185	0.0716	0.0126	1.25E-08	0.1152	0.0714	0.0263	0.006545	++
rs11642841	16	52402988	FTO	a	c	0.1221	0.071	0.0132	7.79E-08	0.1184	0.0812	0.0267	0.002361	++
rs1075184	17	67363529	\N	a	c	0.7837	0.0434	0.0114	0.0001416	0.7872	0.1015	0.0287	0.0003945	++
rs1539952	18	55917492	\N	a	g	0.7811	-0.0573	0.0094	1.00E-09	0.7779	-0.0419	0.0205	0.04078	--
rs6567160	18	55980115	\N	t	c	0.793	-0.0621	0.0098	2.41E-10	0.7768	-0.045	0.021	0.03235	--
rs492443	18	56009782	\N	a	g	0.5051	-0.0294	0.0078	0.0001578	0.5093	-0.0458	0.0176	0.009077	--
rs11663816	18	56027207	\N	t	c	0.8912	-0.0559	0.0125	7.38E-06	0.8881	-0.0594	0.0274	0.03034	--
rs12964203	18	56054584	\N	t	c	0.8931	-0.0621	0.0126	7.92E-07	0.8837	-0.0628	0.0277	0.02331	--
rs2168708	18	56058291	\N	t	g	0.1069	0.0622	0.0126	7.60E-07	0.1105	0.0546	0.0267	0.04043	++
rs2287019	19	50894012	QPCTL	t	c	0.1165	-0.0659	0.0149	0.0000101	0.1132	-0.0579	0.0288	0.04419	--
rs6039485	20	9448584	C20orf103	t	c	0.3568	-0.0343	0.009	0.0001277	0.3489	-0.0435	0.0197	0.02739	--
rs6014231	20	53009476	\N	a	g	0.1642	-0.0418	0.0104	0.0000629	0.1604	-0.0468	0.0234	0.04563	--
rs6023745	20	53009511	\N	t	g	0.8358	0.0418	0.0104	0.0000629	0.8398	0.0469	0.0234	0.04505	++
rs990571	20	53010252	\N	t	c	0.1635	-0.0412	0.0104	0.0000801	0.1579	-0.0471	0.0222	0.03396	--

Nature Genetics: doi:10.1038/ng.2608

rs1108356	20	53011027	\N	t	c	0.164	-0.0417	0.0104	0.0000655	0.1594	-0.0464	0.0234	0.0468	--

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 4, continued

						Stage 1				Stage 2				
SNP	Chr	Position (Build 36)	Gene	Coded allele	Other allele	Coded allele freq	Beta	SE	P-value	Coded allele freq	Beta	SE	P-value	Dir
rs6014233	20	53011496	\N	a	g	0.164	-0.0417	0.0104	0.0000655	0.1593	-0.0464	0.0234	0.04698	--
rs878694	20	53011629	\N	c	g	0.836	0.0417	0.0104	0.0000655	0.8407	0.0465	0.0234	0.04652	++
rs6014234	20	53012830	\N	a	c	0.164	-0.0417	0.0104	0.0000655	0.1593	-0.0464	0.0234	0.04678	--
rs6069071	20	53014114	\N	a	g	0.8363	0.041	0.0104	0.0000867	0.8416	0.0473	0.0234	0.0433	++
rs991212	20	53014846	\N	a	g	0.1632	-0.0419	0.0104	0.0000604	0.1581	-0.0468	0.0234	0.04526	--
rs135571	22	44907872	\N	a	g	0.3981	-0.0399	0.0097	0.000039	0.4037	-0.0461	0.0197	0.0195	--
rs6520141	22	48674349	\N	a	g	0.3481	-0.0305	0.0083	0.000244	0.3549	-0.0403	0.018	0.02521	--

SNP, single nucleotide polymorphism; Chr, chromosome; SE, standard error; Dir, direction of effect

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 5. Study- and stage-specific results for novel BMI SNP rs10261878[a] discovered via meta-analysis of the top 1500 African ancestry BMI hits with GWAS data of European ancestry[b]

Study	N[c]	β (SE)	p-value
African ancestry			
African ancestry Stage 1	39,101	0.0301 (0.0080)	1.66E-04
African ancestry Stage 2	6,817	0.0170 (0.0174)	0.3302
African ancestry Stage 3	25,308	0.0290 (0.0089)	1.01E-03
African ancestry Stages 1+2	45,918	0.0278 (0.0073)	1.30E-04
African ancestry Stages 1+2+3	71,226	0.0280 (0.0056)	4.57E-07
GIANT (European ancestry)			
GIANT	123,706	0.0453 (0.0107)	2.21E-05
Combined (African + European)			
GIANT + African ancestry (S1+S2)	169,624	0.0333 (0.0060)	2.95E-08
GIANT + African ancestry (S1+S2+S3)	194,931	0.0320 (0.0050)	1.23E-10

[a]SNP rs10261878 is on chromosome 7, position (Build 36) 25917070; effect/other alleles C/A; effect allele frequencies 0.94 (GIANT)/0.44 (S1 African descent); *MIR148A* is closest gene (~72 kb away); [b]European ancestry data from the GIANT consortium[30]; [c]Ns reported reflect individuals with data for SNP rs10261878 (may differ from total stage-specific sample size); SNP, single nucleotide polymorphism; β (beta estimate) reported in inverse-normally transformed units; SE, standard error

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 6. Generalization of two novel and one suggestive SNPs discovered through African ancestry GWAS in individuals of European and Asian ancestry

			African ancestry[a]				European ancestry[b]				Asian ancestry[c]			
SNP (Locus)	**Alleles (effect/other)**	**EAF (AA,EA,AS)[d]**	**N**	**β (SE)**	**P**	**Explained variance (%)[e]**	**N**	**β (SE)**	**P**	**Explained variance (%)**	**N**	**β (SE)**	**P**	**Explained variance (%)**
rs7708584 (*GALNT10*)	A/G	0.31,0.42,0.96	70373	0.040 (0.006)	3.37E-11	0.07%	123,864	0.021 (0.0047)	1.20E-05	0.02%	68,906	0.048 (0.016)	0.002	0.02%
rs974417 (*KLHL32*)	C/T	0.65,0.85,0.80	71387	0.031 (0.006)	6.88E-08	0.04%	123,860	0.0049 (0.0065)	0.45	0.01%	68,914	0.015 (0.006)	0.023	0.01%
rs10261878 (*MIR148A/NFE2L3*)	C/A	0.44,0.94,0.96	71226	0.028 (0.006)	4.57E-07	0.04%	123,706	0.0453 (0.0107)	2.21E-05	0.02%	60,063	-0.030 (0.015)	0.053	0.01%
Sum of variance explained						**0.15%**				**0.05%**				**0.03%**

[a]African ancestry BMI results from Stage 1+Stage2+Stage3; [b]European ancestry results from GIANT [30]; [c]Asian ancestry results from meta-analysis of [31,32]; [d]African ancestry (AA) effect allele frequency from Stage 1, European ancestry (EA) effect allele frequency from [30], Asian ancestry (AS) effect allele frequency from [32]; [e]The total fraction of variance explained was calculated using the formula $2f(1-f)*a^2$, where f is the frequency of the variant and a is the additive effect of the variant (see reference [33]), African ancestry values taken from Stage 1+2+3; AA, African ancestry; EA, European ancestry, AS, Asian ancestry; EAF, effect allele frequency

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 7. Associations of two novel and one suggestive SNPs discovered through African ancestry GWAS in 3751 African American children from the Pediatric Research Consortium (PeRC) at the Children's Hospital of Philadelphia.

SNP	Locus	Alleles (effect/other)	EAF	β (SE)	P	Directionally consistent?
rs7708584	*GALNT10*	A/G	0.30	0.0461 (0.0253)	0.0686	Yes
rs974417	*KLHL32*	C/T	0.65	0.0212 (0.0245)	0.3853	Yes
rs10261878	*MIR148A/NFE2L3*	C/A	0.42	0.0056 (0.0232)	0.8091	Yes

EAF, effect allele frequency

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 8. Two novel and one suggestive SNPs for BMI discovered through African ancestry GWAS, adjusted and unadjusted for local ancestry estimates in ARIC, CARDIA, JHS, MESA, WHI, AABC and AAPC studies (n=27,011, 69% of stage 1 sample)

			Pre-adjustment		Post-adjustment	
SNP	**Locus**	**Alleles**	**β (SE)**	**p-value**	**β (SE)**	**p-value**
rs7708584	*GALNT10*	A/G	0.0535 (0.0097)	2.97E-08	0.0647 (0.0116)	2.64E-08
rs974417	*KLHL32*	C/T	0.0446 (0.0092)	1.36E-06	0.0441 (0.0129)	6.59E-04
rs10261878	*MIR148A/NFE2L3*	C/A	0.0341 (0.009)	1.48E-04	0.0466 (0.0129)	2.90E-04

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 9. BMI association results for the two novel and one suggestive variants stratified by the number of European chromosomes in each region

SNP	Alleles (test/reference)	No. of Eur chr	N	β (SE)	P-value
rs7708584	A/G	2	6221	0.0728 (0.017)	2.1E-05
		1	6524	0.0503 (0.021)	0.015
		0	9030	0.0564 (0.22)	0.012
Test of heterogeneity p=0.71					
rs974417	T/C	2	3964	-0.0228 (0.07)	0.74
		1	8042	-0.0395 (0.023)	0.085
		0	9766	-0.0315 (0.015)	0.04
Test of heterogeneity p=0.75					
rs10261878	A/C	2	1732	-0.1131 (0.12)	0.36
		1	8116	-0.0456 (0.024)	0.06
		0	11927	-0.0337 (0.015)	0.024
Test of heterogeneity p=0.95					

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 10. P-values for the association of the two novel and one suggestive SNPs discovered through African ancestry GWAS with waist circumference and waist-hip ratio (unadjusted and adjusted for BMI) using data from the COGENT African ancestry waist consortium (N~20,000) and with height using data from the African Ancestry Height Consortium (N ~37,000)

SNP (locus)	Alleles (effect/other)	Waist Circumference		Waist-Hip Ratio		Height
		Unadjusted	Adjusted for BMI	Unadjusted	Adjusted for BMI	
rs7708584 (*GALNT10*)	A/G	1.14E-04	0.892	0.0953	0.8375	0.1984
rs974417 (*KLHL32*)	C/T	9.20E-06	0.4881	0.0569	0.9365	0.811
rs10261878 (*MIR148A/NFE2L3*)	C/A	4.97E-04	0.1747	0.1272	0.5278	0.0708

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 11. Pleiotropy analysis of the two novel and one suggestive variants in meta-analyses of GWAS for 13 related traits

		rs7708584 *GALNT10*		rs974417 *KLHL32*		rs10261878 *MIR148A/NFE2L3*	
Phenotype	**Reference**	**P-value**	**Direction of effect for A allele[a]**	**P-value**	**Direction of effect for C allele[a]**	**P-value**	**Direction of effect for C allele[a]**
Waist-Hip Ratio	Heid et al, Nature Genetics, 2010	0.72	**+**	**0.01**	+	0.91	+
Height	Lango Allen et al. Nature 2010	0.30	-	0.41	-	0.93	+
Systolic BP	Georg B. Ehret et al. Nature 2011	0.26	+	0.85	+	**0.03**	+
Diastolic BP	Georg B. Ehret et al. Nature 2011	0.10	+	0.47	-	0.19	+
HDL-Cholesterol	Teslovich et al. Nature 2010	0.10	-	0.99	+	0.64	-
LDL -Cholesterol	Teslovich et al. Nature 2010	0.32	+	0.29	**+**	**0.005**	**+**
Total cholesterol	Teslovich et al. Nature 2010	0.65	-	0.31	+	0.14	+
Triglycerides	Teslovich et al. Nature 2010	0.99	+	0.64	+	0.34	-
Fasting glucose	Saxena R et al. Nature Genetics 2010	0.54	+	0.80	**-**	**0.002**	**+**
2 hr glucose	Dupuis J et al. Nature Genetics 2010	0.36	**-**	**0.02**	-	0.90	-
Fasting Insulin	Dupuis J et al. Nature Genetics 2010	0.08	**+**	**0.03**	-	0.25	+
HOMA-B	Dupuis J et al. Nature Genetics 2010	0.44	+	0.37	-	0.60	-
HOMA-IR	Dupuis J et al. Nature Genetics 2010	0.21	+	0.10	-	0.38	+

[a]BMI-increasing allele

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 12. Associations across populations of European ancestry, African ancestry, and Asian ancestry with the 36 previou

						European ancestry[a]				African ancestry[b]				
Locus	SNP	Proxy r^2 (CEU)	Chr	Alleles (effect/other)	EAF[d] (CEU, AA, CHB+JPT)	N	β (SE)	P	Explained variance (%)[e]	N	β (SE)	P	Explained variance (%)	N
NEGR1	rs2815752 rs2568958 (Asian proxy)	1.0	1	A/G A/G	0.67, 0.54, 0.90 0.68, 0.54, 0.90	198,380 --	0.0351 (0.0036) --	1.61E-22 --	0.055% --	38,921 --	0.0045 (0.0078) --	0.563 --	1.01E-03% --	-- 64,40
TNNI3K	rs1514175		1	A/G	0.41, 0.66, 0.82	227,900	0.0240 (0.0032)	8.16E-14	0.028%	39,094	-0.0074 (0.0081)	0.361	2.46E-03%	64,40
PTBP2	rs1555543		1	C/A	0.43, 0.43, 0.11	243,013	0.0198 (0.0032)	3.68E-10	0.019%	39,088	0.0031 (0.0078)	0.690	4.71E-04%	64,40
SEC16B	rs543874 rs574367 (Asian proxy)	0.97	1	G/A T/G	0.28, 0.25, 0.18 0.26, 0.11, 0.18	179,414 --	0.0470 (0.0047) --	3.56E-23 --	0.089% --	38,899 --	0.0571 (0.0090) --	1.80E-10 --	0.122% --	-- 64,40
TMEM18	rs2867125 rs11127485 (Asian proxy)	0.95	2	C/T T/C	0.86, 0.88, 0.91 0.86, 0.88, 0.90	197,806 --	0.0681 (0.0046) --	2.77E-49 --	0.112% --	38,994 --	0.0450 (0.0119) --	1.64E-04 --	0.0428% --	-- 64,40
RBJ	rs713586		2	C/T	0.51, 0.84, 0.51	230,748	0.0309 (0.0032)	6.17E-22	0.048%	38,764	0.0378 (0.0111)	6.50E-04	0.0384%	64,40
FANCL	rs887912 rs2192497 (AA proxy)	1.0	2	T/C C/T	0.33, --, 0 0.33, 0.09, 0	242,807 --	0.0242 (0.0034) --	1.79E-12 --	0.026% --	-- 38,630	-- -0.0054 (0.0138)	-- 0.695	-- 4.78E-04%	
LRP1B	rs2890652 rs1533942 (AA proxy)	0.70	2	C/T C/T	0.18, --, 0 0.12, 0.22, 0	209,068 --	0.0282 (0.0044) --	1.35E-10 --	0.024% --	-- 39,135	-- -0.0004 (0.0094)	-- 0.966	-- 5.49E-06%	
CADM2	rs13078807		3	G/A	0.23, 0.06, 0	237,404	0.0260 (0.0039)	3.94E-11	0.024%	37,375	0.0070 (0.0169)	0.6796	5.53E-04%	
ETV5	rs9816226 rs7647305 (Asian proxy)	0.85	3	T/A C/T	0.84, 0.80, 1 0.79, 0.61, 0.97	196,221 --	0.0405 (0.0046) --	1.69E-18 --	0.044% --	39,136 --	0.0424 (0.0102) --	3.42E-05 --	0.0575% --	-- 64,40
GNPDA2	rs10938397		4	G/A	0.45, 0.25, 0.28	197,008	0.0423 (0.0036)	3.78E-31	0.089%	38,221	0.0534 (0.0095)	1.81E-08	0.107%	64,40
SLC39A8	rs13107325		4	T/C	0.09, 0.02, 0	245,378	0.0512 (0.0069)	1.50E-13	0.043%	19,615	0.0530 (0.0431)	0.218	0.011%	
FLJ35779	rs2112347		5	T/G	0.67, 0.49, 0.47	231,729	0.0241 (0.0033)	2.17E-13	0.026%	39,110	0.0021 (0.0077)	0.784	2.2E-04%	64,40
ZNF608	rs4836133		5	A/C	0.48, 0.80, 0.23	241,999	0.0192 (0.0032)	1.97E-09	0.018%	21,202	0.0139 (0.0141)	0.323	6.18E-03%	
PCSK1	rs261967		5	C/A	0.39, 0.40, 0.47	123,864	0.0150 (0.0047)	0.0014	0.011%	39,120	0.0262 (0.0079)	8.95E-04	0.0329%	64,40
NUDT3	rs206936		6	G/A	0.20, 0.53, 0.47	249,777	0.0208 (0.0038)	3.02E-08	0.014%	39,142	0.0060 (0.0078)	0.441	1.79E-03%	64,40
TFAP2B	rs987237		6	G/A	0.08, 0.10, 0.13	195,776	0.0418 (0.0045)	2.90E-20	0.026%	38,725	0.0510 (0.0128)	6.68E-05	0.0468%	64,40
CDKAL1	rs2206734		6	C/T	0.87, 0.77, 0.54	123,852	0.0172 (0.0061)	0.0047	0.007%	39,121	0.0110 (0.0092)	0.2301	4.29E-03%	64,40

Nature Genetics: doi:10.1038/ng.2608

LRRN6C	rs10968576		9	G/A	0.36, 0.17, 0.23	216,916	0.0257 (0.0035)	2.65E-13	0.030%	39,117	0.0374 (0.0103)	2.98E-04	0.0395%	64,40
KLF9	rs11142387		9	C/A	0.55, 0.37, 0.44	123,853	0.0033 (0.0048)	0.4937	0.0005%	39,096	0.0028 (0.0081)	0.7296	3.66E-04%	64,40
RPL27A	rs4929949 rs9300092 (AA proxy)	1.0	11	C/T G/A	0.59, --, 0.51 0.60, 0.41, 0.51	249,791 --	0.0183 (0.0031) --	2.80E-09 --	0.016% --	-- 38,882	-- 0.0012 (0.0079)	-- 0.879	-- 6.97E-05%	64,40 --
BDNF	rs10767664 rs6265 (Asian proxy)	0.77	11	A/T G/A	0.76, 0.91, 0.54 0.81, 0.96, 0.54	204,158 --	0.0445 (0.0042) --	4.69E-26 --	0.072% --	37,781 --	0.0576 (0.0157) --	2.37E-04 --	0.0543% --	-- 64,40
MTCH2	rs3817334		11	T/C	0.41, 0.27, 0.38	191,943	0.0256 (0.0036)	1.59E-12	0.032%	39,049	0.0083 (0.0087)	0.342	2.72E-03%	64,40
FAIM2	rs7138803		12	A/G	0.37, 0.18, 0.43	200,064	0.0303 (0.0036)	1.82E-17	0.043%	38,918	0.0470 (0.0102)	4.37E-06	0.0652%	64,40
MTIF3	rs4771122		13	G/A	0.27, 0.20, 0.18	198,577	0.0258 (0.0042)	9.48E-10	0.026%	39,127	0.0010 (0.0103)	0.923	3.2E-05%	64,40
PRKD1	rs11847697		14	T/C	0.03, 0.70, 0	241,667	0.0554 (0.0085)	5.76E-11	0.018%	38,217	-0.0041 (0.0092)	0.655	7.06E-04%	
NRXN3	rs10150332		14	C/T	0.27, 0.39, 0	183,022	0.0296 (0.0044)	2.75E-11	0.035%	38,218	0.0344 (0.0083)	3.51E-05	0.0563%	
MAP2K5	rs2241423		15	G/A	0.81, 0.63, 0.45	227,950	0.0347 (0.0039)	1.19E-18	0.037%	39,130	0.0281 (0.0080)	4.40E-04	0.0368%	64,40
GPRC5B	rs12444979		16	C/T	0.85, 0.91, 1	239,715	0.0435 (0.0046)	2.91E-21	0.048%	37,955	0.0125 (0.0148)	0.399	2.56E-03%	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 12, continued

Locus	SNP	Proxy r^2 (CEU)	Chr	Alleles (effect/other)	EAF[d] (CEU, AA, CHB+JPT)	European ancestry[a]				African ancestry[b]				
						N	β (SE)	P	Explained variance (%)[e]	N	β (SE)	P	Explained variance (%)	N
SH2B1	rs7359397 rs4788102 (Asian proxy)	0.99	16	T/C A/G	0.37, 0.09, 0.16 0.36, 0.27, 0.16	204,309 --	0.0325 (0.0035) --	1.88E-20 --	0.049% --	18,554 --	0.0531 (0.0199) --	0.008 --	0.0462% --	-- 64,40
FTO	rs1558902 rs17817449 (Asian proxy)	0.93	16	A/T G/T	0.46, 0.11, 0.15 0.45, 0.40, 0.15	192,344 --	0.0851 (0.0037) --	4.8E-120 --	0.360% --	39,086 --	0.0658 (0.0128) --	2.69E-07 --	0.0848% --	-- 64,40
GP2	rs12597579		16	C/T	0.91, 0.90, 0.77	123,662	0.0286 (0.0110)	0.0095	0.013%	38,909	0.0369 (0.0132)	0.0052	0.0245%	64,40
MC4R	rs571312 rs6567160 (Asian proxy)	1.0	18	A/C C/T	0.24, 0.33, 0.13 0.23, 0.21, 0.13	203,600 --	0.0557 (0.0041) --	6.43E-42 --	0.113% --	38,848 --	0.0001 (0.0083) --	0.990 --	4.42E-07% --	-- 64,40
KCTD15	rs29941		19	G/A	0.67, 0.83, 0.27	192,872	0.0224 (0.0038)	3.01E-09	0.022%	37,592	0.0002 (0.0107)	0.985	1.13E-06%	64,40
QPCTL	rs2287019		19	C/T	0.87, 0.88, 0.82	194,564	0.0370 (0.0045)	1.88E-16	0.031%	38,140	0.0659 (0.0149)	1.01E-05	0.0917%	64,40
TMEM160	rs3810291		19	A/G	0.66, 0.20, 0.17	233,512	0.0249 (0.0035)	1.64E-12	0.028%	39,062	0.0189 (0.0102)	0.065	0.0114%	64,40
Sum of explained variance									**1.674%**				**0.993%**	

[a]European ancestry results from GIANT population published in Speliotes *et al*, 2010[30]; [b]African ancestry results from Stage 1; [c]Asian ancestry results from Wen *et all*, 2012 (Stage 1+2)[32] and Okada
according to European population, frequency for African ancestry (AA) from Stage 1, frequency for European ancestry (CEU) and Asian ancestry (CHB+JPT) from HapMap; [e]The total fraction of vari
formula $2f(1-f)*a^2$, where f is the frequency of the variant and a is the additive effect of the variant (see reference [33]); SNP, single nucleotide polymorphism; Chr, chromosome; EAF, effect allele fre
estimate) reported in inverse-normally transformed units; SE, standard error

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 13. Index SNP from referent population (listed first) and stronger marker ("alternative SNP", listed second) of the index SNP in African ancestry GWAS data as identified by lowest p-value within 500 kb on either side of the index SNP

									European ancestry[d]			African ancestry[e]			
Region	Chr	SNP[a]	r^2 EUR	r^2 AFR	Alleles (effect/other)	EAF (EA, AA)[b]	Position (Build 36)	Locus-Specific Alpha[c]	N	Beta (SE)	P-value	N	Beta (SE)	P-value	Explained variance (%)[f]
PTBP2	1	rs1555543	0.40	0.09	C/A	0.59, 0.43	96717385	0.002	243013	0.0198 (0.0032)	3.69E-10	39088	0.0031 (0.0078)	0.6903	0.01%
		rs12562499			C/G	0.37, 0.09	96720084		123791	0.0203 (0.0048)	2.55E-05	35940	0.0482 (0.0150)	1.34E-03	0.04%
TMEM18	2	rs2867125	0.99	1.00	C/T	0.83, 0.88	612827	0.003	197806	0.0681 (0.0046)	2.77E-49	38994	0.0450 (0.0119)	1.64E-04	0.04%
		rs1320330			G/T	0.82, 0.90	612225		123856	0.0592 (0.0062)	1.31E-21	39002	0.0612 (0.0129)	2.05E-06	0.07%
RBJ	2	rs713586	0.54	0.31	C/T	0.46, 0.84	25011512	0.002	230748	0.0309 (0.0032)	6.17E-22	38764	0.0378 (0.0111)	6.50E-04	0.04%
		rs7586879			T/C	0.34, 0.77	24970481		123852	0.0194 (0.0049)	8.50E-05	38948	0.0418 (0.0095)	1.05E-05	0.06%
NUDT3	6	rs206936	0.62	0.10	G/A	0.21, 0.53	34410847	0.004	249777	0.0208 (0.0038)	3.02E-08	39142	0.0060 (0.0078)	0.441	0.01%
		rs3798554			C/T	0.13, 0.14	34376662		123856	0.0222 (0.0071)	1.82E-03	39142	0.0357 (0.0112)	1.42E-03	0.03%
BDNF	11	rs10767664	0.91	0.40	A/T	0.79, 0.91	27682562	0.003	204158	0.0445 (0.0042)	4.69E-26	37781	0.0576 (0.0157)	2.37E-04	0.05%
		rs10501087			T/C	0.79, 0.93	27626684		123845	0.0414 (0.0059)	1.54E-12	38745	0.0808 (0.0157)	2.51E-07	0.09%
FTO	16	rs1558902	0.93	0.89	A/T	0.42, 0.11	52361075	0.003	192344	0.0851 (0.0037)	4.74E-120	39086	0.0658 (0.0128)	2.69E-07	0.09%
		rs17817964			T/C	0.41, 0.12	52385567		123860	0.0779 (0.0048)	1.03E-58	39080	0.0739 (0.0124)	2.28E-09	0.12%
MC4R	18	rs571312	1.0	0.03	A/C	0.23, 0.33	55990749	0.001	203600	0.0557 (0.0041)	6.43E-42	38848	0.0001 (0.0083)	0.9904	4.42E-07%
		rs6567160			C/T	0.24, 0.21	55980115		167066	0.0552 (0.0046)	1.24E-32	39103	0.0621 (0.0098)	2.41E-10	0.13%
Sum of explained variance in individuals of African ancestry using index SNP															**0.22%**
Sum of explained variance in individuals of African ancestry using stronger marker of the index SNP															**0.53%**

[a]All reference index SNPs from population of European ancestry [30]; [b]European ancestry (EA) allele frequencies from Speliotes *et* al., 2010 [30], African ancestry (AA) allele frequencies from Stage 1; [c]Significance criteria for each region that is based on the number of tag SNPs in the HapMap YRI population that capture ($r^2 \geq 0.8$) all common SNPs (MAF≥0.05) that were correlated with the index signal in the HapMap CEU; [d]European ancestry results from Speliotes *et al*, 2010[30]; [e]African ancestry results from Stage 1; [f]The total fraction of variance explained was calculated using the formula $2f(1-f)*a^2$, where f is the frequency of the variant and a is the additive effect of the variant (see reference [33]). Criteria for inclusion: (1) r^2 between index and alternative SNPs >0.4 in referent population, and (2) p-value in African ancestry GWAS at least one order of magnitude

Nature Genetics: doi:10.1038/ng.2608

smaller for alternative SNP vs. index SNP. Chr, chromosome; EA, European ancestry; AA, African ancestry; EAF, effect allele frequency. Variant p-values for heterogenetity of effect by study were all >0.1.

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 14. Associations between the three BMI SNPs and cis gene expression (cis-eQTLs) in liver, adipose and brain tissues.

Transcript	Gene symbol for transcript	Tissue	Effect (beta)	SE for the effect (beta)	P-value	Source
Association between rs974417 (KLHL32) and gene transcripts below (effect of T-allele vs C-allele)						
AK055697	AK055697	Liver	0.038	0.037	0.301	Reference [34]
AK055697	AK055697	Omental fat	-0.013	0.017	0.443	
AK055697	AK055697	Subcutaneous fat	-0.016	0.020	0.429	
AK091365	AK091365	Liver	-0.009	0.005	0.093	
AK091365	AK091365	Omental fat	0.001	0.002	0.681	
AK091365	AK091365	Subcutaneous fat	0.000	0.003	0.972	
AL080093	AL080093	Liver	-0.001	0.006	0.873	
AL080093	AL080093	Omental fat	0.003	0.002	0.099	
AL080093	AL080093	Subcutaneous fat	-0.004	0.002	0.108	
AL117528	AL117528	Liver	-0.024	0.011	0.029	
AL117528	AL117528	Omental fat	0.003	0.002	0.167	
AL117528	AL117528	Subcutaneous fat	-0.009	0.006	0.147	
ENST00000275053	C6orf167	Liver	0.009	0.017	0.579	
ENST00000275053	C6orf167	Omental fat	-0.002	0.009	0.816	
ENST00000275053	C6orf167	Subcutaneous fat	0.014	0.010	0.178	
NM_014165	C6orf66	Liver	-0.001	0.010	0.941	
NM_014165	C6orf66	Omental fat	-0.003	0.006	0.640	
NM_014165	C6orf66	Subcutaneous fat	-0.002	0.007	0.724	
GI_7661785	C6orf66 (NDUFAF4)	Brain (Cortex)	-0.597	3.779	0.874	
NM_020482	FHL5	Liver	0.002	0.004	0.645	
NM_020482	FHL5	Omental fat	-0.006	0.006	0.265	
NM_020482	FHL5	Subcutaneous fat	-0.001	0.010	0.938	
ILMN_1878007	FUT9	Brain (Frontal Cortex)	-0.017	0.037	0.641	Reference [35]
ILMN_1878007	FUT9	Brain (Cerebellum)	0.015	0.041	0.718	
NM_006581	FUT9	Liver	0.003	0.014	0.819	Reference [34]
NM_006581	FUT9	Omental fat	-0.002	0.003	0.544	
NM_006581	FUT9	Subcutaneous fat	-0.005	0.005	0.323	
NM_030784	GPR63	Liver	-0.005	0.008	0.524	
NM_030784	GPR63	Omental fat	0.003	0.004	0.391	
NM_030784	GPR63	Subcutaneous fat	0.002	0.004	0.535	
hCT1639952.3	hCT1639952.3	Liver	-0.002	0.008	0.778	
hCT1639952.3	hCT1639952.3	Omental fat	-0.004	0.008	0.652	
hCT1639952.3	hCT1639952.3	Subcutaneous fat	0.014	0.008	0.091	
GI_37059722	KIAA0776	Brain (Cortex)	-0.253	2.947	0.931	Reference [36]
NM_015323	KIAA0776	Liver	-0.004	0.012	0.762	Reference [34]
NM_015323	KIAA0776	Omental fat	-0.007	0.007	0.317	
NM_015323	KIAA0776	Subcutaneous fat	0.004	0.007	0.524	
NM_052904	KIAA1900	Liver	0.011	0.011	0.304	
NM_052904	KIAA1900	Omental fat	0.004	0.005	0.385	
NM_052904	KIAA1900	Subcutaneous fat	0.002	0.005	0.744	
GI_24308379	KIAA1900 (KLHL32)	Brain (Cortex)	-7.899	4.339	0.069	Reference [36]
RSE_00000578308	RSE_00000578308	Liver	0.003	0.008	0.654	Reference [34]
RSE_00000578308	RSE_00000578308	Omental fat	-0.004	0.005	0.358	
RSE_00000578308	RSE_00000578308	Subcutaneous fat	-0.006	0.003	0.071	
Association between rs7708584 (GALNT10) and gene transcripts below (effect of G-allele vs A-allele)						
AF007137	AF007137	Liver	-0.008	0.014	0.548	Reference [34]
AF007137	AF007137	Omental fat	-0.004	0.006	0.504	
AF007137	AF007137	Subcutaneous fat	0.003	0.009	0.737	

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 14, continued

Transcript	Gene symbol for transcript	Tissue	Effect (beta)	SE for the effect (beta)	P-value	Source
AF147432	AF147432	Liver	0.009	0.011	0.415	Reference [34]
AF147432	AF147432	Subcutaneous fat	-0.002	0.005	0.684	
BC017920	BC017920	Liver	0.008	0.008	0.315	
BC017920	BC017920	Omental fat	0.005	0.005	0.333	
BC017920	BC017920	Subcutaneous fat	0.000	0.007	0.971	
BC035378	BC035378	Liver	0.004	0.006	0.498	
BC035378	BC035378	Omental fat	0.001	0.002	0.767	
BC035378	BC035378	Subcutaneous fat	0.001	0.004	0.789	
NM_018691	C5orf3	Liver	0.004	0.006	0.518	
NM_018691	C5orf3	Omental fat	-0.002	0.004	0.518	
NM_018691	C5orf3	Subcutaneous fat	0.004	0.004	0.378	
ILMN_1728742	C5orf4	Brain (Frontal Cortex)	-0.003	0.041	0.939	Reference [35]
ILMN_1728742	C5orf4	Brain (Cerebellum)	-0.008	0.035	0.812	
NM_016348	C5orf4	Liver	-0.023	0.013	0.075	Reference [34]
NM_016348	C5orf4	Omental fat	0.007	0.005	0.136	
NM_016348	C5orf4	Subcutaneous fat	0.005	0.006	0.347	
ILMN_2174612	CNOT8	Brain (Frontal Cortex)	-0.003	0.018	0.846	Reference [35]
ILMN_2174612	CNOT8	Brain (Cerebellum)	-0.048	0.02	**0.015**	
NM_004779	CNOT8	Liver	0.000	0.007	0.946	Reference [34]
NM_004779	CNOT8	Omental fat	-0.006	0.003	0.050	
NM_004779	CNOT8	Subcutaneous fat	-0.003	0.005	0.541	
Contig36261_RC	Contig36261_RC	Liver	-0.008	0.012	0.502	
Contig36261_RC	Contig36261_RC	Omental fat	0.004	0.004	0.329	
Contig36261_RC	Contig36261_RC	Subcutaneous fat	-0.004	0.006	0.442	
Contig41110_RC	Contig41110_RC	Liver	0.004	0.004	0.362	
Contig41110_RC	Contig41110_RC	Omental fat	-0.002	0.002	0.308	
Contig41110_RC	Contig41110_RC	Subcutaneous fat	0.003	0.003	0.370	
Contig719_RC	Contig719_RC	Liver	0.003	0.007	0.683	
Contig719_RC	Contig719_RC	Omental fat	0.005	0.004	0.200	
Contig719_RC	Contig719_RC	Subcutaneous fat	-0.004	0.005	0.353	
GI_42476025	FLJ11526 (SAP30L)	Brain (Cortex)	1.1454	1.75488	0.514	Reference [36]
hCT1971097	GALNT10	Liver	0.008	0.004	0.032	Reference [34]
hCT1971097	GALNT10	Omental fat	0.001	0.002	0.715	
hCT1971097	GALNT10	Subcutaneous fat	0.005	0.003	0.067	
NM_017540	GALNT10	Liver	-0.013	0.007	0.087	
NM_017540	GALNT10	Omental fat	-0.016	0.004	0.00010	
NM_017540	GALNT10	Subcutaneous fat	-0.022	0.006	0.00017	
NM_024564	GALNT10	Liver	-0.011	0.005	0.048	
NM_024564	GALNT10	Omental fat	-0.013	0.004	0.00024	
NM_024564	GALNT10	Subcutaneous fat	-0.004	0.005	0.428	
NM_015465	GEMIN5	Liver	-0.004	0.005	0.421	
NM_015465	GEMIN5	Omental fat	0.000	0.003	0.914	
NM_015465	GEMIN5	Subcutaneous fat	0.005	0.004	0.156	
GI_6552333	GRIA1	Brain (Cortex)	0.632	6.434	0.921	Reference [36]
NM_000827	GRIA1	Liver	0.003	0.007	0.680	Reference [34]
NM_000827	GRIA1	Omental fat	-0.001	0.002	0.610	
NM_000827	GRIA1	Subcutaneous fat	-0.004	0.003	0.215	
NM_004821	HAND1	Liver	0.002	0.004	0.606	
NM_004821	HAND1	Omental fat	-0.003	0.005	0.490	
NM_004821	HAND1	Subcutaneous fat	-0.005	0.003	0.111	
hCT1833870.1	hCT1833870.1	Liver	-0.024	0.014	0.084	

Nature Genetics: doi:10.1038/ng.2608

hCT1833870.1	hCT1833870.1	Omental fat	-0.014	0.007	0.038	
hCT1833870.1	hCT1833870.1	Subcutaneous fat	-0.008	0.012	0.508	

Supplementary Table 14, continued

Transcript	Gene symbol for transcript	Tissue	Effect (beta)	SE for the effect (beta)	P-value	Source
hCT28036.2	hCT28036.2	Liver	-0.002	0.004	0.691	Reference [34]
hCT28036.2	hCT28036.2	Omental fat	-0.002	0.001	0.168	
hCT28036.2	hCT28036.2	Subcutaneous fat	0.000	0.001	0.834	
hCT28230.2	hCT28230.2	Liver	0.001	0.006	0.906	
hCT28230.2	hCT28230.2	Omental fat	0.000	0.003	0.947	
hCT28230.2	hCT28230.2	Subcutaneous fat	0.001	0.004	0.761	
HSS00298890	HSS00298890	Liver	0.000	0.003	0.872	
HSS00298890	HSS00298890	Omental fat	0.003	0.001	0.014	
HSS00298890	HSS00298890	Subcutaneous fat	-0.001	0.002	0.539	
ILMN_1681590	LARP1	Brain (Frontal Cortex)	0.007	0.015	0.640	Reference [35]
ILMN_1681590	LARP1	Brain (Cerebellum)	-0.008	0.013	0.519	
NM_015315	LARP1	Liver	-0.006	0.006	0.286	Reference [34]
NM_015315	LARP1	Omental fat	0.001	0.003	0.769	
NM_015315	LARP1	Subcutaneous fat	-0.003	0.005	0.561	
NM_033551	LARP1	Liver	-0.005	0.005	0.341	
NM_033551	LARP1	Omental fat	0.002	0.003	0.439	
NM_033551	LARP1	Subcutaneous fat	-0.001	0.004	0.817	
NM_005927	MFAP3	Liver	0.004	0.003	0.216	
NM_005927	MFAP3	Omental fat	-0.004	0.003	0.216	
NM_005927	MFAP3	Subcutaneous fat	0.009	0.004	**0.023**	
ILMN_1663220	MRPL22	Brain (Frontal Cortex)	-0.011	0.021	0.591	Reference [35]
ILMN_1663220	MRPL22	Brain (Cerebellum)	-0.016	0.024	0.509	
ILMN_1748819	MRPL22	Brain (Frontal Cortex)	-0.013	0.015	0.401	
ILMN_1748819	MRPL22	Brain (Cerebellum)	0.002	0.018	0.891	
NM_014180	MRPL22	Liver	-0.003	0.005	0.515	Reference [34]
NM_014180	MRPL22	Omental fat	-0.004	0.003	0.132	
NM_014180	MRPL22	Subcutaneous fat	0.002	0.003	0.604	
ILMN_1843932	SAP30L	Brain (Frontal Cortex)	-0.024	0.015	0.123	Reference [35]
ILMN_1843932	SAP30L	Brain (Cerebellum)	-0.009	0.014	0.520	
NM_024632	SAP30L	Liver	0.001	0.005	0.845	Reference [34]
NM_024632	SAP30L	Omental fat	0.001	0.003	0.804	
NM_024632	SAP30L	Subcutaneous fat	-0.001	0.003	0.871	
Association between rs10261878 (NFE2L3) and gene transcripts below (effect of C-allele vs A-allele)*						
AK057379	AK057379	Liver	-0.0198	0.0463437	0.670	Reference [34]
AK057379	AK057379	Omental fat	0.004	0.0186603	0.832	
AK057379	AK057379	Subcutaneous fat	-0.0447	0.0287446	0.121	
AK093321	AK093321	Liver	-0.0333	0.0450316	0.459	
AK093321	AK093321	Omental fat	0.0013	0.0189349	0.945	
AK093321	AK093321	Subcutaneous fat	-0.0206	0.0301675	0.495	
BC040837	BC040837	Liver	0.0433	0.1067298	0.685	
BC040837	BC040837	Omental fat	-0.0401	0.0450302	0.373	
BC040837	BC040837	Subcutaneous fat	-0.0107	0.0548954	0.845	
NM_138811	C7orf31	Liver	0.0263	0.0277061	0.342	
NM_138811	C7orf31	Omental fat	0.0032	0.0117421	0.784	
NM_138811	C7orf31	Subcutaneous fat	-0.0153	0.0136022	0.261	
Contig30434_RC	CBX3	Liver	0.0083	0.0232826	0.723	
Contig30434_RC	CBX3	Omental fat	-0.0096	0.0117899	0.415	
Contig30434_RC	CBX3	Subcutaneous fat	-0.0076	0.012463	0.544	
NM_016587	CBX3	Liver	0.0341	0.0235421	0.148	
NM_016587	CBX3	Omental fat	-0.0114	0.01051	0.277	
NM_016587	CBX3	Subcutaneous fat	0.0025	0.0124888	0.843	

Nature Genetics: doi:10.1038/ng.2608

Contig16476_RC	Contig16476_RC	Liver	-0.0122	0.0126577	0.334

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 14, continued

Transcript	Gene symbol for transcript	Tissue	Effect (beta)	SE for the effect (beta)	P-value	Source
Contig16476_RC	Contig16476_RC	Omental fat	0.0012	0.0047619	0.801	Reference [34]
Contig16476_RC	Contig16476_RC	Subcutaneous fat	0.0064	0.0077512	0.406	
Contig27623_RC	Contig27623_RC	Liver	0.0218	0.0207945	0.295	
Contig27623_RC	Contig27623_RC	Omental fat	0.022	0.0121867	0.071	
Contig27623_RC	Contig27623_RC	Subcutaneous fat	0.0215	0.0254668	0.398	
Contig34004_RC	Contig34004_RC	Liver	-0.0168	0.008334	0.044	
Contig34004_RC	Contig34004_RC	Omental fat	-0.0068	0.0134513	0.611	
Contig34004_RC	Contig34004_RC	Subcutaneous fat	0.0019	0.0163785	0.909	
Contig8885_RC	CYCS	Liver	0.0174	0.0197791	0.381	
Contig8885_RC	CYCS	Omental fat	0.007	0.0135229	0.604	
Contig8885_RC	CYCS	Subcutaneous fat	-0.0018	0.0155314	0.908	
NM_018947	CYCS	Liver	0.0088	0.0274656	0.749	
NM_018947	CYCS	Omental fat	0.0272	0.0166107	0.101	
NM_018947	CYCS	Subcutaneous fat	0.0108	0.0184228	0.557	
ENST00000258728	ENST00000258728	Liver	0.0249	0.0227957	0.274	
ENST00000258728	ENST00000258728	Omental fat	-0.0061	0.0125833	0.626	
ENST00000258728	ENST00000258728	Subcutaneous fat	0.0027	0.0141295	0.848	
AK026373	HNRPA2B1	Liver	0.0319	0.0393675	0.418	
AK026373	HNRPA2B1	Omental fat	-0.0103	0.0210981	0.626	
AK026373	HNRPA2B1	Subcutaneous fat	-0.0061	0.0215378	0.778	
NM_002137	HNRPA2B1	Liver	0.0099	0.0241771	0.682	
NM_002137	HNRPA2B1	Omental fat	0.006	0.0146666	0.683	
NM_002137	HNRPA2B1	Subcutaneous fat	0.0297	0.0174656	0.090	
NM_031243	HNRPA2B1	Liver	0.0059	0.0247716	0.813	
NM_031243	HNRPA2B1	Omental fat	-0.0047	0.0142978	0.740	
NM_031243	HNRPA2B1	Subcutaneous fat	0.0231	0.0165293	0.162	
HSS00053669	HSS00053669	Liver	0.002	0.0120743	0.866	
HSS00053669	HSS00053669	Omental fat	-0.0029	0.0058868	0.626	
HSS00053669	HSS00053669	Subcutaneous fat	-0.0114	0.0111662	0.306	
HSS00083037	HSS00083037	Liver	-0.0204	0.0280679	0.468	
HSS00083037	HSS00083037	Omental fat	-0.0002	0.0085329	0.984	
HSS00083037	HSS00083037	Subcutaneous fat	0.0016	0.0118815	0.891	
HSS00145953	HSS00145953	Liver	0.0329	0.0666154	0.622	
HSS00145953	HSS00145953	Omental fat	0.03	0.0293219	0.307	
HSS00145953	HSS00145953	Subcutaneous fat	0.0105	0.0277194	0.706	
HSS00173673	HSS00173673	Liver	-0.0044	0.0102891	0.672	
HSS00173673	HSS00173673	Omental fat	0.0084	0.0055557	0.129	
HSS00173673	HSS00173673	Subcutaneous fat	-0.007	0.0090065	0.434	
HSS00386131	HSS00386131	Liver	0.0023	0.0547936	0.967	
HSS00386131	HSS00386131	Omental fat	-0.0008	0.0112164	0.945	
HSS00386131	HSS00386131	Subcutaneous fat	0.0158	0.0153213	0.303	
Contig37281_RC	NFE2L3	Liver	0.0533	0.0322676	0.099	
Contig37281_RC	NFE2L3	Omental fat	-0.039	0.0159468	0.015	
Contig37281_RC	NFE2L3	Subcutaneous fat	-0.0399	0.0189934	0.036	
NM_004289	NFE2L3	Liver	0.0468	0.0225744	0.039	
NM_004289	NFE2L3	Omental fat	-0.0249	0.0136421	0.068	
NM_004289	NFE2L3	Subcutaneous fat	-0.0133	0.0151091	0.381	
NM_014769	NM_014769	Liver	-0.0133	0.0245859	0.590	
NM_014769	NM_014769	Omental fat	-0.0013	0.0087728	0.881	
NM_014769	NM_014769	Subcutaneous fat	-0.0167	0.0182394	0.359	
NM_022150	NPVF	Liver	-0.0038	0.0138616	0.787	

Nature Genetics: doi:10.1038/ng.2608

NM_022150	NPVF	Omental fat	-0.002	0.0057083	0.720

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 14, continued

Transcript	Gene symbol for transcript	Tissue	Effect (beta)	SE for the effect (beta)	P-value	Source
NM_022150	NPVF	Subcutaneous fat	-0.0108	0.0085922	0.211	Reference [34]
NM_015550	OSBPL3	Liver	0.0654	0.0424851	0.124	
NM_015550	OSBPL3	Omental fat	-0.0321	0.014787	0.030	
NM_015550	OSBPL3	Subcutaneous fat	0.0033	0.0174293	0.850	
Contig58029_RC	SKAP2	Liver	0.0372	0.0268163	0.166	
Contig58029_RC	SKAP2	Omental fat	-0.0005	0.0151661	0.975	
Contig58029_RC	SKAP2	Subcutaneous fat	-0.0159	0.0149345	0.288	
NM_003930	SKAP2	Liver	0.0131	0.0169339	0.438	
NM_003930	SKAP2	Omental fat	-0.0205	0.0168832	0.224	
NM_003930	SKAP2	Subcutaneous fat	0.013	0.016287	0.427	
NM_013322	SNX10	Liver	0.0529	0.0373254	0.157	
NM_013322	SNX10	Omental fat	0.0233	0.0267568	0.385	
NM_013322	SNX10	Subcutaneous fat	0.0013	0.0315403	0.967	
XM_166493	XM_166493	Liver	-0.0077	0.0167022	0.644	
XM_166493	XM_166493	Omental fat	0.008	0.0055506	0.151	
XM_166493	XM_166493	Subcutaneous fat	0.0135	0.010323	0.191	
XM_170128	XM_170128	Liver	0.0471	0.0408954	0.250	
XM_170128	XM_170128	Omental fat	-0.0074	0.0100698	0.463	
XM_170128	XM_170128	Subcutaneous fat	0.0073	0.0168512	0.667	

*No data were available for association between rs10261878 and expression in brain tissues.

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 15. Results from conditional analysis of the two novel BMI SNPs (*GALNT10* and *NFE2L3* loci) with cis gene expression (cis-eQTLs) in liver, adipose, and brain tissues (see also Supplementary Figure 4).

		Novel BMI-associated SNP and p-value for cis-association		SNP showing most significant cis-association with transcript and p-value		LD (YRI) between both SNPs*	Conditional analysis results**			
Transcripts	**Tissue**	**SNP**	**P-value**	**SNP**	**P-value**		**SNP**	**P-value**	**SNP**	**P-value**
GALNT10										
NM_017540	Liver	rs7708584	0.087	rs6886795	1.8E-04	†	rs7708584	0.12	rs6886795	2.5E-04
	Subcutaneous fat	rs7708584	1.7E-04	rs988024	2.5E-12	0.004	rs7708584	0.35	rs988024	2.9E-08
	Omental fat	rs7708584	1.0E-04	rs4958719	2.4E-09	0.005	rs7708584	0.19	rs4958719	1.4E-05
hCT1971097	Liver	rs7708584	0.032	rs10476991	3.7E-06	0.002	rs7708584	0.053	rs10476991	5.7E-06
	Subcutaneous fat	rs7708584	0.067	rs11750554	1.2E-05	0.023	rs7708584	0.20	rs11750554	3.4E-05
	Omental fat	rs7708584	0.72	rs283439	3.0E-05	0.014	rs7708584	0.69	rs283439	3.0E-05
NFE2L3										
Contig37281_RC	Liver	rs10261878	0.099	rs11973430	2.7E-09	0.09	rs10261878	0.011	rs11973430	5.981E-10
	Subcutaneous fat	rs10261878	0.036	rs11973455	5.9E-13	0.09	rs10261878	0.31	rs11973455	5.112E-12
	Omental fat	rs10261878	0.015	rs11973455	7.9E-15	0.09	rs10261878	0.10	rs11973455	5.225E-14
NM_004289	Liver	rs10261878	0.039	rs17502705	3.0E-15	0.04	rs10261878	0.071	rs17502705	5.061E-15
	Subcutaneous fat	rs10261878	0.38	rs11973455	2.8E-22	0.09	rs10261878	0.50	rs11973455	1.00E-21
	Omental fat	rs10261878	0.068	rs11973455	5.6E-19	0.09	rs10261878	0.39	rs11973455	2.93E-18

*LD estimated from the HapMap YRI release 23; **Statistical model includes both the novel BMI-associated SNP and the SNP showing the most significant cis-association; †Unable to be estimated because rs6886795 is monomorphic in the data

Nature Genetics: doi:10.1038/ng.2608

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 16. A summary of the biofeature analysis and putative functional SNPs at 5q33.

SNPs correlated with index (r^2>0.5) in 1KGP AFR	Tested in Stage 1	Correlated SNP tested in Stage 1 (r^2 in AFR)	SNP	Chr	Position (Hg19)	STAGE1 beta	STAGE1 SE	STAGE1 P-value	STAGE1 N	No. of overlapping Biofeatures[a]
rs17625484	Yes		rs17625484	5	153513403	0.0427	0.0103	3.63E-05	38214	1
rs67791094	No	rs7707654 (0.74)		5	153525840					1
hrs7715256	No	rs7708584 (0.59)	rs7715256	5	153537893	-0.041	0.0082	5.87E-07	38177	2
rs7719067	No	rs7708584 (0.77)	rs7719067	5	153538241	0.0445	0.0083	8.68E-08	38198	2
rs55776503	No	rs7707654 (0.86)		5	153539912					2
rs4569924	No	rs7708584 (0.77)	rs4569924	5	153540025	0.0445	0.0083	8.68E-08	38215	2
rs7708584 - Index	Yes		rs7708584	5	153543466	0.0498	0.0086	8.02E-09	38219	0
rs1366219	Yes		rs1366219	5	153545698	0.0498	0.0086	8.02E-09	38219	8
rs2099044	Yes		rs2099044	5	153548206	0.0509	0.0086	3.74E-09	38212	3
rs113042334	No	rs7707654 (0.86)		5	153548856					1
rs56968070	No	rs7707654 (0.98)		5	153550752					4
rs7707654	Yes		rs7707654	5	153555871	0.0502	0.0108	3.12E-06	38221	15
rs58253496	No	rs7707654 (0.98)		5	153560421					3
rs17115711	Yes		rs17115711	5	153572862	0.0443	0.0102	1.50E-05	39140	22
rs17115719	Yes		rs17115719	5	153574772	0.043	0.0102	2.64E-05	39094	15
rs55946756	No	rs17115719 (1.0)		5	153577099					39
rs6865687	Yes		rs6865687	5	153578190	0.0434	0.0102	2.22E-05	39134	17
rs58711554	No	rs6865687 (0.97)		5	153579425					8

[a]73 different chromatin features, which annotate regulatory elements in brain and adipose tissues were obtained from NIH Epigenomics Roadmap and known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types were collected from the ENCODE data (see Online Methods). Each correlated SNP is color coded in Supplementary Figure 5a to reflect the number of biofeatures with which it overlaps.

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 17. A summary of the biofeature analysis and putative functional SNPs at 6q16.

SNPs correlated with index (r^2>0.5) in 1KGP AFR	Tested in Stage 1	Correlated SNP tested in Stage 1 (r^2 in AFR)	SNP	Chr	Position (Hg19)	STAGE1 beta	STAGE1 SE	STAGE1 P-value	STAGE1 N	No. of overlapping Biofeatures[a]
rs13203153	Yes		rs13203153	6	97374850	-0.0341	0.0088	0.000116	38221	16
rs2206568	No	rs7761614 (0.88)		6	97382338					6
rs13198696	No	rs7761614 (0.96)		6	97386246					19
rs7761614	Yes		rs7761614	6	97388179	0.0338	0.0084	5.97E-05	39140	9
rs5025221	No	rs6937736 (0.99)		6	97388516					6
rs35104491	No	rs7769945 (0.98)		6	97390463					8
rs2223614	No	rs6917254 (0.91)		6	97391272					13
rs6917254	Yes		rs6917254	6	97391853	-0.0344	0.0086	6.77E-05	39103	10
rs6937736	Yes		rs6937736	6	97399772	-0.0346	0.0086	6.14E-05	39134	3
rs6937950	Yes		rs6937950	6	97399927	-0.0346	0.0086	6.14E-05	39136	3
rs7769945	Yes		rs7769945	6	97400839	0.0347	0.0086	5.84E-05	39127	2
rs6568674	Yes		rs6568674	6	97402165	-0.0345	0.0099	0.0005008	39072	1
rs6910273	No	rs6568674 (1.0)		6	97403783					1
rs34663994	No	rs6937736 (0.97)		6	97404417					2
rs13211612	Yes		rs13211612	6	97406847	-0.0364	0.0087	3.12E-05	39140	5
rs17057164	Yes		rs17057164	6	97410536	0.0393	0.0082	1.68E-06	39125	3
rs71562315	No	rs13192767 (0.99)		6	97410922					1
rs13192767	Yes		rs13192767	6	97411114	0.0366	0.0087	2.82E-05	38896	2
rs7743034	Yes		rs7743034	6	97411894	-0.024	0.0079	0.002344	39140	4
rs11153285	Yes		rs11153285	6	97413151	-0.0386	0.008	1.38E-06	39100	8
rs11153286	No	rs11153285 (1.0)		6	97413636					8
rs10485381	Yes		rs10485381	6	97413827	0.038	0.008	2.00E-06	39110	9
rs13205380	No	rs13192767 (0.96)		6	97414063					9
rs11753834	No	rs7759572 (1.0)		6	97414756					12

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 17, continued

SNPs correlated with index (r^2>0.5) in 1KGP AFR	Tested in Stage 1	Correlated SNP tested in Stage 1 (r^2 in AFR)	SNP	Chr	Position (Hg19)	STAGE1 beta	STAGE1 SE	STAGE1 P-value	STAGE1 N	No. of overlapping Biofeatures[a]
rs11754336	No	rs7759572 (1.0)		6	97414775					12
rs7758454	No	rs13211612 (0.98)		6	97415116					12
rs7759572	Yes		rs7759572	6	97415135	0.0387	0.008	1.29E-06	39093	11
rs7758623	No	rs7759572 (1.0)		6	97415238					9
rs34398537	No	rs13192767 (1.0)		6	97416270					9
rs974417 - Index	Yes		rs974417	6	97419598	-0.0395	0.0082	1.49E-06	39120	0
rs7762215	No	rs2179537 (1.0)		6	97421139					7
rs2143389	Yes		rs2143389	6	97422681	-0.0391	0.0082	1.90E-06	39078	10
rs9791288	No	rs2179537 (1.0)		6	97422915					11
rs2179537	Yes		rs2179537	6	97422930	0.0241	0.0079	0.002248	39123	11
rs11756272	Yes		rs11756272	6	97423683	0.0209	0.0079	0.008056	38497	13
rs2387762	No	rs974417 (1.0)		6	97424463					8
rs7770535	No	rs974417 (1.0)		6	97425095					1
rs13196401	No	rs13195937 (1.0)		6	97427757					1
rs13195937	Yes		rs13195937	6	97428128	-0.0265	0.0079	0.0007803	39117	2

[a]73 different chromatin features, which annotate regulatory elements in brain and adipose tissues were obtained from NIH Epigenomics Roadmap and known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types were collected from the ENCODE data (see Online Methods). Each correlated SNP is color coded in Supplementary Figure 5b to reflect the number of biofeatures with which it overlaps.

Nature Genetics: doi:10.1038/ng.2608

Supplementary Table 18. A summary of the biofeature analysis and putative functional SNPs at 7p15.

SNPs correlated with index (r^2>0.5) in 1KGP AFR	Tested in Stage 1	Correlated SNP tested in Stage 1 (r^2 in AFR)	SNP	Chr	Position (Hg19)	STAGE1 beta	STAGE1 SE	STAGE1 P-value	STAGE1 N	No. of overlapping Biofeatures[a]
rs10261878 - Index	Yes		rs10261878	7	25950545	-0.0301	0.008	0.0001664	39101	0
rs4722543	Yes		rs4722543	7	25953720	0.0276	0.0081	0.0006567	39078	2
rs9655264	Yes		rs9655264	7	25955324	-0.0281	0.0081	0.0005228	39141	1
rs10223942	No	rs4722544 (0.95)		7	25955552					3
rs4722544	Yes		rs4722544	7	25956167	0.0664	0.0382	0.08185	2101	7
rs998357	Yes		rs998357	7	25979526	-0.0312	0.0082	0.0001439	39084	2
rs7803374	Yes		rs7803374	7	25979695	-0.0304	0.0083	0.0002557	38180	2
rs2893222	No	rs7803374 (1.0)		7	25980231					2
rs6959363	No	rs6977848 (0.83)		7	25984565					2
rs6947509	No	rs10261878 (0.77)		7	25987772					8
rs6953596	No	rs10261878 (0.82)		7	25989298					20
rs6977848	Yes		rs6977848	7	25989520	0.0279	0.0084	0.000922	38059	37
rs62446277	No	rs6977848 (0.86)		7	25989960					35

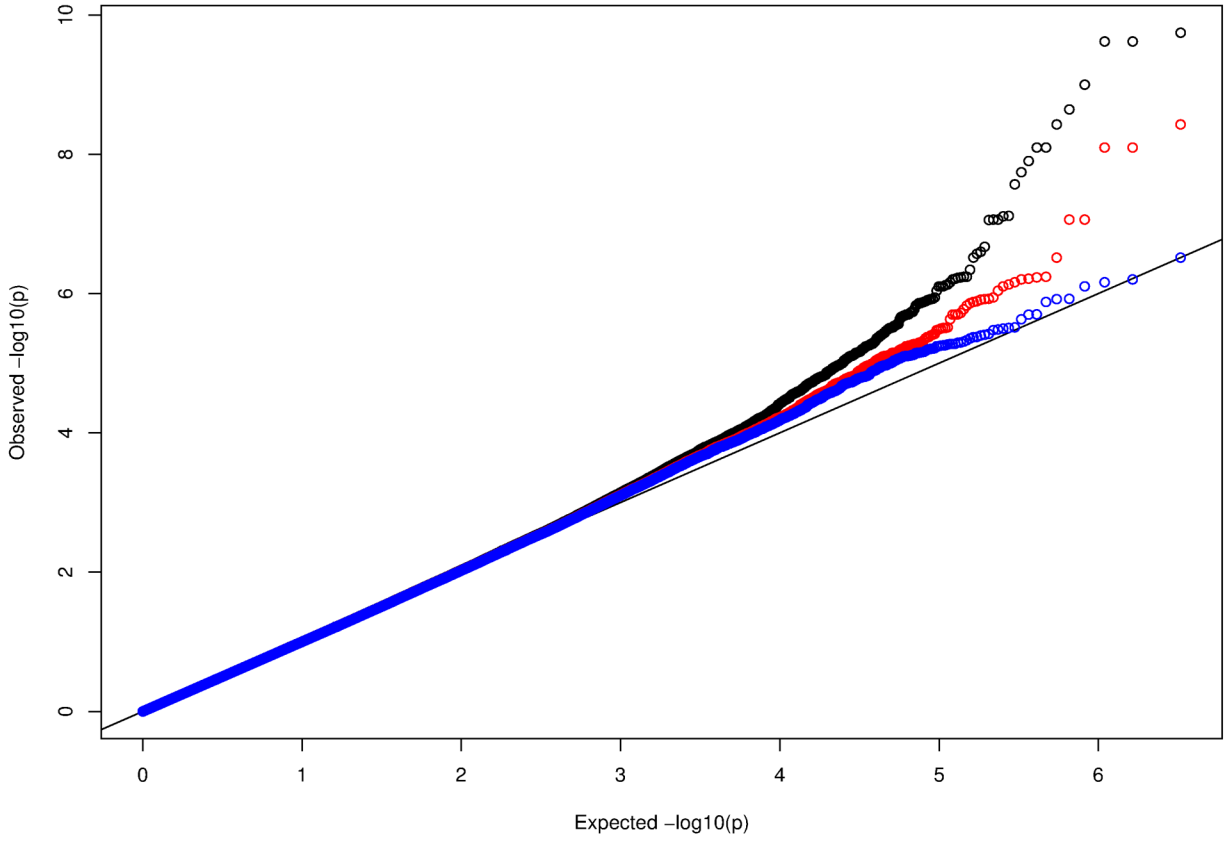
[a]73 different chromatin features, which annotate regulatory elements in brain and adipose tissues were obtained from NIH Epigenomics Roadmap and known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types were collected from the ENCODE data (see Online Methods). Each correlated SNP is color coded in Supplementary Figure 5c to reflect the number of biofeatures with which it overlaps.

Supplementary Table 19. Comprehensive results of the biofeature analysis for the three novel loci.

Comprehensive results for the biofeature analysis are available online as an Excel document titled "Supplementary Table 19-Comprehensive Biofeature Results"

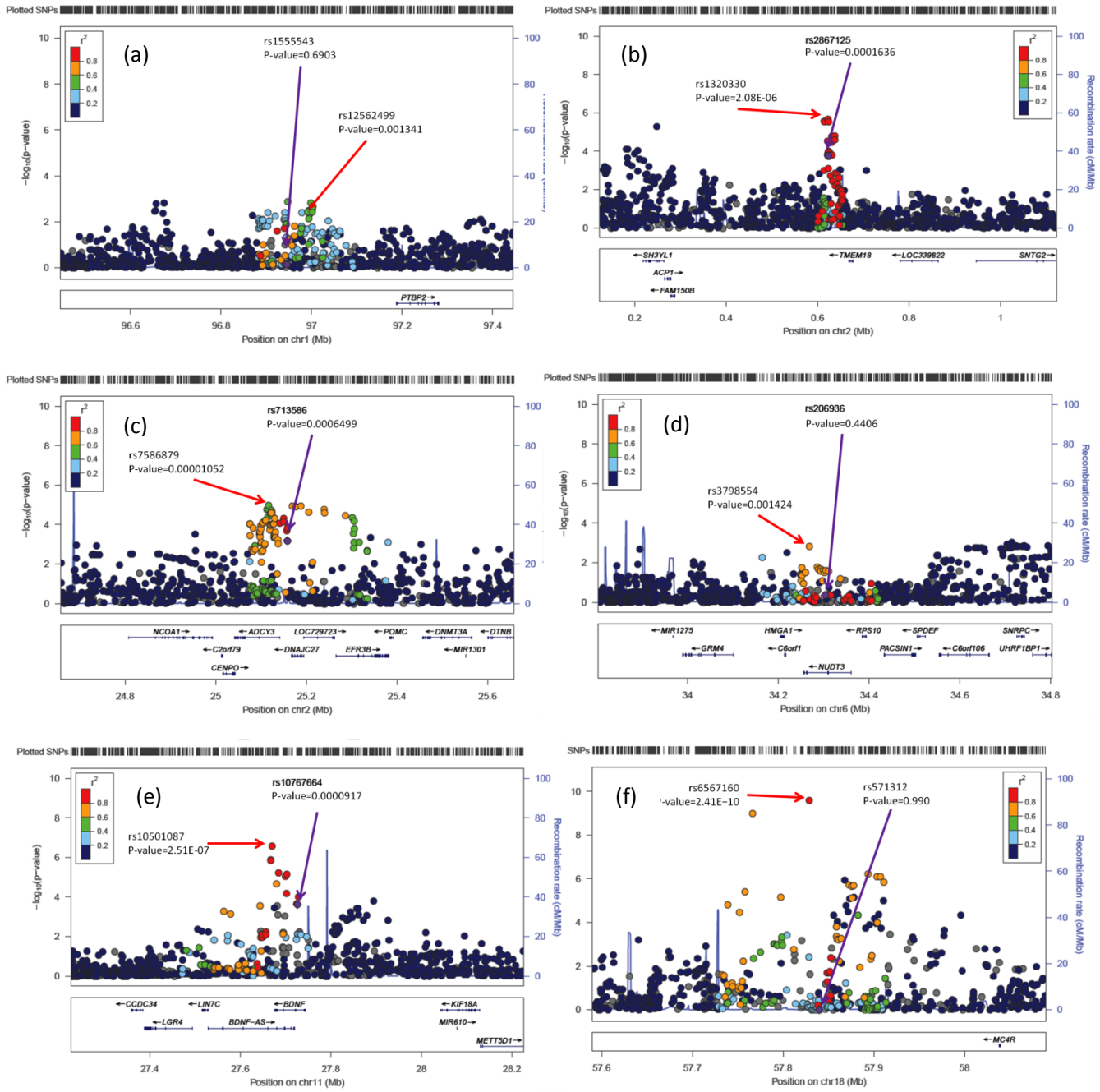
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 1. Quantile-quantile plot of SNPs for the Stage 1 meta-analysis of BMI (black line), after removing SNPs within 1 MB of the 36 BMI SNPs reported in individuals of European and Asian ancestry (red line), and additionally removing the two novel and one suggestive BMI SNPs discovered herein (blue line)



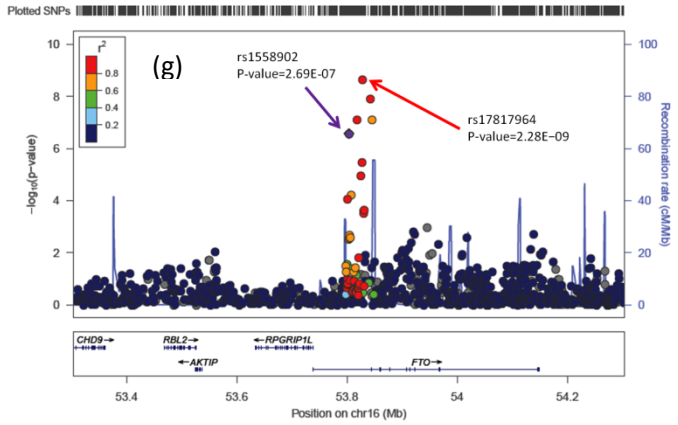
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 2. Regional plots of seven regions harboring better marker of the index SNP in population of African ancestry* (a) *PTBP2*, (b) *TMEM18*, (c) *RBJ*, (d) *NUDT3*, (e) *BDNF*, (f) *FTO*, (g) *MC4R*.



Nature Genetics: doi:10.1038/ng.2608

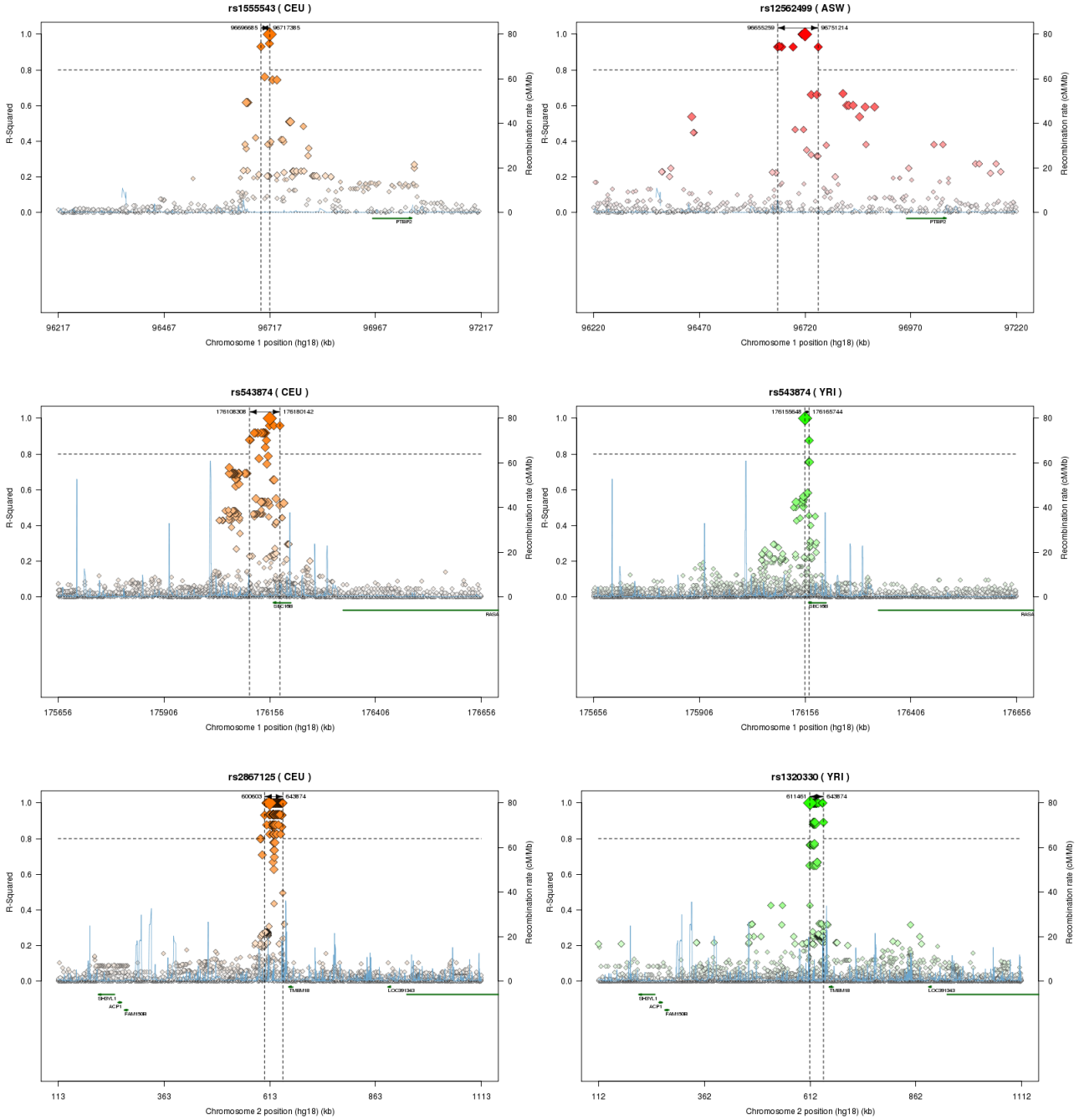
Supplementary Figure 2, continued



*SNPs are plotted by their position 500 kb on either side of the index SNP on the chromosome against their association (-$\log_{10}$ P) with BMI using African ancestry GWAS Stage 1 data. SNPs surrounding the top SNPs are colored to indicate the local LD structure using pairwise r^2 data from the November 2011 EUR panel of the 1000 Genomes.

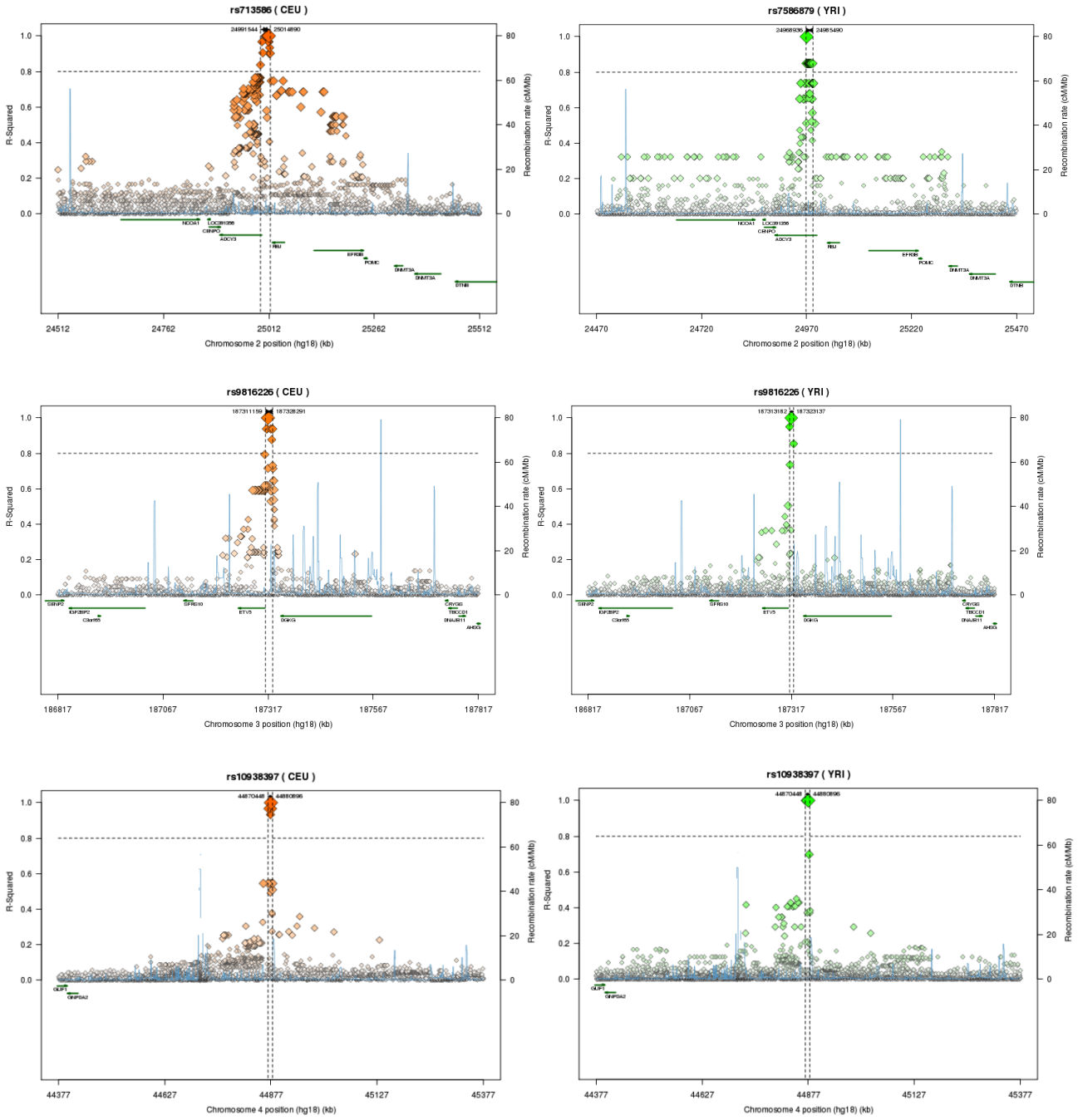
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3. Regional plots* showing the number of SNPs correlated with the index SNP in populations of European ancestry (left column) and African ancestry (right column). SNPs shown are those that were nominally statistically significant ($p<0.05$) in the African ancestry sample and with effects that were directionally consistent with those in the European population. In regions where a stronger marker was identified in the African ancestry sample (Supplemental Table 13) the better marker is shown in the right column.



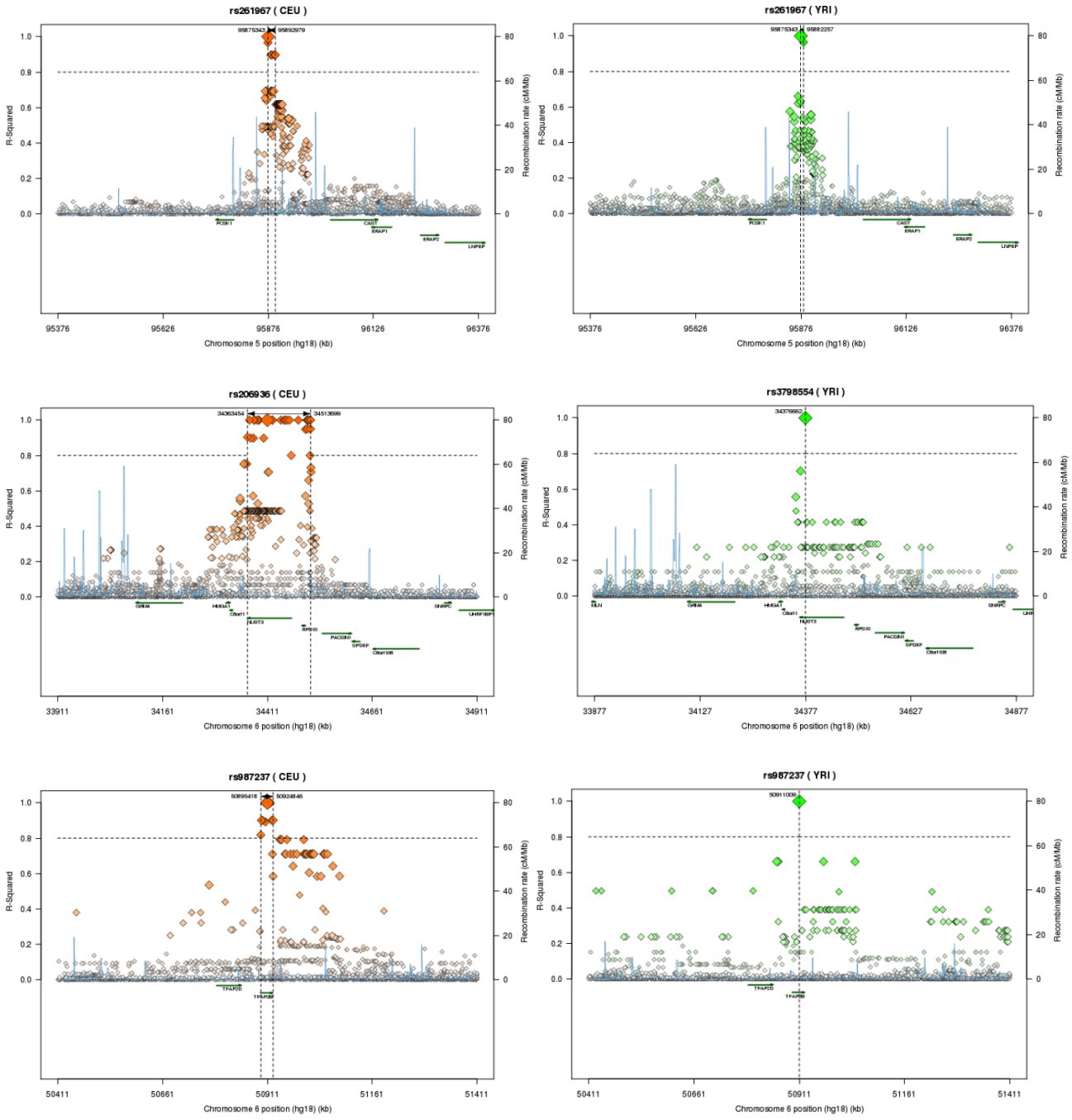
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3, continued



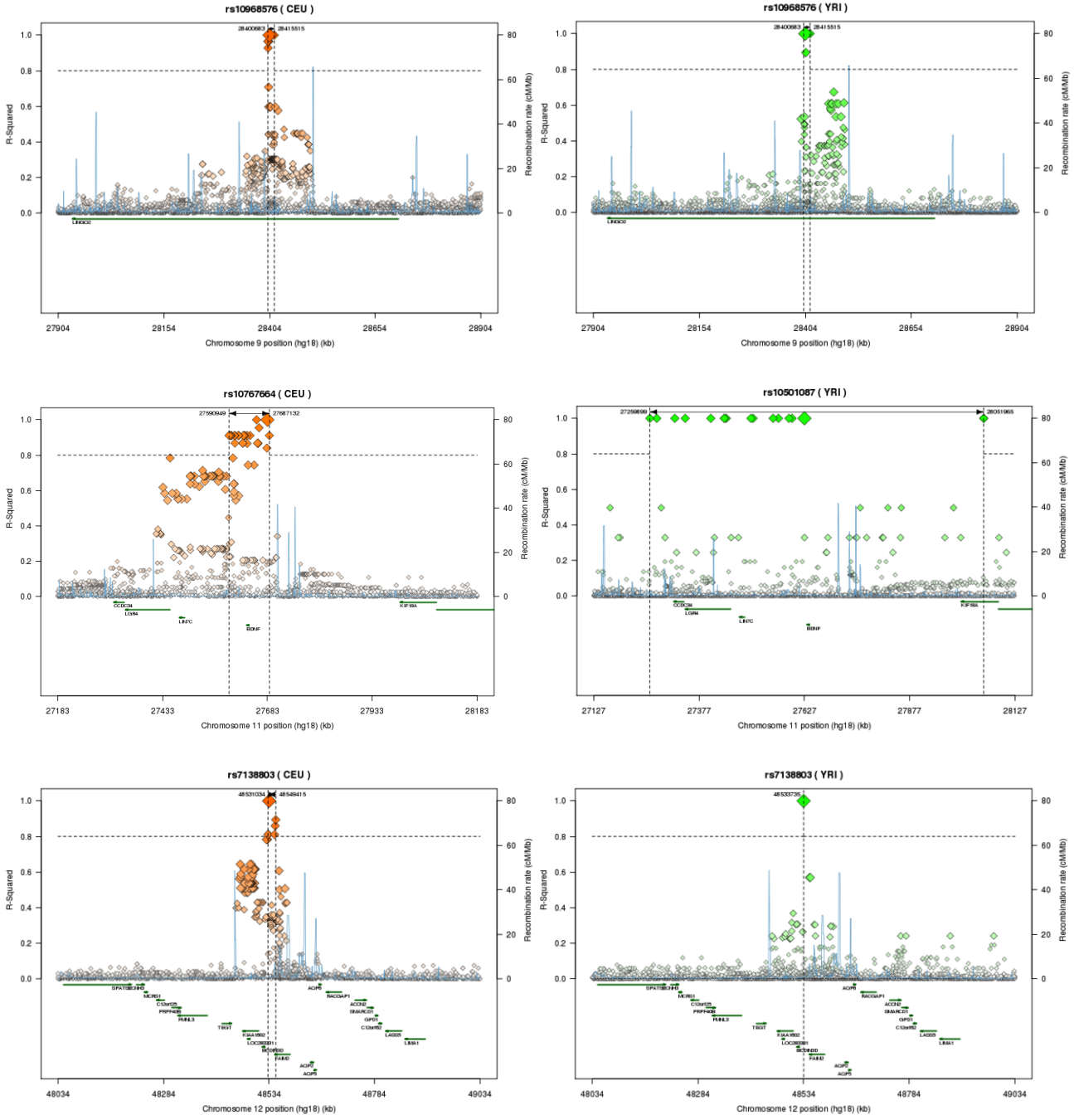
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3, continued



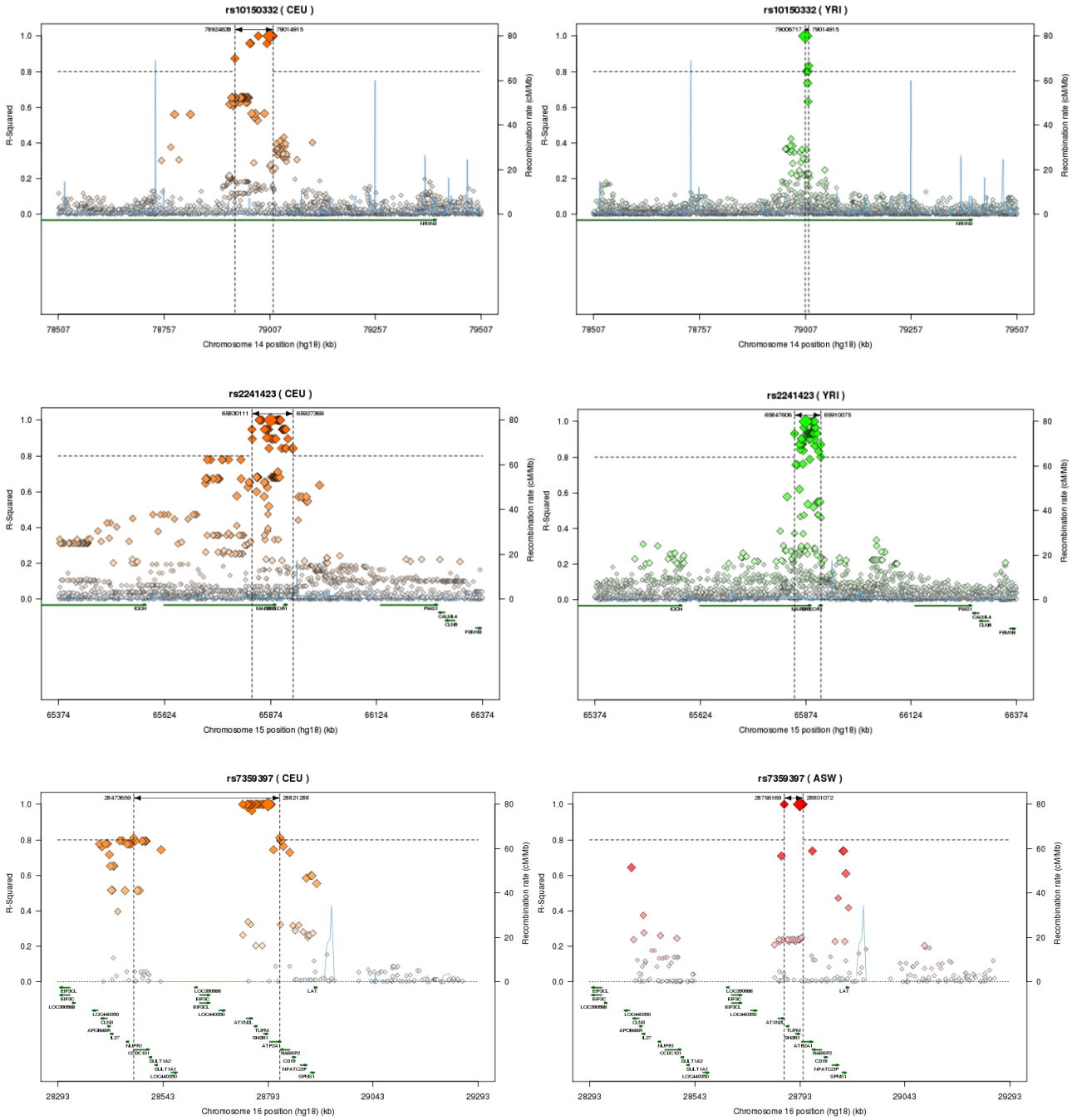
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3, continued



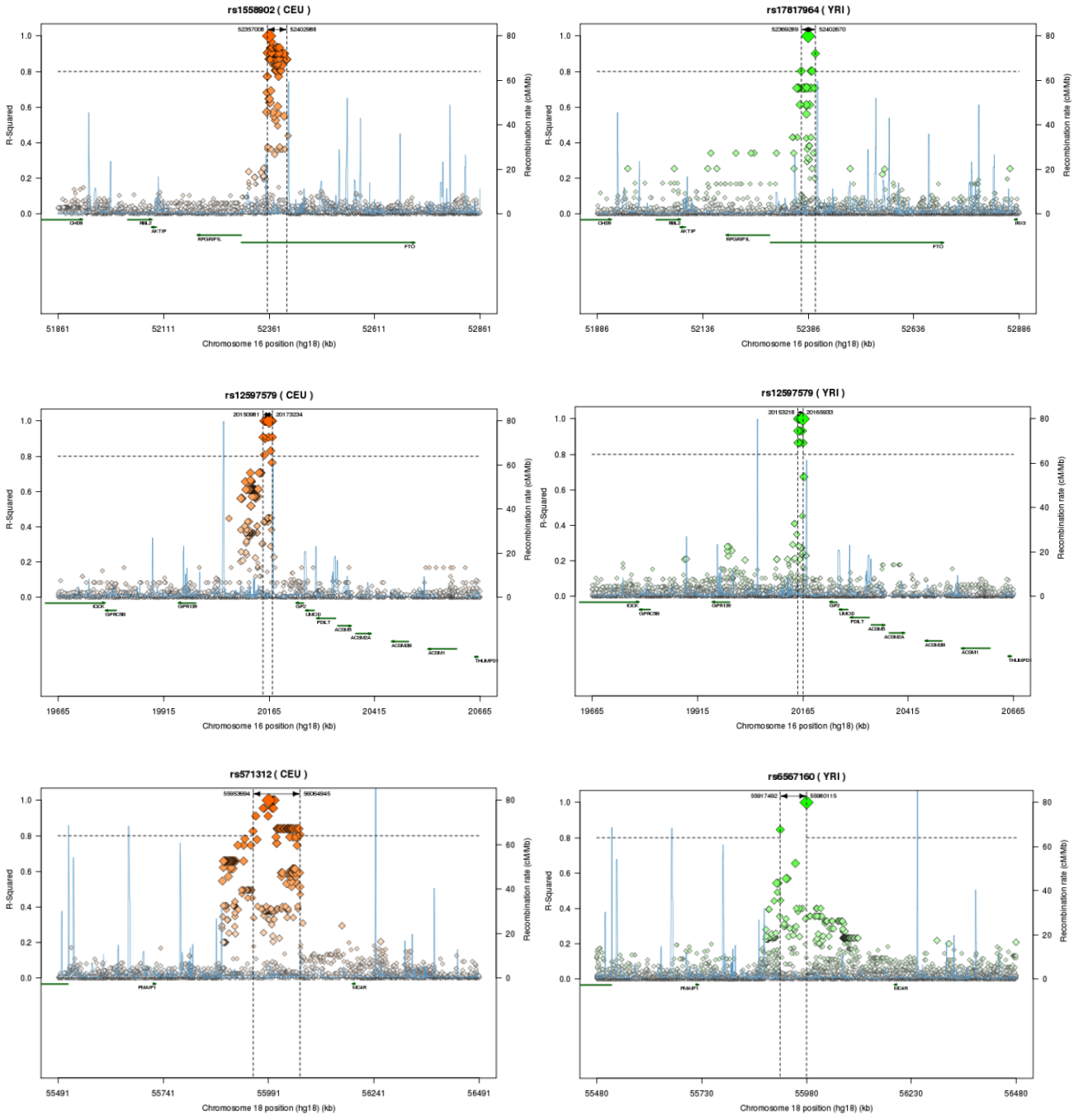
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3, continued



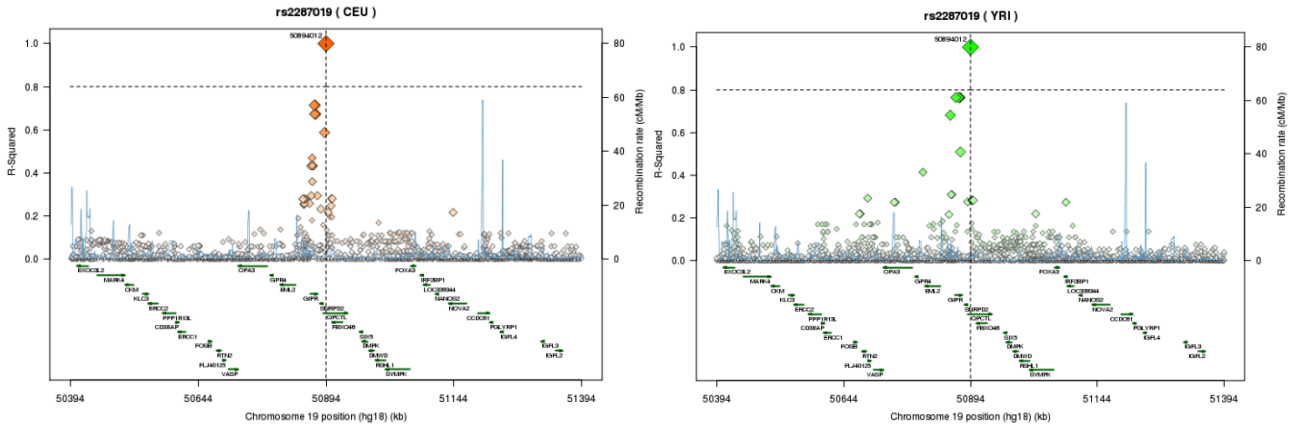
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 3, continued



Nature Genetics: doi:10.1038/ng.2608

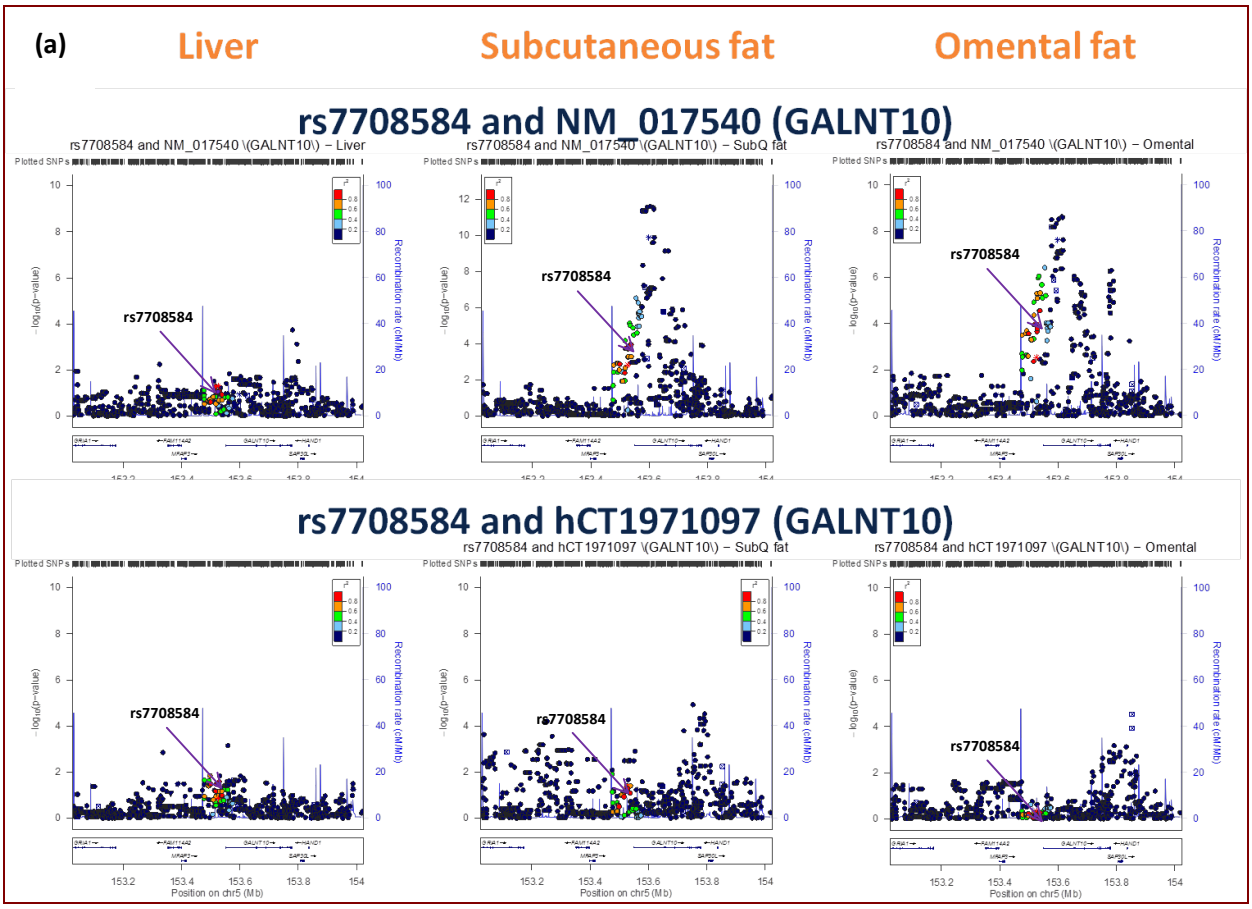
Supplementary Figure 3, continued



*Plots were generated in SNAP (http://www.broadinstitute.org/mpg/snap/index.php). Correlation (r^2) is based on 1000 Genomes except for SNPs rs7359397, rs1555543, and rs12562499 (HapMap Phase 3).

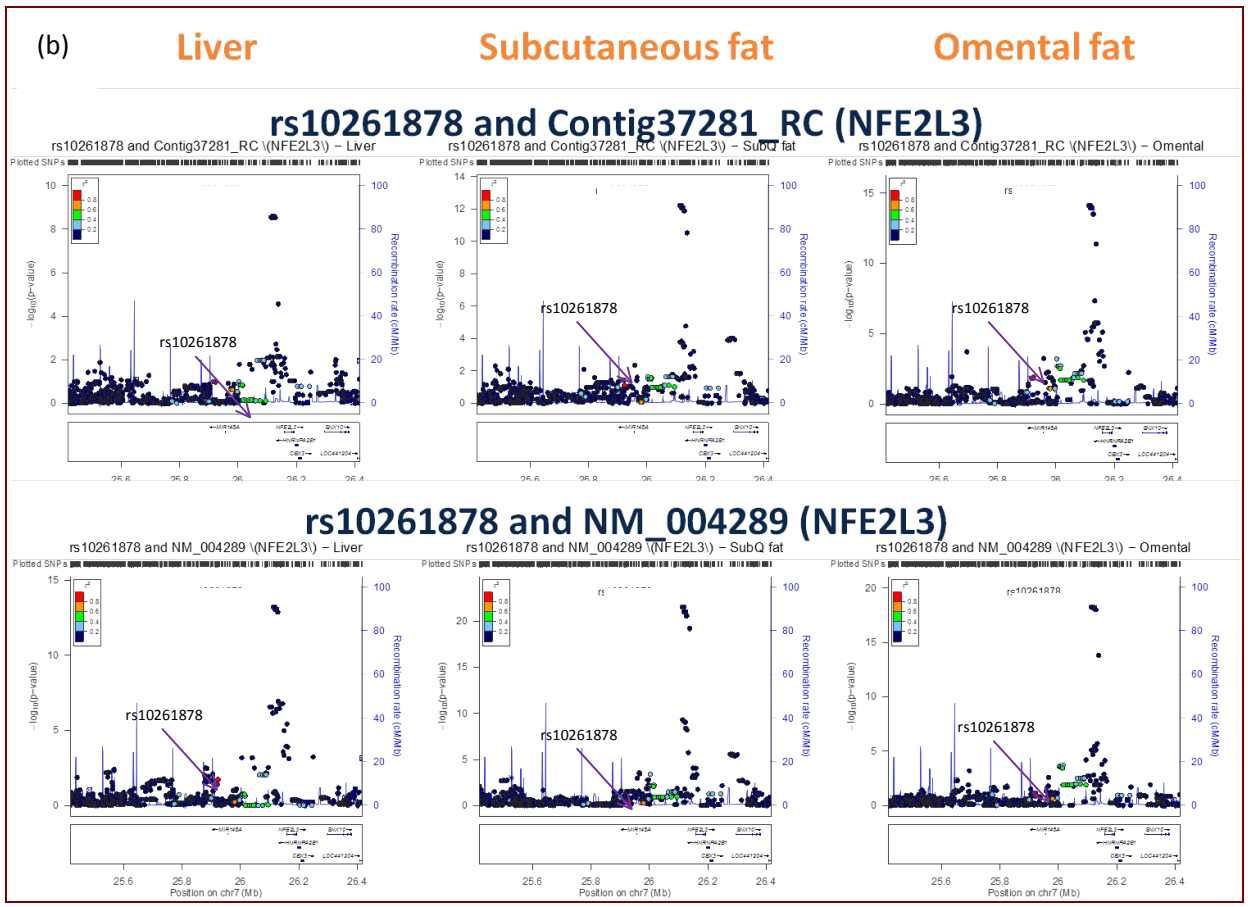
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 4. Regional plots of the cis-associations between SNPs in the GALNT10 (panel a) and NFE2L3 (panel b) loci (lead SNP +/- 500kb) and their respective cis-transcripts (NM_017540 and hCT1971097 for GALNT10; Contig37281_RC and NM_004289 for NFE2L3) (See also Supplementary Table 14 and Online Methods)*



Nature Genetics: doi:10.1038/ng.2608

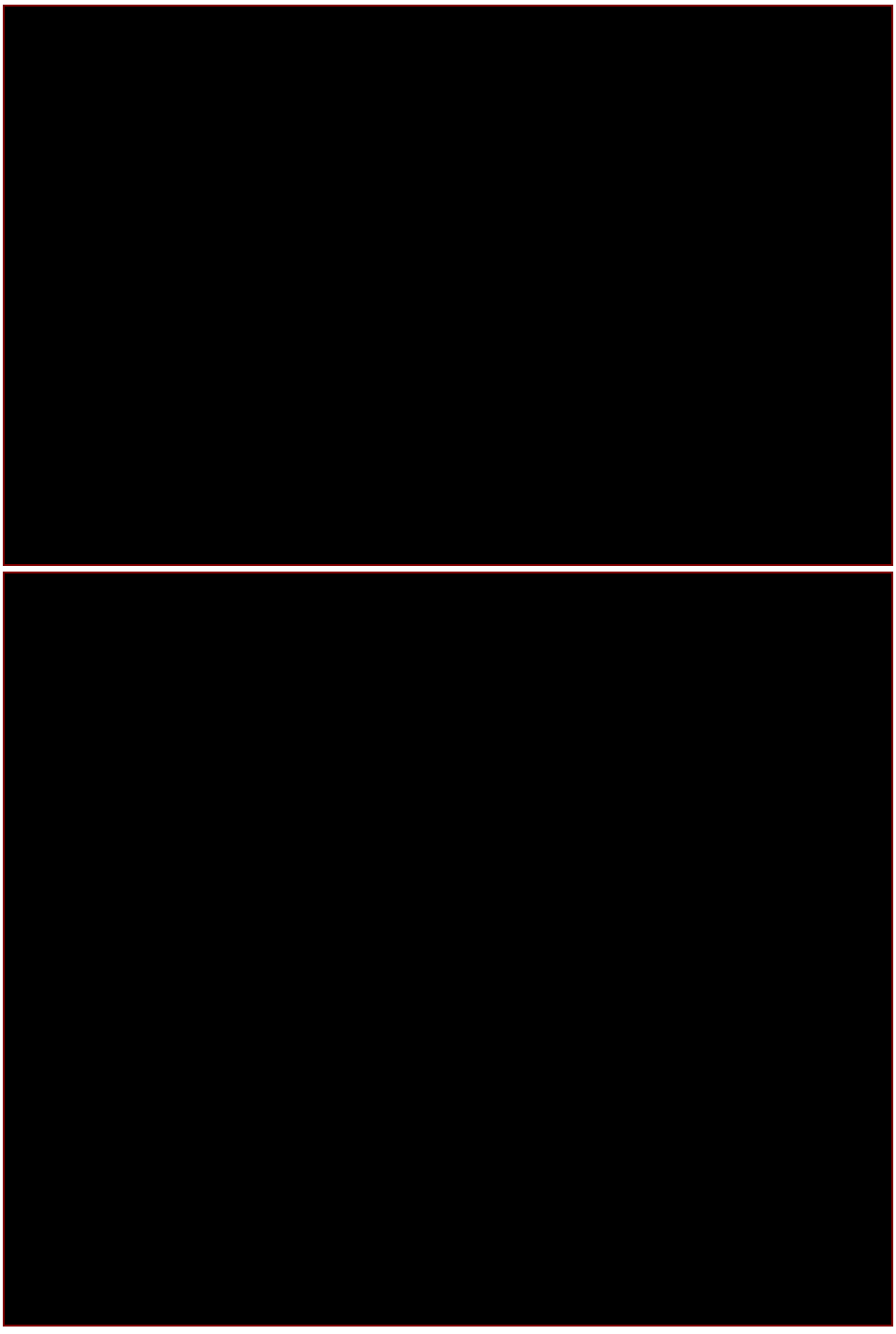
Supplementary Figure 4, continued



*SNPs are plotted by their position, 500 kb on either side of the index SNP on the chromosome against their association ($-\log_{10}$ P) with expression levels. SNPs surrounding the top SNPs are colored to indicate the local LD structure using pairwise r^2 data from the HapMap hg18 CEU version.

Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 5. UCSC Genome Browser view of the three novel loci with FunciSNP results. (a) Genome browser view for rs7708584 at 5q33.2, (b) rs974417 at 6q16.1, (c) rs10261878 at 7p15.2*



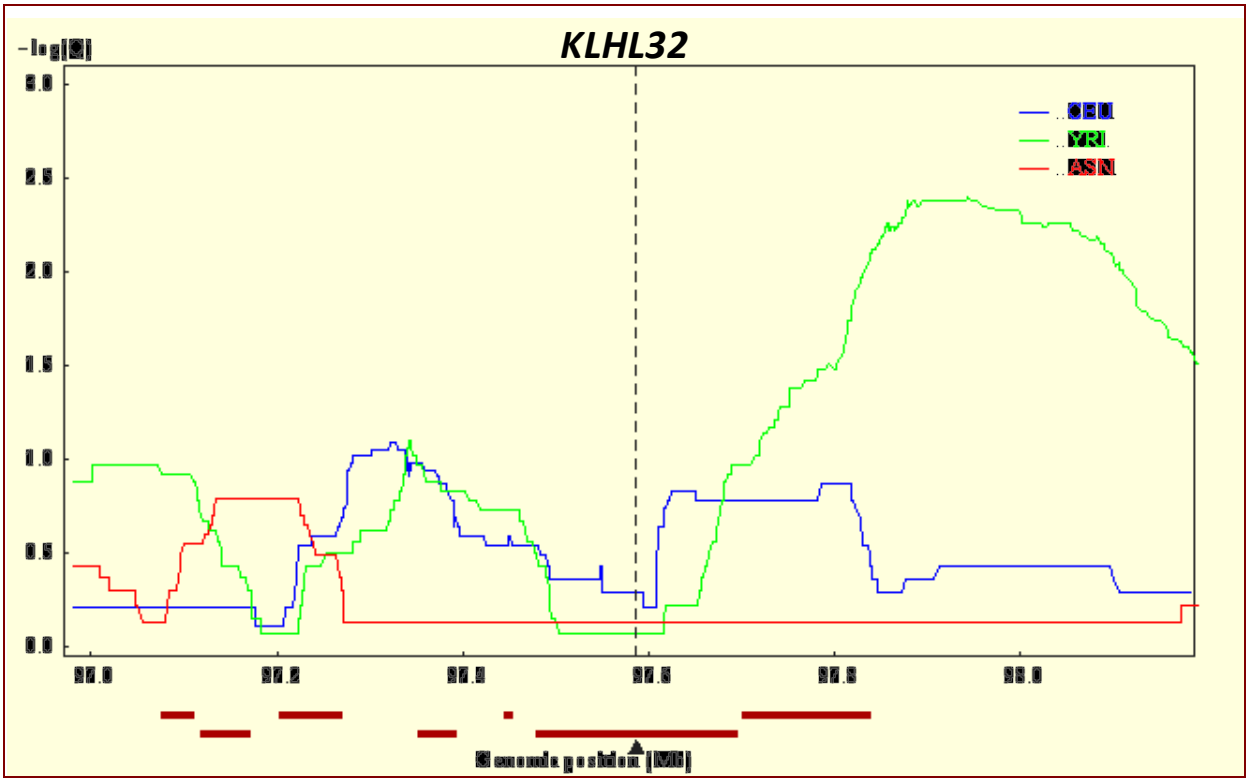
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 5, continued

*Genome browser tracks are ordered in the following manner: Refseq genes, UCSC Genes, Human mRNAs from GenBank, Human spliced ESTs, regulatory elements histone marks from 7 cell lines (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, and NHLF) from ENCODE, and FunciSNP results. In the FunciSNP result track, each index SNP is shown in black and each correlated SNP is color coded to reflect the number of biofeatures with which it overlaps. The color key is shown at the bottom. The color ranges from blue (low number of biofeature overlap) to red (high number of biofeature overlap). The r^2 value of each correlated SNP is also shown to the right of the rsID. We used 73 different chromatin features, which annotate regulatory elements (i.e. enhancers, promoters) from brain and adipose tissues from the NIH Epigenomics Roadmap[37] as well as known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types, which were collected from the ENCODE data[38] ENREF 38. More information regarding each chromatin mark and overlapping SNPs are provided in Supplementary Tables 16-19).

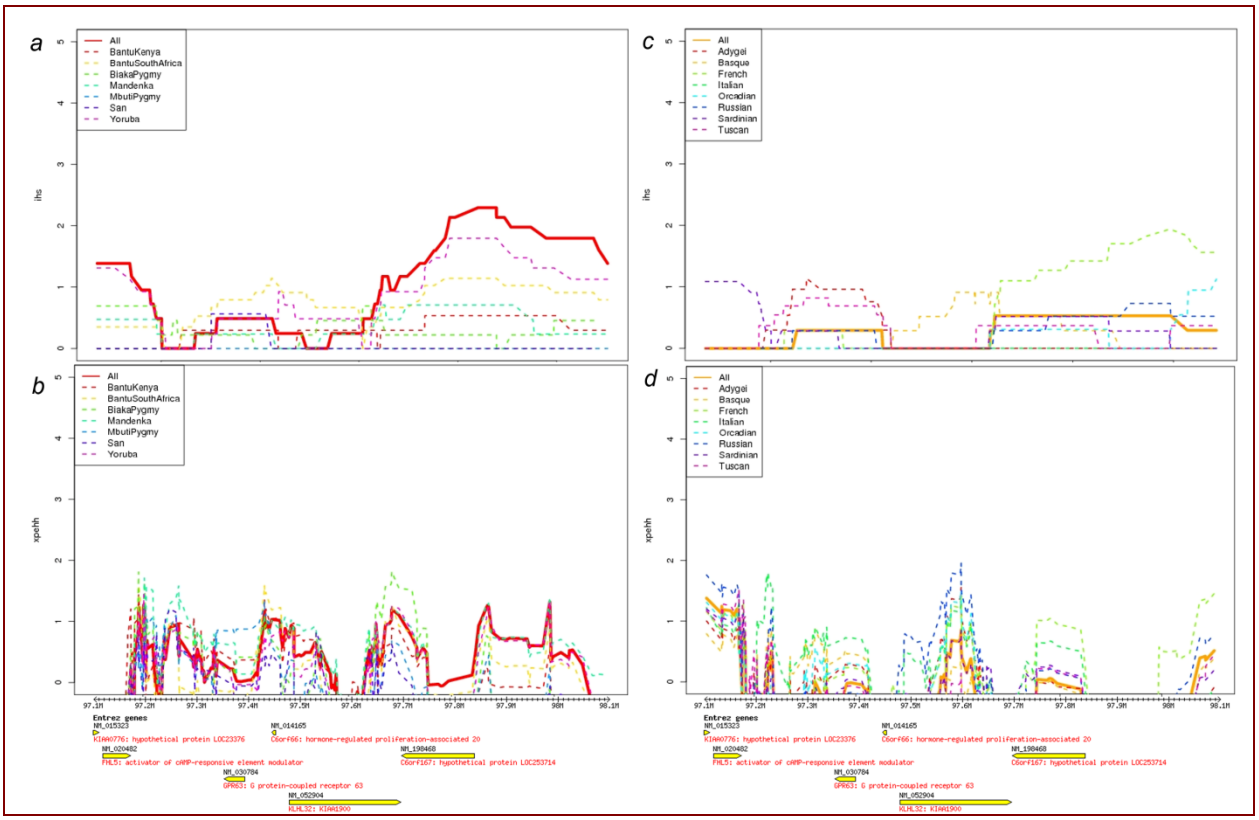
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 6. Evidence for recent positive selection from the Haplotter tool downstream of the *KLHL32* gene in HapMap YRI data where the dotted line indicates the location of *KLHL32* from 97.5Mb to 97.7Mb on chromosome 6. The empirical p-value for this selection peak is 4.0×10^{-2}.



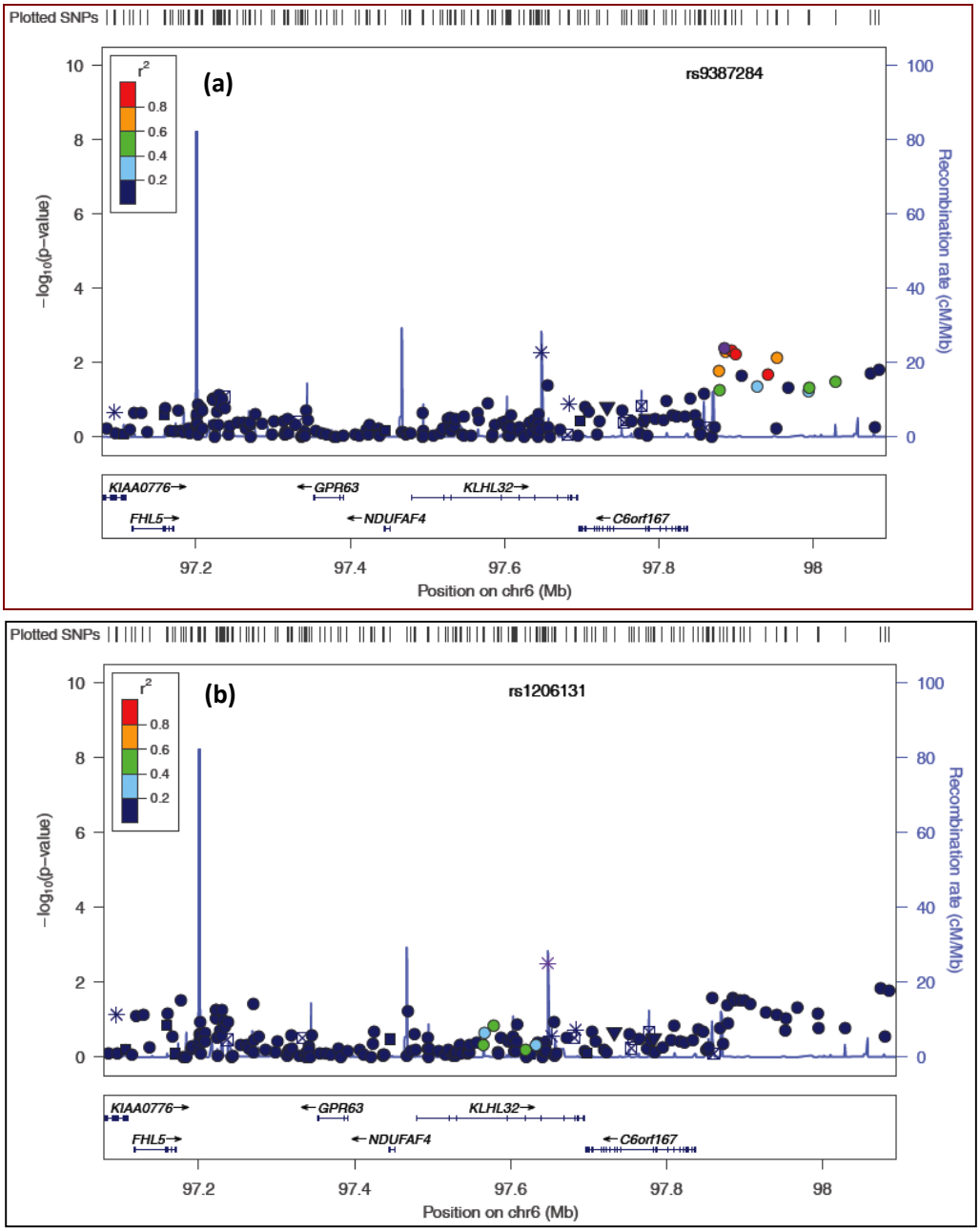
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 7. Evidence for recent selection in African HGDP participants at the *KLHL32* locus. (a) iHS statistics in African populations; (b) XP-EHH statistics in African populations; (c) iHS statistics in European populations; (d) XP-EHH statistics in European populations



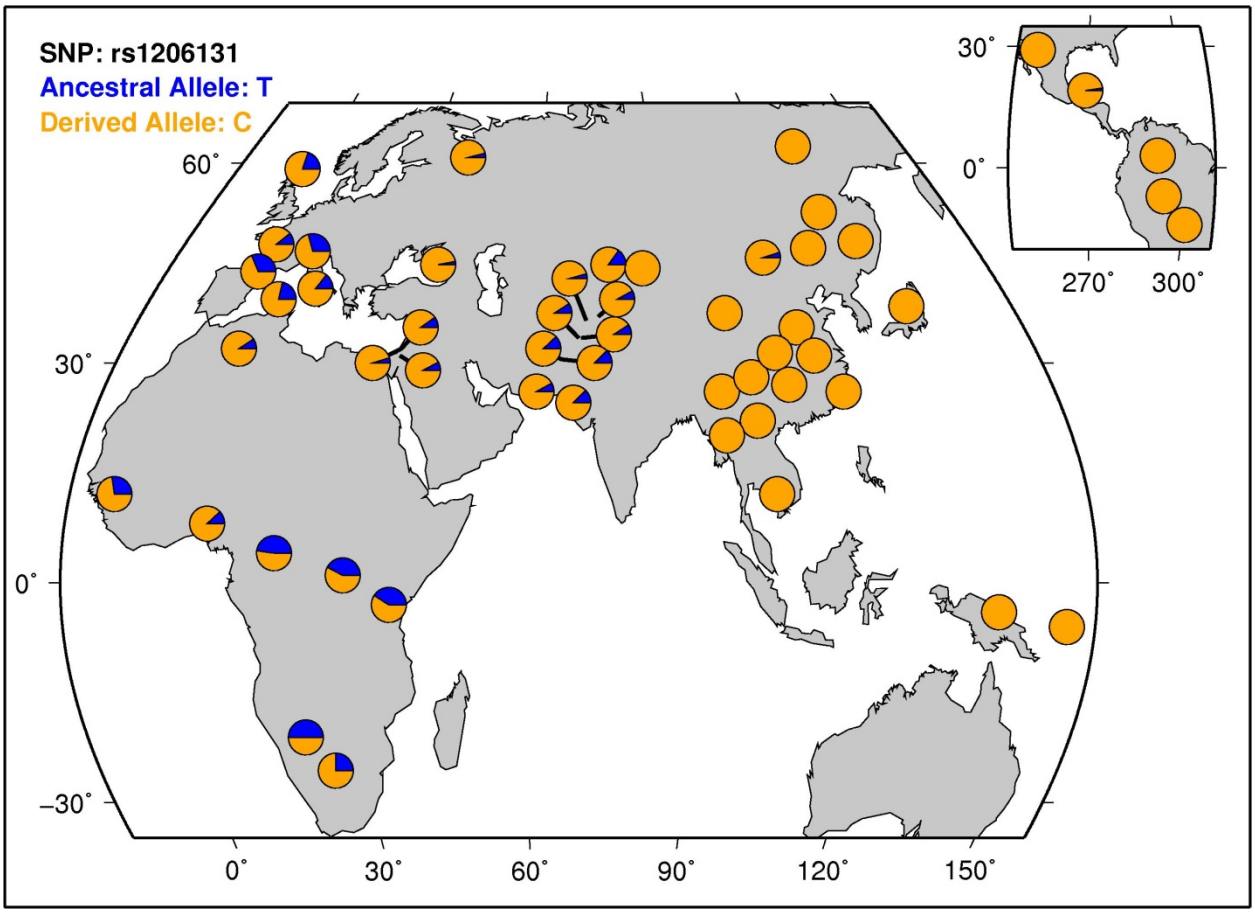
Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 8. LocusZoom Plot of –log(p-values) from Treeselect analysis along the (a) YRI and (b) AA branches of an unrooted tree with LWK, YRI and AA samples in *KLHL32* region



Nature Genetics: doi:10.1038/ng.2608

Supplementary Figure 9. Map of allele frequencies from the HGDP for rs1206131, near *KLHL32*, with geographic patterns of adaptation with lower frequency of the T allele comparing East and South African to East Asian and American populations



Nature Genetics: doi:10.1038/ng.2608

Supplementary Note

Description of Stage 1 Studies

Genome-wide association study of breast cancer in African American women (AABC).
AABC consists of individuals from the nine studies described below[1]. In each study, weight and height values were based on self-report.

The Carolina Breast Cancer Study (CBCS): The CBCS is a population-based case-control study conducted between 1993 and 2001 in 24 counties of central and eastern North Carolina [39]. Cases were identified by rapid case ascertainment system in cooperation with the North Carolina Central Cancer Registry and controls were selected from the North Carolina Division of Motor Vehicle and United States Health Care Financing Administration beneficiary lists. Participants' ages ranged from 20 to 74 years. For Stage 1, 656 African American cases with invasive breast cancer and 608 African American controls were included.

The Los Angeles component of The Women's Contraceptive and Reproductive Experiences Study (CARE): The NICHD Women's CARE Study is a large multi-center population-based case-control study that was designed to examine the effects of oral contraceptive use on invasive breast cancer risk among African American women and white women ages 35-64 years in five U.S. locations [40]. Cases in Los Angeles County were diagnosed from July 1, 1994 through April 30, 1998, and controls were sampled by random-digit dialing (RDD) from the same population and time period; 380 African American cases and 224 African American controls were included in Stage 1 of the scan.

The Multiethnic Cohort Study (MEC): The MEC is a prospective cohort study of 215,000 men and women in Hawaii and Los Angeles [41] between the ages of 45 and 75 years at baseline (1993-1996). Through December, 31 2007, a nested breast cancer (BC) case-control study in the MEC included 556 African American cases (544 invasive and 12 in situ) and 1,003 African American controls, all of which are included in the Stage 1 samples.

The Nashville Breast Health Study (NBHS): The NBHS is a population-based case-control study of breast cancer conducted in Tennessee [42]. The study was initiated in 2001 to recruit patients with invasive breast cancer or ductal carcinoma in situ, and controls, recruited through RDD between the ages of 25 and 75

Nature Genetics: doi:10.1038/ng.2608

years. NBHS contributed 310 African American cases (57 in situ), and 186 African American controls to Stage 1 of the GWAS.

The Northern California Breast Cancer Family Registry (NC-BCFR): The NC-BCFR is a population-based family study conducted in the Greater San Francisco Bay Area, and is one of 6 sites collaborating in the Breast Cancer Family Registry (BCFR), an international consortium funded by NCI [43]. African American breast cancer cases in NC-BCFR were diagnosed after January 1, 1995 and between the ages of 18 and 64 years; population controls were identified through RDD. Stage 1 genotyping was conducted for 440 invasive African American cases and 53 African American controls.

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) Cohort: PLCO, coordinated by the U.S. National Cancer Institute (NCI) in 10 U.S. centers, enrolled during 1993 - 2001 approximately 155,000 men and women, aged 55-74 years, in a randomized, two-arm trial to evaluate the efficacy of screening for these four cancers [44]. A total of 64 African American invasive breast cancer cases and 133 African American controls contributed to Stage 1 of the GWAS.

The San Francisco Bay Area Breast Cancer Study (SFBCS): The SFBCS is a population-based case-control study of invasive breast cancer in Hispanic, African American and non-Hispanic White women conducted between 1995 and 2003 in the San Francisco Bay Area [45]. African American cases, ages 35-79 years, were diagnosed between April 1, 1995 and April 30, 1999, with controls identified through RDD. Stage 1 included 172 invasive African American cases and 231 African American controls from the SFBCS.

Wake Forest University Breast Cancer Study (WFBC): African American breast cancer cases and controls in WFBC were recruited at Wake Forest University Health Sciences from November 1998 through December 2008 [46]. Controls were recruited from the patient population receiving routine mammography at the Breast Screening and Diagnostic Center. Age range of participants was 30-86 years. WFBC contributed 125 cases (116 invasive and 9 in situ) and 153 controls to the Stage 1 analysis.

The Women's Circle of Health Study (WCHS): The WCHS is a case-control study of breast cancer initiated in the New York City boroughs (Manhattan, the Bronx, Brooklyn and Queens) and now in seven counties in New Jersey (Bergen, Essex, Hudson, Mercer, Middlesex, Passaic, and Union)[47]. Eligible cases included women with invasive breast cancer between 20 and 74 years of age; controls were identified

Nature Genetics: doi:10.1038/ng.2608

through RDD. The WCHS contributed 272 invasive African American cases and 240 African American controls to Stage 1 of the GWAS.

Genome-wide association study of prostate cancer in men of African ancestry, part I (AAPC1). AAPC1 includes individuals from the 10 case-control studies of prostate cancer in men of African ancestry described below that are included in Stage 1[2]. In all AAPC1 studies information on weight and height was based on self-report.

The Cancer Prevention Study II Nutrition Cohort (CPS-II): The CPS-II Nutrition Cohort includes over 86,000 men and 97,000 women from 21 US states who completed a mailed questionnaire in 1992 (aged 40-92 years at baseline) [48]. Starting in 1997, follow-up questionnaires were sent to surviving cohort members every other year to update exposure information and to ascertain occurrence of new cases of cancer; a >90% response rate has been achieved for each follow-up questionnaire. From 1998-2001, blood samples were collected in a subgroup of 39,376 cohort members. To further supplement the DNA resources, during 2000-2001, buccal cell samples were collected by mail from an additional 70,000 cohort members. Incident cancers are verified through medical records, or through state cancer registries or death certificates when the medical record cannot be obtained. Genomic DNA from 76 African American prostate cancer cases and 152 age-matched controls were included in Stage 1 of the scan.

Case-Control Study of Prostate Cancer among African Americans in Washington, DC (DCPC): Unrelated men self-described as African American were recruited for several case-control studies on genetic risk factors for prostate cancer between the years 2001 and 2005 from the Division of Urology at Howard University Hospital (HUH) in Washington, DC. Control subjects unrelated to the cases and matched for age (± 5 years) were also ascertained from the prostate cancer screening population of the Division of Urology at HUH [49]. These studies included 292 cases and 359 controls.

The Gene-Environment Interaction in Prostate Cancer Study (GECAP): The Henry Ford Health System (HFHS) recruited cases diagnosed with adenocarcinoma of the prostate of Caucasian or African-American race, less than 75 years of age, and living in the metropolitan Detroit tri-county area [50]. Controls were randomly selected from the same HFHS population base from which cases were drawn. The control sample was frequency matched at a ratio of 3 enrolled cases to 1 control based on race and five-year age stratum. In total, 637 cases and 244 controls were enrolled between January 2002 and

Nature Genetics: doi:10.1038/ng.2608

December 2004. Of study enrollees, DNA for 234 African Americans cases and 92 controls were included in Stage 1 of the scan.

King County (Washington) Prostate Cancer Studies (KCPCS): The study population consists of participants from one of two population-based case-control studies among residents of King County, Washington [51,52]. Incident Caucasian and African American cases with histologically confirmed prostate cancer were ascertained from the Seattle-Puget Sound SEER cancer registry during two time periods, 1993-1996 and 2002-2005. Age-matched (5-year age groups) controls were men without a self-reported history of being diagnosed with prostate cancer and were identified using one-step random digit telephone dialing. Controls were ascertained during the same time periods as the cases. Detailed in-person interviews collected self-reported data on weight one year prior to reference date (date of diagnosis for cases and a randomly assigned date for controls that approximated the distribution of diagnosis dates of cases) and maximum adult height. A total of 145 incident African American cases and 81 African American controls were included from these studies.

The Los Angeles Study of Aggressive Prostate Cancer (LAAPC): The LAAPC is a population-based case-control study of aggressive prostate among African Americans in Los Angeles County [53]. Cases were identified through the Los Angeles County Cancer Surveillance Program rapid case ascertainment system and eligible cases included African American men diagnosed with a first primary prostate cancer between January 1, 1999 and December 31, 2003. Eligible cases also had either tumor extension outside the prostate, metastatic prostate cancer in sites other than prostate, or needle biopsy of the prostate with Gleason grade 8 or higher, or Gleason grade 7 and tumor in more than 2/3 of the biopsy cores. Controls were identified by a neighborhood walk algorithm and were men never diagnosed with prostate cancer, and were frequency matched to cases on age (±5 years). For this study, genomic DNA was included for 296 cases and 140 controls. We also included an additional 163 African American controls from the MEC that were frequency matched to cases on age.

The Multiethnic Cohort (MEC): The MEC is a prospective cohort study of 215,000 men and women in Hawaii and Los Angeles[1] between the ages of 45 and 75 years at baseline (1993-1996). Through January 1, 2009 the African American prostate cancer (PC) case-control study in the MEC included 1,094 cases and 1,096 controls, all of which contributed to the Stage 1 GWAS.

Prostate Cancer Genetics Study (CaP Genes): The African-American component of this study population comprised 160 men: 75 cases diagnosed with more aggressive prostate cancer and 85 age-matched

Nature Genetics: doi:10.1038/ng.2608

controls [54]. All subjects were recruited and frequency-matched on the major medical institutions in Cleveland, Ohio (i.e., the Cleveland Clinic, University Hospitals of Cleveland, and their affiliates) between 2001 and 2004. The cases were newly diagnosed with histologically confirmed disease: Gleason score 7; tumor stage T2c; or a prostate-specific antigen level >10 ng/ml at diagnosis. Controls were men without a prostate cancer diagnosis who underwent standard annual medical examinations at the collaborating medical institutions.

Prostate Cancer Case-Control Studies at MD Anderson (MDA): Participants in this study were identified from epidemiological prostate cancer studies conducted at the University of Texas M.D. Anderson Cancer Center in the Houston Metropolitan area since 1996. Cases were accrued from six institutions in the Houston Medical Center and were not restricted with respect to Gleason score, stage or PSA. Controls were identified via random-digit-dialing or among hospital visitors and they were frequency matched to cases on age and race. Lifestyle, demographic, and family history data were collected using a standardized questionnaire. These studies contributed 543 African American cases and 474 controls to this study [55].

*The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO):*The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial is discussed above [56]. Participants for this study completed a baseline risk factor questionnaire, provided either a blood sample (intervention arm) or buccal cell sample (control arm), and consented to participate in genetic etiologic studies. Included in this study are 286 African American prostate cancer cases and 269 controls without a history of prostate cancer, matched on age at randomization, year of randomization, and study year of the trial.

The Southern Community Cohort Study (SCCS): The SCCS is a prospective cohort of African Americans and non-African Americans which during 2002-2009 enrolled approximately 86,000 residents aged 40-79 years across 12 southern states [57]. Recruitment occurred mainly at community health centers, institutions providing basic health services primarily to the medically uninsured, so that the cohort includes many adults of lower income and educational status. Each study participant completed a detailed baseline questionnaire, and nearly 90% provided a biologic specimen (approximately 45% a blood sample and 45% buccal cells). Follow-up of the cohort is conducted by linkage to national mortality registers and to state cancer registries. Height and weight self-reported for the majority of study participants. For those who had clinic appointments on the day of interview, weight and height were abstracted from their medical record for that day and were found to be highly correlated (>.95) with self-reported values. For approximately 10% of those enrolled, the study interviewers measured height, weight and waist and hip

Nature Genetics: doi:10.1038/ng.2608

circumference. Included in Stage 1 of this study are 212 incident African American prostate cancer cases and a matched stratified random sample of 419 African American male cohort members without prostate cancer at the index date selected by incidence density sampling.

BioVU, the Vanderbilt DNA Databank project: BioVU, the Vanderbilt DNA Databank, is a DNA databank collected from patients at Vanderbilt University Medical Center linked to de-identified electronic medical records. A major goal of the resource is to generate datasets that incorporate de-identified information derived from medical records and genotype information to identify factors that affect disease susceptibility, disease progression, and/or drug response. Vanderbilt's unique algorithm to de-identify the samples led to a non-human subject designation from the Institutional Review Board, allowing the use of blood samples collected for clinical care that otherwise would be discarded. The program has received approval from the IRB and was reviewed in detail by the federal Office for Human Research Protections (OHRP), who agreed with the non-human subjects regulatory designation for both the resource and subsequent research. Program planning described in the proposal (e.g. Community and Ethics Committee involvement, sample handling routines) started in 2004, and sample accrual started at the end of February 2007. Weight and height were measured during clinical examinations of study subjects, using either mechanical or digital scales for weight and a stadiometer for height. We used the first recorded measurements of weight and height in the de-identified electronic medical records as the baseline measures for all eligible subjects. In stage 1, we used GWAS data from 1,237 unrelated African American BioVU participants, of which 1,136 passed quality control criteria.

The NHLBI Candidate gene Association Resource (CARe) Project [58]. The CARe cohort consisted of five studies described below [59].

Atherosclerosis Risk in Communities (ARIC): The ARIC study is a prospective population-based study of atherosclerosis and cardiovascular diseases in 15,792 men and women, including 11,478 non-Hispanic whites and 4,314 African Americans, drawn from 4 U.S. communities (suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi)[3]. In the first three communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Because of the design and focus of the CARe Project, only self-reported African-American participants are included in this analysis. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing. After taking into account availability of adequate amounts of high quality DNA, appropriate informed consent and genotyping quality control and assurance

Nature Genetics: doi:10.1038/ng.2608

procedures, genotype data were available on 2,989 African-American individuals. Weight was measured to the nearest pound using a balance scale. Height was measured to the nearest centimeter using a wall-mounted ruler. Both weight and height were measured without shoes.

Coronary Artery Risk Development in young Adults (CARDIA): The CARDIA study is a prospective, multi-center investigation of the natural history and etiology of cardiovascular disease in African Americans and whites 18-30 years of age at the time of initial examination. The initial examination included 5,115 participants selectively recruited to represent proportionate racial, gender, age, and education groups from four communities: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. Participants from the Birmingham, Chicago, and Minneapolis centers were recruited from the total community or from selected census tracts. Participants from the Oakland center were randomly recruited from the Kaiser-Permanente health plan membership. Details of the study design have been published [5]. From the time of initiation of the study in 1985-1986, five follow-up examinations have been conducted at years 2, 5, 7, 10, 15, and 20. DNA extraction for genetic studies was performed at the Y10 examination. After taking into account availability of adequate amounts of high quality DNA, appropriate informed consent and genotyping quality control and assurance procedures, genotype data were available on 955 African-American individuals. Each participant's age, race, and sex were self-reported during the recruitment phase and verified during the baseline clinic visit. Weight was measured using a *Detecto* Scale (Model #68965) with additional weights (#68967) to allow weighing up to 450 pounds. Height was measured using an anthropometric ruler in centimeters or stadiometer. Measurements were made while participants were wearing minimal clothing (short sleeve shirt, shorts, and socks) and without shoes.

Cleveland Family Study (CFS): The Cleveland Family Study (CFS) is a family-based, longitudinal study designed to characterize the genetic and non-genetic risk factors for sleep apnea[6]. In total, 2534 individuals (46% African American) from 352 families were studied on up to 4 occasions over a period of 16 years (1990-2006). The initial aim of the study was to quantify the familial aggregation of sleep apnea. Over time, the aims were expanded to characterize the natural history of sleep apnea, sleep apnea outcomes, and to identify the genetic basis for sleep apnea. With subsequent exams, the cohort was expanded to include increased minority representation and additional family members. The total sample included index probands (N=275) who were recruited from 3 area hospital sleep centers if they had a confirmed diagnosis of sleep apnea and at least 2 first-degree relatives available to be studied. In the first 5 years of the study, neighborhood control probands (N=87) with at least 2 living relatives available for study were selected at random from a list provided by the index family. All available first-degree relatives

Nature Genetics: doi:10.1038/ng.2608

and spouses of the case and control probands were recruited. Second-degree relatives, including half-sibs, aunts, uncles and grandparents, were also included if they lived near the first degree relatives (cases or controls), or if the family had been found to have two or more relatives with sleep apnea. The sample, which is enriched for individuals with sleep apnea, also contains a high prevalence of individuals with sleep apnea-related traits, including: obesity, impaired glucose tolerance, and hypertension. Data that were used for the CARe analyses were for individuals in whom DNA had been collected (i.e., over the last two exam cycles (N=1447). Height was measured using a rigid stadiometer. Weight was measured with a calibrated digital scale. Height and weight data used in the current analysis were from the last exam conducted in 736 African American subjects.

Jackson Heart Study (JHS): The Jackson Heart Study (JHS) is a prospective population-based study to seek the causes of the high prevalence of common complex diseases among African Americans in the Jackson, Mississippi metropolitan area, including cardiovascular disease, type-2 diabetes, obesity, chronic kidney disease, and stroke [60]. During the baseline examination period (2000-2004) 5,301 self-identified African Americans were recruited from four sources, including (1) randomly sampled households from a commercial listing; (2) ARIC participants; (3) a structured volunteer sample that was designed to mirror the eligible population; and (4) a nested family cohort. Unrelated participants were between 35 and 84 years old, and members of the family cohort were ≥ 21 years old when consent for genetic testing was obtained and blood was drawn for DNA extraction. Based on DNA availability, appropriate informed consent, and genotyping results that met quality control procedures, genotype data were available for 3,030 individuals, including 885 who are also ARIC participants. In the current study, JHS participants who were also enrolled in the ARIC study were analyzed with the ARIC dataset – for this reason, the JHS dataset analyzed here had 2,145 individuals. Weight was measured using a calibrated balance scale (Detector model #437), wearing light clothing and no shoes. Height was measured without shoes using a vertical ruler.

Multi-Ethnic Study of Atherosclerosis (MESA): The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease[20]. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84 at baseline. Approximately 38% of the recruited participants are white, 28% African-American, 22% Hispanic, and 12% Asian, predominantly of Chinese descent. For the current study, after taking into account availability of adequate amounts of high quality DNA, appropriate informed consent and genotyping quality control and

Nature Genetics: doi:10.1038/ng.2608

assurance procedures, genotype data were available on 1,646 African-American individuals. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California, Los Angeles. Each participant received an extensive physical exam to determine coronary calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness, lower extremity vascular insufficiency, arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, socio-demographic factors, lifestyle factors, and psychosocial factors. DNA is extracted and lymphocytes immortalized for study of candidate genes and genome-wide scanning. Participants are followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. In addition to the six Field Centers, MESA involves a Coordinating Center, a Central Laboratory, and Central Reading Centers for Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, and Electrocardiography (ECG). Protocol development, staff training, and pilot testing were performed in the first 18 months of the study. The first examination took place over two years, from July 2000 - July 2002. It was followed by three additional examination periods: September 2002 – January 2004, February 2004 - July 2005 September 2005 - May 2007. Participants are contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality. NHLBI recently funded MESA II, which brought back all participants back for a fifth exam, starting in April 2010. Height and weight were measured with participants wearing light clothing and no shoes.

The Cardiovascular Health Study (CHS): The CHS is a prospective population-based cohort study of risk factors for CHD and stroke in adults 65 years and older. In June 1990, four Field Centers completed the recruitment of 5201 participants[7]. Between November 1992 and June 1993, an additional 687 African Americans were recruited using similar methods. The Field Centers are located in Forsyth County, NC; Sacramento County, CA; Washington County, MD; and Pittsburgh, PA. The baseline examinations consisted of a home interview and a clinic examination. Weight was measured to the nearest 0.1 Lb with the participant wearing light clothes. Height was measured to the nearest 0.1 cm. Body Mass Index was computed as weight (in kg) divided by height squared (in meters).

The Family Heart Study (FamHS): The Family Heart Study began in 1992, and subjects were recruited for an extensive clinical examination during the years 1994-96[9]. A second visit, approximately 8 years later, was conducted, in which a sample of African-American families (N=624) was recruited at the University of Alabama in Birmingham field center. The primary purpose of the visit was to obtain two

Nature Genetics: doi:10.1038/ng.2608

sequential cardiac multidetector CT (MDCT) exams for individual coronary arteries (i.e., left main, left anterior descending, diagonals, circumflex and right coronary arteries) and the aorta to estimate subclinical atherosclerosis burden as calcification. Measurements of the most important CHD risk factors in the lipid, glucose metabolism, blood pressure, and anthropometry domains also were assessed, along with medical history and medication use. Anthropometric measurements were performed with the participants wearing a scrub suit or examination gown, non-constricting underwear and no shoes. Participants were given the opportunity to empty their bladder before taking measurements. Weight was measured to the nearest pound on a calibrated balance scale. Height was measured without shoes to the nearest centimeter with a stadiometer. Body Mass Index (BMI) was computed as weight (in kg) divided by height squared (in meters).

Genetic Epidemiology Network of Arteriopathy (GENOA): GENOA is one of the four networks in the Family Blood Pressure Program[61]. GENOA recruited hypertensive African-American and non-Hispanic white sibships for linkage and family-based association studies to investigate genetic contributions to blood pressure and the cardiac and renal complication of hypertension in multiple racial groups[12]. Participant recruitment for GENOA (1995-2000) was conducted in Jackson, Mississippi and Rochester, Minnesota. African Americans in the study were located solely at the GENOA Jackson field center. Hypertensive probands were ascertained from the Jackson cohort of the Atherosclerosis Risk in Communities (ARIC) Study if they were in a sibship with ≥ 2 individuals with essential hypertension diagnosed prior to age 60 years (prior diagnosis of hypertension by a physician and use of prescription antihypertensive medication reported at the study visit; or systolic BP $\geq$140 mm Hg or diastolic BP $\geq$90 mm Hg on the second and third clinic visit), and were willing to be recruited. Exclusion criteria included secondary hypertension, alcoholism or drug abuse, insulin-dependent diabetes mellitus, active malignancy, or pregnancy or lactation. All available siblings of the index sibling pairs were invited to participate. Height was measured by a wall stadiometer. Participants in light clothing and without shoes were weighed with an electronic balance. BMI was calculated in units of kg/m^2. In the current study, GENOA participants who were also participants in the ARIC Study were analyzed in the ARIC dataset. A total of 996 individuals were included in the current analyses. The institutional review board at each of the study sites approved the study protocols, and written informed consent was obtained from all participants.

Health, Aging, and Body Composition (Health ABC) Study: The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults_ENREF_33. Health ABC enrolled well-functioning, community-dwelling black (N=1281) and white (N=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of white and all black

Nature Genetics: doi:10.1038/ng.2608

Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Participants have undergone annual exams and semi-annual phone interviews. Height and weight were measured with participants wearing light clothing and no shoes. BMI was computed as weight in kilograms divided by height in meters squared.

The Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) study: The HANDLS study is an interdisciplinary, community-based, prospective longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among socioeconomically diverse African Americans and whites in Baltimore[13] ENREF 10. A total of 3,722 participants were recruited from Baltimore, MD with mean age 47.7 (range 30-64) years. We used a Health O Meter Digital Lithium Scale (Model #HDL 976) to measure weight in kg and a Novel Products Inc. Height Meter (Model #DES 290237) to measure height in cm. Participants stood erect on the floor with their heels placed together and touching the wall behind them and their back against the vertical height meter, which was mounted securely on the wall. Technician recorded centimeters when the height meter level was parallel to the floor. Participants stood on the scale, which was placed flat on a level surface with nothing underneath. The scale was zeroed each morning. Readings were taken in kilograms.

Howard University Family Study (HUFS): The HUFS is a population-based study of African American families enrolled from the Washington, D.C. metropolitan area[14]. In the first phase of recruitment, a randomly ascertained cohort of 350 African American families with members in multiple generations from the Washington, D.C. metropolitan area were enrolled and examined. Families were not ascertained based on any phenotype. In a second phase of recruitment, additional unrelated individuals from the same geographic area were enrolled. The total number of recruited individuals was 2,028, of which 1,976 remained after data cleaning. Weight was measured on an electronic scale to the nearest 0.1 kg with the participant wearing light clothes. Height was measured with a stadiometer to the nearest 0.1 cm with participants in bare feet. Ethical approval was obtained from the Howard University Institutional Review Board and written informed consent was obtained from each participant.

Hypertensive Genetic Epidemiology Network (HyperGEN): HyperGEN is part of the Family Blood Pressure Program funded by the National Heart Lung and Blood Institute and was designed to study the genetics of hypertension and related conditions[62] . Participants were recruited from multiply-affected hypertensive sibships ascertained through population-based cohorts or from the community-at-large. The study was later extended to include siblings and offspring of the original sibpair. Probands were

Nature Genetics: doi:10.1038/ng.2608

identified by the onset of hypertension before age 60 and the presence of at least one additional hypertensive sibling who was willing to participate. Participants with type 1 diabetes or advanced renal disease were excluded from HyperGEN. Recruitment, clinical measurement, and DNA isolation were completed in 2003. Two of four centers (AL, NC) recruited 1,277 African Americans, while three centers (NC, MN, and UT) recruited Caucasians. The study was approved by the University of Alabama's Internal Review Board for Human Use, the Washington University Human Research Production Office, the University of North Carolina's Office of Human Research Ethics and the Medical College of Wisconsin's Office of Human Research Protection Program. All subjects provided written informed consent. Body mass index was computed as weight (in kg) divided by height squared (in meters).

Maywood Cohort: The Maywood cohort consisted of African Americans from Maywood, IL, who were enrolled in studies of blood pressure at the Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA[19]. Study protocols were reviewed and approved by the institutional review board at the Loyola University Chicago Stritch School of Medicine prior to all recruitment activities. Written informed consent was obtained from each participant. Phenotype measurements were performed using a standardized protocol [63,64]. Body weight was measured to the nearest 0.2 kg on calibrated electronic scales, while height was obtained using a stadiometer consisting of a steel tape attached to a straight wall and a wooden headboard. The headboard was positioned with the participant shoeless, feet and back against the wall, and head held in the Frankfort horizontal plane and measurement taken to the nearest 0.1 cm [65]. Body mass index was calculated as the ratio of weight in kilograms to the square of height in meters. A total of 775 unrelated participants with GWAS data (4) were included in the present study.

Mesa Air/Family: The goal of the Multi-Ethnic Study of Atherosclerosis and Air Pollution ('MESA Air') is to prospectively examine the relation between individual level assessment of long-term ambient air pollution exposure (including PM2.5 and gaseous co-pollutants) and the progression of subclinical cardiovascular disease in a multi-city, multi-ethnic cohort. MESA Air will also prospectively examine the relationship between individual level assessment of long-term ambient air pollution exposure and the incidence of cardiovascular disease, including myocardial infarction and cardiovascular death. MESA Air is built on the foundation of the ongoing MESA study, a large NIH/NHLBI cohort study. MESA uses state-of-the-art tools to assess the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and risk factors that predict progression to clinically overt cardiovascular disease, and that predict progression of sub-clinical disease itself, in a diverse, population-based sample of men and women aged 45-85. The cohort for the MESA Air study includes 6226 subjects: 5479 enrolled in the parent MESA study; 257 recruited specifically for

Nature Genetics: doi:10.1038/ng.2608

this study, and 490 recruited for another ancillary MESA study (MESA family), the MESA Family study. The entire MESA Air cohort will be followed over the 10-year project period for the occurrence of cardiovascular disease events.

The general goal of the MESA Family Study is to locate and identify genes contributing to the genetic risk for cardiovascular disease, by looking at the early changes of atherosclerosis within families (mainly siblings)[20]. The MESA Family cohort was recruited from the six MESA Field Centers. MESA Family participants underwent the same examination as MESA participants during May2004 - May 2007. In a small proportion of subjects, parents of MESA index subjects participating in MESA Family were studied but only to have blood drawn for genotyping. DNA was extracted and lymphocytes immortalized for 1633 non-classic MESA family members (950 African Americans and 683 Hispanic-Americans) from 594 families, yielding 3,026 sibpairs.

Nigerian Cohort: The population sampling frame for the Nigerian cohort was provided by the International Collaborative Study on Hypertension in Blacks (ICSHIB) as described in detail elsewhere [66]. Participants of the Nigerian cohort were recruited from Igbo-Ora and Ibadan in southwest Nigeria as a long-term study on the environmental and genetic factors underlying hypertension[19] _ENREF_38. The study protocol was reviewed and approved by the Institutional Review Board at Loyola University Medical Center, and the Joint Ethical Committee of the University of Ibadan/University College Hospital, Ibadan, Nigeria. All participants gave written informed consent administered in either English or Yoruba. Phenotype measurements were performed by trained research staff using a standardized protocol [63,64]. Body weight was measured to the nearest 0.2 kg on calibrated electronic scales, whereas height was obtained using a stadiometer consisting of a steel tape attached to a straight wall and a wooden headboard. The headboard was positioned with the participant shoeless, feet and back against the wall and head held in the Frankfort horizontal plane and measurement taken to the nearest 0.1 cm. Body mass index was calculated as the ratio of weight in kilograms to the square of height in meters. A total of 1,188 unrelated adults with GWAS data_ENREF_39 were evaluated in the present analysis.

Wake Forest University School of Medicine (WFSM) Study: The WFU GWAS study is a case-control study for type 2 diabetes and nephropathy[67]. Non-diabetic subjects who reported no history of diabetes and renal disease were recruited from the community and internal medicine clinics at WFU. Diabetic subjects with type 2 diabetes and end stage renal disease were recruited from dialysis facilities in the southeastern U.S. Patients with type 2 diabetes were diagnosed after the age of 25 and did not receive only insulin therapy since diagnosis. In addition, cases had to have at least one of the following three

Nature Genetics: doi:10.1038/ng.2608

criteria for inclusion: a) type 2 diabetes diagnosed at least 5 years before initiating renal replacement therapy, b) background or greater diabetic retinopathy and/or c) ≥100 mg/dl proteinuria on urinalysis in the absence of other causes of nephropathy. All subjects were recruited in North Carolina, South Carolina, Georgia, Tennessee, or Virginia. Informed consent was obtained from all study participants. Recruitment and sample collection procedures were approved by the Institutional Review Boards at WFU. Among 1994 participants in the GWAS study, 816 non-diabetic subjects and 899 subjects with type 2 diabetes and end stage renal disease who have information on weight and height were included in this study. Measurements are either by self-report or by stadiometer for height and calibrated scale for weight.

Women's Health Initiative – SNP Health Association Resource (WHI-SHARe). WHI is one of the largest (n=161,808) studies of women's health ever undertaken in the US[29]. There are two major components of WHI: (1) a Clinical Trial (CT) that enrolled and randomized 68,132 women ages 50 – 79 into at least one of three placebo-control clinical trials (hormone therapy, dietary modification, and calcium/vitamin D); and (2) an Observational Study (OS) that enrolled 93,676 women of the same age range into a parallel prospective cohort study. A diverse population including 26,045 (17%) women from minority groups were recruited from 1993-1998 at 40 clinical centers across the U.S. Of the CT and OS minority participants enrolled in WHI, 12,157 (including 8,515 self-identified African American and 3,642 self-identified Hispanic subjects) who had consented to genetic research were eligible for the WHI SHARe GWAS project. DNA was extracted by the Specimen Processing Laboratory at the Fred Hutchinson Cancer research Center (FHCRC) using specimens that were collected at the time of enrollment. Specimens were stored at -80°C. Blood samples for WBC analyses were collected at baseline. Weight was taken after removing shoes, heavy clothing, and pocket contents on a calibrated digital scale and recorded to the nearest one-tenth kilogram. Height was taken using a wall-mounted stadiometer and recorded to nearest one-tenth centimeter. BMI was calculated from measured height and weight. Study protocols and consent forms were approved by the Institutional Review Boards at all participating institutions. Height was measured at an in-person visit with a stadiometer and weight was measured at an in-person visit with a calibrated scale.

Nature Genetics: doi:10.1038/ng.2608

Description of Studies in Stages 2 and 3

Genome-wide association study of prostate cancer in men of African ancestry, part II (AAPC2). AAPC2 includes individuals from the 4 case-control studies of prostate cancer in men of African ancestry described below that were included in follow-up analyses of the top hits from Stage 1[2]. Individuals in AAPC2 were genotyped at a later date and thus were not available for inclusion in Stage 1. In all AAPC2 studies information on weight and height was based on self-report (except PCBP).

The Multiethnic Cohort (MEC): The MEC is discussed above. An additional 705 prostate cancer cases from the MEC diagnosed after January 1, 2009 and 605 controls with GWAS data and BMI information were included as part of the Stage 2 sample. In Stage 3, we also genotyped the SNPs rs7708584, rs974417 and rs10261878 in an additional 4,327 African American samples in the MEC from studies of type 2 diabetes (male cases: 159, male controls: 179 female cases: 640 and female controls: 587) as well as an additional 446 male and 2,316 female samples unselected for any phenotype. Principle components were estimated for these samples using AIMs from the Metabochip.

Prostate Cancer in a Black Population (PCBP): The PCBP is a population-based case-control study of prostate cancer conducted in Barbados, West Indies. The study (2002-2011) included all incident, histologically-confirmed cases of prostate cancer ascertained from the Pathology Department of the Queen Elizabeth Hospital, Bridgetown, the only institution on the island where specimens are evaluated. Controls were randomly selected from a national database and frequency matched (by 5-year age groups) to the cases. Weight was measured in pounds using a beam balance scale and height was measured in centimeters using a metric rule attached to a wall and a right-angled wood block. In Stage 2, we included an additional 224 prostate cancer cases and 234 controls with GWAS data.

Selenium and Vitamin E Cancer Prevention Trial (SELECT): SELECT is a phase III, placebo-controlled trial that tested whether selenium and vitamin E alone or in combination, might reduce the risk of developing prostate cancer[68]. A total of 35,534 men 55 and older (50 years and older for African Americans) without a history of prostate cancer were enrolled between 2001 and 2004. About 12% of the SELECT participants are African American. A case-cohort study has been established in SELECT and, as of December 31, 2009, includes 263 African American prostate cancer cases and 789 African American non-cases, all of which are included in the Stage 2 analysis.

Nature Genetics: doi:10.1038/ng.2608

The Southern Community Cohort Study (SCCS): The SCCS is discussed above. In Stage 2 we included an additional 353 incident and prevalent cases from the SCCS (incident cases diagnosed after June 1, 2006) and 706 controls with GWAS data. Height and weight self-reported for the majority of study participants. For those who had clinic appointments on the day of interview, weight and height were abstracted from their medical record for that day and were found to be highly correlated (>.95) with self-reported values. For approximately 10% of those enrolled, the study interviewers measured height, weight and waist and hip circumference.

BioVU, the Vanderbilt DNA Databank project: BioVU, the Vanderbilt DNA Databank, is a DNA databank collected from patients at Vanderbilt University Medical Center linked to de-identified electronic medical records. A major goal of the resource is to generate datasets that incorporate de-identified information derived from medical records and genotype information to identify factors that affect disease susceptibility, disease progression, and/or drug response. Vanderbilt's unique algorithm to de-identify the samples led to a non-human subject designation from the Institutional Review Board, allowing the use of blood samples collected for clinical care that otherwise would be discarded. The program has received approval from the IRB and was reviewed in detail by the federal Office for Human Research Protections (OHRP), who agreed with the non-human subjects regulatory designation for both the resource and subsequent research. Program planning described in the proposal (e.g. Community and Ethics Committee involvement, sample handling routines) started in 2004, and sample accrual started at the end of February 2007. Weight and height were measured during clinical examinations of study subjects, using either mechanical or digital scales for weight and a stadiometer for height. We used the first recorded measurements of weight and height in the de-identified electronic medical records as the baseline measures for all eligible subjects. In stage 3, we used GWAS data to examine the three variants in an additional 668 African American BioVU participants who were unrelated to stage 1 BioVU participants at the second cousin level, as estimated by IBS analysis of GWAS data in PLINK.

The Black Women's Health Study (BWHS): The BWHS is an ongoing prospective cohort study of 59,000 African American women[69]. In 1995, 59,000 African-American women 21-69 years of age from across the U.S. enrolled in the BWHS by completing a 14-page postal health questionnaire. The median age at entry was 38, and participants were residents of 17 states in mainland U.S. The baseline questionnaire elicited information on a wide range of variables and biennial follow-up questionnaires are used to identify new cases of disease outcomes and to update covariate information. Participants reported weight in pounds and adult height in feet and inches. DNA samples were obtained from BWHS participants by the mouthwash-swish method. Approximately 50% of participants, 27,800 women,

Nature Genetics: doi:10.1038/ng.2608

provided a sample. Women who provided samples were slightly older than women who did not, but the two groups were closely similar with regard to educational level, geographic region, body mass index, and reproductive factors. Replication genotyping was carried out in samples from a nested case-control study of breast cancer, with controls matched to the cases on year of birth, geographic region of residence, and country of birth (U.S. or other). Included were 838 incident cases and 1,872 controls.

Genetic Study of Atherosclerosis Risk (GeneSTAR): GeneSTAR is an ongoing prospective study begun in 1983 designed to determine environmental, phenotypic, and genetic causes of premature cardiovascular disease[10,11]. Participants came from families identified from 1983-2006 from probands with a premature coronary disease event prior to 60 years of age who were identified at the time of hospitalization in any of 10 Baltimore area hospitals. Their apparently healthy 30-59 year old siblings without known CAD were recruited and underwent phenotypic measurement and characterization between 1983 and 2006. From 2003-2006, adult offspring over 21 years of age of all participating siblings and probands and the coparent of the offspring were recruited and underwent phenotypic measurement and characterization. Weight was measured in pounds using a clinical balance scale with participants wearing light, indoor clothing. Height was measured in inches using a stadiometer.

Ghana: Participants were recruited into the Ghana study between May 2002 and October 2003 by word-of-mouth at churches and at the local market in Sunyani, Ghana[70,71]. Exclusion criteria were age less than 18, prior enrollment of a first or second degree relative, and any acute illness such as malaria that might affect levels of tissue-type plasminogen activator (t-PA) or plasminogen activator inhibitor-1 (PAI-1). Participants were examined at the Regional Hospital, Sunyani after a 10-hour fast. All participants provided written informed consent or fingerprint consent, and all forms were approved by the institutional review boards at Vanderbilt University and Regional Hospital, Sunyani. All participants provided medical history and standard demographic data including age, sex, education, smoking status, alcohol consumption, current medications, cardiovascular disease, diabetes, and cancers. Height, weight and blood pressure were measured. Blood pressures were measured using an Omron HEM-705c instrument (Omron Healthcare Corp., Bannockburn, Ill., USA). Participants were seated in a quiet room and two measures of blood pressure were taken from the left arm. The average of the two measures was used in the analysis. All blood pressure measures were taken prior to blood draws. Height and weight were measured without shoes using a height scale on the wall in the examining room, and using a commercial scale calibrated for accuracy. Three tubes of blood were taken from each participant, stored in liquid nitrogen, and shipped to Vanderbilt University. Genotyping of the SNPs rs7708584 and rs974417 for Stage 3 of this project was conducted using the Sequenom genotyping system at the Vanderbilt DNA

Nature Genetics: doi:10.1038/ng.2608

Resources Core. Candidate SNPs were genotyped and analyzed according to the analytic protocol for the discovery GWAS analysis, with the exception that principal components summarizing ancestry were not adjusted for due to the fact that these participants are not admixed with Europeans. 3,419 participants (1,452 males and 1,967 females) were available for the Stage 3 analyses after quality control.

Charles R. Bronfman Institute for Personalized Medicine (IPM) BioBank Genome Wide Association Study of Cardiovascular, Renal and Metabolic Phenotypes: The Institute for Personalized Medicine (IPM) Biobank Project is a consented, EMR-linked medical care setting biorepository of the Mount Sinai Medical Center (MSMC) drawing from a population of over 70,000 inpatients and 800,000 outpatient visits annually. MSMC serves diverse local communities of upper Manhattan, including Central Harlem (86% African American), East Harlem (88% Hispanic Latino), and Upper East Side (88% Caucasian/white) with broad health disparities. IPM Biobank populations include 28% African American, 38% Hispanic Latino predominantly of Caribbean origin, 23% Caucasian/White. IPM Biobank disease burden is reflective of health disparities with broad public health impact. Biobank operations are fully integrated in clinical care processes, including direct recruitment from clinical sites waiting areas and phlebotomy stations by dedicated Biobank recruiters independent of clinical care providers, prior to or following a clinician standard of care visit. Recruitment currently occurs at a broad spectrum of over 30 clinical care sites. Height and Weight data are from questionnaires submitted at the time of Biobank enrollment. Samples were genome-wide scanned in two phases: 887 samples which contributed to Stage 2 and 3,983 samples which were genotyped as part of Stage 3.

The Johnston County Osteoarthritis Project (JoCo): JoCo is an ongoing population-based prospective cohort study of the occurrence of knee and hip OA in African Americans and Caucasians in Johnston County, North Carolina[18]. This project was designed as a long-term study of ethnic differences in OA occurrence and progression. The samples were collected from six townships among the 17 townships in Johnston County because they contained the largest proportion of African American residents. The participants were initially recruited at the baseline between 1990 and 1997 and were followed up between 1999 and 2004. Additional new individuals were enrolled in 2003–2004 to enrich the sample for African Americans and younger individuals who were deliberately targeted for inclusion. A total of 2583 participants from the Johnston County Cohort were selected from the total study population for genotyping. The group selected for inclusion in the present association study was comprised of both Caucasians of European ancestry (68%) and African Americans (32%), of both sexes (35% men), and were all above 45 years of age. Purified DNA was extracted from whole blood samples, and genotyping

Nature Genetics: doi:10.1038/ng.2608

was done using Illumina Infinium 1M-Duo bead arrays. BMI was calculated from measured height and weight. Only African American subjects with non-missing covariates were used for this analysis.

The NCI Lung Cancer in African Americans GWAS consortium (NCI Lung). The five studies below are included in the NCI Lung.

MD Anderson Lung Cancer Epidemiology Study: African American lung cancer cases and controls were obtained from MD Anderson protocol CPN91-001 "Ecogenetics of Lung Cancer" (PI: Xifeng Wu). The study participants for this case-control study were a consecutive series of lung cancer cases recruited for an ongoing lung cancer study that has been accruing participants at The University of Texas MD Anderson Cancer Center since 1995. The control subjects were recruited from the Kelsey-Seybold Clinic, Houston's largest multidisciplinary physician practice. The control subjects were frequency matched to the cases on age (±5 years) and sex. Height and weight were self-reported. All participants provided informed consent, and the study was approved by the MD Anderson Institutional Review Board.

NCI-MD Lung Cancer Case Control Study: The NCI-MD Lung Cancer Case Control Study is an ongoing observational study in the metropolitan area of Baltimore, MD. Patients with histologically confirmed non-small-cell lung cancer (NSCLC) are recruited from seven area hospitals. Population controls are identified from the Department of Motor Vehicles, and matched to cases by age, race and gender. To be eligible, participants must be United States citizens; residing in Maryland; English-speaking and non-institutionalized. Cases may not have been interviewed previously as a control for the study and population controls may not have a history of cancer other than skin cancer. Written informed consent is obtained from all participants and the study is approved by the Institutional Review Boards of the participating institutions. Upon recruitment, cases and controls receive a structured, in person interview assessing prior medical and cancer history, tobacco use, alcohol use, current medications, occupational history, family medical history, menstrual history and estrogen use, recent nutritional supplements and caffeine intake, and socioeconomic status. Height and weight are self-reported. Participants provide a blood sample and/or mouthwash to collect cheek cells (oral cells). Follow-up is conducted through annual updates from the National Death Index (NDI).

Northern California Lung Cancer Study: The Northern California Lung Cancer Study was approved by the Committee on Human Research of the University of California, San Francisco and by the Institutional Review Boards of all collaborating institutions. The study obtained informed consent from participants. In this study, African American cases and controls older than 18 years of age were identified during two collection periods spanning September 1998-March 2003 and July 2005-March 2008. Cases in the first

Nature Genetics: doi:10.1038/ng.2608

accrual period were identified primarily through the Northern California Cancer Center (NCCC) rapid case ascertainment program. Cases in the second accrual period were identified through both the NCCC and the Kaiser Permanente Medical Care Program (KPMCP). Cases were Northern California residents presenting with previously untreated, histologically confirmed lung cancer[72]. Height and weight measurements were obtained by interviewer administered questionnaire. All UCSF samples included in the NCI GWAS of African American Lung Cancer were genomic DNA extracted from whole blood.

Project CHURCH (Creating a Higher Understanding of Cancer Research & Community Health): Project CHURCH is an ongoing community-based participatory research cohort study with local Houston churches and MD Anderson Cancer Center. The cohort study was developed to: 1) to better understand factors associated with cancer and cancer-related risk behaviors among African-American adults, and 2) to help prevent cancer and cancer-related health disparities among this population. We recruited 1501 participants and have followed them for 3 years (Baseline, Year 1 and Year 2). Questionnaires include lifestyle factors (i.e. diet, physical activity, tobacco), cancer screening and family history, mental health, social environment, neighborhood/environmental measures, socio-demographics, and access to health care. Height and weight were also measured. 1266 (91%) participants provided an optional saliva sample to be banked for future genetic analyses. Participants receive compensation, quarterly newsletters, and a summary of their current fruit and vegetable intake, physical activity levels, and weight status along with current medical recommendations for those assessments. Project CHURCH achieved a very high retention rate; 93% and 90% have returned to complete their Year 1 and Year 2 surveys, respectively.

The Southern Community Cohort Study (SCCS): The SCCS is described above. Included in this study are 224 incident African American lung cancer cases and a matched stratified random sample of 448 African American cohort members without lung cancer at the index date selected by incidence density sampling. Height and weight self-reported for the majority of study participants. For those who had clinic appointments on the day of interview, weight and height were abstracted from their medical record for that day and were found to be highly correlated (>.95) with self-reported values. For approximately 10% of those enrolled, the study interviewers measured height, weight and waist and hip circumference.

A Genome-Wide Association Study of Breast Cancer in the African Diaspora (ROOT): The following studies are included in the ROOT consortium.

The Nigerian Breast Cancer Study (NBCS): The NBCS is an ongoing case-control study of breast cancer in Ibadan, Nigeria initiated in 1998[73,74]. Breast cancer cases were 20 years or older, ascertained at the University College Hospital, Ibadan, which is the oldest tertiary hospital in Nigerian with a catchment

Nature Genetics: doi:10.1038/ng.2608

population of approximate three million. Controls were recruited from a randomly selected community in one of the communities adjoining the hospital. The majority of the study subjects were Yoruba and Yoruba is one of the populations selected by the International HapMap Project to represent African continent. Included in this study were 692 cases and 621 controls recruited between 1998 and 2009.

The Barbados National Cancer Study (BNCS): BNCS is a population-based case-control study designed to evaluate risk factors for incident breast and prostate cancer in the predominantly African population of Barbados, West Indies[75]. Cases were identified through the only pathology department on the island, located at the Queen Elizabeth Hospital, and represented all histologically confirmed incident cases of breast cancers between July 2002 and March 2006. Controls were selected from a national database provided by the Barbados Statistical Services Department, and were frequency matched to breast cancer cases at a 2:1 ratio and by 5-year age groups. Genotypes were conducted from 93 cases and 244 controls.

The Racial Variability in Genotypic Determinants of Breast Cancer Risk Study (RVGBC): RVGBC is a hospital-based genetic epidemiologic study conducted in Philadelphia and Detroit metropolitan areas from 1999 to 2003. Breast cancer cases were identified in the University of Pennsylvania Health System and Karmanos Cancer Institute. Local advertisement was also put to recruit breast cancer cases living in the Philadelphia and Detroit area. Controls were recruited in the same fashion as cases in these institutions except that they do not have breast cancer. Patients with breast cancer had to be diagnosed within 18 months of recruitment and have invasive ductal cancer. The study was designed to over-represent women diagnosed under age 40. The RVGBC contributed 151 African American cases and 272 African American controls.

The Baltimore Breast Cancer Study (BBCS): BBCS is a case control study of breast cancer designed to identify and characterize markers of disease aggressiveness and poor outcome. Incident breast cancer cases and controls were recruited between February of 1993 and August of 2003 in six hospitals in the greater Baltimore area, including the University of Maryland Medical Center, the Baltimore Veterans Affairs Medical Center, Union Memorial Hospital, Mercy Medical Center, and the Sinai Hospital. Controls were frequency-matched to cases by race and age. A total of 117 African Americans incident cases and 111 African Americans controls were included in this study.

The Chicago Cancer Prone Study (CCPS): CCPS is an ongoing hospital-based case-control study designed to investigate the genetics of young-onset breast cancer. Cases with histologically confirmed breast cancer were enrolled through the Cancer Risk Clinic at the University of Chicago. Young-onset cases and African Americans were oversampled. Controls were gender- and age-matched with cases and

Nature Genetics: doi:10.1038/ng.2608

enrolled from patients who visited the same hospital and were willing to donate blood for genetic studies. CCPS contributed 268 cases and 261 controls to this study.

The Southern Community Cohort (SCCS): The SCCS is a prospective cohort study initiated in 2002 focusing on investigating racial disparities in cancer risk[57]. A nested case-control study of breast cancer is included in this project. Cases were 215 women who were diagnosed with breast cancer after the entered the cohort. Control subjects (n =424) were selected randomly from those who were cancer-free at the time of the study and frequency-matched to case patients in a 1 to 2 ratio on age at enrollment , recruitment method, and sample type (blood/buccal cell).

SAPPHIRE: The Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE) is an ongoing prospective cohort study to determine the genetic determinants of inhaled corticosteroid response and other asthma-related phenotypes[76]. Participants were age 12-56 years at the time of enrollment, had at least one physician diagnosis of asthma, and had no prior diagnosis (in the electronic medical record or by patient report) of chronic obstructive pulmonary disease (COPD) or congestive heart failure (CHF). Controls were recruited from the same geographic area and were also age 12-56 years at the time of enrollment, but these individuals had no prior diagnosis of asthma, COPD, or CHF. All participants received care from a single, large health system covering southeast Michigan and metropolitan Detroit. At enrollment, all participants underwent an evaluation which included a staff-administered survey, lung function testing, and measurement of vital signs, height, and weight. Genomic DNA was isolated from a blood specimen obtained at the initial visit, and genotyping was performed using the Affymetrix Axiom Genome-Wide Population-Optimized Human Array.

SIGNET: The Sea Islands Genetics Network (SIGNET) study consists of the REasons for Geographic And Racial Differences in Stroke (REGARDS), the Sea Islands Genetic African American Registry (Project SuGAR), the COBRE for Oral Health project, and the Systemic Lupus Erythematosus in Gullah Health study (SLEIGH). All subjects are African Americans (AA), and all provided written informed consent.

All SIGNET samples (n= 4,298) were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Imputation was performed using MACH (version 1.0.16) to impute all autosomal SNPs using the CEU+YRI reference panel (as supplied by Goncalo Abecasis) from build 36 (2,318,207 SNPs in total).

REGARDS is an observational cohort of 30,239 AA and white men and women enrolled in their homes after a telephone interview in 2003-7[24]. Participants were a national sample oversampled from the southeastern stroke belt (56%) and were 58% female and 42% black by design. Participants were

Nature Genetics: doi:10.1038/ng.2608

followed every 6 months by telephone to ascertain health outcomes, with validation of stroke, coronary heart disease, death and other ancillary study endpoints. For SIGNET, we selected all AA REGARDS type 2 diabetes (T2D) cases recruited from SC, GA, NC, and AL, and an equivalent number of race, sex, and age-strata matched diabetes-free controls. We also included all participants not already included that were current residents of the 15-county "Low Country" region of SC and GA (SC counties Beaufort, Berkeley, Charleston, Colleton, Dorchester, Georgetown, Hampton, Horry, Jasper; GA counties Bryan, Camden, Chatham, Glynn, Liberty, McIntosh). The subset of REGARDS participants genotyped under SIGNET are referred to as SIGNET-REGARDS. GWAS genotyping was completed among 2398 SIGNET-REGARDS AA participants, including 1149 with diabetes and 1249 without diabetes.

Project SuGAR, COBRE and SLEIGH participants combined are referred to as SIGNET-Sea Islands. Project SuGAR (Sea Islands Genetic African American Registry) enrolled patients with type 2 diabetes (T2D) in AA families with multiple affected members living on the Sea Islands[25,27]. Inclusion criteria included at least one affected sibling pair, no more than one parent with T2D, and at least one parent alive. All consenting members of families meeting these criteria were enrolled. Medical, anthropometric, family and medical history, physical examination and laboratory testing were obtained. GWAS genotyping was completed in 1,176 participants; 967 with diabetes and 193 without. All Project SuGAR participants were genotyped under SIGNET.

The COBRE (Center of Biomedical Research Excellence) Oral Health pilot project, "An Epidemiological Study of Periodontal Disease and Diabetes: Cytokine Genes and Inflammation Factors for Oral Health" enrolled 226 AA participants with T2D ≥18 years old and not edentulous[26]. Participants lived along coastal South Carolina or 30 miles inland. Subjects answered a detailed questionnaire focusing on medical and dental history and underwent an oral examination to document periodontal health. All COBRE participants were included in SIGNET.

SLEIGH (Systemic lupus erythematosus in Gullah Health) is a population based case-control study of risk factors for systemic lupus erythematosus (SLE). Inclusion criteria were: 1) age 2 years or older, 2) self-identification as AA Gullah from the Sea Islands of South Carolina, with no known ancestors who were not of Gullah lineage, 3) at least 4 of the 11 American College of Rheumatology classification criteria for SLE, 4) and being able to speak and understand English. First-degree relatives of SLE probands were invited to enroll. Healthy AA control subjects without evidence of autoimmune or connective tissue disease, or of family history of SLE, from the Sea Islands were age- and sex-matched to cases. Under SIGNET, genotype data was generated in 93 SLEIGH non-SLE controls, including 15 with diabetes and 77 without.

Nature Genetics: doi:10.1038/ng.2608

The Women's Health Initiative – GARNET Study (WHI-GARNET): WHI-GARNET (WHI-Genomics and Randomized Clinical Network) sample is composed of individuals from the study participants from the WHI randomized, placebo-controlled clinical trial of hormone therapy in post-menopausal women, funded by NHGRI_ENREF_42. The WHI hormone trial enrolled women aged 50-79 years between 1993-1998, primarily by population-based direct mail strategies, at 40 clinical centers in 24 states and the District of Columbia in the U.S. Details of the study design, recruitment, data collection methods, follow-up, intensive validation and tabulations of baseline data have been published in detail[29]. Genome-wide genotyping for WHI-GARNET was performed at the Broad Institute using the Human Omni 1M Quad v1_B SNP array. Samples were excluded from the dataset for the reasons of genotyping failure, genotypic sex mismatch, and first degree relative of an included individual based on genotype data. A total of 942,499 high-quality SNPs passed QC filters. Genotype imputation was performed at the GARNET Data Coordinating Center (University of Washington) using BEAGLE software and the European continental reference panels selected from the 1000 Genomes Project. Uncertainty in genotype prediction was accounted for by utilizing the dosage information from BEAGLE. Genotyping data were available on 1,803 Caucasian participants with SHBG measurements. For the SHBG assay, Fasting blood samples were collected from each study participant prior to entry into WHI. Women who were using hormones at initial contact completed a 3-month washout period prior to the baseline blood collection. Blood was drawn following at least a 12-hour fast, centrifuged at 1300*g* for 10 minutes, and the separated sera were stored in aliquots at −70°C within 2 hours of blood collection. Serum SHBG was assayed at the Reproductive Endocrine Research Laboratory (University of Southern California, Los Angeles, CA, USA), via a solid-phase, two-site chemiluminescent immunoassay using the Immulite analyzer (Siemens Medical Solutions, Malvern, PA, USA) that consists of polystyrene beads coupled to a mouse monoclonal antibody specific for SHBG (Siemens Medical Solutions, USA). The SHGB assay had a sensitivity of 0.2 nmol/L. The intra-assay CVs for the SHBG assay were 2.5% at 21 nmol/L, 2.7% at 63 nmol/L, and 5.3% at 80 nmol/L. The interassay CVs were 5.2% at 21 nmol/L, 5.2% at 63 nmol/L, and 6.6% at 80 nmol/L.

The Women's Health Initiative – PAGE Study (WHI-PAGE): The WHI study was previously described above. We genotyped the African Americans samples from WHI that were included as part of the Population Architecture Using Genomics and Epidemiology (PAGE) consortium. A subset of 3,262 WHI women were selected for genotyping and inclusion in these PAGE analyses. *Data Collection:* BMI was calculated from measured weight and height at time of enrollment. Race/ethnicity was self-reported as black. Principal components were derived from genome-wide markers to control for global ancestry and identify ancestry outliers for removal. *Genotyping:* The three novel variants SNPs were genotyped at the Translational Genomics Research Institute (TGen) (Phoenix, AZ) on TaqMan. Study protocol

Nature Genetics: doi:10.1038/ng.2608

included calculation of concordance rates among duplicates and other extensive QA procedures. All SNPs passed QA with individual and SNP call rates exceeding 95% and 97%, respectively.

Description of Studies used for Look-ups

Pediatric Research Consortium (PeRC): All subjects were consecutively recruited from the Greater Philadelphia area from 2006 to 2010 at the Children's Hospital of Philadelphia. Our study cohort consisted of 3,751 children of African ancestry. All of these participants had their blood drawn in to an 8ml EDTA blood collection tube and were subsequently DNA extracted for genotyping. All subjects were biologically unrelated and were aged between 2 and 18 years old. This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia. Parental informed consent was given for each study participant for both the blood collection and subsequent genotyping. Self-reported ethnicity was confirmed by multidimensional scaling methodologies. BMI z-score was defined using the Center for Disease control (CDC) growth curves (http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/datafiles.htm). We performed high throughput genome-wide SNP genotyping, using the Illumina Infinium™ II HumanHap550 or 610 BeadChip technology (Illumina, San Diego), at the Center for Applied Genomics at CHOP. We used 750ng of genomic DNA to genotype each sample, according to the manufacturer's guidelines.

Summary of Literature Search on Genes nearest the Two Novel Loci and One Suggestive Locus

We utilized SNIPPER (http://csg.sph.umich.edu/boehnke/snipper/) and SNAP to derive potential biological links of genes in the proximity of the novel association signals (500kb of the index SNP), and present a summary in this section.

1. **rs7708584** ***GALNT10:*** A total of 3 genes are found within 500 kb of the lead marker rs7708584, located 5q33.2. Approximately 26.8 kb upstream of this SNP is the *GALNT10* gene, which belongs to the polypeptide N-acetylgalactosaminyltransferase (pp-GalNAc-T) gene family. The GALNT10 protein catalyzes the first step in the synthesis of mucin-type *O*-glycosylation, which is a common post-translational modification of secreted and membrane-associated proteins[77].

Nature Genetics: doi:10.1038/ng.2608

The *GALNT10* protein is expressed at high level in the small intestine, and at intermediate levels in the stomach, pancreas, ovary, thyroid gland and spleen[78]. Expression of Galnt10 was reported in several distinct hypothalamic, thalamic and amygdaloid nuclei in the mouse central nervous system using in situ hybridization analysis[79]. A suggestive association (p=5.57E10-6) on 5q32 was previously reported with BMI in African Americans [28]. However the index SNP rs2033195 is only modestly correlated with the present SNP rs7708584 (r^2= 0.269 using YRI from HapMap). This may suggest that independent loci on 5q33 chromosome region influence the levels of BMI, or that the large sample from our study helped to narrow down the association on 5q33 with BMI. The other two nearest genes in the region are *MFAP3* and *FAM114A2*. *MFAP3* (microfibrillar-associated protein 3, 5q32-q33.2) is 106 kb from our index signal and encodes a protein involved in the component of the elastin-associated microfibrils. Located 125 kb upstream is *FAM114A2* (family with sequence similarity 114, member A2, 5q31-q33). However, these genes have not been reported to be associated with obesity-adiposity traits or involved in the biological pathways of these traits. A GWA study has reported an association with ventricular conduction in the 1 MB region of rs7708584 (PMID: 21076409) although this signal appears to be independent of our lead SNP signal ($r^2 < 0.2$).

2. **rs974417 *KLHL32*:** SNP rs974417 is intronic of *KLHL32* (kelch-like 32) at 6q16.1. The kelch-like proteins are a large group, with approximately 71 kelch-repeat proteins encoded in the human genome[80]. The exact function of *KLHL32* protein is unknown. However, other kelch proteins have been shown to bind substrate protein in ubiquitination, which has diverse functions, including targeting proteins for proteasomal degradation as well as non-degradative roles such as modulation of protein activity, interaction and localization[81]. Mutations in KLHL9 have been associated with early onset autosomal dominant distal myopathy and skeletal muscle atrophy[82], and hypertension[81]. Another plausible biological candidate within the association signal region is the *GPR63* (G protein-coupled receptor 63 gene, 6q16.1-q16.3) which is 134 kb upstream of rs10261878. *GPR63* may play a role in brain function. Its expression has been detected in the frontal cortex, with lower levels in the thalamus, caudate, hypothalamus and midbrain. The SNP rs12200560 nearby this locus was recently nominally associated with coronary artery disease in a European cohort[83]. This region has also been implicated in linkage studies of atrial fibrillation and dilated cardiomyopathy[84,85].

3. **rs10261878 *MIR148A / NFE2L3*:** Approximately 39 kb upstream of the index SNP is the *MIR148A* (microRNA 148a gene). MicroRNAs (miRNAs) are short (20-24 nt) non-coding

Nature Genetics: doi:10.1038/ng.2608

RNAs highly conserved evolutionarily. MiRNAs are involved in post-transcriptional regulation of gene expression in multicellular organisms by affecting both the stability and translation of mRNAs. Several microRNAs, including 148a, have been shown to be significantly up-regulated during differentiation *in vitro* adipogenesis of 3T3-L1 preadipocytes[86] as well as in human adipocytes[87]. It was noted that there was an inverse correlation of miRNA expression during adipogenesis in obese mice such that miR-148a was induced during adipogenesis but was downregulated in obese adipocytes[86]. Furthermore, human miR-148a has been associated with panic disorder and regulates an anxiety candidate gene *CCKBR* (cholecystokinin B receptor) which may play a regulatory role in the control of food intake[88]. In addition, human miR-148a exerts various biological functions in tumors by targeting oncogenes or tumor suppressors. MiR-148a has been associated and/or involved in esophageal carcinoma[89], colorectal cancer[90] , prostate cancer[91], ovarian cancer[92], melanoma cells[93], gastric cancer[94] and HIV control[95], among other diseases. About 241 kb upstream of SNP rs10261878 is *NFE2L3* (nuclear factor (erythroid-derived 2)-like 3) which is also a strong biological candidate. This gene is a member of the cap "n" collar subfamily of basic-region leucine zipper transcription factors and is presumably involved in the regulation of antioxidant and detoxification genes, such as those for the glutamate cysteine ligase catalytic and modifier subunits, glutathione S-transferase (GST) isoenzymes and NAD(P)H:quinone oxidoreductase 1 (NQO1)[96]. Genomewide association studies have identified loci at 7p15.2, near to *MIR148A and NFE2L3,* associated with waist-hip ratio (rs1055144)[97], endometriosis (rs12700667)[98], type 1 diabetes (rs7804356)[99], and smoking behavior (rs886716 and rs4722613)[100]. However, these loci have low correlations with the locus detected in our study (rs10261878, r^2 <0.1), which may suggest independent association signals.

Nature Genetics: doi:10.1038/ng.2608

Acknowledgements

STAGE 1:

AABC-MEC: The MEC and genotyping of samples for the GWAS of breast and prostate cancer was supported by National Institutes of Health grants CA63464, CA54281, CA164973, CA1326792, CA148085, HG004726 and a Department of Defense Breast Cancer Research Program Era of Hope Scholar Award to CAH (W81XWH-08-1-0383) and the Norris Foundation.

AABC-CARE: CARE was supported by National Institute for Child Health and Development contract NO1-HD-3-3175.

AABC-WCHS: WCHS is supported by a U.S. Army Medical Research and Material Command (USAMRMC) grant DAMD-17-01-0-0334, National Institutes of Health grant CA100598, and the Breast Cancer Research Foundation

AABC-SFBCS: SFBCS was supported by National Institutes of Health grant CA77305 and United States Army Medical Research Program grant DAMD17-96-6071.

AABC-NC-BCFR: The Breast Cancer Family Registry (BCFR) was supported by the National Cancer Institute, National Institutes of Health under RFA CA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry and Principal Investigators, including the Cancer Prevention Institute of California (U01 CA69417). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the BCFR.

AABC-CBCS: CBCS is supported by National Institutes of Health Specialized Program of Research Excellence in Breast Cancer grant CA58223 and Center for Environmental Health and Susceptibility, National Institute of Environmental Health Sciences, National Institutes of Health, grant ES10126.

AABC-PLCO: Genotyping of the PLCO samples was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, the National Institutes of Health. The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the screening center investigators and staff of the PLCO Cancer Screening Trial, Mr. Thomas Riley and staff at Information Management Services, Inc., and Ms. Barbara O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial.

Nature Genetics: doi:10.1038/ng.2608

AABC-NBHS: NBHS is support by National Institutes of Health grant CA100374. NBHS sample preparation was conducted at the Biospecimen Core Lab that is supported in part by the Vanderbilt-Ingram Cancer Center (CA68485).

AABC-WFBC: WFBC is supported by National Institutes of Health grant CA73629.

AAPC1-CPS-II: CPSII is supported by the American Cancer Society.

AAPC1-KCPCS: KCPCS was supported by National Institutes of Health grants CA056678, CA082664 and CA092579, with additional support from the Fred Hutchinson Cancer Research Center.

AAPC1-LAAPC: LAAPC was funded by grant 99-00524V-10258 from the Cancer Research Fund, under Interagency Agreement #97-12013 (University of California contract #98-00924V) with the Department of Health Services Cancer Research Program.

AAPC1-DCPC: DCPC was supported by National Institutes of Health grant S06GM08016 and Department of Defense grants DAMD W81XWH-07-1-0203 and DAMD W81XWH-06-1-0066.

AAPC1-MEC: The MEC and genotyping of samples for the GWAS of prostate cancer was supported by National Institutes of Health grants CA63464, CA54281, CA164973, CA1326792, CA148085 and HG004726.

AAPC1-PLCO: See AABC-PLCO

AAPC1-GECAP: GECAP was supported by National Institutes of Health grant ES011126.

AAPC1-SCCS: SCCS is funded by National Institutes of Health grant CA092447. SCCS sample preparation was conducted at the Biospecimen Core Lab that is supported in part by the Vanderbilt-Ingram Cancer Center (CA68485).

AAPC1-MDA: MDA was support by grants CA68578, ES007784, DAMD W81XWH-07-1-0645, and CA140388

AAPC1-CaP Genes: CaP Genes was supported by CA88164 and CA127298.

BIO-VU: The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant 1UL1RR024975-01 from NCRR and the National Institutes of Health.

CARE: CARE was supported through a contract with National Institute of Child Health and Human Development (NO1-HD-3-3175), by National Heart, Lung, and Blood Institute contract HHSN26800625226C, and by National Institute of Diabetes and Digestive and Kidney Diseases contract R01DK075787.

ARIC: ARIC is a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN26820110005C, HHSN26820110006Cm HHSN26820110007C, HHSN26820110008C, HHSN26820110009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, and HHSN26800625226C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contribution. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and the National Institutes of Health Roadmap for Medical Research.

CARDIA: CARDIA was supported by grants to the following institutions: University of Alabama at Birmingham (N01-HC-48047), University of Minnesota (N01-HC-48048), Northwestern University (N01-HC-48049), Kaiser Foundation Research Institute (N01-HC-48050), University of Alabama at Birmingham (N01-HC-95095), Tufts-New England Medical Center (N01-HC-45204), Wake Forest University (N01-HC-45205), Harbor-UCLA Research and Education Institute (N01-HC-05187), University of California, Irvine (N01-HC-45134, N01-HC-95100), and by National Heart, Lung, and Blood Institute contract HHSN26800625226C.

CFS: CFS was support by a grant to Case Western Reserve University (National Institutes of Health grants HL 46380 and M01RR00080)

JHS: JHS was supported by a grant to Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), and Tougaloo College (N01-HC-95172), and by National Heart, Lung, and Blood Institute contract HHSN26800625226C.

MESA/MESA Air/MESA Family: MESA, MESA Family, and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159 through N01-HC-95169 and UL1-RR-024156. Funding for MESA Family is provided by grants R01-HL-071051, R01-HL-071205, R01-HL-071250, R01-HL-071251, R01-HL-071252, R01-HL-071258, R01-HL-071259,UL1-RR-025005. MESA Air is funded by the US EPA - Science to Achieve Results (STAR) Program Grant #RD831697. Funding for genotyping was provided by National Heart, Lung, and Blood Institute contracts N02-HL-6-4278, N01-HC-65226, and HHSN26800625226C.

CHS: CHS research was supported by National Heart, Lung, and Blood Institute contracts N01-HC-85239, N01-HC-85079, N01-HC-85080, N01-HC-85081, N01-HC-85082, N01-HC-85083, N01-HC-85084, N01-HC-85085, N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-

Nature Genetics: doi:10.1038/ng.2608

75150, N01-HC-45133, HHSN268201200036C and National Heart, Lung, and Blood Institute grants HL080295, HL085251, HL087652, HL105756, and HHSN26800625226C with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute of Aging. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

FamHS: The FamHS was supported by National Institutes of Health grants R01-HL-087700 and R01-HL-088215 (Michael A. Province, PI) from the National Heart, Lung, and Blood Institute; and R01-DK-8925601 and R01-DK-075681 (Ingrid B. Borecki, PI) from the National Institute of Diabetes and Digestive and Kidney Disease.

GENOA: Genetic Epidemiology Network of Arteriopathy (GENOA) study is supported by the National Institutes of Health, grant numbers HL087660, HL101161, and HL100245 from National Heart, Lung, Blood Institute. We thank Eric Boerwinkle, PhD from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA and Julie Cunningham, PhD from the Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA for their help with genotyping.

Health ABC: This research was supported by the National Institute of Aging contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by the National Institute of Aging grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging.

HANDLS: This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol # 2009-149). Data analyses for the HANDLS study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

HUFS: The study was supported by grants S06GM008016-320107 to CR and S06GM008016-380111 to AA, both from the NIGMS/MBRS/SCORE Program. Participant enrollment was carried out at the Howard University General Clinical Research Center (GCRC), which is supported by grant number 2M01RR010284 from the National Center for Research Resources, a component of the National Institutes of Health. Additional support was provided by the Coriell Institute for Biomedical Sciences. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, in the Center for Research in Genomics and Global Health (Z01HG200362). The authors would like to thank participants in the Howard University Family Study (HUFS) and the HUFS field staff. The role of the Howard University General Clinical Research Center (GCRC) in participant enrollment is hereby acknowledged.

HyperGen: HyperGen is supported by National Heart, Lung, and Blood Institute R01 HL55673 and cooperative agreements (U10) with the National Heart, Lung, and Blood Institute: HL54471, HL54515 (UT); HL54472, HL54496 (MN); HL54473 (DCC); HL54495 (AL); HL54509 (NC).

Maywood: Maywood cohort study was supported by the National Institutes of Health grant numbers R37-HL045508, R01-HL074166, R01-HL086718 and R01-HG003054.

Nigeria: The Nigerian cohort study was supported by National Institutes of Health grant numbers R37-HL045508, R01-HL053353, R01-DK075787 and U01-HL054512. The authors acknowledge the assistance of the research staff and participants in Igbo-Ora, Oyo State, Nigeria.

WFSM: Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSC268200782096C. This work was supported by National Institutes of Health grants R01 DK087914, R01 DK066358, R01 DK053591 and by the Wake Forest School of Medicine grant M01 RR07122 and Venture Fund.

WHI-SHARe: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. This manuscript was prepared in collaboration with investigators of the WHI, and has been reviewed and/or approved by the Women's Health Initiative (WHI). WHI investigators are listed at https://cleo.whi.org/researchers/SitePages/WHI%20Investigators.aspx. Funding for WHI SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The datasets used for the analyses

Nature Genetics: doi:10.1038/ng.2608

described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000386.v3.p2.

STAGE 2 and STAGE 3 STUDIES:

AAPC2-MEC: See AABC-MEC

AAPC2-SCCS: See AAPC1-SCCS

AAPC2-PCBP: PCBP was supported by a contract from the National Human Genome Research Institute (N01HG25487) and a grant from the National Cancer Institute (R01CA114379).

AAPC2-SELECT: SELECT is coordinated by SWOG (formerly the Southwest Oncology Group) and funded by Public Health Service cooperative Agreement grant CA37429 awarded by the National Cancer Institute, National Institute of Health, and in part by the National Center for Complementary and Alternative Medicine.

BIO-VU: The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant 1UL1RR024975-01 from NCRR and the National Institutes of Health.

BWHS: BWHS research was supported by R01 CA098663 and R01 CA058420 from the National Cancer Institute and by a grant from the Susan G. Komen for the Cure Foundation.

GeneSTAR: GeneSTAR was supported by grants HL072518, HL087698-01, HL58625-01A1, HL59684, and HL071025-01A1 from the National Heart, Lung, and Blood Institute, grant NR0224103 from the National Institue of Nursing Research, and M01-RR00052 from the National Center for Research Resources, National Institutes of Health, Bethesda, MD.

Ghana: The Ghana study was supported by grant LM010098 from the National Library of Medicine.

IPM: The Mount Sinai Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

JoCo: The Johnston County Osteoarthritis Project is supported in part by cooperative agreements S043, S1734, and S3486 from the Centers for Disease Control and Prevention/Association of Schools of Public Health; the NIAMS Multipurpose Arthritis and Musculoskeletal Disease Center grant 5-P60-AR30701; and the NIAMS Multidisciplinary Clinical Research Center grant 5 P60 AR49465-03. Genotyping services were provided by Algynomics company.

NCI Lung, MDA: Supported by CA121197 and S2, CA133996, CA141716, CA148127, CA70907, CA55769, CA127219, CA111646, Personalized Risk Prediction Program, Duncan Family Institute, Cancer Prevention and Research Institute of Texas RP100443.

NCI Lung, NCI-MD: Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research

NCI Lung, Northern California Lung Cancer Study: Support by a grant from the National Institute of Environmental Health Sciences (ES06717).

NCI Lung, Project CHURCH: University Cancer Foundation and the Duncan Family Institute Center for Community, Implementation and Dissemination Research

Pediatric Research Consortium (PeRC): We thank the network of primary care clinicians, their patients and families for their contribution to this project and clinical research facilitated through the Pediatric Research Consortium (PeRC) at The Children's Hospital of Philadelphia. Rosetta Chiavacci, Elvira Dabaghyan, Hope Thomas, Kisha Harden, Andrew Hill, Kenya Fain, Crystal Johnson-Honesty, Cynthia Drummond, Shanell Harrison and Sarah Wildrick, Cecilia Kim, Frank Mentch, Edward Frackelton, George Otieno, Kelly Thomas, Cuiping Hou, Kelly Thomas and Maria L. Garris provided expert assistance with genotyping or data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnssonof Raförninn ehf for their extensive software design and contribution. This research was financially supported by an Institute Development Award from the Children's Hospital of Philadelphia, a Research Development Award from the Cotswold Foundation and NIH grant R01 HD056465.

ROOT: Support was provided by National Cancer Institute grants (R01 CA142996, R01 CA89085, R01 CA141712, P01 CA82707, and P50 CA125183), the Noreen Fraser Foundation, the Breast Cancer Research Foundation, and the American Cancer Society (MRSG-13-063-01-TBG)

SAPPHIRE: Support for this study was provided by the Strategic Program for Asthma Research from the American Asthma Foundation, the Fund for Henry Ford Hospital, and the National Institute of Allergy and Infectious Diseases (R01AI079139, R01AI061774), National Heart Lung and Blood Institute (R01HL079055), and National Institute of Diabetes and Digestive and Kidney Diseases (R01DK064695), NIH to Dr. Williams.

SIGNET: The Sea Islands Genetics Network (SIGNET) is supported by R01 DK084350 (M Sale), and consists of data from the REasons for Geographic And Racial Differences in Stroke (REGARDS) cohort (U01 NS041588, G Howard), Project SuGAR (Sea Islands Genetic African American Registry, W.M. Keck Foundation, WT Garvey), the Systemic Lupus Erythematosus in Gullah Health (SLEIGH) study

Nature Genetics: doi:10.1038/ng.2608

(PI: G Gilkeson; K23 AR052364, D. Kamen; UL1 RR029882), and a South Carolina Center of Biomedical Research Excellence (COBRE) in Oral Health Project P20 RR017696 (PI: Kirkwood; Sub-award: JK Fernandes).

WHI-GARNET: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. This manuscript was prepared in collaboration with investigators of the WHI, and has been reviewed and/or approved by the Women's Health Initiative (WHI). WHI investigators are listed at https://cleo.whi.org/researchers/SitePages/WHI%20Investigators.aspx. Funding support for WHI GARNET was provided through the NHGRI Genomics and Randomized Trials Network (GARNET) (Grant Number U01 HG005152). Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GARNET Coordinating Center (U01 HG005157). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004424). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000315.v3.p2.

WHI-PAGE: The WHI-PAGE component is supported by NHGRI grant U01 HG004790.

EKS is supported by K23DK080145, the Doris Duke Charitable Foundation (Grant 2012067), and the Department of Internal Medicine at the University of Michigan.

TWW is supported by grant NGFNPLUS 01GS0823.

JHM is supported by NIH grants R01 LM010098 and HL65234.

AGEN Consortium Author Acknowledgements

Last	First	Affiliation
Adair	Linda S.	Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA.
Aung	Tin	Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.
Cai	Qiuyin	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Chang	Li-Ching	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.
Chen	Chien-Hsiun	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; School of Post-Baccalaureate Chinese Medicine, China Medical University, Taichung, Taiwan.
Chen	Gary K.	Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA.
Chen	Yuan-Tsong	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; Department of Pediatrics, Duke University Medical Center, Durham, NC, USA
Cho	Yoon Shin	Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Republic of Korea.
Delahanty	Ryan J.	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Dorajoo	Rajkumar	Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, United Kingdom.
Fann	Cathy S.J.	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.
Gao	Yu-Tang	Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China.
Go	Min Jin	Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Republic of Korea.
Gu	Dongfeng	Cardiovascular Institute and Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.
Haiman	Christopher A.	Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA.
Han	Bok-Ghee	Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Republic of Korea.

Nature Genetics: doi:10.1038/ng.2608

AGEN Consortium Author Acknowledgements, continued

Last	First	Affiliation
Hao	Mei	Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA.
He	Jiang	Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA.
He	Meian	MOE Key Lab of Environment and Health, School of Public Heath, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, 430030, Hubei, China.
Henderson	Brian E.	Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA.
Hu	Yanling	Medical Scientific Research Center, Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, China.
Hu	Frank B.	Departments of Epidemiology and Nutrition, Harvard University School of Public Health, Boston, Massachusetts, USA.
Iwai	Naoharu	Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Suita, Japan.
Kamatani	Naoyuki	Laboratory for Statistical Analysis, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan.
Kato	Norihiro	Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan.
Kokubo	Yoshihiro	Department of Preventive Cardiology, National Cerebral and Cardiovascular Center, Suita, Japan.
Kubo	Michiaki	Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan.
Le Marchand	Loic	Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA.
Lee	Jong-Young	Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Republic of Korea.
Lee	Nanette R.	Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines.
Li	Chun	Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Li	Haixing	Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 294 Tai-Yuan Road, Shanghai, 200031, People's Republic of China
Liang	Jun	Department of Endocrinology, the Central Hospital of Xuzhou, Affiliated Hospital of Southeast University, Xuzhou, Jiangsu, China.
Lin	Xu	Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 294 Tai-Yuan Road, Shanghai, 200031, People's Republic of China
Liu	Jian Jun	Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.
Long	Jirong	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center,

Nature Genetics: doi:10.1038/ng.2608

		Vanderbilt University School of Medicine, Nashville, Tennessee, USA.

AGEN Consortium Author Acknowledgements, continued

Last	**First**	**Affiliation**
Lu	Wei	Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China.
Maeda	Shiro	Laboratory for Endocrinology and Metabolism, CGM, RIKEN, Yokohama, Japan.
Miki	Tetsuro	Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, Japan.
Mo	Zengnan	Institute of Urology and Nephrology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, China.
Mohlke	Karen L.	Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
Nakamura	Yusuke	Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.
Ng	Daniel P.-K.	Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore.
Okada	Yukinori	Laboratory for Statistical Analysis, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan; Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.
Ong	Rick T.H.	Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore.
Qi	Lu	Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA.
Saw	Seang Mei	Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore.
Seielstad	Mark	Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; Department of Genome Science, Aichi-Gakuin University, School of Dentistry, Nagoya, Japan.
Shi	Jiajun	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Shu	Xiao Ou	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Sim	Xueling	Centre for Molecular Epidemiology, National University of Singapore, Singapore, Singapore.
Song	Huaidong	Ruijin Hospital, State Key Laboratory of Medical Genomics, Molecular Medical Center, Shanghai Institute of Endocrinology, Shanghai Jiao Tong University School of Medicine, Shanghai, China.
Su	Mei-Hsin	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.
Tabara	Yasuharu	Department of Basic Medical Research and Education, Ehime

Nature Genetics: doi:10.1038/ng.2608

		University Graduate School of Medicine, Toon, Japan.

Nature Genetics: doi:10.1038/ng.2608

AGEN Consortium Author Acknowledgements, continued

Last	First	Affiliation
Tai	E. Shyong	Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; Laboratory for Cardiovascular Diseases, CGM, RIKEN, Yokohama, Japan; Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.
Takeuchi	Fumihiko	Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan.
Tan	Aihua	Center for Metabolic Disease and Diabetes, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, China.
Tanaka	Toshihiro	Laboratory for Cardiovascular Diseases, CGM, RIKEN, Yokohama, Japan.
Tsai	Fuu-Jen	School of Post-Baccalaureate Chinese Medicine, China Medical University, Taichung, Taiwan; Department of Pediatrics, Duke University Medical Center, Durham, NC, USA.
Tsunoda	Tatsuhiko	Laboratory for Medical Informatics, CGM, RIKEN, Yokohama, Japan.
Wen	Wanqing	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Wong	Tien Yin	Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore; Center for Eye Research Australia, University of Melbourne, East Melbourne, Australia.
Wu	Jer-Yuarn	Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; School of Post-Baccalaureate Chinese Medicine, China Medical University, Taichung, Taiwan.
Wu	Ying	Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
Wu	Tangchun	MOE Key Lab of Environment and Health, School of Public Heath, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, 430030, Hubei, China.
Xiang	Yong-Bing	Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China.
Xu	Jianfeng	Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC, USA.
Yamamoto	Ken	Division of Genome Analysis, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan.
Yang	Handong	Dongfeng Central Hospital, Dongfeng Motor Corporation and Hubei University of Medicine, Shiyan, 442008, Hubei, China.
Yokota	Mitsuhiro	Department of Genome Science, Aichi-Gakuin University, School of Dentistry, Nagoya, Japan.

Nature Genetics: doi:10.1038/ng.2608

AGEN Consortium Author Acknowledgements, continued

Last	First	Affiliation
Zhang	Yi	State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; Shanghai Institute of Hypertension, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; Sino-French Research Center for Life Science and Genomics, Shanghai, China.
Zheng	Wei	Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
Zhu	Dingliang	State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; Shanghai Institute of Hypertension, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; Sino-French Research Center for Life Science and Genomics, Shanghai, China; Shanghai Key Laboratory of Vascular Biology, Shanghai, China.
Zhu	Jiang	Dongfeng Central Hospital, Dongfeng Motor Corporation and Hubei University of Medicine, Shiyan, 442008, Hubei, China.

Nature Genetics: doi:10.1038/ng.2608

BioBank Japan Project Author Acknowledgements

Last	First	Affiliation
Okada	Yukinori	Laboratory for Statistical Analysis, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan; Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.
Kubo	Michiaki	Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan.
Takahashi	Atsushi	Laboratory for Statistical Analysis, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan.
Nakamura	Yusuke	Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.
Tanaka	Toshihiro	Laboratory for Cardiovascular Diseases, CGM, RIKEN, Yokohama, Japan.

Nature Genetics: doi:10.1038/ng.2608

The North American Brain Consortium (NABEC) and The UK Human Brain Expression Consortium (UKBEC) Author Acknowledgements

Last	First	Affiliation
Singleton	Andrew	Laboratory of Neurogenetics, National Institute on Aging,National Institutes of Health, Bethesda, MD, USA
Cookson	Mark	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Gibbs	Raphael	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Hernandez	Dena	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Dillman	Allissa	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
		Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden
Nalls	Michael	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Zonderman	Alan	Research Resources Branch, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Sampth	Arepalli	Research Resources Branch, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Ferrucci	Luigi	Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA
Johnson	Robert	NICHD Brain and Tissue Bank of Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland, USA
Longo	Dan	Lymphocyte Cell Biology Unit, Laboratory of Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA
O'Brien	Richard	Brain Resource Center, Johns Hopkins University, Baltimore, MD, USA
Traynor	Bryan	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Troncoso	Juan	Brain Resource Center, Johns Hopkins University, Baltimore, MD, USA
Van der Brug	Marcel	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; ITGR Biomarker Discovery Group, Genentech, South San Francisco, CA, USA
Zielke	Ronald	NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, MD,USA
Hardy	John	Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Weale	Michael	Department of Medical and Molecular Genetics, King's College London, 8th Floor, Tower Wing, Guy's Hospital, London SE1 9RT, UK

Nature Genetics: doi:10.1038/ng.2608

NABEC and UKBEC Author Acknowledgements, continued

Last	First	Affiliation
Ryten	Mina	Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Ramasamy	Adaikalavan	Department of Medical and Molecular Genetics, King's College London, 8th Floor, Tower Wing, Guy's Hospital, London SE1 9RT, UK; Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
Trabzuni	Daniah	Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK; Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
Smith	Colin	Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Wilkie Building, Teviot Place, Edinburgh EH8 9AG
Walker	Robert	Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Wilkie Building, Teviot Place, Edinburgh EH8 9AG

Nature Genetics: doi:10.1038/ng.2608

References

1. Chen, F. et al. Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet* **20**, 4491-503 (2011).
2. Haiman, C.A. et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* **43**, 570-3 (2011).
3. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
4. Roden, D.M. et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* **84**, 362-9 (2008).
5. Friedman, G.D. et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* **41**, 1105-16 (1988).
6. Tishler, P.V., Larkin, E.K., Schluchter, M.D. & Redline, S. Incidence of sleep-disordered breathing in an urban adult population: the relative importance of risk factors in the development of sleep-disordered breathing. *JAMA* **289**, 2230-7 (2003).
7. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
8. Heard-Costa, N.L. et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet* **5**, e1000539 (2009).
9. Higgins, M. et al. NHLBI Family Heart Study: objectives and design. *Am J Epidemiol* **143**, 1219-28 (1996).
10. Bordeaux, B.C. et al. Effect of obesity on platelet reactivity and response to low-dose aspirin. *Prev Cardiol* **13**, 56-62 (2010).
11. Mora, S. et al. Interaction of body mass index and framingham risk score in predicting incident coronary disease in families. *Circulation* **111**, 1871-6 (2005).
12. Daniels, P.R. et al. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med* **116**, 676-81 (2004).
13. Evans, M.K. et al. Healthy aging in neighborhoods of diversity across the life span (HANDLS): overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethn Dis* **20**, 267-75 (2010).
14. Adeyemo, A. et al. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* **5**, e1000564 (2009).
15. Arnett, D.K. et al. Genetic variation in NCAM1 contributes to left ventricular wall thickness in hypertensive families. *Circ Res* **108**, 279-83 (2011).
16. Tayo, B.O. et al. Genetic background of patients from a university medical center in Manhattan: implications for personalized medicine. *PLoS One* **6**, e19166 (2011).
17. Sempos, C.T., Bild, D.E. & Manolio, T.A. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. *Am J Med Sci* **317**, 142-6 (1999).
18. Jordan, J.M. et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. *J Rheumatol* **34**, 172-80 (2007).
19. Kang, S.J. et al. Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum Mol Genet* **19**, 2725-38 (2010).
20. Bild, D.E. et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-81 (2002).
21. Cohen, M.A. et al. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environ Sci Technol* **43**, 4687-93 (2009).

Nature Genetics: doi:10.1038/ng.2608

22. Huo, D. et al. Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* **33**, 835-40 (2012).
23. Torgerson, D.G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
24. Howard, V.J. et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* **25**, 135-43 (2005).
25. Sale, M.M. et al. Genome-wide linkage scan in Gullah-speaking African American families with type 2 diabetes: the Sea Islands Genetic African American Registry (Project SuGAR). *Diabetes* **58**, 260-7 (2009).
26. Bandyopadhyay, D., Marlow, N.M., Fernandes, J.K. & Leite, R.S. Periodontal disease progression and glycaemic control among Gullah African Americans with type-2 diabetes. *J Clin Periodontol* **37**, 501-9 (2010).
27. Divers, J. et al. The genetic architecture of lipoprotein subclasses in Gullah-speaking African American families enriched for type 2 diabetes: the Sea Islands Genetic African American Registry (Project SuGAR). *J Lipid Res* **51**, 586-97 (2010).
28. Ng, M.C. et al. Genome-wide association of BMI in African Americans. *Obesity (Silver Spring)* **20**, 622-7 (2012).
29. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* **19**, 61-109 (1998).
30. Speliotes, E.K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
31. Okada, Y. et al. Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* **44**, 302-6 (2012).
32. Wen, W. et al. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* **44**, 307-11 (2012).
33. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41**, 18-24 (2009).
34. Zhong, H., Yang, X., Kaplan, L.M., Molony, C. & Schadt, E.E. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* **86**, 581-91 (2010).
35. Nalls, M.A. et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641-9 (2011).
36. Myers, A.J. et al. A survey of genetic human cortical gene expression. *Nat Genet* **39**, 1494-9 (2007).
37. Bernstein, B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).
38. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
39. Newman, B. et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* **35**, 51-60 (1995).
40. Polly, A.M. et al. The NICHD Women's Contraceptive and Reproductive Experiences Study: Methods and Operational Results. *Annals of epidemiology* **12**, 213-221 (2002).
41. Kolonel, L.N. et al. A Multiethnic Cohort in Hawaii and Los Angeles: Baseline Characteristics. *American Journal of Epidemiology* **151**, 346-357 (2000).
42. Zheng, W. et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* **18**, 2761-4 (2009).

Nature Genetics: doi:10.1038/ng.2608

43. John, E.M. et al. Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA* **298**, 2869-76 (2007).
44. Prorok, P.C. et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* **21**, 273S-309S (2000).
45. John, E.M., Schwartz, G.G., Koo, J., Wang, W. & Ingles, S.A. Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population. *Am J Epidemiol* **166**, 1409-19 (2007).
46. Smith, T.R. et al. Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. *Carcinogenesis* **29**, 2132-8 (2008).
47. Ambrosone, C.B. et al. Conducting Molecular Epidemiological Research in the Age of HIPAA: A Multi-Institutional Case-Control Study of Breast Cancer in African-American and European-American Women. *J Oncol* **2009**, 871250 (2009).
48. Calle, E.E. et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**, 500-11 (2002).
49. Robbins et al. *Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus*, 6 (Cold Spring Harbor Laboratory Press, Woodbury, NY, ETATS-UNIS, 2007).
50. Rybicki, B.A. et al. Prostate cancer risk from occupational exposure to polycyclic aromatic hydrocarbons interacting with the GSTP1 Ile105Val polymorphism. *Cancer Detect Prev* **30**, 412-22 (2006).
51. Agalliu, I., Salinas, C.A., Hansten, P.D., Ostrander, E.A. & Stanford, J.L. Statin Use and Risk of Prostate Cancer: Results from a Population-based Epidemiologic Study. *American Journal of Epidemiology* **168**, 250-260 (2008).
52. Stanford, J.L., Wicklund, K.G., McKnight, B., Daling, J.R. & Brawer, M.K. Vasectomy and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* **8**, 881-6 (1999).
53. Ingles, S.A. et al. Association of prostate cancer with vitamin D receptor haplotypes in African-Americans. *Cancer Res* **58**, 1620-3 (1998).
54. Liu, X., Plummer, S.J., Nock, N.L., Casey, G. & Witte, J.S. Nonsteroidal antiinflammatory drugs and decreased risk of advanced prostate cancer: modification by lymphotoxin alpha. *Am J Epidemiol* **164**, 984-9 (2006).
55. Strom, S.S. et al. Androgen receptor polymorphisms and risk of biochemical failure among prostatectomy patients. *The Prostate* **60**, 343-351 (2004).
56. Gohagan, J.K., Prorok, P.C., Hayes, R.B. & Kramer, B.S. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clin Trials* **21**, 251S-272S (2000).
57. Signorello, L.B. et al. Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc* **97**, 972-9 (2005).
58. Musunuru, K. et al. Candidate Gene Association Resource (CARe) / Clinical Perspective. *Circulation: Cardiovascular Genetics* **3**, 267-275 (2010).
59. Lettre, G. et al. Genome-Wide Association Study of Coronary Heart Disease and Its Risk Factors in 8,090 African Americans: The NHLBI CARe Project. *PLoS Genet* **7**, e1001300 (2011).
60. Taylor, H.A., Jr. et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis* **15**, S6-4-17 (2005).
61. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39**, 3-9 (2002).
62. Williams, R.R. et al. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. *Ann Epidemiol* **10**, 389-400 (2000).

Nature Genetics: doi:10.1038/ng.2608

63. Cooper, R. et al. Evaluation of an electronic blood pressure device for epidemiological studies. *Blood pressure monitoring* **2**, 35-40 (1997).
64. Ataman, S.L. et al. Standardization of blood pressure measurement in an international comparative study. *Journal of clinical epidemiology* **49**, 869-77 (1996).
65. Tayo, B.O., Harders, R., Luke, A., Zhu, X. & Cooper, R.S. Latent common genetic components of obesity traits. *International journal of obesity* **32**, 1799-806 (2008).
66. Cooper, R. et al. The prevalence of hypertension in seven populations of west African origin. *American journal of public health* **87**, 160-8 (1997).
67. McDonough, C.W. et al. A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int* **79**, 563-72 (2011).
68. Lippman, S.M. et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA* **301**, 39-51 (2009).
69. Palmer, J.R., Wise, L.A., Horton, N.J., Adams-Campbell, L.L. & Rosenberg, L. Dual effect of parity on breast cancer risk in African-American women. *J Natl Cancer Inst* **95**, 478-83 (2003).
70. Williams, S.M. et al. A population-based study in Ghana to investigate inter-individual variation in plasma t-PA and PAI-1. *Ethn Dis* **17**, 492-7 (2007).
71. Schoenhard, J.A. et al. Male-female differences in the genetic regulation of t-PA and PAI-1 levels in a Ghanaian population. *Hum Genet* **124**, 479-88 (2008).
72. Cabral, D.N. et al. Population- and community-based recruitment of African Americans and Latinos: the San Francisco Bay Area Lung Cancer Study. *Am J Epidemiol* **158**, 272-9 (2003).
73. Huo, D. et al. Parity and breastfeeding are protective against breast cancer in Nigerian women. *Br J Cancer* **98**, 992-6 (2008).
74. Huo, D. et al. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and breast cancer risk in Africans. *Breast Cancer Res Treat* **110**, 367-76 (2008).
75. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
76. Rumpel, J.A. et al. Genetic ancestry and its association with asthma exacerbations among African American subjects with asthma. *J Allergy Clin Immunol* **130**, 1302-6 (2012).
77. Perrine, C.L. et al. Glycopeptide-preferring polypeptide GalNAc transferase 10 (ppGalNAc T10), involved in mucin-type O-glycosylation, has a unique GalNAc-O-Ser/Thr-binding site in its catalytic domain not found in ppGalNAc T1 or T2. *J Biol Chem* **284**, 20387-97 (2009).
78. Cheng, L. et al. Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T10. *FEBS Lett* **531**, 115-21 (2002).
79. Nelson, P.A., Sutcliffe, J.G. & Thomas, E.A. A new UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase mRNA exhibits predominant expression in the hypothalamus, thalamus and amygdala of mouse forebrain. *Brain Res Gene Expr Patterns* **1**, 95-9 (2002).
80. Prag, S. & Adams, J.C. Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinformatics* **4**, 42 (2003).
81. Boyden, L.M. et al. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482**, 98-102 (2012).
82. Cirak, S. et al. Kelch-like homologue 9 mutation is associated with an early onset autosomal dominant distal myopathy. *Brain* **133**, 2123-35 (2010).
83. Davies, R.W. et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet* **5**, 217-25 (2012).
84. Ellinor, P.T., Shin, J.T., Moore, R.K., Yoerger, D.M. & MacRae, C.A. Locus for atrial fibrillation maps to chromosome 6q14-16. *Circulation* **107**, 2880-3 (2003).

Nature Genetics: doi:10.1038/ng.2608