

RESEARCH ARTICLE

Open Access



Multiethnic genome-wide association study identifies ethnic-specific associations with body mass index in Hispanics and African Americans

Yasmmyn D. Salinas, Leyao Wang and Andrew T. DeWan*

Abstract

Background: Genome-wide association studies of obesity have typically assumed fixed genetic effects across ethnicities, rarely attempting to thoroughly compare and contrast findings across various ethnic groups. Therefore, our study aimed to identify novel genetic associations with body mass index (BMI), a common measure of obesity, and explore their cross-ethnic generalizability in a multiethnic population. To that end, we conducted ethnic-specific genome-wide association analyses among 1235 Hispanic, 706 Asian, 1549 African American, and 2395 European American subjects from the Multi-ethnic Study of Atherosclerosis (MESA). We compared findings across ethnicities and investigated single-nucleotide polymorphisms (SNPs) with suggestive BMI-association p -values among 3379 Hispanic and 6871 African American subjects from the Women's Health Initiative (WHI).

Results: We identified a genome-wide significant association in MESA Hispanics—rs12253976 in *KLF6* (beta = 5.792 kg/m² per-allele, 95 % confidence interval (CI): 3.885, 7.698; $p = 3.43 \times 10^{-9}$)—and suggestive SNPs with $p < 5 \times 10^{-6}$ in MESA Hispanics, European Americans and African Americans that display ethnic-specific effects on BMI. Of these suggestive SNPs, Hispanic SNP rs12255372 and African American SNP rs6435678 had the most evidence of replication in WHI. rs12255372 (in *TCF7L2*) was associated with lower BMI in both MESA (beta = -1.111 kg/m², 95 % CI: -1.578, -0.645; $p = 3.33 \times 10^{-6}$) and WHI Hispanics (beta = -0.304 kg/m², 95 % CI: -0.613, 0.006; $p = 0.054$). This *TCF7L2* intronic region contains several SNPs (rs7901695, rs4506565, rs4132670, and rs12243326) with low p -values ($p < 10^{-3}$) in MESA and betas of similar magnitude and direction in MESA and WHI, but only rs12243326 is in strong linkage disequilibrium with rs12255372 in our Hispanic populations, suggesting independent signals in this region. rs6435678 (in *ERBB4*) was associated with greater BMI in both MESA (beta = 1.104 kg/m², 95 % CI: 0.643, 1.564; $p = 2.85 \times 10^{-6}$) and WHI African Americans (beta = 0.219 kg/m², 95 % CI: -0.021, 0.460; $p = 0.074$).

Conclusions: Two BMI-association signals are present in the *TCF7L2* intronic region of Hispanics, one of which is tagged by rs12255372. *ERBB4* rs6435678 is a novel BMI-association signal in African Americans. Overall, our data suggest that ethnic-specific associations are involved in the genetic determination of BMI. Ethnic-specificity has potential implications for the development of gene-based therapies for obesity.

Keywords: Obesity, Body mass index, Ethnic differences, Genetic epidemiology

* Correspondence: andrew.dewan@yale.e
Department of Chronic Disease Epidemiology, Yale School of Public Health,
60 College Street, New Haven, CT 06520, USA

Background

Obesity is one of the most pressing health problems in the United States (U.S.). It affects nearly 35 % of adults and 17 % of children [1], predisposing them to many chronic conditions including type 2 diabetes (T2D), cardiovascular disease (CVD), and several cancers [2]. The cost of treating obesity-related conditions places great financial burden on the healthcare system [3]. Consequently, understanding the etiology of obesity and developing interventions to prevent its comorbidities are critical public health concerns.

The etiology of obesity is multifactorial [4], but family studies suggest that 40-70 % of the variation in body mass index (BMI), a common measure of obesity, is explained by genetic factors [5–7]. Genome-wide association studies (GWAS), which have identified over 100 loci associated with BMI and other obesity-related traits, have greatly expanded our understanding of the genetic basis of obesity [8]. However, further investigation is warranted for several reasons. First, known BMI loci account for only a fraction of the estimated variation in BMI [9]. Second, previous GWAS have primarily relied on data from subjects of European ancestry [10]. Hispanics and African Americans are underrepresented in GWAS, and it is precisely these populations that are overburdened by obesity in the U.S. [11]. Third, obesity GWAS have either analyzed a single ethnic group in isolation or pooled multiethnic data in cross-ethnic meta-analyses, assuming that genetic effects are *fixed* across ethnic groups, and have rarely attempted to thoroughly compare and contrast findings across ethnic groups.

As noted in [12], to more fully gauge the clinical and public health implications of genetic associations with BMI, studies should not only focus on the replication of genetic loci identified in European populations; they should also evaluate the cross-ethnic generalizability of genetic associations in multiethnic populations. Making unbiased cross-ethnic comparisons of genetic effects is facilitated by the availability of data from multiple ethnic groups sampled in the same fashion from the same underlying source population. Conducting ethnic-specific GWAS within such multiethnic populations could also reveal loci not readily detectable in Europeans due to cross-ethnic differences in allele frequencies and haplotype structures [13].

For these reasons, we used an ethnic-specific GWAS approach to examine genetic associations with BMI in the Multi-ethnic Study of Atherosclerosis (MESA), which includes subjects of four ethnicities: Hispanic, Asian, African American, and European American. We identified the top BMI-associated single-nucleotide polymorphisms (SNPs) in each ethnicity and evaluated whether those SNPs were associated with BMI to a similar extent in the other ethnicities. We then sought to

replicate the top SNPs in Hispanics and African Americans in an independent cohort consisting of multiethnic subjects from the Women's Health Initiative (WHI).

Results

Discovery sample characteristics and model covariates

Descriptive statistics for MESA are shown in Table 1. Unadjusted associations between participant characteristics and BMI are summarized in Additional file 1: Table S1. After model building, ethnic-specific covariates were: age, sex, smoking, diabetes, and arthritis in Hispanics; age, sex, education, diabetes, and arthritis in Asians and African Americans; and age, sex, income, education, smoking, physical activity, diabetes and arthritis in European Americans.

Population stratification

Additional file 2: Figures S1-S2 shows quantile-quantile plots of observed vs. expected p-values in the discovery and replication datasets, before and after adjustment for population stratification. Before adjustment, there was evidence of genomic inflation in MESA Hispanics ($\lambda = 1.019$), WHI Hispanics ($\lambda = 1.158$), and WHI African Americans ($\lambda = 1.662$). After systematic adjustment for the first two ethnic-specific principal components (PCs) in linear models, these λ estimates were significantly improved ($\lambda = 1.000$ for MESA Hispanics, $\lambda = 1.035$ for WHI Hispanics, and $\lambda = 1.034$ for WHI African Americans), and all λ values were below our predetermined threshold of 1.05. Adjustment for additional PCs did not materially alter these estimates. We note that no evidence of genomic inflation was observed in MESA Asians, European Americans, and African Americans. Systematic adjustment for ethnic-specific PC1 and PC2 did not greatly influence the magnitude of the observed p-values in these populations.

BMI-associated regions

Following SNP quality control (QC) in MESA, 853,278 SNPs in Hispanics, 683,998 in Asians, 871,948 in African Americans, and 749,659 in European Americans were analyzed (Additional file 3: Table S2). The top SNPs ($p < 5 \times 10^{-6}$) in the MESA ethnic groups are displayed in Table 2.

In linear regression analyses, adjusted for all covariates and assuming an additive mode of inheritance, the most significant SNP in Hispanics was rs12253976 (10p15.1) ~8 kb upstream of *KLF6* (beta = 5.792 kg/m² per-allele, 95 % confidence interval (CI): 3.885, 7.698; $p = 3.43 \times 10^{-9}$). This SNP was the only variant in ethnic-specific analyses to achieve genome-wide significance after Bonferroni adjustment for multiple comparisons. In African Americans, the most significant SNP was rs7763896 (6q23.2) ~7 kb upstream of *CTGF* (beta = 3.140 kg/m²

Table 1 Description of the MESA study population by ethnicity^a

Characteristic ^b	Hispanics (n = 1235) ^{c,d}	Asians (n = 706) ^{c,d}	African Americans (n = 1549) ^{c,d}	European Americans (n = 2395) ^{c,d}
Age (years)	61.26 ± 10.27	62.50 ± 10.36	62.35 ± 10.10	62.57 ± 10.19
Sex				
Female	624 (50.5)	357 (50.6)	827 (53.4)	1246 (52.0)
Male	611 (49.5)	349 (49.4)	722 (46.6)	1149 (48.0)
Income				
Low	894 (74.0)	468 (66.7)	733 (51.2)	756 (31.6)
High	314 (26.0)	234 (33.3)	700 (48.9)	1639 (68.4)
Education				
<12 years	577 (46.7)	172 (24.4)	188 (12.1)	115 (4.8)
12-15 years	532 (43.1)	257 (36.4)	839 (54.2)	1074 (44.8)
≥16 years	126 (10.2)	277 (39.2)	522 (33.7)	1206 (50.4)
Smoking				
Ever	563 (45.6)	176 (24.9)	852 (55.0)	1331 (55.6)
Never	672 (54.4)	530 (75.1)	697 (45.0)	1064 (44.4)
Physical Activity (met-min/wk)	5939.53 ± 6002.02	3764.16 ± 3915.99	6542.21 ± 6937.06	5699.50 ± 5383.03
Diabetes				
Yes	207 (16.8)	77 (10.9)	242 (15.6)	124 (5.2)
No	1028 (83.2)	629 (89.1)	1307 (84.4)	2271 (94.8)
Arthritis				
Yes	432 (35.0)	181 (25.6)	675 (43.6)	849 (35.5)
No	803 (65.0)	525 (74.4)	874 (56.4)	1546 (64.6)
Body Mass Index (kg/m ²)	29.35 ± 5.18	24.02 ± 3.27	30.11 ± 5.86	27.75 ± 5.07

Abbreviation: met-min/wk metabolic minutes per week

^a Table values are mean ± SD for continuous variables and n (column %) for categorical variables

^b Characteristics included are covariates in at least one ethnic-specific linear regression model

^c Sample size represents number of individuals that passed quality control and have complete data for all ethnic-specific linear regression model covariates

^d Percentages may not sum to 100 % due to rounding and n's may not sum to sample size due to missing values for variables not included as covariates in ethnic-specific regression models

per-allele, 95 % CI: 1.927, 4.352; $p = 4.35 \times 10^{-7}$). In European Americans, the most significant SNP was rs6866721 (5q23.1; intergenic) near *SEMA6A* (beta = 0.758 kg/m² per-allele, 95 % CI: 0.478, 1.039; $p = 1.26 \times 10^{-7}$). No SNP in Asians achieved $p < 5 \times 10^{-6}$. The strength of associations and estimated per-allele effect sizes for these SNPs were relatively consistent across the unadjusted, minimally-adjusted, and fully-adjusted models (Additional file 4: Table S3).

Regional plots visualizing association results for rs12253976, rs7763896, and rs6866721 and their respective flanking region (±500 kb) SNPs are shown in Additional file 5: Figures S3-S5. The chromosome 5 region of European American subjects contains SNPs with low p-values ($p < 10^{-5}$) and in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with rs6866721; these include SNPs rs1672492, rs1672491, and rs7704264 (Table 2). In contrast, the plots for rs7763896 in African Americans and rs12253976 in Hispanics did not show evidence of association for SNPs in their respective flanking regions.

Ethnic-specificity

The associations between the top SNPs and BMI were generally ethnic-specific (Table 2). Two exceptions were Hispanic SNP rs12255372 and European American SNP rs7926805, whose associations with lower BMI achieved nominal significance ($p < 0.05$) in African Americans. Nevertheless, the I^2 statistics (from concurrent cross-ethnic meta-analysis in MESA) for all top SNPs were > 50 %, indicating substantial cross-ethnic heterogeneity. Moreover, regional association plots showed that no variants in the vicinity of those SNPs were significantly associated with BMI in the other ethnicities.

Replication analyses

Additional file 6: Tables S4a-b describes the WHI samples, and Additional file 7: Tables S5a-b shows unadjusted associations between the examined subject characteristics and BMI.

The results of replication analyses in WHI Hispanics and African Americans are shown in Table 3. Three

Table 2 Top SNPs ($p < 5 \times 10^{-6}$) across the MESA ethnic groups ^{a,b}

	SNP	Allele ^c	Chr	Bp ^d	I ² ^e	Hispanics			African Americans			European Americans			Asians		
						MAF	beta (95% CI)	p ^f	MAF	beta (95% CI)	p ^f	MAF	beta (95% CI) ^g	p ^{fg}	MAF	beta (95% CI) ^h	p ^{fh}
Hispanics	rs2081325	T	2	118,205,061	93.6	0.015	3.99 (2.31, 5.67)	3.67x10 ⁻⁰⁶	0.085	0.30 (-0.41, 1.02)	0.404	0.002	MAF	MAF	0.000	MONO	MONO
	rs12253976	G	10	3,825,732	97.0	0.011	5.79 (3.89, 7.70)	3.43x10⁻⁰⁹	0.066	-0.22 (-1.01, 0.56)	0.575	0.001	MAF	MAF	0.000	MONO	MONO
	rs1409874	T	10	6,690,432	81.9	0.322	-1.05 (-1.49, -0.60)	4.22x10 ⁻⁰⁶	0.119	-0.46 (-1.07, 0.16)	0.148	0.184	-0.12 (-0.47, 0.24)	0.521	0.259	0.11 (-0.27, 0.49)	0.579
	rs1249269	G	10	28,660,149	88.4	0.414	1.03 (0.64, 1.43)	3.79x10 ⁻⁰⁷	0.360	-0.03 (-0.44, 0.39)	0.893	0.437	-0.12 (-0.40, 0.16)	0.414	0.113	-0.31 (-0.86, 0.24)	0.274
	rs11007044	A	10	28,680,671	85.9	0.171	1.29 (0.77, 1.81)	1.19x10 ⁻⁰⁶	0.142	0.29 (-0.27, 0.85)	0.308	0.267	0.14 (-0.18, 0.45)	0.395	0.006	MAF	MAF
	rs12255372	T	10	114,798,892	78.3	0.244	-1.11 (-1.58, -0.65)	3.33x10 ⁻⁰⁶	0.293	-0.52 (-0.96, -0.09)	0.019	0.290	-0.25 (-0.55, 0.06)	0.110	0.006	MAF	MAF
	rs16943469	T	17	54,799,958	93.9	0.015	3.98 (2.36, 5.61)	1.68x10 ⁻⁰⁶	0.070	0.25 (-0.54, 1.03)	0.537	0.000	MAF	MAF	0.000	MONO	MONO
	rs6038725	A	20	7,081,039	91.9	0.426	-1.04 (-1.43, -0.65)	2.24x10 ⁻⁰⁷	0.383	0.11 (-0.30, 0.52)	0.600	0.317	0.20 (-0.10, 0.49)	0.185	0.219	0.57 (0.17, 0.97)	0.006
	rs179747	G	20	7,107,995	90.9	0.444	-1.07 (-1.46, -0.68)	1.01x10 ⁻⁰⁷	0.341	0.19 (-0.22, 0.61)	0.362	0.321	0.18 (-0.11, 0.48)	0.227	0.236	0.31 (-0.08, 0.70)	0.119
	rs6085916	G	20	7,112,725	90.7	0.451	-1.03 (-1.42, -0.64)	2.28x10 ⁻⁰⁷	0.322	0.25 (-0.17, 0.68)	0.244	0.323	0.20 (-0.09, 0.49)	0.181	0.242	0.29 (-0.10, 0.68)	0.143
rs2326897	C	20	7,112,922	91.0	0.449	-1.05 (-1.44, -0.66)	1.39x10 ⁻⁰⁷	0.321	0.27 (-0.15, 0.70)	0.211	0.322	0.20 (-0.10, 0.49)	0.187	0.242	0.29 (-0.09, 0.67)	0.140	
rs13036410	T	20	7,122,578	90.7	0.445	-1.05 (-1.43, -0.66)	1.29x10 ⁻⁰⁷	0.337	0.19 (-0.23, 0.61)	0.376	0.318	0.18 (-0.11, 0.48)	0.225	0.242	0.29 (-0.09, 0.68)	0.139	
African Americans	rs7602754	C	2	19,315,168	79.8	0.148	0.27 (-0.28, 0.82)	0.332	0.356	-0.96 (-1.36, -0.55)	4.41x10 ⁻⁰⁶	0.234	-0.19 (-0.52, 0.13)	0.241	0.057	-0.61 (-1.36, 0.13)	0.107
	rs6435678	C	2	212,418,713	77.9	0.116	0.09 (-0.53, 0.71)	0.783	0.231	1.10 (0.64, 1.56)	2.85x10 ⁻⁰⁶	0.116	0.01 (-0.43, 0.45)	0.956	0.041	0.05 (-0.82, 0.93)	0.908
	rs6739663	A	2	240,878,904	84.8	0.456	-0.01 (-0.4, 0.38)	0.952	0.499	-0.93 (-1.32, -0.53)	4.45x10 ⁻⁰⁶	0.391	0.08 (-0.21, 0.36)	0.596	0.399	0.10 (-0.25, 0.44)	0.589
	rs7763896	C	6	132,321,497	91.5	0.057	0.16 (-0.67, 0.99)	0.707	0.026	3.14 (1.93, 4.35)	4.35x10 ⁻⁰⁷	0.114	-0.04 (-0.47, 0.40)	0.864	0.000	MONO	MONO
	rs1567101	A	10	30,901,089	81.8	0.130	0.09 (-0.50, 0.67)	0.776	0.386	1.03 (0.63, 1.43)	5.96x10 ⁻⁰⁷	0.106	0.22 (-0.22, 0.66)	0.334	0.285	-0.07 (-0.45, 0.32)	0.741
	rs7092615	T	10	30,902,186	81.3	0.127	0.09 (-0.51, 0.69)	0.759	0.375	1.03 (0.62, 1.43)	8.36x10 ⁻⁰⁷	0.106	0.18 (-0.26, 0.62)	0.411	0.284	-0.05 (-0.44, 0.33)	0.786
	rs12826956	G	12	73,122,021	88.0	0.140	0.16 (-0.41, 0.72)	0.587	0.150	1.30 (0.75, 1.85)	3.48x10 ⁻⁰⁶	0.233	-0.32 (-0.65, 0.00)	0.053	0.050	0.11 (-0.62, 0.85)	0.764
	rs2822787	A	21	14,876,531	83.3	0.167	0.31 (-0.20, 0.83)	0.235	0.215	1.15 (0.67, 1.64)	3.12x10 ⁻⁰⁶	0.180	-0.13 (-0.49, 0.23)	0.471	0.499	0.17 (-0.17, 0.51)	0.334
European Americans	rs6899277	G	5	78,881,387	82.3	0.383	0.07 (-0.33, 0.48)	0.725	0.177	0.25 (-0.28, 0.78)	0.349	0.380	-0.70 (-0.99, -0.41)	1.75x10 ⁻⁰⁶	0.425	-0.04 (-0.37, 0.29)	0.818
	rs6866721	C	5	116,240,680	85.8	0.447	0.03 (-0.36, 0.42)	0.887	0.343	-0.19 (-0.6, 0.22)	0.358	0.434	0.76 (0.48, 1.04)	1.26x10 ⁻⁰⁷	0.295	-0.06 (-0.42, 0.3)	0.755
	rs1672492	G	5	116,250,736	82.0	0.498	-0.10 (-0.49, 0.30)	0.632	0.484	0.33 (-0.07, 0.72)	0.108	0.368	0.68 (0.39, 0.97)	3.79x10 ⁻⁰⁶	0.227	-0.20 (-0.59, 0.20)	0.323
	rs1672491	C	5	116,250,866	81.9	0.500	-0.09 (-0.48, 0.30)	0.653	0.484	0.32 (-0.08, 0.72)	0.114	0.367	0.69 (0.40, 0.97)	3.04x10 ⁻⁰⁶	0.223	-0.19 (-0.59, 0.20)	0.340
	rs7704264	A	5	116,252,052	85.9	0.396	0.11 (-0.29, 0.50)	0.605	0.241	-0.03 (-0.5, 0.45)	0.907	0.445	0.71 (0.44, 0.99)	5.66x10 ⁻⁰⁷	0.433	-0.26 (-0.59, 0.08)	0.133
	rs2497155	T	6	84,920,682	59.2	0.070	-0.51 (-1.27, 0.25)	0.187	0.025	0.04 (-1.21, 1.29)	0.949	0.106	-1.05 (-1.49, -0.61)	3.56x10 ⁻⁰⁶	0.051	0.05 (-0.75, 0.84)	0.906
	rs7926805	C	11	17,715,163	83.6	0.242	0.06 (-0.41, 0.53)	0.797	0.480	-0.48 (-0.88, -0.09)	0.017	0.233	-0.80 (-1.11, -0.48)	9.93x10 ⁻⁰⁷	0.160	0.29 (-0.17, 0.75)	0.222

Abbreviations: Chr Chromosome, Bp base pair, CI confidence interval, MAF minor allele frequency

^a No SNP in Asians with $p < 5 \times 10^{-6}$

^b Boxed values indicate the top SNPs within each ethnic group; p-value in bold indicates genome-wide significant result

^c Minor allele

^d Base pair positions reported refer to the March 2006 human genome reference assembly (NCBI36/hg18)

^e I² values from cross-ethnic meta-analysis

^f p-values adjusted for first two principal components and all covariates in ethnic-specific linear regression models

^g MAF indicates SNP had a MAF < 0.01 in that particular ethnic group

^h MONO indicates SNP is monomorphic in that particular ethnic group

Table 3 Top Hispanic and African American SNPs ($p < 5 \times 10^{-6}$) across MESA and WHI ^a

SNP	Minor Allele	Chr	Gene	Bp	MESA Hispanics				WHI Hispanics			
					MAF	beta	95 % CI	p ^b	MAF	beta	95 % CI	p ^b
rs2081325	T	2	intergenic	118,205,061	0.015	3.987	(2.307, 5.667)	3.67×10^{-06}	0.009	0.155	(-1.174, 1.483)	0.820
rs12253976	G	10	<i>KLF6</i>	3,825,732	0.011	5.792	(3.885, 7.698)	3.43×10^{-09}	0.008	0.598	(-0.837, 2.032)	0.414
rs1409874	T	10	intergenic	6,690,432	0.322	-1.045	(-1.488, -0.602)	4.22×10^{-06}	0.312	-0.089	(-0.380, 0.202)	0.550
rs1249269	G	10	intergenic	28,660,149	0.414	1.030	(0.635, 1.425)	3.79×10^{-07}	0.410	-0.197	(-0.472, 0.077)	0.159
rs11007044	A	10	intergenic	28,680,671	0.171	1.290	(0.772, 1.808)	1.19×10^{-06}	0.176	-0.113	(-0.465, 0.239)	0.530
rs12255372	T	10	<i>TCF7L2</i>	114,798,892	0.244	-1.111	(-1.578, -0.645)	3.33×10^{-06}	0.249	-0.304	(-0.613, 0.006)	0.054
rs16943469	T	17	<i>YPEL2</i>	54,799,958	0.015	3.984	(2.362, 5.607)	1.68×10^{-06}	0.006	0.229	(-1.475, 1.932)	0.793
rs6038725	A	20	intergenic	7,081,039	0.426	-1.038	(-1.429, -0.647)	2.24×10^{-07}	0.384	0.218	(-0.054, 0.489)	0.116
rs179747	G	20	intergenic	7,107,995	0.444	-1.065	(-1.455, -0.676)	1.01×10^{-07}	0.392	0.228	(-0.044, 0.501)	0.101
rs6085916	G	20	intergenic	7,112,725	0.451	-1.028	(-1.416, -0.641)	2.28×10^{-07}	0.399	0.197	(-0.078, 0.472)	0.160
rs2326897	C	20	intergenic	7,112,922	0.449	-1.051	(-1.439, -0.662)	1.39×10^{-07}	0.397	0.197	(-0.073, 0.467)	0.153
rs13036410	T	20	intergenic	7,122,578	0.445	-1.047	(-1.433, -0.661)	1.29×10^{-07}	0.391	0.219	(-0.050, 0.488)	0.111
SNP	Minor Allele	Chr	Gene	Bp	MESA African Americans				WHI African Americans			
					MAF	beta	95 % CI	p ^b	MAF	beta	95 % CI	p ^b
rs7602754	C	2	intergenic	19,315,168	0.356	-0.956	(-1.362, -0.549)	4.41×10^{-06}	0.340	0.099	(-0.117, 0.315)	0.370
rs6435678	C	2	<i>ERBB4</i>	212,418,713	0.231	1.104	(0.643, 1.564)	2.85×10^{-06}	0.235	0.232	(-0.009, 0.476)	0.061
rs6739663	A	2	intergenic	240,878,904	0.499	-0.926	(-1.319, -0.532)	4.45×10^{-06}	0.494	-0.194	(-0.399, 0.008)	0.062
rs7763896	C	6	<i>CTGF</i>	132,321,497	0.026	3.140	(1.927, 4.352)	4.35×10^{-07}	0.026	-0.175	(-0.806, 0.460)	0.589
rs1567101	A	10	intergenic	30,901,089	0.386	1.030	(0.627, 1.433)	5.96×10^{-07}	0.392	0.033	(-0.178, 0.242)	0.761
rs7092615	T	10	intergenic	30,902,186	0.375	1.025	(0.619, 1.431)	8.36×10^{-07}	0.386	0.061	(-0.149, 0.271)	0.567
rs12826956	G	12	intergenic	73,122,021	0.150	1.300	(0.753, 1.847)	3.48×10^{-06}	0.152	0.010	(-0.277, 0.292)	0.945
rs2822787	A	21	<i>SAMSN1</i>	14,876,531	0.215	1.153	(0.670, 1.636)	3.12×10^{-06}	0.202	-0.067	(-0.318, 0.186)	0.601

Abbreviations: Chr Chromosome, Bp base pair, CI confidence interval, MAF minor allele frequency

^a Bolded SNPs have beta in the same direction and suggestive p-values in MESA and WHI

^b p-values adjusted for all ethnic-specific linear regression model covariates; only the MESA analyses are sex-adjusted

Table 4 Association with rs12255372 across the MESA and WHI Hispanics linear regression models

	MESA Hispanics		WHI Hispanics	
	beta (95 % CI)	p	beta (95 % CI)	p
Unadjusted model	-1.144 (-1.623, -0.666)	3.09×10^{-6}	-0.409 (-0.718, -0.101)	0.009
+ PC 1 and 2	-1.034 (-1.519, -0.548)	3.24×10^{-5}	-0.342 (-0.655, -0.028)	0.033
+ age and sex ^a	-1.014 (-1.494, -0.534)	3.75×10^{-5}	-0.334 (-0.647, -0.021)	0.037
Full model ^{a,b}	-1.111 (-1.578, -0.645)	3.33×10^{-6}	-0.304 (-0.613, 0.006)	0.054

Abbreviations: CI confidence interval, PC principal components

^a Only the MESA analyses are sex-adjusted

^b Full model includes all ethnic-specific model covariates

SNPs—rs12255372 (10q25.2; *TCF7L2*), rs6435678 (2q.34; *ERBB4*), and rs6739663 (2q37.3; intergenic)—approached nominal significance in WHI, with betas in the same direction as in MESA. *TCF7L2* and *ERBB4* were associated with BMI in a recent large-scale GWAS meta-analysis [8]. Therefore, we examined these loci more thoroughly.

TCF7L2

rs12255372 is the Hispanic SNP with most suggestive evidence of replication in WHI ($p = 3.33 \times 10^{-6}$ in MESA, and $p = 0.037$ and $p = 0.054$ respectively in the WHI age-adjusted and fully-adjusted models; Table 4). Minor allele frequencies (MAFs) for rs12255372 were similar across both Hispanic populations (0.244 in

MESA and 0.249 in WHI; Table 3 and Additional file 8: Table S6).

Figure 1 is a regional plot visualizing association results for rs12255372 and its flanking region (± 500 kb) markers in both Hispanic populations. This plots shows four other SNPs—rs7901695, rs4506565, rs4132670, and rs12243326—with low p-values ($p < 10^{-3}$) in MESA. These SNPs approached or achieved nominal significance in WHI and had betas of similar magnitude and direction in both Hispanic populations (Table 5). However, only rs12243326 is in strong LD with rs12255372 in our Hispanic populations ($r^2 = 0.89$ in both MESA and WHI). The other three (rs7901695, rs4506565, and rs4132670) have r^2 values of 0.37-0.54 in MESA and 0.54-0.63 in WHI Hispanics.

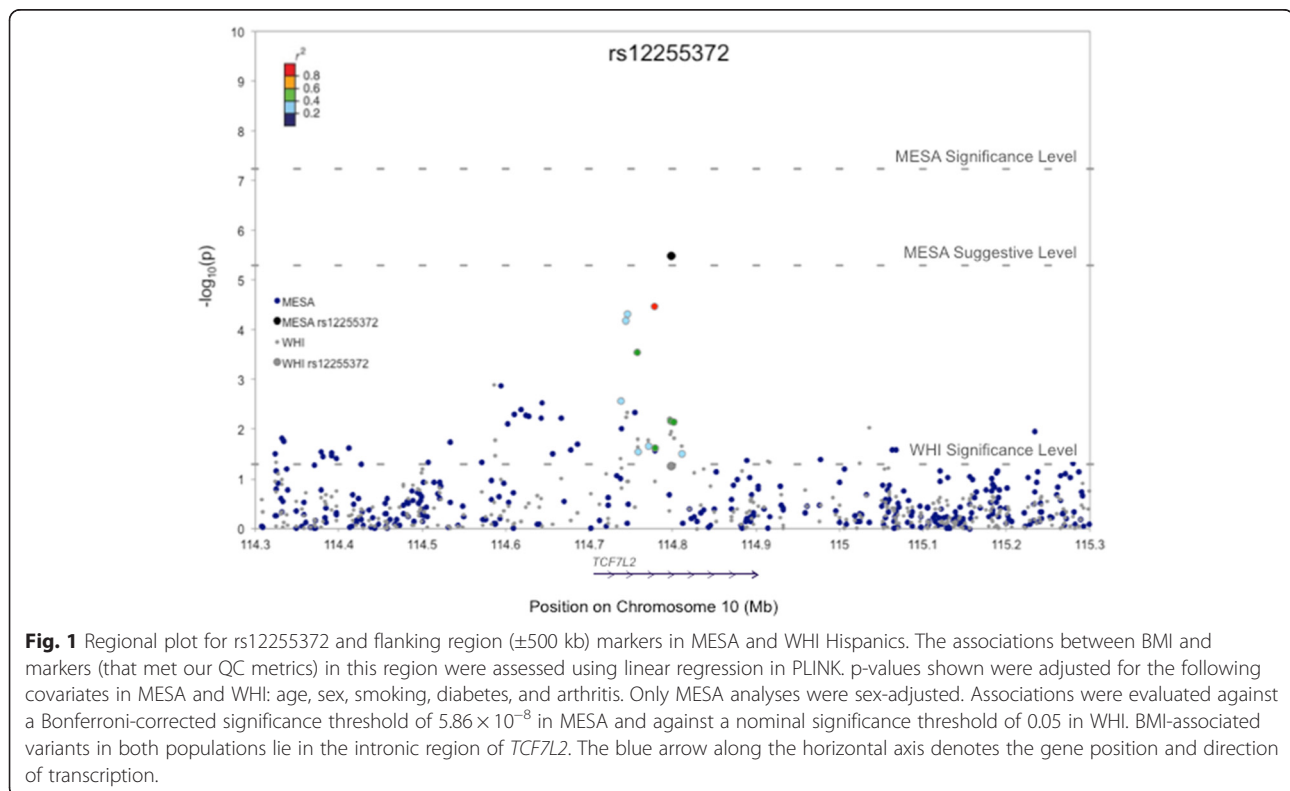


Table 5 BMI-associated SNPs in TCF7L2 in MESA and WHI Hispanics

SNP	Bp	Minor Allele	MESA Hispanics				WHI Hispanics			
			Full model ^a		Full model ^a w/o diabetes		Full Model ^a		Full model ^a w/o diabetes	
			beta (95 % CI)	p	beta (95 % CI)	p	beta (95 % CI)	p	beta (95 % CI)	p
rs7901695	114,744,078	C	-0.867 (-1.291, -0.442)	6.60 × 10 ⁻⁰⁵	-0.812 (-1.244, -0.380)	2.41 × 10 ⁻⁰⁴	-0.403 (-0.690, -0.116)	0.006	-0.381 (-0.669, -0.092)	0.010
rs4506565	114,746,031	A	-0.885 (-1.310, -0.460)	4.81 × 10 ⁻⁰⁵	-0.835 (-1.267, -0.402)	1.64 × 10 ⁻⁰⁴	-0.415 (-0.703, -0.128)	0.005	-0.396 (-0.686, -0.107)	0.007
rs4132670	114,757,761	A	-0.816 (-1.256, -0.377)	2.86 × 10 ⁻⁰⁴	-0.793 (-1.241, -0.345)	0.001	-0.337 (-0.628, -0.047)	0.023	-0.321 (-0.614, -0.028)	0.032
rs12243326	114,778,805	G	-0.990 (-1.457, -0.523)	3.46 × 10 ⁻⁰⁵	-0.991 (-1.467, -0.515)	4.72 × 10 ⁻⁰⁵	-0.251 (-0.560, 0.059)	0.112	-0.243 (-0.554, 0.069)	0.127
rs12255372	114,798,892	T	-1.111 (-1.578, -0.645)	3.33 × 10 ⁻⁰⁶	-1.103 (-1.578, -0.628)	5.82 × 10 ⁻⁰⁶	-0.304 (-0.613, 0.006)	0.054	-0.296 (-0.608, 0.015)	0.063

Abbreviations Bp = base pair, CI = Confidence Interval; w/o = without

^a Full model includes: PC1 and PC2, age, sex, and ethnic-specific model covariates; only MESA analyses are sex-adjusted

ERBB4

rs6435678 is the African American SNP with most suggestive evidence of replication in WHI ($p = 2.85 \times 10^{-6}$ in MESA, and $p = 0.051$ and $p = 0.061$ respectively in the WHI age-adjusted and fully-adjusted models; Table 6). MAFs for rs6435678 were similar across both African American populations (0.231 in MESA and 0.235 in WHI; Table 3 and Additional file 9: Table S7).

Figure 2 is a regional plot visualizing association results for rs6435678 and its flanking region (± 500 kb) markers in both African American populations. This plot shows another BMI-associated SNP at this locus: rs16847102 ($p = 1.12 \times 10^{-5}$ in MESA, and $p = 0.053$ and $p = 0.074$ respectively in the WHI age-adjusted and fully-adjusted model). rs16847102 is in strong LD with rs6435678 in our African American populations ($r^2 = 0.90$ and 0.87 respectively in MESA and WHI), and the strength, magnitude, and direction of its BMI-association across both populations mirror that of rs6435678 across the unadjusted, minimally-adjusted, and fully-adjusted models (Table 6).

Discussion

Ethnic-specific associations with BMI in MESA

In this study, we investigated genetic associations with BMI via an ethnic-specific GWAS approach. Using data from MESA, we identified suggestive SNPs ($p < 5 \times 10^{-6}$) displaying ethnic-specific effects on BMI. These include rs12253976 (10p15.1; *KLF6*) in Hispanics, rs7763896 (6q23.2; *CTGF*) in African Americans, and rs6866721 (5q23.1; intergenic) in European Americans. The I^2 values from concurrent cross-ethnic meta-analyses in MESA provided statistical evidence of substantial cross-ethnic heterogeneity, suggesting that the top SNP effects were not generalizable across ethnicities. Combining results across ethnicities would have masked the SNP effects at these loci (Additional file 10: Table S8). Our findings hence provide support for the hypothesis that ethnic-specific associations are involved in the genetic determination of BMI and highlight the importance of using ethnic-specific approaches for discovery of genetic associations with obesity-related traits.

Ethnic-specificity, which we define as heterogeneity of SNP effects across ethnicities, likely explains some previous failed replications of candidate obesity loci. Examples include an association in the *SIMI* intronic region, discovered in Pima Indians but not generalizable to French Europeans [14], as well as functional coding variant W64R in *ADRB3*, associated with BMI in East Asians but not in Europeans [15]. Ethnic-specificity also has important implications for the evaluation of genetic loci as potential therapeutic agents for obesity: it emphasizes the need for a personalized medicine approach that focuses on identifying the most effective therapies for subjects of different ethnicities.

Table 6 BMI-associated SNPs in *ERBB4* in MESA and WHI African Americans

SNP	Bp	Minor Allele	Model	MESA African Americans		WHI African Americans	
				beta (95 % CI)	p	beta (95 % CI)	p
rs6435678	212,418,713	C	Unadjusted	1.106 (0.647, 1.565)	2.58×10^{-06}	0.350 (0.098, 0.602)	0.006
			+ PC 1 and 2	1.012 (0.526, 1.498)	4.71×10^{-05}	0.238 (-0.014, 0.489)	0.064
			+ age and sex ^a	1.121 (0.652, 1.591)	3.10×10^{-06}	0.249 (-0.001, 0.500)	0.051
			Full model ^{a,b}	1.104 (0.643, 1.564)	2.85×10^{-06}	0.232 (-0.009, 0.476)	0.061
rs16847102	212,431,233	A	Unadjusted	0.938 (0.458, 1.418)	1.33×10^{-04}	0.349 (0.100, 0.599)	0.006
			+ PC 1 and 2	0.911 (0.430, 1.393)	2.15×10^{-04}	0.238 (-0.012, 0.487)	0.062
			+ age and sex ^a	1.041 (0.576, 1.506)	1.25×10^{-05}	0.245 (-0.003, 0.493)	0.053
			Full model ^{a,b}	1.026 (0.570, 1.483)	1.12×10^{-05}	0.219 (-0.021, 0.460)	0.074

Abbreviations: Bp base pair, CI Confidence Interval, PC principal components

^a Only the MESA analyses are sex-adjusted

^b Full model includes all ethnic-specific model covariates

Replication in WHI

We sought to replicate the genome-wide significant association for rs12253976 in *KLF6* using data from WHI Hispanics. However, no association signal was detected in this independent population. Since the WHI is an all-female cohort, we explored whether this failed replication could be partially explained by a difference of the SNP effect between men and women. For this purpose, we conducted the following analyses in MESA: a formal test for heterogeneity of the SNP effect by sex [with a cross-product SNP-by-sex interaction term added to the ethnic-specific model] and sex-stratified analyses. As

shown in Additional file 11: Table S9, the p-value for the SNP-by-sex interaction for rs12253976 in MESA Hispanics was 2.51×10^{-4} . While not genome-wide significant, this result prompted us to explore the results of sex-stratified analyses, which revealed that the BMI association signal for rs12253976 in MESA Hispanics was actually stronger in women. Therefore, we concluded that our failure to replicate the findings for this SNP in WHI Hispanics was likely not due to an initial male-driven association in MESA.

We also explored the following explanations for this failed replication: heterogeneity in the BMI distributions

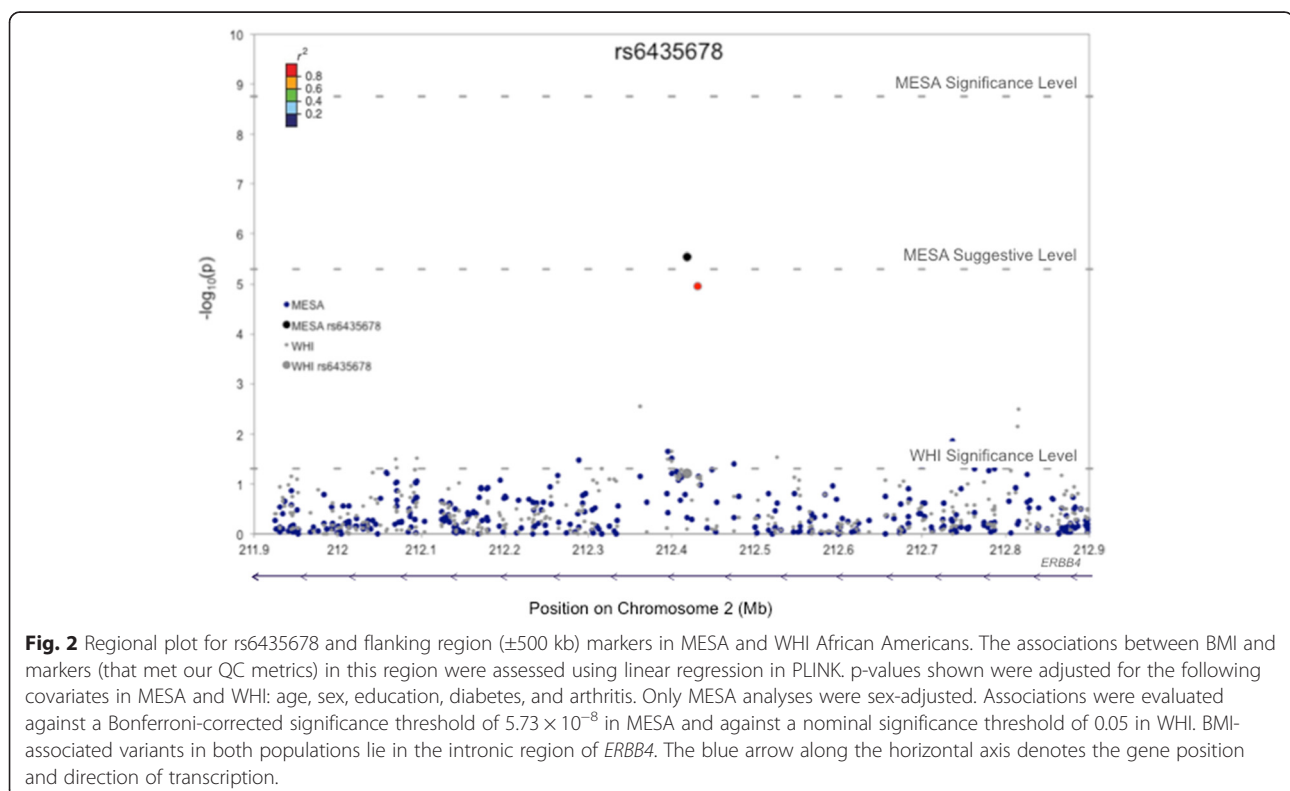


Fig. 2 Regional plot for rs6435678 and flanking region (± 500 kb) markers in MESA and WHI African Americans. The associations between BMI and markers (that met our QC metrics) in this region were assessed using linear regression in PLINK. p-values shown were adjusted for the following covariates in MESA and WHI: age, sex, education, diabetes, and arthritis. Only MESA analyses were sex-adjusted. Associations were evaluated against a Bonferroni-corrected significance threshold of 5.73×10^{-8} in MESA and against a nominal significance threshold of 0.05 in WHI. BMI-associated variants in both populations lie in the intronic region of *ERBB4*. The blue arrow along the horizontal axis denotes the gene position and direction of transcription.

of the two Hispanic populations and genetic heterogeneity between the two populations at this locus. As shown in Additional file 12: Figures S6-S7, the BMI distributions of MESA and WHI are similar. Furthermore, the MAFs for this SNP were only slightly different in MESA and WHI (0.011 vs. 0.008; Table 3). However, the genotype distributions were different across the two populations. There were no MESA Hispanic minor allele homozygotes; in fact, the observed association was driven by 27 heterozygotes, who, on average, were 5.8 BMI units heavier than major allele homozygotes (Additional file 13: Table S10). In contrast, all three genotypes were represented in WHI, and the estimated per-allele effect size was more modest. While these differences may be due to sample size differences, they may also indicate true heterogeneity between MESA and WHI. Therefore, the locus containing rs12253976 merits further investigation in other Hispanic populations.

Association with *TCF7L2*

We also investigated all other suggestive SNPs ($p < 5 \times 10^{-6}$) in MESA Hispanics in WHI. The Hispanic SNP with most suggestive evidence of replication in WHI was rs12255372, an intronic variant in *TCF7L2*. The regional association plot for this locus displayed four other BMI-associated variants (rs7901695, rs4506565, rs4132670, and rs12243326). Of these, only rs12243326 was in strong LD with rs12255372. The other three SNPs (rs7901695, rs4506565, rs4132670) were in weak LD with rs12255372 in our Hispanic populations, but are expected to be in strong LD with the previously-reported BMI-associated *TCF7L2* variant, rs7903146 [8] [estimated $r^2 = 0.72-1.00$ across the representative International HapMap Project (HapMap phase 3 [16]) populations of Mexican ancestry in Los Angeles, California (MEX), Utah Residents of Northern and Western European Ancestry (CEU), and Chinese in Metropolitan Denver, Colorado (CHD); Additional file 14: Table S11]. Therefore, rs7901695, rs4506565, rs4132670 may serve as proxies for rs7903146 (not genotyped in the Affymetrix 6.0 SNP array), and we conclude that the rs12255372 BMI signal is independent from that of rs7903146. Altogether, this suggests that there are two BMI-associated regions at the *TCF7L2* locus in our Hispanic populations, one tagged by the rs7903146 proxy SNPs and another by rs12255372.

The minor alleles of rs12255372 and rs7903146 have been consistently associated with an increased risk of T2D [17, 18] and have thus been studied extensively in that context. Studies in populations of European ancestry [19], where the two SNPs are in strong LD (Additional file 14: Table S11 and Additional file 15: Figures S8-S13), have proposed rs7903146 as the causal *TCF7L2* variant, given the stronger T2D-association signal at that SNP. However, studies in Hispanic and

African American populations [20, 21], where LD between the two SNPs is weak or non-existent (Additional file 14: Table S11 and Additional file 15: Figures S8-S13), have reported association signals at both SNPs, with one study [21] showing that, in Hispanics, rs12255372 yields a stronger T2D-signal than rs7903146. These studies suggest a role for rs12255372 as an independent T2D-signal in *TCF7L2*; and both SNPs may be functionally significant, as both reside in independent, predicted enhancer sites [22].

Our findings suggest a similar story in the context of BMI determination. Locke *et al.* [8], whose most significant *TCF7L2* analysis only included subjects of European ancestry, proposed rs7903146 as a causal variant in this region. Our study, on the other hand, was able to detect two *TCF7L2* signals, since LD does not mask the rs12255372 signal in Hispanic populations.

We note that our study did not find a significant association between the rs7903146 proxy variants (rs7901695, rs4506565, and rs4132670) and BMI in MESA European Americans. We propose two possible explanations for this: that our study was insufficiently powered to detect the purported effect size for rs7903146 in Europeans (-0.02 kg/m^2 per (minor) allele [8]), and/or that, unlike our study, which consisted entirely of population-based samples, the Locke *et al.* meta-analysis also included case-control studies of T2D. Regarding the latter, Locke *et al.* detected evidence of systematic ascertainment bias at this locus (stronger effects in T2D case-control studies than in population-based studies) [8]. This is in line with candidate gene investigations in population-based samples of European ancestry, such as DESIR [23] and the Framingham Heart Study [24], which refuted prior claims of *TCF7L2* BMI-associations made by studies examining this relationship only among individuals with T2D [25].

Nonetheless, in our Hispanic population-based samples, we find that the minor alleles of *TCF7L2* intronic variants are associated with lower BMI. Since our regression models had adjusted for diabetes, we examined the effect of removing this variable from the models in *ad hoc* analyses. Table 5 shows that the associations with lower BMI at this locus were either attenuated or unchanged after removing this variable. Thus, in our Hispanic populations, *TCF7L2* intronic variants are associated with lower BMI independently of T2D.

Association with *ERBB4*

We also investigated all other suggestive SNPs ($p < 5 \times 10^{-6}$) in MESA African Americans in WHI. The SNP with the most suggestive p-values across both African American populations was rs6435678, an intronic variant in *ERBB4*. The regional plot for this locus showed that the association pattern of the rs6435678 flanking region markers reflected the LD structure of the MESA

and WHI African American populations, thus lending additional support to our finding. We note that, upon inspecting this region across the other ethnicities, we found no significant evidence of BMI-associations. Thus, our data suggest that the BMI-effect of rs6435678 may be specific to African Americans, though further investigation in independent multiethnic samples is necessary to substantiate this finding.

ERBB4 was previously linked to BMI in populations of European ancestry via an association with rs7599312 [8], located ~10 kb upstream of this gene. rs6435678, which resides in intron 3, is not in LD with rs7599312 in our African American samples ($r^2 = 0.00$ and 0.01 respectively in MESA and WHI). Thus, we conclude that rs6435678 represents a novel signal in *ERBB4*. We note that no association with rs7599312 was detected in MESA European Americans. However, this was not surprising because the purported effect size for this SNP is only 0.02 kg/m^2 per-allele in populations of European ancestry [8], which our study was not powered to detect. We also note that there is no LD between rs7599312 and rs6435678 in European Americans ($r^2 = 0.00$ in MESA).

An association between variants in *ERBB4* and BMI is biologically plausible. *ERBB4* encodes a receptor tyrosine kinase expressed in various tissues, including liver and pancreas. In the liver, *ERBB4* regulates lipogenesis by binding to Neuregulin 4, an epidermal growth factor secreted by brown adipose tissue [26]. In the pancreas, *ERBB4* is involved in the epidermal growth factor receptor signaling pathway, which regulates islet cell differentiation [27, 28] and β -cell signal transduction, and whose disruption has been linked to impaired glucose tolerance and reduced insulin response in mice [29].

Strengths and limitations

Conducting our GWAS analyses within MESA gave us the unique opportunity to compare genetic associations across four ethnic groups that were sampled in the same fashion from the same underlying source population. However, we note that stratifying MESA by ethnicity limited our statistical power to detect variants with small effect sizes: for variants with $\text{MAF} \geq 0.2$, we had $\geq 80\%$ power to detect effect sizes of ≥ 1.4 , ≥ 1.7 , ≥ 1.2 , and $\geq 1.7 \text{ kg/m}^2$ in Asians, Hispanics, European Americans, and African Americans, respectively (Additional file 16: Tables S12). This could explain why only one SNP achieved genome-wide significance; why no suggestive SNPs ($p < 5 \times 10^{-6}$) were identified in Asians; or why *FTO* was only nominally associated in European Americans and had inconsistent effects in other ethnicities. We acknowledge that a fixed-effects meta-analysis of ethnic-specific GWAS data would be better powered than our ethnic-stratified approach. However, in our own meta-

analysis in MESA, we detected substantial evidence of cross-ethnic heterogeneity. Thus, we concluded that pooled effect estimates across ethnicities should not be presented; they would be meaningless, since the effect of the SNPs is not common to all ethnicities.

Another limitation is that many of the SNPs that we identified in MESA did not show evidence of replication in WHI. The WHI Hispanic and African American populations—though similar to their MESA counterparts with respect to BMI distribution (Additional file 12: Figures S6-S7) and MAFs at the evaluated loci (Table 3)—are composed entirely of women, and this had the potential to affect our ability to replicate our findings. Given the results of our formal tests of heterogeneity of the SNP effects by sex and the accompanying sex-stratified analyses in MESA (Additional file 11: Table S9), there is no evidence to suggest that failure to replicate our findings is due to initial male-driven associations in MESA.

A final limitation is that the genotyping platform used by the MESA and WHI studies (the Affymetrix 6.0 SNP array) was designed to optimize coverage of common genetic variants ($\text{MAF} \geq 0.1$). As noted in [30], lower frequency variants are more likely to be ethnic-specific. Therefore, in multiethnic studies, the use of custom arrays optimized for minority populations would be optimal.

Conclusions

By employing an ethnic-specific GWAS approach, we identified suggestive BMI-associated SNPs in Hispanics, African Americans, and European Americans that can be explored in future studies. The Hispanic and African American SNPs directed us to *TCF7L2* and *ERBB4*. We show that the *TCF7L2* intronic region contains two BMI-association signals in Hispanics, one of which (rs12255372) would have likely gone undetected had we not employed an ethnic-specific analytic approach. We also show that the *ERBB4* intronic region contains a novel BMI association signal (rs6435678) that may be specific to African Americans.

Overall, our data suggest that ethnic-specific associations are involved in the genetic determination of BMI. The existence of heterogeneous SNP effects across ethnicities highlights the need for utilizing ethnic-specific approaches for discovery of genetic associations and may have important implications for the development of gene-based therapies for common diseases such as obesity.

Methods

Discovery phase

Subjects providing data for the discovery phase included 1235 Hispanic, 706 Asian, 1549 African American, and 2395 European American subjects recruited into

MESA, a multi-center, prospective study of risk factors affecting CVD progression. Recruitment has been described elsewhere [31]. Briefly, 6814 men and women aged 45–84 years were recruited from six U.S. field centers in 2000–2002. MESA ascertained subject race and ethnicity via a standard questionnaire that adopted the definitions used by the U.S. Office of Management and Budget (OMB). [For simplicity, our present study uses the term ‘ethnicity’ to refer to the four racial-ethnic groups defined in MESA]. MESA recruited overlapping ethnic groups among field centers to minimize confounding by ethnicity by site [31]. Blood was collected from each subject, and DNA samples were genotyped for 909,622 SNPs using the Affymetrix 6.0 SNP array. Samples were required to have a call rate > 95 %. Further details of sample preparation and genotyping are described elsewhere [31].

Genotype and phenotype information for MESA were obtained from the National Center for Biotechnology Information's database of Genotypes and Phenotypes (NCBI dbGaP study accession: phs000209.v11.p3 MESA SNP Health Association Resource (SHARe)). For the present analysis, MESA was stratified into four ethnic-specific samples. EIGENSTRAT [32] analyses verified that the MESA ethnic groups were clustering together based on genotype data (Additional file 17: Figures S14-S15). Subjects analyzed met the QC thresholds described below and summarized in Additional file 3: Table S2 and had complete data for all ethnic-specific model covariates.

All phenotypic data used herein were obtained at the MESA baseline examination. The primary outcome variable was BMI (kg/m^2), calculated from height and weight measurements collected by trained staff at the field centers. All genotyped subjects had baseline BMI data. Variables examined as potential covariates due to their previously-reported associations with BMI were sex, baseline age, education, income, smoking, arthritis, diabetes, and physical activity. Details regarding covariate measurement are provided in Additional file 18: Supplemental Methods.

Ethnic-specific associations between BMI and potential covariates were examined in SAS 9.3 (SAS Institute, Cary NC). Ethnic-specific linear models were then built, with the most parsimonious models selected via backwards elimination of covariates with $p > 0.05$ until model adjusted- r^2 values were maximized. Age and sex were retained in the models regardless of the statistical significance of their associations with BMI. Individuals with missing values for any covariate included in the final ethnic-specific models were removed from the analyses. Details of variable parameterization and evaluation of the appropriateness of using linear regression on these data are provided in Additional file 18: Supplemental Methods.

Ethnic-specific SNP QC analyses were performed within PLINK [33]. We excluded SNPs based on low call rate (< 98 %), low MAF (< 0.01), and significant deviation from Hardy-Weinberg equilibrium ($p \leq 5.5 \times 10^{-8}$) (Additional file 3: Table S2).

Genetic QC procedures also included assessments for cryptic relatedness and population stratification. Cryptic relatedness between subjects in each ethnicity was examined within PLINK using pair-wise identity-by-descent (IBD) estimation. Pairs with $\hat{\pi}$ (estimated proportion of genome shared IBD) > 0.2 were inspected, and only one subject from each family was included. Population stratification was assessed by calculating genomic inflation factors (λ) in PLINK and conducting ethnic-specific PC analysis in EIGENSTRAT. The first two ethnic-specific PCs were systematically added as covariates to each ethnic-specific linear model; and these PCs were sufficient to control for genomic inflation (all λ below a pre-determined threshold of 1.05).

Associations between individual SNPs and BMI were tested in PLINK using ethnic-specific linear regression analyses, with initial adjustment for the first two PCs; additional adjustment for age and sex; and full adjustment for ethnic-specific covariates. Genome-wide significance was evaluated against ethnic-specific Bonferroni-corrected thresholds, as performed in [34]. The significance thresholds were 5.86×10^{-8} (0.05/853,278), 7.31×10^{-8} (0.05/683,998), 5.73×10^{-8} (0.05/871,948), and 6.67×10^{-8} (0.05/749,659) in Hispanics, Asians, African Americans, and European Americans, respectively.

After identifying SNPs with suggestive p-values ($p < 5 \times 10^{-6}$) in each MESA ethnic group, we evaluated the generalizability of these associations to other ethnic groups. The ± 500 kb flanking regions of the top SNPs were examined across all ethnicities to account for potential ethnic differences in LD patterns.

Lastly, we conducted cross-ethnic meta-analyses using an inverse-variance method in PLINK and calculated the I^2 statistic for each SNP. I^2 values quantify the percentage of variability in effect estimates attributable to heterogeneity rather than to chance alone [35, 36], and, in this context, they can be interpreted as a measure of cross-ethnic heterogeneity.

Replication phase

Suggestive SNPs ($p < 5 \times 10^{-6}$) identified in MESA Hispanics and African Americans were investigated in 3379 Hispanic and 6871 African American women from WHI. Recruitment and selection criteria for WHI have been described previously [37]. Briefly, WHI recruited postmenopausal women aged 50–79 years at 40 U.S. field centers in 1993–1998. WHI also ascertained subject race and ethnicity via a standard questionnaire that adopted

OMB definitions. [As for MESA, our present study uses the term ‘ethnicity’ to refer to the racial-ethnic groups defined in WHI]. Of 161,808 subjects recruited, 12,008 were genotyped for the WHI SNP SHARe project (NCBI dbGaP study accession: phs000200.v10.p3) using the Affymetrix 6.0 SNP array. Of these, 3560 Hispanics and 8359 African Americans had baseline BMI data.

All measurements used herein were obtained at the WHI baseline examination. BMI was calculated from height and weight measurements collected by trained staff. Covariates were selected based on ethnic-specific regression models built for MESA Hispanics and African Americans (Additional file 18: Supplemental Methods.).

SNPs and subjects were excluded from genome-wide SNP analyses if they did not meet the QC criteria outlined previously (Additional file 3: Table S2). Replication results were evaluated against a nominal significance level of 0.05 and same direction of effect.

Additional files

Additional file 1: Table S1. Unadjusted associations between participant characteristics and BMI across the MESA ethnic groups. (XLSX 51 kb)

Additional file 2: Figures S1-S2. Quantile-quantile p-value plots for MESA and WHI, pre- and post- adjustment for population stratification. (DOCX 255 kb)

Additional file 3: Table S2. Summary of Subject and SNP Quality Control Procedures. (XLSX 10 kb)

Additional file 4: Table S3. Top candidate SNP in each MESA ethnic group across all regression models. (XLSX 46 kb)

Additional file 5: Figures S3-S5. Regional association plots for top candidate SNPs and flanking region markers (± 500 kb) in MESA. (DOCX 257 kb)

Additional file 6: Tables S4a-b. Description of the WHI SHARe Hispanic and African American study populations. (XLSX 41 kb)

Additional file 7: Tables S5a-b. Unadjusted associations between participant characteristics and BMI in WHI Hispanics and African Americans. (XLSX 44 kb)

Additional file 8: Table S6. Average BMI by genotype for rs12255372 in MESA and WHI Hispanics. (XLSX 39 kb)

Additional file 9: Table S7. Average BMI by genotype for rs6435678 in MESA and WHI African Americans. (XLSX 40 kb)

Additional file 10: Table S8. Meta-analysis for top candidate SNP in MESA Hispanics, African Americans, and European Americans. (XLSX 46 kb)

Additional file 11: Table S9. Evaluation of heterogeneity of SNP effects by sex for top MESA Hispanic and African American SNPs ($p < 5 \times 10^{-6}$). (XLSX 53 kb)

Additional file 12: Figures S6-S7. Histograms and quantile-quantile plots for ethnic-specific BMI distributions in MESA and WHI. (DOCX 377 kb)

Additional file 13: Table S10. Average BMI by genotype for the top candidate SNP in each MESA ethnic group. (XLSX 41 kb)

Additional file 14: Table S11. Summary of association results and LD structure analysis of *TCF7L2* variants. (XLSX 59 kb)

Additional file 15: Figures S8-S13. Ethnic-specific LD structures of *TCF7L2* in MESA and WHI. (DOCX 739 kb)

Additional file 16: Table S12. Power estimations in MESA and WHI. (XLSX 50 kb)

Additional file 17: Figures S14-S15. PC analyses across the four MESA ethnic groups and across the WHI ethnic groups, plotted in the first two dimensions. (DOCX 622 kb)

Additional file 18: Supplemental Methods. Details regarding ethnic-specific linear regression model covariates and checks for normality of the response variable. (DOCX 184 kb)

Abbreviations

Not applicable.

Acknowledgements

We thank Dr. Michael Bracken (Professor of Chronic Disease Epidemiology, Yale School of Public Health) for helpful discussions throughout the completion of this project.

Funding

Dr. DeWan is supported by grant 1R01HL116742 from the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health. Ms. Salinas is supported by a Diversity Supplement to grant 1R01HL116742. Dr. Wang is supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health and a Ruth L. Kirschstein National Research Service Award Individual Postdoctoral Fellowship (F32; award number 1F32AI114097-01A1).

Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, and CTSA UL1-RR-024156. Funding for SHARe genotyping was provided by NHLBI contract N02-HL-64278.

The WHI program is funded by the National Heart, Lung, and Blood Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. Funding for WHI SHARe genotyping was provided by NHLBI contract N02-HL-64278. The MESA and MESA SHARe project are conducted and supported by the NHLBI in collaboration with the MESA investigators.

None of the funding institutions listed above played any role in the design, analysis or preparation of the manuscript. In addition, this manuscript was not prepared in collaboration with either the MESA or WHI investigators and does not necessarily reflect the opinions of the MESA or WHI investigators or the NHLBI.

Availability of data and material

Genotype and phenotype information used in this study for MESA and WHI are available from the National Center for Biotechnology Information's database of Genotypes and Phenotypes (NCBI dbGaP study accession: phs000209.v11.p3 for MESA SNP Health Association Resource (SHARe) and phs000200.v10.p3 for WHI SNP SHARe).

Author's contributions

YDS and ATD conceived the experiments. YDS and LW carried out the analyses. YDS drafted the manuscript. All authors edited and approved the final version of the manuscript.

Authors' information

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethical approval and consent to participate

This study was approved by the Yale University Human Investigation Committee (HIC) under protocol number 1410014735. This secondary data analysis project was granted expedited approval by the HIC and informed consent was not required.

Received: 20 November 2015 Accepted: 1 June 2016

Published online: 13 June 2016

References

- Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA*. 2014;311(8):806–14.
- Bray GA. Medical consequences of obesity. *J Clin Endocrinol Metab*. 2004;89(6):2583–9.
- Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*. 2011;378(9793):815–25.
- Grundy SM. Multifactorial causation of obesity: implications for prevention. *Am J Clin Nutr*. 1998;67(3):563S–72.
- Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997;27(4):325–51.
- Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet*. 2013;9(5):e1003520.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90(1):7–24.
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197–206.
- Bogardus C. Missing heritability and GWAS utility. *Obesity*. 2009;17(2):209–10.
- Xia Q, Grant SF. The genetics of human obesity. *Ann N Y Acad Sci*. 2013;1281(1):178–90.
- Wang Y, Beydoun MA. The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiol Rev*. 2007;29(1):6–28.
- Fesinmeyer MD, North KE, Ritchie MD, Lim U, Franceschini N, Wilkens LR, Gross MD, Bůžková P, Glenn K, Qibrera PM. Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. *Obesity*. 2013;21(4):835–46.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576):2225–9.
- Holder JL, Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Hum Mol Genet*. 2000;9(1):101–8.
- Kurokawa N, Young EH, Oka Y, Satoh H, Wareham NJ, Sandhu M, Loos RJ. The ADRB3 Trp64Arg variant and BMI: a meta-analysis of 44 833 individuals. *Int J Obes*. 2008;32(8):1240–9.
- Consortium IH. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52–8.
- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320–3.
- Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, Nejjari C, Patsch W, Chikri M, Meyre D. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med*. 2007;85(7):777–82.
- Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, Howson JM, Auton A, Myers S, Morris A. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet*. 2012;44(12):1294–301.
- Sale MM, Smith SG, Mychaleckyj JC, Keene KL, Langefeld CD, Leak TS, Hicks PJ, Bowden DW, Rich SS, Freedman BI. Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes*. 2007;56(10):2638–42.
- Parra E, Cameron E, Simmonds L, Valladares A, McKeigue P, Shriver M, Wacher N, Kumate J, Kittles R, Cruz M. Association of TCF7L2 polymorphisms with type 2 diabetes in Mexico City. *Clin Genet*. 2007;71(4):359–66.
- Ruiz-Narváez EA. Redundant enhancers and causal variants in the TCF7L2 gene. *Eur J Hum Genet*. 2014;22:1243–6.
- Cauchi S, Meyre D, Choquet H, Dina C, Born C, Marre M, Balkau B, Froguel P. TCF7L2 Variation Predicts Hyperglycemia Incidence in a French General Population The Data From an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) Study. *Diabetes*. 2006;55(11):3189–92.
- Stolerman E, Manning A, McAteer J, Fox C, Dupuis J, Meigs J, Florez J. TCF7L2 variants are associated with increased proinsulin/insulin ratios but not obesity traits in the Framingham Heart Study. *Diabetologia*. 2009;52(4):614–20.
- Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, Balkau B, Charpentier G, Pattou F, Stetsyuk V. Transcription Factor TCF7L2 Genetic Study in the French Population Expression in Human β -Cells and Adipose Tissue and Strong Association With Type 2 Diabetes. *Diabetes*. 2006;55(10):2903–8.
- Wang GX, Zhao XY, Meng ZX, Kern M, Dietrich A, Chen Z, Cozocov Z, Zhou D, Okunade AL, Su X, et al. The brown fat-enriched secreted factor Nrg4 preserves metabolic homeostasis through attenuation of hepatic lipogenesis. *Nat Med*. 2014;20(12):1436–43.
- Huotari MA, Miettinen PJ, Palgi J, Koivisto T, Ustinov J, Harari D, Yarden Y, Otonkoski T. ErbB signaling regulates lineage determination of developing pancreatic islet cells in embryonic organ culture. *Endocrinology*. 2002;143(11):4437–46.
- Kritzik MR, Krahl T, Good A, Gu D, Lai C, Fox H, Sarvetnick N. Expression of ErbB receptors during pancreatic islet development and regrowth. *J Endocrinol*. 2000;165(1):67–77.
- Miettinen PJ, Ustinov J, Ormio P, Gao R, Palgi J, Hakonen E, Juntti-Berggren L, Berggren PO, Otonkoski T. Downregulation of EGF receptor signaling in pancreatic islets causes diabetes due to impaired postnatal beta-cell growth. *Diabetes*. 2006;55(12):3299–308.
- Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, Cawley S, Chung E, Connell S, Eshragh J. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*. 2011;98(2):79–89.
- Bild DE, Bluemke DA, Burke GL, Detrano R, Roux AVD, Folsom AR, Greenland P, Jacobs Jr DR, Kronmal R, Liu K. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871–81.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
- Zhao L, Bracken MB, DeWan AT. Genome-wide association study of pre-eclampsia detects novel maternal single nucleotide polymorphisms and copy-number variants in subsets of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study cohort. *Ann Hum Genet*. 2013;77(4):277–87.
- Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–58.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557.
- WHI-Study-Group. Design of the Women's Health Initiative Clinical Trial and Observational Study—examples from the Women's Health Initiative. *Control Clin Trials*. 1998;19(1):61–109.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

