

# UCSF

## UC San Francisco Previously Published Works

### Title

A Systematic Mapping Approach of 16q12.2/FTO and BMI in More Than 20,000 African Americans Narrows in on the Underlying Functional Variation: Results from the Population Architecture using Genomics and Epidemiology (PAGE) Study

### Permalink

<https://escholarship.org/uc/item/0p75173q>

### Journal

PLOS Genetics, 9(1)

### ISSN

1553-7390

### Authors

Peters, Ulrike  
North, Kari E  
Sethupathy, Praveen  
et al.

### Publication Date

2013

### DOI

10.1371/journal.pgen.1003171

Peer reviewed

# A Systematic Mapping Approach of 16q12.2/*FTO* and BMI in More Than 20,000 African Americans Narrows in on the Underlying Functional Variation: Results from the Population Architecture using Genomics and Epidemiology (PAGE) Study

Ulrike Peters<sup>1\*</sup>, Kari E. North<sup>2,3</sup>, Praveen Sethupathy<sup>2,4</sup>, Steve Buyske<sup>5,6</sup>, Jeff Haessler<sup>1</sup>, Shuo Jiao<sup>1</sup>, Megan D. Fesinmeyer<sup>1</sup>, Rebecca D. Jackson<sup>7</sup>, Lew H. Kuller<sup>8</sup>, Aleksandar Rajkovic<sup>9</sup>, Unhee Lim<sup>10</sup>, Iona Cheng<sup>9</sup>, Fred Schumacher<sup>11</sup>, Lynne Wilkens<sup>10</sup>, Rongling Li<sup>12</sup>, Keri Monda<sup>2,3</sup>, Georg Ehret<sup>13,14</sup>, Khanh-Dung H. Nguyen<sup>13</sup>, Richard Cooper<sup>15</sup>, Cora E. Lewis<sup>16</sup>, Mark Leppert<sup>17</sup>, Marguerite R. Irvin<sup>18</sup>, C. Charles Gu<sup>19</sup>, Denise Houston<sup>20</sup>, Petra Buzkova<sup>21</sup>, Marylyn Ritchie<sup>22</sup>, Tara C. Matise<sup>5</sup>, Loic Le Marchand<sup>10</sup>, Lucia A. Hindorff<sup>12</sup>, Dana C. Crawford<sup>23,24</sup>, Christopher A. Haiman<sup>11</sup>, Charles Kooperberg<sup>1</sup>

**1** Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **2** Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **5** Department of Genetics, Rutgers University, Piscataway, New Jersey, United States of America, **6** Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, United States of America, **7** Department of Internal Medicine, Ohio State Medical Center, Columbus, Ohio, United States of America, **8** Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **9** Department of Obstetrics and Gynecology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **10** Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, United States of America, **11** Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, United States of America, **12** Office of Population Genomics, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **13** Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **14** Division of Cardiology, Geneva University Hospital, Geneva, Switzerland, **15** Preventive Medicine and Epidemiology, Loyola University, Chicago, Illinois, United States of America, **16** Department of Medicine, University of Alabama, Birmingham, Alabama, United States of America, **17** Department of Human Genetics, University of Utah, Salt Lake City, Utah, United States of America, **18** Department of Biostatistics, University of Alabama, Birmingham, Alabama, United States of America, **19** Department of Biostatistics, Washington University, St. Louis, Missouri, United States of America, **20** Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **21** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **22** Pennsylvania State University, State College, Pennsylvania, United States of America, **23** Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **24** Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

## Abstract

Genetic variants in intron 1 of the fat mass- and obesity-associated (*FTO*) gene have been consistently associated with body mass index (BMI) in Europeans. However, follow-up studies in African Americans (AA) have shown no support for some of the most consistently BMI-associated *FTO* index single nucleotide polymorphisms (SNPs). This is most likely explained by different race-specific linkage disequilibrium (LD) patterns and lower correlation overall in AA, which provides the opportunity to fine-map this region and narrow in on the functional variant. To comprehensively explore the 16q12.2/*FTO* locus and to search for second independent signals in the broader region, we fine-mapped a 646-kb region, encompassing the large *FTO* gene and the flanking gene *RPGRIP1L* by investigating a total of 3,756 variants (1,529 genotyped and 2,227 imputed variants) in 20,488 AAs across five studies. We observed associations between BMI and variants in the known *FTO* intron 1 locus: the SNP with the most significant p-value, rs56137030 ( $8.3 \times 10^{-6}$ ) had not been highlighted in previous studies. While rs56137030 was correlated at  $r^2 > 0.5$  with 103 SNPs in Europeans (including the GWAS index SNPs), this number was reduced to 28 SNPs in AA. Among rs56137030 and the 28 correlated SNPs, six were located within candidate intronic regulatory elements, including rs1421085, for which we predicted allele-specific binding affinity for the transcription factor *CUX1*, which has recently been implicated in the regulation of *FTO*. We did not find strong evidence for a second independent signal in the broader region. In summary, this large fine-mapping study in AA has substantially reduced the number of common alleles that are likely to be functional candidates of the known *FTO* locus. Importantly our study demonstrated that comprehensive fine-mapping in AA provides a powerful approach to narrow in on the functional candidate(s) underlying the initial GWAS findings in European populations.

**Citation:** Peters U, North KE, Sethupathy P, Buyske S, Haessler J, et al. (2013) A Systematic Mapping Approach of 16q12.2/*FTO* and BMI in More Than 20,000 African Americans Narrows in on the Underlying Functional Variation: Results from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS Genet* 9(1): e1003171. doi:10.1371/journal.pgen.1003171

**Editor:** Mark I. McCarthy, University of Oxford, United Kingdom

**Received:** September 12, 2012; **Accepted:** October 22, 2012; **Published:** January 17, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802). The MEC study is funded through the National Cancer Institute (R37CA54281, R01 CA63, P01CA33619, U01CA136792, and U01CA98758). Funding support for the "Epidemiology of putative genetic variants: The Women's Health Initiative" study is provided through the NHGRI PAGE program (U01HG004790). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: [http://www.whiscience.org/publications/WHI\\_investigators\\_shortlist.pdf](http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf). The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022. PS was supported in part by NIDDK/NIH K99 grant 1K99DK091318-01. Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801). The National Institutes of Mental Health also contributes to the support for the Coordinating Center. The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript, except for RL and LAH, who as part of their role in the NHGRI PAGE program (U01HG004802, U01HG004790, U01HG004801) collaboratively worked on this manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [upeters@fhcrc.org](mailto:upeters@fhcrc.org)

## Author Summary

Genetic variants within the fat mass- and obesity-associated (*FTO*) gene are associated with increased risk of obesity. To better understand which specific genetic variant(s) in this genetic region is associated with obesity risk, we attempt to genotype or impute all known genetic variants in the region and test for association with body mass index as a measurement of obesity in over 20,000 African Americans. We identified 29 potential candidate variants, of which one variant (rs1421085) is a particularly interesting candidate for future functional follow-up studies. Our example shows the powerful approach of studying a large African American population, substantially reducing the number of possible functional variants compared with European descent populations.

## Introduction

The association between variants in the fat mass and obesity associated (*FTO*) gene on chromosome 16q12.2 and body mass index (BMI) is well-established in populations of European descent. Genome-wide association studies (GWAS) and subsequent replication studies have identified several strongly correlated single nucleotide polymorphisms (SNPs) located in intron 1 of *FTO* associated with increased BMI and increased risk of obesity [1–13]. With an observed effect size of 0.35 kg/m<sup>2</sup> (0.1 z-score units of BMI) per risk allele, the *FTO* locus has a substantially stronger effect on BMI than any other identified common locus [14]. While this impact on BMI may seem small, it has a potential public health bearing, as even a 1 unit increase in BMI results in an estimated 8% increase in coronary heart disease [15], and excess weight in midlife is associated with increased mortality [16]. Thus, a seemingly small increase in BMI can have a marked impact, particularly in countries with an increasing burden of excess weight, such as the US, where an estimated 68% of adults were overweight or obese in 2008 [17].

Studies in non-European populations have had varied success in replicating the findings at the *FTO* locus. While several studies showed an association between *FTO* SNPs and obesity-related phenotypes in Hispanic [5,6,18,19] and Asian populations [20–26], studies of African or African American (AA) subjects showed limited support for some of the most consistent *FTO* GWAS

findings initially identified in subjects of European descent [2,5,6,14,18,19,27–32]. This lack of generalization in AA may be attributable to the lower levels of linkage disequilibrium (LD) with the underlying functional variant(s) at the 16q12.2/*FTO* locus, as compared with European Americans (EAs). SNPs discovered in GWAS (i.e. "index SNPs") are often not the functional variants; however, they do tag genomic regions harboring strongly correlated variants, one or more of which are the potentially functional variant(s). Because different ancestral populations differ in their LD patterns, an index SNP discovered in one ancestral group (e.g. EA) may or may not be strongly correlated with the functional variant(s) in a different ancestral group (e.g. AA). Thus, the index SNP may not show evidence for replication in AAs; however, other SNPs in the region may be in high LD with the functional variant(s), and, hence, measuring these SNPs may further characterize associations within the genomic region. Therefore, a full exploration of potential replication/generalizability of GWAS findings in other ancestral groups requires investigating not only at the index SNP(s), but also examining, if possible, all variants of the region tagged by the index SNP(s). African populations are particularly suited for these studies because the LD pattern between SNPs tends to be substantially weaker than in other ancestral groups, as has been demonstrated for the *FTO* gene [30]. This process can reduce the number of potential functional variants for follow-up molecular investigation [33]. Given that functional studies can be labor- and cost-intensive, narrowing the associated region is an important step toward elucidating the underlying molecular mechanism. While molecular evaluation of the 16q12.2/*FTO* locus provides some promising leads [34], the putative functional variant(s) in this locus remain under investigation, and fine-mapping studies have been limited with respect to the number of tested variants, sample sizes and inclusion of non-EA populations. Fine-mapping studies, particularly when conducted within a broader region, may also identify additional independent signal(s) implicating multiple functional variants in the region. To better understand the relationship of genetic variation at this locus and BMI in AAs, we comprehensively assessed the association of BMI with variants in the 16q12.2/*FTO* region and flanking gene *RPGRIPL* in over 20,000 AA samples from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. We used the Metachip as genotyping platform [35], which included all suitable SNPs discovered in the 1000 Genomes Project Pilot 1 for the 16q12.2/*FTO* region and led to successfully genotyping of over

**Table 1.** Distribution of study characteristics by study and overall.

Characteristics	ARIC (n = 3,297)		GenNet (n = 517)		HyperGen (n = 1,171)		MEC (n = 3,865)	
	Male	Female	Male	Female	Male	Female	Male	Female
<b>Sex</b>	1,224 (37)	2,073 (63)	221(43)	296 (57)	391 (33)	780 (67)	1,037 (27)	2,828 (73)
<b>Age</b>								
<b>Mean (+/-SD)</b>	53.8 (6.0)	53.4 (5.7)	38.2(7.1)	38.4(8.0)	47.6 (12.5)	48.0 (12.4)	61.9 (8.0)	58.6 (8.8)
<b>Range (max-min)</b>	44–66	44–66	22–64	20–62	21–85	20–81	45–76	45–77
<b>Height (m)</b>								
<b>Mean (+/-SD)</b>	176.2 (6.7)	163.2 (6.2)	176.9(7.2)	164.3(6.4)	175.8 (7.0)	162.4(6.2)	175.0 (6.5)	161.3 (6.4)
<b>Range (max-min)</b>	153–197	125–188	155–196	138–183	155–198	145–183	153–198	133–198
<b>Weight (kg)</b>								
<b>Mean (+/-SD)</b>	86.7 (16.5)	82.4 (17.8)	86.8(22.8)	88.8(25.2)	93.0(21.9)	89.3(22.0)	86.3 (14.3)	77.6 (16.6)
<b>Range (max-min)</b>	50.9–165.9	44.5–177.3	54.9–204.3	49.5–175.2	54.9–181.4	45.4–210.0	49.0–179.3	43.6–181.6
<b>BMI (kg/m<sup>2</sup>)</b>								
<b>Mean (+/-SD)</b>	27.9 (4.9)	31.0 (6.5)	27.4(6.9)	32.9(9.2)	30.0(6.3)	33.7(7.6)	27.3 (4.2)	28.9 (6.0)
<b>Range (max-min)</b>	18.6–54.4	18.5–65.9	18.8–62.8	18.9–63.2	18.6–52.4	19.4–62.7	18.6–60.2	18.6–64.7
<b>Obesity (BMI ≥ 30), n (%)</b>	360 (29)	1010 (49)	63 (29)	161 (54)	181(46)	499(64)	212 (20)	990 (35)
Characteristics	WHI genotyped		WHI imputed		All (n = 20,488)			
	Female (n = 5,312)		Female (n = 6,326)		Male	Female	Both	
<b>Sex</b>	5,312 (100)		6,326 (100)		2,873 (14)	17,615 (86)	20,488 (100)	
<b>Age</b>								
<b>Mean (+/-SD)</b>	61.2(7.2)		61.8 (7.1)		54.7 (8.0)	59.1 (7.6)	58.5 (7.7)	
<b>Range (max-min)</b>	50–79		50–79		22–85	20–81	20–85	
<b>Height (m)</b>								
<b>Mean (+/-SD)</b>	162.6 (6.3)		162.4 (6.4)		175.7 (6.7)	162.4 (6.3)	164.3 (6.4)	
<b>Range (max-min)</b>	100–192		117–194		153–198	100–198	100–198	
<b>Weight (kg)</b>								
<b>Mean (+/-SD)</b>	83.3 (18.4)		81.9 (17.7)		87.4 (17.2)	82.1 (18.1)	82.9 (18.0)	
<b>Range (max-min)</b>	45.0–187.0		39–181		49.0–204.3	39–210	39.0–210	
<b>BMI (kg/m<sup>2</sup>)</b>								
<b>Mean (+/-SD)</b>	31.5 (6.8)		31.0 (6.4)		27.9 (5.1)	31.0 (6.6)	30.5 (6.4)	
<b>Range (max-min)</b>	18.6–68.3		18.6–69.2		18.6–62.8	18.5–69.2	18.5–69.2	
<b>Obesity (BMI ≥ 30), n (%)</b>	2806 (53)		3185 (50)		816 (28)	8651 (49)	9467 (46)	

doi:10.1371/journal.pgen.1003171.t001

1,500 SNPs. This together with imputation into the updated 1000 Genomes Project allowed us to densely fine-mapping this region. In addition, we performed a detailed bioinformatic analysis to propose candidate polymorphisms for follow-up functional evaluation.

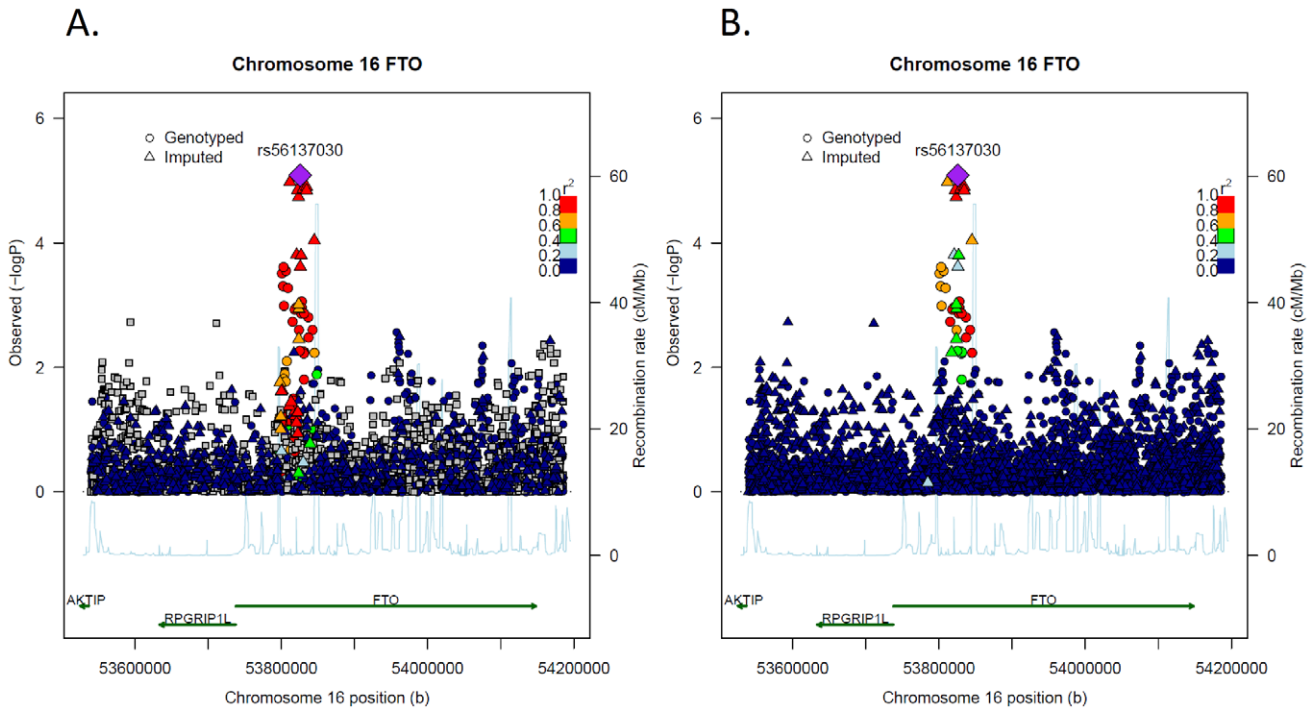
## Results

The age of the 20,488 participants ranged from 20 to 85 years with an average age of 58.5 years across the cohorts (Table 1). The fraction of men included in each cohort varied from 0% to 43%. In all studies that included both genders, men had on average a lower BMI than the women. Participants in the Hypertensive Genetic Epidemiology Network (HyperGEN) had the highest BMI while participants in the Multiethnic Cohort (MEC) had the lowest BMI. The obesity rate (defined as BMI ≥ 30 kg/m<sup>2</sup>) across the studies was 46% ranging from 31% to 58%.

The targeted 16q12.2/*FTO* fine-mapping region spans 646 kb from 53,539,509 to 54,185,773 (build 37) on the long arm of chromosome 16 (16q12.2), including the large *FTO* gene (411 kb

as well as 198 kb downstream of *FTO*, which includes the *RPGRIPII* gene and 37 kb upstream of *FTO* (Figure 1). On average, we successfully genotyped or imputed one SNP per 172 bp (= 646,264 bp/3,756 SNPs). The allele frequency distribution of the 3,756 SNPs in this region is shown in Table 2. In contrast to GWAS platforms, a large fraction of SNPs have allele frequencies <5% (52.5%), including 16.7% with allele frequency <1%. Twenty-one SNPs are within the exons of *FTO* and *RPGRIPII*, of which 7 are synonymous and 14 are missense (2 synonymous and 4 missense are located in *FTO*).

The most significant SNPs in the 16q12.2/*FTO* fine-mapping region was rs56137030 (p-value  $8.3 \times 10^{-6}$ ), which showed no evidence for heterogeneity (p = 0.13; Table 3). Each A allele of this variant (allele frequency = 0.12) increased BMI by 1.35% (95% confidence interval (0.76%–1.95%). Table 3 also displays results for the three next most significant SNPs showing similar significant associations (p-values  $1.4 \times 10^{-5}$  to  $1.1 \times 10^{-5}$ ) and Table S1 shows results separately for each study. All three SNPs were correlated with rs56137030 ( $r^2 \geq 0.73$  in AA and  $r^2 \geq 0.91$  in EA; Table S2).



**Figure 1. Association results (p-values) and correlation structure for all SNPs in the 16q12.2/*FTO* region and lnBMI among African Americans using rs56137030 to calculate correlation among SNPs (LocusZoom plots).** The top half of each figure has physical position along the x axis, and the  $-\log_{10}$  of the meta-analysis p-value on the y-axis. Each dot on the plot represents the p-value of the association for one SNP with lnBMI across all studies. The most significant SNP (rs56137030) is marked as a purple diamond. The color scheme represents the pairwise correlation ( $r^2$ ) for the SNPs across the 16q12.2/*FTO* region with the most significant SNP (rs56137030). Gray squares indicate that correlation was missing for this p-value because the variant was monomorphic in EA. The bottom half of the figure shows the position of the genes across the region. A and B show the same region and results. The only difference between A and B is that in A correlation with the most significant SNP (rs56137030) was calculated based on EAs, specifically based on data from 65 European Americans (Utah residents with Northern and Western European ancestry from the CEPH collection, CEU) sequenced as part of the 1000 Genomes Project and B correlation was based on 61 African Americans from the South-west (ASW) and sequenced as part of the 1000 Genomes Project. doi:10.1371/journal.pgen.1003171.g001

rs56137030 and the three next most significant SNPs are located in intron 1 of the *FTO* gene, approximately in the middle of the region (Figure 1). The nine GWAS index SNPs previous highlighted in EA studies are also located in the same intron 1 *FTO* region and results for these nine GWAS index SNPs are provided in Table 3 showing p-values < 0.05 for five out of the nine index SNPs ( $0.03$  to  $3.0 \times 10^{-4}$ ).

The most significant variant, rs56137030, was correlated at  $r^2 > 0.5$  with 103 SNPs in Europeans (Figure 1a), including eight of the nine GWAS index SNPs (Table S2). In contrast in AAs rs56137030 was correlated with only 28 SNPs at  $r^2 > 0.5$  (Table S2, Figure 1b). All 28 variants correlated with rs56137030 in AA showed some evidence of association with BMI (p-values 0.0057 to  $1.1 \times 10^{-5}$ ) and no or limited evidence of heterogeneity (all p for heterogeneity > 0.04). To investigate if any of these correlated SNPs were associated with BMI independently from rs56137030 we adjusted each SNP for rs56137030 (including rs56137030 and a second SNP simultaneously in one model). None of these variants remained significant at  $p < 0.05$  (Table S2). As expected, the p-values of rs56137030 were also less significant in these conditional analyses, particularly for SNPs highly correlated in AA, demonstrating that these findings are not independent. All 28 SNPs correlated with rs56137030 ( $r^2 > 0.5$ ) in AA are located between 53,800,954 and 53,845,487, spanning a 44.5 kb region about 104.8 kb downstream of the exon 1 boundary, and ending about 1.4 Kb after exon 2 ends [no variant in exon 2 was genotyped or imputed and, to our knowledge, only 3 variants (rs116753298,

**Table 2. Distribution of allele frequency of SNPs in the *FTO* fine-mapping region.**

SNP	n	%
<b>Allele frequency<sup>a</sup></b>		
0.14%–1%	627	16.7%
>1%–5%	1346	35.8%
>5%–10%	431	11.5%
>10%–25%	679	18.1%
>25%	673	17.9%
Total	3756	100.0%
<b>Exonic SNPs<sup>b</sup></b>		
<u>SNPs in <i>RPGRIP1L</i> exons</u>		
synonymous SNPs in <i>RPGRIP1L</i>	3	0.1%
missense SNPs in <i>RPGRIP1L</i>	12	0.3%
<u>SNPs in <i>FTO</i> exons</u>		
synonymous SNPs in <i>FTO</i>	4	0.1%
missense SNPs in <i>FTO</i>	2	0.1%

<sup>a</sup>Based on allele frequency for all AA participants in this study.

<sup>b</sup>Based on data from the Exome Variation Server (<http://snp.gs.washington.edu/EVS/> accessed August 2012).

doi:10.1371/journal.pgen.1003171.t002

**Table 3.** Association between SNPs in the *FTO* region and BMI for all studies combined.

SNP rs#	Position <sup>a</sup>	Alleles <sup>b</sup>		CAF <sup>c</sup>	% change in BMI per coding allele		nom.p	Rsq <sup>d</sup>	p.het	
		Coding <sup>b</sup>	Baseline <sup>b</sup>		Beta estimate	95%CI				
<b>Five most significant variants</b> (all in the <i>FTO</i> intron 1 region)										
rs56137030	53825905	A	G	0.12	1.35	(0.76,1.95)	8.3E-06	0.98	1.3E-01	
rs62033400	53811788	G	A	0.12	1.35	(0.76,1.95)	1.1E-05	0.99	1.3E-01	
rs7188250	53834607	C	T	0.12	1.34	(0.75,1.94)	1.3E-05	0.97	8.3E-02	
rs62033413	53830055	G	C	0.12	1.33	(0.74,1.93)	1.4E-05	0.98	4.5E-02	
<b>Index SNPs of GWAS</b> (variants highlighted in previous studies of EA; all in <i>FTO</i> intron 1 region)										
rs9939609	53820527	T	A	0.52	0.04	(-0.33,0.41)	8.2E-01	1.00	5.5E-01	
rs8050136	53816275	A	C	0.44	0.42	(0.03,0.82)	3.2E-02	1.00	4.9E-02	
rs1421085	53800954	G	A	0.12	1.11	(0.49,1.72)	3.0E-04	1.00	2.6E-01	
rs17817449	53813367	C	A	0.39	0.37	(-0.02,0.77)	5.9E-02	1.00	5.6E-02	
rs1121980	53809247	A	G	0.47	0.35	(-0.02,0.73)	7.2E-02	1.00	2.4E-01	
rs1558902	53803574	T	A	0.12	1.13	(0.46,1.80)	1.0E-03	1.00	2.4E-01	
rs6499640	53769677	A	G	0.65	0.05	(-0.34,0.44)	8.1E-01	1.00	1.6E-01	
rs9930506	53830465	G	A	0.22	0.66	(0.19,1.14)	5.5E-03	1.00	1.6E-01	
rs9941349	53825488	A	G	0.12	0.69	(0.20,1.19)	5.4E-03	1.00	5.7E-01	

<sup>a</sup>SNPposition based on build 37.

<sup>b</sup>Coding = coding allele, Base = baseline allele (risk estimates provide the log additive effect per copy of the coding allele).

<sup>c</sup>CAF = coding allele frequency.

<sup>d</sup>Rsq = measurement of imputation accuracy, ranging from 0 (low) to 1 (high).

doi:10.1371/journal.pgen.1003171.t003

rs149393601, and chr16:53844100) all with allele frequency <0.1% have been reported in *FTO* exon 2 so far [36,37]].

To predict the molecular mechanisms underlying the genetic association signals, and to identify candidate variants for functional follow-up, rs56137030 and variants in LD ( $r^2 > 0.5$ ;  $n = 28$ ) were assessed for overlap with eleven different genome-wide functional annotation datasets (Table S3, Material and Methods). Among these 29 variants, six (rs11642015, rs17817497, rs3751812, rs17817964, rs62033408, and rs1421085) were located within candidate intronic regulatory elements, including two (rs3751812 and rs1421085) that were within highly sequence-conserved elements among vertebrates, and two (rs11642015 and rs1421085) that were predicted to have allele-specific binding affinities for different transcription factors. Specifically, we predicted that only the T allele at rs11642015 binds Paired box protein 5 (*PAX5*) and that the C allele at rs1421085 has a substantially reduced binding affinity for Cut-like homeobox 1 (*CUX1*; Figure S1).

Outside the known *FTO* intron 1 region, we observed no strong evidence for a second independent signal: When we adjusted each SNP for rs56137030 the most significant SNPs outside of the *FTO* intron 1 region was a SNP located at position 53710931 (no rs number reported) intronic of the neighbor gene *RPGRIP1L* (conditional p-value  $7.7 \times 10^{-4}$ ) followed by rs8051873 in intron 8 of *FTO* (conditional p-value 0.0011).

Table S4 shows results for variants highlighted in previous AA studies. Of these nine SNPs only two SNPs had a p-values < 0.05 (rs3751812, p value 0.0012).

## Discussion

In this large study of over 20,000 AAs we densely fine mapped the entire *FTO* gene and adjacent *RPGRIP1L* gene, spanning a total region

of almost 650 kb. We observed significant associations for variants in the known locus in intron 1 of *FTO*. Due to reduced correlation in AA compared with EA with the most significant SNPs, we were able to substantially reduce the number of functional candidates. Six SNPs were located within candidate intronic regulatory elements, including rs1421085, for which we predicted allele-specific binding affinity for the transcription factor *CUX1*. Because we did not focus solely on the known *FTO* intron 1 region, we were able to comprehensively investigate the region; however, this approach revealed no evidence for a second independent signal in AA.

This is one of the first studies of the MetaboChip in an ancestral group that is particularly well suited for fine-mapping GWAS loci, due to its distinct linkage disequilibrium (LD) patterns and lower LD overall. Our example clearly shows the powerful approach of studying a large AA population, substantially reducing the number of possible functional variants compared with European descent populations. While very large number of Europeans and EAs are genotyped on the MetaboChip, the number of Minority populations genotyped on this chip is substantially smaller; however, this is the focus of the PAGE Study. As many papers using the MetaboChip in European populations will be published over the next years, our results show the important contribution that Minority populations and, in particularly AA, will have for systematic mapping of GWAS loci.

The importance of the 16q12.2/*FTO* locus for obesity-related traits was identified in genome-wide scans of Europeans. These scans highlighted several variants within the *FTO* intron 1, all of which, except for rs6499640, are in high LD with each other in EA [1–7]. Consistently these studies showed an increase in BMI (~1.1% to 1.3% per risk allele; [7]). However, AA studies showed very limited or no evidence for an association with rs9939609 [2,5,6,14,18,28–31], rs1121980 [18,19,28,30,31], rs17817449 [18,19,28–31], or the previously reported functional variant

rs8050136 [2,6,19,27–29,31,32]. We observed nominal evidence for association with BMI for rs17817449 and rs8050136, but results were not among the most significant associations. rs1421085 was significantly associated in our study ( $p = 3 \times 10^{-4}$ ; Table S2) and was also found to be associated in the study from Nock et al. [29] ( $n = 469$ ,  $p\text{-value} = 7 \times 10^{-4}$ ), Hassanein et al. [31] ( $n = 4,217$ ,  $p\text{-value} = 3 \times 10^{-4}$ ), and Hester et al. [32] ( $n = 4,992$ ,  $p\text{-value} = 0.07$ ) but not with four smaller AA studies ( $\leq 1000$  AA subjects) [5,18,28,38]. Consistent with our finding Hassanein et al. [31] also observed a significant association with rs9941349 and rs1558902 ( $n = 9,881$ ,  $p\text{-value} = 4 \times 10^{-6}$  and  $n = 4,217$ ,  $p\text{-value} = 2 \times 10^{-5}$ , respectively), these SNPs have not been genotyped in other AA studies. However, to put these findings into context with other variants in this region a comprehensive evaluation of all variants is needed.

To conduct a more comprehensive evaluation of the *FTO* locus, some AA studies extended the SNPs list from the EA index SNPs described above (Table S4). Grant et al. [27] analyzed eleven *FTO* SNPs genotyped as part of their GWAS in about 2,000 AA children and only rs3751812 showed a marginally significant association with BMI ( $p\text{-value} = 0.02$ ). Wing et al. [18,19] genotyped up to 27 SNPs in the intron 1 of *FTO* in a cohort study and family study including 288 and 604 AA, respectively, and observed an association of BMI with rs1108102 ( $p\text{-value} = 5 \times 10^{-4}$ ; [18]); however, this finding was not confirmed in their cohort study or in our larger meta-analysis (Table S4). A fine mapping study of 47 tagging SNPs in 497 AA children [28] identified an association of rs8057044 ( $p\text{-value} = 5 \times 10^{-4}$ ) with BMI, which was not replicated by the current study. In two fine-mapping studies [5,14], no *FTO*-BMI associations were noted, possibly because the majority of subjects were lean ( $\sim 75\%$  had a BMI between 18–25) in one study [14], or because the sample size was small in both studies (about 1,100). Hassanein et al. [31] genotyped 34 tagging SNPs in the *FTO* intron 1 region in 4,217 AA and followed up findings from two variants (rs3751812,  $p\text{-value} = 4 \times 10^{-4}$  and rs9941349,  $p\text{-value} = 6 \times 10^{-5}$ ) in four additional studies ( $n = 5,664$ ), adding some support ( $p\text{-values}$  ranging from 0.016 to 0.64) which resulted in overall  $p\text{-values}$  of  $2.6 \times 10^{-6}$  and  $3.6 \times 10^{-6}$ , respectively. The authors concluded that this finding reduced the potential functional variants to those correlated with these two variants. Both variants (rs3751812 and rs9941349) were also associated in our meta-analysis ( $p\text{-value}$  0.001 and 0.005, respectively; Table S4), although they were not among the most significant findings. In summary, except for Hassanein et al. [31], studies in AA were relatively small and showed mixed results that were mainly not replicated in our study. As the number of variants genotyped in any AA study was limited (10 to 50 SNPs), we were able to investigate if genotyping or imputing additional variants in this region (in total 3,756) may even further reduce the list of possible functional variants and search for second independent signal(s).

In our analysis, rs56137030 was most significantly associated with BMI. In EA, rs56137030 has a similarly high allele frequency as the previously reported GWAS index SNPs (A allele frequency = 0.42) and was highly correlated with the GWAS index SNPs ( $r^2 \geq 0.87$ ), except for rs6499640 ( $r^2 = 0.12$ ), which is also not strongly correlated with any of the other GWAS index SNPs in EA or AA. In AA, the correlation between rs56137030 and GWAS index SNPs varied substantially ( $r^2 = 0.001$  to 0.73), demonstrating that studying AA can substantially reduce the bin of correlated SNPs defined by the index signal identified in EA. Specifically, the number of SNPs correlated with rs56137030 at  $r^2 > 0.5$  was 103 in EA, but only 28 in AA. Including these 28 SNPs together with rs56137030 in conditional analyses showed that the significance of

each of the SNPs, as well as rs56137030, was substantially reduced, supporting the idea that any of the highly correlated SNPs is a potential functional candidate. rs56137030 and all 28 highly correlated SNPs are non-coding, suggesting that the functional variant is likely to have a *cis*-regulatory effect. Among these variants, six are located within the candidate intronic regulatory elements, two of which are highly conserved, and two that are predicted to confer allele-specific binding affinities for transcription factors. The variant rs1421085 is within a highly conserved element and may be particularly interesting, because the C allele has a substantially reduced binding affinity for *CUX1*, which has been previously implicated in the transcriptional regulation of *FTO* [34]. Accordingly, this variant is a compelling candidate for follow-up functional evaluation, though outside the scope of the present study.

We did not observe any evidence for a second independent signal within the broader 16q12.2/*FTO* region. This finding is consistent with the only other AA study that extended the fine-mapping approach beyond intron 1 to the entire *FTO* gene including 262 tagSNPs in 1,485 subjects [30]. However, even within our substantially larger study of over 20,000 AAs, power to identify second independent signals may still have been limited, particularly for less frequent variants or variants with weaker effects, given the increased burden of multiple comparisons that needs to be adjusted for when testing all SNPs across the entire region.

Several limitations warrant consideration to inform fine-mapping and functional characterization studies. To comprehensively evaluate the region we not only included directly genotyped SNPs, which included all SNPs known at the time of the chip development and suitable for genotyping SNPs but we also imputed to the most recent version of the 1000 Genomes Project. While this approach provides a rather complete list of variants imputed SNPs tend to be called with varying accuracy. To account for the imputation accuracy we used the dosage, which we showed results in unbiased estimates [39]. However, we also note that the overall  $p\text{-value}$  of a SNP is impacted by the imputation accuracy (lower imputation accuracy results in higher  $p\text{-values}$ ). Accordingly, it is important for the interpretation of the results that not only the most significant SNPs will be considered as functional variants but also those correlated with the most significant SNPs as done in this paper. Second, for a part of the WHI samples directly genotyped SNPs were only available from a smaller subset of SNPs as genotyping in this subset was based on a GWAS platform and not the dense MetaboChip. However, the imputation  $R_{sq}$  as a measurement of the accuracy was very high (Table S1). Third, despite the relative large samples size of over 20,000 AA the statistical significance of the finding is relative weak compared to previous studies in European descent populations for the *FTO* region. We note that the relative weak power of our study is not due to differences in the observed effect size, e.g. we observed a 1.35% change in BMI per allele for the most significant SNP while the replication stage of Willer et al. [7] observed a 1.25% change in BMI per allele of their most significant *FTO* SNP (rs9939609). However, the substantially lower allele frequency in AA compared with EA (12% vs. 41%) and the larger variance of BMI in AA populations (e.g. the standard deviation in our study was 6.4 kg/m<sup>2</sup> compared with 4.2 kg/m<sup>2</sup> in European populations [7]) explains the reduce power. Fourth, our functional characterization is based on in silico analyses and requires experimental validation. Finally, the majority of study participants were female and it is unclear how a predominantly female population may have influenced the results.

To our knowledge, this is the largest and most comprehensive fine-mapping study conducted to date in AA. Our findings likely rule out that several of the EA index SNPs in intron 1 of *FTO* such as rs9939609, as well as a large fraction of SNPs correlated in EA but not in AA are the underlying functional variants. With rs56137030 and its correlated SNPs, our finding points us closer to the functional variant(s). Among these, rs1421085 is the most compelling candidate for follow-up functional evaluation. Importantly, our study demonstrated that comprehensive fine-mapping in AA provides a powerful approach to narrow in on the functional candidate(s) underlying the initial GWAS findings in EA.

## Materials and Methods

### Ethics statement

All studies were approved by Institutional Review Boards at their respective sites, and all participants provided informed consent.

### Study populations

PAGE involves several studies, described briefly below and in more detail in Text S1 as well as at the PAGE website (<https://www.pagestudy.org>) [40]. In brief, participants were recruited from Atherosclerosis Risk in Communities Study (ARIC), GenNet, Hypertension Genetic Epidemiology Network (HyperGEN), Multiethnic Cohort (MEC), and Women's Health Study (WHI). ARIC randomly selected and recruited 15,792 participants aged 45–64 at four U.S. communities [41]. GenNet and HyperGEN are two family-based studies designed to investigate the genetics of hypertension and related conditions [42]. The Multiethnic Cohort (MEC) is a population-based prospective cohort study of over 215,000 men and women in Hawaii and California aged 45–75 at baseline (1993–1996) and primarily of five ancestries [43]. The WHI encompasses four randomized clinical trials as well as a prospective cohort study of 161,808 post-menopausal women aged 50–79, recruited (1993–1998) and followed up at 40 centers across the US [44]. All studies collected self-identified racial/ethnic group via questionnaire. We selected all AA participants from ARIC, HyperGEN, and GenNet for genotyping. In MEC, a subset of AAs was selected based on availability of biomarker or as controls for nested case-control studies. WHI included all AAs who provided consent for DNA analysis. We excluded underweight ( $BMI < 18.5 \text{ kg/m}^2$ ) and extremely overweight ( $BMI > 70 \text{ kg/m}^2$ ) individuals with the assumption that these extremes could be attributable to data coding errors, an underlying illness or possibly to a familial syndrome and hence, a rare mutation. We also limited analysis to adults defined as age  $> 20$  years.

### Anthropometric measurements

For ARIC HyperGEN, GenNet and WHI, BMI was calculated from height and weight measured at time of study enrollment. In MEC, self-reported height and weight were used to calculate baseline BMI. A validation study within MEC has shown high validity of self-reported height and weight. Specifically this study showed that BMI was under-estimated based on self-reported compared to measured weight, but the difference was small ( $< 1$  BMI unit) and comparable to the findings from national surveys [45].

### SNP selection for genotyping

SNPs were included as part of the Metachip, a 200 k Illumina customized iSelect array developed through the collaborative efforts of several consortia working on metabolic syndrome related

diseases. Details on the design can be found elsewhere (<http://www.sph.umich.edu/csg/kang/MetaboChip/>). In brief, SNPs within the 16q12.2/*FTO* region were selected based on 1000 Genomes Pilot 1 and HapMap phase 2. The boundaries around each GWAS index SNP were determined by identifying all SNPs with  $r^2 \geq 0.5$  with the index SNP, and then expanding the initial boundaries by 0.02 cM in either direction using the HapMap-based genetic map [46]. The total interval size of the 16q12.2/*FTO* region was 646 kb. All 1000 Genomes Pilot 1 SNPs obtained from Sanger Institute (August 12, 2009) and the Broad Institute (August 11, 2009) were considered as potential fine mapping SNPs, unless SNP allele frequency was  $< 0.01$  in all three HapMap samples (CEU, YRI and HBC/JPT). SNPs were excluded if (a) the Illumina design score was  $< 0.5$  or (b) there were SNPs within 15 bp in both directions of the SNP of interest with allele frequency of  $> 0.02$  among Europeans (CEU). SNPs annotated as nonsynonymous, essential splice site, or stop codon were included regardless of allele frequency, design score, or nearby SNPs in the primer.

### Genotyping and quality control

Samples were genotyped on the Metachip at the Human Genetics Center of the University of Texas-Houston (ARIC, GenNet and HyperGEN), the University of Southern California Epigenome Center (MEC), and the Translational Genomics Research Institute (WHI). Each center also genotyped 90 HapMap YRI (Yoruba in Ibadan, Nigeria) samples to facilitate cross-study quality control (QC), as well as 2–3% study-specific blinded replicates to assess genotyping reproducibility. Genotypes were called separately for each study using GenomeStudio with the GenCall 2.0 algorithm. Samples were called using study-specific cluster definitions (based on samples with call rate  $> 95\%$ , ARIC, MEC, WHI) or cluster definitions provided by Illumina (GenNet, HyperGEN) and kept in the analysis if call rate was  $> 95\%$ . We excluded SNPs with GenTrain score  $< 0.6$  (ARIC, MEC, WHI) or  $< 0.7$  (GenNet, HyperGEN), cluster separation score  $< 0.4$ , call rate  $< 0.95$ , and Hardy-Weinberg Equilibrium  $p < 1 \times 10^{-6}$ . We also excluded SNPs based on Mendelian errors in 30 YRI trios  $> 1$ , replication errors  $> 2$  with discordant calls (when comparing across studies) in 90 YRI samples  $> 3$ , and discordant calls for 90 YRI genotyped in PAGE versus HapMap database  $> 3$ . In total, we successfully genotyped 1,694 out of 1,818 variants in the 16q12.2/*FTO* region. After excluding 165 SNPs that were monomorphic or had very low allele frequency ( $< 0.01\%$ ) we included 1,529 variants in the analysis.

For ARIC, MEC and WHI we identified related persons using PLINK by estimating identical-by-descent (IBD) statistics for all pairs. When apparent first-degree relative pairs were identified, we excluded from each pair the member with the lower call rate. We excluded from further analysis samples with an inbreeding coefficient (F) above 0.15 (ARIC, MEC, WHI) [47]. We determined principal components of ancestry in each study separately using EIGENSOFT [48,49] and excluded apparent ancestral outliers from further analysis as described elsewhere [50]. In total 240 subjects failed genotyping (ARIC = 27, GenNet = 9, HyperGEN = 26, MEC = 140, and WHI = 27). After further excluding subjects based on age and BMI (see above), a total of 14,162 subjects with Metachip data were included (3,297 from ARIC, 517 from GenNet, 1,171 from HyperGEN, 3,865 from MEC, and 5,312 from WHI). In addition we included 6,326 WHI participants genotyped as part of the SNP Health Association Resource (SHARe) on the Affymetrix 6.0 platform. Details can be found elsewhere [51].



## Imputation to 1000 Genomes Project

To impute to the 1000 Genomes Project we used as the reference panel the haplotypes of the 1092 samples (all populations) from release version 2 of the 1000 Genomes Project Phase I (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20110521>) [52]. Combining reference data from all populations has been found to improve imputation accuracy of the low-frequency variants [53], and, hence, is recommended. We built the target panel by combining all genotype data in the *FTO* region from all studies. We used genotyped data from the MetaboChip for all studies, except for WHI SNP Health Association Resource (SHARe,  $n=6,326$  samples) where we used genotype data from the Affymetrix 6.0 platform. The target panel was phased using Beagle [54]. We then performed a haplotype-to-haplotype imputation to estimate genotypes (as allele dosages) at 1000 Genomes Project variants. The phased target panel was imputed to the interval 53.5–54.2 Mb on chromosome 16 of the 1000 genomes reference panel using minimac [55]. To evaluate the quality of each imputed SNP we calculated  $R_{sq}$ . We excluded imputed SNPs with  $R_{sq}<0.9$  for SNPs with allele frequency  $<0.5\%$ ,  $R_{sq}<0.8$  for SNPs with allele frequency  $>0.5\text{--}1\%$ ,  $R_{sq}<0.7$  for SNPs with allele frequency  $>1\text{--}3\%$ ,  $R_{sq}<0.6$  for SNPs with allele frequency  $>3\text{--}5\%$ , and  $R_{sq}<0.5$  for SNPs with allele frequency  $>5\%$ . Given the large reference panel this resulted in high imputation quality [51]. To calculate pairwise correlation between variants we used the 1000 Genomes Project data, specifically we used 61 African Americans from the South-west (ASW) and 65 European Americans (Utah residents with Northern and Western European ancestry from the CEPH collection, CEU).

## Statistical analysis

The association between each SNP and natural log-transformed BMI ( $\ln$ BMI) was estimated using linear regression. SNP genotypes were coded assuming an additive genetic model (i.e., 0, 1, or 2 copies of the coded allele). All analyses were adjusted for age (continuous), sex, and study site (as applicable). All models (except WHI) included  $\text{sex} \times \text{age}$  interaction terms to account for possible effect modification by sex. In addition, we adjusted for the top two principal components of ancestry. Family data from GenNet and HyperGEN was analyzed using mixed models (variance component models) to account for relatedness.

Results ( $\beta$  and SEs for  $\ln$ BMI) were combined with fixed-effects meta-analysis weighting the effect size estimates ( $\beta$ -coefficients) by their estimated standard errors, using METAL [56]. We evaluated  $Q$ -statistic and  $I^2$  as a measure of heterogeneity [57,58], to describe the presence or absence of excess variation between the PAGE cohorts. For ease of interpretation, we calculated the % change in BMI per copy of the effect allele based on the beta for the  $\ln$ BMI. To graphically display the results, we used LocusZoom [59]. We tested for independence of findings by including the most significant variant and each of the other variant into the same model (i.e. we included 2 variants simultaneously in one model). We evaluated if SNPs are independent by investigating the  $p$ -value. For an independent SNP the  $p$ -value would remain low/similar after adjusted for the most significant SNP.

## Bioinformatics

We conducted a bioinformatic characterization for the most significant SNP and all SNPs correlated with the most significant

SNP ( $r^2>0.5$ ). We implemented in-house Perl scripts to query bioinformatic databases, and assigned each of the 16 SNPs to one or more of the functional annotation datasets listed in Table S3. These datasets are not mutually exclusive. For example, a SNP can be located within both a candidate regulatory element (dataset #7) and a CTCF binding site (dataset #10). Because *FTO* is expressed and may have functional relevance in a wide array of tissues, we defined candidate *cis*-regulatory elements (dataset #7) as DNaseI hypersensitive sites (open chromatin loci) that are present in at least one human cell type. For SNPs that occur within predicted transcription factor binding sites (datasets #3 and #8), we computed transcription factor binding affinity for each SNP allele using the PWM-scan algorithm [60], as described previously [61].

## Supporting Information

**Figure S1** Predicted CUX1 binding site at the rs1421085 locus. (DOCX)

**Table S1** Association between SNPs in the *FTO* region and BMI for each study separately. (DOCX)

**Table S2** Risk estimates for SNPs correlated with rs56137030 and combined analyses with rs56137030 for all studies combined. (DOCX)

**Table S3** Functional annotation datasets. (DOCX)

**Table S4** Association between *FTO* SNPs and BMI for SNPs highlighted in previous studies of African Americans. (DOCX)

**Text S1** Description of each study. (DOC)

## Acknowledgments

We thank the MetaboChip design team lead by Goncalo Abecasis, David Altshuler, Michael Boehnke, and Mark McCarthy for their work to design the chip. As part of this we thank each of the participating meta-analysis Consortia for providing SNP and locus lists. We thank Quang Le and Richard Durbin from the Sanger Institute and Jared Maguire and Mark Daly from the Broad Institute for providing August 1000 Genomes genotype calls prior to their public release. Most of all, we thank a small set of young scientists whose creative thinking and long hours turned the MetaboChip concept into a real product: in particular Noel Burt, Ben Voight, and Cameron Palmer at the Broad Institute and Hyun Min Kang, Jun Ding, and Yanming Li at the University of Michigan. The complete list of PAGE members can be found at <http://www.pagestudy.org>.

## Author Contributions

Conceived and designed the experiments: UP KEN SB RDJ LLM LAH DCC CAH CK. Performed the experiments: UP KEN PS SB SJ MDF RDJ UL IC FS LW RL KM GE K-DHN RC CEL ML MRI CCG DH PB MR TCM LLM LAH DCC CAH CK. Analyzed the data: UP KEN PS SB JH SJ FS CK. Contributed reagents/materials/analysis tools: RDJ LHK AR GE K-DHN RC CEL ML MRI CCG DH LLM DCC CAH CK. Wrote the paper: UP KEN PS GE LAH CAH CK.

## References

1. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, et al. (2007) A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316: 889–894.
2. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, et al. (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41: 18–24.

3. Hinney A, Nguyen TT, Scherag A, Friedel S, Bronner G, et al. (2007) Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS ONE* 2: e1361. doi:10.1371/journal.pone.0001361
4. Renstrom F, Payne F, Nordstrom A, Brito EC, Rolandsson O, et al. (2009) Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Hum Mol Genet* 18: 1489–1496.
5. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, et al. (2007) Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet* 3: e115. doi:10.1371/journal.pgen.0030115
6. Song Y, You NC, Hsu YH, Howard BV, Langer RD, et al. (2008) *FTO* polymorphisms are associated with obesity but not diabetes risk in postmenopausal women. *Obesity* 16: 2472–2480.
7. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, et al. (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41: 25–34.
8. Dina C, Meyre D, Gallina S, Durand E, Korner A, et al. (2007) Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nat Genet* 39: 724–726.
9. Wang K, Li WD, Zhang CK, Wang Z, Glessner JT, et al. (2011) A genome-wide association study on obesity and obesity-related traits. *PLoS ONE* 6: e18939. doi:10.1371/journal.pone.0018939
10. Cotsapas C, Speliotes EK, Hatoum IJ, Greenawald DM, Dobrin R, et al. (2009) Common body mass index-associated variants confer risk of extreme obesity. *Hum Mol Genet* 18: 3502–3507.
11. Meyre D, Delplanque J, Chevre JC, Lecocur C, Lobbens S, et al. (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 41: 157–159.
12. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, et al. (2008) Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat Genet* 40: 768–775.
13. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 42: 937–948.
14. Hennig BJ, Fulford AJ, Sirugo G, Rayco-Solon P, Hattersley AT, et al. (2009) *FTO* gene variation and measures of body mass in an African population. *BMC Med Genet* 10: 21.
15. Li TY, Rana JS, Manson JE, Willett WC, Stampfer MJ, et al. (2006) Obesity as compared with physical activity in predicting risk of coronary heart disease in women. *Circulation* 113: 499–506.
16. Adams KF, Schatzkin A, Harris TB, Kipnis V, Mouw T, et al. (2006) Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med* 355: 763–778.
17. Flegal KM, Carroll MD, Ogden CL, Curtin LR (2010) Prevalence and trends in obesity among US adults, 1999–2008. *JAMA* 303: 235–241.
18. Wing MR, Ziegler JM, Langefeld CD, Roh BH, Palmer ND, et al. (2010) Analysis of *FTO* gene variants with obesity and glucose homeostasis measures in the multiethnic Insulin Resistance Atherosclerosis Study cohort. *Int J Obes (Lond)*.
19. Wing MR, Ziegler J, Langefeld CD, Ng MC, Haffner SM, et al. (2009) Analysis of *FTO* gene variants with measures of obesity and glucose homeostasis in the IRAS Family Study. *Hum Genet* 125: 615–626.
20. Dorajoo R, Blakemore AI, Sim X, Ong RT, Ng DP, et al. (2011) Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int J Obes (Lond)*.
21. Cha SW, Choi SM, Kim KS, Park BL, Kim JR, et al. (2008) Replication of genetic effects of *FTO* polymorphisms on BMI in a Korean population. *Obesity (Silver Spring)* 16: 2187–2189.
22. Chang YC, Liu PH, Lee WJ, Chang TJ, Jiang YD, et al. (2008) Common variation in the fat mass and obesity-associated (*FTO*) gene confers risk of obesity and modulates BMI in the Chinese population. *Diabetes* 57: 2245–2252.
23. Tan JT, Dorajoo R, Scielstad M, Sim XL, Ong RT, et al. (2008) *FTO* variants are associated with obesity in the Chinese and Malay populations in Singapore. *Diabetes* 57: 2851–2857.
24. Ng MC, Park KS, Oh B, Tam CH, Cho YM, et al. (2008) Implication of genetic variants near *TCF7L2*, *SLC30A8*, *HHEX*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, and *FTO* in type 2 diabetes and obesity in 6,719 Asians. *Diabetes* 57: 2226–2233.
25. Al-Attar SA, Pollex RL, Ban MR, Young TK, Bjerregaard P, et al. (2008) Association between the *FTO* rs9939609 polymorphism and the metabolic syndrome in a non-Caucasian multi-ethnic sample. *Cardiovasc Diabetol* 7: 5.
26. Omori S, Tanaka Y, Takahashi A, Hirose H, Kashiwagi A, et al. (2008) Association of *CDKAL1*, *IGF2BP2*, *CDKN2A/B*, *HHEX*, *SLC30A8*, and *KCNJ11* with susceptibility to type 2 diabetes in a Japanese population. *Diabetes* 57: 791–795.
27. Grant SF, Li M, Bradfield JP, Kim CE, Annaiah K, et al. (2008) Association analysis of the *FTO* gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP. *PLoS ONE* 3: e1746. doi:10.1371/journal.pone.0001746
28. Bollepalli S, Dolan LM, Deka R, Martin LJ (2010) Association of *FTO* gene variants with adiposity in African-American adolescents. *Obesity (Silver Spring)* 18: 1959–1963.
29. Nock NL, Plummer SJ, Thompson CL, Casey G, Li L (2011) *FTO* polymorphisms are associated with adult body mass index (BMI) and colorectal adenomas in African-Americans. *Carcinogenesis* 32: 748–756.
30. Adeyemo A, Chen G, Zhou J, Shriner D, Dumaty A, et al. (2010) *FTO* genetic variation and association with obesity in West Africans and African Americans. *Diabetes* 59: 1549–1554.
31. Hassanein MT, Lyon HN, Nguyen TT, Akyzbekova EL, Waters K, et al. (2010) Fine mapping of the association with obesity at the *FTO* locus in African-derived populations. *Human molecular genetics* 19: 2907–2916.
32. Hester JM, Wing MR, Li J, Palmer ND, Xu J, et al. (2011) Implication of European-derived adiposity loci in African Americans. *International journal of obesity*.
33. McKenzie CA, Abecasis GR, Keavney B, Forrester T, Ratcliffe PJ, et al. (2001) Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). *Hum Mol Genet* 10: 1077–1084.
34. Stratigopoulos G, LeDuc CA, Cremona ML, Chung WK, Leibel RL (2011) Cut-like homeobox 1 (*CUX1*) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (*RPGRIP1L*) genes and coordinates leptin receptor signaling. *The Journal of biological chemistry* 286: 2155–2170.
35. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, et al. (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 8: e1002793. doi:10.1371/journal.pgen.1002793
36. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. *Genome Res* 12: 996–1006.
37. (ESP) NESP (January 2012) Exome Variant Server.
38. Deo RC, Reich D, Tandon A, Akyzbekova E, Patterson N, et al. (2009) Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. *PLoS Genet* 5: e1000342. doi:10.1371/journal.pgen.1000342
39. Jiao S, Hsu L, Hutter CM, Peters U (2011) The use of imputed values in the meta-analysis of genome-wide association studies. *Genet Epidemiol* 35: 597–605.
40. Matisse TC, Ambite JL, Buyske S, Carlson CS, Cole SA, et al. (2011) The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. *Am J Epidemiol* 174: 849–859.
41. (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 129: 687–702.
42. (2002) Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* 39: 3–9.
43. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, et al. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 151: 346–357.
44. (1998) 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.
45. Gorber SC, Tremblay MS (2010) The bias in self-reported obesity from 1976 to 2005: a Canada-US comparison. *Obesity (Silver Spring)* 18: 354–361.
46. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
47. Weale ME (2010) Quality control for genome-wide association studies. *Methods Mol Biol* 628: 341–372.
48. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2: e190. doi:10.1371/journal.pgen.0020190
49. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909.
50. Buyske S, Wu Y, Carty CL, Cheng I, Assimes TL, et al. (2012) Evaluation of the Metabochip Genotyping Array in African Americans and Implications for Fine Mapping of GWAS-Identified Loci: The PAGE Study. *PLoS ONE* 7: e35651. doi:10.1371/journal.pone.0035651
51. Liu Y AA, Buyske S, Peters U, Boerwinkle E, Carlson C, Carty C, Crawford DC, Haessler J, Haiman C, Le Marchand L, Hindorf L, Manolio T, Matisse T, Wang W, Kooperberg CL, North KE, Li Y. (In Press) Genotype Imputation of Metabochip SNPs Using study specific reference panel of ~4,000 haplotypes in African Americans. *Genet Epidemiol*.
52. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–1073.
53. Howie B, Marchini J, Stephens M (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1: 457–470.
54. Browning SR, Browning BL (2007) 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.
55. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44: 955–959.
56. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.
57. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7: 177–188.

58. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327: 557–560.
59. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26: 2336–2337.
60. Levy S, Hannenhalli S (2002) Identification of transcription factor binding sites in the human genome sequence. *Mamm Genome* 13: 510–514.
61. Sethupathy P, Giang H, Plotkin JB, Hannenhalli S (2008) Genome-wide analysis of natural selection on human cis-elements. *PLoS ONE* 3: e3137. doi:10.1371/journal.pone.0003137