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

PETERS, Ulrike, et al.

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 × 10(-6)) had not been highlighted in previous studies. While rs56137030 was correlated at r(2)>0.5 with [...]
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 Peters1*, Kari E. North2,3, Praveen Sethupathy2,4, Steve Buyske5,6, Jeff Haessler1, Shuo Jiao1, Megan D. Fesinmeyer1, Rebecca D. Jackson7, Lew H. Kuller8, Aleksandar Rajkovic9, Unhee Lim10, Iona Cheng9, Fred Schumacher11, Lynne Wilkens10, Rongling Li12, Keri Monda12,3, Georg Ehret13,14, Khanh-Dung H. Nguyen13, Richard Cooper15, Cora E. Lewis16, Mark Leppert17, Marguerite R. Irvin18, C. Charles Gu19, Denise Houston20, Petra Buzkova21, Marylyn Ritchie22, Tara C. Matise5, Loic Le Marchand10, Lucia A. Hindorff12, Dana C. Crawford23,24, Christopher A. Haiman11, Charles Kooperberg1

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, McKusik-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-kt 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 (8.3×10−10) had not been highlighted in previous studies. While rs56137030was correlated at r²>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.


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

Received September 12, 2012; Accepted October 22, 2012; Published January 17, 2013
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/m² (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 molecular evaluation of the 16q12.2/FTO gene [30]. This process can reduce the number of potential functional variants for follow-up studies and helps identify additional independent signal(s) implicating multiple 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/generaliability 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 signals 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 RPGRIP1L in over 20,000 AA samples from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. We used the Metaobichip 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

Fine-Mapping of FTO Locus in African Americans

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, R01CA63, P01CA33619, U01CA136792, and U01CA8758). Funding support for the “Epidemiology of putative genetic variants: The Women’s Health Initiative” study is provided through the NHGRI PAGE program (U01HG004790). The WHO 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, 32102-0, 32105-6, 32108-7, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHO investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHO investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf. The Atherosclerosis Risk in Communities (ARC) 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 1K99DK09318. 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 U01HG04803 (CALICO), U01HG04798 (EAGLE), U01HG04802 (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

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.
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 $\geq 30$ kg/m$^2$) 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 *RPGRIP1L* 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 *RPGRIP1L*, 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 $8.3 \times 10^{-26}$), 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 $1.4 \times 10^{-25}$ to $1.1 \times 10^{-25}$) 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).
rs56137030 and the three next most significant SNPs are located in intron 1 of the \textit{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 \textit{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 3.610^{-2}).

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 1.610^{-2}) 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, rs116751859, and rs116752985) have been genotyped in this study].

Figure 1. Association results (p-values) and correlation structure for all SNPs in the 16q12.2/\textit{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/\textit{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

<table>
<thead>
<tr>
<th>SNP</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele frequency$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.14%–1%</td>
<td>627</td>
<td>16.7%</td>
</tr>
<tr>
<td>&gt;1%–5%</td>
<td>1346</td>
<td>35.8%</td>
</tr>
<tr>
<td>&gt;5%–10%</td>
<td>431</td>
<td>11.5%</td>
</tr>
<tr>
<td>&gt;10%–25%</td>
<td>679</td>
<td>18.1%</td>
</tr>
<tr>
<td>&gt;25%</td>
<td>673</td>
<td>17.9%</td>
</tr>
<tr>
<td>Total</td>
<td>3756</td>
<td>100.0%</td>
</tr>
<tr>
<td>Exonic SNPs$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNPs in \textit{RPGRIP1L} exons</td>
<td>15</td>
<td>0.4%</td>
</tr>
<tr>
<td>synonymous SNPs in \textit{RPGRIP1L}</td>
<td>3</td>
<td>0.1%</td>
</tr>
<tr>
<td>missense SNPs in \textit{RPGRIP1L}</td>
<td>12</td>
<td>0.3%</td>
</tr>
<tr>
<td>SNPs in \textit{FTO} exons</td>
<td>6</td>
<td>0.2%</td>
</tr>
<tr>
<td>synonymous SNPs in \textit{FTO}</td>
<td>4</td>
<td>0.1%</td>
</tr>
<tr>
<td>missense SNPs in \textit{FTO}</td>
<td>2</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

$^a$Based on allele frequency for all AA participants in this study.

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 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 [2,3,6,14,18,28–31], rs1121980 [18,19,28,30,31], rs17817449 [18,19,28–31], or the previously reported functional variant 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.  

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

<table>
<thead>
<tr>
<th>SNP#</th>
<th>Position</th>
<th>Coding</th>
<th>Baseline</th>
<th>CAF</th>
<th>% change in BMI per coding allele</th>
<th>Beta estimate</th>
<th>95%CI</th>
<th>nom.p</th>
<th>Rsq</th>
<th>p.het</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs56137030</td>
<td>53825905</td>
<td>A</td>
<td>G</td>
<td>0.12</td>
<td>1.35</td>
<td>(0.76,1.95)</td>
<td>8.3E-06</td>
<td>0.98</td>
<td>1.3E-01</td>
<td></td>
</tr>
<tr>
<td>rs62033400</td>
<td>53811788</td>
<td>G</td>
<td>A</td>
<td>0.12</td>
<td>1.35</td>
<td>(0.76,1.95)</td>
<td>1.1E-05</td>
<td>0.99</td>
<td>1.3E-01</td>
<td></td>
</tr>
<tr>
<td>rs7188250</td>
<td>53834607</td>
<td>C</td>
<td>T</td>
<td>0.12</td>
<td>1.34</td>
<td>(0.75,1.94)</td>
<td>1.3E-05</td>
<td>0.97</td>
<td>8.3E-02</td>
<td></td>
</tr>
<tr>
<td>rs62033413</td>
<td>53800055</td>
<td>G</td>
<td>C</td>
<td>0.12</td>
<td>1.33</td>
<td>(0.74,1.93)</td>
<td>1.4E-05</td>
<td>0.98</td>
<td>4.5E-02</td>
<td></td>
</tr>
</tbody>
</table>

*SNP position based on build 37.*

*Coding = coding allele, Base = baseline allele (risk estimates provide the log additive effect per copy of the coding allele).*

*CAF = coding allele frequency.*

*Rsq = measurement of imputation accuracy, ranging from 0 (low) to 1 (high).*

**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
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 = 5 × 10^{-4}; Table S2) and was also found to be associated in the study from Nock et al. [29] (n = 469, p-value = 7 × 10^{-4}), Hassanein et al. [31] (n = 4,217, p-value > 3 × 10^{-5}), and Hester et al. [32] (n = 4,992, p-value = 0.07) but not with four smaller AA studies (≤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-value = 4 × 10^{-6} and n = 4,217, p-value = 2 × 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-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 rs1081012 (p-value = 5 × 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-value = 5 × 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 (~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-value = 4 × 10^{-4} and rs9941349, p-value = 6 × 10^{-5}) in four additional studies (n = 5,664), adding some support (p-values ranging from 0.016 to 0.64) which resulted in overall p-values of 2.6 × 10^{-6} and 3.6 × 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-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 = 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 tagsSNPs in 1,405 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 led 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-value of a SNP is impacted by the imputation accuracy (lower imputation accuracy results in higher p-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^2 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 1.33% 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^2 compared with 4.2 kg/m^2 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 rs368170309 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 overweight (BMI<18.5 kg/m²) or extremely overweight (BMI>70 kg/m²) 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 Metaobochip, 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 rs≥0.3 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 Metaobochip 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<1×10^-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 Metaobochip 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 (SHAREs, 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 Rsq. We excluded imputed SNPs with Rsq<0.9 for SNPs with allele frequency <0.5%, Rsq<0.8 for SNPs with allele frequency >0.5–1%, Rsq<0.7 for SNPs with allele frequency >1–3%, Rsq<0.6 for SNPs with allele frequency >3–5%, and Rsq<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 (lnBMI) 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 sex*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 [β and SEs for lnBMI] were combined with fixed-effects meta-analysis weighting the effect size estimates (β-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 lnBMI. 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 change in lnBMI. To graphically display the results, we used LocusZoom [59].

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 #9), 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 Gonzalo 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 Burtt, 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. 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.


