Assessing efficiency of fine-mapping obesity associated variants through leveraging ancestry architecture and functional annotation using PAGE and UKBB Cohorts

Mohammad Yaser Anwar (manwaryaser@gmail.com)
University of North Carolina at Chapel Hill https://orcid.org/0000-0002-2349-8214

Mariaelisa Graff
UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

Heather M. Highland
UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

Roelof Smit
Icahn School of Medicine at Mount Sinai

Zhe Wang
Icahn School of Medicine at Mount Sinai

Victoria L. Buchanan
UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

Kristina L. Young
UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

Eimear E. Kenny
Icahn School of Medicine at Mount Sinai

Lindsay Fernandez-Rhodes
The Pennsylvania State University

Simin Liu
Brown University School of Public Health

Themistocles Assimes
Stanford University School of Medicine

David O. Garcia
The University of Arizona

Kim Daeeun

UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

Christopher R. Gignoux
University of Colorado Anschutz Medical Campus School of Medicine

Anne E. Justice
Geisinger Health

Christopher A. Haiman
University of Southern California Keck School of Medicine

Steve Buyske
Rutgers University: Rutgers The State University of New Jersey

Ulrike Peters
Fred Hutch Cancer Center: Fred Hutchinson Cancer Center

Ruth Loos
Icahn School of Medicine at Mount Sinai

Charles Kooperberg
Fred Hutchinson Cancer Research Center: Fred Hutchinson Cancer Center

Kari E. North
UNC Gillings School of Global Public Health: The University of North Carolina at Chapel Hill Gillings School of Global Public Health

---

Research Article

Keywords: Fine-mapping, obesity, multi-ethnic, efficiency, functional variants

Posted Date: March 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2608695/v1

Version of Record: A version of this preprint was published at Human Genetics on September 1st, 2023. See the published version at https://doi.org/10.1007/s00439-023-02593-7.

Abstract

Inadequate representation of non-European ancestry populations in genome-wide association studies (GWAS) has limited opportunities to isolate functional variants. Fine-mapping in multi-ancestry populations should improve the efficiency of prioritizing variants for functional interrogation. To evaluate this hypothesis, we leveraged ancestry architecture to perform comparative GWAS and fine-mapping of obesity related phenotypes in European ancestry populations from the UK Biobank (UKBB) and multi-ancestry samples from the Population Architecture for Genetic Epidemiology (PAGE) consortium with comparable sample sizes. In 10 of the investigated regions with genome wide significant associations for obesity related traits, fine-mapping in our ancestrally diverse sample led to 95% and 99% credible sets (CS) with fewer variants than in the European ancestry sample. Lead fine-mapped variants in PAGE regions had higher average coding scores, and higher average posterior probabilities for causality compared to UKBB. Importantly, 99% CS in PAGE loci contained strong expression quantitative trait loci (eQTLs) in adipose tissues or harbored more variants in tighter linkage disequilibrium (LD) with eQTLs. Results also suggested three novel candidates for functional effect on waist-to-hip ratio adjusted for BMI (WHRBMI-adj) (rs5781117 near gene *RP11-392O17.1*, rs10187501 in gene *COBLL1*, and rs1964599 near gene *CCDC92*), all within the 99% CS. Leveraging ancestrally diverse populations with heterogeneous ancestry architectures, coupled with functional annotation, increased fine mapping efficiency and performance, and reduced the set of candidate variants for consideration for future functional studies. Significant overlap in genetic causal variants across populations suggest generalizability of genetic mechanisms underpinning obesity related traits across populations.

Background

Genome-wide association studies (GWAS) (Caballero et al. 2015; Visscher et al. 2012; Zhang et al. 2019) have uncovered > 1000 independent genomic regions associated with body mass index (BMI) and waist-to-hip ratio adjusted for BMI (WHRBMIadj). We expect further discoveries to follow as data from large cohorts with millions of participants are becoming available. Yet biological translation of GWAS remains a challenge given extensive linkage disequilibrium (LD) across the genome (Riancho 2012; Tam et al. 2019). Except for relatively rare monogenic cases inherited as Mendelian traits (Challis et al. 2004; Huszar et al. 1997; Krude et al. 1998), obesity is a complex polygenic phenotype (Hinney et al. 2010) involving large and disparate genomic regions. Considerable LD among associated variants (Pritchard and Przeworski 2001), presence of multiple independent causal variants in a locus (Flister et al. 2013), and generally small effect sizes (Hodge and Greenberg 2016; Yengo et al. 2018) create methodological barriers to identifying functional genes and variants for polygenic traits. Consequently, only a handful of studies of rare nonsynonymous variants (i.e., minor allele frequency (MAF) < 1%) with large effect sizes have established causality with reasonable certainty (Emdin et al. 2018; Yengo et al. 2018).

Study population composition (Magavern et al. 2022) also complicates identification of functional genes and variants. Genetic studies predominantly involve self-identified Non-Hispanic White populations, groups with limited representation of global ancestral diversity (Price et al. 2008; Ralph and Coop 2013).

Identifying causal variants is arduous when obesity influencing variants are in long haplotype blocks that segregate together. Therefore, traditional fine-mapping methods that assign a posterior probability of causality to each mapped variant in a region (Chen et al. 2015; Hormozdiari et al. 2014; Wakefield 2007) solely based on European populations reference LD architecture (Benner et al. 2016; Pruim et al. 2010) rarely leads to unambiguous characterization of a potential causal variant (Chen et al. 2019; Tam et al. 2019; Witte 2010).

Hence, it has been proposed that fine-mapping in multi-ancestry and admixed populations with shorter haplotype blocks (Consortium 2012; Mao et al. 2017) will narrow the number of candidate variants and increase statistical power to identify likely functional variants (Kichaev and Pasaniuc 2015; van de Bunt et al. 2015). Additionally, with a greater proportion of discovered variants being mapped to non-coding genomic regions (Giral et al. 2018), often with regulatory effect on genes that are distant to them in DNA sequence (Vance et al. 2014), functional annotation may facilitate prioritization of causal variants in GWAS regions. GWAS fine-mapping coupled with functional annotation show substantial improvements over traditional methods in identifying likely functional variants and are increasingly adopted in large scale studies that involve multiple study sites and tens of thousands of participants (Willems et al. 2017; Zhao et al. 2021).

In this study, we performed fine-mapping of obesity-related phenotypes using samples from two studies, 1) Europeans from the UK Biobank (UKBB) and 2) multi-ancestry samples from the Population Architecture for Genetic Epidemiology (PAGE) consortium. For approximate power comparability, we considered a random subset of UKBB participants, to match the multi-ancestry sample sizes available in PAGE. Our main goal was to leverage ancestry architecture followed by functional annotation to narrow the list of possible causative genes and functional variants underlying GWAS signals. We additionally hypothesized that applying fine-mapping approaches in ancestrally diverse populations would reveal more variants with compelling evidence for functionality than in the European UKBB subset we interrogated.

Methods

Study Cohorts

The PAGE consortium includes several studies. In brief, PAGE consists of multiple populations grouped by self-identified race and ethnicity, European Americans (EA), African Americans (AA), Hispanic Americans (HA), Native Americans (NAm), East Asians (ASN), and Native Hawaiians/Pacific Islander (NH) (CALiCo-SOL and Fernandez-Rhodes ; Manolio 2009; Matise et al. 2011). All participating sites in PAGE ascertained both males and females except for the females only Women's Health Initiative (WHI). Analyses were performed in all populations combined via meta-analyses.

The UKBB is a population-based study of citizens of the United Kingdom (Sudlow et al. 2015). Approximately half a million individuals, primarily of European ancestry, were recruited between 2006 to

2010. Genetic and phenotypic characteristics of UKBB individuals used for GWAS were defined previously (Bycroft et al. 2018). We randomly selected a subset of unrelated European ancestry individuals for GWAS to be comparable to our sample sizes in PAGE, thereby removing sample size as a factor contributing to differential results **(Sup. Table 1)**.

Phenotyping And Quality Control

We studied two anthropometric phenotypes: BMI (kg/m2; a measure of overall adiposity) and waist-to-hip ratio adjusted for BMI (WHRBMI-adj), a proxy of central obesity. In 16 of 17 studies that contribute to PAGE, height and weight were measured by study staff at study enrollment, to calculate BMI (weight/height2) **(Sup. Table 1)**. In the remaining Multi-ethnic Cohort (MEC), BMI is based on self-reported height and weight at enrollment. Pilot analyses of BMI in MEC illustrated a comparable distribution to national surveys (Gorber and Tremblay 2010). Waist circumference was measured at the level of the natural waist in horizontal plane to the nearest 0.5 cm (Carty et al. 2014); no waist or hip circumference measurement was available for the BioMe sub-cohort.

Recruitment and data collection in the UKBB sample has been previously described (Biobank 2007; Bycroft et al. 2017). UKBB participants were randomly selected from those study participants that self-identified as European and that clustered within the 1000 Genomes Europeans (EUR) ancestry population when applying a k-means clustering approach to genotype data (n = 451,337). As we grouped study participants by self-reported race /ethnicity, we additionally excluded those UKBB participants that did not self-report as European (n = 32). We excluded women who were pregnant or unsure if they were pregnant. We removed any BMI or WHR measures ± 6 SD from the mean by sex.

Genotyping And Quality Control

In PAGE, approximately 50,000 individuals were genotyped using the Multiethnic Genotyping Array (MEGA) panel, as previously described (Bien et al. 2016). The remaining PAGE individuals were genotyped with Affymetrix arrays. After quality control, genomic imputation was performed using the 1000 Genomes Phase 3 reference population; details are accessible here (Hu et al. 2021).

In the UKBB, genotyping was performed using either the Applied Biosystems UKBB Lung Exome Variant Evaluation (UK BiLEVE) Axiom Array or UKB Axiom Array (Bycroft et al. 2018). The genotypes were imputed using IMPUTE4 with a combination of reference panels: i) the Haplotype Reference Consortium and ii) UK10K and the 1000 Genomes Phase 3 (Bycroft et al. 2017). For this study, we excluded non-autosomal genetic variants, those with poor imputation (R2 < 0.4), effective sample size < 30, or MAF of < 0.05. Criteria used for calculating effective sample size for each variant are defined in **Appendix**. Approximately 32 million of the 60 million variants had MAF < 0.01 and were removed from analyses.

Gwas And Meta-analyses

In PAGE, GWAS were performed with *SUGEN* (Lin et al. 2014). In UKBB, GWAS were performed using SAIGE (Zhou et al. 2018). Age, sex (BMI only), study center (PAGE only) and the first 10 principal components (PC) of ancestry were included as covariates. In PAGE, GWAS were initially performed in sex-combined (BMI only) or sex-stratified self-identified race/ethnic group separately, and subsequently meta-analyzed across sex (WHRBMI-adj) and self-identified race/ethnicity groups (both BMI and WHRBMI-adj) using METAL (Willer et al. 2010). In UKBB, GWAS were similarly performed in a subset with approximate sample size to match PAGE **(Sup. Table 1)**. For WHR, we additionally adjusted associations for BMI (WHRBMIadj).

Selection And Configuration Of Genomic Regions For Fine-mapping

For each trait/group combination, variants exceeding GWAS significance level ($p < 5x10\text{-}8$) were extracted from PAGE and UKBB summary statistics and compiled into a single list **(Sup. Table 2)**. GWAS variants were then LD pruned (threshold $R2 > 0.1$) and grouped into independent clusters each representing a potential functional region. Variants with the lowest p-value in each cluster were selected as the index variant for that cluster (see Box 1), irrespective of whether this association was observed in the UKBB or PAGE results). We further restricted analyses to loci where both PAGE and UKBB overlapping regions harbored at least one variant exceeding GWAS significance level in both populations. The minimum base pair (bp) distance between each pair of adjacent index variants was ≥ 300 kbp. Therefore, each functional locus was defined as the set of variants that were located ≤ 150 kbp from each index variant for that region **(Sup. Figure 1)**. This method was applied to each phenotype and ancestry group combination. Hence, each genomic region probed in UKBB GWAS was similarly fine-mapped in PAGE in the same overlapping genomic region (i.e., in pairs).

Fine-mapping And Sensitivity Analyses

For both fine-mapping and sensitivity analyses, we used summary statistics and *SLALOM* (Kanai et al. 2022). This method incorporates an Approximate Bayesian Factor (ABF) (Wakefield 2007, 2009) for fine-mapping which estimates a posterior inclusion probability (PIP) for each variant and derives the smallest possible 95% and 99% causal set (CS) (i.e., set of variants whose cumulative posterior probability is 95% or 99%) based on p-value, and LD assuming one functional variant per locus. Additionally, GWAS statistics from PAGE are from a meta-analysis of populations with distinct patterns of ancestry and admixture, genotyped on distinct platforms and imputed separately. To assess if the heterogeneous characteristics of the contributing cohorts (e.g., difference in patterns of admixture and ancestry, sample size, genotyping, or imputation) may have affected fine-mapping outputs in PAGE, *SLALOM* performed *DENTIST* (Chen et al. 2021) based *DENTIST-Simplified* (*DENTIST-S*) test to flag loci with suspect GWAS results. According to *DENTIST-S*, observed statistical significance of a variant is expected to be proportional to its LD with the lead variant (variant with highest posterior probability) assuming both belong to the set representing the same signal. The presence of variant(s) in tighter LD with the lead but higher than expected p-value suggest association outlier(s), and hence the quality of fine-mapping in that

locus would be questionable. For fine-mapping, *SLALOM* inferred LD from *gnomAD* (Koch 2020) reference African, admixed American, East Asian, and European populations averaged by each populations' study sample size when testing PAGE loci, but utilized UKBB-specific reference for these regions in UKBB. It also inferred functional annotation from Variant Effect Predictor (VEP) for each fine-mapped variant. VEP is a toolset for prioritization, and functional annotation of genomic variants in coding and non-coding regions. (McLaren et al. 2016)

Assessing Fine-mapping Efficiency

We conducted two stages of analyses as follows:

Stage 1. We extracted the variants with the highest posterior probabilities in each region (termed the lead fine-mapped variant, see Box1) within the 99% CS from both PAGE and UKBB. Using the 1000 Genome multi-ancestry and EUR reference populations, we then estimated pairwise LD between the lead fine mapped variants in the 99% CS and their LD proxies from both PAGE and UKBB, to determine if UKBB and PAGE data were representing the same signal (threshold $R2 > 0.1$). We also assessed whether the lead fine mapped variants were previously reported in the literature for obesity related traits using the GWAS-Catalogue (Welter et al. 2014) and *PhenoScanner* (Staley et al. 2016). To further assess the efficiency of fine-mapping, we performed functional annotation of all variants in the 99% CS in each region using the Combined Annotation Dependent Depletion (CADD) tool for scoring deleteriousness of both coding and noncoding variants (Kircher et al. 2014). Negative log CADD score values > 10 suggest a high probability of functionality and > 20 also have experimental evidence for functionality. We defined a best non-lead fine-mapped variant (defined here as a variant present in the 99% CS but not the lead fine-mapped variant, but with the highest CADD, see Box1) for each region. The purpose of this exercise was to compare *SLALOM* generated functional annotations from VEP to those we generated with LD proxies, and consideration of CADD score.

Stage 2. We performed variant annotation using expression trait quantitative loci (eQTL) evidence from obesity relevant tissues (whole blood, adipose, brain, liver, and skeletal muscle tissues) utilizing GTeX version 8 (gtexportal.org) (Consortium 2020). Complementary to fine-mapping, CADD scores, and VEP annotations, eQTLs were used to narrow and characterize likely causal variants or their close proxies. We extracted eQTL summary statistics for the lead fine-mapped and index variants in each region and assessed if they were included in 99% CS in PAGE or UKBB fine-mapping results. In regions where no significant eQTL existed in obesity relevant tissues, we alternatively searched for splice QTLs (sQTL), eQTL in other tissues, and eQTL from the literature.

Results

>

GWAS

We conducted GWAS of BMI and WHRBMI-adj in the 6 separate PAGE populations and in our single UKBB sample. Population and study-specific GWAS results in PAGE were subsequently meta-analyzed together. Manhattan and QQ plots for GWAS results are presented in **Appendix**. GWAS significant variants are reported in **Sup. Table 2**.

Fine-mapping

Stage 1. We selected 10 pairs of regions for fine-mapping where at least 1 variant was associated with BMI or WHRBMIadj traits at GWAS significance level ($p < 5\times10 - 8$) in both PAGE and UKBB **(Sup. Table 3)**. We then extracted 99% CS from each region in both cohorts, and performed VEP-based functional annotation and CADD scoring. In all 10 regions, the most statistically significant index variant in UKBB GWAS was also the lead fine-mapped variants (i.e., variants with highest posterior probability); comparatively, in PAGE, only 8 of 10 regions identified the same index and lead fine-mapped variant (**Table 1**). Also, 9 of 10 lead fine-mapped variants in UKBB were previous reported as the index SNP associated with BMI or WHRBMIadj in the literature versus 6 of 10 in PAGE. Lead variants in UKBB and PAGE in three regions (chr1:177739839–178039226 for BMI, chr1:219490672–219790221 and chr12:124365252–124665152 for WHRBMIadj respectively) displayed moderate to low LD with each other ($R2 < 0.5$).

Overall, the average number of variants with posterior probability > 0.1 in the PAGE fine-mapped regions was 1.9 (vs. 2.4 in UKBB), and the median 99% CS was 15.5 (vs. 23 in UKBB). For the lead variants in PAGE, the average posterior probability was 0.35 (vs. 0.28 in UKBB), and average CADD was 8.61 (vs. 7.23 in UKBB) (**Table 2).**

In pairwise comparison between the lead (**Table 1**) and the best non-lead variants (**Sup. Table 4**) in each region, we observed that in 6 of the 10 UKBB regions, the non-lead variants demonstrate a higher potential for functionality (or were likely closer proxies to causal variants) (highlighted rows, **Sup. Table 4**), due to higher CADD scores, or being previously reported in multiple populations. In contrast, in only 3 of the 10 PAGE loci (**Sup. Table 4**, highlighted rows), the non-lead variant displayed a higher protentional of functionality compared to the lead variants for their respective regions (**Table 1**) (comparing CADD scores).

Stage 2. Functional annotation was completed in all target loci to characterize likely causal variants operating through gene expression **(Sup. Table 5)**. Overall, in 9 out of 10 regions, the strongest candidates for eQTL were present in PAGE 99% CS (vs 8 in UKBB) **(Table 2)**. Similarly, the most significant eQTLs in all 9 regions were in LD with the lead fine-mapped variants in PAGE (or were the same as lead variant), but the same was observed in 8 regions for UKBB (excluding WHRBMIadj associated region chr12:123880705–124180606 where UKBB and PAGE signals were in loose LD with each other and no strong eQTL with either lead variants, therefore inconclusive) **(Table 2)**.

For BMI, in associated region chr1:177739839–178039226, rs543874 (nearest gene *SEC16B*) had a high CADD score (18.5) and is in LD with 2 subcutaneous adipose tissue eQTL variants (rs6679120 and rs6682862, gene *SEC16B*) making this a good candidate for functional studies. Importantly, this variant is also the lead fine-mapped variant in the PAGE data. In region chr11:27535512–27834103, variant rs10767664 (gene *BDNF-AS*) is the strongest candidate due to tight LD with the index variants, inclusion in both cohorts 99% CS, and a high CADD score (22.1). Other strong QTLs in the region are independent. Finally, for locus chr16:53650985–53916489, the lead variant (in both cohorts) rs1421085 (*FTO*) is the best functional candidate because other significant QTLs in the region are independent *(R2* < 0.1). Additionally, previously reported GWAS significant obesity-associated variants in LD with rs1421085 (*R2* > 0.1), were neither significant QTLs nor had CADD scores as high as rs1421085 (19.58) **(Sup. Table 6)**.

For WHRBMIadj, variant rs10753805 (nearest gene *KIFAP3*) in region chr1:170202330–170501746 **(Sup. Table 5)**, may be the best variant for further functional interrogation, with collective evidence from a strong CADD score (17.4), tight LD with lead variants in both PAGE and UKBB, and being an eQTL with adipose tissue. Other significant eQTLs in the region are functionally independent (i.e., no LD with the index variant). In region chr 1:219317330–219616879, variant rs5781117 (near gene *RP11-392O17.1*) is prioritized through its inclusion in the 99% CS (UKBB) and tight LD with the lead index variant. The other variant in the region with a strong eQTL (rs748273) was not in the 99% CS and had a smaller CADD score (15.5 vs 21.1 for rs5781117). For the region chr2:164522659–164822197, variant rs10187501 (in gene *COBLL1*) was in tight LD with the lead fine mapped variant, was included in the 99% CS (PAGE), and displayed a high CADD score of 19.04. At locus chr3:64567951–64867428, the fine-mapped lead variant rs66815886 (both cohorts) is an interesting candidate for future functional interrogation. Although not significantly associated with gene expression levels in key obesity related tissues in GTEx (only eQTL observed in testes tissue), this variant is in LD with rs76329608 (eQTL in adipose tissue for the same gene [*ADAMTS9-AS2*]). For the region chr6:43640429–43940152, the UKBB lead variant is rs998584 which was included in the 99% CS in both cohorts but with otherwise limited evidence elsewhere. For the locus chr6:126984129–127283578, the lead variant rs577721086 (in both cohorts) is a strong functional candidate given its CADD score of 20.4. Finally, both rs10773049 (lead variant in UKBB) and rs1964599 (lead in PAGE) [near gene *CCDC92*] could be close proxies to functional variants in this region because both displayed associations with adipose tissue expression, not in LD with other observed strong eQTL in the region or each other **(Sup. Table 5)**.

Sensitivity Analyses

Finally, sensitivity analyses implemented in *DENTIST-S* suggested that one fine-mapped region in PAGE (chr16:53650985–53950401) harbored an association outlier **(Sup. Figures 2)**, where a variant (rs113008794) illustrated a higher-than-expected p-value (or lower -log p-value) for statistical association with BMI despite high LD (*R2* > 0.8) with the lead fine mapped variant. Such an association outlier suggests that the quality of fine-mapping of the *FTO* locus in PAGE should be interpreted with caution. Nonetheless, such association outliers were not observed in the UKBB data for the same region, and both

populations prioritized the same variant. No association outliers were observed in WHRBMIadj associated loci in either UKBB or PAGE.

Discussion

In this study we performed fine-mapping of loci associated with BMI and WHRBMIadj in both the multi-ancestry PAGE sample and a similarly sized subset of European ancestry UKBB participants. As complex traits are influenced by a large number of genetic variants (Pasaniuc and Price 2017), functional annotations help facilitate the prioritization of genes and variants when multiple variants are statistically significant in a region (Trynka et al. 2013) There was significant overlap in obesity associated signals across studies, in that lead signals were often shared (same variant or in tight LD ($R2 > 0.9$) with one another). Substantial overlap in likely causal variants across populations suggest a shared relevance of genetic mechanisms underpinning anthropometric traits.

Fine-mapping in the ancestrally diverse PAGE populations produced 95% and 99% CS with fewer variants. Importantly, 99% CS in PAGE loci contained more variants with strong evidence for eQTL effects in obesity relevant tissues when compared to the UKBB 99% CS. Furthermore, more lead candidate functional variants were identified in the PAGE 99% CS in comparison to UKBB data. Taken together, these results are markers of more efficient fine-mapping in multi-ancestry samples and are largely consistent with prior observations (Asimit et al. 2016; Kanai et al. 2021).

For BMI associated region chr1:177770704–178070091, variant rs543874 (*SEC16B*) has been previously linked to obesity (Costa-Urrutia et al. 2020; Lv et al. 2015; Mei et al. 2022); this variant was the lead fine-mapped variant in PAGE and has been reported as the lead variant in at least two previous studies (Fernández-Rhodes et al. 2017; Sahibdeen et al. 2018). In the region chr11:27535512–27834103, rs7929344 (Fernández-Rhodes et al. 2017)(gene *BDNF-AS*) and rs1519480 (Gong et al. 2013) were previously fine-mapped, but our 99% CS in both cohorts contained two additional variants of interest, rs10767664 and rs6265 (*BDNF-AS*). The latter variant is a missense and has been associated with energy intake (Daily and Park 2017), but rs10767664 has been frequently linked to glucose regulation and insulin resistance (de Luis et al. 2017; de Luis et al. 2018), with presumed regulatory effects in adipose tissue, and a high CADD score **(Sup. Table 4)**, and therefore more likely the functional variant in the region. Finally, in region chr16:53650985–53916489, variant rs62048402 was previously fine-mapped (Daily and Park 2017; Gong et al. 2013), but our proposed lead fine-mapped and index variant, rs1421085 (*FTO*), has strong evidence for functionality for multiple reasons. First, none of the previously reported FTO index variants in the region were in LD with rs1421085 ($R2 > 0.1$) and none displayed high CADD scores. Second, the intronic variant rs1421085 has been fine mapped in previous multi-ancestry analyses (Akiyama et al. 2014). Third, studies in mice demonstrated that variants in this region interacted with the promoter of *Irx3* (Smemo et al. 2014). In addition, functional characterization of rs1421085 in the *FTO* locus demonstrated a doubling of IRX3 expression during early adipocyte differentiation (Claussnitzer et al. 2015). Lastly, in a very recent study, human to mice conservation of the rs1421085 regulatory effect on *IRX3* gene expression has been reported (Laber et al. 2021).

For WHRBMIadj, in associated region chr1:170202330–170501746, we propose rs10753805 as the most likely functional variant. We did not find any other tightly linked fine-mapped variants within this region in the literature. In region chr1:219317330–219616879, we propose variant rs5781117 as the most likely functional variant. Previously reported index variant rs12025363 (Zhang et al. 2022) is in moderate LD with our suggested fine-mapped lead variants (R2 ~ 0.4), and was not significantly associated with gene expression in adipose tissue (*RP11-392O17.1* gene). Similarly, for the region chr2:164522659–164822197, rs10187501 (*COBLL1*) is also a novel observation and a candidate for causal association with obesity (previously linked to type 2 diabetes (Ocvirk 2020), elevated CADD score, and a significant sQTL for *COBLL1* in the adipose tissue). For the region chr3:64567951–64867428, the lead fine-mapped and proposed functional variant rs66815886 (*ADAMTS9-AS2*) is consistent with prior findings (Ng et al. 2017). In the region chr6:43640429–43940152, lead fine-mapped variant rs998584 (*VEGF* gene) has been previously suggested as functional (Wu et al. 2019). In the region chr6:126984129–127283578, while rs72959041 had been previously fine mapped as possibly functional for the *RSPO3* gene (Horikoshi et al. 2015), fine-mapping in both PAGE and the UKBB highlighted another known variant, rs577721086 (Christakoudi et al. 2021) and this conclusion corroborates functional evidence. Finally, in locus 12:123880705–124180606, while rs10773049 (the lead UKBB variant) was previously fine-mapped (Wu et al. 2019), rs1964599 (the lead in PAGE) has low LD with the UKBB lead fine-mapped variant (R2 ~ 0.1) and also displays strong evidence of functionality.

While illustrating a distinct improvement in fine-mapping efficiency in ancestrally diverse populations, there were notable limitations in our study. First, the fine-mapping method employed in this study assumes one causal variant per locus. Therefore, we could not characterize additional signals (if present) and investigate the fine-mapping efficiency in such a setting, which likely represents the biological reality of many GWAS signals. Another limiting factor was possible heterogeneity in variants' effect size within PAGE (Martin et al. 2019), leading to inconsistent GWAS statistical significance across populations (Huang et al. 2022; Langlois et al. 2016; Liu et al. 2019; Tan et al. 2014) in PAGE. This has been previously evidenced in lower predictive power of polygenic risk scores across populations (Duncan et al. 2019; Kamiza et al. 2022), and partly attributed to heterogeneity in causal architecture and gene-environment interactions (Galinsky et al. 2019). However, in 7 of 10 loci interrogated, the lead variant in PAGE was the same or in very tight LD with lead variants in UKBB demonstrating limited discordance. Additionally, PAGE cohorts were genotyped using multiple genotyping arrays (Bien et al. 2016; Matise et al. 2011) resulting in power loss in GWAS meta-analysis (due to differential variant missingness) (2 et al. 2021), and increased risk of errors further confounding GWAS (Maier et al. 2020; Wei and Nielsen 2019). Yet, an average of 61% of variants overlapped in PAGE and UKBB and complementary functional assessment, incorporation of expression QTLs, and literature review identified almost a complete overlap of the candidate functional variants in the 99% CS, except for rs5781117 (where it was present in the GWAS set but not in 99% CS). Finally, sensitivity analyses with DENTIST-S showed only one locus (chr16:53650985–53916489, *FTO*) harbored association outliers, where a variant in tight LD with the lead fine-mapped illustrated higher than expected association p-value with BMI, and therefore the fine-

mapping results could be affected. Nonetheless, our suggested prioritized variant, rs1421085, is now a known compelling functional variant in this region.

Nonetheless, our study has its strength. First, this was the first comprehensive fine-mapping study of obesity associated loci that was conducted in parallel in both EA and diverse ancestry population, purposely matched by sample size to remove methodological limitations inherent in such comparisons. Indeed, harmonization of phenotypes and use of sample size averaged LD minimized the bias in calibration of meta-analyzed fine-mapping. Second, Incorporation of sensitivity analysis enabled detection of associations' outliers and identified loci with questionable fine-mapping accuracy. Third, comparative fine-mapping in conjunction with comprehensive functional annotation interrogation improved fine mapping resolution to identify potentially functional variants in regions where previous studies have been inconclusive. Finally, we encourage functional studies for 3 WHRBMIadj loci, where we identified strong functional candidates within 99% CS in both UKBB and PAGE.

In conclusion, results illustrated improved efficiency in fine-mapping functional variants in multi-ancestry samples in obesity associated genomic regions. While it is methodologically challenging to pinpoint causal variants with small effect sizes in complex traits, incorporation of ancestrally diverse populations with distinct genetic architectures and functional annotations has reduced the set of candidate variants for causal assessment in future molecular studies.

Declarations

Acknowledgement

We thank Dr. Masahiro. Kanai (mkanai@broadinstitute.org) for his contribution to data analyses, particularly with implementation of SLALOM.

Conflict of Interest

No conflict of interest.

Data Availability

PAGE accession numbers are: phs000356 PAGE collectively; phs000223 PAGE-ARIC and phs000280 ARIC Cohort; phs000555 PAGE-HCHS/SOL and phs000810 HCHS/SOL Cohort; phs000220 PAGE-MEC; phs000227 PAGE-WHI and phs000200 WHI Cohort; phs000925 PAGE-IPM-BioMe.

Funding

This work was supported by UK Biobank application 25953. The PAGE Study is funded by the National Human Genome Research Institute with co-funding from the National Institute on Minority Health and Health Disparities. Assistance with data management, data integration, data dissemination, genotype imputation, ancestry deconvolution, population genetics, analysis pipelines and general study

coordination was provided by the PAGE Coordinating Center (NI-HU01HG007419). Genotyping services were provided by the Center for Inherited Disease Research, which is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University, contract number HHSN268201200008I. Genotype data quality control and quality assurance services were provided by the Genetic Analysis Center in the Biostatistics Department of the University of Washington, through support provided by the Center for Inherited Disease Research contract. PAGE was also funded by grants R56HG010297 and R01HG010297. PAGE data and materials included in this report were funded through the following studies and organizations: The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies. We thank all participants and all our recruiters who have assisted and continue to assist in data collection and management. We are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sina. The MEC characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded by the National Cancer Institute (R37CA54281, R01CA63, P01CA33619, U01CA136792 and U01CA98758). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, and US Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004 and 75N92021D00005. The HCHS/SOL is a collaborative study supported by contracts from the National Heart, Lung and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC 65235), University of Illinois at Chicago (HHSN268201300003I / N01-HC-65236 Northwestern University), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centres/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities; National Institute on Deafness and Other Communication Disorders; National Institute of Dental and Craniofacial Research; National Institute of Diabetes and Digestive and Kidney Diseases; National Institute of Neurological Disorders and Stroke; and NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03. The Atherosclerosis Risk in Communities (ARIC) study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The Coronary Artery Risk Development in Young Adults Study (CARDIA): The Coronary Artery Risk Development in Young Adults Study (CARDIA) is supported by contracts HHSN268201800003I, HHSN268201800004I, HHSN268201800005I, HHSN268201800006I, and HHSN268201800007I from the National Heart, Lung, and Blood Institute (NHLBI). CARDIA is also partially supported by the Intramural Research Program of

the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). GWAS genotyping and data analyses were funded in part by grants U01-HG004729 and R01-HL093029 from the National Institutes of Health to Dr. Myriam Fornage. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005. MYA is funded by NIDDK grant # 3R01DK122503 – 02W1. Kari North is supported by R01HD057194, R01 DK122503, R01HG010297, R01HL142302, R01HL143885, R01HG009974, and R01DK101855. The following grants supported this study: R01HL151152 (North, Graff), R01HG011345 (North, Graff).

Contributions

MYA, MG, KEN participated in the study conception. MYA and MG performed data analyses. MYA, MG, and KEN drafted the manuscript. All co-authors performed critical reviews. KEN and MG supervised the study. All authors read and approved the final manuscript.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Corresponding author

Correspondence to Mohammad Yaser Anwar.

Ethics Approval & Consent to Participate

All study participants provided written informed consent and each study was approved by relevant institutional review board.

All methods were performed in accordance with the relevant guidelines and regulations set by Declaration of Helsinki.

HCHS/SOL- The institutional review board at the coordinating center (University of North Carolina Office of Human Research Ethics, 07-1003) and Board Office, 200601-0471; University of California-San Diego Human Research Protection Program, 3677; University of Miami Human Subject Research Office, FWA00002247) approved study protocols. All participants gave informed consent.

BioMe- Program for the Protection of Human Subjects, Mount Sinai Health System, Icahn School of Medicine at Mount Sinai

CARDIA- University of Texas Health Science Center at Houston

WHI- Fred Hutchison Cancer Research Center

ARIC- The Johns Hopkins Medicine Institutional Review Board

Cameron County- Committee for the Protection of Human Subjects at the University of Teas Health Sciences Center at Houston; Human Research Protections Program at Vanderbilt University

MEC- USC Institutional Review Board

Consent to publish

Not applicable.

References

1. 2 WgWglPGAASJKM leaderM, 29 MatmgpFH (2021) 1 Matmm-aMJ, Manuscript analyses team member: heritability m, supplements, PHEWAS Matm, randomization MatmM, projection MatmP, prioritization g, Mapping the human genetic architecture of COVID-19. Nature 600: 472–477
2. Akiyama K, Takeuchi F, Isono M, Chakrawarthy S, Nguyen QN, Wen W, Yamamoto K, Katsuya T, Kasturiratne A, Pham ST (2014) Systematic fine-mapping of association with BMI and type 2 diabetes at the FTO locus by integrating results from multiple ethnic groups. PLoS ONE 9:e101329
3. Asimit JL, Hatzikotoulas K, McCarthy M, Morris AP, Zeggini E (2016) Trans-ethnic study design approaches for fine-mapping. Eur J Hum Genet 24:1330–1336
4. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32:1493–1501
5. Bien SA, Wojcik GL, Zubair N, Gignoux CR, Martin AR, Kocarnik JM, Martin LW, Buyske S, Haessler J, Walker RW (2016) Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. PLoS ONE 11:e0167758
6. Biobank U (2007) Protocol for a large-scale prospective epidemiological resource
7. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J (2017) Genome-wide genetic data on ~ 500,000 UK Biobank participants. BioRxiv: 166298
8. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209
9. Caballero A, Tenesa A, Keightley PD (2015) The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. Genetics 201:1601–1613
10. CALiCo-SOL LF-R, Fernandez-Rhodes LPopulation Architecture using Genomics and Epidemiology (PAGE)
11. Carty CL, Bhattacharjee S, Haessler J, Cheng I, Hindorff LA, Aroda V, Carlson CS, Hsu C-N, Wilkens L, Liu S (2014) Analysis of metabolic syndrome components in > 15 000 african americans identifies

pleiotropic variants: results from the population architecture using genomics and epidemiology study. Circulation: Cardiovasc Genet 7:505–513

12. Challis B, Coll A, Yeo G, Pinnock S, Dickson S, Thresher R, Dixon J, Zahn D, Rochford J, White A (2004) Mice lacking pro-opiomelanocortin are sensitive to high-fat feeding but respond normally to the acute anorectic effects of peptide-YY3-36. Proceedings of the National Academy of Sciences 101: 4695–4700
13. Chen H-H, Petty LE, Bush W, Naj AC, Below JE (2019) GWAS and Beyond: Using Omics Approaches to Interpret SNP Associations. Curr genetic Med Rep 7:30–40
14. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, Schaid DJ (2015) Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. Genetics 200:719–736
15. Chen W, Wu Y, Zheng Z, Qi T, Visscher PM, Zhu Z, Yang J (2021) Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. Nat Commun 12:1–10
16. Christakoudi S, Evangelou E, Riboli E, Tsilidis KK (2021) GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. Sci Rep 11:1–18
17. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V (2015) FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med 373:895–907
18. Consortium G (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369:1318–1330
19. Consortium GP (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56
20. Costa-Urrutia P, Abud C, Franco-Trecu V, Colistro V, Rodríguez-Arellano ME, Alvarez-Fariña R, Acuna Alonso V, Bertoni B, Granados J (2020) Effect of 15 BMI-associated polymorphisms, reported for Europeans, across ethnicities and degrees of Amerindian ancestry in Mexican children. Int J Mol Sci 21:374
21. Daily JW, Park S (2017) Interaction of BDNF rs6265 variants and energy and protein intake in the risk for glucose intolerance and type 2 diabetes in middle-aged adults. Nutrition 33:187–194
22. de Luis DA, Aller R, Izaola O, Primo D, Romero E (2017) rs10767664 gene variant in brain-derived neurotrophic factor is associated with diabetes mellitus type 2 in Caucasian females with obesity. Annals of Nutrition and Metabolism 70:286–292
23. de Luis DA, Ovalle HF, Izaola O, Primo D, Aller R (2018) RS 10767664 gene variant in Brain Derived Neurotrophic Factor (BDNF) affect metabolic changes and insulin resistance after a standard hypocaloric diet. J Diabetes Complicat 32:216–220
24. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B (2019) Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 10:1–9

25. Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, Haas M, Bick A, Zekavat SM, Nomura A (2018) Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. Nat Commun 9:1–8
26. Fernández-Rhodes L, Gong J, Haessler J, Franceschini N, Graff M, Nishimura KK, Wang Y, Highland HM, Yoneyama S, Bush WS (2017) Trans-ethnic fine-mapping of genetic loci for body mass index in the diverse ancestral populations of the Population Architecture using Genomics and Epidemiology (PAGE) Study reveals evidence for multiple signals at established loci. Hum Genet 136:771–800
27. Flister MJ, Tsaih S-W, O'Meara CC, Endres B, Hoffman MJ, Geurts AM, Dwinell MR, Lazar J, Jacob HJ, Moreno C (2013) Identifying multiple causative genes at a single GWAS locus. Genome Res 23:1996–2002
28. Galinsky KJ, Reshef YA, Finucane HK, Loh PR, Zaitlen N, Patterson NJ, Brown BC, Price AL (2019) Estimating cross-population genetic correlations of causal effect sizes. Genet Epidemiol 43:180–188
29. Giral H, Landmesser U, Kratzer A (2018) Into the wild: GWAS exploration of non-coding RNAs. Front Cardiovasc Med 5:181
30. Gong J, Schumacher F, Lim U, Hindorff LA, Haessler J, Buyske S, Carlson CS, Rosse S, Bůžková P, Fornage M (2013) Fine mapping and identification of BMI loci in African Americans. Am J Hum Genet 93:661–671
31. Gorber SC, Tremblay MS (2010) The bias in self-reported obesity from 1976 to 2005: a Canada–US comparison. Obesity 18:354–361
32. Hinney A, Vogel CI, Hebebrand J (2010) From monogenic to polygenic obesity: recent advances. Eur Child Adolesc Psychiatry 19:297–310
33. Hodge SE, Greenberg DA (2016) How can we explain very low odds ratios in GWAS? I. Polygenic models. Human Hered 81:173–180
34. Horikoshi M, MÓ“gi R, van de Bunt M, Surakka I, Sarin A-P, Mahajan A, Marullo L, Thorleifsson G, HÓ“gg S, Hottenga J-J (2015) Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. PLoS Genet 11:e1005230
35. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E (2014) Identifying causal variants at loci with multiple signals of association. Genetics 198:497–508
36. Hu Y, Bien SA, Nishimura KK, Haessler J, Hodonsky CJ, Baldassari AR, Highland HM, Wang Z, Preuss M, Sitlani CM (2021) Multi-ethnic genome-wide association analyses of white blood cell and platelet traits in the Population Architecture using Genomics and Epidemiology (PAGE) study. BMC Genomics 22:1–11
37. Huang QQ, Sallah N, Dunca D, Trivedi B, Hunt KA, Hodgson S, Lambert SA, Arciero E, Wright J, Griffiths C (2022) Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistani and Bangladeshi individuals. Nat Commun 13:1–11
38. Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, Berkemeier LR, Gu W, Kesterson RA, Boston BA, Cone RD (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell 88:131–141

39. Kamiza AB, Toure SM, Vujkovic M, Machipisa T, Soremekun OS, Kintu C, Corpas M, Pirie F, Young E, Gill D (2022) Transferability of genetic risk scores in African populations.Nature Medicine:1–4
40. Kanai M, Elzur R, Zhou W, Wu K-HH, Rasheed H, Tsuo K, Hirbo JB, Wang Y, Bhattacharya A, Zhao H (2022) Meta-analysis fine-mapping is often miscalibrated at single-variant resolution.Cell Genomics:100210
41. Kanai M, Ulirsch JC, Karjalainen J, Kurki M, Karczewski KJ, Fauman E, Wang QS, Jacobs H, Aguet F, Ardlie KG (2021) Insights from complex trait fine-mapping across diverse populations. medRxiv
42. Kichaev G, Pasaniuc B (2015) Leveraging functional-annotation data in trans-ethnic fine-mapping studies. Am J Hum Genet 97:260–271
43. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315
44. Koch L (2020) Exploring human genomic diversity with gnomAD. Nat Rev Genet 21:448–448
45. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A (1998) Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet 19:155–157
46. Laber S, Forcisi S, Bentley L, Petzold J, Moritz F, Smirnov KS, Al Sadat L, Williamson I, Strobel S, Agnew T (2021) Linking the FTO obesity rs1421085 variant circuitry to cellular, metabolic, and organismal phenotypes in vivo. Sci Adv 7:eabg0108
47. Langlois C, Abadi A, Peralta-Romero J, Alyass A, Suarez F, Gomez-Zamudio J, Burguete-Garcia AI, Yazdi FT, Cruz M, Meyre D (2016) Evaluating the transferability of 15 European-derived fasting plasma glucose SNPs in Mexican children and adolescents. Sci Rep 6:1–8
48. Lin D-Y, Tao R, Kalsbeek WD, Zeng D, Gonzalez IIF, Fernández-Rhodes L, Graff M, Koch GG, North KE, Heiss G (2014) Genetic association analysis under complex survey sampling: the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet 95:675–688
49. Liu HY, Alyass A, Abadi A, Peralta-Romero J, Suarez F, Gomez-Zamudio J, Audirac A, Parra EJ, Cruz M, Meyre D (2019) Fine-mapping of 98 obesity loci in Mexican children. Int J Obes 43:23–32
50. Lv D, Zhang D-D, Wang H, Zhang Y, Liang L, Fu J-F, Xiong F, Liu G-L, Gong C-X, Luo F-H (2015) Genetic variations in SEC16B, MC4R, MAP2K5 and KCTD15 were associated with childhood obesity and interacted with dietary behaviors in Chinese school-age population. Gene 560:149–155
51. Magavern EF, Gurdasani D, Ng FL, Lee SSJ (2022) Health equality, race and pharmacogenomics. Br J Clin Pharmacol 88:27–33
52. Maier R, Akbari A, Wei X, Patterson N, Nielsen R, Reich D (2020) No statistical evidence for an effect of CCR5-Δ 32 on lifespan in the UK Biobank cohort. Nat Med 26:178–180
53. Manolio TA (2009) Collaborative genome-wide association studies of diverse diseases:. programs of the NHGRI's office of population genomics
54. Mao L, Fang Y, Campbell M, Southerland WM (2017) Population differentiation in allele frequencies of obesity-associated SNPs. BMC Genomics 18:1–16

55. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51:584–591
56. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, Haiman CA, Heiss G, Kooperberg C, Marchand LL (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
57. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The ensembl variant effect predictor. Genome Biol 17:1–14
58. Mei H, Yin B, Yang W, Zhang J, Lu H, Qi X, Mei W, Zhang H, Zhang J (2022) Associations between Gene-Gene Interaction and Overweight/Obesity of 12-Month-Old Chinese Infants. BioMed research international 2022.
59. Ng MC, Graff M, Lu Y, Justice AE, Mudgal P, Liu C-T, Young K, Yanek LR, Feitosa MF, Wojczynski MK (2017) Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. PLoS Genet 13:e1006719
60. Ocvirk V (2020) Molecular and cellular mechanisms underlying the GRB14/COBLL1 diabetes risk locus. Technische Universität, München
61. Pasaniuc B, Price AL (2017) Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet 18:117–127
62. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P (2008) Discerning the ancestry of European Americans in genetic association studies. PLoS Genet 4:e236
63. Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14
64. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26:2336–2337
65. Ralph P, Coop G (2013) The geography of recent genetic ancestry across Europe. PLoS Biol 11:e1001555
66. Riancho JA (2012) Genome-wide association studies (GWAS) in complex diseases: advantages and limitations. Reumatol Clin 8:56–57
67. Sahibdeen V, Crowther NJ, Soodyall H, Hendry LM, Munthali RJ, Hazelhurst S, Choudhury A, Norris SA, Ramsay M, Lombard Z (2018) Genetic variants in SEC16B are associated with body composition in black South Africans. Nutr diabetes 8:1–10
68. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507:371–375

69. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J (2016) PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics 32:3207–3209
70. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12:e1001779
71. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D (2019) Benefits and limitations of genome-wide association studies. Nat Rev Genet 20:467–484
72. Tan L-J, Zhu H, He H, Wu K-H, Li J, Chen X-D, Zhang J-G, Shen H, Tian Q, Krousel-Wood M (2014) Replication of 6 obesity genes in a meta-analysis of genome-wide association studies from diverse ancestries. PLoS ONE 9:e96149
73. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet 45:124–130
74. van de Bunt M, Cortes A, Consortium I, Brown MA, Morris AP, McCarthy MI (2015) Evaluating the performance of fine-mapping strategies at common variant GWAS loci. PLoS Genet 11:e1005535
75. Vance KW, Sansom SN, Lee S, Chalei V, Kong L, Cooper SE, Oliver PL, Ponting CP (2014) The long non-coding RNA P aupar regulates the expression of both local and distal genes. EMBO J 33:296–311
76. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90:7–24
77. Wakefield J (2007) A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 81:208–227
78. Wakefield J (2009) Bayes factors for genome-wide association studies: comparison with P-values. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society 33:79–86
79. Wei X, Nielsen R (2019) CCR5-Δ 32 is deleterious in the homozygous state in humans. Nat Med 25:909–910
80. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001–D1006
81. Willems SM, Wright DJ, Day FR, Trajanoska K, Joshi PK, Morris JA, Matteini AM, Garton FC, Grarup N, Oskolkov N (2017) Large-scale GWAS identifies multiple loci for hand grip strength providing biological insights into muscular fitness. Nat Commun 8:1–12
82. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191
83. Witte JS (2010) Genome-wide association studies and beyond. Annu Rev Public Health 31:9

84. Wu Y, Broadaway KA, Raulerson CK, Scott LJ, Pan C, Ko A, He A, Tilford C, Fuchsberger C, Locke AE (2019) Colocalization of GWAS and eQTL signals at loci with multiple signals identifies additional candidate genes for body fat distribution. Hum Mol Genet 28:4161–4172
85. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM (2018) Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. Hum Mol Genet 27:3641–3649
86. Zhang X, Li T-Y, Xiao H-M, Ehrlich KC, Shen H, Deng H-W, Ehrlich M (2022) Epigenomic and transcriptomic prioritization of candidate obesity-risk regulatory GWAS SNPs. Int J Mol Sci 23:1271
87. Zhang Y-M, Jia Z, Dunwell JM (2019) The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, vol 10. Frontiers Media SA, p 100
88. Zhao B, Zhang J, Ibrahim JG, Luo T, Santelli RC, Li Y, Li T, Shan Y, Zhu Z, Zhou F (2021) Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). Mol Psychiatry 26:3943–3955
89. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50:1335–1341

Tables

Table 1 and 2 are available in the Supplementary Files section.

Box

Box 1 is available in the Supplementary Files section.

Appendix

The Appendix is not available with this version

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Fig.1.updated.jpg
- Supp.Fig.2.updated.jpg
- Supple.Tables.updated.xlsx
- Box1.png
- Table1.xlsx
- Table2.xlsx