Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits

Anne E. Justice et al.#

Few genome-wide association studies (GWAS) account for environmental exposures, like smoking, potentially impacting the overall trait variance when investigating the genetic contribution to obesity-related traits. Here, we use GWAS data from 51,080 current smokers and 190,178 nonsmokers (87% European descent) to identify loci influencing BMI and central adiposity, measured as waist circumference and waist-to-hip ratio both adjusted for BMI. We identify 23 novel genetic loci, and 9 loci with convincing evidence of gene-smoking interaction (GxSMK) on obesity-related traits. We show consistent direction of effect for all identified loci and significance for 18 novel and for 5 interaction loci in an independent study sample. These loci highlight novel biological functions, including response to oxidative stress, addictive behaviour, and regulatory functions emphasizing the importance of accounting for environment in genetic analyses. Our results suggest that tobacco smoking may alter the genetic susceptibility to overall adiposity and body fat distribution.
Recent genome-wide association studies (GWAS) have described loci implicated in obesity, body mass index (BMI) and central adiposity. Yet most studies have ignored environmental exposures with possibly large impacts on the trait variance. Variants that exert genetic effects on obesity through interactions with environmental exposures often remain undiscovered due to heterogeneous main effects and stringent significance thresholds. Thus, studies may miss genetic variants that have effects in subgroups of the population, such as smokers. It is often noted that currently smoking individuals display lower weight/BMI and higher waist circumference (WC) as compared to nonsmokers. Smokers also have the smallest fluctuations in weight over ~20 years compared to those who have never smoked or have stopped smoking. Also, heavy smokers (>20 cigarettes per day [CPD]) and those that have smoked for more than 20 years are at greater risk for obesity than non-smokers or light to moderate smokers (≤20 CPD)9,10. Men and women gain weight rapidly after smoking cessation and although studies suggest that tobacco use suppresses appetite, smoking may result in an increased metabolic rate. Identifying genes that influence adiposity and interact with smoking may help us clarify pathways through which smoking influences weight and central adiposity.

A comprehensive study that evaluates smoking in conjunction with genetic contributions is warranted. Using GWAS data from the Genetic Investigation of Anthropometric Traits (GIANT) Consortium, we identified 23 novel genetic loci, and 9 loci with convincing evidence of gene-smoking interaction (GxSMK) on obesity, assessed by BMI and central obesity independent of overall body size, assessed by WC adjusted for BMI (WCadjBMI) and waist-to-hip ratio adjusted for BMI (WHRadjBMI). By accounting for smoking status, we focus both on genetic variants observed through their main effects and GxSMK effects to increase our understanding of their action on adiposity-related traits. These loci highlight novel biological functions, including response to oxidative stress, addictive behaviour and regulatory functions emphasizing the importance of accounting for environment in genetic analyses. Our results suggest that smoking may alter the genetic susceptibility to overall adiposity and body fat distribution.

Results

GWAS discovery overview. We meta-analysed study-specific association results from 57 Hapmap-imputed GWAS and 22 studies with Metabochip, including up to 241,258 (87% European descent) individuals (51,080 current smokers and 190,178 nonsmokers) while accounting for current smoking (SMK) (Methods section, Supplementary Fig. 1, Supplementary Tables 1–4). For primary analyses, we conducted meta-analyses across ancestries and sexes. For secondary analyses, we conducted meta-analyses in European-descent studies alone and sex-specific meta-analyses (Tables 1–4, Supplementary Data 1–6). We considered four analytical approaches to evaluate the effects of smoking on genetic associations with adiposity traits (Fig. 1, Methods section). Approach 1 (SNPadjSMK) examined genetic associations after adjusting for SMK. Approach 2 (SNPpoint) considered the joint impact of main effects adjusted for SMK + interaction effects. Approach 3 focused on interaction effects (SNPint); Approach 4 followed up loci from Approach 1 for interaction effects (SNPscreen). Results from Approaches 1–3 were considered genome-wide significant (GWS) with a P-value <5 × 10^{-8} while Approach 4 used Bonferroni adjustment after screening. Lead variants >500 kb from previous associations with BMI, WCadjBMI, and WHRadjBMI were considered novel. All association results are reported with effect estimates oriented on the trait increasing allele in the current smoking stratum.

Across the three adiposity traits, we identified 23 novel associated genetic loci (6 for BMI, 11 for WCadjBMI, 6 for WHRadjBMI) and nine having significant GxSMK interaction effects (2 for BMI, 2 for WCadjBMI, 5 for WHRadjBMI; Fig. 1, Tables 1–4, Supplementary Data 1–6). We provide a comprehensive comparison with previously-identified loci by trait in supplementary material (Supplementary Data 7, Supplementary Note 1).

Accounting for smoking status. For primary meta-analyses of BMI (combined ancestries and sexes), 58 loci reached GWS in Approach 1 (SNPadjSMK; Supplementary Data 1, Supplementary Figs 2 and 3), including two novel loci near SOX11 and SRRM1P2 (Table 1). Three more BMI loci were identified using Approach 2 (SNPint), including a novel locus near CCDC93 (Supplementary Figs 4 and 5). For WCadjBMI, 62 loci reached GWS for Approach 1 (SNPadjSMK) and two more for Approach 2 (SNPpoint), including eight novel loci near KIF1B, HDLBP, DOCK3, ADAATM3, CDK6, GSDMC, TMEM38B and ARFGF2 (Table 1, Supplementary Data 2, Supplementary Figs 2–5). Lead variants near PSMB10 from Approaches 1 and 2 (rs14178 and rs113090, respectively) are >500 kb from a previously-identified WCadjBMI-associated variant (rs16957304); however, after conditioning on the known variant, our signal is attenuated (P_{conditional} = 3.02 × 10^{-2} and P_{conditional} = 5.22 × 10^{-3}), indicating that this finding is not novel. For WHRadjBMI, 32 loci were identified in Approach 1 (SNPadjSMK), including one novel locus near HLA-C, with no additional loci in Approach 2 (SNPpoint; Table 1, Supplementary Data 3, Supplementary Figs 2–5).

We used GCTA to identify loci from our primary meta-analyses that harbour multiple independent SNPs (Methods section, Supplementary Tables 5–7). Conditional analyses revealed no secondary signals within 500 kb of our novel lead SNPs. Additionally, we performed conditional association analyses to determine whether our novel variants were independent of previous GWAS loci within 500 kb that are associated with related traits of interest. All BMI-associated SNPs were independent of previously identified GWS associations with anthropometric and obesity-related traits. Seven novel loci for WCadjBMI were near previous associations with related anthropometric traits. Of these, association signals for rs6743226 near HDLBP, rs10269774 near CDK6, and rs6012558 near ARFGF2 were attenuated (P_{conditional} > 1E^{-5} and β decreased by half) after conditioning on at least one nearby height and hip circumference adjusted for BMI (HIPadjBMI) SNP, but association signals remained independent of other related SNP-trait associations. For WHRadjBMI, our GWAS signal was attenuated by conditioning on two known height variants (rs6457374 and rs2247056), but remained significant in other conditional analyses. Given high correlations among waist, hip and height, these results are not surprising.

Several additional loci were identified for Approaches 1 and 2 in secondary meta-analysis (Table 2, Supplementary Data 1–6, Supplementary Fig. 6). For BMI, 2 novel loci were identified by Approach 1, including 1 near EPHA3 and 1 near INADL. For WCadjBMI, 2 novel loci were identified near RAII4 and PRNP. For WHRadjBMI, five novel loci were identified in secondary meta-analyses near BBX, TRBI1, EHMT2, SMIM2 and EYA4. A comprehensive summary of nearby genes for all novel loci and their potential biological relevance is available in Supplementary Note 2.
Modification of genetic predisposition by smoking. Approach 3 directly evaluated GxSMK interaction (SNPint; Table 3, Supplementary Data 1–6, Fig. 2, Supplementary Figs 9 and 10). For primary meta-analysis of BMI, two loci reached GWS including a previously identified GxSMK interaction locus near CHRN4B (ref. 3), and a novel locus near INPP4B. Both loci exhibit GWS effects on BMI in smokers and no effects in nonsmokers. For CHRN4B (cholinergic nicotine receptor B4), the variant minor allele (G) exhibits a decreasing effect on BMI in smokers and no effect in nonsmokers (b\textsubscript{smk} = −0.047) but no effect in nonsmokers (b\textsubscript{nonsmk} = 0.002). Previous studies identified nearby SNPs in high LD associated with smoking (nonsynonymous, rs16969968 in CHRN4A)\textsuperscript{3} and arterial calcification (rs3825807, a missense variant in ADAMTS7)\textsuperscript{18}. Conditioning on these variants attenuated our interaction effect but did not eliminate it (Supplementary Table 7), suggesting a complex relationship between smoking, obesity, heart disease, and genetic variants in this region. Importantly, the CHRNA5-CHRAS3-CHRN4B gene cluster has been associated with lower BMI in current smokers\textsuperscript{3}, but with higher BMI in never smokers\textsuperscript{3}, evidence supporting the lack of association in nonsmokers as well as a lack of previous GWAS findings on 15q25 (Supplementary Data 8)\textsuperscript{1}. The CHRNA5-CHRAS3-CHRN4B genes encode the nicotinic acetylcholine receptor (nAChR) subunits \textalpha\textsubscript{3}, \textalpha\textsubscript{5} and \beta\textsubscript{4}, which are expressed in the central nervous system\textsuperscript{17}. Nicotine has differing effects on the body and brain, causing changes in metabolism and feeding behaviours\textsuperscript{18}. These findings suggest smoking exposure may modify genetic effects on 15q24-25 to influence smoking-related diseases, such as obesity, through distinct pathways.

In primary meta-analyses of WCadjBMI, one novel GWS locus (near GRIN2A) with opposite effect directions by smoking status was identified for Approach 3 (SNPint; Table 3, Supplementary Data 2, Fig. 2, Supplementary Figs 9 and 10). The T allele of rs14141488 increases WCadjBMI in current smokers and decreases it in nonsmokers (b\textsubscript{smk} = 0.037, b\textsubscript{nonsmk} = −0.015). In secondary meta-analysis of European women-only, we identified an interaction between rs6076699, near PRNP, and SMK on WCadjBMI (Table 4, Supplementary Data 5, Supplementary Fig. 6), a locus also identified in Approach 2 (SNPjoint) for European women. The major allele, A, has a positive effect on current smokers as compared to a weaker and negative effect on WC in nonsmokers (b\textsubscript{smk} = 0.169, b\textsubscript{nonsmk} = −0.070), suggesting why this variant remained undetected in previous GWAS of WCadjBMI (Supplementary Data 8).

Approach 4 (SNPscreen; Fig. 1, Methods section) evaluated GxSMK interactions after screening SNPadjSMK results (from

<table>
<thead>
<tr>
<th>Trait</th>
<th>Approach 1</th>
<th>Approach 2</th>
<th>Approach 3</th>
<th>Approach 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>65 / 4</td>
<td>57 / 2</td>
<td>2 / 1</td>
<td>1 / 0</td>
</tr>
<tr>
<td>WCadjBMI</td>
<td>72 / 9</td>
<td>66 / 5</td>
<td>2 / 2</td>
<td>0</td>
</tr>
<tr>
<td>WHRadjBMI</td>
<td>44 / 5</td>
<td>24 / 2</td>
<td>0</td>
<td>5 / 0</td>
</tr>
</tbody>
</table>

Figure 3 presents analytical power for Approaches 1 and 2 while Supplementary Table 8 and Supplementary Fig. 7 present simulation results to evaluate type 1 error (Methods section). A heat map cross-tabulates P-values for Approaches 1 and 2 along with Approach 3 examining interaction only (Supplementary Fig. 8). We demonstrate that the two approaches yield valid type 1 error rates and that Approach 1 can be more powerful to find associations given zero or negligible quantitative interactions, whereas Approach 2 is more efficient in finding associations when interaction exists.
Approach 1) using Bonferroni-correction (Methods section, Tables 3–4, Supplementary Data 1–6). We identified two SNPs, near LYPLAL1 and RSPO3, with significant interaction; both have previously published main effects on anthropometric traits. These loci exhibit effects on WHRadjBMI in nonsmokers, but not in smokers (Fig. 2). In secondary meta-analyses, we identified three known loci with significant GxSMK interaction effects on WHRadjBMI near MAP3K1, HOXC4-HOXC6 and JUND (Table 4, Supplementary Data 3 and 6). We identified rs1809420, near CHRNA5-CHRNA3-CHRNB4, for BMI in the men-only, combined-ancestries meta-analysis (Supplementary Data 1).

Power calculations demonstrate that Approach 4 has increased power to identify SNPs that show (i) an effect in one stratum (smokers or nonsmokers) and a less pronounced but concordant effect in the other stratum, or (ii) an effect in the larger nonsmoker stratum and no effect in smokers (Fig. 3). In contrast, Approach 3 has increased power for SNPs that show (i) an effect in the smaller smoker stratum and no effect in nonsmokers, or (ii) an opposite effect between smokers and nonsmokers (Fig. 3). In secondary meta-analyses, we identified *loci identified for Approach 3) remain to be discovered as more variance in the trait can be explained as we drop the threshold for significance.

**Enrichment of genetic effects by smoking status.** When examining the smoking specific effects for BMI and WCadjBMI loci in our meta-analyses, no significant enrichment of genetic effects by smoking status were noted. (Fig. 2, Supplementary Figs 11 and 12). However, our results for WHRadjBMI were enriched for loci with a stronger effect in nonsmokers as compared to smokers, with 35 of 45 loci displaying numerically larger effects in nonsmokers ($P_{\text{binomial}} = 1.2 \times 10^{-4}$).

![Forest plot for novel and GxSMK loci stratified by smoking status.](Image1)

We calculated the variance explained by subsets of SNPs selected on 15 significance thresholds for Approach 1 from $P_{\text{SNPadjSMK}} = 1 \times 10^{-8}$ to $P_{\text{SNPadjSMK}} = 0.1$ (Supplementary Table 9, Fig. 4). Differences in variance explained between smokers and nonsmokers were significant ($P_{\text{diff}} < 0.003 = 0.05/15$, Bonferroni-corrected for 15 thresholds) for BMI at each threshold, with more variance explained in smokers. For WCadjBMI, the difference was significant for SNP sets beginning with $P_{\text{SNPadjSMK}} \geq 3.16 \times 10^{-4}$, and for WHRadjBMI at $P_{\text{SNPadjSMK}} \geq 1 \times 10^{-6}$. In contrast to BMI, SNPs from Approach 1 explained a greater proportion of the variance in nonsmokers for WHRadjBMI. Differences in variance explained were greatest for BMI (differences ranged from 1.8 to 21% for smokers) and lowest for WHRadjBMI (ranging from 0.3 to 8.8% for nonsmokers).

These results suggest that smoking may increase genetic susceptibility to overall adiposity, but attenuate genetic effects on body fat distribution. This contrast is concordant with phenotypic observations of higher overall adiposity and lower central adiposity in smokers$^4,5,7$. Additionally, smoking increases oxidative stress and general inflammation in the body$^9$ and may exacerbate weight gain$^{20}$. Many genes implicated in BMI are involved in appetite regulation and feeding behaviour$^{1}$. For waist traits, our results adjusted for BMI likely highlight distinct pathways through which smoking alters genetic susceptibility to body fat distribution. Overall, our results indicate that more loci remain to be discovered as more variance in the trait can be explained as we drop the threshold for significance.

**Functional or biological role of novel loci.** We conducted thorough searches of the literature and publicly available bioinformatics databases to understand the functional role of all genes within 500 kb of our lead SNPs. We systematically explored the potential role of our novel loci in affecting gene expression both with and without accounting for the influence of smoking behaviour (Methods section, Supplementary Note 3, Supplementary Tables 10–12).

We found the majority of novel loci are near strong candidate genes with biological functions similar to previously identified adiposity-related loci, including regulation of body fat/weight, angiogenesis/adipogenesis, glucose and lipid homeostasis, general growth and development. (Supplementary Notes 1 and 3).

We identified rs17396340 for WCadjBMI (Approaches 1 and 2), an intronic variant in the KIF1B gene. This variant is associated with expression of KIF1B in whole blood with and without accounting for SMK (GTEx and Supplementary Tables 10 and 12) and is highly expressed in the brain$^{21}$. Knockout and mutant forms of KIF1B in mice resulted in multiple brain abnormalities, including hippocampus morphology$^{22}$, a region involved in (food) memory and cognition$^{23}$. Variant rs17396340 is associated with expression levels of ARSA in LCL tissue. Human adipocytes express functional ARSA, which turns dopamine sulfate into active dopamine. Dopamine regulates appetite through leptin.
and adiponectin levels, suggesting a role for ARSA in regulating appetite.\textsuperscript{24}

Expression of CD47 (CD47 molecule), near rs670752 for WHRadjBMI (Approach 1, women-only), is significantly decreased in obese individuals and negatively correlated with BMI, WC and Hip circumference.\textsuperscript{25} Conversely, in mouse models, CD47-deficient mice show decreased weight gain on high-fat diets, increased energy expenditure, improved glucose profile and decreased inflammation.\textsuperscript{26}

Several novel loci harbour genes involved in unique biological functions and pathways including addictive behaviours and response to oxidative stress. These potential candidate genes near our association signals are highly expressed in relevant tissues for regulation of adiposity and smoking behaviour (for example, brain, adipose tissue, liver, lung and muscle; Supplementary Note 2, Supplementary Table 10).

The $\text{CHRNA5-CHRNA3-CHRNB4}$ cluster is involved in the eNOS signalling pathway (Ingenuity KnowledgeBase, http://www.ingenuity.com) that is key for neutralizing reactive oxygen species introduced by tobacco smoke and obesity.\textsuperscript{27} Disruption of this pathway has been associated with dysregulation of adiponectin in adipocytes of obese mice, implicating this pathway in downstream effects on weight regulation.\textsuperscript{27,28} This finding is especially important due to the compounded stress adiposity places on the body as it increases chronic oxidative stress itself.\textsuperscript{28}

INPP4B has been implicated in the regulation of the PI3K/Akt signalling pathway\textsuperscript{29} that is important for cellular growth and proliferation, but also eNOS signalling, carbohydrate metabolism, and angiogenesis.\textsuperscript{30}

GRIN2A, near rs4141488, controls long-term memory and learning through regulation and efficiency of synaptic transmission\textsuperscript{31} and has been associated with heroin addiction.\textsuperscript{32} Nicotine increases the expression of GRIN2A in the prefrontal cortex in murine models.\textsuperscript{33} There are no established relationships between GRIN2A and obesity-related phenotypes in the literature, yet memantine and ketamine, pharmacological antagonists of

---

**Figure 3 | Power comparison across Approaches.** Shown is the power to identify adjusted (Approach 1, dashed black lines), joint (Approach 2, dotted green lines) and interaction (Approach 3 and 4, solid magenta and orange lines) effects for various combinations of SMK- and NonSMK-specific effects and assuming 50,000 smokers and 180,000 nonsmokers. For (a,c,e), the effect in smokers was fixed at a small ($R^2_{\text{SMK}} = 0.01\%$, similar to the realistic NUDT3 effect on BMI), medium ($R^2_{\text{SMK}} = 0.07\%$, similar to the realistic BDNF effect on BMI) or large ($R^2_{\text{SMK}} = 0.34\%$, similar to the realistic FTO effect on BMI) genetic effect, respectively, and varied in nonsmokers. For (b,d,f), the effect in nonsmokers was fixed to the small, medium and large BMI effects, respectively, and varied in smokers.
GRIN2A activity\textsuperscript{34,35}, are implicated in treatment for obesity-associated disorders, including binge-eating disorders and morbid obesity (ClinicalTrials.gov identifiers: NCT00330655, NCT02334059, NCT01997515, NCT01724983). Memantine is under clinical investigation for treatment of nicotine dependence (ClinicalTrials.gov identifiers: NCT01535040, NCT00136786 and NCT00136747). While our lead SNP is not within a characterized gene, rs4141488 and variants in high LD ($r^2 \approx 0.7$) are within active enhancer regions for several tissues, including liver, fetal leg muscle, smooth stomach and intestinal muscle, cortex and several embryonic and pluripotent cell types (Supplementary Note 2), and therefore may represent an important regulatory region for nearby genes like GRIN2A.

In secondary meta-analysis of European women-only, we identified a significant GxSMK interaction for rs6076699 on WCadjBMI (Table 4, Supplementary Data 4, Supplementary Fig. 6). This SNP is 100 kb upstream of PRNP (prion protein), a signalling transducer involved in multiple biological processes related to the nervous system, immune system, and other cellular functions (Supplementary Note 2)\textsuperscript{36}. Alternate forms of the oligomers may form in response to oxidative stress caused by copper exposure\textsuperscript{37}. Copper is present in cigarette smoke and...
Influence of novel loci on related traits

In a look-up in existing GWAS of smoking behaviours (Ever/Never, Current/Not-Current, Smoking Quantity (SQ)) Supplementary Data 8, eight of our 26 SNPs were nominally associated with at least one smoking trait. After multiple test correction ($P_{\text{Regression}} < 0.005/26 = 0.0019$), only one SNP remains significant: rs12902602, identified for Approaches 2 (SNPjoint) and 3 (SNPint) for BMI, showed association with SQ ($P = 1.45 \times 10^{-9}$).

We conducted a search in the NHGRI-EBI GWAS Catalog to determine if any of our newly identified loci are in high LD with variants associated with related cardiometabolic and behavioural traits or diseases. Of the seven novel BMI SNPs, only rs12902602 was in high LD ($r^2 > 0.7$) with SNPs previously associated with smoking-related traits (for example, nicotine dependence), lung cancer, and cardiovascular diseases (for example, coronary heart disease; Supplementary Table 13). Of the 12 novel WCadjBMI SNPs, 5 were in high LD with previously reported GWAS variants for mean platelet volume, height, infant length, and melanoma. Of the six novel WHRadBMI SNPs, three were near previously associated variants, including cardiometabolic traits (for example, LDL cholesterol, triglycerides and measures of renal function).

Given high phenotypic correlation between WC and WHR with height, and established shared genetic associations that overlap our adiposity traits and height, we expect cross-trait associations between our novel loci and height. Therefore, we conducted a look-up of all of our novel SNPs to identify overlapping association signals (Supplementary Data 8). No novel BMI loci were significantly associated with height ($P_{\text{Regression}} < 0.002(0.05/24)$ SNPs). However, there are additional variants that may be associated with height, but not previously reported in GWAS examining height, including two for WHRadBMI near EYA4 and TRIB1, and two for WCadjBMI near KIF1B and HDLBP ($P_{\text{Regression}} < 0.002$).

Finally, as smoking has a negative (weight decreasing) effect on BMI, it is likely that smoking-associated genetic variants have an effect on BMI in current smokers. Therefore, we expected that smoking-associated SNPs exhibit some interaction with smoking on BMI. We looked up published smoking behaviour SNPs in our novel loci, in our own data, two variants reached nominal significance ($P_{\text{SNPint}} < 0.05$) for GxSMK interaction BMI (Supplementary Data 14), but only one reached Bonferroni-corrected significance ($P < 0.005$). No smoking-associated SNPs exhibited GxSMK interaction. Therefore, we did not see a strong enrichment for low interaction $P$ values among previously identified smoking loci.

Validation of novel loci. We pursued validation of our novel and interaction SNPs in an independent study sample of up to 119,644 European adults from the UK Biobank study (Tables 1–4, Supplementary Table 15, Supplementary Fig. 9). We found consistent directions of effects in smoking strata (for Approaches 2 and 3) and in SNPadjSMK results (Approach 1) for each locus examined (Supplementary Fig. 13). For BMI, three SNPs were not GWS ($P_{\text{SNPadjSMK}} < 5 \times 10^{-8}$) following meta-analysis with our GIANT results:

Table 1: Summary of association results for novel loci reaching genome-wide significance in Approach (App) 1 ($P_{\text{SNPjoint}} < 5 \times 10^{-8}$) or Approach 2 ($P_{\text{SNPint}} < 5 \times 10^{-8}$) for our primary meta-analysis in combined ancestries and combined sexes.

<table>
<thead>
<tr>
<th>App</th>
<th>Marker</th>
<th>Chr-Pos (kb/10)</th>
<th>Nearest Gene</th>
<th>N</th>
<th>EAF</th>
<th>E/O</th>
<th>Smokers</th>
<th>Non-smokers</th>
<th>Main and interaction effects</th>
<th>GIANT = UKBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMII</td>
<td>rs10929925</td>
<td>2:6355557</td>
<td>SOX1T</td>
<td>225,067</td>
<td>0.55</td>
<td>C/A</td>
<td>0.019</td>
<td>7.80E-03</td>
<td>0.02</td>
<td>8.40E-08</td>
</tr>
<tr>
<td>WCadjBMI</td>
<td>rs6749480</td>
<td>3:8443102</td>
<td>SMRM2</td>
<td>186,968</td>
<td>0.85</td>
<td>A/G</td>
<td>0.025</td>
<td>2.30E-02</td>
<td>0.027</td>
<td>3.90E-06</td>
</tr>
<tr>
<td>WCadjBMI</td>
<td>rs13069244</td>
<td>3:18044172</td>
<td>CDCC39</td>
<td>233,776</td>
<td>0.08</td>
<td>A/G</td>
<td>0.016</td>
<td>1.80E-05</td>
<td>0.021</td>
<td>6.60E-05</td>
</tr>
<tr>
<td>WHRadBMI</td>
<td>rs1049281</td>
<td>6:31266567</td>
<td>HLA-C</td>
<td>149,205</td>
<td>0.66</td>
<td>C/T</td>
<td>0.022</td>
<td>1.30E-02</td>
<td>0.027</td>
<td>2.00E-08</td>
</tr>
</tbody>
</table>

Adj, adjusted for smoking; app, approach; int, interaction; chr, chromosome; EAF, effect allele frequency; E/O, effect/other; Pos, position (bp). Significant $P$-values that reach genome-wide significance ($P < 5 \times 10^{-8}$) threshold are in bold.
This phenomenon is of particular concern when the harmonization of smoking exposure.

However, WC and WHR may not behave in the same manner definition allowed us to maximize sample size, as many remained GWS for SNPadjSMK, but was only nominally (Approach 3); and rs6012558 near RAI14 rs1545348 near ART14

Another limitation may be potential bias in our effect estimates potential collider bias and postulate true gain in power through are associated with any smoking behaviour trait and that exhibit there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by (for example, cigarette, pipe) for two reasons. First, focusing on smoking status without consideration of type of smoking harmonization of smoking status. We chose to stratify on current smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent estimates should be interpreted with caution. Additionally, we adjusted for smoking; W, men only.

To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. To assess how much additional information is provided by accounting for environmental exposures in GWAS. A possible limitation of our study may be the definition and harmonization of smoking status. We chose to stratify on current smoking status without consideration of type of smoking (for example, cigarette, pipe) for two reasons. First, focusing on smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study’s European descent participants (−0.13, 0.08 and 0.12, respectively) and in the Framingham Heart Study (−0.05, 0.08 and 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that were significant associations with smoking (Approach 3) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes.
Table 4 | Summary of association results for loci showing significance for interaction with smoking in Approach 3 (SNPint) and/or Approach 4 (SNPscreen) in our secondary meta-analyses not identified in primary meta-analyses.

<table>
<thead>
<tr>
<th>Approach/Strata</th>
<th>Marker</th>
<th>Chr/Pos (hg19)</th>
<th>Nearest Gene</th>
<th>N</th>
<th>EAF</th>
<th>SnpType</th>
<th>Smokers</th>
<th>Non-smokers</th>
<th>Main and interaction effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM1</td>
<td>6:AM</td>
<td>rs1809420*</td>
<td>1579056769</td>
<td>57,081</td>
<td>0.59</td>
<td>T/C</td>
<td>0.074</td>
<td>9.8E-04</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WCadjBMI</td>
<td>SEV</td>
<td>rs6076699</td>
<td>20,4566688</td>
<td>76,930</td>
<td>0.97</td>
<td>A/G</td>
<td>0.169</td>
<td>1.4E-06</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHRadjBMI</td>
<td>4:EM</td>
<td>rs30000*</td>
<td>5.55063533</td>
<td>71,424</td>
<td>0.27</td>
<td>G/A</td>
<td>0.002</td>
<td>7.8E-01</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>6:AM</td>
<td>rs598933*</td>
<td>80,852</td>
<td>0.27</td>
<td>A/G</td>
<td>0.004</td>
<td>5.0E-01</td>
<td>0.034</td>
<td>4.0E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>6:AM</td>
<td>rs20744*</td>
<td>12,54438101</td>
<td>70,868</td>
<td>0.37</td>
<td>A/C</td>
<td>0.003</td>
<td>6.0E-01</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>6:AM</td>
<td>rs20743*</td>
<td>12,5443920</td>
<td>71,136</td>
<td>0.36</td>
<td>A/G</td>
<td>0.006</td>
<td>6.2E-01</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>6:AM</td>
<td>rs26068504*</td>
<td>19,38893859</td>
<td>80,087</td>
<td>0.37</td>
<td>A/G</td>
<td>0.006</td>
<td>2.6E-01</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>6:AM</td>
<td>rs26068504*</td>
<td>19,38893859</td>
<td>80,087</td>
<td>0.37</td>
<td>A/G</td>
<td>0.006</td>
<td>2.6E-01</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A, all ancestries; Adj, adjusted for smoking; app, approach; C, combined sexes; Chr, chromosome; E, European-only; EAF, effect allele frequency; E/O, effects/other; int, interaction; M, men only; Pos, position (bp); W, women only.

All estimates are from the stratum specified in the Approach/Strata column. The I² between the ADAMTS7 (rs1809420) and CHRNA3 variant (rs1290362) in Table 3 is 0.72 (HapMap 2, CEU).

Additionally, the PRNP variant (rs6076699) is the same as the variant that came up from Approach 2 (Table 2). *Known locus.

Significant P-values after multiple test correction are italicized.

Significant P-values that reach genome-wide significance (P < 10⁻⁵) threshold are in bold.

1.53 to 2.11%, PMGRdiff = 4.3 × 10⁻⁵, by 27% for WCadjBMI (from 2.59 to 3.29%, PMGRdiff = 3.9 × 10⁻⁵) and by 168% for WHRadjBMI (from 0.82 to 2.20%, PMGRdiff = 3.2 × 10⁻¹¹). Therefore, despite potential limitations, much is gained by accounting for environmental exposures in GWAS studies.

Discussion

To better understand the effects of smoking on genetic susceptibility to obesity, we conducted meta-analyses to uncover genetic variants that may be masked when the environmental influence of smoking is not considered, and to discover genetic loci that interact with smoking on adiposity-related traits. We identified 161 loci in total, including 23 novel loci (6 for BMI, 11 for WCadjBMI, and 6 for WHRadjBMI). While many of our newly identified loci support the hypothesis that smoking may influence weight fluctuations through appetite regulation, these novel loci also have highlighted new biological processes and pathways implicated in the pathogenesis of obesity.

Importantly, we identified nine loci with convincing evidence of GxSMK interaction on obesity-related traits. We were able to replicate the previous GxSMK interaction with BMI within the CHRNA5-CHRNA3-CHRN4B gene cluster. One novel BMI-associated locus near INPP4B and two novel WCadjBMI-associated loci near GRIN2A and PRNP displayed significant GxSMK interaction. We were also able to identify significant GxSMK interaction for one known BMI-associated locus near ADAMTS7 and for five known WHRadjBMI-associated loci near LPLAL1, RSPO3, MAP3K1, HOXC4-HOXC6 and JUND. The majority of these loci harbor strong candidate genes for adiposity with a possible role for the modulation of effects through tobacco use.

We identified 18 new loci in Approach 1 (P_withinSNP) by adjusting for current smoking status. Our analyses did not allow us to determine whether these discoveries are due to different subsets of subjects included in the analyses compared to previous studies or due only to adjusting for current smoking. Adjustment for current smoking in our analyses, however, did reveal novel associations. Specifically after accounting for smoking in our analyses, all novel BMI loci exhibit P-values that are at least one order of magnitude lower than in previous GIANT investigations, despite smaller samples in the current analysis.

While sample sizes for both WCadjBMI and WHRadjBMI are comparable with previous GIANT investigations, our P values for variants identified in Approach 1 are at least two orders of magnitude lower than previous findings. Thus, adjustment for smoking may have indeed revealed new loci. Further, loci identified in Approach 2, including nine novel loci, suggest that accounting for interaction improves our ability to detect these loci even in the presence of only modest evidence of GxSMK interaction.

There are several challenges in validating genetic associations that account for environmental exposure. In addition to exposure harmonization and potential bias due to adjustment for smoking exposure, differences in trait distribution, environmental exposure frequency, ancestry-specific LD patterns and allele frequency across studies may lead to difficulties in replication, especially for gene-by-environment studies. Furthermore, the ‘winner’s curse’ (inflated discovery effects estimates) requires larger sample sizes for adequate power in replication studies. Despite these challenges, we were able to detect consistent direction of effect in an independent sample for all novel loci. Some results that did not remain GWS in the GIANT + UKBB meta-analysis had results that were just under the threshold for significance, suggesting that a larger sample may be needed to confirm these results, and thus the associations near INPP4B, GRIN2A, RA114, PRNP and JUND should be interpreted with caution.

While we found that effects were not significantly enriched in smokers for BMI, there is a greater proportion of variance in BMI explained by variants that are significant for Approach 1 (SNPadjSMK), which may be expected given that there are a greater number of variants with higher effect estimates in smokers. For WCadjBMI, there was no enrichment for stronger effects in one stratum compared to the other for our significant loci; however, there was a greater proportion of explained variance in WCadjBMI for loci identified in Approach 1 (SNPadjSMK) in nonsmokers. For WHRadjBMI, there were significantly more loci that exhibit greater effects in nonsmokers, and this pattern was mirrored in the variance explained analysis. The large difference between effects in smokers and nonsmokers likely explains the sub-GWS levels of our loci in previous GIANT investigations. For example, the T allele of rs7697556, 81kb from the ADAMTS3 gene, was associated with increased WCadjBMI and exhibits a sixfold greater effect in nonsmokers compared to smokers, although the interaction effect was only nominal; in previous GWAS this variant was nearly GWS. These differences in effect estimates between smokers and nonsmokers may help explain inconsistent findings in previous analyses that show central adiposity increases with increased smoking, but is associated with decreased weight and BMI5,9,10.

Our results support previous findings that implicate genes involved in transcription and gene expression, appetite regulation, macronutrient metabolism, and glucose homeostasis. Several of our novel loci have candidate genes within 500 kb of our tag
variants that are highly expressed and/or active in brain tissue (BBX, KIF1B, SOX11 and EPAH3) and, like other obesity-associated genes, may be involved in previously-identified pathways linked to neuronal regulation of appetite (KIF1B, GRIN2A and SLC23A2), adipogenesis (ANGPTL3 and TNF) and glucose, lipid and energy homeostasis (CD47, STK25, STK19, RAGE, AIF1, LYPLAL1, HDLBP, ANGPTL3, Dock7, KIF1B, PREX1 and RPS12).

Many of our newly identified loci highlight novel biological functions and pathways where dysregulation may lead to increased susceptibility to obesity, including response to oxidative stress, addictive behaviour, and newly identified regulatory functions. There is a growing body of evidence that supports the notion that exposure to oxidative stress leads to increased adiposity, risk of obesity, and poor cardiometabolic outcomes27,56. Our results for BMI and WCadjBMI, specifically STK19 and TNF pathways linked to neuronal regulation of appetite (sex in the model. Phenotype residuals were obtained from the adjustment models appropriate) for study site and principal components to account for ancestry.

10 Study design overview

We applied four approaches to identify genetic loci that influence adiposity traits by accounting for current tobacco smoking status (Fig. 1). We defined smokers as those who responded that they were currently smoking; not current smokers were those that responded ‘no’ to currently smoking. We evaluated three traits: body mass index (BMI), waist circumference adjusted for BMI (WCadjBMI), and highlight novel biological processes and regulatory functions for genes implicated in increased obesity risk. Eighteen of these remained significant in our validation with the UK Biobank sample. We confirmed most established loci in our analyses after adjustment for smoking status in smaller samples than were needed in previous discovery analyses. A typical approach in large-scale GWAS meta-analyses is not to adjust for covariates such as current smoking; our findings highlight the importance of accounting for environmental exposures in genetic analyses.

Methods

Study design overview. We applied four approaches to identify genetic loci that influence adiposity traits by accounting for current tobacco smoking status (Fig. 1). We defined smokers as those who responded that they were currently smoking; not current smokers were those that responded ‘no’ to currently smoking. We evaluated three traits: body mass index (BMI), waist circumference adjusted for BMI (WCadjBMI), and waist-to-hip ratio adjusted for BMI (WHRadjBMI). Our first two meta-analytical approaches were aimed at determining whether there are novel genetic variants that affect adiposity traits by adjusting for SMK (SNPadjSMK), or by jointly accounting for SMK and for interaction with SMK (SNPjoint); while Approaches 3 and 4 aimed to determine whether there are genetic variants that affect adiposity traits through interaction with SMK (SNPint and SNPscreen) (Fig. 1). Our primary meta-analyses focused on results from all ancestries, sexes combined. Secondary meta-analyses were performed using the European-descent populations only, as well as stratified by sex (men-only and women-only) in all ancestries and in European-descent study populations.

Cohort descriptions and sample sizes. The GIANT consortium was formed by an international group of researchers interested in understanding the genetic architecture of anthropometric traits (Supplemental Tables 1–4 for study sample sizes and descriptive statistics). In total, we included up to 79 studies comprising up to 241,258 individuals for BMI (51,080 smokers, 190,178 non-smokers), 208,176 for WCadjBMI (43,226 smokers, 164,950 non-smokers), and 189,180 for WHRadjBMI (40,543 smokers, 148,637 non-smokers) with HapMap II imputed genome-wide chip data (up to 2.8M SNPs in association analyses), and/or with genotyped MetaboChip data (~195K SNPs in association analyses). In instances where studies submitted both MetaboChip and GWAS data, these were for non-overlapping individuals. Each study’s Institutional Review Board has approved this research and all study participants have provided written informed consent.

Phenotype descriptions. Our study highlights three traits of interest: BMI, WCadjBMI and WHRadjBMI. Height and weight, used to calculate BMI (kg m−2), were measured in all studies; waist and hip circumferences were measured in the vast majority. For each sex, traits were adjusted using linear regression for age and age² (as well as for BMI for WCadjBMI and WHRadjBMI), and (when appropriate) for study site and principal components to account for ancestry. Family data were used in a linear mixed effects model to account for familial relationships and also conducted analyses for men and women combined including sex in the model. Phenotype residuals were obtained from the adjustment models and were inverse normally transformed subsequently to facilitate comparability across studies and with previously published analyses. The trait transformation was necessary for smokers and non-smokers for the SMK-stratified model and using all individuals for the SMK-adjusted model.

Defining smokers. The participating studies have varying levels of information on smoking, some with a simple binary variable and others with repeated, precise data. Since the effects of smoking cessation on adiposity appear to be immediate7,8,32, a binary smoking trait (current smoker versus not current smoker) is used for the analyses as most studies can readily derive this variable. We did not use a variable of ‘ever smoker vs. never’ as it increases heterogeneity across studies, thus adding noise; also this definition would make harmonization across studies difficult.

Genotype identification and imputation. Studies with GWAS array data or Metabochip array data contributed to the results. Each study applied study-specific standard exclusions for sample call rate, gender checks, sample heterogeneity and ethnic group outliers (Supplementary Table 2). For each studies (except those that employed directly typed MetaboChip genotypes), genome-wide chip data was imputed to the HapMap II reference data set.

Study level analyses. To obtain study-specific summary statistics used in subsequent meta-analyses, the following linear models (or linear mixed effects models for studies with families/related individuals) were run separately for men and women and controls for case-control studies using phenotype residuals from the models described above. Studies with family data also conducted analyses with these models for men and women combined after accounting for dependency among family members as a function of their kinship correlations. We assumed an additive genetic model. The analyses were run using various GWAS software Supplementary Table 2.

SMK-adjusted : TRAIT = β0 + β1 SNP + β2 SMK

SMK-stratified : TRAIT = β0 + β1 SNP (run in current smokers and non-smokers separately)

Quality control of study-specific summary statistics. The aggregated summary statistics were quality-controlled according to a standardized protocol57. These included checks for issues with trait transformations, allele frequencies and strand. Low quality SNPs in each study were excluded for the following criteria: (i) SNPs with low minor allele count (MAC < 5, MAG = MAC X N) and monomorphic SNPs, (ii) genotyped SNPs with low SNP call-rate (<95%) or low Hardy-Weinberg equilibrium P value (<10−6), (iii) imputed SNPs with low imputation quality (MACH-Rsq or OVAR < 0.3, or information score < 0.4 for SNPTEST/IMPUTE/ IMPUTE2, or < 0.8 for PLINK). To test for issues with relatedness or overlapping samples and to correct for potential population stratification, the study-specific summary statistics were used to test for relatedness using lambda factors (Supplementary Fig. 1). GC correction for GWAS data used all SNPs, but GC correction for MetaboChip data were restricted to chip QT interval SNPs only as the chip was enriched for associations with obesity-related traits. Any study-level GWAS file with a lambda > 1.5 was removed from further analyses. While we established this criterion, no study results were removed for this reason.

Meta-analyses. Meta-analyses used study-specific summary statistics for the phenotype associations for each of the above models. We used a fixed-effects inverse variance weighted method for the SNP main effect analyses. All meta-analyses were run in METAL58. As study results came in two separate batches (Stage 1 and Stage 2), meta-analyses from the two stages were further meta-analysed (Stage 1 + Stage 2). A second GC correction was applied to all SNPs when combining Stage 1 and Stage 2 meta-analyses in the final meta-analysis. First, Hapmap-imputed GWAS data were meta-analysed together, as were Metabochip studies. This step was followed by a combined GWAS + Metabochip meta-analysis. For primary analyses, we conducted meta-analyses across ancestries and sexes. For secondary meta-analyses, we conducted meta-analyses in European-descent studies alone, and sex-specific meta-analyses. There were two reasons for conducting secondary meta-analyses, First, both WCadjBMI and WHRadjBMI have been shown to display sex-specific genetic effects2,59,60. Second, by including populations from multiple ancestries in our primary meta-analyses, we may be inflated estimate heterogeneity due to differences in effect sizes, allele frequencies, and patterns of linkage disequilibrium across ancestries, potentially decreasing power to detect genetic effects. See Supplementary Fig. 1 for a summary of the primary meta-analysis study design. The obtained SMK-stratified summary statistics were later used to calculate summary SNPpoint and SNPint statistics using EasyStrata61. Briefly, this software implements a two-sample, large sample test of equal regression parameters between smokers and nonsmokers58 for SNPint and the two degree of freedom test of main and interaction effects for SNPpoint14.

Lead SNP selection. Before selecting a lead SNP for each locus, SNPs with high heterogeneity I2 ≥ 0.75 or a minimum sample size below 50% of the maximum
Approaches. Figure 1 outlines the four approaches that we used to identify novel SNPs. The left side of Fig. 1 focuses on the first hypothesis that examines the effect of SNPs on adiposity traits. Approach 1 considered a linear regression model that includes the SNP and SMK, thus adjusting for SMK (SNPadjSMK). Summary SNPadjSMK results were obtained from the SMK-adjusted meta-analysis. Approach 2 used summary SMK-stratified meta-analysis results to consider the joint hypothesis that this variant has an effect/interaction on outcomes as a 2 degree of freedom test (SNPjoint). For this approach, the null hypothesis was that there is no main and no interaction effect on the outcome. Thus, rejection of this hypothesis could be due to either a main effect or an interaction effect or both.

The second hypothesis of interest focuses on the second hypothesis, testing for interaction of a variant with SMK on adiposity traits as outcomes. Approach 3 used the SMK-stratified results to directly contrast the regression coefficients for a test of interaction (SNPint)19. Approach 4 used a screening strategy to evaluate interaction, whereby the SMK-adjusted main effect results (Approach 1) were screened using a significant main effect at the P = 0.05 level and then carried forward for a test of interaction, comparing the SMK-stratified specific regression coefficients in the second step (SNPscreen). In Approaches 1–3 variants significant at P < 5 × 10^{-8} were considered GWAS. In Approach 4 (SNPscreen) variants for which the P value of the test of interaction is less than 0.05 and the number of variants forwarded for analysis were considered significant for interaction. We performed analytical power computations to demonstrate the usefulness and characteristic of the two interaction Approaches.

Locuszoom plots. Regional association plots were generated for novel loci using the program Locuszoom (http://locuszoom.sph.umich.edu/) . For each plot, LD was calculated using a multiethnic sample of the 1000 Genomes Phase 1 reference panels22, including EUR, AFR, EAS and AMR. Previous SNP-trait associations highlighted within the plots include traits of interest (for example, cardiometabolic, smoking and cancer) and identified here are independent of the previously reported SNP-trait associations. The program Locuszoom (http://locuszoom.sph.umich.edu/) was used to identify any genes and nearby candidate genes (within 10 kilobase pairs). In addition, we performed in-depth literature searches on nearby candidate genes. Snipper v1.2 (http://csg.sph.umich.edu/boehnke/snipper/) was used to identify any genes and cis- or trans-eQTLs within 500 kb of our lead SNPs. All genes identified by Snipper were manually curated and examined for evidence of relationship with smoking and/or adiposity. To explore any potential regulatory or function role of the association regions, loci were also examined using several online bioinformatic tools/databases, including HaploReg v4.1 (ref. 63), UCSC Genome Browser (http://genome.ucsc.edu/), GTEx Portal (http://www.gtexportal.org), and RegulomeDB44.

eQTL analyses. We used two approaches to systematically explore the role of novel loci in regulating gene expression. First, to gain a general overview of the regulatory role of newly identified GWAS regions, we conducted an eQTL lookup using > 50 eQTL studies59, with specific citations for > 100 data sets included in the current query for blood cell related eQTL studies and relevant non-blood cell tissue (for example, brain tissue). These data sets were integrated from online sources including ScanDB, the Broad Institute GTEx Portal, and the Pritchard Lab (eqtl.ucucho.edu). Additional details on the methods, including study references can be found in Supplementary Note 3. Only significant cis-eQTLs in high LD with our novel lead SNPs (r2 > 0.9, calculated in the 500 KB either side of each lead SNP, and imputation was carried out using IMPUTE2 (ref. 70) with the 1000 Genomes Phase 3 data set as a reference panel.

Conditional analyses. To determine if multiple association signals were present within a single locus, we used GCTA20 to perform approximate joint conditional analyses on the SNPadjSMK and SMK-stratified data. The following criteria were used to select candidate loci for conditional analyses: nearby SNP (± 500 kb) with an R2 > 0.4 and an association P < 10^{-8} for any of our primary analyses. GCTA uses associations from our meta-analyses and LD estimates from reference data sets containing individual-level genotypic data to perform the conditional analyses. To calculate the LD structure, we used two U.S. cohorts, the Atherosclerosis Risk in Communities (ARIC) consisting of 15,798 Caucasians and 1580 individuals of African American descent, and the Framingham Heart Study (FramHS) consisting of 8,481 individuals of European ancestry, both studies imputed to HapMap r2. However, because our primary analyses were conducted in multiple ancestries, each study supplemented the genetic data using HapMap reference populations so that the final reference panel was composed of about 1–3% Asians (CHB + JPT) and 4–6% Africans (YRI for the FramHS) for the entire reference sample. We extracted each 1 MB region surrounding our candidate SNPs, performed joint approximate conditional analyses, and then repeated the steps for the appropriate Approach to identify additional association signals.

Many of the SNPs identified in the current analyses were nearby SNPs previously associated with related anthropometric and obesity traits (for example, height, visceral adipose tissue). For all lead SNPs near a SNP previously associated with these traits, GCTA was also used to perform approximate conditional analyses on the SNPadjSMK and SMK-stratified data in order to determine if the loci identified here are independent of the previously identified SNP-trait associations.

Power and type I error. In order to illustrate the validity of the approaches with regards to type I error, we conducted simulations. For two MA-F, we assumed standardized stratum-specific outcomes for 50,000 smokers and 180,000 nonsmokers and generated 10,000 simulated stratum-specific effect sizes under the stratum-specific null hypotheses of ‘no stratum-specific effects’. We applied the four approaches to the simulated stratum-specific association results and inferred type 1 error of each approach by visually examining QQ plots and by calculating type 1 error rates. The type 1 error rates shown reflect the proportion of nominally significant simulation results for the respective approach. Analytical power calculations to identify effects for various combinations of SMK- and NonSMK-specific effects in the Approach 3 and 4 were conducted for 20 SNPs at each 100 smokers and 180,000 nonsmokers. We first assumed three different fixed effect estimates in smokers that were small (RSMK = 0.01%, similar to the realistic NUDT3 effect on BMI), medium (RSMK = 0.07%, similar to the realistic BDNF effect on BMI) or large (RSMK = 0.34%, similar to the realistic FTO effect on BMI) genetic effects, and varied the effect in nonsmokers. Second, we assumed fixed (small, medium and large) effects in nonsmokers and varied the effect in smokers.

Biological summaries. To identify genes that may be implicated in the association between our lead SNPs (Tables 1–3) and BMI, WHRadjBMI and WCDadjBMI, and to shed light on the complex relationship between genetic variants, SMK and adiposity, we performed in-depth literature searches on nearby candidate genes. Snipper v1.2 (http://csg.sph.umich.edu/boehnke/sniper/) was used to identify any genes and cis- or trans-eQTLs within 500 kb of our lead SNPs. All genes identified by Snipper were manually curated and examined for evidence of relationship with smoking and/or adiposity. To explore any potential regulatory or function role of the association regions, loci were also examined using several online bioinformatic tools/databases, including HaploReg v4.1 (ref. 63), UCSC Genome Browser (http://genome.ucsc.edu/), GTEx Portal (http://www.gtexportal.org), and RegulomeDB44.

Variance-explained estimates. We estimated the phenotypic variance in smokers and nonsmokers explained by the association signals. For each associated region, we selected subsets of SNPs within 500 kb of our lead SNPs and based on varying P value thresholds (ranging from 1 × 10^{-8} to 0.1) from Approach 1 (SNPadjSMK model). First, each subset of SNPs was clumped into independent regions to identify the lead SNP for each region. The variance explained by each subset of SNPs in the SMK and non-SMK strata was estimated by summing the variance explained by the individual lead SNPs. Then, we tested for the significance of the differences across the two strata assuming that the weighted sum of chi-squared distributed variables tend to a Gaussian distribution ensured by Lyapunov’s central limit theorem67.68.

Smoking behaviour lookups. In order to determine if any of the loci identified in the current study are associated with smoking behaviour, we conducted a look-up of smoking behaviour lookups from novel lead SNPs from Approach 3 in eQTL and eQTLXSMK interaction while accounting for current smoking status, using the following four approaches similar to those used in our primary analyses of our traits: (1) eQTL adjusted for SMK, (2) eQTL stratified by SMK, (3) eQTL > SMK interaction and (4) joint main + eQTLXSMK interaction. Significance level was evaluated by FDR < 5% per eQTL analysis and across all loci identified for that model in the primary meta-analysis. Additional details can be found in Supplementary Note 3.

N for each strata (for example, N > max(N women smokers)/2) were excluded. Lead SNPs that met significance criteria were selected based on distance (± 500 kb), and we defined the SNP with the lowest P value as the top SNP for a locus. SNPs that reached genome-wide significance (GWAS), but had no other SNPs within 500 kb with a P < 1 × 10^{-5} (loosely SNPs), were excluded from the SNP selection process. Two variants were excluded from Approach 2 based on this criterion, rs2149656 for WCDadjBMI and rs3266276 for WHRadjBMI.
Main effects lookup in previous GIANT investigations. To better understand why our novel variants remained undiscovered in previous investigations that did not take SMK into account, we also conducted a lookup of our novel variants in published GWAS results examining genetic main effects on BMI, WC, WCadjBMI, WHR, WHRadjBMI, and height.1,2,5,11

GWAS catalog lookups. To further investigate the identified genetic variants in this study and to gain additional insight into their functionality and possible effects on related cardiometabolic traits, we searched for previous SNP-trait associations nearby our lead SNPs. PLINK was used to find all SNPs within 500 kb of any of our lead SNPs and calculate r^2 values using a combined ancestry (AMR, AFR, EUR, ASN) 1000 Genomes Phase 1 reference panel62 to allow for LD calculation for SNPs on the Illumina Metabochip and to best estimate LD in our multiethnic GWAS. All SNPs within the specified regions were compared with the NHGRI-EBI (National Human Genome Research Institute, European Bioinformatics Institute) GWAS Catalog v1.0 (www.ebi.ac.uk/gwas/).62 For overlap, and distances between the two SNPs were calculated using STATATA v14, for the chromosome and base pair positions based on human genome reference build 19. All previous associations within 500 kb and with an R^2 > 0.5 with our lead SNP were retained for further interrogation.

Genetic risk score calculation. We calculated several unweighted genetic risk scores (GRSs) for each individual in the population-based KORA-S3 and KORA-S4 studies (total N = 3,457). We compared GRSs limited to previously known lead SNPs (see Supplementary Tables 1–4 for lists of novel lead SNPs). Risk scores were tested for association with the obesity trait using the following linear regression models: The unadjusted GRS model (TRAIT = β_0 + β_1GRS), the adjusted GRS model (TRAIT = β_0 + β_1GRS + β_2SMK) and the GRSxSMK interaction model (TRAIT = β_0 + β_1GRS + β_2SMK + β_3GRSxSMK). Additionally, we used an F statistic to test whether the residual sum of squares (RSS) for the full model including GRSxSMK interaction was significantly different from the reduced model.

Data availability. Summary statistics of all analyses are available at https://www.broadinstitute.org/collaboration/giant/.

References


& L. Adrienne Cupples4,10,**

1 Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. 2 Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany. 3 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine; St. Louis, Missouri 63108, USA. 4 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. 5 Population Health Research Institute, St. George’s, University of London, London SW17 ORE, UK. 6 TransMed Systems, Inc., Cupertino, California 95014, USA. 7 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. 8 The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 9 Steno Diabetes Center, Gentofte, Denmark. 10 NHLBI Framingham Heart Study, Framingham, Massachusetts 01702, USA. 11 Division of Preventive Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, USA. 12 Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. 13 Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, The Framingham Heart Study, Framingham, Massachusetts, USA. 14 Institute of Social and Preventive Medicine, University of Lausanne, Lausanne, Switzerland. 15 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. 16 Swiss institute of Bioinformatics, 1015 Lausanne, Switzerland. 17 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 18 Department of Biobank Research, Umeå University, Umeå, Sweden. 19 Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, SE-205 02 Malmö, Sweden. 20 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA. 21 Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA. 22 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, USA. 23 The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, USA. 24 Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3015GE, The Netherlands. 25 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. 26 University of Lille, CNRS, Institut Pasteur de Lille, UMR 8199 - EGIS, Lille, France. 27 Internal Medicine - Nephrology, University of Michigan, Ann Arbor, Michigan, USA. 28 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, 48109, USA. 29 Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley 6009, Australia. 30 Department of Health Sciences, University of Milan, Via A. Di Rudini 8, 20142, Milano, Italy. 31 Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. 32 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. 33 Department of Psychiatry, Dokuz Eylul University, Izmir, Turkey. 34 Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden. 35 Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. 36 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. 37 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. 38 Epidemiology Department, Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore. 39 MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland. 40 William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. 41 Department of Health, National Institute for Health and Welfare, Helsinki FI-00271, Finland. 42 Vh Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. 43 Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. 44 IER, University of Essex, Colchester CO435Q, UK. 45 Department of Epidemiology and Public Health, UCL, London, WC1E 6BT, UK. 46 Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. 47 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge CB2 0QQ, UK. 48 Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland. 49 Department of Clinical Medicine, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland.
Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, The Netherlands. 242 Center for Medical Systems Biology, Leiden, The Netherlands. 243 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg 85764, Germany. 244 Department of Statistics, University of Oxford, Oxford, UK. 245 Mount Sinai School of Medicine, New York 10029, USA. 246 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. 247 Department of Preventive Medicine, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. * These authors contributed equally to this work. ** These authors jointly supervised this work.

† Deceased.