Predicting mechanisms of action at genetic loci associated with discordant effects on type 2 diabetes and abdominal fat accumulation

Yonathan Tamrat Aberra[1,2], Lijiang Ma[3], Johan L.M. Björkegren[3,4], Mete Civelek.[1,2]

1. Department of Biomedical Engineering, University of Virginia. Charlottesville, Virginia.
2. Center for Public Health Genomics, University of Virginia. Charlottesville, Virginia.
3. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai. New York, New York.
4. Department of Medicine, Karolinska Institutet, Huddinge, Stockholm, Sweden

Corresponding author: Mete Civelek, PhD
Center for Public Health Genomics
PO Box 800717
Old Med School 3836
Charlottesville, VA 22908-0717
Phone Number: 434-243-1669
Fax Number: 434-982-1815
E-mail: mete@virginia.edu

ABSTRACT

Obesity is a major risk factor for cardiovascular disease, stroke and type 2 diabetes (T2D). One risk factor that couples the adipose tissue accumulation characteristic of obesity to increased T2D risk is abdominal fat accumulation, measured by calculating the ratio of waist-to-hip circumference adjusted for the body-mass index (WHRadjBMI). Body fat distribution is estimated to have 40-80% heritability, which suggests a strong genetic component. Genetic loci associated with WHRadjBMI identified in genome-wide association studies (GWAS) are predicted to act through adipose tissues, but the molecular mechanisms regulating the accumulation of adipose tissue in various fat depots and their consequences for T2D risk are poorly understood. By integrating multi-omic data with colocalization analysis and Tissue of Action score analysis, we predicted colocalization of 79 genetic loci associated with both T2D adjusted for BMI (T2DadjBMI) and WHRadjBMI. Six of these loci were significantly associated with discordant effects on T2DadjBMI and WHRadjBMI, meaning that the alleles of the lead genetic variants in these loci that were protective for T2D were associated with increased abdominal obesity. Genes in these loci could explain the metabolically healthy obesity and lean insulin resistance (MHO/LIR) phenotype; therefore, we predicted functional genes and tissues of action at the discordant loci, and investigated how expression of these genes was linked to metabolic phenotypes. Through this rigorous consensus analysis, we provide evidence for the hypothesis that the genes encoding Thyroid Adenoma Associated Protein Antisense RNA (*THADA-AS*), Gypsy Retrotransposon Integrase 1 (*GIN1*), and Peptidylglycine Alpha-Amidating Monooxygenase (*PAM*) are involved in molecular mechanisms that uncouple abdominal fat accumulation from risk for T2D.

INTRODUCTION

Obesity has reached high levels of prevalence, and is known to be a major risk factor for cardiometabolic complications such as coronary artery disease, stroke, and type 2 diabetes (T2D)[1]. Recently developed therapeutics such as SGLT2 inhibitors and GLP-1 agonists, in combination with lifestyle interventions, have been reported to aid patients lose 2-6 kg of body weight and reduce insulin resistance[2]. Despite the overall increased risk for T2D and therapeutic potential of weight loss, there exists significant risk stratification within obese patient population, and lean individuals are known to develop T2D[3]. Recently, known phenomena such as the metabolically healthy obesity and lean insulin resistance (MHO/LIR) phenotypes have garnered increased interest for their uncoupling of obesity with risk for cardiometabolic complications[3,4]. Several mechanisms have been proposed to mediate the MHO/LIR phenomena, but the precise genetic molecular, cellular, and physiological mechanisms underlying this phenomena remain unknown. Characterizing these mechanisms is key in efforts to develop novel and personalized therapeutic interventions. One mechanism thought to mediate this risk is body fat distribution. Excessive abdominal fat accumulation, as quantified by the ratio of waist-to-hip circumference adjusted for body mass index (WHRadjBMI), is associated with a stronger increase of T2D risk than overall obesity as approximated by the BMI[5]. Additionally, WHRadjBMI is more genetically heritable than BMI, with heritability estimates ranging from 40-80%[6].

Because genetic heritability is integral to the etiology of these complex traits, they have been the subject of many genome-wide association studies (GWAS). The most recent T2DadjBMI and WHRadjBMI GWAS each included more than a million participants, giving them the power to discover novel associations at hundreds of genetic loci[6,7]. Despite the investigational and statistical power GWAS provide, individual genetic association studies are unable to predict the mechanisms of action associating a locus with a trait of interest. A number of methods have been developed to follow up on GWAS loci by incorporating multi-modal data to predict mechanisms of action underlying each locus. One such group of methods, colocalization analysis (COLOC), uses summary-level genetic association data for two traits to test the hypothesis that a genetic locus contains a single variant responsible for both traits. COLOC has been used to predict mechanisms of action connecting molecular traits to disease and to discover shared pathogenesis for complex diseases[8,9]. Another recent paradigm, Tissue of ACTion scores for Investigating Complex trait-Associated Loci (TACTICAL)[10], was first developed to investigate T2D and predict tissues of action (TOA) whose function underlies associations at GWAS loci. TACTICAL incorporates genetic expression data, epigenetic annotations, and GWAS associations to inform TOA predictions. Using the consensus of these methods to inform high-confidence data-driven predictions can prioritize hypotheses for functional annotation and experimental validation.

To advance the understanding of mechanisms linking body fat distribution to T2D risk independent of overall obesity, we leveraged these prediction methods to identify the mechanisms of action at genetic loci associated with T2D and WHR, both adjusted for the BMI. Using the most recent GWAS results and high-resolution epigenetic annotations of metabolically active tissues, we predicted colocalization of 79 genetic loci associated with both T2DadjBMI and WHRadjBMI. Of these colocalized loci, 6 were associated with discordant effects protecting against T2D but increasing abdominal fat accumulation. We integrated epigenomic annotations, genetic variant annotations, expression and splicing quantitative trait loci (QTL) data, and gene-tissue expression data to predict the functional genes within the discordant loci and their tissues of action. These heretofore uncharacterized genetic association signals with discordant metabolic effects can illuminate mechanisms that lead to MHO/LIR, and inform personalized patient risk stratification and therapeutic strategies.

METHODS

GWAS-GWAS Colocalization Analysis. GWAS results for T2DadjBMI and WHRadjBMI were obtained from Mahajan et al[7] and Pulit et al[6]. The set of single nucleotide polymorphisms (SNPs) within 200kb of a genome-wide significant SNP in either GWAS were included in the colocalization analysis. SNPs reported to have minor allele frequencies of 0% or 100%, indicating very rare variants, in either GWAS were excluded from analysis. After this filtering, genetic loci colocalization analysis windows, hereinafter referred to as analysis windows, were created by merging with other nearby genome-wide significant signals to form genetic loci for colocalization analysis. We performed colocalization analysis on these genetic loci with Coloc.abf[11], Hyprcoloc[12], and visual inspection using LocusCompare[13].

The default priors were used for Hyprcoloc. In Coloc.abf, the default priors were used for the individual GWAS hypotheses, p1 and p2. The prior for single variant colocalization, p12, was set to 5e-06 as prescribed by Wallace et. al[14]. to balance false negative and positive results. Loci were considered colocalized if the regional probability of colocalization was greater than 0.70. In Coloc.abf, this was the sum of the PPH3 and PPH4 statistics, and in Hyprcoloc this was the regional probability statistic. Loci that met colocalization criteria in either method were plotted using LocusCompare with the default European ancestry linkage disequilibrium (LD) data from 1000Genomes and with genome build hg19. This resulted in 121 LocusCompare plots on which visual inspection was performed in order to identify colocalized genetic association signals. If genetic loci were considered colocalized by at least two of the three colocalization analysis methods, we considered these consensus genetic loci. We termed this consensus analysis 'COLOC'.

Discordant Locus Identification. We obtained the 99% credible set of SNPs from the results of Bayesian Factor Analysis implemented through Coloc.abf at each locus. We calculated the Z-scores for the association test of each genetic variant, and if the Z-score associated with SNP had the opposite sign for association with WHRadjBMI and T2DadjBMI with respect to the same allele, we considered it discordant. We then identified which loci the SNPs were located, and used the haploR package in R[15] to separate signals in the same loci using LD clumping (R^2>0.50) on the discordant variants.

GWAS-QTL Colocalization Analysis. We obtained expression quantitative trait locus (eQTL) data from the GenoType-Expression Project (GTEx) for 49 tissues[16], the Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) cohort for 6 tissues[17], and the Metabolic Syndrome in Men (METSIM) for subcutaneous adipose tissue[18]. We also obtained subcutaneous adipose tissue splice QTL (sQTL) results from the METSIM cohort[19]. Data sources and further information are detailed in Supplementary Table 1. We extracted the QTL data for each gene or transcript within 1Mb of a discordant locus start or end site, and independently colocalized with the T2DadjBMI and WHRadjBMI GWAS using Coloc.abf. We considered a locus as colocalized if PPH4 was greater than 0.50, a threshold used for GWAS-eQTL colocalization in admixed populations[20], and if they passed visual inspection.

fGWAS Annotation Enrichment analysis. We used the functional GWAS (fGWAS) command-line tool to compute the enrichment of the credible set of SNPs from each of the 79 colocalized loci. We mapped SNPs to annotations with coordinates in adipose, liver, pancreatic islet, and skeletal muscle cell chromatin-state bed files[21], as well as mapping SNPs to coding regions, and performed enrichment analysis with the complete model of all annotations with the -fine and -xv flags on fGWAS. We used the natural-log of the Bayes Factor of the colocalization test, and computed the log-base enrichment for genetic and epigenetic annotations. From the log-base enrichment, we computed annotation enrichment.

Tissue of Action Analysis. We conducted tissue-of-action (TOA) score analysis using the credible set of SNPs from each of the 79 colocalized loci. We used the Coloc.abf PPH4 scores, genetic variant coordinates, and HaploReg LD data to separate independent association signals in the same loci[10]. We used the enrichment scores calculated using fGWAS to weight signal score assignment. We used TACTICAL to integrate the credible set of SNPs with the genomic annotation weights, and genomic annotation maps for adipose, pancreatic islet, skeletal muscle, and liver tissues. We used the default tissue classification thresholds of .20 to classify signals as belonging to a TOA, with a .10 shared threshold for shared TOA assignment.

Genetic Expression – Phenotype Correlation Analysis. We calculated the biweight midcorrelation and its significance, as implemented by the Weighted Genetic Coexpression Network Analysis (WGCNA) package[22], between genes of interest and metabolic phenotypes measured in the METSIM cohort[23].

RESULTS

Colocalization analysis of genetic loci associated with Type 2 Diabetes and Body Fat Distribution predicts colocalization of discordant T2DadjBMI and WHRadjBMI association signals at six loci.

In order to predict genetic loci where a single variant was significantly associated with both T2DadjBMI and WHRadjBMI, we selected analysis windows using genetic variants within 200kB of a SNP that reached genome-wide significance. Proximal windows within 100kB of another window were merged to form our genetic loci for COLOC (Figure 1A). This analysis yielded 79 genetic loci where a single variant was significantly associated with both T2DadjBMI and WHRadjBMI. We obtained the 99% credible SNP set (Supplementary Table 2) and discovered that 143 variants in colocalized loci were associated with discordant effects on T2DadjBMI and WHRadjBMI (Figure 1B) with respect to the same allele of a SNP. These discordant variants represented 6 independent loci in five independent cytoband regions. No secondary signal was apparent in these loci.

To investigate the genetic variants associated with discordant effects on WHRadjBMI and T2DadjBMI, we performed a phenome-wide association study (PheWAS) for metabolically related traits and queried HaploReg to discover genetic variant annotations. We used the highest powered GWAS available for our PheWAS, and searched for significantly associated traits with each discordant SNP (Figure 1C). We observed that lipidemic and glycemic trait associations clustered together, as expected. Five of the six discordant association signals were associated with discordant effects on lipid and glycemic traits. At the second association signal in the 5q11.2 region, association signals exemplified this metabolic discordance. Represented by genetic variant rs459193, the association signal was associated with protection for T2D, but with higher fasting glucose[24], fasting insulin[24], increased Hemoglobin A1C[24], and lower glucose tolerance[24]. rs459193 was associated with higher abdominal fat accumulation[6], but also with lower HDL levels[25], circulating triglycerides[25], and total cholesterol[25], but with higher LDL[25]. Discordant variants overwhelmingly lie in noncoding regions of the genome, with only one missense variant in a coding region (Figure 1D). Because the vast majority of discordant variants lie in noncoding regions, predicting the functional genetic effect predisposing carriers to decreased T2D risk and increased abdominal fat accumulation is not trivial. Therefore, we further investigated these variants with functional prediction methods.

Integration of molecular QTLs and genomic annotations to predict functional genes in tissues of action at discordant genetic loci.

To predict the genes and tissues of function at discordant loci, we used publicly available multi-omic data from metabolically relevant tissue-specific resources to predict functional mechanisms underlying associations. We first interrogated where the 143 discordant variants in the credible set were located in relation to tissue-specific epigenomic annotations in pancreatic islet, adipose, liver, and skeletal muscle tissues[21]. Variants were overwhelmingly located in regions of weak transcriptional activity of all four tissues, followed by pancreatic islet and liver enhancer regions. Using these discordant variant maps, we calculated the enrichment of these variants for tissue-specific annotations (Figure 2A). Using variant annotation maps, the enrichment scores, and gene expression data from the GTEx project[26], we implemented Tissue of Action (TOA) score analysis in TACTICAL to partition Bayesian posterior probabilities of the credible set of colocalized variants and predicted functional tissues underlying associations at all 100 genetic association signals in colocalized loci (Supplementary Table 3). Adipose tissues were classified as contributors at all 6 discordant loci. Adipose tissue was classified as the sole TOA at four loci, and skeletal muscle and liver tissue shared classification with adipose tissue at the remaining two discordant loci (Figure 2B). Within the entire set of classified TOA scores, we found that signal TOA assignment did not group along tissue assignments (Supplementary Figure 1). The tissue assignments of four genetic association signals in concordant loci were separated from the other 96 along the first primary component, but discordant loci were grouped with the other loci.

Orthogonally, we used Coloc.abf to predict colocalization of discordant genetic associations with quantitative trait loci (QTL) summary statistics from multiple cohorts and tissue groups (Supplementary Table 1). We extracted expression QTL (eQTL) and splicing QTL (sQTL) summary statistics for all genes within 1Mb of the lead variant of all discordant colocalized loci from adipose, pancreatic, skeletal muscle, and liver tissues. Using Coloc.abf, we predicted colocalization of QTLs with the lead genetic variant at two of the six discordant loci (Figure 2C). We predicted four potentially causal effector transcripts in these two loci. These genes are *GIN1*,

PAM, *PPIP5K2*, and *THADA-AS* (Supplementary Table 4). Using both Coloc.abf and TACTICAL enabled us to make high-confidence consensus predictions of tissues and genes of action at discordant loci.

Correlation analysis of genetic expression reveals physiological mechanisms at play in discordant genetic loci.

Significant correlations between tissue-specific gene expression and metabolic phenotypes can reveal biologically-relevant functional relationships. We used subcutaneous adipose tissue transcriptomic data from the 426 men of the METSIM cohort to investigate how adipose tissue transcription was related to 23 metabolic phenotypes underlying T2D and abdominal fat accumulation[18,19,23]. We extracted gene-expression data for genes with an eQTL or sQTL colocalization within 1Mb of the start or end position of a discordant locus. We then extracted transcript-specific expression data for *PAM*, for which three splice clusters colocalized with discordant genetic loci. We then calculated the biweight midcorrelation for adipose tissue gene and transcript expression with 23 metabolic phenotypes. Gene expression data for all four of potentially causal genes was available. We found significant (FDR < 0.05) correlations of adipose tissue gene expression of three genes with thirteen phenotypes (Figure 3A). *THADA-AS* expression, the predicted functional gene within 2p21 locus, was significantly associated with a metabolically unhealthy profile. Subcutaneous adipose tissue *THADA-AS* expression was positively correlated with insulin resistance, abdominal fat accumulation, serum triglyceride levels, but with higher levels of plasma Interleukin-1 receptor antagonist (IL-1Ra) and C-reactive protein (CRP). While IL-1Ra plays a protective role in inflammation[27], CRP has been used as a biomarker of increased inflammation in acute injuries and chronic disease[28]. Expression of *GIN1* was correlated with higher plasma adiponectin and lower plasma insulin levels. Adiponectin, secreted by adipocytes, increases insulin-sensitivity, and this provides a mechanism for protection from T2D. We additionally found significant positive correlations of subcutaneous adipose tissue expression of one *PAM* transcript variant, *ENST00000504691.1*, with abdominal fat accumulation and an inflammatory profile (Figure 3B). Together, the colocalization evidence and correlation evidence supports the hypothesis that the same genetic association signal is responsible for a functional *GIN1* eQTL and *PAM* sQTL that promotes obesity and protects against T2D risk. Through this correlation analysis, we were able to predict physiological consequences of genes genetically regulated through QTLs and predict the causal genes at two of the six discordant genetic loci.

Integration of multi-omic data to predict the functional gene and tissue of action at the discordant 2p21 locus

By predicting the mechanisms of action at discordant loci, we were able to generate specific hypotheses about the genes at each locus that underlie GWAS associations. We were able to achieve a consensus prediction at the 2p21 locus containing the gene encoding Thyroid Adenoma Associated Protein (THADA). THADA was first metabolically characterized in drosophila melanogaster, where investigators found that *Thada* knockout resulted in increased triglyceride accumulation and adipogenesis. Through interactions with Sarco-endoplasmic Reticulum Ca^{2+} ATPase, THADA uncouples endoplasmic reticulum Ca^{2+} transport from ATP expenditure, which results in increased non-shivering thermogenesis and energy expenditure[29]. Therefore, it has been hypothesized that developing THADA agonists may provide a new therapeutic avenue for obesity. The lead association signal in the 2p21 locus is associated with increased abdominal fat accumulation, but decreased T2D risk (Figure 4A). This association signal is also associated with increased expression of *THADA* Antisense RNA *AC010883.5* (*THADA-AS*) in adipose tissues specifically (Figure 4B, 4C). Additionally, TACTICAL classified the association signal at the 2p21 locus to have a likely adipose tissue mechanism of action (Figure 2B). To investigate this, we obtained ATAC-Sequencing of preadipocytes and mature adipocytes[30]. Chromatin accessibility in the 2p21 locus is significantly increased in the locus following differentiation (Figure 4D). This consensus of results predicts that the downregulation of *THADA* expression by its antisense RNA *THADA-AS* in subcutaneous adipose tissue promotes phenotypes associated with MHO/LIR.

DISCUSSION

We report here the integration of multi-omic data spanning the genome, transcriptome, and epigenome to predict functional genes and tissues underlying association signals protecting carriers from T2D while increasing abdominal fat accumulation. Using colocalization analysis, we predicted colocalization of T2DadjBMI and WHRadjBMI association signals at 79 genetic loci. The protective allele of 6 association signals were associated with lower T2D risk but higher abdominal fat accumulation, independent of overall obesity (Figure 1). By analyzing colocalization with molecular QTLs and partitioning posterior probabilities using enrichment of genomic annotations, we were able to predict genes and tissues of action that could contribute to MHO/LIR (Figure 2). We predicted that the protective allele of a genetic variant in the 2p21 locus resulted in higher expression of *THADA-AS* in adipose tissues specifically, and that increased expression of *THADA-AS* results in increased abdominal fat accumulation but decreased T2D risk (Figure 3). *THADA* and *THADA-AS* adipose tissue expression was positively correlated with metabolically unfavorable obesity phenotypes, but negatively correlated with insulin-related phenotypes (Figure 4). These results are the first comprehensive prediction of genetic mechanisms that uncouple obesity from T2D risk.

The six genetic loci associated with discordant metabolic phenotypes offer insight into genetic mechanisms promoting MHO/LIR, an understudied group of phenomena uncoupling obesity and diabetes risk. Because MHO/LIR encompasses a number of phenotypes with multiple definitions, studying and identifying cases of it is a major undertaking. MHO has been defined as obesity with less than 3 components of the Metabolic Syndrome (MetSyn) and as obesity with healthy HOMA-IR. Because of this lack of precision, MHO/LIR has been controversial and termed an intermediate state[31], but a growing body of evidence has accumulated providing evidence that genetic mechanisms influence predisposition to it. In Samoans, the common *CREBRF* coding variant rs12513649 increases BMI and overall adiposity but protects from insulin resistance[32]. Additionally, *IRS1*, *COBLL1*, *PLA2G6*, and *TOMM40* have been associated with higher BMI but with protective lipidemic and glycemic traits[33]. The physiological functions of these genes have been proposed to involve adipose tissue caloric load capacity and body fat distribution[32,33].

While abdominal fat accumulation has been proposed as the strongest predictor of obesity-related complications[4,5,31,33,34], our findings point to mechanisms independent of both overall obesity and abdominal obesity that contradict this trend. Further, this investigation predicted specific hypotheses linking functional genes at discordant loci to tissues of action for experimental follow up *in vitro* and *in vivo*. If these loci are functionally annotated and fully characterized, they will additionally provide medical benefit to risk allele carriers. Discovering mechanisms uncoupling obesity from T2D can aid in personalized therapeutic strategies and in understanding personalized risk stratification. Risk-stratified personalized obesity treatment could prioritize patients that would benefit from weight-loss interventions, and use genotype as a biomarker for patients who would benefit from other therapeutic strategies[3]. Thus, the importance of personalized risk stratification will only increase as abdominal obesity becomes more prevalent, and will benefit from precise diagnoses that incorporate the specific molecular, cellular, and physiological mechanisms that result in the symptoms that lead to diagnosis.

We predicted tissues and mechanisms of action at six loci that are associated with abdominal obesity and protection for type 2 diabetes. A particular example of a peculiar metabolic discordance was revealed at the 2p21 locus containing *THADA* and *THADA-AS*, represented by variant rs6752964. The associations have been replicated multiple times[7,35], but the exact mechanisms underlying this association are unknown. *THADA* plays an evolutionarily conserved role in intracellular calcium signaling and consequently non-shivering thermogenesis. In Drosophila melanogaster, *Thada* knockout flies developed obesity and hyperphagia without altered circulating glucose levels[29]. Follow up studies using Mendelian Randomization have likewise found consistent links between THADA and adiposity, but failed to link THADA with diabetic traits such as insulin resistance[36]. Our investigation was able to find significant correlations between *THADA-AS* expression with diabetic and obesity-related traits. *THADA-AS* expression was positively correlated with hyperinsulinemia and hyperglycemia, but surprisingly also with insulin sensitivity. Consistent with the increased abdominal fat accumulation, *THADA-AS* expression was positively correlated with inflammation, unhealthy lipid profiles, and hypertensive traits. We also predicted that *GIN1*, *PPIP5K2*, and *PAM* are regulated by the genetic association signal represented by variant rs74567345, and act through subcutaneous adipose tissue (Figure 2) to increase abdominal fat accumulation and decrease risk for type 2 diabetes (Figure 1). PAM has additionally been found to impact pancreatic β-cell insulin release[37] , but haploinsufficiency of *Pam* in mice did not accelerate development of diabetes and obesity in mice[38].

We additionally found significant correlations with inflammatory markers and intermediate abdominal obesity phenotypes with increased transcription of *PAM* transcript variant *ENST00000504691.1*, but protection from intermediate diabetic phenotypes (Figure 3), further providing evidence for this uncoupling role. Further experimental characterization of genes within discordant loci in animal models is necessary to elucidate the specific mechanisms by which genetic factors impact MHO/LIR.

Although our analysis incorporated genome, transcriptome, epigenome, and phenome data in multiple cohorts, and used the consensus of orthogonal methods to predict the mechanisms of action at discordant loci, follow up is required to validate each prediction. Additionally, our genetic expression data used single strand sequencing, and therefore parsing out the associated effects of sense and antisense transcripts is difficult. Finally, it is critical to the discovery of more genetic loci underlying MHO/LIR to diversify ancestry in genetic association studies. Without experimental follow up and extensive clinical studies, genotype should not be used as a diagnostic metric. This followup and further experimental characterization is critical to placing our results in the proper context and providing the basis for personalized interventions for T2D. The predictions at these six loci provide specific hypotheses to be tested, and should they be validated experimentally provide knowledge of the precise mechanisms uncoupling obesity from T2D risk.

DATA AVAILABILITY

Our analysis pipeline is publicly available on Github (https://github.com/aberrations/predicting-functional-mechanisms-discordant-loci), and we include source data information in Supplementary Table 1.

SOURCES OF FUNDING

This work was supported by R01 DK118287 (to M.C.) from the National Institute of Diabetes and Digestive and Kidney Diseases, T32 HL007284 (to Y.T.A.) from the National Heart Lung and Blood Institute, 1-19-IBS-105 (to M.C.) from the American Diabetes Association, and the The Louis Stokes Alliances for Minority Participation Bridge-to-the-Doctorate Virginia-North Carolina Alliance Fellowship (to Y.T.A.) from the National Science Foundation.

REFERENCES

1. McCarthy, M. I. Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* **363**, 2339–2350 (2010).
2. Brown, E., Heerspink, H. J. L., Cuthbertson, D. J. & Wilding, J. P. H. SGLT2 inhibitors and GLP-1 receptor agonists: established and emerging indications. *The Lancet* **398**, 262–276 (2021).
3. Blüher, M. Metabolically Healthy Obesity. *Endocr. Rev.* **41**, bnaa004 (2020).
4. Iacobini, C., Pugliese, G., Blasetti Fantauzzi, C., Federici, M. & Menini, S. Metabolically healthy versus metabolically unhealthy obesity. *Metabolism.* **92**, 51–60 (2019).
5. Emdin, C. A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626–634 (2017).
6. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
7. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
8. Wallace, C. *et al.* Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet.* **21**, 2815–2824 (2012).
9. Fortune, M. D. *et al.* Statistical Colocalization of Genetic Risk Variants for Related Autoimmune Diseases in the Context of Common Controls. *Nat. Genet.* **47**, 839–846 (2015).
10. Torres, J. M. *et al.* A Multi-omic Integrative Scheme Characterizes Tissues of Action at Loci Associated with Type 2 Diabetes. *Am. J. Hum. Genet.* **107**, 1011–1028 (2020).
11. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).
12. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
13. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* **51**, 768–769 (2019).
14. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS Genet.* **16**, e1008720 (2020).
15. Zhbannikov, I. Y., Arbeev, K., Ukraintseva, S. & Yashin, A. I. haploR: an R package for querying web-based annotation tools. *F1000Research* **6**, 97 (2017).

16. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

17. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).

18. Raulerson, C. K. *et al.* Adipose Tissue Gene Expression Associations Reveal Hundreds of Candidate Genes for Cardiometabolic Traits. *Am. J. Hum. Genet.* **105**, 773–787 (2019).

19. Brotman, S. M. *et al.* Subcutaneous adipose tissue splice quantitative trait loci reveal differences in isoform usage associated with cardiometabolic traits. *Am. J. Hum. Genet.* **109**, 66–80 (2022).

20. Gay, N. R. *et al.* Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. *Genome Biol.* **21**, 233 (2020).

21. Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2301–2306 (2017).

22. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

23. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).

24. Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).

25. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).

26. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

27. Volarevic, V., Al-Qahtani, A., Arsenijevic, N., Pajovic, S. & Lukic, M. L. Interleukin-1 receptor antagonist (IL-1Ra) and IL-1Ra producing mesenchymal stem cells as modulators of diabetogenesis. *Autoimmunity* **43**, 255–263 (2010).

28. Herwald, H. & Egesten, A. C-Reactive Protein: More than a Biomarker. *J. Innate Immun.* **13**, 257–258 (2021).

29. Moraru, A. *et al.* THADA Regulates the Organismal Balance between Energy Storage and Heat Production. *Dev. Cell* **41**, 72-81.e6 (2017).

30. Perrin, H. J. *et al.* Chromatin accessibility and gene expression during adipocyte differentiation identify context-dependent effects at cardiometabolic GWAS loci. *PLOS Genet.* **17**, e1009865 (2021).

31. Smith, G. I., Mittendorfer, B. & Klein, S. Metabolically healthy obesity: facts and fantasies. *J. Clin. Invest.* **129**, 3978–3989 (2019).

32. Minster, R. L. *et al.* A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat. Genet.* **48**, 1049–1054 (2016).

33. Genes that make you fat, but keep you healthy. https://onlinelibrary.wiley.com/doi/epdf/10.1111/joim.12827.

34. Denis, G. V. & Obin, M. S. 'Metabolically healthy obesity': Origins and implications. *Mol. Aspects Med.* **34**, 59–70 (2013).

35. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).

36. Grarup, N. *et al.* Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* **57**, 2534–2540 (2008).

37. Thomsen, S. K. *et al.* Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic β-cells. *Nat. Genet.* **50**, 1122–1131 (2018).

38. Chen, Y.-C. *et al.* PAM haploinsufficiency does not accelerate the development of diet- and human IAPP-induced diabetes in mice. *Diabetologia* **63**, 561–576 (2020).

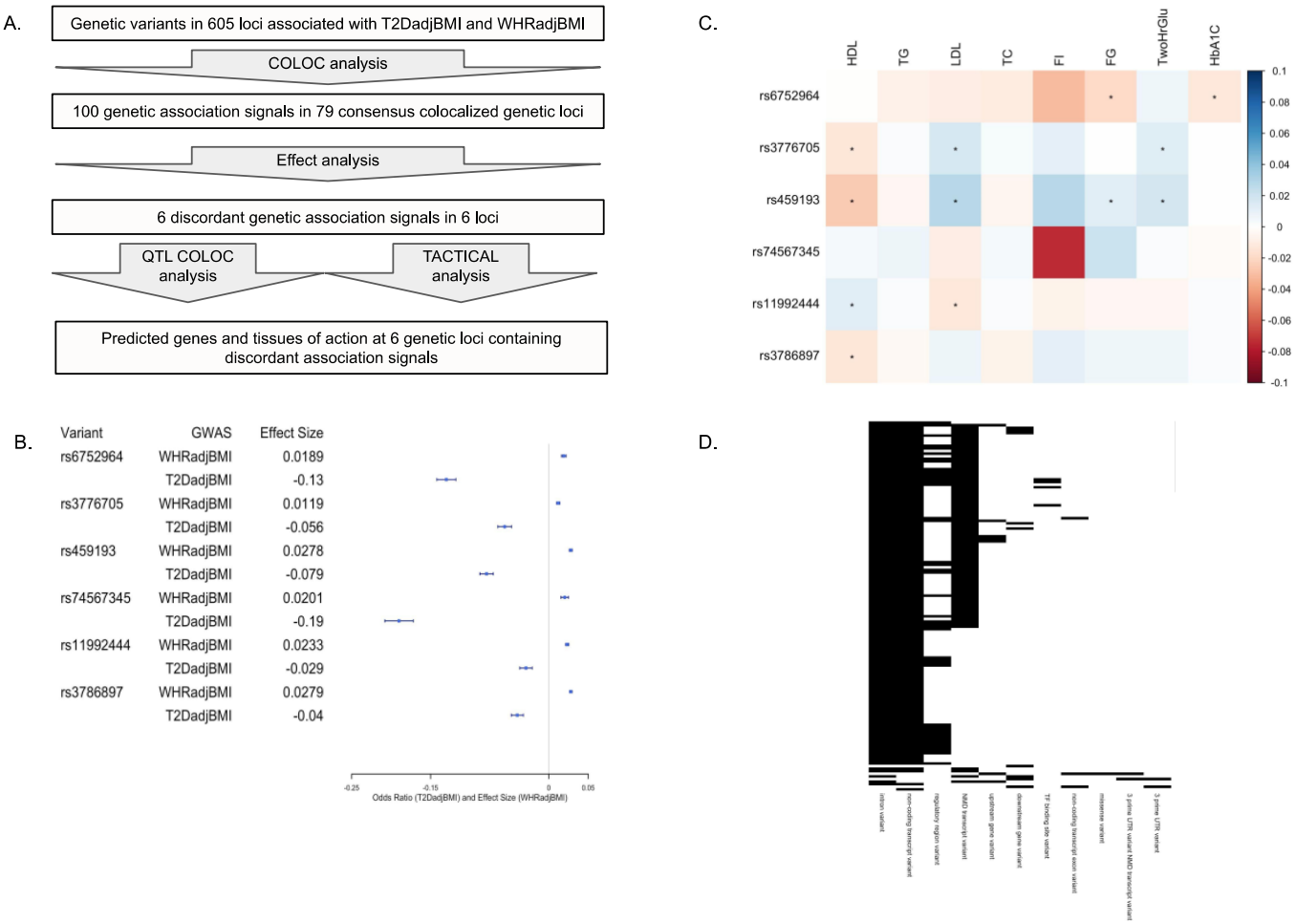
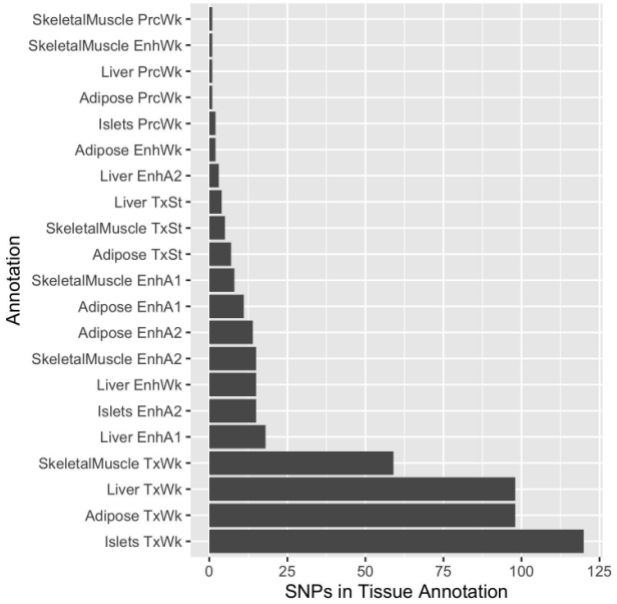
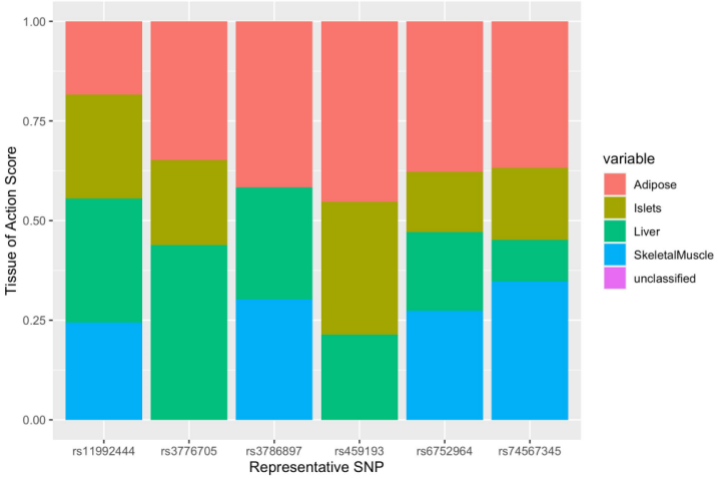


Figure 1. Analysis summary and discordant variant characteristics. (A) Summary of analysis pipeline and generated results. Details of data sources are available in Supplementary Table 1. (B) Effect size (WHRadjBMI) and Odds Ratio (T2DadjBMI) of lead genetic variant at discordant association signals. (C) PheWAS of lead discordant genetic variant effect sizes on High Density Lipoprotein (HDL), Triglyceride (TG), Low Density Lipoprotein (LDL), Total Cholesterol (TC), Fasting Insulin (FI), Fasting Glucose (FG), Two-Hour Glucose Tolerance Test (TwoHrGlu), and Glycated hemoglobin (HbA1C). Asterisk denote associations that reach genome-wide significance level (p-value < 5e-08). (D) Characteristics of 99% credible set variants in discordant genetic loci. Black color indicates the presence of variant annotation.

A.



B.



C.

SNP	Chr	Position	QTL Gene	Colocalized Tissues: Subcutaneous Adipose	Colocalized Tissues: Visceral Adipose	Colocalized Tissues: Skeletal Muscle	Expression Effect
rs6752964	2	43754655	THADA-AS	●	● ●		↓
rs74567345	5	102338811	PAM	▲		●	↑
rs74567345	5	102338811	GIN1	●			↑
rs74567345	5	102338811	PPIP5K2	●			↓

● eQTL ▲ sQTL T2DadjBMI WHRadjBMI

Figure 2. Functional annotation of discordant loci. (A) Tissue-specific epigenomic annotation of discordant variants in the 99% credible set. Epigenomic annotations are Weak Transcription (TxWk), Enhancer Activity (EnhA1/2), Weak Enhancer Activity (EnWk), Weak Polycomb Repressed Complex (PrcWk), and Strong Transcription (TxSt).(B) Tissue of Action scores for association signals in the six discordant loci. (C) Summary table of the eQTL and sQTL colocalizations with WHRadjBMI and T2DadjBMI for discordant loci. Expression effect is with respect to the protective type 2 diabetes allele.

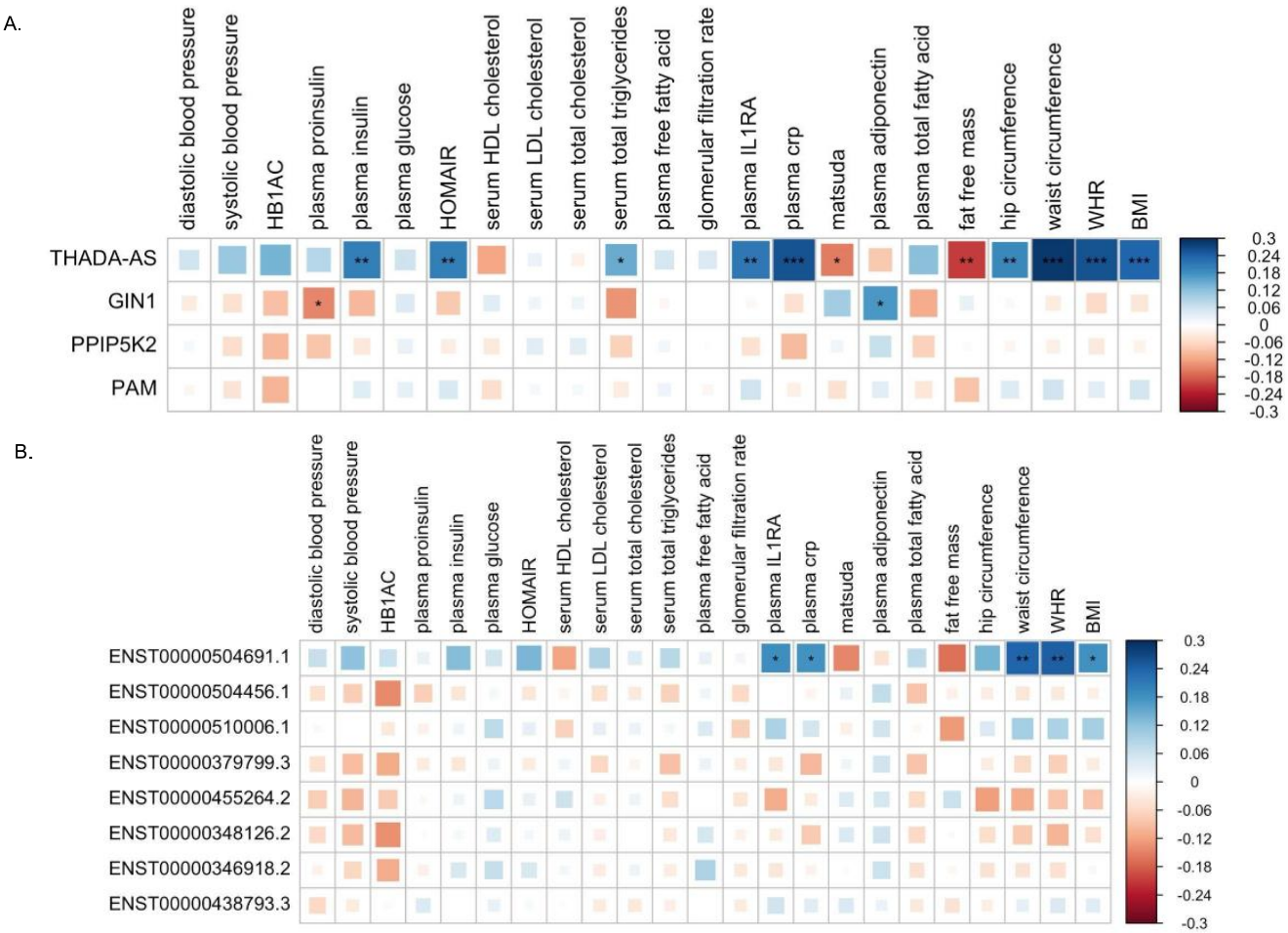


Figure 3. Correlations of colocalized transcripts with metabolic phenotypes. Biweight midcorrelation of (A) colocalizaed eQTL gene expression and (B) transcripts of the PAM sQTL gene with metabolic phenotypes measured in 426 participants of the METSIM cohort.

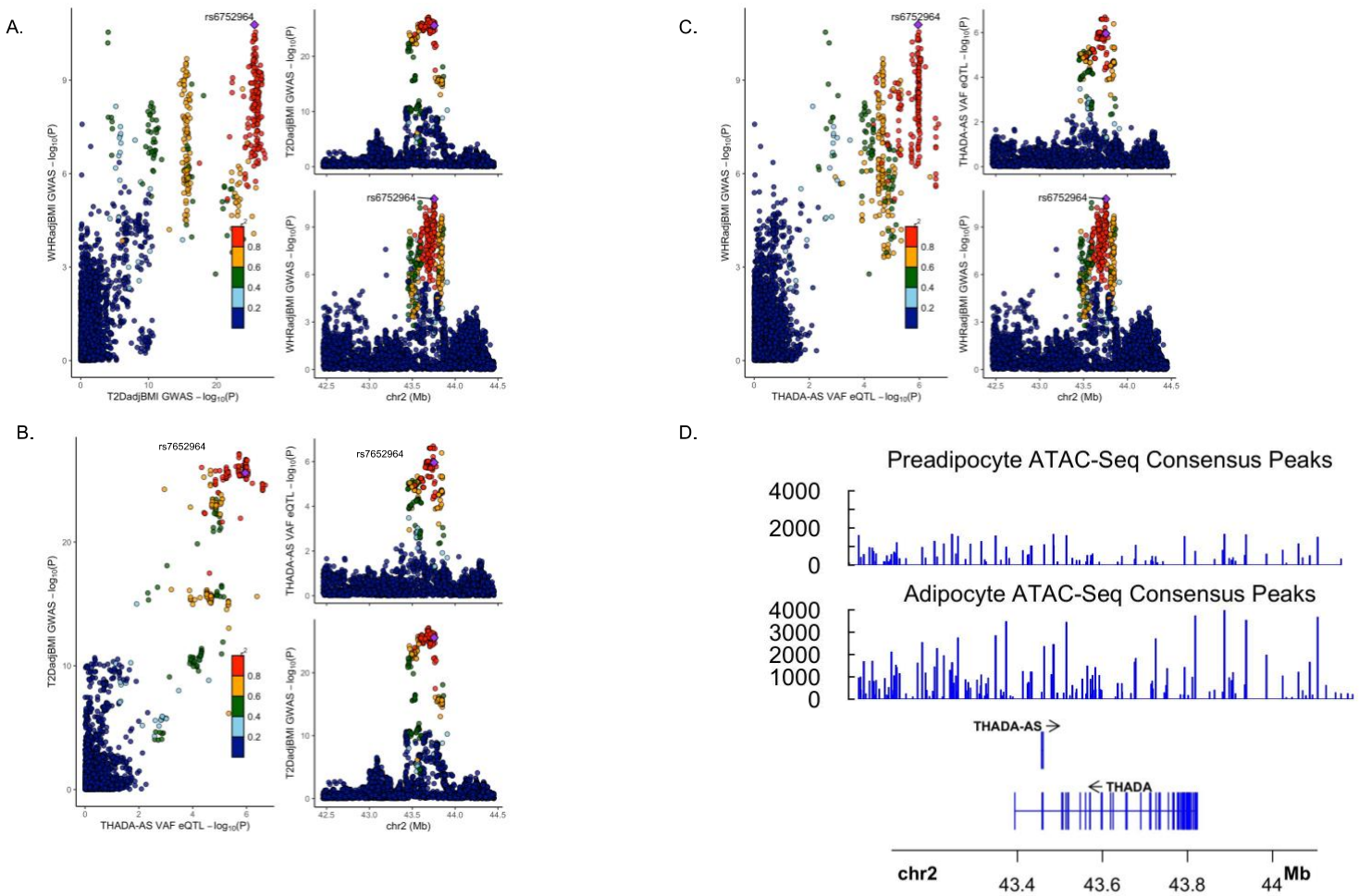


Figure 4. THADA-AS as a potential causal gene at the 2p21 locus. Colocalization of (A) T2DadjBMI and WHRadjBMI GWAS signals as well as (B) T2DadjBMI and (C) WHRadjBMI GWAS and THADA-AS eQTL signal at the 2p21 locus. Chromatin accessibility peaks of preadipocytes and mature adipocytes at the 2p21 locus. ATAC-seq data obtain from GEO accession number GSE178796[30].