

UC San Diego

UC San Diego Previously Published Works

Title

Using metabolite profiling to construct and validate a metabolite risk score for predicting future weight gain

Permalink

<https://escholarship.org/uc/item/2j71878v>

Journal

PLOS ONE, 14(9)

ISSN

1932-6203

Authors

Geidenstam, Nina
Hsu, Yu-Han H
Astley, Christina M
[et al.](#)

Publication Date

2019

DOI

10.1371/journal.pone.0222445

Peer reviewed

RESEARCH ARTICLE

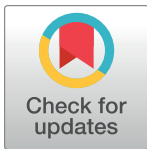
Using metabolite profiling to construct and validate a metabolite risk score for predicting future weight gain

Nina Geidenstam¹, Yu-Han H. Hsu^{2,3,4}, Christina M. Astley^{3,4}, Josep M. Mercader^{4,5}, Martin Ridderstråle¹, Maria E. Gonzalez^{6,7}, Clicerio Gonzalez^{6,7}, Joel N. Hirschhorn^{2,3,4}, Rany M. Salem⁸*

1 Clinical Obesity, Institution of Clinical Sciences, Malmö, Lund University, Sweden, **2** Department of Genetics, Harvard Medical School, Boston, MA, United States of America, **3** Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, United States of America, **4** Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, United States of America, **5** Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States of America, **6** Instituto Nacional de Salud Publica, Cuernavaca, Morelos, Mexico, **7** Centro de Estudios en Diabetes, Mexico City, Mexico, **8** Department of Family Medicine and Public Health, UC San Diego, San Diego, CA, United States of America

* These authors contributed equally to this work.

* rsalem@ucsd.edu (RMS); Joel.Hirschhorn@childrens.harvard.edu (JNH)



OPEN ACCESS

Citation: Geidenstam N, Hsu Y-HH, Astley CM, Mercader JM, Ridderstråle M, Gonzalez ME, et al. (2019) Using metabolite profiling to construct and validate a metabolite risk score for predicting future weight gain. PLoS ONE 14(9): e0222445. <https://doi.org/10.1371/journal.pone.0222445>

Editor: Robin D. Clugston, University of Alberta, CANADA

Received: February 11, 2019

Accepted: August 29, 2019

Published: September 27, 2019

Copyright: © 2019 Geidenstam et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data from the Framingham Heart Study is available from dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v30.p11). We do not have ownership of the MCDS data and obtained permission to analyze the data from the SIGMA T2D Consortium. Previously published MCDS data can be accessed through the T2D Knowledge Portal (<http://www.type2diabetesgenetics.org>) and inquiries for unpublished data should be directed to Jose Florez

Abstract

Background

Excess weight gain throughout adulthood can lead to adverse clinical outcomes and are influenced by complex factors that are difficult to measure in free-living individuals. Metabolite profiling offers an opportunity to systematically discover new predictors for weight gain that are relatively easy to measure compared to traditional approaches.

Methods and results

Using baseline metabolite profiling data of middle-aged individuals from the Framingham Heart Study (FHS; $n = 1,508$), we identified 42 metabolites associated ($p < 0.05$) with longitudinal change in body mass index (BMI). We performed stepwise linear regression to select 8 of these metabolites to build a metabolite risk score (MRS) for predicting future weight gain. We replicated the MRS using data from the Mexico City Diabetes Study (MCDS; $n = 768$), in which one standard deviation increase in the MRS corresponded to ~ 0.03 increase in BMI (kg/m^2) per year (i.e. ~ 0.09 kg/year for a 1.7 m adult). We observed that none of the available anthropometric, lifestyle, and glycemic variables fully account for the MRS prediction of weight gain. Surprisingly, we found the MRS to be strongly correlated with baseline insulin sensitivity in both cohorts and to be negatively predictive of T2D in MCDS. Genome-wide association study of the MRS identified 2 genome-wide ($p < 5 \times 10^{-8}$) and 5 suggestively ($p < 1 \times 10^{-6}$) significant loci, several of which have been previously linked to obesity-related phenotypes.

(jcflores@mg.harvard.edu) and Dorothy Pazin (dorothy@broadinstitute.org).

Funding: We thank the Broad Metabolomics Platform and SIGMA T2D Consortium for sharing data resources. This work was supported by the National Heart, Lung, and Blood Institute grants F31HL126581 (Y.H.H.) and R00HL122515 (R.M.S.), National Institute of Diabetes and Digestive and Kidney Diseases grants T32DK110919 (Y.H.H.), K12DK094721 (C.M.A.), and R01DK075787 (J.N.H.), Endocrine Scholars Award (C.M.A.), and Doris Duke Charitable Foundation (<http://www.ddcf.org>) grant 215205 (J.N.H.). Genome-wide genotyping and metabolite profiling of the MCDS cohort was performed as part of the Slim Initiative in Genomic Medicine for the Americas (SIGMA) T2D Consortium. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: J.N.H. serves on the Scientific Advisory Board of Camp4 Therapeutics.

Conclusions

We have constructed and validated a generalizable MRS for future weight gain that is an independent predictor distinct from several other known risk factors. The MRS captures a composite biological picture of weight gain, perhaps hinting at the anabolic effects of preserved insulin sensitivity. Future investigation is required to assess the relationships between MRS-predicted weight gain and other obesity-related diseases.

Introduction

Obesity is a global epidemic [1] associated with adverse clinical outcomes including type 2 diabetes (T2D) [2], cardiovascular disease (CVD) [3], non-alcoholic fatty liver disease [4], and death [5]. Population studies demonstrate that adults, on average, continue to gain weight at a rate of 0.4–1 kg per year through the middle decades of life [6–8]. Likewise, the percentage of adults who are overweight or obese (body mass index, BMI ≥ 25 kg/m²) has increased from around 29% to 37% globally over the past three decades [1].

Excess weight gain can be intuitively explained by an excess of energy intake relative to energy expenditure and energy partitioning within an individual; however, the components leading to energy imbalance are very complex, highly interactive, and hard to measure in free-living populations over extended time periods [9]. Therefore, even though studies have identified various risk factors for weight gain, such as unhealthy diet, reduced physical activity, change in smoking status, and glucose-stimulated insulin secretion, the biological mechanisms linking these risk factors to weight gain have yet to be fully elucidated [8,10–12]. In terms of genetic predisposition, while recent genome-wide association studies (GWAS) have identified hundreds of BMI-associated loci [13,14], little is known about the genetics of longitudinal change in BMI, which do not always agree with that of cross-sectional BMI [15]. Furthermore, while recent studies have developed multivariate models that incorporate different types of risk factors to predict weight gain [16,17], these models often include variables that may not be readily available in many research or clinical settings (e.g. recalled lifestyle and dietary variables derived from detailed questionnaires and specialized calculations). Overall, the etiology of obesity remains incompletely understood and there is a lack of easily generalizable prediction models for weight gain, thus non-surgical interventions to modify weight have largely been met with modest effect [18,19].

Metabolite profiling systematically measures metabolite levels in biological samples and offers a new, powerful approach for studying both intrinsic (i.e. genetic) and extrinsic (i.e. environmental) influences on obesity phenotypes in a data-driven, hypothesis-free manner that does not depend on prior knowledge. Novel biomarkers for a trait could be identified by testing for correlation between metabolites and the trait measured at a single time point; furthermore, if longitudinal data are available, predictive biomarkers could also be identified by associating metabolites with a future outcome. Numerous studies have successfully applied this approach to identify correlated or predictive biomarkers for obesity and cardiometabolic traits [20–27]. For instance, branched-chain amino acids (BCAAs), aromatic amino acids, and lipids with low carbon number and double bond content have been shown to correlate with cross-sectional BMI [21,22] and even predict the risk of developing insulin resistance and/or T2D [24–26].

In contrast, there are limited studies of metabolite profiling with longitudinal weight trajectories. A few studies have identified lipids, BCAAs, and metabolites involved in energy

metabolism and/or oxidative stress to be correlated with past weight change [28,29]. While these correlated biomarkers provide insights into the metabolic consequences of weight change, they are not as useful as predictive biomarkers in terms of informing clinical obesity prevention strategies. On the other hand, studies that looked at association between metabolites and future weight change have identified acetylcholine, leucine, hippuric acid, acetylglycine, urate, and xanthine to be predictive biomarkers for significant weight gain [29,30]. These metabolites may be used as clinical risk factors and may even point to causal biological pathways for weight gain. However, none of these studies address how the individual metabolites may be combined into a replicable and generalizable risk score for predicting weight gain, nor do the studies compare the metabolites against conventional obesity risk factors to evaluate their independence and predictive power. These additional analyses are important for assessing the potential usefulness of metabolite predictors for weight gain.

The objectives of this study were to construct and validate a metabolite risk score (MRS) consisting of multiple metabolites for predicting future weight gain and to investigate the underlying biology captured by the MRS. Towards this end, we used longitudinal BMI measurements and baseline metabolite profiling data of 1,508 middle-aged individuals from the Framingham Heart Study (FHS) to construct an MRS consisting of 8 metabolites. The association of the MRS with future weight gain was replicated using data from 768 individuals in the Mexico City Diabetes Study (MCDS), demonstrating its validity and generalizability across study populations. To understand the biological information contained within the MRS, we investigated its association with other obesity-associated factors and performed a GWAS for the MRS. Our findings indicate that the components of the MRS reflect diverse biological factors related to weight gain, possibly including anabolic effects of preserved insulin sensitivity and protection from T2D, which are unlikely to be captured by single metabolites or other obesity-associated factors.

Subjects and methods

Study samples and datasets

Discovery cohort: Data for the Framingham Heart Study (FHS), a prospective cohort study of CVD risk factors, were obtained through the NCBI database of Genotypes and Phenotypes (dbGaP) with accession number phs000007.v19.p7 [24,31]. Metabolite profiling was performed on 2,016 fasting plasma samples from the FHS Offspring Cohort collected at Exam 5 (1991–1995) using liquid chromatography-mass spectrometry (LC-MS) [32]. We filtered samples with quality controlled metabolite data ($n = 2,015$; see *Data processing*) to exclude samples with missing BMI values at Exam 5 or Exam 7 (1998–2002) ($n = 271$) and samples with clinical diagnosis of T2D ($n = 141$), CVD ($n = 154$) and/or renal disease ($n = 53$) at Exam 5, resulting in a total of 1,508 samples. T2D was defined as having fasting plasma glucose ≥ 126 mg/dl or using insulin or other anti-diabetic medications; CVD and renal disease status were ascertained using ‘Survival File and Follow-up for Cardiovascular Events’ (pht003316) and Exam 5 clinical exam (pht000034). For genome-wide association analyses, we further restricted to a subset of 1,349 samples with available genetic data (see **S1 Text** for details).

Replication cohort: The Mexico City Diabetes Study (MCDS), a prospective cohort study for T2D, has been described previously [33,34]. Briefly, LC-MS profiling was performed on 865 fasting plasma samples collected at the 1998 visit. Starting with 841 samples with quality controlled metabolite data (see *Data processing*), we excluded one sample with incorrect height measurement, samples with missing BMI values at the 1998 or 2008 visit ($n = 3$), and samples with CVD ($n = 68$) or renal disease ($n = 1$) at the 1998 visit, resulting in 768 samples for use in analysis. No T2D cases were present at the 1998 visit. T2D was defined as having fasting

glucose \geq 126 mg/dl, 2-hour glucose \geq 200 mg/dl in oral glucose tolerance test, or using anti-diabetic medication; CVD was defined as receiving a possible or probable myocardial infarction diagnosis in electrocardiograph results; renal disease was defined as having history of receiving kidney dialysis, transplant, or medication. For genome-wide association analyses, we used a subset of 588 samples with available genetic data (see [S2 Text](#) for details).

Data collection for both cohorts were conducted at external facilities by researchers not involved in the design and analysis of this study. All participants provided informed consent. This study was approved by the Boston Children's Hospital and UC San Diego Institutional Review Boards.

Data processing

Change in BMI: The rate of change in BMI was calculated as difference in BMI divided by difference in age between the baseline (b) and follow-up (f) time points: $(BMI_f - BMI_b) / (Age_f - Age_b)$, with age expressed in years. For FHS, we used Exams 5 and 7 as the baseline and follow-up time points, respectively; for MCDS, we used the 1998 and 2008 visits as the two time points. Height and weight values measured during the exam visits were used to derive BMI at each time point for both cohorts. We adjusted the rate of change in BMI for baseline age and sex and refer to the adjusted values as Δ BMI. As a sensitivity analysis, we performed rank-based inverse normal transformation on Δ BMI and confirmed that the transformed vs. untransformed Δ BMI values produced comparable results. To retain interpretability, we used the untransformed Δ BMI in units of $\text{kg/m}^2/\text{yr}$ in downstream analyses.

Metabolite data: Baseline metabolite profiling data for each cohort were quality controlled and standardized using the following steps: (1) remove samples and metabolites with $>$ 25% missing data, (2) log-transform the remaining metabolites, (3) adjust for covariates (age, sex, and fasting time for FHS; age and sex for MCDS), (4) impute missing values using a multiple imputation approach implemented in the MICE R package (v2.25; see [S3 Text](#) for details) [35], and (5) perform rank-based inverse normal transformation to calculate metabolite abundance z-scores. After these steps, we limited all subsequent analyses to 183 metabolites shared between the FHS and MCDS datasets.

Genetic risk score for BMI: We used 97 previously identified BMI-associated variants [13] and effect size estimates for their BMI-increasing alleles to calculate an effect size-weighted genetic risk score (GRS_{BMI}) for both FHS and MCDS samples. Large-scale Mexican-based BMI GWAS was not available, so European-based effect estimates were used for both cohorts despite ancestry differences.

Smoking status: Self-reported smoking information was available for FHS at both baseline and follow-up time points. We constructed a binary "Quit Smoking" variable, where subjects that were smokers at baseline but not at follow-up were categorized as "true" ($n = 97$) and subjects whose smoking status did not change were categorized as "false" (reference category).

Physical activity: Physical activity (PA) was calculated as the sum of time spent on moderate and heavy physical activities per day (hours/day), using self-reported questionnaires. PA data were not available at baseline, so we used the PA data collected prior to baseline time points, i.e. Exam 4 (mean 3.65 years before baseline) for FHS and the 1991 visit (mean 6.30 years before baseline) for MCDS. For FHS, which had follow-up PA data, we were able to calculate change in PA (Δ PA) as PA at Exam 7 minus PA at Exam 4.

Dietary intake: Daily macronutrient values (grams/day) derived from self-reported food frequency questionnaires were used to calculate total caloric intake (calories/day) and the percentage of energy intake from carbohydrate (Carb Intake), fat (Fat Intake), and protein (Protein Intake). For FHS, dietary data collected at baseline were used; for MCDS, data collected at

the 1991 visit were used due to lack of data available at baseline. For FHS, dietary fat intakes (monounsaturated, polyunsaturated, and saturated fats; grams/day) were also available.

Glycemic measures: Plasma fasting glucose (mg/dl) and insulin (mIU/l) concentrations, as well as 2-hour concentrations after oral glucose tolerance test, were available at baseline for both cohorts. The homeostasis model assessment of insulin resistance (HOMA-IR), Matsuda insulin sensitivity index (Matsuda ISI), and quantitative insulin sensitivity check index (QUICKI) were also available and calculated as described previously [36–38].

Constructing the MRS for weight gain

A schematic overview of the protocol for constructing, validating, and evaluating the MRS is provided in [S1 Fig](#).

Identifying Δ BMI-associated metabolites (S1A(i) Fig): First, to identify metabolites associated with Δ BMI, we performed univariate linear regression of Δ BMI on each metabolite in FHS. As both Δ BMI and metabolites were pre-adjusted for covariates (see *Data processing* above), no additional variables were included in the regression models. Since FHS contains families of related individuals ([S2 Fig](#)), we also performed sensitivity analysis using a linear mixed model approach implemented in the lmerTest R package (v2.0–29) to account for sample relatedness as estimated by pedigree. We observed that the association statistics obtained using simple linear regression vs. linear mixed model were highly consistent ([S1 Table](#) and [S3 Fig](#)), and therefore used linear regression in all subsequent analyses. For metabolites that showed nominally significant ($p < 0.05$) association, we tested for their replication in MCDS (although this replication was not required for including metabolites in the MRS as described below). Clustered metabolite correlation heat maps were created using FHS data and the heatmap.2() function in the gplots R package (v3.0.1).

Building the MRS model (S1 A(ii) Fig): To build an optimal multivariate model for predicting weight gain using FHS data, metabolites nominally associated ($p < 0.05$) with Δ BMI in FHS were used as input variables into a stepwise regression procedure implemented in the step() function in R (v3.2.1). In order to determine which direction of variable selection (forward, backward, or bidirectional) and model evaluation metric (Akaike information criterion, AIC or Bayesian information criterion, BIC) to use for running the stepwise procedure, we performed 100 repeated 10-fold cross validations ([S4 Fig](#)). Based on the cross-validation results, the AIC metric combined with a bidirectional selection approach was used to train the final stepwise model using all FHS data. We estimated partial R^2 of each metabolite (i.e. proportion of variance uniquely explained by the metabolite) in the final model by comparing the full model against a reduced model without the metabolite: $partial R^2 = (SSE_{reduced} - SSE_{full}) / SSE_{reduced}$, where $SSE = error\ sum\ of\ squares$.

Calculating the MRS z-scores (S1A(iii) and S1B(i) Fig): We calculated a single MRS value for each subject in either FHS or MCDS as: $\sum \beta_i z_i$, where β_i is the effect size estimate for each metabolite in the FHS stepwise model and z_i is the corresponding metabolite abundance z-score for the sample. Rank-based inverse normal transformation was performed on the MRS values for each cohort and the resulting MRS z-scores were used for all subsequent analyses.

Association analyses for Δ BMI, MRS, and other obesity-related phenotypes

Linear regression was performed to test for association between Δ BMI, MRS (or individual metabolites in the MRS), and other obesity-related variables, including baseline BMI and waist-to-hip ratio (WHR), GRS_{BMI} , smoking status, physical activity, dietary intakes, and glycemic measures ([S1A\(iv\)](#), [S1B\(ii\)](#) and [S1C\(i\) Fig](#)). Logistic regression was performed to associate future T2D status with MRS, baseline anthropometric (BMI and WHR), or glycemic

measures (**S1C(ii) Fig**). When testing for independence of MRS and another variable in terms of association with Δ BMI, we included the variable as a covariate in regression analysis to see whether it would affect the statistical significance of the association observed between Δ BMI and MRS. When applicable, we estimated partial R^2 of each variable (i.e. proportion of variance uniquely explained by the variable) in Δ BMI multivariate models, using the same formula described for the MRS model above.

GWAS for MRS and MRS metabolites

Genome-wide association analyses were performed in both FHS and MCDS to identify genetic loci associated with MRS and with the individual metabolites in the MRS (**S1C(iii) Fig**). To account for sample relatedness, we used the linear mixed model and kinship matrix calculation methods in EPACTS (v3.2.6) [39] to perform the GWAS in each cohort. For MCDS, we also included genotyping platform as a covariate (since genotype imputation was performed separately for each platform). We separately analyzed genetic variants with minor allele count $> = 3$ in FHS or MCDS, and meta-analyzed the two cohorts using the inverse variance weighted method in METAL (2011-03-25 version) [40]. To identify independent loci, we “clumped” the meta-analysis results, grouping together sets of nearby, correlated variants that are within 500kb and have pairwise $r^2 > 0.1$. We used PLINK (v1.9) [41] and 1000 Genomes phase 1 reference panel [42] to generate the clumps of correlated variants and selected the “lead” (best-associated) variant from each clump. We used the Variant Effect Predictor [43] to identify nearest genes (± 5 kb) for the lead variants. For MRS-associated lead variants reaching suggestive significance ($p < 1 \times 10^{-6}$), we also searched the NHGRI-EBI GWAS catalog (v1.0.2, 2019-05-03) [44] for traits previously associated ($p < 5 \times 10^{-8}$) with the variants (± 50 kb).

Results

Metabolites associated with longitudinal change in BMI

We used longitudinal BMI measurements to calculate age- and sex-adjusted change in BMI over time (Δ BMI) for 1,508 samples from FHS (demographic and clinical characteristics shown in **Table 1**). The mean Δ BMI is 0.127 kg/m²/year (standard deviation, SD = 0.320) with mean follow-up interval of 6.9 years. Using baseline metabolite profiling data, we identified 42 metabolites that have nominally significant association ($p < 0.05$) with Δ BMI (**S1 Table**). Increased levels of 35 metabolites are associated with less weight gain (i.e. lower Δ BMI). These include tyrosine, malate, alanine, niacinamide, hydroxyglutarate, glutamic acid, xanthine, quinolinate, lactate, many lipids with relatively lower double bond content (i.e. between 0–3 double bonds; except for 2 triacylglycerides, TAG 56:10 and 58:12). On the other hand, 7 metabolites show the opposite trend, where increased levels are associated with more weight gain (i.e. higher Δ BMI). These include uridine, pyruvate, and 5 lipids with higher double bond content (TAG 56:6, 56:6, 58:7, 58:8, and a cholesteryl ester, CE 20:5). Overall, the Δ BMI-associated metabolites are spread out in different correlation clusters among all tested metabolites, with lipids of similar saturation level (i.e. double bond content) showing especially strong correlation with each other (**Fig 1 and S5 Fig**).

MRS model for weight gain

In order to combine the predictive power of individual metabolites, we performed stepwise regression using the 42 Δ BMI-associated metabolites in FHS to build an MRS model for weight gain. The resulting stepwise model includes eight metabolites that each contributes to the prediction of weight gain: TAG 56:6, malate, niacinamide, sphingomyelin (SM) 24:0, uridine,

Table 1. Descriptive characteristics of the study cohorts.

Characteristics	Framingham Heart Study (FHS)		Mexico City Diabetes Study (MCDS)	
Number of samples	1508		768	
Number of female samples	804 (53.3)		470 (61.2)	
Δ BMI (kg/m ² /yr)	0.127 \pm 0.320		0.000 \pm 0.220	
Characteristics by Time Point	Baseline (Exam 5)	Follow-up (Exam 7)	Baseline (1998 visit)	Follow-up (2008 visit)
Age (yr)	54.1 \pm 9.52	61.0 \pm 9.43	52.2 \pm 7.59	62.4 \pm 7.56
BMI (kg/m ²)	27.4 \pm 4.87	28.3 \pm 5.33	28.5 \pm 4.43	29.5 \pm 4.80
Fasting Glucose (mg/dl)*	94.4 \pm 9.46	97.8 \pm 10.1	90.7 \pm 10.1	99.7 \pm 32.2
T2D	0 (0)	105 (6.96)	0 (0)	144 (18.8)

Showing mean \pm standard deviation for quantitative traits or sample count (%) for binary traits.

* Fasting glucose: excluded T2D cases in both cohorts and missing measurements in MCDS (n = 16).

<https://doi.org/10.1371/journal.pone.0222445.t001>

TAG 56:2, tyrosine, and xanthine (Table 2). These metabolites show relatively weak correlation with each other ($r = -0.097$ to 0.319 , S2 Table and S6 Fig) and represent distinct clusters formed by the 42 Δ BMI-associated metabolites (Fig 1). Using these 8 metabolites in the stepwise model, we calculated an effect size-weighted MRS score for all FHS samples and observed that this MRS is strongly associated with Δ BMI ($\beta = 0.0612$, $p = 1.82 \times 10^{-14}$) and explains 3.83% of the observed Δ BMI variance (S3 Table and S7 Fig). However, this estimate is likely inflated due to overfitting and the winner's curse [45], as indicated by our cross validation results (i.e. MRS models generated using 10-fold cross validations all had $R^2 < 0.383$; S4 Fig).

Replication in an independent cohort

To replicate the associations of individual metabolites and the MRS in an independent cohort, and to obtain unbiased estimates of effect size, we first analyzed the 42 Δ BMI-associated metabolites in 768 samples from MCDS. The mean Δ BMI in MCDS is 0 kg/m²/year (SD = 0.320; mean follow-up interval = 10.7 years; Table 1), which does not differ significantly from that of FHS ($p = 0.0913$ in Welch's two-sample *t*-test). In MCDS, 36 of the 42 (85.7%) metabolites show directionally consistent Δ BMI associations compared to FHS, with 24 (57.1%) also reaching nominal significance, providing strong evidence of replication in an independent cohort drawn from a different population (S1 Table). Next, we calculated the effect size-weighted MRS score for all MCDS samples using the FHS stepwise model and validated that the MRS significantly predicts weight gain in MCDS ($\beta = 0.0298$, $p = 1.63 \times 10^{-4}$; S3 Table and S7 Fig), where 1.84% of Δ BMI variation is explained by the score. The observed effect size estimate indicates that one SD increase in MRS predicts 0.0298 kg/m²/year increase in Δ BMI (i.e. ~ 0.9 kg over 10 years for a 1.7 m adult).

Anthropometric and lifestyle risk factors do not mediate prediction of Δ BMI by the MRS

We investigated the relationships between Δ BMI, MRS, and various obesity-related risk factors, both to gain insights regarding the underlying biology captured by the MRS and to evaluate whether the MRS is independent of these factors. First, we assessed two baseline anthropometric measures: BMI and WHR. Both measures are inversely correlated with Δ BMI in MCDS (but show no association in FHS) and with MRS in both cohorts (S4 Table). However, conditional analyses using BMI or WHR as covariate showed that the association between Δ BMI and MRS remains strongly significant after adjusting for either variable (Fig

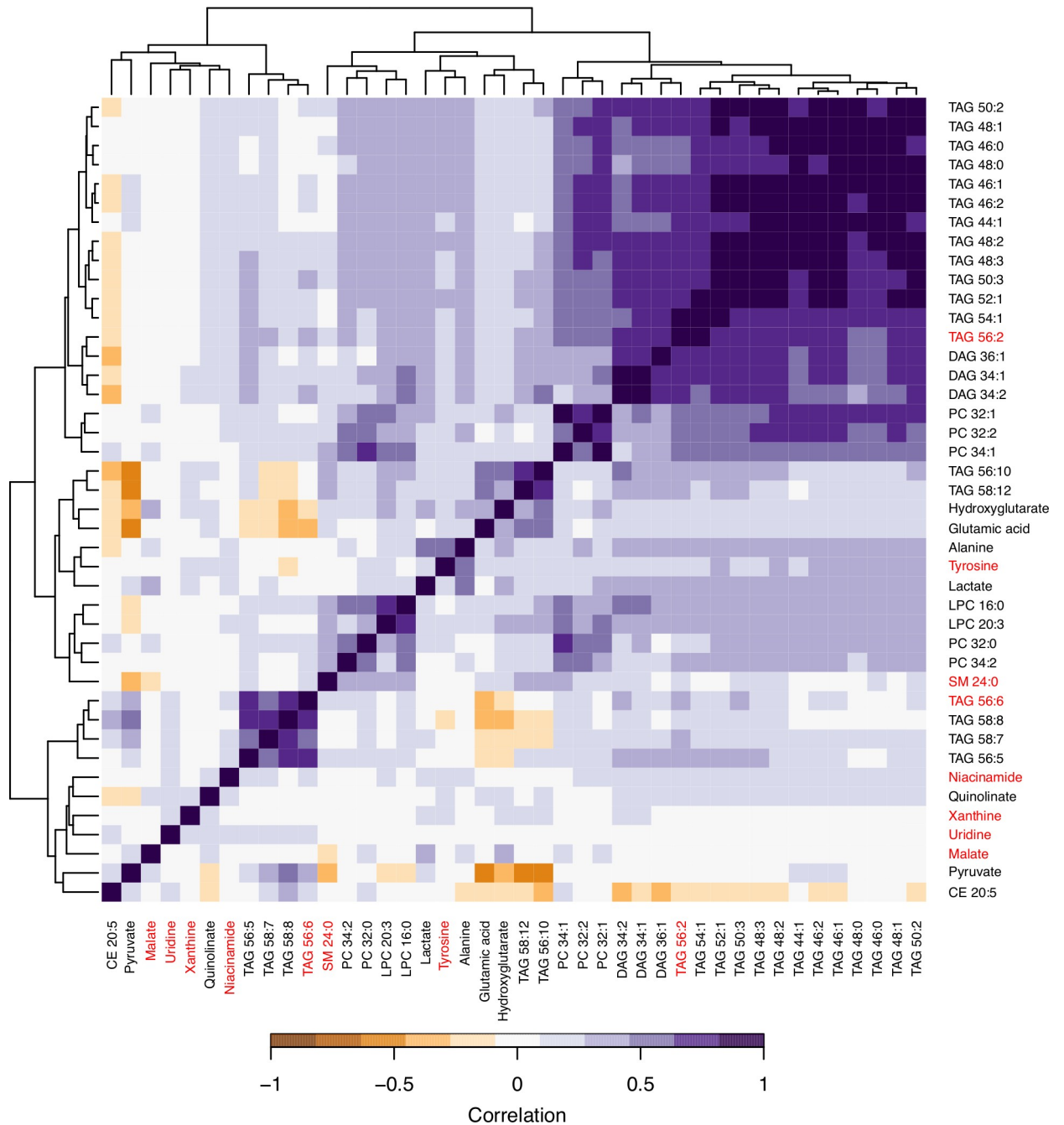


Fig 1. Clustered correlation heat map of the 42 Δ BMI-associated ($p < 0.05$) metabolites in FHS. The 8 metabolites included in the MRS are labeled in red. TAG, triacylglyceride; DAG, diacylglyceride; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; SM, sphingomyelin; CE, cholesteryl ester.

<https://doi.org/10.1371/journal.pone.0222445.g001>

2A and S3 Table). We also asked whether a genetic risk score for cross-sectional BMI (GRS_{BMI} , see Methods) could predict Δ BMI. As expected, GRS_{BMI} is significantly associated with baseline age and sex-adjusted BMI in both cohorts ($p = 2.37 \times 10^{-6}$ in FHS; $p = 5.51 \times 10^{-3}$ in MCDS); the degree of association is similar between FHS and MCDS (i.e. overlapping 95% confidence intervals for β ; S4C Table) despite differences in ancestry.

Table 2. Metabolite risk score model for predicting weight gain in FHS.

Metabolite	β	SE	CI 2.5	CI 97.5	P-value	Partial R ²
TAG 56:6	0.0310	0.0088	0.0138	0.0482	4.17E-04	0.0083
Malate	-0.0241	0.0082	-0.0401	-0.0080	3.27E-03	0.0058
Niacinamide	-0.0226	0.0084	-0.0389	-0.0062	6.98E-03	0.0048
SM 24:0	-0.0209	0.0081	-0.0367	-0.0050	9.88E-03	0.0044
Uridine	0.0211	0.0084	0.0046	0.0375	1.20E-02	0.0042
TAG 56:2	-0.0203	0.0089	-0.0378	-0.0027	2.38E-02	0.0034
Tyrosine	-0.0177	0.0084	-0.0342	-0.0013	3.49E-02	0.0030
Xanthine	-0.0145	0.0082	-0.0306	0.0017	7.88E-02	0.0021

β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β ; Partial R², proportion of variance uniquely explained by the metabolite (see [Methods](#)); TAG, triacylglyceride; SM, sphingomyelin.

<https://doi.org/10.1371/journal.pone.0222445.t002>

However, GRS_{BMI} is not associated with ΔBMI or with MRS ([S4 Table](#)). In summary, none of the anthropometric-related risk factors we tested account for the predictive power of MRS for weight gain, even though the MRS can be associated with baseline BMI and WHR.

We also looked at several lifestyle factors that can influence weight gain: smoking status, physical activity, and dietary intake measures. Quitting smoking is strongly associated with ΔBMI but not with MRS ([S4 Table](#)), and adjusting for this variable did not affect the MRS prediction of ΔBMI ([S3 Table](#)). To further assess confounding due to smoking, we performed a sensitivity analysis by regressing ΔBMI against MRS using only baseline non-smokers in FHS. The association between MRS and ΔBMI in non-smoking samples is highly significant ($n = 1,243$, $\beta = 0.045$, $p = 1.96 \times 10^{-7}$, $R^2 = 0.021$), albeit with a slightly diminished effect size estimate (empirical $p < 0.0061$, calculated using 10,000 random samplings of 1,243 FHS samples) compared to association in all samples ($n = 1,508$, $\beta = 0.055$, $p = 5.80 \times 10^{-12}$, $R^2 = 0.031$). Overall, these results indicate that while smoking is associated with weight gain in our data, it does not explain the observed association between ΔBMI and MRS.

In contrast to smoking status, neither physical activity nor dietary intake variables are significantly correlated with ΔBMI and MRS in our data ([S4 Table](#)). Furthermore, based on results of conditional analyses, none of these variables account for the association between ΔBMI and MRS ([S3 Table](#)). More detailed analysis of the FHS dietary data, however, showed that several MRS metabolites are nominally associated with different types of fat intake (TAG 56:6 with saturated and monounsaturated fats, malate with polyunsaturated fat, uridine with saturated and polyunsaturated fats, and xanthine with saturated fat; [S5 Table](#)), suggesting that these metabolites may be influenced by dietary patterns. In general, our results demonstrate that the MRS is a stronger predictor for weight gain compared to lifestyle factors that measure energy intake (i.e. diet) or expenditure (i.e. physical activity) in our datasets, and the MRS has the additional advantage of not relying on participant recall.

Insulin resistance measures and future T2D outcome are negatively associated with MRS

Next, to better understand the physiological correlates of the information contained within the MRS, we tested whether baseline clinical measures of dysglycemia and insulin resistance are associated with ΔBMI and MRS. In both FHS and MCDS, ΔBMI and MRS are generally negatively correlated with variables that assessed insulin resistance (i.e. fasting/2-hour glucose and insulin, and HOMA-IR) and positively correlated with those that marked insulin sensitivity (i.e. Matsuda ISI and QUICKI) ([S4 Table](#)). Analysis of individual metabolites within the MRS

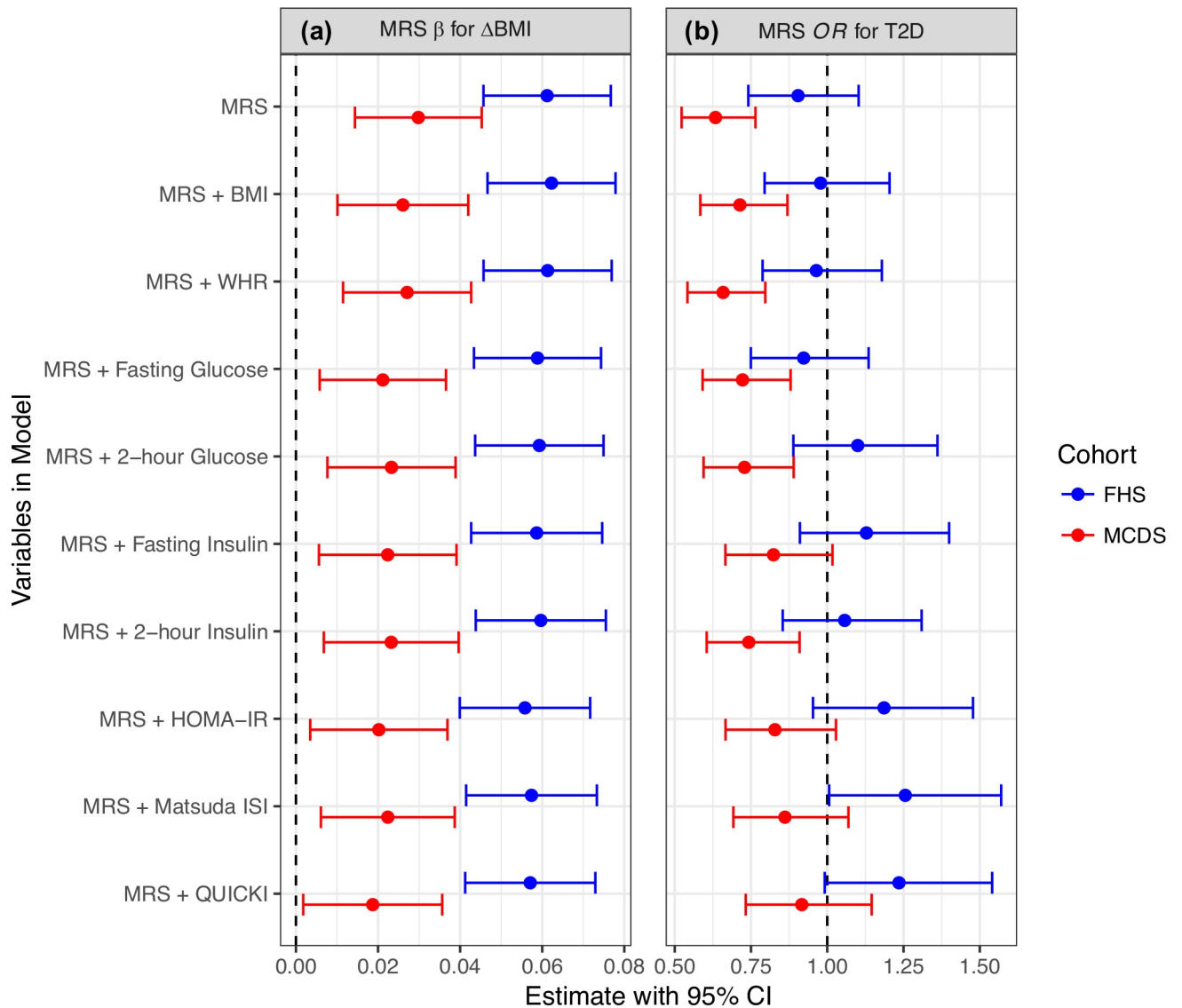


Fig 2. Association between Δ BMI or future T2D status and MRS with adjustment for baseline anthropometric or glycemic measures. (a) Effect size estimates (β) of MRS for Δ BMI. (b) Odds ratios (OR) of MRS for T2D status at follow-up. Variables in Model, which variables were included in the linear (for Δ BMI) or logistic (for T2D) regression model when estimating the effect of MRS. Error bars indicate 95% confidence interval (CI). Vertical dashed black lines indicate null effect ($\beta = 0$ or OR = 1).

<https://doi.org/10.1371/journal.pone.0222445.g002>

indicated that TAG 56:2 and tyrosine are the strongest drivers of the associations between MRS and glycemic response in both cohorts, with the remaining metabolites showing less consistent associations (S6 Table). The percentage of Δ BMI variance explained by the MRS (1.84% in MCDS) is on the same order of magnitude as the percentage explained by the various glycemic traits (1.38 to 5.21%; S4 Table) and BMI GRS. In conditional analyses, the relationship between Δ BMI and MRS remained nominally significant after adjusting for each of the glycemic variables (Fig 2A and S3 Table), implying that the MRS is not solely a proxy for insulin sensitivity.

In both FHS and MCDS, the baseline anthropometric and glycemic measures are, as expected, significantly associated with future incident T2D ($p = 1.93 \times 10^{-29}$ to 1.10×10^{-3} ; S7 Table). Therefore, we tested whether the MRS can be used to predict future T2D using logistic

regression and assessed whether its predictive power is independent from these baseline variables. Unexpectedly, the MRS is negatively associated ($p = 2.57 \times 10^{-6}$) with T2D in MCDS, where one SD increase in MRS translates into 36.6% reduction in risk of developing T2D over a ~10-year period; in FHS, the MRS also has a negative but nonsignificant association with T2D (Fig 2B and S8 Table). However, when we included specific glycemic variables (i.e. fasting insulin, HOMA-IR, Matsuda ISI, and QUICKI) as covariates during conditional analyses, the association between T2D status and MRS in MCDS became nonsignificant ($p > 0.05$; Fig 2B and S8 Table). Thus, measured baseline glycemic variables could account for the observed association between MRS and lower incident T2D, but not continued rise in BMI as shown in Fig 2A.

Genetic loci associated with MRS

Finally, we utilized genetic data to begin to dissect the underlying biology reflected by the MRS. We performed separate GWAS for the MRS in FHS and MCDS and meta-analyzed the association results. We identified 2 genome-wide significant loci ($p < 5 \times 10^{-8}$; Table 3). The most significant locus (rs174565; $p = 5.97 \times 10^{-10}$) lies in the *FADS1/FADS2* gene cluster, both of which encode fatty acid desaturases that introduce double bonds into fatty acids [46]. This locus has been associated with a wide range of traits (S9 Table), including lipid-related metabolites [47,48], fasting glucose [49], liver enzyme levels [50], and heart function traits [51–53]. The second locus (rs192386132; $p = 2.80 \times 10^{-8}$) is near *TACC2*, which encodes a transforming acidic coiled-coil containing protein that has been linked to cancer [54]; no previous GWAS associations have been reported near (± 50 kb) this locus. We also identified five additional loci at suggestive significance ($p < 1 \times 10^{-6}$; Table 3), 2 of which overlap with BMI loci reported in the GWAS catalog (rs12325540 and rs146167165; S9 Table).

Since the MRS is a composite phenotype generated using 8 individual metabolites, we also performed GWAS separately for each of the metabolites and identified a total of 37 loci reaching at least suggestive significance (S10 Table). Only 3 of these loci are genome-wide significant (2 for TAG 56:6 and 1 for TAG 56:2), one of which overlaps with the *FAD1/FADS2* locus for MRS. Interestingly, the *TACC2* locus, which is genome-wide significant for MRS, does not show up as significant for any of the MRS metabolites. Together, these findings suggest that while the MRS can reflect biology driven by individual metabolites in the MRS (e.g. TAG 56:6), it might also be capturing more complex biological patterns that are not specific to a single metabolite.

Table 3. Genetic loci reaching suggestive significance ($p < 1 \times 10^{-6}$) in MRS GWAS meta-analysis.

Variant ID	Chr	Position	EA	OA	EAF	β	SE	P-value*	R ²	Nearest Gene (± 5 kb)	In GWAS Catalog
rs174565	11	61591636	C	G	0.701	0.267	0.043	5.97E-10	0.030	<i>FADS1, FADS2</i>	Yes
rs192386132	10	123937855	A	G	0.003	1.981	0.357	2.80E-08	0.025	<i>TACC2</i>	No
rs12325540	16	73380363	C	T	0.788	0.236	0.044	8.63E-08	0.019	-	Yes
rs72848293	17	72314849	A	G	0.052	0.444	0.087	3.17E-07	0.019	<i>DNAI2</i>	No
rs113940640	9	37228390	A	T	0.023	0.657	0.130	4.33E-07	0.019	<i>ZCCHC7</i>	Yes
rs146167165	6	163054293	T	C	0.013	0.986	0.197	5.26E-07	0.024	<i>PARK2</i>	Yes
rs188635171	12	7378287	A	G	0.009	1.087	0.219	7.07E-07	0.020	-	Yes

* Genome-wide significant p -values are in bold. Chr, chromosome; Position, hg19 genomic position; EA, effect allele (i.e. MRS-increasing allele); OA, other allele; EAF, effect allele frequency; β , effect size estimate; SE, standard error of β ; R², proportion of variance explained, estimated as: $\beta^2 \times 2f(1-f)$, where $f = \text{EAF}$; In GWAS Catalog, whether the locus (± 50 kb) has been linked to traits in the GWAS catalog (see S9 Table for list of associated traits).

<https://doi.org/10.1371/journal.pone.0222445.t003>

Discussion

In this study, we defined a metabolite risk score (MRS) incorporating eight individual metabolites, which predicts change in BMI over a span of 7–10 years. This MRS was constructed and validated using two independent cohorts with demographic differences (e.g. geography, ancestry, lifestyle, and dietary patterns), indicating the generalizability of the score. Based on our results, we estimated that adults of average height (~1.6 m for women to ~1.8 m for men) with high MRS (+1 SD above mean) would gain about 0.15 to 0.19 kg more than those with low MRS (-1 SD below mean) per year. Given that the average person gains 0.4–1.0 kg/year in adulthood, identifying individuals with high-risk MRS could be useful on both the individual and population levels, especially if additional metabolites are identified to improve predictive power. Importantly, none of the routinely measured anthropometric, lifestyle, and glycemic risk factors we tested fully account for the association between Δ BMI and MRS. The proportion of Δ BMI variance explained by the MRS is comparable to that for the best predictive risk factors (i.e. glycemic measures). Therefore, the MRS can be used as an independent predictor for studying weight gain.

The MRS consists of eight metabolites, including three lipids (TAG 56:6, SM24:0, and TAG 56:2), malate, niacinamide, uridine, tyrosine, and xanthine, and may point to independent biological processes related to weight gain. For the lipids, we observed that TAG 56:6 is positively correlated with Δ BMI, while SM 24:0 and TAG 56:2, which have lower double bond content, show the opposite trend. This pattern is in accordance with previous studies that found lipids with varying saturation levels to have opposite association with obesity-related phenotypes. For instance, lipids of lower vs. higher double bond content were previously associated with increased vs. decreased risk of diabetes in FHS, respectively [25]. Here, we observed that TAG 56:2 and SM 24:0 are linked to decreased weight gain in both FHS and MCDS and increased T2D risk in MCDS (i.e. they have negative effect sizes in the MRS that is protective against diabetes), while TAG 56:6 has the opposite direction of effect. Among the non-lipid metabolites, xanthine has been previously identified as a predictor for significant future weight gain [30]; and uridine is closely related to both xanthine and urate [55], which was another predictor identified in the same study. Niacinamide is involved in nicotinamide adenine dinucleotide biosynthesis and has been connected to obesity and weight regulation through its role in energy metabolism [56,57]. Tyrosine has been associated with past weight gain, insulin resistance, dyslipidemia, and T2D [24,26,29,58]. Finally, though malate has not been linked directly to weight gain, it is involved in several metabolic pathways that are essential for energy metabolism, including the TCA cycle, pyruvate/malate-cycle, and malate/aspartate shuttle [59,60]. Further investigation is necessary to test whether these MRS metabolites exert causal influence on weight gain.

When we investigated whether the MRS might be tagging specific aspects of obesity-related physiology, we found it to be a strong marker of preserved insulin sensitivity. This finding is consistent with the notion that insulin effect can drive obesity and agrees with previous studies that found insulin to be associated with increased weight gain [61,62]. The unanticipated negative correlations between the MRS and baseline insulin resistance or future T2D in our data have important implications for interpreting our results: historically, excess weight (or BMI) has been used as a proxy for metabolic dysfunction associated with disease outcomes, under the assumption that higher weight gain is associated with higher risk of developing diseases; however, part of our MRS actually captures a “healthy” component of weight gain with respect to diabetes risk and identified weight-gainers who had more robust insulin secretion. Thus, for clinical disease prevention efforts related to T2D, it might be more fruitful to target individuals with low MRS who are less likely to gain weight but are more metabolically unhealthy with

respect to T2D. Future investigation is needed to test if this type of inverse association also exists between the MRS and other obesity-related diseases such as CVD, liver disease, cancer and mortality. In any case, we note that the MRS-T2D association we observed loses statistical significance after adjustment for some of the glycemic measures, indicating that the MRS and these measures are capturing the same insulin effect signature responsible for T2D prediction. In contrast, the MRS- Δ BMI association appears to be partially independent from insulin effect, supporting the MRS as a composite marker encompassing both known (insulin-related) and potentially novel aspects of obesogenic physiology.

Some limitations of this report should be recognized. First, the discovery and replication cohorts differ on risk factors such as ethnicity and lifestyle, thus the participants may have important biological differences (e.g. Δ BMI is not correlated with baseline BMI or WHR in FHS, but shows significant association in MCDS). While we were able to replicate the MRS despite these caveats, future replication in additional cohorts could yield a better estimate of the predictive power of MRS on weight gain across different settings. Next, by excluding samples with metabolic diseases (T2D, CVD, and renal disease) at baseline, we may have introduced selection bias, whereby subsetting data using a variable caused by two independent variables induces an association between them. Nevertheless, since these diseases are known to cause drastic changes in the metabolome, it was prudent to exclude them to reduce the risk of capturing a skewed profile. In addition, a similar bias may be caused by the fact that we needed to subtract out (i.e. “condition on”) baseline BMI in order to calculate Δ BMI, therefore potentially inducing associations between Δ BMI or MRS and other variables correlated with baseline BMI. This bias may contribute to our seeming paradoxical observation that the MRS is correlated with glycemic measures and negatively predictive of T2D. However, we note that (1) the MRS is more strongly associated with several glycemic variables compared to with baseline BMI ([S4 Table](#)) and (2) the MRS is independent from baseline BMI in conditional analysis to predict T2D in MCDS ([S8 Table](#)). These results provide some evidence that the association between the MRS and insulin response/T2D is not entirely a statistical artifact induced by our definition of Δ BMI. Finally, while we observed little correlation between Δ BMI or MRS and lifestyle factors that measure energy intake and expenditure in this report, lifestyle data were not available at every visit for both cohorts, and our investigations were also limited by the methods used to assess lifestyle in these cohorts. Therefore, future investigation is required to determine if the lack of correlation is a reflection of limited sample size and/or data availability. Additional data types (e.g. dietary quality measures or calorimetry-derived metabolic rate) may also help us assess if and how the MRS is linking imbalance in energy metabolism to weight gain.

In conclusion, we have systematically constructed an MRS that is a generalizable weight gain predictor and is independent from traditional risk factors. Interestingly, we found the MRS to be a strong marker for insulin sensitivity (even though they independently contribute to weight gain prediction) and showed that individuals with higher MRS have lower risk of developing T2D in one cohort, indicating that it may be more clinically important to track individuals with lower MRS who are less likely to gain weight but more metabolically unhealthy with respect to T2D. We were not able to test association of MRS with other obesity-associated outcomes, such as cancer, which may be related to unchecked weight gain and anabolic effects of insulin. Overall, our study showcases the power of leveraging metabolite profiling to elucidate and identify the role of intrinsic and extrinsic factors on weight change. Further studies of metabolite profiling and weight gain are required to replicate and expand these findings in independent cohorts. For instance, untargeted profiling methods may be used to discover many more unknown metabolites predictive of weight gain, which can greatly improve the power of the MRS model. Another critical future direction is to leverage genetic

data to use Mendelian randomization approaches to establish if and how the MRS is capturing the causal biology of weight gain.

Supporting information

S1 Fig. Schematic overview for constructing, validating, and understanding the MRS for weight gain. The MRS was constructed using FHS data and tested for replication in MCDS. After validating that the MRS was associated with Δ BMI in both cohorts, we used phenotype and genetic data from both cohorts to study the relationships between MRS and various obesity-related risk factors, future T2D outcome, and genetic variants.
(PDF)

S2 Fig. Sample relatedness in FHS. (a) Distribution of pedigree size (left) and number of individuals in pedigrees of different sizes ($n = 1,508$; right). (b) Distribution of identity by descent measure (PI_HAT) for all pairs of individuals with genetic data ($n = 1,349$; left) or subset of pairs with $PI_HAT > 0.05$ (right). PI_HAT was calculated using 92,210 independent genetic markers using PLINK (v1.9).
(PDF)

S3 Fig. Comparison of Δ BMI-metabolite association statistics derived using simple linear model (LM) vs. linear mixed model (LMM) in FHS. Effect size estimates (a) and negative $\log_{10} p$ -value (b) of the 183 metabolites calculated using each approach are plotted. R^2 , squared correlation between the LM and LMM statistics.
(PDF)

S4 Fig. Summary of cross validation results for the stepwise regression procedure in FHS. 100 repeated 10-fold cross validations were performed to determine the optimal direction of variable selection (DIRECTION: forward, backward, or both/bidirectional) and model evaluation metric (METRIC: AIC or BIC) for building a stepwise MRS model for predicting Δ BMI. R^2 (a) and root-mean-squared error (b) statistics across the 100 cross validations are plotted as notched boxplots, with the notches indicating 95% confidence interval of the median.
(PDF)

S5 Fig. Clustered correlation heat map of all 183 metabolites analyzed in FHS. The 8 MRS metabolites and the other 34 Δ BMI-associated ($p < 0.05$) metabolites are labeled in red and blue, respectively. CE, Cholesteryl esters; DAG, diacylglyceride; F1P/F6P/G1P/G6P, fructose-1-phosphate/fructose-6-phosphate/glucose-1-phosphate/glucose-6-phosphate; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; PC, phosphatidylcholine; SM, sphingomyelin; TAG, triacylglyceride.
(PDF)

S6 Fig. Density histogram of pairwise correlation between the 42 Δ BMI-associated metabolites (“Nominal”) vs. the subset of 8 metabolites included in the MRS (“MRS”) in FHS. Dashed vertical lines indicate mean correlation of each group.
(PDF)

S7 Fig. Δ BMI vs. MRS for FHS (red) and MCDS (blue) samples. Linear regression lines (with shaded 95% confidence regions) and corresponding statistics (bottom right text labels) are shown.
(PDF)

S1 Table. FHS and MCDS association statistics for Δ BMI-associated ($p < 0.05$) metabolites in FHS. * As a sensitivity analysis in FHS, LMM (linear mixed model) regression was

performed to account for sample relatedness (see [Methods](#)); ** Nominally significant p -values are in bold. β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β ; LMM R^2 (i.e. proportion of variance explained) estimated as: $\beta^{2*} \text{var}(\text{Metabolite})/\text{var}(\Delta\text{BMI})$; Directional Consistency, whether the directions of effect in FHS and MCDS agree with each other; TAG, triacylglyceride; SM, sphingomyelin; PC, phosphatidylcholine; DAG, diacylglyceride; LPC, lysophosphatidylcholine; CE, cholesteryl ester.
(XLSX)

S2 Table. Correlation and correlation p -value between the 8 MRS metabolites in FHS. Correlation (red) is shown in upper left triangle and correlation p -value (blue) is shown in lower right triangle. Metabolites are tabulated in order to match [Fig 1](#) heat map. Nominally significant p -values are in bold.

(XLSX)

S3 Table. Association between ΔBMI and MRS with adjustment for obesity-related risk factors. * Association Model: $Y \sim X$ indicates linear regression with Y as dependent variable and X as independent variable(s). ** Nominally significant p -values are in bold. N , number of samples; β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β ; R^2 , proportion of variance explained, partial R^2 is shown for models with multiple independent variables (see [Methods](#)).

(XLSX)

S4 Table. Association between ΔBMI or MRS and obesity-related risk factors. * All risk factors were assessed at or prior to baseline, except for "Quit Smoking" and " ΔPA " (see [Methods](#)). ** Nominally significant p -values are in bold. Greyed out rows indicate that data were not available. N , number of samples; β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β .

(XLSX)

S5 Table. Association between MRS or MRS metabolites and dietary fats in FHS. * Nominally significant p -values are in bold. $N = 1387$ for all analyses. β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β ; TAG, triacylglyceride; SM, sphingomyelin.

(XLSX)

S6 Table. Association between MRS metabolites and baseline anthropometric or glycemic measures. * Nominally significant p -values are in bold. N , number of samples; β , effect size estimate; SE, standard error of β ; CI 2.5—CI 97.5, 95% confidence interval for β ; TAG, triacylglyceride; SM, sphingomyelin.

(XLSX)

S7 Table. Association between future T2D status and baseline anthropometric or glycemic risk factors. * Nominally significant p -values are in bold. N total, number of samples; N cases, number of T2D cases (by follow-up); OR, odds ratio; CI 2.5—CI 97.5, 95% confidence interval for OR.

(XLSX)

S8 Table. Association between future T2D status and MRS with adjustment for baseline anthropometric or glycemic measures. * Association Model: $Y \sim X$ indicates logistic regression with Y as dependent variable and X as independent variable(s). ** Nominally significant p -values are in bold. N total, number of samples; N cases, number of T2D cases (by follow-up); OR, odds ratio; CI 2.5—CI 97.5, 95% confidence interval for OR.

(XLSX)

S9 Table. List of GWAS catalog associations ($p < 5E-8$) that overlap ($\pm 50\text{kb}$) with suggestive MRS lead variants. PMID, PubMed ID of study that reported the association; Chr, chromosome; Position_hg38, hg38 genomic position; β or OR, effect size or odds ratio estimate for the effect allele.

(XLSX)

S10 Table. Genetic loci reaching suggestive significance ($p < 1E-6$) in MRS metabolite GWAS meta-analysis. * Genome-wide significant p -values are in bold. ** Loci boundaries were defined by linkage disequilibrium proxies ($\leq 500\text{kb}$ and $r^2 > 0.1$) with minimum and maximum chromosomal positions. Chr, chromosome; Position, hg19 genomic position; EA, effect allele (i.e. metabolite-increasing allele); OA, other allele; EAF, effect allele frequency; β , effect size estimate; SE, standard error of β ; R^2 , proportion of variance explained, estimated as: $\beta^{2*}2f(1-f)$, where $f = \text{EAF}$; TAG, triacylglyceride; SM, sphingomyelin.

(XLSX)

S1 Text. Details of FHS genetic data.

(DOCX)

S2 Text. Details of MCDS genetic data.

(DOCX)

S3 Text. Missing value imputation for metabolite data.

(DOCX)

Acknowledgments

We thank the Broad Metabolomics Platform and SIGMA T2D Consortium for sharing data resources. This work was supported by the National Heart, Lung, and Blood Institute grants F31HL126581 (Y.H.H.) and R00HL122515 (R.M.S), National Institute of Diabetes and Digestive and Kidney Diseases grants T32DK110919 (Y.H.H.), K12DK094721 (C.M.A), and R01DK075787 (J.N.H.), Endocrine Scholars Award (C.M.A.), and Doris Duke Charitable Foundation (<http://www.ddcf.org>) grant 215205 (J.N.H.). Genome-wide genotyping and metabolite profiling of the MCDS cohort was performed as part of the Slim Initiative in Genomic Medicine for the Americas (SIGMA) T2D Consortium.

Author Contributions

Conceptualization: Nina Geidenstam, Yu-Han H. Hsu, Christina M. Astley, Joel N. Hirschhorn, Rany M. Salem.

Data curation: Christina M. Astley, Josep M. Mercader, Maria E. Gonzalez, Clicerio Gonzalez.

Formal analysis: Nina Geidenstam, Yu-Han H. Hsu, Rany M. Salem.

Investigation: Nina Geidenstam, Yu-Han H. Hsu, Martin Ridderstråle, Maria E. Gonzalez, Clicerio Gonzalez, Joel N. Hirschhorn, Rany M. Salem.

Methodology: Yu-Han H. Hsu, Christina M. Astley, Josep M. Mercader, Joel N. Hirschhorn, Rany M. Salem.

Resources: Josep M. Mercader, Martin Ridderstråle, Maria E. Gonzalez, Clicerio Gonzalez.

Supervision: Christina M. Astley, Joel N. Hirschhorn, Rany M. Salem.

Validation: Nina Geidenstam, Yu-Han H. Hsu.

Visualization: Nina Geidenstam, Yu-Han H. Hsu, Rany M. Salem.

Writing – original draft: Nina Geidenstam, Yu-Han H. Hsu, Christina M. Astley, Rany M. Salem.

Writing – review & editing: Christina M. Astley, Josep M. Mercader, Martin Ridderstråle, Maria E. Gonzalez, Clicerio Gonzalez, Joel N. Hirschhorn, Rany M. Salem.

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014; 384: 766–781. [https://doi.org/10.1016/S0140-6736\(14\)60460-8](https://doi.org/10.1016/S0140-6736(14)60460-8) PMID: 24880830
2. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*. 2006; 444: 840–846. <https://doi.org/10.1038/nature05482> PMID: 17167471
3. Mandviwala T, Khalid U, Deswal A. Obesity and Cardiovascular Disease: a Risk Factor or a Risk Marker? *Curr Atheroscler Rep*. 2016; 18: 21. <https://doi.org/10.1007/s11883-016-0575-4> PMID: 26973130
4. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology*. 2010; 51: 679–689. <https://doi.org/10.1002/hep.23280> PMID: 20041406
5. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*. 2013; 309: 71–82. <https://doi.org/10.1001/jama.2012.113905> PMID: 23280227
6. Williamson DF. Descriptive epidemiology of body weight and weight change in U.S. adults. *Ann Intern Med*. 1993; 119: 646–649. Available: <https://www.ncbi.nlm.nih.gov/pubmed/8363190> https://doi.org/10.7326/0003-4819-119-7_part_2-199310011-00004 PMID: 8363190
7. Lewis CE, Jacobs Jr. DR, McCreath H, Kiefe CI, Schreiner PJ, Smith DE, et al. Weight gain continues in the 1990s: 10-year trends in weight and overweight from the CARDIA study. *Coronary Artery Risk Development in Young Adults*. *Am J Epidemiol*. 2000; 151: 1172–1181. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10905529> <https://doi.org/10.1093/oxfordjournals.aje.a010167> PMID: 10905529
8. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. *N Engl J Med*. 2011; 364: 2392–2404. <https://doi.org/10.1056/NEJMoa1014296> PMID: 21696306
9. Hall KD, Heymsfield SB, Kemnitz JW, Klein S, Schoeller DA, Speakman JR. Energy balance and its components: implications for body weight regulation. *Am J Clin Nutr*. 2012; 95: 989–994. <https://doi.org/10.3945/ajcn.112.036350> PMID: 22434603
10. Kimokoti RW, Newby PK, Gona P, Zhu L, Jajuja GK, Pencina MJ, et al. Diet quality, physical activity, smoking status, and weight fluctuation are associated with weight change in women and men. *J Nutr*. 2010; 140: 1287–1293. <https://doi.org/10.3945/jn.109.120808> PMID: 20484553
11. Chaput JP, Tremblay A, Rimm EB, Bouchard C, Ludwig DS. A novel interaction between dietary composition and insulin secretion: effects on weight gain in the Quebec Family Study. *Am J Clin Nutr*. 2008; 87: 303–309. Available: <https://www.ncbi.nlm.nih.gov/pubmed/18258618> <https://doi.org/10.1093/ajcn/87.2.303> PMID: 18258618
12. Astley CM, Todd JN, Salem RM, Vedantam S, Ebbeling CB, Huang PL, et al. Genetic Evidence That Carbohydrate-Stimulated Insulin Secretion Leads to Obesity. *Clin Chem*. 2018; 64: 192–200. <https://doi.org/10.1373/clinchem.2017.280727> PMID: 29295838
13. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518: 197–206. <https://doi.org/10.1038/nature14177> PMID: 25673413
14. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum Mol Genet*. 2018; <https://doi.org/10.1093/hmg/ddy271> PMID: 30124842
15. Rukh G, Ahmad S, Ericson U, Hindy G, Stocks T, Renstrom F, et al. Inverse relationship between a genetic risk score of 31 BMI loci and weight change before and after reaching middle age. *Int J Obes*. 2016; 40: 252–259. <https://doi.org/10.1038/ijo.2015.180> PMID: 26374450
16. Steffen A, Sørensen TIA, Knüppel S, Travier N, Sánchez MJ, Huerta JM, et al. Development and Validation of a Risk Score Predicting Substantial Weight Gain over 5 Years in Middle-Aged European Men and Women. *PLoS One*. 2013; <https://doi.org/10.1371/journal.pone.0067429> PMID: 23874419

17. Bachlechner U, Boeing H, Haftenberger M, Schienkiewitz A, Scheidt-Nave C, Vogt S, et al. Predicting risk of substantial weight gain in German adults—A multi-center cohort approach. *Eur J Public Health*. 2017; <https://doi.org/10.1093/eurpub/ckw216> PMID: 28013243
18. Douketis JD, Macie C, Thabane L, Williamson DF. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. *Int J Obes*. 2005; 29: 1153–1167. <https://doi.org/10.1038/sj.ijo.0802982> PMID: 15997250
19. Dombrowski SU, Knittle K, Avenell A, Araujo-Soares V, Snihotta FF. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. *BMJ*. 2014; 348: g2646. <https://doi.org/10.1136/bmj.g2646> PMID: 25134100
20. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009; 9: 311–326. <https://doi.org/10.1016/j.cmet.2009.02.002> PMID: 19356713
21. Wurtz P, Wang Q, Kangas AJ, Richmond RC, Skarp J, Tiainen M, et al. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med*. 2014; 11: e1001765. <https://doi.org/10.1371/journal.pmed.1001765> PMID: 25490400
22. Ho JE, Larson MG, Ghorbani A, Cheng S, Chen MH, Keyes M, et al. Metabolomic Profiles of Body Mass Index in the Framingham Heart Study Reveal Distinct Cardiometabolic Phenotypes. *PLoS One*. 2016; 11: e0148361. <https://doi.org/10.1371/journal.pone.0148361> PMID: 26863521
23. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation*. 2012; 125: 2222–2231. <https://doi.org/10.1161/CIRCULATIONAHA.111.067827> PMID: 22496159
24. Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011; 17: 448–453. <https://doi.org/10.1038/nm.2307> PMID: 21423183
25. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest*. 2011; 121: 1402–1411. <https://doi.org/10.1172/JCI44442> PMID: 21403394
26. Wurtz P, Soininen P, Kangas AJ, Ronnema T, Lehtimäki T, Kahonen M, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care*. 2013; 36: 648–655. <https://doi.org/10.2337/dc12-0895> PMID: 23129134
27. Geidenstam N, Al-Majdoub M, Ekman M, Spégel P, Ridderstråle M. Metabolite profiling of obese individuals before and after a one year weight loss program. *Int J Obes*. Nature Publishing Group; 2017; 41: 1369–1378. <https://doi.org/10.1038/ijo.2017.124> PMID: 28529327
28. Wahl S, Vogt S, Stuckler F, Krumsiek J, Bartel J, Kacprowski T, et al. Multi-omic signature of body weight change: results from a population-based cohort study. *BMC Med*. 2015; 13: 48. <https://doi.org/10.1186/s12916-015-0282-y> PMID: 25857605
29. Menni C, Migaud M, Kastenmuller G, Pallister T, Zierer J, Peters A, et al. Metabolomic Profiling of Long-Term Weight Change: Role of Oxidative Stress and Urate Levels in Weight Gain. *Obes (Silver Spring)*. 2017; 25: 1618–1624. <https://doi.org/10.1002/oby.21922> PMID: 28758372
30. Zhao H, Shen J, Djukovic D, Daniel-MacDougall C, Gu H, Wu X, et al. Metabolomics-identified metabolites associated with body mass index and prospective weight gain among Mexican American women. *Obes Sci Pr*. 2016; 2: 309–317. <https://doi.org/10.1002/osp4.63> PMID: 27708848
31. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979; 110: 281–290. Available: <http://www.ncbi.nlm.nih.gov/pubmed/474565> <https://doi.org/10.1093/oxfordjournals.aje.a112813> PMID: 474565
32. Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab*. Elsevier Inc.; 2013; 18: 130–143. <https://doi.org/10.1016/j.cmet.2013.06.013> PMID: 23823483
33. Williams Amy AL, Jacobs Suzanne SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, Márquez-Luna C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014; 506: 97–101. <https://doi.org/10.1038/nature12828> PMID: 24390345
34. Hsu YHH, Churchhouse C, Pers TH, Mercader JM, Metspalu A, Fischer K, et al. PAIRUP-MS: Pathway analysis and imputation to relate unknowns in profiles from mass spectrometry-based metabolite data. *PLoS Comput Biol*. 2019; <https://doi.org/10.1371/journal.pcbi.1006734> PMID: 30640898
35. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011; 45: 1–67. <https://doi.org/10.18637/jss.v045.i03>
36. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin

- concentrations in man. *Diabetologia*. 1985; 28: 412–419. Available: <https://www.ncbi.nlm.nih.gov/pubmed/3899825> <https://doi.org/10.1007/bf00280883> PMID: 3899825
37. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care*. 1999; 22: 1462–1470. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10480510> <https://doi.org/10.2337/diacare.22.9.1462> PMID: 10480510
 38. Katz A, Nambi SS, Mather K, Baron AD, Follmann DA, Sullivan G, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J Clin Endocrinol Metab*. 2000; 85: 2402–2410. <https://doi.org/10.1210/jcem.85.7.6661> PMID: 10902785
 39. Kang HM. EPACTS (Efficient and Parallelizable Association Container Toolbox). Available: <http://genome.sph.umich.edu/wiki/EPACTS>
 40. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26: 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340> PMID: 20616382
 41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4: 7. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852
 42. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012/11/07. 2012; 491: 56–65. <https://doi.org/10.1038/nature11632> PMID: 23128226
 43. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016; 17: 122. <https://doi.org/10.1186/s13059-016-0974-4> PMID: 27268795
 44. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017; 45: D896–D901. <https://doi.org/10.1093/nar/gkw1133> PMID: 27899670
 45. Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav Ecol Sociobiol*. 2011; 65: 47–55. <https://doi.org/10.1007/s00265-010-1038-5> PMID: 21297852
 46. Marquardt A, Stohr H, White K, Weber B+H. cDNA cloning, genomic structure, and chromosomal localization of three members of the human fatty acid desaturase family. *Genomics*. 2000; 66: 175–183. <https://doi.org/10.1006/geno.2000.6196> PMID: 10860662
 47. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010/08/06. 2010; 466: 707–713. <https://doi.org/10.1038/nature09270> PMID: 20686565
 48. Mozaffarian D, Kabagambe EK, Johnson CO, Lemaitre RN, Manichaikul A, Sun Q, et al. Genetic loci associated with circulating phospholipid trans fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *Am J Clin Nutr*. 2015; 101: 398–406. <https://doi.org/10.3945/ajcn.114.094557> PMID: 25646338
 49. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010; 42: 105–116. <https://doi.org/10.1038/ng.520> PMID: 20081858
 50. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011; 43: 1131–1138. <https://doi.org/10.1038/ng.970> PMID: 22001757
 51. den Hoed M, Eijgelsheim M, Esko T, Brundel BJ, Peal DS, Evans DM, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet*. 2013; 45: 621–631. <https://doi.org/10.1038/ng.2610> PMID: 23583979
 52. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet*. 2014; 46: 826–836. <https://doi.org/10.1038/ng.3014> PMID: 24952745
 53. Verweij N, Mateo Leach I, van den Boogaard M, van Veldhuisen DJ, Christoffels VM, LifeLines Cohort S, et al. Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet*. 2014; 7: 475–481. <https://doi.org/10.1161/CIRCGENETICS.113.000373> PMID: 24850809
 54. Lauffart B, Gangisetty O, Still IH. Molecular cloning, genomic structure and interactions of the putative breast tumor suppressor TACC2. *Genomics*. 2003; [https://doi.org/10.1016/S0888-7543\(02\)00039-3](https://doi.org/10.1016/S0888-7543(02)00039-3)
 55. Albrecht E, Waldenberger M, Krumsiek J, Evans AM, Jeratsch U, Breier M, et al. Metabolite profiling reveals new insights into the regulation of serum urate in humans. *Metabolomics*. 2014; <https://doi.org/10.1007/s11306-013-0565-2> PMID: 24482632

56. Kraus D, Yang Q, Kong D, Banks AS, Zhang L, Rodgers JT, et al. Nicotinamide N-methyltransferase knockdown protects against diet-induced obesity. *Nature*. 2014; 508: 258–262. <https://doi.org/10.1038/nature13198> PMID: 24717514
57. Yang SJ, Choi JM, Kim L, Park SE, Rhee EJ, Lee WY, et al. Nicotinamide improves glucose metabolism and affects the hepatic NAD-sirtuin pathway in a rodent model of obesity and type 2 diabetes. *J Nutr Biochem*. 2014; 25: 66–72. <https://doi.org/10.1016/j.jnutbio.2013.09.004> PMID: 24314867
58. Yamakado M, Nagao K, Imaizumi A, Tani M, Toda A, Tanaka T, et al. Plasma Free Amino Acid Profiles Predict Four-Year Risk of Developing Diabetes, Metabolic Syndrome, Dyslipidemia, and Hypertension in Japanese Population. *Sci Rep*. 2015; 5: 11918. <https://doi.org/10.1038/srep11918> PMID: 26156880
59. Jitrapakdee S, Wutthisathapornchai A, Wallace JC, MacDonald MJ. Regulation of insulin secretion: role of mitochondrial signalling. *Diabetologia*. 2010; 53: 1019–1032. <https://doi.org/10.1007/s00125-010-1685-0> PMID: 20225132
60. Jensen M V, Joseph JW, Ronnebaum SM, Burgess SC, Sherry AD, Newgard CB. Metabolic cycling in control of glucose-stimulated insulin secretion. *Am J Physiol Endocrinol Metab*. 2008; 295: E1287–97. <https://doi.org/10.1152/ajpendo.90604.2008> PMID: 18728221
61. Swinburn BA, Nyomba BL, Saad MF, Zurlo F, Raz I, Knowler WC, et al. Insulin resistance associated with lower rates of weight gain in Pima Indians. *J Clin Invest*. 1991; 88: 168–173. <https://doi.org/10.1172/JCI115274> PMID: 2056116
62. Wedick NM, Snijder MB, Dekker JM, Heine RJ, Stehouwer CD, Nijpels G, et al. Prospective investigation of metabolic characteristics in relation to weight gain in older adults: the Hoorn Study. *Obes (Silver Spring)*. 2009; 17: 1609–1614. <https://doi.org/10.1038/oby.2008.666> PMID: 19197256