Supplementary Text 1. List of Supplementary Figures, Tables, and Text

Supplementary Figures:

Supplementary Figure 1. PAIRUP-MS pathway annotation and analysis framework.

Supplementary Figure 2. Scheme for generating null metabolite groups used in pathway analyses.

Supplementary Figure 3. Clustered correlation heat map of 577 BMI-associated metabolites in OE and MCDS.

Supplementary Figure 4. Level of association between the 324 G_M SNPs and the 324 BMI-associated metabolites in OE and MCDS.

Supplementary Tables:

Supplementary Table 1. Association summary statistics for 324 BMI-associated metabolites with independent genetic instruments (G_M).

Supplementary Table 2. UKB BMI GWAS summary statistics for the 97 BMI SNPs identified in GIANT consortium.

Supplementary Table 3. Association between BMI or the 324 BMI-associated metabolites and G_B in OE, MCDS, and their meta-analysis.

Supplementary Table 4. HMDB m/z query results for matched unknown metabolites in cause, effect, or bidirectional group.

Supplementary Table 5. Known metabolites in cause, effect, or bidirectional group with genome-wide significant published instruments.

Supplementary Table 6. MR-PRESSO global test results using BMI as exposure and each of the 324 BMI-associated metabolites as outcome.

Supplementary Table 7. Nominally enriched ($p < 0.05$) metabolite sets in pathway analyses.

Supplementary Text:

Supplementary Text 2. Details of BMI GWAS in UKB.

Supplementary Text 3. Sensitivity analyses of the metabolite classification scheme.

Supplementary Text 4. Analysis software.

Supplementary Text 5. Data availability.

Supplementary Text 6: Supplementary discussion of Cirulli *et al*.

Supplementary Text 7: Supplementary discussion of IV analysis limitations.

Supplementary Text 2. Details of BMI GWAS in UKB.

GWAS was performed using 453,397 European samples as determined by both genetic and self-reported ancestry. Quality control of the samples have been described previously[1]. Individuals flagged by UKB as having excess heterozygosity, excess missingness, an inferred versus genetic sex mismatch, putative aneuploidy, or currently pregnant were additionally excluded. Average BMI from up to three assessment center visits of the male and female samples were separately adjusted for covariates (average age in months, average age in months squared, genotyping platform, assessment year, and the first ten genetic principal components), followed by inverse normal transformation to calculate BMI *z*-scores. BOLT-LMM[2] (v2.3.2) was used to perform linear mixed model GWAS on sex-combined BMI *z*-scores.

Supplementary Text 3. Sensitivity analyses of the metabolite classification scheme.

To assess how sensitive our genetic IV analysis and classification scheme would be to weak instrument and pleiotropy bias, we evaluated external metabolite instruments and performed a test for pleiotropy for the BMI instrument. First, to address weak instrument bias that could be caused by our G_M, which were selected using relatively small GWAS, we searched previously published GWAS to obtain external instruments for the 10 known metabolites classified in our causality groups. Eight out of the 10 known metabolites have genome-wide significant ($p < 5 \times 10^{-8}$) published instruments (**Supplementary Table 5**). Four of these metabolites have published instruments in the same loci (i.e. < 100 kb) as our G_M, indicating that a good portion of our G_M is replicable in larger cohorts. When looking at only the top published instruments (i.e. SNP with smallest *p*-value) for the cause and bidirectional metabolites, in all 5 cases, the metabolites showed consistent direction of causal effect compared to results obtained using our G_M (**Supplementary Table 5A**). Similarly, when we looked at top published instruments for the effect metabolites (i.e. metabolites that lack causal metabolite-to-BMI evidence in our data), 2 out of 3 metabolites had metabolite-to-BMI IV $p > 0.05$, thus generally agreeing with results from our G_M (**Supplementary Table 5B**). The only exception is valine, which we explored in more detail in Discussion of main text.

Next, to estimate horizontal pleiotropy among individual BMI SNPs (G_b) contained in G_B, we performed the MR-PRESSO global test for each BMI-associated metabolite (**Supplementary Table 6**). Overall, only 17 out of 324 (5.25%) metabolites had MR-PRESSO $p < 0.05$ in the pleiotropy tests, and only two of those were in our causality groups (both unknowns, with OE names HILIC-pos_4106 and C8-pos_871). Thus, horizontal pleiotropy of the BMI SNPs is unlikely to bias the majority of our metabolite classifications using G_B. However, due to insufficient numbers of strong instruments available in our metabolite GWAS or in the literature, a comparable analysis could not be performed to assess pleiotropy in the metabolite instruments.

Supplementary Text 4. Analysis software.

Unless otherwise indicated in the main text, data analyses were performed using R (v3.2) and Python (v2.7). Mapping of unknown metabolites across datasets and pathway analyses were performed using PAIRUP-MS (https://github.com/yuhanhsu/PAIRUP-MS). Inverse variance weighted meta-analyses of BMI-metabolite and metabolite-G_B associations in OE and MCDS were performed using the meta R package (v4.7.0). Plots in various figures were generated using default R functions, ggplot2 (v2.2.1), or the heatmap.2() function in gplots (v3.0.1). For main analysis steps described in this paper that were not performed using pre-existing software, we provide example scripts and documentation on GitHub (https://github.com/yuhanhsu/ObesityMetabolome).

Supplementary Text 5. Data availability.

Individual-level metabolite data for the OE cohort are available in the PAIRUP-MS GitHub (https://github.com/yuhanhsu/PAIRUP-MS); other individual-level OE and BioAge data may be requested through the Estonian Biobank (https://www.geenivaramu.ee/en/biobank.ee/data-access) and related questions should be directed to Tonu Esko (tesko@broadinstitute.org). We obtained permission to analyze the MCDS 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 (jcflorez@mgh.harvard.edu) and Dorothy Pazin (dorothy@broadinstitute.org). UKB data access can be requested through the UK Biobank Access Management System (https://bbams.ndph.ox.ac.uk/ams/). Meta-analyzed metabolite GWAS summary statistics with $p < 1 \times 10^{-5}$ are available on GitHub (https://github.com/yuhanhsu/ObesityMetabolome) and requests for the full summary statistics should be directly to Joel Hirschhorn (Joel.Hirschhorn@childrens.harvard.edu).

Supplementary Text 6. Supplementary discussion of Cirulli *et al*.

The observational association of many BMI-associated metabolites in our study generally agree with that reported by Cirulli *et al*[3] in a recent survey of the obesity metabolome in a larger number of individuals including positive (e.g. kynurenine, leucine, glutamate, valine, alanine, tyrosine, propionylcarnitine) and negative associations (e.g. asparagine, glycine, serine). Their data from longitudinal weight change metabolomics indicate that changes in the metabolome track with future changes in BMI, in a way that is at least partly independent from baseline BMI. We did not examine longitudinal weight changes, but rather attempted to identify metabolites with greater evidence of being causal for adult BMI. The authors specifically examined variants associated with BMI and their association with metabolites, both individually and as a group. Broadly, they did not identify evidence that BMI instruments (equivalent to G_B and G_b in our study) mediated the effect on BMI via metabolites. In contrast, our approach provides another means to test the directionality of effect by ranking metabolites as more or less likely to be causes of or effects of BMI using a combination of instruments for BMI and the metabolite. Finally, we did identify some metabolites with observational associations that differed from the IV association (e.g. valine is an effect of BMI with BMI IV effect directionally consistent with the crude association, while tyrosine is bidirectional with BMI vs. metabolite IV effect estimates in opposing directions).

Supplementary Text 7. Supplementary discussion of IV analysis limitations.

The presence of a limited number of G_M SNPs associated multiple metabolites could be evidence of potential pleiotropy bias whereby a non-causal metabolite may appear to be causal for a trait, because of the metabolite has shared genetic instrument with a true causal metabolite. It is also plausible, however, that these G_M are causal for multiple metabolites along a potentially causal metabolic pathway (e.g. variant for an enzyme in the pathway), which is vertical pleiotropy and not a violation of the exclusion criterion. Furthermore, because we combined metabolites into causality groups to identify enriched metabolite sets in pathway analyses, the influence of misclassification of a single metabolite is diminished (i.e. an enriched metabolite set is not driven by a single metabolite in the set).

We sometimes identified cause and effect metabolite associations to be in opposite directions from each other and/or the observational BMI-metabolite association. This may be a byproduct of our *p*-value ranking method in a relatively small dataset. Weak instrument bias in a two-sample design will bias effect estimates towards the null, thus limiting our ability to discriminate metabolites with stronger or weaker evidence of a cause or effect using the top and bottom quartiles of *p*-values. Thus, the inversion of observational and cause/effect associations may be the result of weak instrument bias, the result of pleiotropy bias, or a reflection of true complexity of obesity pathophysiology. In addition, the two directions of effect of the bidirectional metabolites could represent an admixture of effects, as pleiotropy and cyclic biological pathways would violate the exclusion criterion. The downstream pathways analyses thus use the classification (i.e. bidirectional) rather than the direction or magnitude of the IV effect estimates to infer interesting biology despite this limitation.

Our study was limited to two discovery metabolomics datasets from disparate ancestral backgrounds and at a single point in an individual's lifetime. These ancestral differences could potentially affect validity of genetic IV analyses which assume random allocation of variants conditional on ancestry. In addition, if the metabolome is ascertained under differing conditions in the

two population, this could also introduce bias (e.g. ascertainment prior to the onset of other disease states in one population versus after the onset in another population). Data from other populations (European and other ancestry), with metabolites ascertained at different times in life history will improve the validity and generalizability of these findings and the methodologic approach.

References

1 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 2017; : 166298.

2 Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; **47**: 284–290.

3 Cirulli ET, Guo L, Leon Swisher C, Shah N, Huang L, Napier LA *et al.* Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. *Cell Metab* 2019. doi:10.1016/j.cmet.2018.09.022.