ScreenDMT reveals linoleic acid diols replicably associate with BMI and stimulate adipocyte calcium fluxes

Jonathan M. Dreyfuss[1], Vera Djordjilovic[2], Hui Pan[1], Valerie Bussberg[3], Allison M. MacDonald[3], Niven R. Narain[3], Michael A. Kiebish[3], Matthias Blüher[4], Yu-Hua Tseng[5,6], Matthew D. Lynes[7,8,9*]

[1]Bioinformatics & Biostatistics Core, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA
[2]Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, Venice, Italy
[3]BPGbio Inc, Framingham, MA, USA
[4]Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG) of the Helmholtz Zentrum München at the University of Leipzig and University Hospital, Leipzig, Germany
[5]Integrative Physiology and Metabolism, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA
[6]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA
[7]Center for Molecular Medicine, MaineHealth Institute for Research, Scarborough, ME, USA
[8]Department of Medicine, MaineHealth, Portland, ME, USA
[9]Roux Institute at Northeastern University, Portland, ME, USA

* Corresponding author:
Matthew D. Lynes, Ph.D.
E-mail: matthew.lynes@mainehealth.org

Abstract

Activating brown adipose tissue (BAT) improves systemic metabolism, making it a promising target for metabolic syndrome. An activator of BAT is 12,13-dihydroxy-9Z-octadecenoic acid (12,13-diHOME), which we previously identified to be inversely associated with body mass index (BMI) and which directly improves metabolism in other tissues. Here we profile plasma lipidomics from a cohort of 83 people and test which lipids' association with BMI replicates in a concordant direction using our novel tool ScreenDMT, whose power and validity we demonstrate via mathematical proofs and simulations. We find that the linoleic acid diols 12,13-diHOME and 9,10-diHOME both replicably inversely associate with BMI and mechanistically activate calcium fluxes in mouse brown and white adipocytes in vitro. ScreenDMT can be applied to test directional replication, directional mediation, and qualitative interactions, and our results identify 9,10-diHOME and the pathway stimulated by linoleic acid diols as candidate therapeutic targets to improve systemic metabolism.

Introduction

The incidence of pre-diabetes, metabolic syndrome, and type 2 diabetes are increasing word-wide and pose a critical need for improved therapeutics. A promising approach is activating brown adipose tissue (BAT), which increases the rate of systemic metabolism by oxidizing metabolic fuels like glucose and fatty acids and secreting signaling molecules. We and others have discovered that BAT can produce omega (ω)-6 and ω-3 polyunsaturated fatty acids (PUFAs)-derived bioactive lipids (i.e. lipokines) such as 12,13-dihydroxy-octadecaenoic acid (12,13-diHOME), which increases BAT activity (*1, 2*), skeletal muscle fatty acid uptake (*3*), cardiac function (*4*), and endothelial function to reduce atherosclerosis (*5*). Furthermore, 12,13-diHOME in mother's milk is associated with lower infant fat mass (*6*) and an open-label clinical trial has found that sustained BAT activation over 4 weeks improved systemic metabolism (*7*).

We had previously identified 12,13-diHOME due to its inverse association with BMI (body mass index) in a cohort of 55 individuals (Study 1) with a broad range of BMI (*8*) and then showed that 12,13-diHOME mechanistically induces translocation of fatty acid transport proteins to the surface of cells (*8*). Other groups also observed a relationship between human obesity and 12,13-diHOME; however, the signaling pathway that drives this effect is unknown. To confirm this association in another study using analogous methods and gain the power to identify additional signaling lipids that benefit systemic metabolism, we recruited a new cohort.

In this work, we profile signaling lipids in a cohort of 83 individuals (Study 2) that was partitioned based on BMI into obese (BMI > 40 kg/m^2) and non-obese (BMI< 30 kg/m^2). We first test if we can replicate the significant, negative rank correlation between plasma concentration of 12,13-diHOME at room temperature and BMI in Study 2. Secondly, we examine if other lipids' association with BMI replicates in both studies (*8*). Replication of two studies tests a mathematically equivalent null hypothesis as mediation analysis, because both settings evaluate two independent hypotheses (mediation tests if each lipid or analyte is both associated with the exposure, and separately if it is associated with the outcome given the exposure). Replication and mediation can be tested by calculating each analyte's p-value as the analyte's maximum p-value over the two studies (i.e. the "joint significance" or "MaxP" method*'s* p-value) and then adjusting these maximum p-values for multiple testing of many analytes (*9*). The MaxP p-values were found via simulations to control the error rate and have more power than many competing approaches (*10, 11*).

However, such application of the MaxP method has two limitations: it does not account for direction, whereas it is natural to require an effect to have the same direction in both studies to be considered replicated, and there are improved methods of adjusting p-values in this context. We previously accounted for directionality in High-throughput mediation analysis (Hitman), which proposed a Directional MaxP Test (DMT) and use of empirical Bayesian linear regression modeling to improve power (*12*), although here we apply non-parametric methods to mirror our previous analysis of 12,13-diHOME (*8*). The directional MaxP test assessed mediation in a direction consistent with the exposure's effect on the outcome and its p-value is half the (non-directional) MaxP p-value if the direction is consistent, whereas otherwise its p-value is one. DMT was mathematically proven to control the error rate and was found to have stronger p-values than the MaxP and other methods (*12*).

We improved MaxP's adjusted p-values by developing ScreenMin (*13*), which reduces the multiple testing burden by filtering out analytes whose minimum p-value is so weak that their MaxP p-value could never be significant. This idea was further developed in AdaFilter (*14*), which also proposes a method to account for direction. There are other methods that improve upon MaxP adjusted p-values and can account for direction, such as radjust-sym (*15*), where analytes with strong p-values from each study are selected and only analytes selected from both studies are tested for replicability, and the empirical Bayesian method RepFdr (*16*).

Results

Development and validation of ScreenDMT

To identify which lipid's association with BMI is replicated, we sought a test that accounted for direction and had the strongest adjusted p-values in terms of the false discovery rate (FDR). Because we cared about directionality it was natural to consider DMT. By comparison, it was recently shown that the MaxP test is the likelihood ratio test (LRT) for the non-directional mediation null hypothesis (*17*), which is mathematically equivalent to testing replication without regard for direction in two studies. The LRT has many beneficial properties such as often being optimal, which is why LRTs so often used in practice, e.g. the *t*-test for means. For directional mediation and replication, we mathematically prove that DMT is the LRT in Text S1. To our knowledge, analyses of replication and mediation had not been previously connected to analysis of qualitative interactions, which occurs when an effect changes sign between groups of subjects, e.g. a treatment benefits one group of patients while it harms another. Surprisingly, we find that directional replication and mediation are connected to qualitative interaction, because applying the DMT in search of effects that have opposite signs between groups is the LRT for qualitative interactions (*18, 19*).

We considered how to improve the adjusted p-values from our directional MaxP test with screening or filtering. The filtering approach in AdaFilter when applied to two studies uses the minimum p-value ($p_{\mathbf{min}}$) per analyte as the filtering p-value and the maximum p-value as the selection p-value. We cannot directly use AdaFilter by naively setting the DMT p-value ($p_{\mathbf{DMT}}$) as the selection p-value, since the selection p-value must exceed the filtering p-value, which would not always be true. Instead, we developed a new way to apply screening to our directional MaxP test, which we term "ScreenDMT." ScreenDMT defines its selection p-value per analyte as the maximum of p_{DMT} and p_{min}. We mathematically prove ScreenDMT to be valid for controlling the FDR and the family-wise error rate (FWER; the rate controlled by the Bonferonni method) in Text S1 for large sample sizes. For small sample sizes, the screening-

based multiple testing adjustment is very slightly too strong, but this is where the DMT p-value, like similar statistical tests, is weak anyway. Like in AdaFilter (*14*), theoretical control of the FWER requires independence of p-values within each study, while control of the FDR for a large number of analytes allows weak dependence between analytes. Such weak dependence typically holds, for example, between gene loci and expression (*14*). Importantly, not all FDR methods theoretically allow dependence, such as RepFdr (*16*).

We compare ScreenDMT's FDR and power via simulation against competing directional replication methods, including the directional approaches of AdaFilter (*14*), radjust-sym (*15*), RepFdr (*16*), and DMT (Figure 1). We also compare against a directional alternative to the MaxP test that was adapted from a meta-analysis due to Pearson (*14, 20*), which applies MaxP for each analyte to right-sided p-values of both studies, then applies MaxP to left-sided p-values, and doubles the minimum from the two applications. This procedure also underlies AdaFilter's directional approach.

We see in Figure 1 that RepFdr often does not control its FDR below the nominal level, and apart from RepFdr, ScreenDMT retains the most power across the scenarios. We also see that DMT has more power than the approach adapted from Pearson (*14, 20*), which helps explain ScreenDMT's power advantage over AdaFilter. We perform the analogous simulation for FWER using methods that account for directionality and provide FWER in Figure S1. We see that all methods properly control FWER, DMT is slightly more powerful than the method adapted from Pearson (*14, 20*), and ScreenDMT's power is very similar but slightly stronger than AdaFilter's.

Inverse association of 12,13-diHOME with BMI replicates

We wanted to test if 12,13-diHOME negatively associates with BMI in the Study 2 cohort of 83 individuals (Table 1). Our previous association of 12,13-diHOME to BMI in Study 1 was assessed using Spearman rank correlation, so we also use a nonparametric test in Study 2 (Figure 2A, 2B). Since Study 2 is split into obese and non-obese subjects, we evaluate whether lipids are differentially abundant between groups with a non-parametric t-test (i.e. Wilcoxon rank sum or Mann-Whitney U test) (Figure 2C, 2D). We apply this test in a one-sided manner to test if 12,13-diHOME is lower in obese subjects. Since 12,13-diHOME had already been prioritized, its p-value does not need to be adjusted for multiple testing. We obtain a one-sided p-value of 0.00093 against a significance p-value threshold of 0.05, which demonstrates replication (Figure 2D).

Application of ScreenDMT to test replication of other lipids' association with BMI

In addition to 12,13-diHOME, we measured a panel of 108 oxidized lipid species, including metabolites of linoleic acid, α-linolenic acid, arachidonic acid, dihomo-γ-linolenic acid, Docosahexaenoic acid and Eicosapentaenoic acid (Figure 3A). To assess if any lipids other than 12,13-diHOME have an association with BMI that replicates, we applied the non-parametric t-test to all the lipids except 12,13-diHOME and then assessed replication with ScreenDMT. The only other lipid that replicated was 9,10-dihydroxy-12Z-octadecenoic acid (9,10-diHOME), which is a regioisomer of 12,13-diHOME (Figure 3B), whose FDR is 5%. Whereas if we had tested replication with DMT, the FDR of 9,10-diHOME would be nearly double at 9%. We show the plots of 9,10-diHOME vs. BMI and of 9,10-diHOME vs. 12,13-diHOME in both cohorts in Figure 2.

9,10-diHOME and 12,13-diHOME activate adipocyte calcium fluxes

12,13-diHOME and its stereoisomer 9,10-diHOME replicably associate with BMI, and we had previously shown that 12,13-diHOME increases cellular fatty acid uptake by inducing the translocation of fatty acid transporters FATP1 and CD36 to the cell surface; however, the signaling pathway that mediates this effect is unknown. 12,13-diHOME and 9,10-diHOME have both been found to activate ion flux in primary neurons and CHO cells (*21*), so we measured calcium flux in adipocytes treated with each diol. We used in vitro differentiated WT1 mouse brown preadipocytes and in vitro differentiated 3T3-L1 mouse white preadipocytes to model brown and white adipocytes, respectively. Both cell lines were differentiated using a standard adipogenic induction media, and then loaded with Fura-2. After we recorded baseline calcium flux, cells were treated with Hanks Buffered Saline Solution (HBSS) containing 1% fatty acid-free bovine serum albumin (BSA) alone or containing different concentrations of linoleic acid diol, and calcium flux was monitored every 50 seconds for approximately 9 minutes in a plate reader (Figure 5). Both 9,10-diHOME and 12,13-diHOME triggered calcium influx in brown and white adipocytes, suggesting the signaling pathway activated downstream of these two lipids overlaps.

Methods

Study 2 Cohort

We included 84 individuals with (n = 41) or without (n = 42) obesity defined by a BMI > 30 kg/m^2 (Table 1). Study participants were consecutively recruited to the Leipzig Obesity BioBank (LOBB). The study was approved by the local Ethics Committee of the University of Leipzig, Germany (Reg. numbers: 363–10-13122010 and 017–12-230112) and all subjects gave written witnessed informed consent before inclusion into the LOBB repository.

Lipidomics

Plasma was analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) to semi-quantitatively measure the concentrations of a panel of 113 signaling lipids. Plasma samples were thawed at room temperature and immediately placed on ice. Aliquots of 100 µL were taken and added to 300 µL of methanol (stored at -20 ºC) for a protein crash. Ten µL of a mixture of 5 deuterated internal standards, each at 100 pg/µL, were spiked into the samples, then samples were vortexed for 10 seconds and stored at -20 ºC overnight. Samples were then subjected to a solid phase extraction. C18 cartridges at 500 mg/6 mL (Biotage, Uppsala, Sweden) were conditioned with 10 mL of methanol followed by 10 mL of water. The samples were centrifuged at 14000 g, and the pH of the supernatant of the samples was adjusted by adding 3 mL of pH 3.5 water before loading the samples onto the C18 cartridges. The cartridges were washed with 5 mL water followed by 5 mL hexanes. Fractions were dried down under a stream of N2 gas, and reconstituted in 50 µL methanol:water (1:1, by vol). Samples were vortexed and then transferred to LC-MS vials for analysis.

Electrospray ionization (ESI) LC-MS/MS was performed on a QTRAP 6500 (Sciex, Framingham, MA, USA) coupled to an Agilent Infinity 1290 (Agilent, Santa Clara, CA, USA) LC system with a InfinityLab Poroshell 120 EC-C18 (4.6 x 100 mm, 2.7 µm; Agilent) analytical LC column with a column oven heated to 60ºC. Ten µL of sample was injected at a flow rate of 400 µL/min and were separated by reversed phase chromatography with mobile phases A

(100% H2O, 0.1% acetic acid) and B (100% MeOH, 0.1% acetic acid). The gradient started at 5% B and increased to 20% B by 3 minutes, then increased to 60% B by 8 minutes, increased to 90% B by 23 minutes, held at 90% B for 3 minutes, then returned to 5% B for the last 3 minutes. The total LC run time was 29 minutes. Samples were only acquired in negative polarity due to the chemical structure of the targeted lipids. The ESI source parameters were ion source gas 1 (GS1) 30, ion source gas 2 (GS2) 30, curtain gas (CUR) 30, ion spray voltage (IS) -4500, and temperature 500. The declustering potential (DP), entrance potential (EP), collision energy (CE), and exit potential (CXP) were tuned for each individual targeted lipid and internal standard. The MS method was a targeted scheduled MRM method. There were 101 MRM scans scheduled for optimized windows ranging from 120–360 seconds. The mass spectrometer acquisition time was set to 29 minutes. The LC-MS/MS data was acquired by Analyst 1.6.2 software (Sciex, Framingham, MA, USA) and processed with MultiQuant 3.0.1 (Sciex, Framingham, MA, USA) for peak integration. The targeted lipid species were measured semi-quantitatively and reported as area ratios of the peak area of the analyte divided by the peak area of the internal standard.

Bioinformatics analysis of lipidomics data

We reanalyzed our previous lipidomics data from Study 1 (*8*) and Study 2's lipidomics data using the R software. Both data sets included values of zero for missing data. Our only processing step was to filter out lipids with 20% or more missing values, which was applied to both datasets. To reproduce our previous study, we tested Spearman rank correlation with R function cor.test, which yielded the same coefficient and p-value as previously reported (*8*). For the Study 2 lipidomics dataset, we tested differential abundance between groups using the R function wilcox.test. Wilcoxon test z-scores were calculated from the wilcoxonZ function in the R package rcompanion. Replication FDRs from ScreenDMT were considered significant using FDR < 15%, as we used previously (*12*).

Cell Culture

WT-1 mouse preadipocytes and 3T3-L1 mouse fibroblasts were cultured in high glucose DMEM supplemented with 10% FBS. When cells reached confluence, adipocyte differentiation was induced by using induction media (DMEM high glucose media with 10% FBS, 33 µM Biotin, 17 µM Pantothenate, 0.5 µM human insulin, 500 µM IBMX, 2 nM T3, 0.1 µM dexamethasone and 30 µM indomethacin) for 2 days, then cells were switched to a differentiation media (DMEM high glucose media with 10% FBS, 0.5 µM human insulin, 2 nM T3) for a final 7 days. After fully differentiated, adipocytes were washed with PBS and starved for 1 hour in DMEM.

Calcium Flux Analysis

FLUOFORTE fluorescence was measured in adipocytes according to manufacturer's directions (Enzo Scientific, Farmingdale, NY, USA). Briefly, starved adipocytes were incubated with FLUOFORTE Dye-loading solution for 1 hour at room temperature, then baseline fluorescence was read on a plate reader. After treatment with vehicle control or different concentrations of lipid diol, FLUOFORTE fluorescence was read every 50 seconds.

Replication simulations

We assessed the performance of methods that provide adjusted p-values by simulating two studies with the same 1,000 features or analytes following the simulations in (*22*). We simulated normally distributed statistics with variance of one per analyte per study. The statistics for analytes without signal were simulated with mean zero in both studies. In the Equal Signal simulation, the statistics for analytes that were truly directionally replicated were simulated with a mean of 3.5 or -3.5 in both studies, and these comprised 5% of the analytes. The statistics for analytes that were not directionally replicated but had some signal ("Percent of null analytes with signal") were simulated with mean zero in one study and mean magnitude of 3.5 in the other, or else they were simulated with a mean of 3.5 in one study and -3.5 in the other study. In the Unequal Signal setting, the mean magnitudes for one study was 2.66 while the other study had mean magnitude of 4.66. Like the AdaFilter simulations (*14*), the statistics were simulated under weak dependence, where analytes were correlated with correlation coefficient of 0.5 to other features in their block with block size of 10, whereas under strong dependence, features were similarly correlated but in block sizes of 100. Features were randomly assorted to blocks.

For each setting 1000 simulations were run. RepFdr using its default parameters was too slow to run this many simulations, so the tolerance specified in RepFdr's package vignette was used. So that RepFdr calculates FDRs when it estimates the fraction of nulls to be one in both studies, we modified it to calculate each analyte's FDR as one instead of producing an error. For the FWER case, we do not compare to (*15*) because its replication implementation radjust-sym does not include a FWER procedure and its mediation implementation does not account for direction (*23*).

Discussion

We test replication which has been defined as obtaining the same result in new data (*24*). Although often used synonymously, "reproducibility" can be defined as obtaining the same result on the same data with the same analysis (*24*), whereas meta-analysis combines statistics from datasets, which can yield significance due to a strong effect in a single study. Replication is fundamental to science and its mathematical treatment has been found to have numerous applications (*9*). We focus on replication in a concordant direction and show that our test for directional replication is mathematically equivalent to testing for qualitative interactions and directional mediation. An example of a qualitative interaction would be a pair of genes that are positively associated in healthy participants and negatively associated in those with disease. Finding such pairs is important for identifying the gene interactions causing disease and other questions in differential network biology (*25*).

We demonstrated the value of mediation accounting for direction in a previous study where Hitman, which is based on DMT, identified growth hormone receptor as a mediator of gastric bypass surgery's reduction of glycemia (*12*). We found that surgery reduced GHR levels, and lower GHR was associated with lower glycemia. Mediation methods that do not account for direction would have called the metabolite retinol as a mediator, even though surgery reduced retinol levels, and lower retinol was associated with glycemia being increased. Thus, surgery's apparent action via retinol would be for glycemia to increase. This type of "inconsistent" mediation is unlikely and does not help explain mechanism of action, so our directional MaxP test saves power by assigning such an analyte a p-value of one. ScreenDMT improved upon the directional MaxP test with more efficient adjusted p-values (*14*). These improved adjusted p-values correspond to selection p-values that are based on, but not

necessarily the same, as the directional MaxP test p-values. Thus, at times, the adjusted p-values may not have the same order as the DMT p-values. As usual when testing many hypotheses, significance should be based on adjusted p-values.

A possible reason that both 12,13-diHOME and 9,10-diHOME levels are associated with BMI is their shared biosynthetic pathways downstream of linoleic acid. Linoleic acid is a polyunsatured omega-6 fatty acid that is one of two essential fatty acids for humans and can be used as an energy source (*26*). However, it can also be converted to arachinodonic acid and metabolized into a diverse group of signaling molecules, including the linoleic acid diols (*27, 28*). Linoleic acid diols are synthesized in a two-step process that begins with oxidation of linoleic acid at either the double bond at the 9,10 position or the 12,13 position, resulting in one of two regioisomer epoxides. This step is catalyzed by Cytochrome P450 oxidases, however little is known regarding the specific identity of the enzyme(s) that catalyze this step or if and how the position of the epoxide moiety is determined. Biosynthesis of 12,13-diHOME from its parent epoxide is mediated by an epoxide hydrolase, most likely soluble epoxide hydrolase encoded by the gene Ephx2, as mice lacking Ephx2 have circulating levels of 12,13-diHOME 8-fold lower than wildtype or Ephx1 knockout controls (*29*). Similarly, 9,10-diHOME is generated from an epoxide precursor called 9,10-epOME, however only genetic ablation of both Ephx2 and Ephx1, which encodes the microsomal isozyme of epoxide hydrolase, is sufficient to lower the levels of circulating 9,10-diHOME (*29*).

In mice and cells, 12,13-diHOME can activate changes in metabolism by recruiting fatty acid transport proteins to the cell surface, however the signaling pathway or pathways that mediate this effect are unknown. A candidate transport protein is CD36 and translocation of CD36 can be activated by calcium flux (*30*). Since both 12,13-diHOME and 9,10-diHOME can activate changes in membrane polarization (*21*) and we have recently shown that 12,13-diHOME can activate calcium flux in bovine endothelial cells (*5*), we tested whether both lipids in their physiological range from 1-100nM (*8, 31, 32*) have a similar effect on adipocyte calcium flux and found that they do. The beneficial effect of 9,10-diHOME is historically surprising, because it was originally tested in the micromolar range where it was found to cause mitochondrial damage in leukocytes, so it was named Leukotoxin diol. However, given our new findings there is an evidence base to test the beneficial effects of 9,10-diHOME on systemic metabolism and given the overlap between signaling that we observed with 12,13-diHOME, the signaling pathway targeted by these two molecules could prove to be a valuable therapeutic target in the treatment of obesity and its sequalae.

Figure Legends

Figure 1. Comparison of ScreenDMT's power and false discovery rate to competing approaches in replication of two datasets via simulation. Simulation of two datasets with strong or weak dependence between analytes when the signal strength or effect size per analyte is equal between the two datasets and when it is unequal. The x-axis represents the percent of analytes that do not truly directionally replicate but show some effect in one of the studies or show effect in both studies of opposite direction. (A) Probability of detecting true associations (power). (B) False discovery rate. The FDR threshold is 5%.

Figure 2. 9,10-diHOME and 12,13-diHOME versus BMI in both cohorts. (A) Circulating 9,10-diHOME versus BMI in patients from the original cohort. (B) Circulating 12,13-diHOME versus BMI in patients from the original cohort. (C) Circulating 9,10-diHOME versus BMI in patients from the second validation cohort. (D) Circulating 12,13-diHOME versus BMI in

patients from the second validation cohort. All lipid values were $\log_2$ transformed for graphing.

Figure 3. Biosynthesis of linoleic acid diols 9,10-diHOME. (A) Lipokines measured in our lipidomic panel shown as a product of their precursor fatty acids and the oxidative pathways that are the first step in their biosynthesis. (B) Biosynthesis of linoleic acid diols.

Figure 4. Linoleic acid diols activate calcium flux in cultured adipocytes. (A) FLUOFORTE quantification in differentiated WT1 brown adipocytes treated after timepoint 2 with different concentrations of 9,10-diHOME or vehicle control. (B) FLUOFORTE quantification in differentiated WT1 brown adipocytes treated after timepoint 2 with different concentrations of 12,13-diHOME or vehicle control. (C) FLUOFORTE quantification in differentiated 3T3-L1 white adipocytes treated after timepoint 2 with different concentrations of 9,10-diHOME or vehicle control. (D) FLUOFORTE quantification in differentiated 3T3-L1 white adipocytes treated after timepoint 2 with different concentrations of 12,13-diHOME or vehicle control. Each timepoint is approximately 50 seconds apart. Data are plotted as the normalized means ± s.e.m.; n = 6 wells per group; *P < 0.05 vs. vehicle by ANOVA with post-hoc Bonferroni test.

Table 1. Anthropometrics of Study 2 cohort. Data are shown as mean for each group with the interquartile range in parentheses.

Figure S1. Comparison of ScreenDMT's power and family-wise error rate to competing approaches in replication of two datasets via simulation. Simulation of two datasets with strong or weak dependence between analytes when the signal strength per analyte is equal between the two datasets and when it is unequal. The x-axis represents the percent of analytes that do not truly directionally replicate but show some effect in one of the studies or show effect in both studies of opposite direction. (A) Probability of detecting true associations (power). (B) Family-wise error rate (FWER). The FWER threshold is 5%.

Author Contributions

Conceptualization: JMD, MDL; Methodology: JMD, VD, HP, MDL; Software: JMD, VD, HP; Validation: JMD, VD, MDL; Investigation: JMD, VD, VB, AMM, MDL; Writing – Original Draft: JMD, MDL. Writing – Review & Editing: VD, HP, Y-HT; Supervision: MAK, NRN, Y-HT.

Funding

This work was supported in part by U.S. National Institutes of Health (NIH) grants R01DK122808 and R01DK102898 (to Y.-H.T.), P30DK036836 (to Joslin Diabetes Center's Diabetes Research Center), and by US Army Medical Research grant W81XWH-17-1-0428 (to Y.-H.T.). M.D.L was supported by NIH fellowships (T32DK007260, F32DK102320 and K01DK111714).

Data Availability Statement

Lipidomics measurements for both studies and sample metadata is available at https://github.com/jdreyf/screendmt-dihome-replication, which also contains R code to reproduce this paper's lipidomics results and simulations. The code relies on the freely

available R package DirectionalMaxPTest (https://github.com/jdreyf/DirectionalMaxPTest), which implements DMT and ScreenDMT. All datasets generated for this study are included in this article.

Conflicts of Interest

VB, AMM, MAK, and NRN are employees of BPGbio Inc. Y-HT and MDL are inventors of US Patent 11,433,042 related to 12,13-diHOME.

References

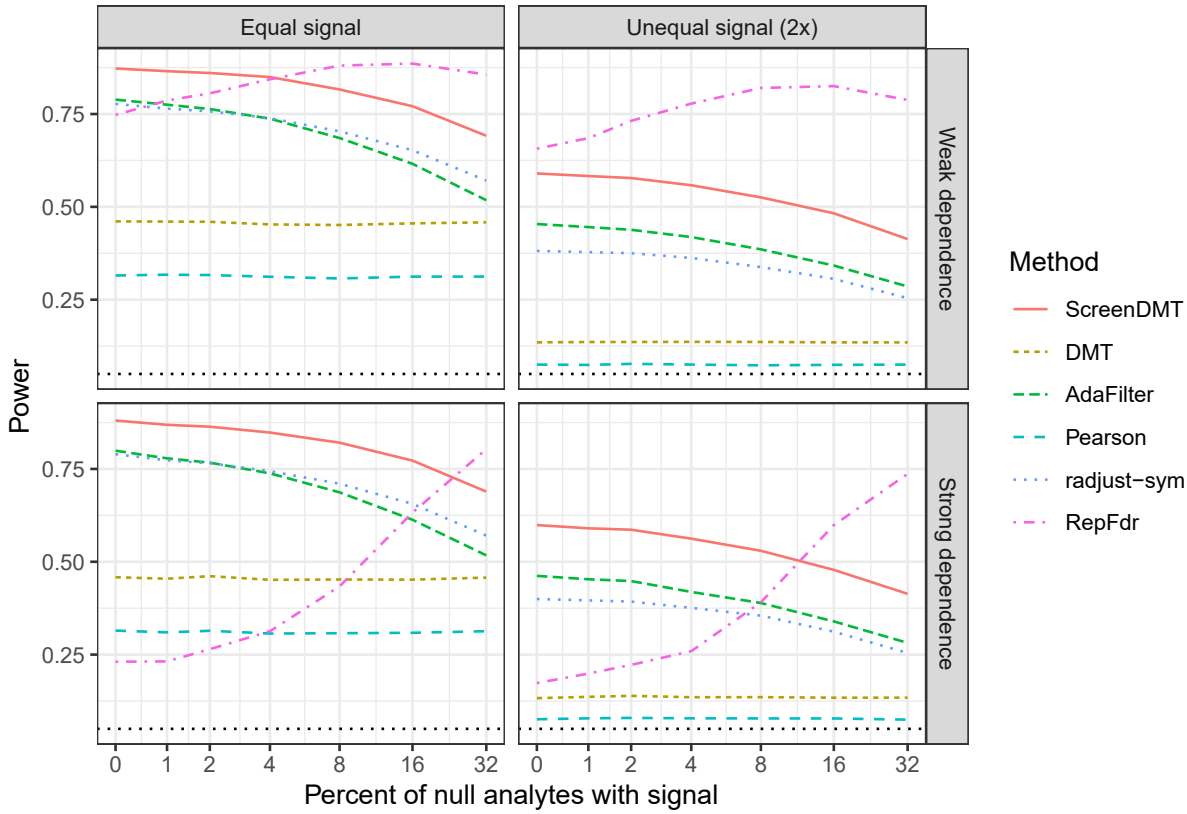
1. F. Shamsi, C. H. Wang, Y. H. Tseng, The evolving view of thermogenic adipocytes - ontogeny, niche and function. *Nat Rev Endocrinol* **17**, 726-744 (2021).
2. L. O. Leiria, C. H. Wang, M. D. Lynes, K. Yang, F. Shamsi, M. Sato, S. Sugimoto, E. Y. Chen, V. Bussberg, N. R. Narain, B. E. Sansbury, J. Darcy, T. L. Huang, S. D. Kodani, M. Sakaguchi, A. L. Rocha, T. J. Schulz, A. Bartelt, G. S. Hotamisligil, M. F. Hirshman, K. van Leyen, L. J. Goodyear, M. Blüher, A. M. Cypess, M. A. Kiebish, M. Spite, Y. H. Tseng, 12-Lipoxygenase Regulates Cold Adaptation and Glucose Metabolism by Producing the Omega-3 Lipid 12-HEPE from Brown Fat. *Cell Metab* **30**, 768-783.e767 (2019).
3. K. I. Stanford, M. D. Lynes, H. Takahashi, L. A. Baer, P. J. Arts, F. J. May, A. C. Lehnig, R. J. W. Middelbeek, J. J. Richard, K. So, E. Y. Chen, F. Gao, N. R. Narain, G. Distefano, V. K. Shettigar, M. F. Hirshman, M. T. Ziolo, M. A. Kiebish, Y.-H. Tseng, P. M. Coen, L. J. Goodyear, 12,13-diHOME: An Exercise-Induced Lipokine that Increases Skeletal Muscle Fatty Acid Uptake. *Cell Metabolism* **27**, 1111-1120.e1113 (2018).
4. K. M. Pinckard, V. K. Shettigar, K. R. Wright, E. Abay, L. A. Baer, P. Vidal, R. S. Dewal, D. Das, S. Duarte-Sanmiguel, D. Hernández-Saavedra, P. J. Arts, A. C. Lehnig, V. Bussberg, N. R. Narain, M. A. Kiebish, F. Yi, L. M. Sparks, B. H. Goodpaster, S. R. Smith, R. E. Pratley, E. D. Lewandowski, S. V. Raman, L. E. Wold, D. Gallego-Perez, P. M. Coen, M. T. Ziolo, K. I. Stanford, A Novel Endocrine Role for the BAT-Released Lipokine 12,13-diHOME to Mediate Cardiac Function. *Circulation* **143**, 145-159 (2021).
5. K. Park, Q. Li, M. D. Lynes, H. Yokomizo, E. Maddaloni, T. Shinjo, R. St-Louis, Q. Li, S. Katagiri, J. Fu, A. Clermont, H. Park, I. H. Wu, M. G. Yu, H. Shah, Y.-H. Tseng, G. L. King, Endothelial cells induced progenitors into brown fat to reduce atherosclerosis. *Circ Res* **131**, 168-183 (2022).
6. D. Wolfs, M. D. Lynes, Y.-H. Tseng, S. Pierce, V. Bussberg, A. Darkwah, V. Tolstikov, N. R. Narain, M. C. Rudolph, M. A. Kiebish, E. W. Demerath, D. A. Fields, E. Isganaitis, Brown Fat-Activating Lipokine 12,13-diHOME in Human Milk Is Associated With Infant Adiposity. *The Journal of Clinical Endocrinology and Metabolism* **106**, e943-e956 (2021).
7. A. E. O'Mara, J. W. Johnson, J. D. Linderman, R. J. Brychta, S. McGehee, L. A. Fletcher, Y. A. Fink, D. Kapuria, T. M. Cassimatis, N. Kelsey, C. Cero, Z. A. Sater, F. Piccinini, A. S. Baskin, B. P. Leitner, H. Cai, C. M. Millo, W. Dieckmann, M. Walter, N. B. Javitt, Y. Rotman, P. J. Walter, M. Ader, R. N. Bergman, P. Herscovitch, K. Y. Chen, A. M. Cypess, Chronic mirabegron treatment increases human brown fat, HDL cholesterol, and insulin sensitivity. *The Journal of Clinical Investigation* **130**, 2209-2219 (2020).
8. M. D. Lynes, L. O. Leiria, M. Lundh, A. Bartelt, F. Shamsi, T. L. Huang, H. Takahashi, M. F. Hirshman, C. Schlein, A. Lee, L. A. Baer, F. J. May, F. Gao, N. R. Narain, E. Y. Chen, M. A. Kiebish, A. M. Cypess, M. Blüher, L. J. Goodyear, G. S. Hotamisligil, K. I. Stanford, Y. H.

Tseng, The cold-induced lipokine 12,13-diHOME promotes fatty acid transport into brown adipose tissue. *Nat Med* **23**, 631-637 (2017).
9. Y. Benjamini, R. Heller, D. Yekutieli, Selective inference in complex research. *Philos Transact A Math Phys Eng Sci* **367**, 4255-4271 (2009).
10. D. P. MacKinnon, C. M. Lockwood, J. M. Hoffman, S. G. West, V. Sheets, A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* **7**, 83-104 (2002).
11. R. Barfield, J. Shen, A. C. Just, P. S. Vokonas, J. Schwartz, A. A. Baccarelli, T. J. VanderWeele, X. Lin, Testing for the indirect effect under the null for genome-wide mediation analyses. *Genet Epidemiol* **41**, 824-833 (2017).
12. J. M. Dreyfuss, Y. Yuchi, X. Dong, V. Efthymiou, H. Pan, D. C. Simonson, A. Vernon, F. Halperin, P. Aryal, A. Konkar, Y. Sebastian, B. W. Higgs, J. Grimsby, C. M. Rondinone, S. Kasif, B. B. Kahn, K. Foster, R. Seeley, A. Goldfine, V. Djordjilović, M. E. Patti, High-throughput mediation analysis of human proteome and metabolome identifies mediators of post-bariatric surgical diabetes control. *Nat Commun* **12**, 6951 (2021).
13. V. Djordjilović, C. M. Page, J. M. Gran, T. H. Nøst, T. M. Sandanger, M. B. Veierød, M. Thoresen, Global test for high-dimensional mediation: Testing groups of potential mediators. *Stat Med* **38**, 3346-3360 (2019).
14. J. Wang, L. Gui, W. J. Su, C. Sabatti, A. B. Owen, Detecting multiple replicating signals using adaptive filtering procedures. *Ann. Statist.* **50**, (2022).
15. M. Bogomolov, R. Heller, Assessing replicability of findings across two studies of multiple features. *Biometrika* **105**, 505-516 (2018).
16. R. Heller, D. Yekutieli, Replicability analysis for genome-wide association studies. *Ann Appl Stat* **8**, 481-498 (2014).
17. Z. Liu, J. Shen, R. Barfield, J. Schwartz, A. A. Baccarelli, X. Lin, Large-Scale Hypothesis Testing for Causal Mediation Effects with Applications in Genome-wide Epigenetic Studies. *J Am Stat Assoc* **117**, 67-81 (2022).
18. A. Hudson, A. Shojaie, Statistical Inference for Qualitative Interactions with Applications to Precision Medicine and Differential Network Analysis. *arXiv*, (2020).
19. M. Gail, R. Simon, Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41**, 361-372 (1985).
20. A. B. Owen, Karl Pearson's meta-analysis revisited. *Ann. Statist.* **37**, (2009).
21. D. Green, S. Ruparel, X. Gao, N. Ruparel, M. Patil, A. Akopian, K. Hargreaves, Central activation of TRPV1 and TRPA1 by novel endogenous agonists contributes to mechanical allodynia and thermal hyperalgesia after burn injury. *Mol Pain* **12**, (2016).
22. V. Djordjilović, J. Hemerik, M. Thoresen, On optimal two-stage testing of multiple mediators. *Biom J*, (2022).
23. J. N. Sampson, S. M. Boca, S. C. Moore, R. Heller, FWER and FDR control when testing multiple mediators. *Bioinformatics* **34**, 2418-2424 (2018).
24. B. A. Nosek, T. E. Hardwicke, H. Moshontz, A. Allard, K. S. Corker, A. Dreber, F. Fidler, J. Hilgard, M. Kline Struhl, M. B. Nuijten, J. M. Rohrer, F. Romero, A. M. Scheel, L. D. Scherer, F. D. Schönbrodt, S. Vazire, Replicability, robustness, and reproducibility in psychological science. *Annu Rev Psychol* **73**, 719-748 (2022).
25. T. Ideker, N. Krogan, Differential network biology. *Molecular systems biology* **8**, n/a-n/a (2012).
26. C. Carneheim, B. Cannon, J. Nedergaard, Rare fatty acids in brown fat are substrates for thermogenesis during arousal from hibernation. *Am J Physiol* **256**, R146-154 (1989).
27. M. D. Lynes, S. D. Kodani, Y. H. Tseng, Lipokines and Thermogenesis. *Endocrinology* **160**,

2314-2325 (2019).
28. I. Abe, Y. Oguri, A. R. P. Verkerke, L. B. Monteiro, C. M. Knuth, C. Auger, Y. Qiu, G. P. Westcott, S. Cinti, K. Shinoda, M. G. Jeschke, S. Kajimura, Lipolysis-derived linoleic acid drives beige fat progenitor cell proliferation. *Dev Cell* **57**, 2623-2637.e2628 (2022).
29. M. L. Edin, B. G. Hamedani, A. Gruzdev, J. P. Graves, F. B. Lih, S. J. Arbes, 3rd, R. Singh, A. C. Orjuela Leon, J. A. Bradbury, L. M. DeGraff, S. L. Hoopes, M. Arand, D. C. Zeldin, Epoxide hydrolase 1 (EPHX1) hydrolyzes epoxyeicosanoids and impairs cardiac recovery after ischemia. *J Biol Chem* **293**, 3281-3292 (2018).
30. Y. Angin, R. W. Schwenk, R. Nergiz-Unal, N. Hoebers, J. W. Heemskerk, M. J. Kuijpers, W. A. Coumans, M. A. van Zandvoort, A. Bonen, D. Neumann, J. F. Glatz, J. J. Luiken, Calcium signaling recruits substrate transporters GLUT4 and CD36 to the sarcolemma without increasing cardiac substrate uptake. *Am J Physiol Endocrinol Metab* **307**, E225-236 (2014).
31. S. K. Vasan, R. Noordam, M. S. Gowri, M. J. Neville, F. Karpe, C. Christodoulides, The proposed systemic thermogenic metabolites succinate and 12,13-diHOME are inversely associated with adiposity and related metabolic traits: evidence from a large human cross-sectional study. *Diabetologia* **62**, 2079-2087 (2019).
32. O. C. Kulterer, L. Niederstaetter, C. T. Herz, A. R. Haug, A. Bileck, D. Pils, A. Kautzky-Willer, C. Gerner, F. W. Kiefer, The Presence of Active Brown Adipose Tissue Determines Cold-Induced Energy Expenditure and Oxylipin Profiles in Humans. *J Clin Endocrinol Metab* **105**, (2020).

A

Probability of detecting true associations



B

False discovery rate

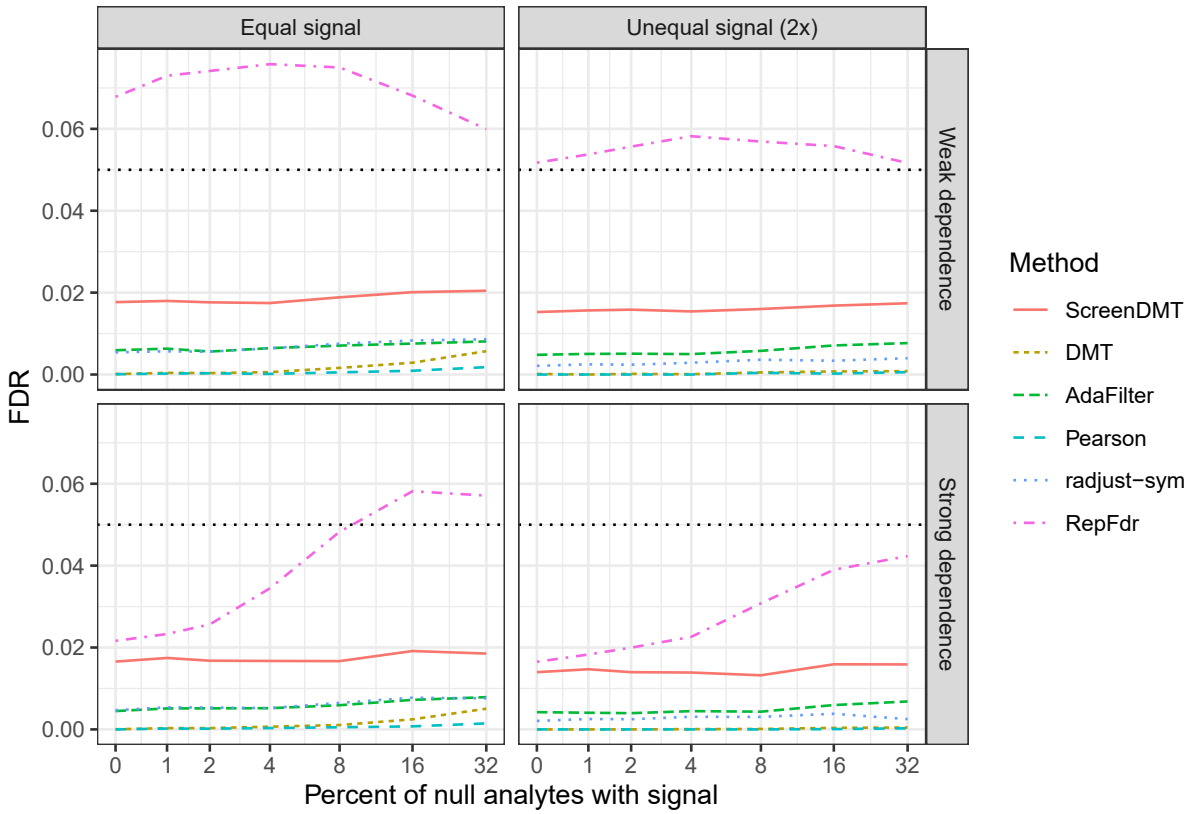


Figure 1

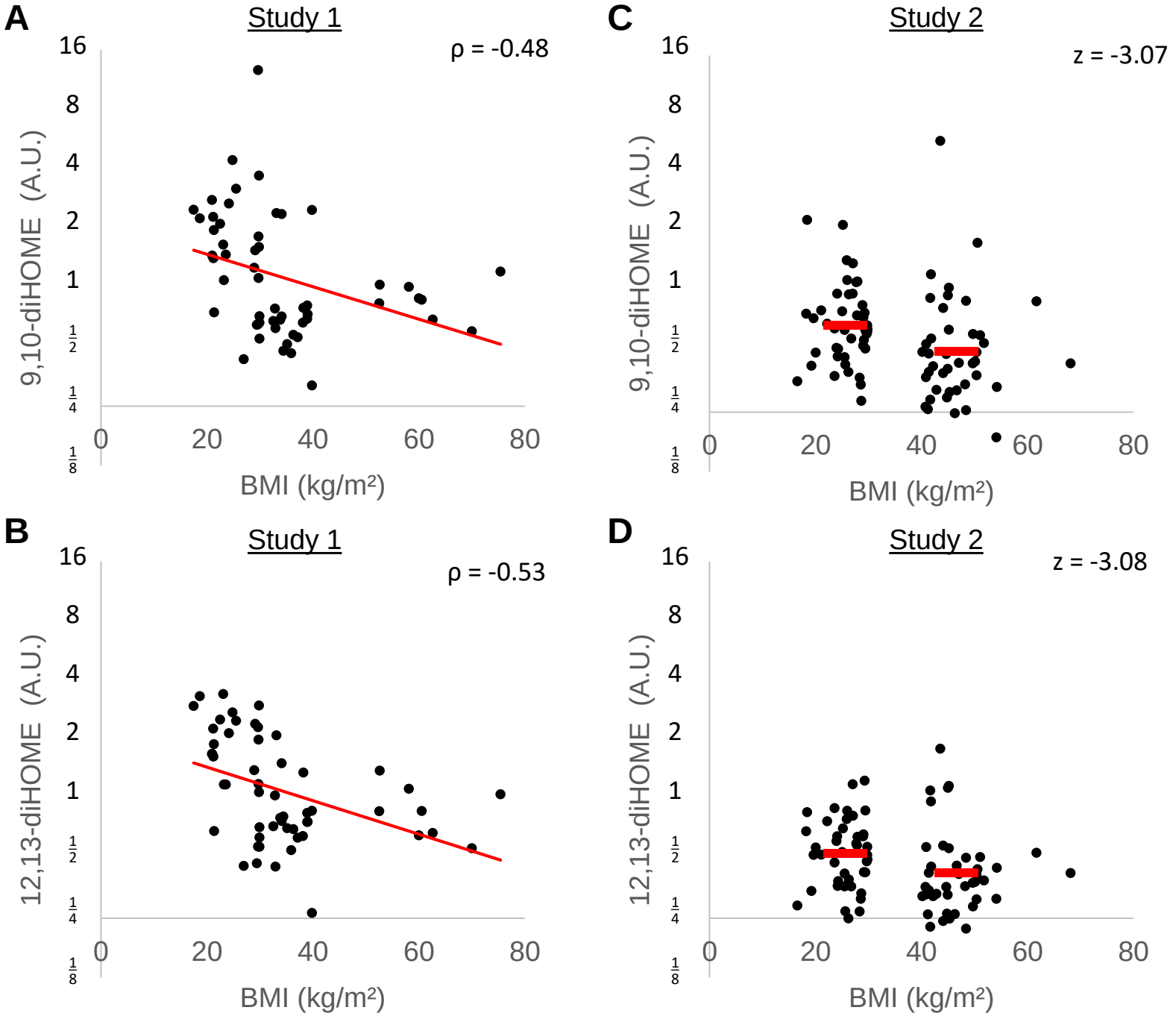

A
Study 1
ρ = -0.48
9,10-diHOME (A.U.)
16
8
4
2
1
1/2
1/4
1/8
0
20
40
60
80
BMI (kg/m²)
C
Study 2
z = -3.07
9,10-diHOME (A.U.)
BMI (kg/m²)
B
Study 1
ρ = -0.53
12,13-diHOME (A.U.)
BMI (kg/m²)
D
Study 2
z = -3.08
12,13-diHOME (A.U.)
BMI (kg/m²)

Figure 2

A

	Cycloxygenase Dioxygenase (FA+O_2)	Lipoxygenase Dioxygenase (FA+O_2)	Cytochrome P450 oxidase Monoxygenase (FA+O)	Non-enzymatic
Linoleic Acid (LA, 18:2 ω-6)		9-oxoODE 13-oxoODE 9-HODE 13-HODE	9(10)-EpOME 12(13)-EpOME 9,10-diHOME 12,13-diHOME	
α-Linolenic acid (ALA, 18:3 ω-3)		9-HOTrE 13-HOTrE/13-HOTrE(r)		
Arachinodonic Acid (AA, 20:4 ω-6)	15-deoxy-delta12,14-PGJ2 2,3-dinor-11beta-PGF2a tetranor-PGFM PGA2/PGJ2 PGB2 15-deoxy-delta121,4-PGD2 1314-dihydro-15-keto PGA2 Bicyclo PGE2 delta12-PGJ2 PGA1 2,3-dinor TxB2 15-keto-PGE2 PGK2 13,14-dihydro-15-keto PGE2 13,14-dihydro-15-keto PGD2 15-keto-PGF2a PGE2/PGD2 PGD2 PGF2a 11-beta-PGF2a/PGF2b 13,14-dihydro-15-keto PGF2a 20-carboxy LTB4 11-dehydro TxB2 19/20-OH PGE2 6-keto PGE1 TxB2 6-keto-PGF1a 19/20-OH PGF2a 6,15-diketo-1314-dihydro PGF1a 1314-dihydro PGE1	Tetranor-12-HETE 15-oxoETE 12-oxoETE 5-oxoETE 15-HETE 12-HETE 5-HETE 20-HETE 12-oxoLTB4 LTB4 5,6-diHETE 5,15-diHETE Hepoxilin A3 LXA4 LXB4 LTE4 LTD4 LTC4	14(15)-EET 11(12)-EET 8(9)-EET 5(6)-EET 16-HETE 17-HETE 18-HETE 5,6-diHETrE 8,9-diHETrE 11,12-diHETrE 14,15-diHETrE	11-HETE 9-HETE 8-HETE 8-iso PGF2a 5-iPF2a-VI
Dihomo-γ-linolenic acid (DGLA, 20:3 ω-6)	PGE1/PGD1 PGF1A	5-HETrE 8-HETrE 15-HETrE		
Docosahexaenoic acid (DHA, 22:6 ω-3)		17-HDHA 14-HDHA 7-HDHA 4-HDHA 8-HDHA 10-HDHA 13-HDHA 16-HDHA 20-HDHA Maresin1 PD1 RvD1 RvD2	19(20)-EpDPE 16(17)-EpDPE 19,20-diHDPA	11-HDHA
Eicosapentaenoic acid (EPA, 20:5 ω-3)	PGD3/PGE3 TxB3	18-HEPE 15-HEPE 12-HEPE 5-HEPE 8-HEPE 14,15-diHETE	14(15)-EpETE 17(18)-EpETE 17,18-diHETE	11-HEPE 9-HEPE

B

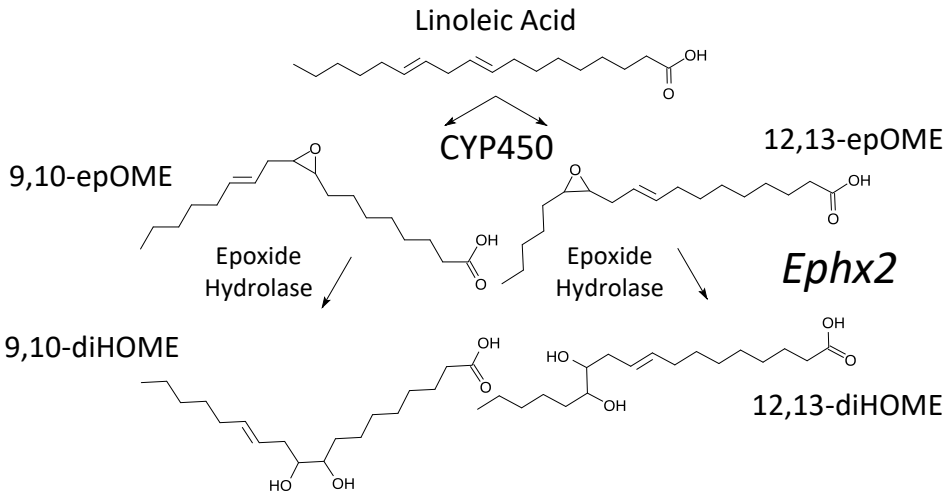


Figure 3

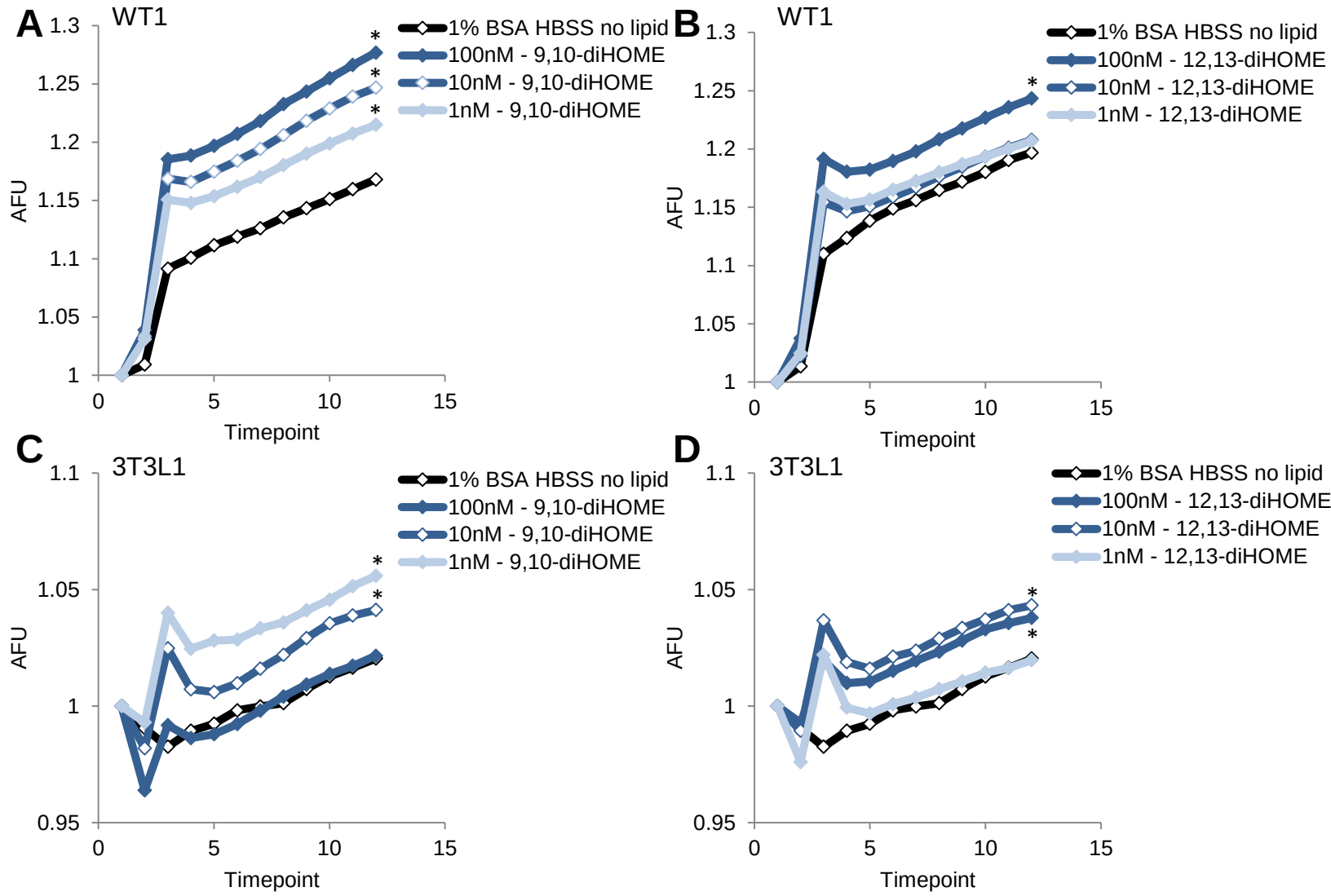
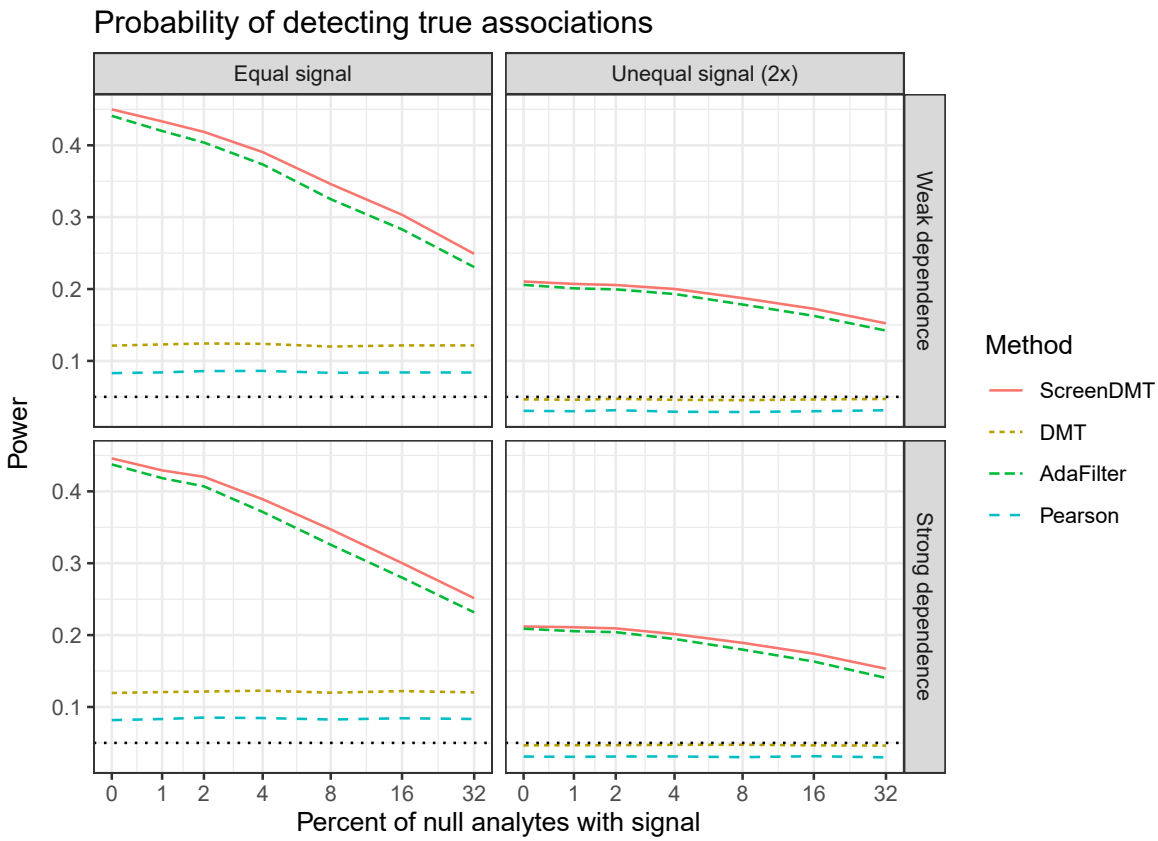

A
WT1
1% BSA HBSS no lipid
100nM - 9,10-diHOME
10nM - 9,10-diHOME
1nM - 9,10-diHOME
AFU
Timepoint
B
WT1
1% BSA HBSS no lipid
100nM - 12,13-diHOME
10nM - 12,13-diHOME
1nM - 12,13-diHOME
AFU
Timepoint
C
3T3L1
1% BSA HBSS no lipid
100nM - 9,10-diHOME
10nM - 9,10-diHOME
1nM - 9,10-diHOME
AFU
Timepoint
D
3T3L1
1% BSA HBSS no lipid
100nM - 12,13-diHOME
10nM - 12,13-diHOME
1nM - 12,13-diHOME
AFU
Timepoint

Figure 4

Study group	Lean	Obese
Age (years)	52.1 (32.5-64.6)	41.6 (31.3-51.9)
BMI (kg/m²)	26.2 (24-28.8)	45.1 (41.8-49.7)
Triglycerides (mmol/l)	0.8 (0.6-03.5)	5.3 (0.6-6.6)
Height (m)	1.7 (1.6-1.8)	1.7 (1.7-1.8)
Body weight (kg)	75.5 (67.9-83)	135.6 (117.9-150)
BMI (kg/m²)	26.2 (24-28.8)	45.1 (41.8-49.7)
Body fat (%)	30.9 (23.1-38.6)	46 (41.3-48.9)
Hemoglobin (mmol/l)	8.3 (7.9-8.9)	8.3 (7.9-8.9)
Erythrocytes (exp12/l)	4.6 (4.3-4.8)	4.7 (4.5-4.9)
Thrombocytes (exp9/l)	259 (230.3-294)	285.5 (244.5-320.3)
Leukocytes (exp9/l)	6.6 (5.5-8)	7.7 (6.4-8.8)
ALAT (µkat/l)	0.4 (0.3-0.5)	0.5 (0.4-0.6)
ASAT (µkat/l)	0.5 (0.4-0.5)	0.5 (0.4-0.6)
gGT (µkat/l)	0.7 (0.3-0.7)	0.7 (0.3-1.1)
Fasting plasma glucose (mmol/l)	5.3 (4.8-5.9)	5.8 (5.2-7.8)
HbA1c (%)	5.6 (5.4-6)	5.8 (5.3-6.6)
Fasting plasma insulin (pmol/l)	31.3 (15.9-83.5)	143.5 (79.2-185.4)
C-Peptid (nmol/l)	1 (0.7-1.3)	1.4 (0.9-2)
Creatinine (µmol/l)	76 (59-85)	67.5 (56.8-79.3)
Cholesterol total (mmol/l)	5.5 (4.5-6.4)	4.9 (4.3-5.3)
HDL-Cholesterol (mmol/l)	1.3 (1.1-1.8)	1.1 (0.9-1.5)
LDL-Cholesterol (mmol/l)	3.1 (2.4-3.9)	2.6 (2.1-2.9)
Total protein (g/l)	72.3 (69.8-78)	75.6 (73.2-84.1)
Albumin (g/l)	45.9 (44.4-46.8)	45.5 (43.6-46.7)
TSH (mU/l)	1.5 (0.9-2)	2.1 (1.4-3.3)
fT3 (pmol/l)	4.8 (4.3-5.2)	4.5 (4.3-5.3)
fT4 (pmol/l)	15.7 (14-17.1)	13.8 (13-15.6)
Cortisol (nmol/l)	354 (212.9-549.7)	396.3 (302.8-500.8)
Testosteron (nmol/l)	1.1 (0.5-9)	1.4 (0.9-5.5)
RRsys (mmHg)	131 (124.3-139.3)	143 (127-162)
RR dia (mmHg)	84.5 (74.5-90)	85 (80-94)
hsCRP (mg/dl)	0.3 (0.2-1)	2.2 (1.1-3.8)

Table 1

A


Probability of detecting true associations
Equal signal
Unequal signal (2x)
Weak dependence
Strong dependence
Power
0.4
0.3
0.2
0.1
0
1
2
4
8
16
32
Percent of null analytes with signal
Method
ScreenDMT
DMT
AdaFilter
Pearson

B

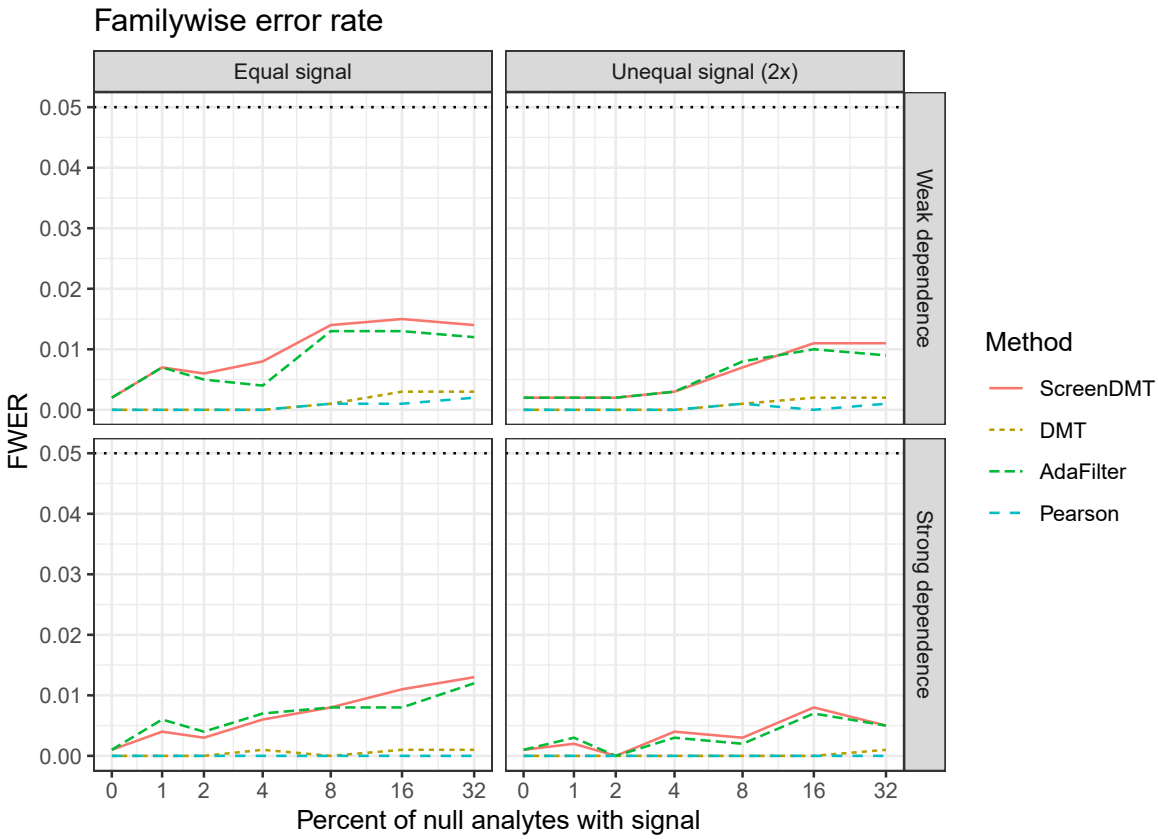

Familywise error rate
Equal signal
Unequal signal (2x)
Weak dependence
Strong dependence
FWER
0.05
0.04
0.03
0.02
0.01
0.00
0
1
2
4
8
16
32
Percent of null analytes with signal
Method
ScreenDMT
DMT
AdaFilter
Pearson

Figure S1

Text S1: ScreenDMT procedure and validation

Problem statement

Let ω_1 and ω_2 be two finite real valued parameters. In the mediation setting ω_1 and ω_2 represent the exposure-mediator and the mediator-outcome effect, while in the replication setting ω_1 and ω_2 represent parameters pertaining to the same biological entity measured in two studies. Whether we are interested in directional mediation or directional replication across two studies, without loss of generality we can consider the problem of detecting mediation in a direction consistent with the exposure increasing the outcome or replication in the same direction across two studies. Thus, we are interested in testing the directional null hypothesis:

$$H_0 : \omega_1\omega_2 \leq 0, \quad \text{against} \quad \bar{H}_0 : \omega_1\omega_2 > 0.$$

The null hypothesis can alternatively be written as:

$$H_0 : \omega_1 = 0 \text{ or } \omega_2 = 0 \text{ or } \operatorname{sign}(\omega_1\omega_2) = -1.$$

We assume that we have $\sqrt{n}_i$-consistent and asymptotically normal estimates $\hat{\omega}_i$ of ω_i for $i = 1, 2$ such that

$$\sqrt{n}_i(\hat{\omega}_i - \omega_i) \xrightarrow{d} N(0, \sigma_i^2)$$

where the convergence is convergence in distribution and $0 < \sigma_i < \infty$ for $i = 1, 2$, where σ_i is known. Then we can define component p-values P_1 and P_2 for testing $H_1 : \omega_1 = 0$ and $H_2 : \omega_2 = 0$, respectively,

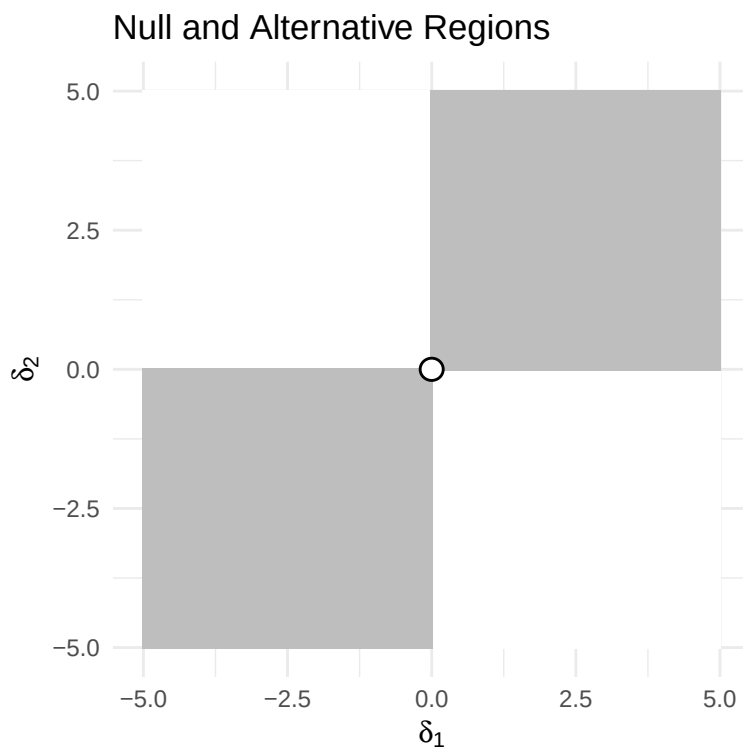
$$P_i = 2\Phi\left(-\frac{\sqrt{n}_i|\hat{\omega}_i|}{\sigma_i}\right) \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The procedures below are valid for large samples if consistent estimates of σ_i are used instead. We further assume that $\hat{\omega}_1$ and $\hat{\omega}_2$ are independent (i.e. $\hat{\omega}_1 \perp\!\!\!\perp \hat{\omega}_2$), a plausible assumption in both the mediation and the replication setting.

To simplify notation we let

$$\delta_i = \frac{\omega_i\sqrt{n}_i}{\sigma_i}$$

for $i = 1, 2$ so that Equation 1 becomes $P_i = 2\Phi\left(-|\hat{\delta}_i|\right)$. This is without loss of generality, since in any dataset $\sqrt{n}_i/\sigma_i$ is a constant coefficient of $\hat{\omega}_i$. The constant is always positive, so δ_i could replace ω_i in the problem statement. We assume ω_i and σ_i are fixed, so $|\delta_i| \to \infty$ implies $\delta_i \neq 0$ and $n_i \to \infty$. This problem's null (white) and alternative (grey) regions are shown in the figure below.



Directional MaxP test

Previous result

The following result given in Dreyfuss et al. (2021) regarding the directional MaxP test introduced in the Hitman method specifies how a p-value for H_0 can be obtained from both two-sided p-values associated to the two parameters under study.

Let P_1 and P_2 be the two valid component p-values for testing $H_1 : \omega_1 = 0$ and $H_2 : \omega_2 = 0$ and let $\bar{P} = \max\{P_1, P_2\}$ and $\underline{P} = \min\{P_1, P_2\}$. Further, let $\hat{S} = \mathbb{I}(\text{sign}(\hat{\omega}_1\hat{\omega}_2) = 1)$ indicate if the signs of the estimates $\hat{\omega}_1$ and $\hat{\omega}_2$ are the same, where $\mathbb{I}(\cdot)$ is the indicator function which is one if its argument is true and zero otherwise. Then

$$P_{DMT} = \left(\frac{\bar{P}}{2}\right)^{\hat{S}}$$

is a valid p-value for testing H_0.

We showed this by demonstrating that $\mathbb{P}(P_{DMT} \leq u) \leq u$ for $0 \leq u \leq 1$. The case of $u = 0$ is trivial and we identified that under the null hypothesis $\mathbb{P}(\hat{S} = 1) \leq \frac{1}{2}$ and then focused on the more challenging case of $0 < u < \frac{1}{2}$. We also showed that $\mathbb{P}(\hat{S} = 1) = \frac{1}{2}$ if and only (δ_1, δ_2) is on the boundary of the parameter space where $\delta_1 = 0 \cup \delta_2 = 0$. Since δ_1 and δ_2 are treated symmetrically, without loss of generality we considered the null region $H_0^\star : \delta_1 \leq 0 \cap \delta_2 \geq 0$. On the boundary of $H_0^\star$, we found that $\mathbb{P}(P_{DMT} \leq u | \delta_1 = 0, \delta_2 = 0) = 2u^2 < u$, whereas if one parameter is zero and the magnitude of the other tends to infinity (i.e. the parameter is nonzero and the sample size tends to infinity), we saw that $\mathbb{P}(P_{DMT} \leq u) \to u$, which is the least favorable null. Gail and Simon (1985) came to the same conclusion, which they proved using Schur-convexity. We also found that when $(\delta_1, \delta_2) \to (-\infty, \infty)$ then $\mathbb{P}(\hat{S} = 1) \to 0$ so $\mathbb{P}(P_{DMT} \leq u) \to 0$, which is the global minimum.

Proposition 1: The directional MaxP test is the Likelihood Ratio Test (LRT)

Proof

This problem is related to the problem of finding subsets of samples with effects of different signs (i.e. qualitative interactions, QI), whose LRT was shown by Gail and Simon (1985) and nicely presented in Hudson and Shojaie (2020). In that problem, $H_0^{QI} = \omega_1\omega_2 \geq 0$ is tested against $\bar{H}_0^{QI} : \omega_1\omega_2 < 0$, so the null region of the parameter space is $\Omega_0^{QI} = \{(\omega_1, \omega_2) : (\omega_1, \omega_2) \in H_0^{QI}\}$. This is also the form of the problem in directional mediation when the exposure decreases the outcome (Dreyfuss et al. 2021).

For our problem, consider applying a LRT based on the asymptotic sampling distribution of $(\hat{\omega}_1, \hat{\omega}_2)$. The LRT rejects H_0 for large values of:

$$\tau = -2 \ln \left[\frac{\max_{(v_1,v_2)\in\Omega_0} \prod_{i\in\{1,2\}} \phi\{\sqrt{n_i}(\hat{\omega}_i - v_i)/\sigma_i\}}{\max_{(u_1,u_2)\in\mathbb{R}^2} \prod_{i\in\{1,2\}} \phi\{\sqrt{n_i}(\hat{\omega}_i - u_i)/\sigma_i\}} \right]$$

where $\Omega_0 = \{(x_1, x_2) : (x_1, x_2) \in H_0\}$ is the null region of the parameter space ($\bar{\Omega}_0$ is the alternative region) and $\phi(\cdot)$ is the standard normal density function. The denominator is maximized at $u_i^\star = \hat{\omega}_i$ for both $i = 1, 2$. Then,

$$\begin{aligned} \tau &= -2 \max_{(v_1,v_2)\in\Omega_0} \sum_{i=1}^{2} -\frac{n_i}{2} \left(\frac{\hat{\omega}_i - v_i}{\sigma_i} \right)^2 \\ &= \min_{(v_1,v_2)\in\Omega_0} \sum_{i=1}^{2} n_i \left(\frac{\hat{\omega}_i - v_i}{\sigma_i} \right)^2 \end{aligned}$$

If $(\hat{\omega}_1, \hat{\omega}_2) \in \Omega_0$, then we will have $(v_1^\star, v_2^\star) = (\hat{\omega}_1, \hat{\omega}_2)$, yielding $\tau = 0$. So we next consider points under the alternative hypothesis. Without loss of generality, consider that $(\hat{\omega}_1, \hat{\omega}_2)$ is in quadrant I, where both parameters are non-negative. Then the closest (v_1, v_2) can be to $(\hat{\omega}_1, \hat{\omega}_2)$ is to be on the boundary of quadrant I, where for one of $i = 1$ or $i = 2$ we have that $v_i = \hat{\omega}_i$ and for the other $i, v_i = 0$. The minimization of the sum will select the study with the smallest $\hat{\delta}$:

$$\tau = \min_{i\in\{1,2\}} \hat{\delta}_i^2 \mathbb{I}[(\hat{\delta}_1, \hat{\delta}_2) \in \bar{\Omega}_0]$$

where the p-value of the LRT is computed as $\mathbb{P}(\chi_1^2 > \tau)/2$ if $\tau > 0$, otherwise the p-value is one. This is equivalent to computing the p-value as

$$\max_{i\in\{1,2\}} \mathbb{P}(\chi_1^2 > \hat{\delta}_i^2)/2$$

if $(\hat{\delta}_1, \hat{\delta}_2) \in \bar{\Omega}_0$ or otherwise one, which is asymptotically equivalent to the directional MaxP test, so the DMT is the LRT.

ScreenDMT: multiple testing of directional hypotheses with screening

When considering multiple hypotheses (e.g. for many analytes, lipids, genes), we propose the following two-step multiple testing procedure. The ScreenDMT procedure includes an additional step based on AdaFilter (Wang et al. 2022) to calculate the adjusted p-values.

1. Apply the directional MaxP test per potential analyte to obtain P_{DMT}.
2. Calculate adjusted p-values using AdaFilter with filtering p-value $\Lambda = \underline{P}$ and selection p-value $\Xi = \max\{P_{DMT}, \underline{P}\}$. This can be used for calculating adjusted p-values that control the FDR or the FWER.

Dreyfuss et al. (2021) provided an option in step 1 to calculate valid p-values of these parameters in omics data using the R package Limma (Ritchie et al. 2015), which has been found to improve power across omics datasets. The same option is available for ScreenDMT.

The key property that AdaFilter relies on is a form of conditional validity (Wang et al. 2022). Assuming that for each analyte P_1 is independent of P_2, the property states that when H_0 is true:

$$\mathbb{P}(\Xi < u | \Lambda < u) \leq u$$

holds for any fixed u > 0 whenever $\mathbb{P}(\underline{P} < u) > 0$.

Proposition 2: ScreenDMT satisfies conditional validity in asymptotic sample sizes

Proof

To demonstrate conditional validity, we must show that this holds on the boundary of the parameter space where $\omega_1 = 0 \cup \omega_2 = 0$ and off the boundary where the first partial derivatives are zero. We have,

$$\begin{aligned}
\mathbb{P}(\Xi < u | \Lambda < u) =& \mathbb{P}(\max\{P_{DMT}, \underline{P}\} < u | \underline{P} < u) \\
=& \mathbb{P}(P_{DMT} < u, \underline{P} < u | \underline{P} < u) \\
=& \mathbb{P}(P_{DMT} < u, \underline{P} < u) / \mathbb{P}(\underline{P} < u) \\
=& \mathbb{P}(P_{DMT} < u | \underline{P} < u)
\end{aligned}$$

For $u \geq 1$, we trivially have $\mathbb{P}(P_{DMT} < u | \underline{P} < u) \leq u$. We know that P_{DMT} and $\underline{P}$ are symmetric in ω_1 and ω_2, so without loss of generality we only consider the null region $H_0^\star : \omega_1 \leq 0 \cap \omega_2 \geq 0$. For $0 < u < 1$, using the law of total probability, we continue

$$\begin{aligned}
=& \mathbb{P}(P_{DMT} < u, \hat{S} = 1 | \underline{P} < u) + \mathbb{P}(P_{DMT} < u, \hat{S} = 0 | \underline{P} < u) \\
=& \mathbb{P}(P_{DMT} < u, \hat{S} = 1 | \underline{P} < u), \text{ since } \hat{S} = 0 \implies P_{DMT} = 1, \\
=& \mathbb{P}(\bar{P} < 2u, \hat{S} = 1 | \underline{P} < u) \\
=& \mathbb{P}(P_1 < 2u, P_2 < 2u, \hat{S} = 1 | P_1 < u \cup P_2 < u) \\
=& \frac{\mathbb{P}(P_1 < 2u, P_2 < 2u, P_1 < u \cup P_2 < u, \hat{S} = 1)}{\mathbb{P}(P_1 < u \cup P_2 < u)} \\
=& \frac{\mathbb{P}(P_1 < u, P_2 < 2u, \hat{S} = 1) + \mathbb{P}(P_1 < 2u, P_2 < u, \hat{S} = 1) - \mathbb{P}(P_1 < u, P_2 < u, \hat{S} = 1)}{\mathbb{P}(P_1 < u) + \mathbb{P}(P_2 < u) - \mathbb{P}(P_1 < u, P_2 < u)}.
\end{aligned} \tag{2}$$

For $u > \frac{1}{2}$ we know that $P_i < 2u$ for $i = 1, 2$, so Equation 2 becomes

$$\begin{aligned}
\mathbb{P}(P_{DMT} < u | \underline{P} < u) &= \mathbb{P}(\hat{S} = 1) \frac{\mathbb{P}(P_1 < u) + \mathbb{P}(P_2 < u) - \mathbb{P}(P_1 < u, P_2 < u)}{\mathbb{P}(P_1 < u) + \mathbb{P}(P_2 < u) - \mathbb{P}(P_1 < u, P_2 < u)}, \text{ as } \hat{S} \perp\!\!\!\perp (P_1, P_2), \\
&\leq \frac{1}{2} \\
&< u
\end{aligned}$$

where $\hat{S} \perp\!\!\!\perp (P_1, P_2)$ because P_i is calculated based on $|\hat{\omega}_i|$, which is independent of $\hat{S}$, and we showed the bound $\mathbb{P}(\hat{S} = 1) \leq \frac{1}{2}$ previously (Dreyfuss et al. 2021).

We now consider $\mathbb{P}(P_{DMT} < u | \underline{P} < u)$ under $H_0^\star$ for $0 < u \leq \frac{1}{2}$ both on the boundary of the parameter space where $\omega_1 = 0 \cup \omega_2 = 0$ and then over the whole parameter space.

Boundary of parameter space

Under $H_0^\star$ for $0 < u \leq \frac{1}{2}$ when $\omega_1 = 0 \cup \omega_2 = 0$ we know that $\mathbb{P}(\hat{S} = 1 | \omega_1 = 0 \cup \omega_2 = 0) = \frac{1}{2}$ and $\mathbb{P}(P_i < u | \omega_i = 0) = u$. Without loss of generality, let $\omega_2 = 0$ and $\omega_1 \leq 0$, so $\mathbb{P}(P_2 < u | \omega_2 = 0) = u$ and $\mathbb{P}(P_1 < u | \omega_1 \leq 0) = F(u) \geq u$. For these parameter values, Equation 2 becomes

$$\begin{aligned}
\mathbb{P}(P_{DMT} < u | \underline{P} < u, \omega_2 = 0) &= u \frac{2F(u) + F(2u) - F(u)}{2[u + F(u) - uF(u)]} \\
&= u \frac{F(u) + F(2u)}{2[u + F(u) - uF(u)]} \\
&\to u, \text{ as } F(u) \to 1 \\
&\text{and} \\
\mathbb{P}(P_{DMT} < u | \underline{P} < u, \omega_1 = \omega_2 = 0) &= \frac{3u}{2(2-u)} \\
&\leq u, \text{ for } 0 < u \leq \frac{1}{2}.
\end{aligned}$$

$F(u) \to 1$ for $\delta_2 = 0$ as $\delta_1 \to -\infty$ as per Equation 1 which occurs as $n_1 \to \infty$.

Whole parameter space

We continue Equation 2 for $0 < u \leq \frac{1}{2}$ using the definition of our component p-values from Equation 1,

$$\begin{aligned}
&\mathbb{P}(P_{DMT} < u | \underline{P} < u) \\
=&\Big\{ \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_1|\right) < u, 2\Phi\left(-|\hat{\delta}_2|\right) < 2u, \hat{S} = 1\right) + \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_1|\right) < 2u, 2\Phi\left(-|\hat{\delta}_2|\right) < u, \hat{S} = 1\right) \\
&- \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_1|\right) < u, 2\Phi\left(-|\hat{\delta}_2|\right) < u, \hat{S} = 1\right) \Big\} / \\
&\Big\{ \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_1|\right) < u\right) + \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_2|\right) < u\right) - \mathbb{P}\left(2\Phi\left(-|\hat{\delta}_1|\right) < u, 2\Phi\left(-|\hat{\delta}_2|\right) < u\right) \Big\} \\
=&\Big\{ \mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|\hat{\delta}_2| < \Phi^{-1}(u), \hat{S} = 1\right) + \mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}(u), -|\hat{\delta}_2| < \Phi^{-1}\left(\frac{u}{2}\right), \hat{S} = 1\right) - \\
&\mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|\hat{\delta}_2| < \Phi^{-1}\left(\frac{u}{2}\right), \hat{S} = 1\right) \Big\} / \\
&\Big\{ \mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right)\right) + \mathbb{P}\left(-|\hat{\delta}_2| < \Phi^{-1}\left(\frac{u}{2}\right)\right) - \mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|\hat{\delta}_2| < \Phi^{-1}\left(\frac{u}{2}\right)\right) \Big\}.
\end{aligned} \tag{3}$$

We calculate these probabilities according to the law of total probability by integrating over the realized values (x_1, x_2) of the random estimates $(\hat{\omega}_1, \hat{\omega}_2)$. This involves the joint density

$$f\left(\hat{\delta}_1 = x_1, \hat{\delta}_2 = x_2\right) = \phi\left(x_1 - \delta_1\right) \phi\left(x_2 - \delta_2\right)$$

which follows from $\hat{\delta}_1 \perp\!\!\!\perp \hat{\delta}_2$, where $\phi(\cdot)$ is the standard normal density. For example, for the first term in the numerator of Equation 3 with fixed δ_i for $i = 1, 2$,

$$
\begin{aligned}
&\mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|\hat{\delta}_2| < \Phi^{-1}(u), \hat{S} = 1\right) \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}\left(-|\hat{\delta}_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|\hat{\delta}_2| < \Phi^{-1}(u), \hat{S} = 1 \middle| \hat{\delta}_1 = x_1, \hat{\delta}_2 = x_2\right) f\left(\hat{\delta}_1 = x_1, \hat{\delta}_2 = x_2\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}\left(-|x_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|x_2| < \Phi^{-1}(u), \hat{S} = 1\right) \phi\left(x_1 - \delta_1\right) \phi\left(x_2 - \delta_2\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}\left(-|x_1| < \Phi^{-1}\left(\frac{u}{2}\right), -|x_2| < \Phi^{-1}(u), \hat{S} = 1\right) \phi\left(x_1 - \delta_1\right) \phi\left(x_2 - \delta_2\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
=&\int_{-\Phi^{-1}\left(\frac{u}{2}\right)}^{\infty} \phi\left(x_1 - \delta_1\right) \mathrm{d}x_1 \int_{-\Phi^{-1}(u)}^{\infty} \phi\left(x_2 - \delta_2\right) \mathrm{d}x_2 + \int_{-\infty}^{\Phi^{-1}\left(\frac{u}{2}\right)} \phi\left(x_1 - \delta_1\right) \mathrm{d}x_1 \int_{-\infty}^{\Phi^{-1}(u)} \phi\left(x_2 - \delta_2\right) \mathrm{d}x_2 \\
=&\lim_{\psi\to\infty}\left\{\Phi\left(x_1 - \delta_1\right)\Big|_{-\Phi^{-1}\left(\frac{u}{2}\right)}^{\psi} \Phi\left(x_2 - \delta_2\right)\Big|_{-\Phi^{-1}(u)}^{\psi} + \Phi\left(x_1 - \delta_1\right)\Big|_{-\psi}^{\Phi^{-1}\left(\frac{u}{2}\right)} \Phi\left(x_2 - \delta_2\right)\Big|_{-\psi}^{\Phi^{-1}(u)}\right\} \\
=&\left[1 - \Phi\left(-\Phi^{-1}(u/2) - \delta_1\right)\right]\left[1 - \Phi\left(-\Phi^{-1}(u) - \delta_2\right)\right] + \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u) - \delta_2\right) \\
=&\Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\Phi\left(\Phi^{-1}(u) + \delta_2\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u) - \delta_2\right), \text{ since } \Phi(-x) = 1 - \Phi(x).
\end{aligned}
$$

The fourth line holds because (x_1, x_2) are realizations of random variables, rather than being random themselves, so the probabilities evaluate to zero where their event is false and one where it is true. Following this example, we can evaluate Equation 3,

$$
\begin{aligned}
&\mathbb{P}(P_{DMT} < u | \underline{P} < u) \\
=&\big\{\Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\Phi\left(\Phi^{-1}(u) + \delta_2\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u) - \delta_2\right) + \\
&\Phi\left(\Phi^{-1}(u) + \delta_1\right)\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) + \Phi\left(\Phi^{-1}(u) - \delta_1\right)\Phi\left(\Phi^{-1}(u/2) - \delta_2\right) - \\
&\Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\big\}/ \\
&\big\{\Phi\left(\Phi^{-1}(u/2) + \delta_1\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_1\right) + \Phi\left(\Phi^{-1}(u/2) + \delta_2\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_2\right) - \\
&\Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u/2) - \delta_2\right) - \\
&\Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\big\} \\
=&\big\{\Phi\left(\Phi^{-1}(u/2) + \delta_1\right)\left[\Phi\left(\Phi^{-1}(u) + \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) + \delta_2\right)\right] + \\
&\Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\left[\Phi\left(\Phi^{-1}(u) - \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\right] + \\
&\Phi\left(\Phi^{-1}(u) + \delta_1\right)\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) + \Phi\left(\Phi^{-1}(u) - \delta_1\right)\Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\big\}/ \\
&\big\{\left[\Phi\left(\Phi^{-1}(u/2) + \delta_1\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_1\right)\right]\left[1 - \Phi\left(\Phi^{-1}(u/2) + \delta_2\right) - \Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\right] + \\
&\Phi\left(\Phi^{-1}(u/2) + \delta_2\right) + \Phi\left(\Phi^{-1}(u/2) - \delta_2\right)\big\}.
\end{aligned} \tag{4}
$$

First partial derivative To see where the maximum of Equation 4 may be for $0 < u \leq \frac{1}{2}$, we first evaluate where

$$
\frac{\partial \mathbb{P}(P_{DMT} < u | \underline{P} < u)}{\partial \delta_1} = 0.
$$

$$
\begin{aligned}
&\frac{\partial \mathbb{P}(P_{DMT} < u | \underline{P} < u)}{\partial \delta_1} \\
=&\{\phi\left(\Phi^{-1}(u/2)+\delta_1\right)\left[\Phi\left(\Phi^{-1}(u)+\delta_2\right)-\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)\right]- \\
&\phi\left(\Phi^{-1}(u/2)-\delta_1\right)\left[\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)-\Phi\left(\Phi^{-1}(u)-\delta_2\right)\right]+ \\
&\phi\left(\Phi^{-1}(u)+\delta_1\right)\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)-\phi\left(\Phi^{-1}(u)-\delta_1\right)\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\}* \\
&\{\left[\Phi\left(\Phi^{-1}(u/2)+\delta_1\right)+\Phi\left(\Phi^{-1}(u/2)-\delta_1\right)\right]\left[1-\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)-\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\right]+ \\
&\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)+\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\}- \\
&\{\Phi\left(\Phi^{-1}(u/2)+\delta_1\right)\left[\Phi\left(\Phi^{-1}(u)+\delta_2\right)-\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)\right]+ \\
&\Phi\left(\Phi^{-1}(u/2)-\delta_1\right)\left[\Phi\left(\Phi^{-1}(u)-\delta_2\right)-\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\right]+ \\
&\Phi\left(\Phi^{-1}(u)+\delta_1\right)\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)+\Phi\left(\Phi^{-1}(u)-\delta_1\right)\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\}* \\
&\{\left[\phi\left(\Phi^{-1}(u/2)+\delta_1\right)-\phi\left(\Phi^{-1}(u/2)-\delta_1\right)\right]\left[1-\Phi\left(\Phi^{-1}(u/2)+\delta_2\right)-\Phi\left(\Phi^{-1}(u/2)-\delta_2\right)\right]\}.
\end{aligned} \tag{5}
$$

For fixed parameters off the boundary under $H_0^\star$ (i.e. $\delta_1 < 0, \delta_2 > 0$), and $0 < x \leq \frac{1}{2}$, we have the following limits for $i = 1, 2$ as the sample size (and thus the magnitude of δ_i) grows without bound,

$$
\begin{aligned}
\lim_{|\delta_i|\to\infty} \Phi\left(\Phi^{-1}(x)+|\delta_i|\right) = 1 \text{ and } \lim_{|\delta_i|\to\infty} \Phi\left(\Phi^{-1}(x)-|\delta_i|\right) = 0 \\
\lim_{|\delta_1|\to\infty} \phi\left(\Phi^{-1}(x)+\delta_1\right) = \lim_{|\delta_1|\to\infty} \phi\left(\Phi^{-1}(x)-\delta_1\right) = 0 \\
\lim_{|\delta_1|\to\infty} \left[\phi\left(\Phi^{-1}(u/2)+\delta_1\right)-\phi\left(\Phi^{-1}(u/2)-\delta_1\right)\right] = 0
\end{aligned} \tag{6}
$$

The second line in Equation 6 can be seen to hold based on the leading terms, which grow as $\exp\left\{-\delta_1^2/2\right\}$. The third line in Equation 6 is proportional to

$$
\begin{aligned}
&\exp\left\{-\frac{1}{2}\left[\Phi^{-1}(u/2)+\delta_1\right]^2\right\} - \exp\left\{-\frac{1}{2}\left[\Phi^{-1}(u/2)-\delta_1\right]^2\right\} \\
=&\exp\left\{-\frac{1}{2}\left[(\Phi^{-1}(u/2))^2+2\Phi^{-1}(u/2)\delta_1+\delta_1^2\right]\right\} - \exp\left\{-\frac{1}{2}\left[(\Phi^{-1}(u/2))^2-2\Phi^{-1}(u/2)\delta_1+\delta_1^2\right]\right\} \\
\propto&\exp\left\{-\delta_1\left[\delta_1/2+\Phi^{-1}(u/2)\right]\right\} - \exp\left\{-\delta_1\left[\delta_1/2+\Phi^{-1}(u/2)\right]\right\}
\end{aligned}
$$

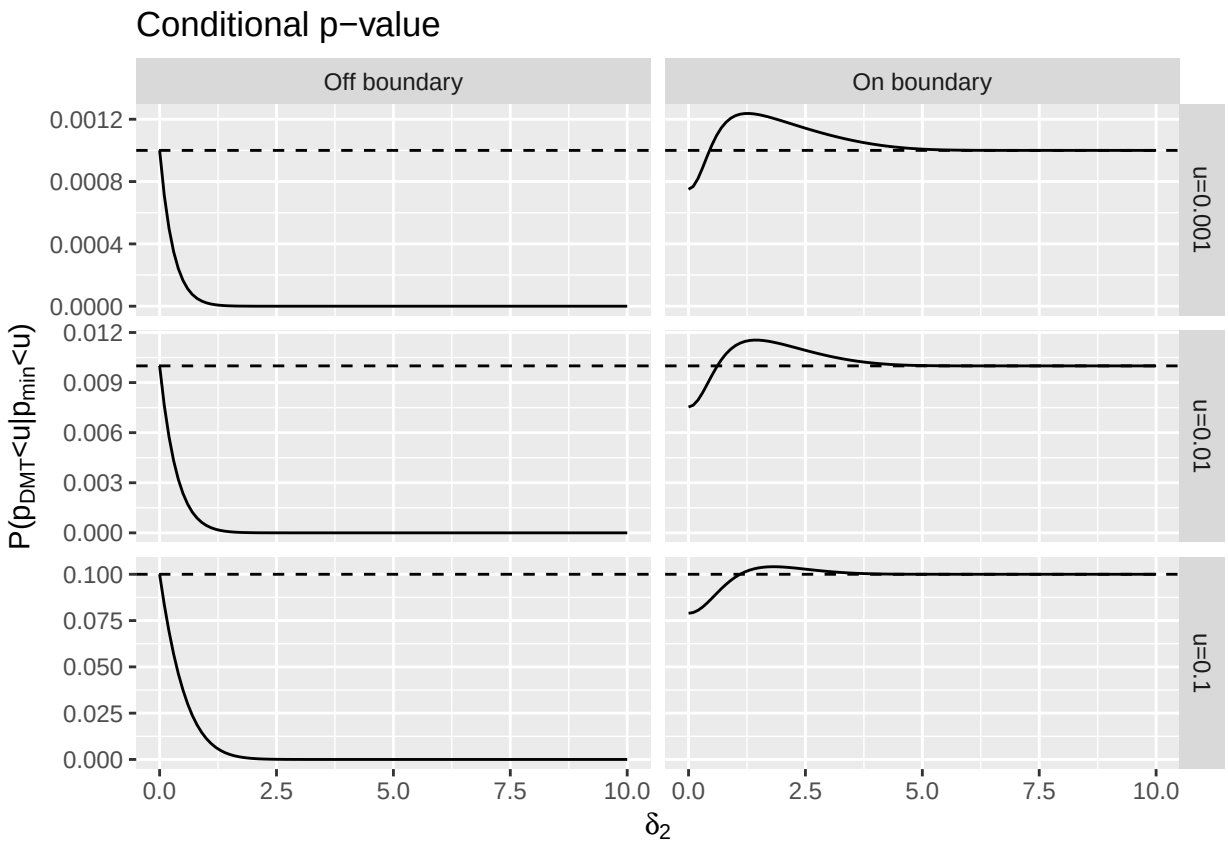
which is also dominated by $\exp\left\{-\delta_1^2/2\right\}$ as $|\delta_i| \to \infty$ or equivalently as $n \to \infty$. Given the limits in Equation 6, for parameters off the boundary under $H_0^\star$ each term in Equation 5 goes to zero as the sample size increases by the product limit rule.

Thus off the boundary, $\frac{\partial \mathbb{P}(P_{DMT}<u|\underline{P}<u)}{\partial \delta_1} \to 0$ when $n_1 \to \infty$ and by symmetry of $\mathbb{P}(P_{DMT} < u|\underline{P} < u)$, also $\frac{\partial \mathbb{P}(P_{DMT}<u|\underline{P}<u)}{\partial \delta_2} \to 0$. Thus, off the boundary both partial derivatives go to zero when $(n_1, n_2) \to (\infty, \infty)$ and $\mathbb{P}(P_i < u) \to 1$ for $i = 1, 2$ because of Equation 1, so $\mathbb{P}(\underline{P} < u) \to 1$. Whereas $\mathbb{P}(P_{DMT} < u|\underline{P} < u) \to \mathbb{P}(\hat{S} = 1)$, and $\mathbb{P}(\hat{S} = 1) \to 0$ as $(n_1, n_2) \to (\infty, \infty)$, so $\mathbb{P}(P_{DMT} < u|\underline{P} < u) \to 0$ and this is the global minimum. However, on the boundary, we have shown that $\mathbb{P}(P_{DMT} < u|\underline{P} < u) \to u$ when only one parameter is zero as the sample size grows unbounded, which is the asymptotic maximum under $H_0^\star$, which completes the proof.

Conditional p-value in finite sample sizes

We have the closed form expression of the conditional p-value $\mathbb{P}(P_{DMT} < u|\underline{P} < u)$ in Equation 4 for finite sample sizes, so we plot it below. We see that when δ_1 is off the boundary in the left panels (here we show $\delta_1 = -10$) and $\delta_2 > 0$, conditional validity is easily satisfied, since the curve is far below the dashed line at u. When $\delta_1 = 0$ is on the boundary in the right panels and δ_2 takes on some small values there is slight

inflation but conditional validity holds when δ_2 is near zero and when $\delta_2 > 0$ for large sample sizes, which is concordant with our mathematical proof.



References

Dreyfuss, J M, Y Yuchi, X Dong, V Efthymiou, H Pan, D C Simonson, A Vernon, et al. 2021. "High-Throughput Mediation Analysis of Human Proteome and Metabolome Identifies Mediators of Post-Bariatric Surgical Diabetes Control." *Nature Communications* 12 (1): 6951. https://doi.org/10.1038/s41467-021-27289-2.

Gail, M, and R Simon. 1985. "Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets." *Biometrics* 41 (2): 361–72. https://doi.org/10.2307/2530862.

Hudson, Aaron, and Ali Shojaie. 2020. "Statistical Inference for Qualitative Interactions with Applications to Precision Medicine and Differential Network Analysis." *arXiv*. https://doi.org/10.48550/arxiv.2010.08703.

Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Res* 43 (7): e47. https://doi.org/10.1093/nar/gkv007.

Wang, Jingshu, Lin Gui, Weijie J. Su, Chiara Sabatti, and Art B. Owen. 2022. "Detecting Multiple Replicating Signals Using Adaptive Filtering Procedures." *The Annals of Statistics* 50 (4). https://doi.org/10.1214/21-AOS2139.