
Faculty Work Comprehensive List

2-2015

Novel Approach to Identify Optimal Metabotypes of Elongase and Desaturase Activities in Prevention of Acute Coronary Syndrome

Nathan L. Tintle

Dordt College, nathan.tintle@dordt.edu

John W. Newman

University of California - Davis

Gregory C. Shearer

Pennsylvania State University - Main Campus

Follow this and additional works at: https://digitalcollections.dordt.edu/faculty_work

 Part of the [Cardiovascular Diseases Commons](#), and the [Genetics and Genomics Commons](#)

Recommended Citation

Tintle, N. L., Newman, J. W., & Shearer, G. C. (2015). Novel Approach to Identify Optimal Metabotypes of Elongase and Desaturase Activities in Prevention of Acute Coronary Syndrome. *Metabolomics*
<https://doi.org/10.1007/s11306-015-0787-6>

This Article is brought to you for free and open access by Dordt Digital Collections. It has been accepted for inclusion in Faculty Work Comprehensive List by an authorized administrator of Dordt Digital Collections. For more information, please contact ingrid.mulder@dordt.edu.

Novel Approach to Identify Optimal Metabotypes of Elongase and Desaturase Activities in Prevention of Acute Coronary Syndrome

Abstract

Both metabolomic and genomic approaches are valuable for risk analysis, however typical approaches evaluating differences in means do not model the changes well. Gene polymorphisms that alter function would appear as distinct populations, or metabotypes, from the predominant one, in which case risk is revealed as changed mixing proportions between control and case samples. Here we validate a model accounting for mixed populations using biomarkers of fatty acid metabolism derived from a case/control study of acute coronary syndrome subjects in which both metabolomic and genomic approaches have been used previously. We first used simulated data to show improved power and sensitivity in the approach compared to classic approaches. We then used the metabolic biomarkers to test for evidence of distinct metabotypes and different proportions among cases and controls. In simulation, our model outperformed all other approaches including Mann-Whitney, *t*-tests, and χ^2 . Using real data, we found distinct metabotypes of six of the seven activities tested, and different mixing proportions in five of the six activity biomarkers: D9D, ELOVL6, ELOVL5, FADS1, and Sprecher pathway chain shortening (SCS). High activity metabotypes of non-essential fatty acids and SCS decreased odds for acute coronary syndrome (ACS), however high activity metabotypes of 20-carbon fatty acid synthesis increased odds. Our study validates an approach that accounts for both metabolomic and genomic theory by demonstrating improved sensitivity and specificity, better performance in real world data, and more straightforward interpretability.

Keywords

acute coronary syndrome, metabolomic, genomic, likelihood ratio test, metabotype estimation, fatty acid activity indices

Disciplines

Cardiovascular Diseases | Genetics and Genomics

Comments

- This is a pre-publication author manuscript of the final, published article. The definitive version is published by Springer and available from <http://dx.doi.org/10.1007/s11306-015-0787-6>
- DOI: 10.1007/s11306-015-0787-6
- Copyright © 2015 by Springer Science+Business Media New York

In preparation for submission to: *Metabolomics*

Title: A novel approach to identify optimal metabotypes of elongase and desaturase activities in prevention of acute coronary syndrome

Authors: Nathan L. Tintle¹, John W. Newman^{2, 3}, Gregory C. Shearer, PhD⁴

Affiliations: ¹Dept of Statistics, Dordt College, Iowa; ²Western Human Nutrition Research Center, United States Department of Agriculture; ³Department of Nutrition, UC Davis, University of California, Davis CA 95616; ⁴Dept. of Nutritional Sciences, The Pennsylvania State University

Abbreviated Title: Mixture modeling identifies optimal metabolism

Word Count: 5,395

Abstract: 232

Tables: 1

Figures: 3

Supplemental Tables: 7

Supplemental Figs: 2

Corresponding Author: Gregory C. Shearer
110 Chandlee Laboratory
Dept. of Nutritional Sciences
The Pennsylvania State University
University Park, PA 16801

Tel: 814 867-3040

E-mail: gcs13@psu.edu

Significance:

This study formulates a novel test that better reflects how distributions would appear in samples having variance resulting from both genetic and environmental sources. The means-testing hypothesis is replaced with a test for different mixing proportions in multi-modal distributions. The study demonstrates superior test performance with simulated data. Estimates of enzyme activity from real world data are demonstrated to better conform to the hypothesized multi-modal distributions and differences between control subjects and those with acute coronary syndrome are used to infer risk.

Abstract

Both metabolomic and genomic approaches are valuable for risk analysis, however typical approaches evaluating differences in means do not model the changes well. Gene polymorphisms that alter function would appear as distinct populations, or metabotypes, from the predominant one, in which case risk is revealed as changed mixing proportions between control and case samples. Here we validate a model accounting for mixed populations using biomarkers of fatty acid metabolism derived from a case/control study of acute coronary syndrome subjects in which both metabolomic and genomic approaches have been used previously. We first used simulated data to show improved power and sensitivity in the approach compared to classic approaches. We then used the metabolic biomarkers to test for evidence of distinct metabotypes and different proportions among cases and controls. In simulation, our model outperformed all other approaches including Mann-Whitney, t -tests, and χ^2 . Using real data, we found distinct metabotypes of six of the seven activities tested, and different mixing proportions in five of the six activity biomarkers: D9D, ELOVL6, ELOVL5, FADS1, and Sprecher pathway chain shortening (SCS). High activity metabotypes of non-essential fatty acids and SCS decreased odds for acute coronary syndrome (ACS), however high activity metabotypes of 20-carbon fatty acid synthesis increased odds. Our study validates an approach that accounts for both metabolomic and genomic theory by demonstrating improved sensitivity and specificity, better performance in real world data, and more straightforward interpretability.

Introduction:

Improving risk prediction is increasingly important for early intervention and prevention. Several well known risk prediction algorithms are employed – such as the Prospective Cardiovascular Munster (PROCAM) study(1), the 3rd Joint European Task Force [Systematic Coronary Risk Evaluation (SCORE)](2), the Atherosclerosis Risk in Communities (ARIC) study(3), the Reynolds Risk Score(4, 5) and finally, the original and most widely used system, the Framingham Risk Score(6, 7) The last was designed to predict the 10-year risk for major coronary events, and it does so with a c-statistic [area under the receiver operating characteristic (ROC) curve] of 0.7-0.8(3, 6, 7). All of these prediction algorithms generally include the following standard risk factors: age, sex, total (or low-density lipoprotein) cholesterol (C), high-density lipoprotein C (HDL-C), blood pressure, and both smoking and diabetes status. Despite the utility of standard factors in coronary heart disease (CHD) risk estimation, there remains an intense interest in finding additional markers that would improve upon this standard(8-10); however, while a number of putative risk factors have been tested, few have added meaningfully(11, 12). Surprisingly, a relationship between leading gene variants related to classical risk markers for acute coronary syndrome was not validated in the data used here (13), demonstrating the challenges in this endeavor. To date such analyses have rarely considered fatty acid metabolism, however we have demonstrated that models based on the erythrocyte fatty-acid profiles can outperform classic risk models in case/control discrimination (14). Here we considered how we could employ an approach that jointly emphasizes the impact of genetic variance on metabolomic datasets.

A typical approach to comparison of phenotypes across case-control populations is to compare the component means using methods like the two-sample *t*-test (with or without pooled variance) and generalizations of these approaches in a regression framework. In case/control settings this translates into testing for evidence of a shift in the means between the two components, with the hypothesis that $\delta \neq 0$, where $\mu_{\text{Case}} = \mu_{\text{Control}} + \delta$ (illustrated in Figure 1A). However considering the metabolic mechanisms that give rise to the relative abundances of fatty acids – *e.g.* genes, diet, and lifestyle patterns – testing for a difference in means may not be optimal, since what appears to be a shift in means could actually be a shift in the relative abundance of distinct *metabotypes*, or metabolic phenotypes in the study case vs. control populations. While the term is most commonly applied to arrays comprising multiple metabolites that together represent a “readout of the metabolic state”(15), here we consider the concept of a metabotype as a phenotypic variation within a single biomarker of enzymatic behavior, or activity. Thus, functionally distinct genotypes should give rise to distinct metabotypes in large population samples. For example, delta-9 desaturase (D9D) is a critical step in desaturation of 16-carbon and 18-carbon fatty acids. It is

critical in the synthesis of palmitoleic and oleic acids from palmitic and stearic acids respectively. *In vivo*, this activity can be estimated by the palmitoleic to palmitic acid or the oleic acid: stearic acid ratio (OA:SA)(16). In a population with two D9D genotypes, having unique activities (k_m , b_{max}), each genotype would define a unique OA:SA mean with variance around the mean determined by diet, lifestyle and other environmental factors. Figure 1B shows how two D9D populations with unique means (μ) and variance (σ^2) would appear in a population sample.

In this case, a simple comparison of the means is likely to ignore the underlying biological cause: that the observed (marginal) distribution of fatty-acid levels in the case and control samples, respectively, represent a mixture of two or more underlying distributions, or components, derived from each genetically determined metabotype.

In this paper we propose a likelihood ratio test (LRT) method for testing case-control metabotype differences. The test estimates the metabotype mean of each sample, and then tests for differences in the relative frequencies of the metabolotypes in the case-control samples. We compare the proposed LRT method to means testing on the marginal distributions of FA levels and a chi-squared (χ^2) test using the estimated metabotype frequencies and find that the LRT is most powerful across a wide variety of situations. We then apply our methodology to the previously analyzed case-control study of 767 ACS-cases and their matched controls(14) to determine whether multiple populations of FA elongase and desaturases can be identified and if so whether the relative abundance of these populations differ between cases and controls. Numerous differences in cases and controls metabotype frequencies are identified and the resulting insights enhance biological interpretability.

Methods:

Summary of approach: We first validated the proposed test using simulated data, then applied it to an established fatty acid dataset where conclusions regarding risk for ACS could be drawn. In validating the novel statistical approach (LRT), two objectives were considered: 1) demonstrating that the test yields an accurate type I (α ; false positive) error rate; and 2) that when an effect is present, the test finds the effect ($1-\beta$ (power); true positive). Since the proposed test more plausibly conforms to biological mechanisms than the current standard, our primary endpoint was to demonstrate power equivalent to current approaches. As a secondary endpoint, we sought to demonstrate that the test outperforms current approaches in some situations.

In applying the test to real world data, we used a fatty acid dataset where prior knowledge of its usefulness existed (14). Here, the LRT test was used to answer two questions: 1) do multiple components,

also contextually referred to as metabolotypes, exist in the sampled populations; 2) is there a difference between the mixing proportions in the control and ACS populations; or more simply: does one metabolotype appear to confer risk for ACS by virtue of greater prevalence among ACS subjects. Although many factors might explain why there are multiple metabolotypes in a sample population, consideration of how genetics-induced differences in enzymatic function might appear in a lipidomic analysis of large cohorts guided our development of this approach. Thus we also utilized genetic models of inheritance (dominant, recessive and additive where appropriate to determine whether the observations of metabolotypes conformed to it.

Ethical Statements and compliance: This research was performed in accordance with the ethical principles for medical research involving human subjects outlined in the Declaration of Helsinki. The study was approved by the Institutional Review Board of Saint Luke's Hospital and the Institutional Review Board of the University of Missouri-Kansas City School of Medicine.

Subject Selection. Patients diagnosed with an ACS were recruited as previously described. A total of 1,661 patients were enrolled and samples obtained. Control subjects having routine blood draws were recruited as described from the same location as >80% of the cases.

Laboratory Methods: RBC-FA composition was measured as previously described(17) using the boron trifluoride methylation method and followed by gas chromatography (GC) analysis using an Agilent 6890 (Agilent Technologies, Palo Alto, CA) equipped with a capillary column (SP2560, 100m., Supelco, Bellefonte, PA).

Activity Biomarkers: We used product:precursor ratios in pathways of human fatty acid metabolism to obtain estimates of bioactivity (Supplemental Figure 1). A summary of each activity biomarker is listed in supplementary figure 1b. Since our intention was to develop an analytical approach combining both genetic and metabolomic considerations, we reference the activity estimates by the predominant human gene responsible for the activity however as noted, there could be other explanations for the presence of multiple components.

Test validation via simulation: To evaluate the performance of different statistical tests for case-control differences in FA levels, we conducted two main simulation studies: 1) data simulated from a two component mixture model; 2) data simulated from a three component mixture model.

In the simulation analysis involving a two-component mixture model we simulated two normal distributions, with a one unit difference in means between the two components. We considered standard deviations in the first group of either $s_1=0.25$ or 0.50 , and standard deviations in the second component

(s_2) either the same, 50% larger or 100% larger than the first component. We simulated 800 cases and 800 controls from the two component normal mixture distribution, where the mixing proportions (p_1 and p_2) may be the same (no association between genotype and disease) or different (association between genotype and disease) in the cases and controls. The proportion of controls in component 1 was simulated to follow either a dominant or recessive genetic mode of inheritance. In particular, we considered risk (minor) allele frequencies of $f=5\%$, 10% , 25% , 33% and 50% . Under an assumption of Hardy-Weinberg Equilibrium, this yields the following proportions of controls in component 2 ($p_{2,controls}$; Dominant: 9.75% , 19% , 43.75% , 55.1% and 75% ; Recessive: 0.25% , 1% , 6.25% , 10.9% and 25% , where these proportions are computed as $p_{2,controls,dominant} = f^2 + 2f(1-f)$ and $p_{2,controls,recessive} = f^2$. The proportion of cases in component 1, $p_{1,cases}$ was set to either be the same, 90% , 75% or 50% the size of $p_{1,controls}$, representing a risk allele with varying amounts of impact on disease risk. Thus, there were a total of 240 simulation settings ($2 (s_1) \times 3 (s_2) \times 10 (p_{2,controls}) \times 4 (p_{2,cases})$) considering a two-component normal distribution. Five-hundred randomly generated sets of data were created for each simulation setting. Empirical type I error and power were both computed as the proportion of the 500 data sets for which the p-value was less than 0.05.

We also simulated data from a three component normal mixture model. We considered the following settings. All simulation settings used a 1-unit difference between each of the three means of the components. The standard deviation for the first component was either 0.25 or 0.50, while the standard deviations of the remaining components were either the same, or different (50% larger in second, 100% larger in third). We simulated 800 cases and 800 controls. Similar to the two-component mixture simulation, we considered risk (minor) allele frequencies of $f=5\%$, 10% , 25% , 33% and 50% . The size of the three components in the controls was again determined using Hardy-Weinberg equilibrium. Case proportions in components one, two and three were computed to represent an additive, increasing risk model. In particular, $Xg_1, (1-X)g_1 + Xg_2, (1-X)g_2 + g_3$, are the three case component proportions respectively, where $X=1, 0.9, 0.75$ or 0.5 , and g_1, g_2 and g_3 represent the component proportions in the controls. Thus, there were a total of 80 simulation settings ($2 (s_1) \times 2 (s_2) \times 5 (p_{2,controls}) \times 4 (p_{2,cases})$) using a three-component normal distribution.

Methods of analyzing metabotype data: When analyzing simulated and real data, we used a variety of methods of analyzing fatty acid levels. To aid in the understanding of these approaches we begin by introducing some notation. Let G_i represent a random variable indicating the observed fatty acid level when an individual is truly in the i^{th} metabotype. We assume that $G_i \sim N(\mu_i, \sigma_i)$, where μ_i, σ_i indicate the mean and standard deviation of the normally distributed fatty acid levels within the i^{th} metabotype.

Then, $X \sim \sum_{i=1}^k p_{x,i} G_i$, where X represents the observed distribution of fatty levels within the case

population, and $p_{x,i}$ is the proportion of cases with the i th metabotype, where $\sum_{i=1}^k p_{x,i} = 1$ and k indicates the number of distinct metabotypes. Similarly, $Y \sim \sum_{i=1}^k p_{y,i} G_i$.

Standard methods of evaluating differences in X and Y include the t -test, which uses $t = \frac{\bar{x} - \bar{y}}{s_{x,y}}$, where \bar{x}, \bar{y} are the sample means of the cases and controls, and $s_{x,y}$, is a function of the sample standard deviations (s_x, s_y) and sample sizes (n_x, n_y) in the case and control samples, respectively, and can be estimated using assumptions of equal or unequal variance. A common non-parametric alternative is the Mann-Whitney test, which first computes the ranks, r_j , of the $n_x + n_y$ combined sample of fatty acid levels, and then evaluates the difference between the average rank of the case and control samples (\bar{r}_x, \bar{r}_y) and the expected average rank if the samples were from the same population.

We consider two methods here which reflect the underlying assumption of observed samples which are mixtures of different metabotypes. The first method using metabotypes is a χ^2 test of independence can be used on the resulting $2 \times k$ table of categorized individuals which occurs from cross-classifying individuals by disease status and estimated metabotype. Metabotypes were estimated by using MClust (18) to fit a 2 or 3 component mixture distributions using an Expectation-Maximization algorithm. For a given mixture model, each individual can be assigned a vector of posterior probabilities corresponding to the individuals likelihood of belonging to each component of the mixture distribution (see details in the following section). An individual can then be classified to the “most likely” metabotype by using the maximum posterior probability. The second method (19) does not require the use of maximum posterior probabilities to classify individuals into metabotypes. Instead, Kim et al. use a likelihood ratio test (LRT) to compare the full distribution of posterior metabotype probabilities between cases and controls. Two different versions of the LRT were considered: (1) one where the number of distinct metabotypes (k) was considered known *a priori* (LRT_k); (2) one where the number of distinct metabotypes (k) was estimated from the data (LRT_{est}).

Statistical analysis of simulated data: Data was simulated as described earlier using R version 3.0.3. Each of the different methods of analyzing fatty acid data (see *Methods of analyzing metabotype data*) was then applied to the data yielding a p-value measuring the strength of evidence of a difference in the means, distributions, or mixing proportions of metabotypes between cases and controls, depending on the test used. Type I error rates and power were estimated by dividing the number of simulations yielding a p-value less than 0.05 by the total number of simulations at that setting. For the LRT test, a single component represents a case where no difference exists.

Statistical analysis of real data: We used a multi-step procedure to analyze the real fatty acid data. First, we used Mclust (18) to fit 1, 2 or 3 component mixtures of equal variance normal distributions to each of the 14 fatty acid ratios, for a total of three different models being estimated on each fatty acid ratio. The model yielding the largest Bayesian Information Criterion (BIC) was deemed the best fitting model for the FA. The guidelines of Raftery (20) were used to ensure the best fitting model was significantly better than other models, namely, that the improvement in BIC is at least 6 when comparing the best to second best fitting model. We also fit models after dropping the most extreme 0.5% of the FA values (top 0.25% and bottom 0.25%), since this class of model fitting techniques can be sensitive to extreme values (21).

We then applied the LRT (see *Methods of analyzing metabotype data* for details) to each data using the estimated number of components. All other tests described above were also applied to the data. Odds ratios were estimated using a logistic regression model predicting case-control status by each individual's metabotype posterior probability for a particular fatty acid. Recent related results have illustrated the near optimality (vs. LRT) of using posterior probabilities to predict case-control status in such models (22). Two sided tests at significance level 5% were used in all cases. Missing FA data was below 0.8% (range 0-0.8%) for all 9 of the FA ratios.

Results:

Two component simulations: The type I error rate (false positives) for the proposed LRT test was well controlled (Supplemental Table 1), and this was the case whether we affirmed prior knowledge of two metatypes (LRT₂) or if the number of metatypes was unknown and the test was required to estimate the metabotype number (LRT_{est}). The type I error rate was comparable and well-controlled by all methods (equal and unequal variance *t*-tests, Mann-Whitney test, χ^2 test, LRT₂ and LRT_{est}) across all simulation settings considered here. In particular, in almost all cases, a 95% CI on the empirical type I error rate contains the nominal type I error rate of 5%. Minor exceptions occurred when the sample size of one component was small, yielding slightly conservative or slightly liberal type I error rates for methods estimating metabotype proportions from the data (χ^2 test, LRT₂ and LRT_{est}).

In addition to control of type 1 errors, both LRT methods provided superior power in most scenarios. Supplemental Table 2 illustrates the power of case-control association tests when there is a difference between component proportions in cases and controls. Ninety-two of 180 simulation settings yielded power greater than 99% for all 6 methods considered here, meaning model selection is not particularly important in such contexts. In 70 out of the remaining 88 settings, the LRT₂ yields power equivalent to or better than the equal variance *t*-test, the unequal variance *t*-test, the Mann-Whitney test

and the χ^2 test. On average, LRT_2 provided an increase in power of 4 percentage points compared to both the equal and unequal variance t -tests (min=0, max 15 point increase), 12.5 percentage points more power than the Mann-Whitney test (min=0, max=44.4 point increase) and 3.9 percentage points more power than the χ^2 test (min=0, max=15.4 percentage points). In the 18 settings when the LRT_2 was not the most powerful test, the equal and unequal variance t -tests were on average 0.7 percentage points better (min= -1 point, max=2.4 points), the Mann-Whitney test was, on average, 2.6 points better (min= -0.8 points, max=8.8 points) and the chi-squared test was, on average, 8.2 points better (min=0, max=21.6 points). LRT_2 tended to be best when the standard deviations were smaller and equal and when differences in the case-control proportions were smaller.

In 59 out of the 88 settings LRT_{est} yielded power equivalent to or better than the equal variance t -test, the unequal variance t -test, the Mann-Whitney test and the χ^2 test. On average across the 59 settings, LRT_{est} provided 4 percentage points more power than both the equal and unequal variance t -tests (min=0, max=15 points), 13.9 points better than the Mann-Whitney test (min=0.2 points, max=44.4 points) and 4 points better than the χ^2 test (min=0, max=15.4 points). In the remaining 29 settings the average amount that both t -tests were better than LRT_{est} was 1.1 points (min= - 2.4 points, max=6.4 points), for Mann-Whitney the average improvement was 2.2 points (min= - 1.8 points, max=8.6 points) and for the χ^2 test the average improvement was 4 points (min= - 15.8 points, max=5.6 points). LRT_{est} tended to be optimal in the same situations as LRT_2 .

Figure 2 illustrates simulation setting where 90.25% of controls are in component 1 and 81.2% of cases are in component 1. The standard deviation is the same in both components and varies on the x-axis; the difference in component means is fixed at 1. The figure illustrates that the LRT_{est} is more powerful than all other tests in this case.

Three component simulations: The type I error rate for the χ^2 test, the LRT_3 , and the LRT_{est} was well controlled for settings with three components (Supplemental Table 3) as long as the sample size in each of the three components was reasonably large (at least three observations). The t -tests and Mann-Whitney tests showed type I error control in all situations. Twenty-one of 60 settings yielded power greater than 99% for all six methods considered here, meaning model selection is not particularly important in such contexts. In 11 of the remaining 39 settings the LRT_3 is the most powerful, with average power gain vs the t -test of 0.3 points (min=0, max=1.6 points), vs. the Mann-Whitney test of 2.6 points (min=0, max=15.4 points) and vs. the χ^2 test of 46.4 points (min=0, max=100 points), however we note that the χ^2 test often suffers from small sample size issues in this simulation, yielding low power. In 24 of the 39 settings LRT_{est} was the most powerful of all methods with an average improvement vs. the t -test of 1 point

(min=0, max=7 points), vs. Mann-Whitney 5.7 points (min=0, max=26.8 points) and 56.7 points vs. χ^2 (min=1.2 points, max=100 points). In general, across the settings we investigated it was much harder to correctly identify the data as being produced from a three-component mixture model, than a two component mixture. We note that in many cases the best model was a two component model because of the small number of individuals present in the third component. As sample size increases the performance of the three component mixture model and test improves (details not shown). In general, the LRT test performs better than other tests in similar ways to those described earlier.

Case and control plasma fatty acids: The means and standard deviations overall, and within each of the case-control samples for each of the fourteen fatty acid ratios considered here are provided in Supplemental Table 5. Correlation between fatty acid ratios are provided in Supplemental Table 6. Notably, correlations within closely related pathways are stronger than among distantly related ones.

Metabotypes estimations: We defined metabotypes to be single components of each mixture, and each component could be comprised of case or control subjects. We used MClust to fit 1, 2 and 3 component mixtures of normal distributions for each of the activity surrogates (*i.e.* fatty acid ratios). Most fatty acids had a clear “best fitting model” that was robust to the potential presence of outliers. Of the 14 activity biomarkers explored, eleven were clearly mixtures of two components. One activity, ELVOL2, was ambiguously either not different or a mixture of three components (see Supplemental Tables 7a and 7b for details), and no further analysis was conducted.

Case/control Metabotype differences: We next tested whether the proportion of metabotypes differed among cases and controls. Table 1 gives the estimated means and standard deviations of each component for the 11 activity biomarkers for which a 2-component normal mixture model fit best. Among enzymes acting on saturated and mono-unsaturated fatty acids, two differences were notable: D9D activity and ELOVL6 activity. For D9D activity on 16- and 18-carbon fatty acids, two metabotypes were present: a more common low activity metabotype, and a metabotype with a higher mean activity. While this high activity metabotype was less common, the higher 18:1n9/18:0 ratio was more abundant in cases (14.8%) than controls (7.1%) ($P_{undadjusted} < 0.0001$). Adjusting for other fatty acids in the pathway nullified the difference however a reduced risk, Odds Ratio (OR)=0.3 [0.2, 0.6; $P_{adjusted} < 0.0001$], persisted for subjects with higher 18:1n9/18:0 based activity than on the 16:1n7/16:0-based activity. Two metabotypes were also observed among ELOVL6 populations. While only 40% of control subjects had the lower activity metabotype, nearly 50% of cases did ($P < 0.0001$) suggesting lower elongation activity is associated with a higher risk.

Similarly, key difference among cases and controls were observed among biomarkers of enzymes acting on polyunsaturated fatty acids. Three differences were noted: ELOVL5 activity on omega-6 fatty acids, Sprecher pathway chain shortening activity, and FADS1 activity; the lattermost only apparent after adjustment. For ELOVL5, only 7% of controls had the high activity metabotype compared to 20% of cases ($P_{adjusted} < 0.0001$), indicating metabotypes having increased elongation of gLA to dgLA can confer risk. The production of 22-carbon fatty acids (n6-DPA and DHA) is a multi-step process sometimes termed Sprecher-pathway chain shortening (SCS). Subjects with high SCS activity on omega-3 fatty acids represented 10% of the controls, however they represented only 4.5% of the cases. Adjustment for other fatty acids ablated this effect, however the effect persisted when the relative SCS activity on omega-3 to omega-6 fatty acids was measured, and subjects with the metabotype associated with relatively more omega-3 (DHA) production had lower risk for ACS, OR= 0.3 [0.1, 0.5]. After adjustment, a third difference emerged for FADS1, where efficient conversion of dgLA to AA was more prevalent among cases than controls.

The different proportions of metabotypes in cases and controls allow for calculation of odds ratios. Since single component populations allows for only a continuously increasing or decreasing odds ratio across sample values, a particular advantage of conceptualizing multiple components is the possibility of identify more biologically intuitive odds patterns. Figure 3 plots the estimated odds from the two-component mixture model of D9D activity when expressed as the OA/SA ratio and the odds ratio is frequency of the cases and controls at each activity, the observed odds, and the odds obtained from a means model. As can be observed in the figure, the 2-component mixture model appears to fit the observed data better than the means model, confirming our intuition.

Discussion:

Identifying disease risk factors and predicting risk have become key elements in modern health care and effective intervention policy. The prevalent approach to risk identification is through means testing of measured metabolic and life style factors between populations of cases and controls. However, an inherent weakness to means testing is the assumption that collected samples represent a selection from a single population, and the impact of this assumption is most apparent with large heterogeneous sample sets. We demonstrate a test that robustly detects populations with multiple components (aka, multi-modal distributions). Using the test, our study population was empirically found to be multi-modal and thus violates this assumption. Both genetic and environmental factors can provide sources of variance responsible for this violation. For instance some gene polymorphisms can result in functional changes, direct or epistatic, that would impart a discontinuous, digital shift in measures of activity. Alternatively,

metabolic variance due to environmental factors, including diet and lifestyle, would impart continuous variance. Thus, a unified explanation of environmental factors and Mendelian genetics implies digital shifts overlain on continuous variation, or multi-modal distributions. Based on these facts, we hypothesize that sample sets with multiple *metabotype* components, as opposed to a single component, should be expected. Here we term these components *metabotypes*. Metabotypes are commonly conceived in metabolomic terms, representing broad fingerprint-like patterns of metabolites that act as phenotypes and so inform diagnostic or prognostic studies (23). Here, we apply the concept of a metabolic phenotype more narrowly to specific metabolic activities and overlay the biomarker concept with the prevailing genetic theory. While studies have merged genetic and metabolic data for risk assessment (24), here we incorporate prior knowledge of diet-gene interaction to develop a likelihood ratio test that more faithfully reflects the biology of both.

Using simulated sample sets, the LRT test was found to not only detect mixed populations, but in doing so provided great power to discriminate true means in nearly every scenario tested. In addition, the LRT_{est} gives added biological insight. By accounting for the expected variance conferred by functional genetic differences as the component means themselves, from that conferred by nutritional status and lifestyle as the variance off means, we give insight into how these axes of risk interact. For instance, because a lifestyle confers risk by elevating a biomarker, it is not clear or necessary that genes which elevate the same biomarker would also confer risk. In cases where genetic risk operates in the opposite direction as nutritional/lifestyle the simpler analyses would obfuscate biomarker interpretation, which would be especially revealed by the LRT_{est} analysis. The LRT_{est} also effectively identifies important functional phenotypes in lieu of null variants, and so could be used to effectively identify the subset of genes whose polymorphism actually results in functional alterations. Since SNP analyses do not measure functional alterations, this approach could provide guidance to such approaches. Here, we have identified distinct fatty acid metabolism metabotypes which are associated with increased risk of ACS. By implication, we have also identified optimal activities (Figure 3), in contrast to means testing techniques which do not allow for local optima as they can only provide monotonic changes.

Metabotypes with greater desaturation activity on stearic acid were at increased risk for ACS compared to activity on palmitic acid, and effect that persists after adjusting for other correlated activities. At least two genes encode enzymes with D9D activity in humans (25), and increasing activity is associated with increase adiposity (26). Prior tracer work confirms higher desaturase activity on SA (27), and this biomarker is a predictor of risk for cardiovascular disease mortality (28). By contrast there is no direct evidence for a role of ELOVL6 in heart disease, however by elongating the SCD precursor PA to

SA, high ELOVL6 activity is thought to contribute to palmitate-induced lipotoxicity (29). ELOVL6 are related to co-morbidities of heart disease including gene polymorphisms in insulin resistance (30, 31) and increased activity in hepatic steatosis (32). We observed that metabolotypes with increased ELOVL6 activity is protective. However, since our analysis focuses on the effect of belonging to a metabolotype, and not deviating from a metabolotype mean, it is not necessary to conclude these observations are conflicting.

We did not find evidence for FADS2 activity in risk, however efficient metabolotypes of ELOVL5 and FADS1 were associated with increased risk, the latter only after adjustment. ELOVL5 polymorphisms were not associated with either risk for myocardial infarction but were associated with risk intermediates such as LDL-cholesterol (33). Such failure to replicate highlights the potential for methodologies that combine approaches to increase analytical power. The delta5-desaturase activity mediates the conversion of DGLA (20:3n6) to AA (20:4n6) as well as eicosatetraenoic acid (20:4n3) to EPA (20:5n3) and has been associated with numerous pathologies. Based on biomarkers from cholesteryl ester pools, Lu *et al* observed a polymorphism of high delta 5-desaturase activity (rs174547) is associated with reduced risk for coronary heart disease (34), and other cases are reported where high δ 5-desaturase activity is associated with reduced risk (35). Such findings identify activities associated with risk, however our model suggests at least part of the risk results from metabolotypes with differing efficiencies, providing insight into possible gene/lifestyle interaction.

We have approached our analysis based on the assumption that metabolotypes result from genetic polymorphisms however other factors, environmental or unknown, could also establish them. In the latter case, environment is common and correlation of metabolotypes is more likely, however in the former case, dependence is unnecessary and less likely. Thus, demonstrating independence is an inference (albeit weak) supporting a genetic basis. We make two findings of independence: lack of correlation between fatty acids of different pathways and an additive effect of metabolotype membership. Notably, the strongest correlation is observed among closely related activities (outlined in bold), less strong as activities become more distant. Further metabolotype information alone, independent of variance off the means predicts a large portion of cases, consistent with independence and direct a role for metabolotypes in ACS risk (Supplemental Figure 2).

Implications: Recent reports have highlighted the difficulties associated with residual risk, including one report from this same study where SNP analyses showed no genome-wide associations with myocardial infarction (13). Our test not only provides greater explanation of the data, it is formed by combining basic principles of genetics and environmental factors and so provides straightforward interpretation. Since it predicts populations with multi-modal properties, it has the added benefit of not being constrained to the

unrealistic prediction that risk is monotonic with respect to activity: risk can either plateau or reach local maxima and minima. While the latter usually requires at least three components (not found here), we identified plateaued risk (Figure 3) where decreasing D9D activity did not monotonically decrease risk. Identifying such locales could guide interventions by providing metabolically ideal targets, or in the example case, a threshold above which measures to decrease activity could be beneficial.

Limitations: We have interpreted most of our results by extrapolating to the expected genetics underlying the metabotypes; however, factors other than genetics may be responsible for their presence. We have additionally inferred that the biomarkers reflect the apparent functional enzyme activities. While common, this presupposes that changes in precursor rate of appearance and product rate of clearance are not responsible for the changes. Thus, while we have robustly detected metabotypes, we cannot make strong inferences to mechanism. It is hoped we can validate their application to risk estimation in future studies. Our method is only likely to detect metabotypes that are widely separated and would be restricted to large datasets. If metabotypes are truly attributable to genetic polymorphisms, the additive genetic model would predict three distinct metabotypes, however in most cases we detect only two. This is likely due to the expected small abundance of recessive metabotypes, and would need to be confirmed in datasets where the recessive phenotype is more abundant.

Conclusions: We have verified a novel approach to understand how fatty acids and other continuous metabolomic datasets could be evaluated for relationships to risk. We further validated this test, in the context of real biological data, to discover risks associated with unique metabotypes and improved biological interpretability.

Acknowledgements: The authors acknowledge John Spertus and William Harris for their work on the parent project and in generating the data. Additional support was provided by the USDA Intramural Projects 5306-51530-019-00D. The USDA is an equal opportunity provider and employer.

Table 1. Evaluating relationships between FA ratio and case-control status

FA Ratio	Best model	Model Details Mean (SD) of each Component	Estimated Mixing Proportions		LRT _{est} p-value	Unadjusted OR (95% CI) ^a	Overall adjusted OR (95% CI) ^b
			Case	Control			
Short-chain desaturase							
D9D C16	2 Metabotype	M1: 0.06 (0.02) M2: 0.12 (0.02)	90.0 10.0	91.5 8.5	0.36	1.3 (0.8, 1.9)	1.0 (0.4, 2.3)
D9D C18	2 Metabotype	M1: 1.17 (0.25) M2: 1.96 (0.25)	85.2 14.8	92.8 7.1	2.6 x10 ⁻⁶	2.6 (1.8, 3.9)***	1.5 (0.8, 2.7)
D9D 16:18	2 Metabotype	M1: 0.05 (0.01) M2: 0.08 (0.01)	95.5 4.5	91.2 8.8	0.001	0.4 (0.3, 0.7)***	0.3 (0.2, 0.6)***
ELOVL6	2 Metabotype	M1: 0.62 (0.12) M2: 0.70 (0.12)	53.3 46.7	39.2 60.8	4.3 x 10 ⁻⁸	0.2 (0.1, 0.3)***	0.2 (0.1, 0.5)***
ELOVL6 mono	2 Metabotype, Normal	M1: 0.08 (0.02) M2: 0.16 (0.02)	92.5 7.5	92.4 7.6	0.98	1.0 (0.6, 1.5)	1.3 (0.5, 3.2)
Arachidonate synthesis pathway							
FADS2	2 Metabotype	M1: 0.03 (0.01) M2: 0.05 (0.01)	95.5 4.5	93.6 6.4	0.13	0.6 (0.4, 1.1)	0.9 (0.5, 1.6)
ELOVL5	2 Metabotype	M1: 4.0 (1.0) M2: 6.4 (1.0)	80.3 19.7	92.6 7.4	4.2x10 ⁻¹²	5.5 (3.6, 8.4)***	5.8 (3.7, 8.9)***
FADS1	2 Metabotype	M1: 7.9 (2.0) M2: 12.3 (2.0)	89.7 10.3	92.0 8.0	0.12	1.5 (0.9, 2.4)	1.8 (1.1, 2.8)***
Docosanoid synthesis pathway							
SCS n6	2 Metabotype	M1: 0.21 (0.04) M2: 0.22 (0.04)	92.2 7.8	94.4 5.6	0.11	1.8 (1.0, 3.0)	1.5 (0.9, 2.6)
SCS n3	2 Metabotype	M1: 1.7 (0.5) M2: 3.2 (0.5)	95.5 4.5	89.7 10.3	1.9x10 ⁻⁵	0.3 (0.2, 0.6)***	0.8 (0.4, 1.5)
SCS n3:n6	2 Metabotype	M1: 7.9 (2.5) M2: 16.6 (2.5)	96.1 3.9	88.2 11.7	2.7x10 ⁻⁸	0.3 (0.1, 0.4)***	0.3 (0.1, 0.5)***

*p<0.05, **p<0.01, ***p<0.001

^aORs are interpreted as the increase in the likelihood of being a case given you are in metabotype 2.^bAdjusted ORs are adjusted for all FAs within the block/pathway.

Figures

Figure 1:

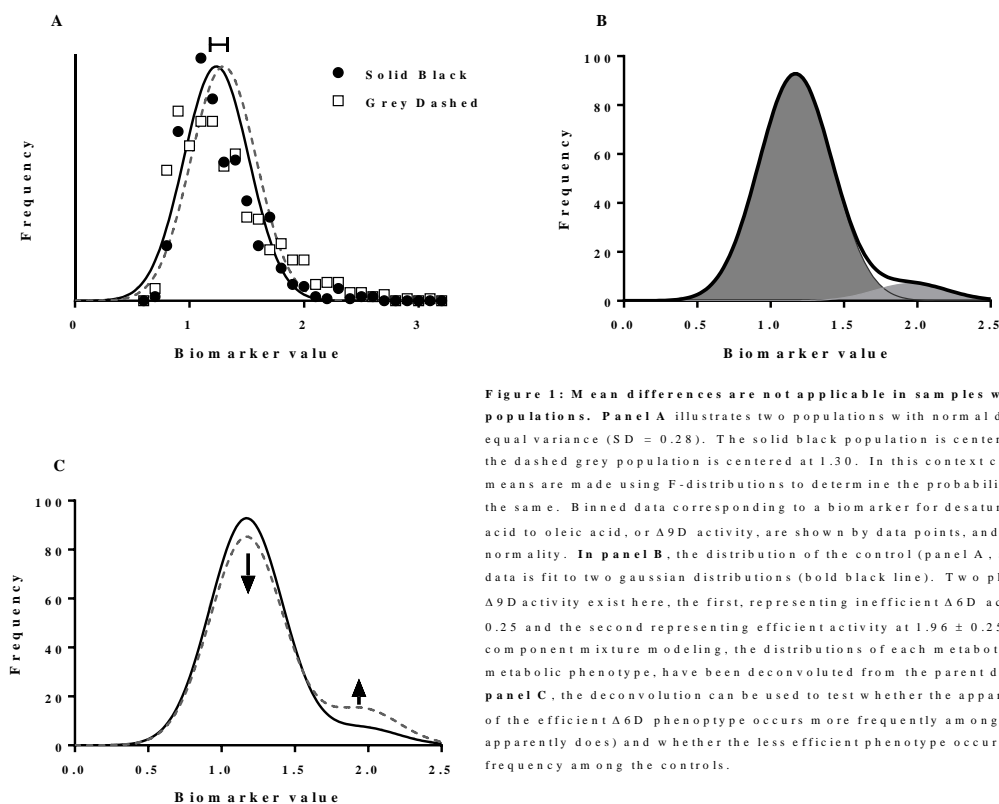


Figure 1: Mean differences are not applicable in samples with multiple populations. Panel A illustrates two populations with normal distribution and equal variance ($SD = 0.28$). The solid black population is centered at 1.23 and the dashed grey population is centered at 1.30. In this context comparison of means are made using F-distributions to determine the probability that they are the same. Binned data corresponding to a biomarker for desaturation of stearic acid to oleic acid, or $\Delta 9D$ activity, are shown by data points, and skewed from normality. In panel B, the distribution of the control (panel A, solid black) data is fit to two gaussian distributions (bold black line). Two phenotypes of $\Delta 9D$ activity exist here, the first, representing inefficient $\Delta 6D$ activity at 1.17 ± 0.25 and the second representing efficient activity at 1.96 ± 0.25 . Using component mixture modeling, the distributions of each metabolite, or metabolic phenotype, have been deconvoluted from the parent distribution. In panel C, the deconvolution can be used to test whether the apparent abundance of the efficient $\Delta 6D$ phenotype occurs more frequently among the cases (as it apparently does) and whether the less efficient phenotype occurs with greater frequency among the controls.

Figure 2

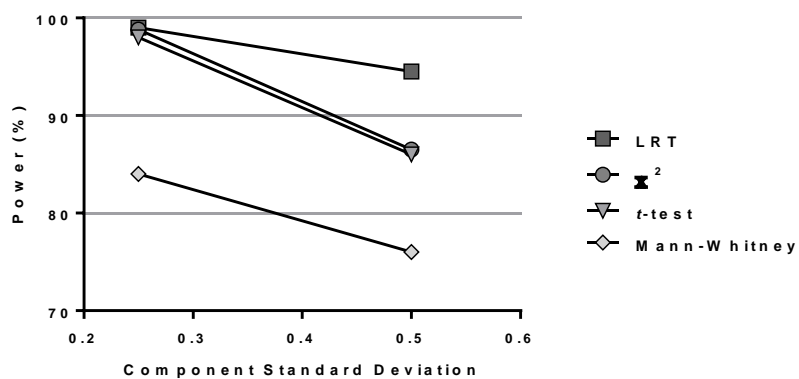


Figure 2: The power of LRT is superior, especially with increased variance. Results for simulations for power to detect differences where 90.25% of controls are in component 1 and 81.2% of cases are in component 1 with the standard deviation on the x-axis; the difference in component means is fixed at 1.

Figure 3

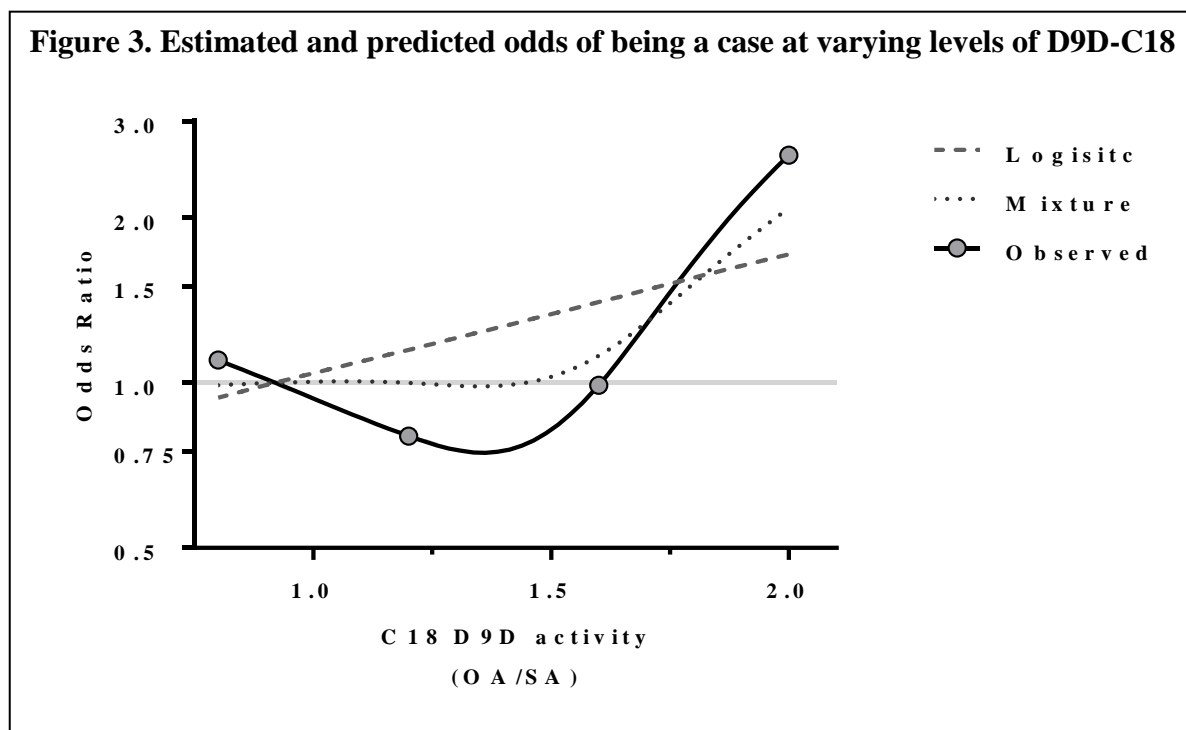


Figure 3: Observed odds more closely resemble estimates of mixture models. Odds estimates for the mixture model are computed by finding the ratio of the estimated density functions for the metabotypes, for the logistic model (predicting case-control status by FA level directly) they are found directly from the model. Observed odds ratios were estimated via binning of the real data.

Supplemental Tables

Supplemental Table 1. Type I error rates for two-component mixture models

St. dev. grp 1	St. dev. grp 2	Prop. of cases in grp 1	Prop. of controls in grp 1	Type I error rate of different statistical analysis methods						Percent of 500 simulations yielding best estimate of...		
				T-test (equal variance)	T-test (unequal variance)	Mann- Whitney Test	χ^2 test	LRT ₂	LRT _{est}	1 comp.	2 comp.	3 comp.
0.25	0.25	0.25	0.25	0.064	0.064	0.064	0.048	0.06	0.06	0	1	0
0.25	0.375	0.25	0.25	0.036	0.036	0.032	0.036	0.052	0.048	0	0.338	0.662
0.25	0.5	0.25	0.25	0.038	0.038	0.042	0.036	0.038	0.04	0	0.018	0.982
0.25	0.25	0.449	0.449	0.056	0.056	0.06	0.05	0.052	0.052	0	1	0
0.25	0.375	0.449	0.449	0.056	0.056	0.062	0.054	0.072	0.068	0	0.056	0.944
0.25	0.5	0.449	0.449	0.042	0.042	0.042	0.038	0.05	0.04	0	0.148	0.852
0.25	0.25	0.5625	0.5625	0.05	0.05	0.042	0.048	0.058	0.058	0	1	0
0.25	0.375	0.5625	0.5625	0.036	0.036	0.036	0.046	0.054	0.054	0	0.962	0.038
0.25	0.5	0.5625	0.5625	0.058	0.058	0.058	0.066	0.064	0.064	0	1	0
0.25	0.25	0.81	0.81	0.062	0.062	0.058	0.048	0.056	0.056	0	0.998	0.002
0.25	0.375	0.81	0.81	0.056	0.056	0.068	0.048	0.056	0.056	0	1	0
0.25	0.5	0.81	0.81	0.034	0.032	0.034	0.026	0.028	0.028	0	1	0
0.25	0.25	0.9025	0.9025	0.074	0.074	0.062	0.06	0.066	0.066	0	0.998	0.002
0.25	0.375	0.9025	0.9025	0.038	0.038	0.046	0.04	0.05	0.05	0	1	0
0.25	0.5	0.9025	0.9025	0.056	0.056	0.05	0.04	0.056	0.056	0	1	0
0.25	0.25	0.75	0.75	0.046	0.046	0.052	0.034	0.032	0.032	0	1	0
0.25	0.375	0.75	0.75	0.04	0.04	0.04	0.04	0.06	0.06	0	1	0
0.25	0.5	0.75	0.75	0.046	0.046	0.048	0.032	0.046	0.046	0	1	0
0.25	0.25	0.891	0.891	0.05	0.05	0.054	0.06	0.064	0.064	0	0.998	0.002
0.25	0.375	0.891	0.891	0.054	0.054	0.054	0.042	0.056	0.056	0	1	0
0.25	0.5	0.891	0.891	0.066	0.066	0.072	0.032	0.046	0.046	0	1	0
0.25	0.25	0.9375	0.9375	0.038	0.038	0.032	0.034	0.05	0.05	0	1	0
0.25	0.375	0.9375	0.9375	0.05	0.05	0.04	0.04	0.048	0.048	0	1	0
0.25	0.5	0.9375	0.9375	0.054	0.054	0.052	0.032	0.032	0.032	0	1	0
0.25	0.25	0.99	0.99	0.044	0.044	0.042	0.028	0.048	0.038	0.02	0.794	0.186

0.25	0.375	0.99	0.99	0.038	0.038	0.036	0.012	0.048	0.038	0.012	0.594	0.394
0.25	0.5	0.99	0.99	0.052	0.052	0.038	0.038	0.06	0.068	0.012	0.408	0.58
0.25	0.25	0.9975	0.9975	0.032	0.032	0.036	0.038	0.058	-- ^a	0.682	0.172	0.146
0.25	0.375	0.9975	0.9975	0.058	0.058	0.056	0.036	0.092	-- ^a	0.578	0.176	0.246
0.25	0.5	0.9975	0.9975	0.062	0.062	0.058	0.034	0.064	-- ^a	0.49	0.132	0.378
0.5	0.5	0.25	0.25	0.036	0.036	0.04	0.048	0.04	0.04	0	1	0
0.5	0.75	0.25	0.25	0.034	0.034	0.04	0.038	0.042	-- ^a	0.24	0.746	0.014
0.5	1	0.25	0.25	0.048	0.048	0.044	0.04	0.056	0.056	0	0.998	0.002
0.5	0.5	0.449	0.449	0.062	0.062	0.058	0.052	0.052	0.05	0.014	0.984	0.002
0.5	0.75	0.449	0.449	0.066	0.066	0.052	0.058	0.058	0.058	0	0.984	0.016
0.5	1	0.449	0.449	0.054	0.054	0.054	0.04	0.046	0.046	0	1	0
0.5	0.5	0.5625	0.5625	0.068	0.068	0.066	0.062	0.08	0.08	0.016	0.984	0
0.5	0.75	0.5625	0.5625	0.042	0.042	0.044	0.038	0.052	0.052	0	1	0
0.5	1	0.5625	0.5625	0.044	0.044	0.042	0.048	0.05	0.05	0	1	0
0.5	0.5	0.81	0.81	0.058	0.058	0.062	0.04	0.052	0.052	0.002	0.998	0
0.5	0.75	0.81	0.81	0.04	0.04	0.058	0.032	0.03	0.03	0	1	0
0.5	1	0.81	0.81	0.048	0.048	0.052	0.042	0.056	0.056	0	1	0
0.5	0.5	0.9025	0.9025	0.044	0.044	0.036	0.034	0.048	0.048	0.012	0.986	0.002
0.5	0.75	0.9025	0.9025	0.046	0.046	0.058	0.024	0.04	0.04	0	1	0
0.5	1	0.9025	0.9025	0.048	0.048	0.046	0.036	0.044	0.044	0	1	0
0.5	0.5	0.75	0.75	0.034	0.034	0.038	0.028	0.028	0.028	0	1	0
0.5	0.75	0.75	0.75	0.048	0.048	0.042	0.038	0.052	0.052	0	1	0
0.5	1	0.75	0.75	0.058	0.058	0.066	0.032	0.046	0.046	0	1	0
0.5	0.5	0.891	0.891	0.05	0.05	0.06	0.028	0.04	0.038	0.01	0.99	0
0.5	0.75	0.891	0.891	0.054	0.054	0.044	0.04	0.048	0.048	0	1	0
0.5	1	0.891	0.891	0.052	0.052	0.056	0.044	0.068	0.068	0	1	0
0.5	0.5	0.9375	0.9375	0.042	0.042	0.034	0.042	0.038	-- ^a	0.162	0.83	0.008
0.5	0.75	0.9375	0.9375	0.046	0.046	0.042	0.038	0.046	0.048	0	0.976	0.024
0.5	1	0.9375	0.9375	0.06	0.06	0.062	0.05	0.066	0.07	0	0.964	0.036
0.5	0.5	0.99	0.99	0.05	0.05	0.056	0.06	0.076	-- ^a	0.994	0.004	0.002
0.5	0.75	0.99	0.99	0.06	0.06	0.056	0.046	0.054	-- ^a	0.858	0.074	0.068
0.5	1	0.99	0.99	0.048	0.048	0.052	0.058	0.07	-- ^a	0.548	0.086	0.366
0.5	0.5	0.9975	0.9975	0.046	0.046	0.048	0.034	0.072	-- ^a	0.998	0.002	0

0.5	0.75	0.9975	0.9975	0.038	0.038	0.044	0.05	0.056	-- ^a	0.996	0.002	0.002
0.5	1	0.9975	0.9975	0.044	0.044	0.044	0.044	0.056	-- ^a	0.946	0.016	0.038

Bolded entries indicate type I error rates where the corresponding 95% CI does not include 5%, indicating empirical inflation of the nominal 5% type I error rate

^a The best fitting model was often a 1 component model in these cases, making the LRT_{best} unusable to test for a difference.

Supplemental Table 2. Power for two-component mixture models

St. dev. grp 1	St. dev. grp 2	Prop. of cases in grp 1	Prop. of controls in grp 1	Power of different statistical analysis methods						Percent of 500 simulations yielding best estimate of...		
				T-test (equal variance)	T-test (unequal variance)	Mann- Whitney Test	χ^2 test	LRT ₂	LRT _{est}	1 comp.	2 comp.	3 comp.
0.25	0.25	0.225	0.25	0.156	0.156	0.142	0.168	0.19	0.19	0	1	0
0.25	0.25	0.1875	0.25	0.756	0.756	0.59	0.802	0.824	0.824	0	1	0
0.25	0.375	0.225	0.25	0.116	0.116	0.122	0.154	0.166	0.14	0	0.37	0.63
0.25	0.375	0.1875	0.25	0.64	0.64	0.56	0.668	0.728	0.69	0	0.438	0.562
0.25	0.375	0.125	0.25	0.996	0.996	0.982	1	1	1	0	0.596	0.404
0.25	0.5	0.225	0.25	0.12	0.12	0.116	0.086	0.108	0.1	0	0.024	0.976
0.25	0.5	0.1875	0.25	0.514	0.514	0.508	0.43	0.55	0.506	0	0.064	0.936
0.25	0.5	0.125	0.25	0.98	0.98	0.978	0.968	0.986	0.986	0	0.154	0.846
0.25	0.25	0.4041	0.449	0.35	0.35	0.326	0.364	0.39	0.39	0	1	0
0.25	0.25	0.33675	0.449	0.986	0.986	0.976	0.992	0.996	0.996	0	0.998	0.002
0.25	0.375	0.4041	0.449	0.31	0.31	0.294	0.302	0.302	0.246	0	0.064	0.936
0.25	0.375	0.33675	0.449	0.968	0.968	0.964	0.97	0.982	0.968	0	0.092	0.908
0.25	0.5	0.4041	0.449	0.314	0.314	0.336	0.236	0.318	0.266	0	0.018	0.982
0.25	0.5	0.33675	0.449	0.93	0.93	0.94	0.892	0.924	0.924	0	0	1
0.25	0.25	0.50625	0.5625	0.506	0.506	0.476	0.534	0.582	0.582	0	1	0
0.25	0.375	0.50625	0.5625	0.444	0.444	0.442	0.47	0.52	0.476	0	0.7	0.3
0.25	0.5	0.50625	0.5625	0.41	0.41	0.406	0.372	0.418	0.418	0	1	0
0.25	0.5	0.421875	0.5625	0.992	0.992	0.992	0.984	0.994	0.994	0	0.888	0.112
0.25	0.25	0.729	0.81	0.908	0.908	0.794	0.958	0.968	0.968	0	0.998	0.002
0.25	0.375	0.729	0.81	0.884	0.884	0.766	0.896	0.904	0.904	0	1	0
0.25	0.5	0.729	0.81	0.868	0.868	0.754	0.854	0.898	0.898	0	1	0
0.25	0.25	0.81225	0.9025	0.98	0.98	0.842	0.992	0.994	0.994	0	1	0
0.25	0.375	0.81225	0.9025	0.982	0.982	0.858	0.992	0.998	0.998	0	1	0
0.25	0.5	0.81225	0.9025	0.974	0.974	0.83	0.982	0.984	0.984	0	1	0
0.25	0.25	0.675	0.75	0.84	0.84	0.778	0.87	0.882	0.882	0	1	0
0.25	0.375	0.675	0.75	0.8	0.8	0.704	0.794	0.85	0.85	0	1	0

0.25	0.5	0.675	0.75	0.75	0.75	0.684	0.722	0.78	0.78	0	1	0
0.25	0.25	0.8019	0.891	0.98	0.98	0.88	0.996	0.996	0.996	0	1	0
0.25	0.375	0.8019	0.891	0.97	0.97	0.842	0.992	0.996	0.996	0	1	0
0.25	0.5	0.8019	0.891	0.982	0.982	0.82	0.98	0.99	0.99	0	1	0
0.25	0.25	0.84375	0.9375	0.998	0.998	0.908	1	1	1	0	1	0
0.25	0.375	0.84375	0.9375	0.994	0.994	0.878	1	1	1	0	1	0
0.25	0.5	0.84375	0.9375	0.99	0.99	0.854	0.996	0.998	0.998	0	1	0
0.25	0.25	0.891	0.99	1	1	0.932	1	1	1	0	1	0
0.25	0.375	0.891	0.99	1	1	0.91	1	1	1	0	1	0
0.25	0.5	0.891	0.99	1	1	0.894	1	1	1	0	1	0
0.25	0.25	0.89775	0.9975	1	1	0.938	1	1	1	0	1	0
0.25	0.375	0.89775	0.9975	1	1	0.93	1	1	1	0	1	0
0.25	0.5	0.89775	0.9975	1	1	0.886	1	1	1	0	1	0
0.5	0.5	0.225	0.25	0.112	0.112	0.102	0.098	0.126	0.126	0	1	0
0.5	0.5	0.1875	0.25	0.478	0.478	0.442	0.45	0.558	0.558	0.002	0.998	0
0.5	0.5	0.125	0.25	0.978	0.978	0.948	0.978	0.994	0.994	0	1	0
0.5	0.75	0.225	0.25	0.084	0.084	0.08	0.072	0.072	-- ^a	0.344	0.642	0.014
0.5	0.75	0.1875	0.25	0.358	0.358	0.378	0.246	0.354	-- ^a	0.452	0.542	0.006
0.5	0.75	0.125	0.25	0.864	0.864	0.886	0.678	0.84	-- ^a	0.708	0.286	0.006
0.5	1	0.225	0.25	0.086	0.086	0.09	0.064	0.08	0.08	0.002	0.986	0.012
0.5	1	0.1875	0.25	0.232	0.232	0.27	0.156	0.22	0.214	0.018	0.978	0.004
0.5	1	0.125	0.25	0.73	0.73	0.804	0.5	0.716	-- ^a	0.074	0.918	0.008
0.5	0.5	0.4041	0.449	0.23	0.23	0.248	0.21	0.262	0.26	0.008	0.992	0
0.5	0.5	0.33675	0.449	0.894	0.894	0.914	0.876	0.93	0.926	0.006	0.994	0
0.5	0.5	0.2245	0.449	1	1	1	1	1	0.988	0.012	0.988	0
0.5	0.75	0.4041	0.449	0.174	0.174	0.186	0.13	0.194	0.196	0	0.978	0.022
0.5	0.75	0.33675	0.449	0.798	0.798	0.83	0.636	0.794	0.794	0	0.986	0.014
0.5	0.75	0.2245	0.449	1	1	1	0.99	1	0.994	0.006	0.958	0.036
0.5	1	0.4041	0.449	0.168	0.168	0.164	0.108	0.156	0.156	0	1	0
0.5	1	0.33675	0.449	0.63	0.63	0.698	0.458	0.612	0.612	0	1	0
0.5	1	0.2245	0.449	0.99	0.99	0.998	0.948	0.994	0.994	0	1	0
0.5	0.5	0.50625	0.5625	0.386	0.386	0.408	0.342	0.426	0.424	0.024	0.976	0
0.5	0.5	0.421875	0.5625	0.97	0.97	0.98	0.96	0.986	0.962	0.024	0.976	0

0.5	0.5	0.28125	0.5625	1	1	1	1	1	0.982	0.018	0.982	0
0.5	0.75	0.50625	0.5625	0.312	0.312	0.338	0.242	0.32	0.32	0	1	0
0.5	0.75	0.421875	0.5625	0.95	0.95	0.96	0.842	0.936	0.936	0	0.99	0.01
0.5	1	0.50625	0.5625	0.22	0.22	0.24	0.164	0.23	0.23	0	1	0
0.5	1	0.421875	0.5625	0.866	0.866	0.88	0.728	0.856	0.856	0	1	0
0.5	0.5	0.729	0.81	0.684	0.684	0.652	0.696	0.774	0.774	0	1	0
0.5	0.5	0.405	0.81	1	1	1	1	1	0.984	0.016	0.984	0
0.5	0.75	0.729	0.81	0.59	0.59	0.512	0.496	0.65	0.65	0	1	0
0.5	1	0.729	0.81	0.538	0.538	0.388	0.454	0.574	0.574	0	1	0
0.5	0.5	0.81225	0.9025	0.86	0.86	0.764	0.87	0.94	0.94	0	1	0
0.5	0.75	0.81225	0.9025	0.806	0.806	0.654	0.788	0.906	0.906	0	1	0
0.5	1	0.81225	0.9025	0.728	0.728	0.47	0.782	0.878	0.878	0	1	0
0.5	0.5	0.675	0.75	0.62	0.62	0.622	0.602	0.686	0.684	0.004	0.996	0
0.5	0.5	0.375	0.75	1	1	1	1	1	-- ^a	0.03	0.97	0
0.5	0.75	0.675	0.75	0.546	0.546	0.482	0.438	0.558	0.558	0	1	0
0.5	1	0.675	0.75	0.486	0.486	0.386	0.41	0.526	0.526	0	1	0
0.5	1	0.5625	0.75	0.99	0.99	0.984	0.964	0.992	0.992	0	1	0
0.5	0.5	0.8019	0.891	0.808	0.808	0.708	0.826	0.916	0.914	0.002	0.998	0
0.5	0.75	0.8019	0.891	0.75	0.75	0.606	0.714	0.836	0.836	0	1	0
0.5	1	0.8019	0.891	0.704	0.704	0.474	0.736	0.842	0.842	0	1	0
0.5	0.5	0.84375	0.9375	0.896	0.896	0.76	0.918	0.976	0.964	0.014	0.986	0
0.5	0.75	0.84375	0.9375	0.846	0.846	0.646	0.854	0.946	0.946	0	1	0
0.5	1	0.84375	0.9375	0.832	0.832	0.548	0.874	0.942	0.942	0	1	0
0.5	0.5	0.891	0.99	0.946	0.946	0.814	0.98	1	-- ^a	0.194	0.804	0.002
0.5	0.75	0.891	0.99	0.932	0.932	0.718	0.988	0.998	0.998	0	0.976	0.024
0.5	1	0.891	0.99	0.894	0.894	0.556	0.986	1	0.994	0	0.944	0.056
0.5	0.5	0.89775	0.9975	0.95	0.95	0.838	0.97	1	-- ^a	0.3	0.69	0.01
0.5	0.75	0.89775	0.9975	0.942	0.942	0.698	0.978	1	0.994	0.006	0.922	0.072
0.5	1	0.89775	0.9975	0.928	0.928	0.556	0.978	1	1	0	0.87	0.13

^a The best fitting model was often a 1 component model in these cases, making the LRT_{est} unusable to test for a difference.

Supplemental Table 3. Type I error rate for 3-component mixtures

St. dev. grp 1	St. dev. grp 2	St. dev. grp 3	Ctrl minor allele frequency	Type I error of different statistical analysis methods						Percent of 500 simulations yielding best estimate of...		
				T-test (equal variance)	T-test (unequal variance)	Mann- Whitney Test	χ^2 test	LRT ₃	LRT _{est}	1 comp.	2 comp.	3 comp.
0.5	0.5	0.5	0.25	0.066	0.066	0.066	0.06	0.06	0.076	0	0.85	0.15
0.5	0.5	0.5	0.33	0.038	0.038	0.04	0.04	0.05	0.04	0.014	0.662	0.324
0.5	0.5	0.5	0.5	0.048	0.048	0.044	0.04	0.064	-- ^a	0.372	0.436	0.192
0.5	0.75	1	0.05	0.046	0.046	0.052	0.046	0.07	0.046	0	0.998	0.002
0.5	0.75	1	0.1	0.06	0.06	0.056	0.046	0.116	0.068	0	1	0
0.5	0.5	0.5	0.05	0.04	0.04	0.046	0.046	0.07	0.056	0.006	0.994	0
0.5	0.5	0.5	0.1	0.06	0.06	0.056	0.056	0.082	0.052	0	1	0
0.5	0.75	1	0.25	0.06	0.06	0.06	0.048	0.058	0.066	0	1	0
0.5	0.75	1	0.33	0.068	0.068	0.058	0.054	0.07	0.05	0	1	0
0.5	0.75	1	0.5	0.046	0.046	0.052	0.056	0.07	0.056	0	0.54	0.46
0.25	0.375	0.5	0.25	0.042	0.042	0.048	0.026	0.038	0.044	0	1	0
0.25	0.375	0.5	0.33	0.042	0.042	0.042	0.048	0.05	0.054	0	0.124	0.876
0.25	0.375	0.5	0.5	0.052	0.052	0.052	0.044	0.048	0.048	0	0	1
0.25	0.25	0.25	0.25	0.032	0.032	0.042	0	0.05	0.052	0	0.996	0.004
0.25	0.25	0.25	0.33	0.046	0.046	0.044	0.048	0.044	0.044	0	0	1
0.25	0.25	0.25	0.5	0.054	0.054	0.056	0.054	0.062	0.062	0	0	1
0.25	0.25	0.25	0.05	0.036	0.036	0.046	0.03	0.068	0.044	0	1	0
0.25	0.25	0.25	0.1	0.036	0.036	0.038	0	0.086	0.058	0	1	0
0.25	0.375	0.5	0.05	0.08	0.08	0.082	0.01	0.15	0.068	0	1	0
0.25	0.375	0.5	0.1	0.05	0.05	0.046	0	0.08	0.056	0	1	0

^a The best fitting model was often a 1 component model in these cases, making the LRT_{est} unusable to test for a difference.

Supplemental Table 4. Power for 3-component mixtures

					Type I error of different statistical analysis methods						Percent of 500 simulations yielding best estimate of...		
St. Dev grp 1	St. dev. grp 2	St. Dev grp 3	Ctrl. Minor allele frequency	Change in cases	T-test (equal variance)	T-test (unequal variance)	Mann-Whitney Test	χ^2 test	LRT ₃	LRT _{est}	1 comp.	2 comp.	3 comp.
0.5	0.5	0.5	0.25	0.9	0.646	0.646	0.59	0.488	0.612	0.642	0.002	0.74	0.258
0.5	0.5	0.5	0.33	0.9	0.602	0.602	0.568	0.474	0.556	0.624	0.006	0.676	0.318
0.5	0.5	0.5	0.33	0.5	1	1	1	1	1	0.986	0.014	0.816	0.17
0.5	0.5	0.5	0.5	0.9	0.446	0.446	0.456	0.336	0.38	0.342	0.196	0.608	0.196
0.5	0.5	0.5	0.5	0.75	0.986	0.986	0.99	0.972	0.988	0.926	0.068	0.694	0.238
0.5	0.75	1	0.05	0.9	0.856	0.856	0.642	0.692	0.75	0.91	0	1	0
0.5	0.75	1	0.05	0.75	1	1	1	0.906	0.99	1	0	1	0
0.5	0.75	1	0.05	0.5	1	1	1	0.89	1	1	0	1	0
0.5	0.75	1	0.1	0.9	0.762	0.762	0.62	0.542	0.614	0.778	0	1	0
0.5	0.75	1	0.1	0.75	1	1	1	0.812	0.968	1	0	1	0
0.5	0.75	1	0.1	0.5	1	1	1	0.838	1	1	0	1	0
0.5	0.5	0.5	0.05	0.9	0.896	0.896	0.784	0.848	0.912	0.966	0	1	0
0.5	0.5	0.5	0.1	0.9	0.816	0.816	0.7	0.664	0.746	0.852	0	1	0
0.5	0.75	1	0.25	0.9	0.546	0.546	0.482	0.29	0.434	0.508	0	1	0
0.5	0.75	1	0.25	0.75	1	1	1	0.934	0.99	0.992	0	1	0
0.5	0.75	1	0.33	0.9	0.462	0.462	0.412	0.298	0.402	0.464	0	0.996	0.004
0.5	0.75	1	0.33	0.75	0.998	0.998	0.996	0.924	0.98	0.984	0	0.99	0.01
0.5	0.75	1	0.5	0.9	0.326	0.326	0.308	0.196	0.238	0.264	0	0.452	0.548
0.5	0.75	1	0.5	0.75	0.938	0.938	0.942	0.828	0.89	0.904	0	0.454	0.546
0.25	0.375	0.5	0.25	0.9	0.736	0.736	0.614	0.236	0.658	0.654	0	1	0
0.25	0.375	0.5	0.25	0.75	1	1	1	0.256	1	1	0	0.928	0.072
0.25	0.375	0.5	0.33	0.9	0.658	0.658	0.566	0.542	0.596	0.596	0	0.002	0.998
0.25	0.375	0.5	0.5	0.9	0.492	0.492	0.49	0.37	0.4	0.4	0	0	1
0.25	0.375	0.5	0.5	0.75	0.998	0.998	0.998	0.986	0.998	0.998	0	0	1
0.25	0.25	0.25	0.25	0.9	0.776	0.776	0.638	0.224	0.792	0.706	0	0.712	0.288

0.25	0.25	0.25	0.33	0.9	0.696	0.696	0.586	0.686	0.692	0.692	0	0	1
0.25	0.25	0.25	0.5	0.9	0.516	0.516	0.5	0.508	0.502	0.502	0	0	1
0.25	0.25	0.25	0.05	0.9	0.992	0.992	0.906	0.086	0.586	1	0	1	0
0.25	0.25	0.25	0.05	0.75	1	1	1	0	0.876	1	0	1	0
0.25	0.25	0.25	0.05	0.5	1	1	1	0.282	1	1	0	1	0
0.25	0.25	0.25	0.1	0.9	0.956	0.956	0.848	0.002	0.862	0.972	0	1	0
0.25	0.25	0.25	0.1	0.75	1	1	1	0	0.978	1	0	1	0
0.25	0.25	0.25	0.1	0.5	1	1	1	0.066	1	1	0	1	0
0.25	0.375	0.5	0.05	0.9	0.984	0.984	0.84	0.01	0.634	0.99	0	1	0
0.25	0.375	0.5	0.05	0.75	1	1	1	0	0.904	1	0	1	0
0.25	0.375	0.5	0.05	0.5	1	1	1	0	1	1	0	1	0
0.25	0.375	0.5	0.1	0.9	0.934	0.934	0.792	0	0.826	0.934	0	1	0
0.25	0.375	0.5	0.1	0.75	1	1	1	0	0.966	1	0	1	0
0.25	0.375	0.5	0.1	0.5	1	1	1	0.206	1	1	0	1	0

Supplemental Table 5. Descriptive statistics on the distributions of the 14 FA ratios

	Overall Mean	Overall SD	Case Mean	Case SD	Control Mean	Control SD
D9D C16	0.07	0.03	0.07	0.03	0.07	0.02
D9D C18	1.27	0.36	1.3	0.4	1.24	0.31
D9D 16:18	0.05	0.01	0.05	0.01	0.05	0.01
ELOVL6	0.66	0.12	0.64	0.13	0.67	0.11
ELOVL6 mono	0.08	0.03	0.08	0.03	0.08	0.03
FADS2	0.03	0.01	0.03	0.01	0.03	0.01
ELOVL5	4.38	1.37	4.75	1.41	4.02	1.22
FADS5	8.43	2.45	8.49	2.63	8.38	2.25
ELOVL2 n3	3.61	1.33	4.04	1.21	3.19	1.3
ELOVL2 n6	0.19	0.04	0.2	0.04	0.18	0.04
ELOVL2 n3:n6	18.93	5.39	20.5	4.8	17.37	5.48
SCSn6	0.22	0.05	0.23	0.05	0.22	0.05
SCSn3	1.86	0.63	1.76	0.57	1.96	0.67
SCS n3:n6	8.71	3.49	8.04	2.94	9.38	3.86

Supplemental Table 6

Supplemental Table 6

			Non-essential FA metabolism					EFA elongation to eicosanoids			Elongation of eicosanoids to docosanoids							
			D9D			ELOVL6					ELOVL 2			Pathway summary				
			D9D C16	D9D C18	D9D 16:18	ELOVL6	ELOVL6 mono	FADS2	ELOVL5	FADS5	ELOVL2 n3	ELOVL2 n6	ELOVL2 n3:n6	SCSn6	SCSn3	SCS n3:n6		
Non-essential FA metabolism	D9D	D9D C16																Key Rho Color
		D9D C18																
		D9D 16:18																
	ELOVL6	ELOVL6																
		ELOVL6 mono																
EFA elongation to eicosanoids			FADS2															
			ELOVL5															
			FADS5															
Elongation of eicosanoids to docosanoids	ELOVL 2	ELOVL2 n3																
		ELOVL2 n6																
		ELOVL2 n3:n6																
	Pathway summary	SCSn6																
		SCSn3																
		SCS n3:n6																

Supplemental Table 7a: Fitting finite normal (log-normal) mixture models using BIC EQUAL VARIANCE

	Norm_1	Norm_2	Norm_3
D9D C16	6840	7104	7090
D9D C18	-1236	-984	-999
D9D 16:18	8911	9122	9107
ELOVL6	2067	2051	2051
ELOVL6 mono	6335	6703	6688
FADS2	10721	10948	10934
ELOVL5	-5287	-5210	-5226
FADS5	-7092	-7034	-7049
ELOVL2 n3	-5217	-5226	-5214
ELOVL2 n6	5745	5729	5722
ELOVL2 n3:n6	-9501	-9515	-9496
SCSn6	4742	4839	4824
SCSn3	-2943	-2743	-2758
SCS n3:n6	-8169	-7857	-7871

MClust (R package) was used to assess the fit (using BIC) of a single or multiple component (up to 3)

Gaussian model for a total of 3 models fit to each of the 14 FA ratios. The model giving the larger BIC is highlighted in yellow.

Using the guidelines of Raftery (1995) for comparing model BIC's

FAs with very strong evidence (diff in BIC of >10; like $p < 0.006$): D9D-C16, D9D-C18, D9D-C16-C18,

ELOVL6, ELOVL6-mono, FADS2, ELOVL5, FADS5, ELOVL2_N6, SCS-N3, SCS-N6, SCS-N3-N6

FAs with strong evidence (diff in BIC 6-10; like $0.006 < p < 0.05$): NONE

FAs without strong evidence (two “best” fitting models:): Go with more parsimonious model:

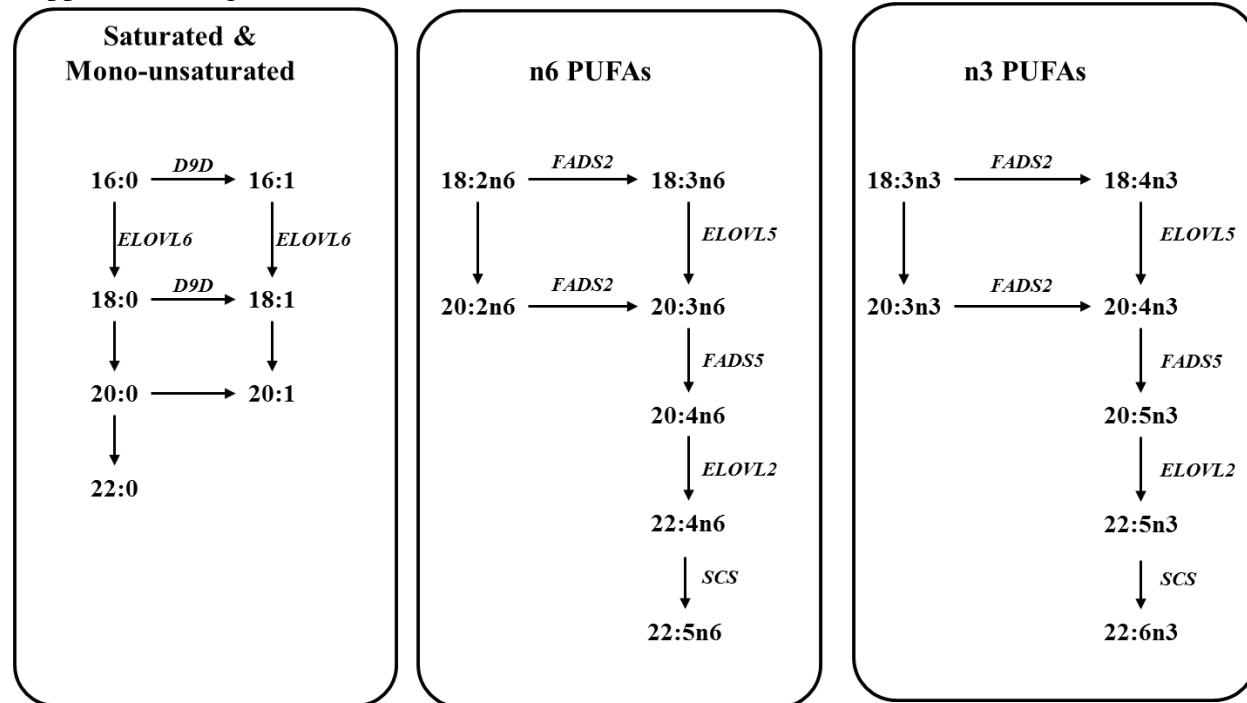
ELOVL2_N3, ELOVL2_N3_N6,

Supplemental Table 7b: Fitting finite normal (log-normal) mixture models using BIC EQUAL VARIANCE (after removal of the most extreme 0.5% of values in the sample)

	Norm_1	Norm_2	Norm_3
D9D C16	6919	7180	7165
D9D C18	-1146	-901	-916
D9D 16:18	8972	9165	9150
ELOVL6	2140	2127	2121
ELOVL6 mono	6425	6777	6762
FADS2	10839	11047	11032
ELOVL5	-5191	-5116	-5133
FADS5	-6939	-6883	-6900
ELOVL2 n3	-5126	-5131	-5124
ELOVL2 n6	5783	5769	5764
ELOVL2 n3:n6	-9349	-9359	-9362
SCSn6	4839	4923	4907
SCSn3	-2833	-2652	-2667
SCS n3:n6	-8018	-7722	-7737

All best fitting models were the same as without the most extreme values.

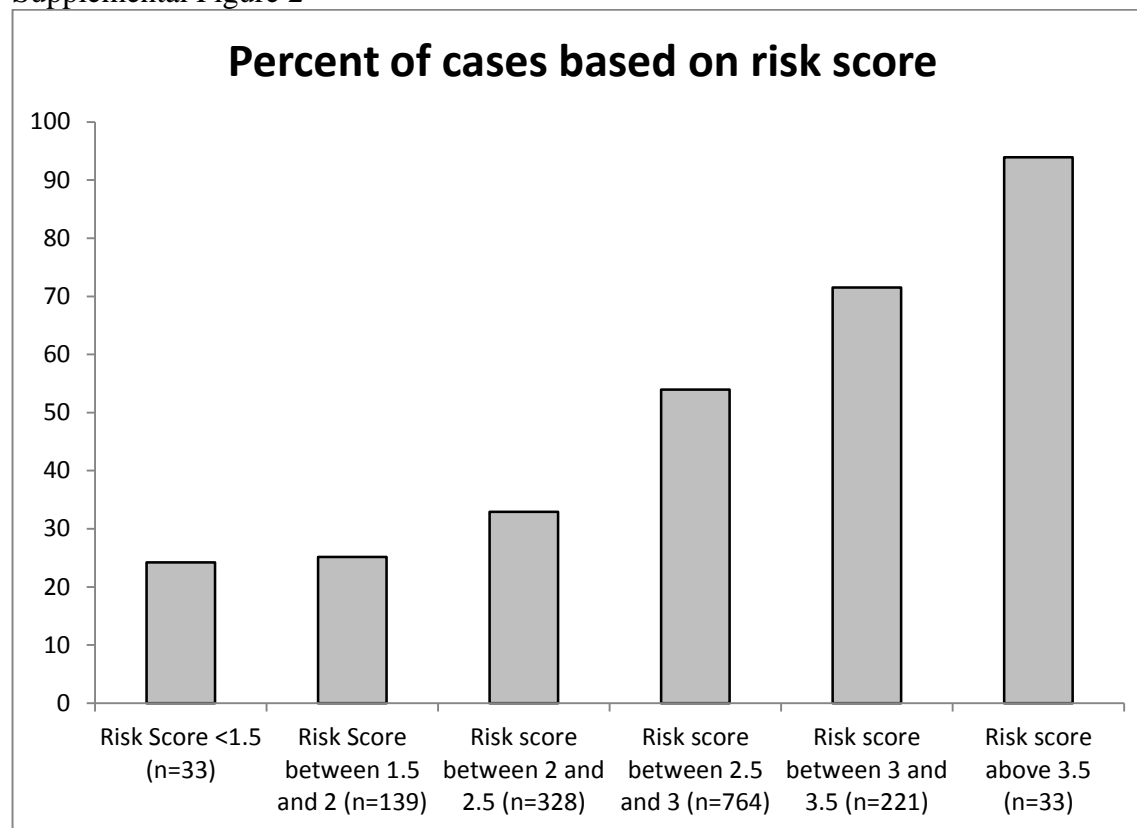
Supplemental Figure 1a



Supplemental Figure 1b

General Activity	Non-essential fatty acid metabolism						Essential fatty acid metabolism								
Activity	desaturase PA: n7 position SA: n9 position			elongase 16 carbon FAs			to eicosanoids			to docosanoids					
Comments	16-C	18-C	Efficiency	Sat	Mono	Efficiency	AA step1	AA step2	AA step3	omega-3	omega-6	Efficiency	omega-6	omega-3	Efficiency
Associated Gene	D9D			ELOVL6			FADS2	ELOVL5	FADS5	ELOVL2			Multiple: Pathway summary		
Activity marker Ratio	POA:PA	OA:SA	[POA:PA]/[OA:SA]	SA:PA	OA:POA	[SA:PA]/[OA:POA]	gLA:LA	dgLA:dgLA	AA:dgLA	n3-DPA:EPA	DTA:AA	n3:n6	n6-DPA:AA	EPA:DHA	n3:n6
Abbreviation	D9D C16	D9D C18	D9D 16:18	ELOVL6	ELOVL6 mono	n/a	FADS2	ELOVL5	FADS5	ELOVL2 n3	ELOVL2 n6	ELOVL2 n3:n6	SCSn6	SCSn3	SCS n3:n6

Supplemental Figure 2



References

1. Assmann G, Cullen P, & Schulte H (2002) Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 105(3):310-315.
2. De Backer G, *et al.* (2004) European guidelines on cardiovascular disease prevention in clinical practice. Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts). *Atherosclerosis* 173(2):381-391.
3. Chambless LE, *et al.* (2003) Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *Journal of clinical epidemiology* 56(9):880-890.
4. Ridker PM, Paynter NP, Rifai N, Gaziano JM, & Cook NR (2008) C-Reactive Protein and Parental History Improve Global Cardiovascular Risk Prediction. The Reynolds Risk Score for Men. *Circulation*.
5. Ridker PM, Buring JE, Rifai N, & Cook NR (2007) Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *Jama* 297(6):611-619.
6. D'Agostino RB, Sr., *et al.* (2008) General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 117(6):743-753.
7. Wilson PW, *et al.* (1998) Prediction of coronary heart disease using risk factor categories. *Circulation* 97(18):1837-1847.
8. Vasan RS (2006) Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* 113(19):2335-2362.
9. Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115(7):928-935.
10. Cook NR (2008) Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* 54(1):17-23.
11. Lloyd-Jones DM, Liu K, Tian L, & Greenland P (2006) Narrative Review: Assessment of C-Reactive Protein in Risk Prediction for Cardiovascular Disease. *Ann Intern Med*.
12. Folsom AR, *et al.* (2006) An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the atherosclerosis risk in communities study. *Archives of internal medicine* 166(13):1368-1373.
13. Morgan TM, Krumholz HM, Lifton RP, & Spertus JA (2007) Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA* 297(14):1551-1561.
14. Shearer GC, Pottala JV, Spertus JA, & Harris WS (2009) Red blood cell fatty acid patterns and acute coronary syndrome. *PloS one* 4(5):e5444.
15. Holmes E, Wilson ID, & Nicholson JK (2008) Metabolic phenotyping in health and disease. *Cell* 134(5):714-717.
16. Vessby B, Gustafsson IB, Tengblad S, Boberg M, & Andersson A (2002) Desaturation and elongation of Fatty acids and insulin action. *Ann N Y Acad Sci* 967:183-195.
17. Block RC, Harris WS, Reid KJ, Sands SA, & Spertus JA (2007) EPA and DHA in blood cell membranes from acute coronary syndrome patients and controls. *Atherosclerosis*.
18. Fraley C, Raferty AE, Murphy TB, & Scrucca L (2012) mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation (Department of Statistics, University of Washington, Seattle, WA), 4.
19. Kim W, Gordon D, Sebat J, Ye KQ, & Finch SJ (2008) Computing power and sample size for case-control association studies with copy number polymorphism: application of mixture-based likelihood ratio test. *PloS one* 3(10):e3475.

20. Raftery AE (1995) Bayesian model selection in social research. *Sociol Methodol* 25:111-163.
21. Mendell NR, Finch SJ, & Thode HC, Jr. (1993) Where is the likelihood ratio test powerful for detecting two component normal mixtures? *Biometrics* 49(3):907-915.
22. Liu K, Luedtke A, & Tintle N (2013) Optimal methods for using posterior probabilities in association testing. *Human heredity* 75(1):2-11.
23. Dumas ME (2012) Metabolome 2.0: quantitative genetics and network biology of metabolic phenotypes. *Molecular bioSystems* 8(10):2494-2502.
24. Shin SY, *et al.* (2014) An atlas of genetic influences on human blood metabolites. *Nat Genet* 46(6):543-550.
25. Flowers MT & Ntambi JM (2008) Role of stearyl-coenzyme A desaturase in regulating lipid metabolism. *Current opinion in lipidology* 19(3):248-256.
26. Jeyakumar SM, *et al.* (2009) Fatty acid desaturation index correlates with body mass and adiposity indices of obesity in Wistar NIN obese mutant rat strains WNIN/Ob and WNIN/GR-Ob. *Nutrition & metabolism* 6:27.
27. Rhee SK, Kayani AJ, Ciszek A, & Brenna JT (1997) Desaturation and interconversion of dietary stearic and palmitic acids in human plasma and lipoproteins. *The American journal of clinical nutrition* 65(2):451-458.
28. Warensjo E, Sundstrom J, Vessby B, Cederholm T, & Riserus U (2008) Markers of dietary fat quality and fatty acid desaturation as predictors of total and cardiovascular mortality: a population-based prospective study. *The American journal of clinical nutrition* 88(1):203-209.
29. Green CD & Olson LK (2011) Modulation of palmitate-induced endoplasmic reticulum stress and apoptosis in pancreatic beta-cells by stearyl-CoA desaturase and Elovl6. *American journal of physiology. Endocrinology and metabolism* 300(4):E640-649.
30. Liu Y, *et al.* (2013) Genetic analysis of the ELOVL6 gene polymorphism associated with type 2 diabetes mellitus. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.]* 46(7):623-628.
31. Morcillo S, *et al.* (2011) ELOVL6 genetic variation is related to insulin sensitivity: a new candidate gene in energy metabolism. *PloS one* 6(6):e21198.
32. Matsuzaka T, *et al.* (2012) Elovl6 promotes nonalcoholic steatohepatitis. *Hepatology* 56(6):2199-2208.
33. Aslibekyan S, *et al.* (2012) Genetic variation in fatty acid elongases is not associated with intermediate cardiovascular phenotypes or myocardial infarction. *European journal of clinical nutrition* 66(3):353-359.
34. Lu Y, *et al.* (2012) Markers of endogenous desaturase activity and risk of coronary heart disease in the CAREMA cohort study. *PloS one* 7(7):e41681.
35. Wang L, Folsom AR, & Eckfeldt JH (2003) Plasma fatty acid composition and incidence of coronary heart disease in middle aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Nutrition, metabolism, and cardiovascular diseases : NMCD* 13(5):256-266.