Bayesian Multivariate Genetic Analysis Boosts Translational Insights

Sarah Urbut (surbut@mgh.harvard.edu)
Massachusetts General Hospital https://orcid.org/0000-0002-1135-9647
Satoshi Koyama
Broad Institute of MIT and Harvard
Sunitha Selvaraj
Broad Institute of MIT and Harvard
Benjamin Neale
Massachusetts General Hospital https://orcid.org/0000-0003-1513-6077
Christopher O'Donnell
Brigham and Women's Hospital
Gina Peloso
Boston University https://orcid.org/0000-0002-5355-8636
Pradeep Natarajan
Massachusetts General Hospital https://orcid.org/0000-0001-8402-7435

Article

Keywords:

Posted Date: February 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1319235/v1

Bayesian Multivariate Genetic Analysis Boosts Translational Insights

Sarah M Urbut, MD PhD[1,2], Satoshi Koyama, MD PhD[1,2,3], Margaret S Selvaraj, PhD[1,2,3], Benjamin Neale, PhD[2,3,4]; Christopher J O'Donnell, MD MPH[3,5], Gina M Peloso, PhD[6], Pradeep Natarajan, MD MMSc[1,2,3*]

Affiliations:

[1] Cardiovascular Research Center, Massachusetts General Hospital; Boston, MA 02114 USA
[2] Program in Medical and Population Genetics, Broad Institute; Cambridge, MA 02142 USA
[3] Department of Medicine Harvard Medical School; Boston, MA 02115 USA
[4] Analytic Translational and Genetics Unit, Massachusetts General Hospital; Boston, MA 02114 USA
[5] VA Boston Dept of Veterans Affairs; Boston, MA 02130 USA
[6] Department of Biostatistics, Boston University School of Public Health; Boston, MA 02218 USA

*Please address correspondence to:
Pradeep Natarajan, MD MMSc
185 Cambridge Street, Boston, MA 02114
Office: 617-726-1843
Fax: 617-726-2203
Email: *pnatarajan@mgh.harvard.edu*
Twitter: @pnatarajanmd

Words: 3200
Figures: 4

Abstract

Conventional genome-wide association studies interrogate individual variants against individual traits. We introduce a multivariate adaptive shrinkage (mash) model that allows each variant to be modeled as a mixture of multivariate normal distributions, boosting power when genetic effects across conditions are shared, using summary statistics. We show that controlling local false sign rates accurately and powerfully identifies replicable genetic associations, and that multivariate control furthers the ability to explain complex disease. We apply this framework to the genetic analyses of blood lipid levels, principal predictors and therapeutic targets for coronary artery disease. Our approach yields high concordance between independent datasets, more accurately prioritizes causal genes, and significantly improves polygenic prediction beyond state-of-the-art methods by up to 59% for lipid traits. Importantly, we describe a framework with important implications for genome-wide association studies and polygenic risk score construction.

Introduction

A principal goal of genome-wide association studies (GWAS) is to accurately identify genetic variants that influence risk of developing a phenotype of interest. For the vast majority of complex traits, the number of significantly associated variants increases with increasing sample size but only the minority of estimated heritability remains explained[1,2]. Novel Bayesian methods leveraging genetic pleiotropy applied to existing samples may improve power for genetic discovery beyond widespread existing univariate approaches [3,4]. As many phenotypes in biology exist on quantitative and continuous spectra describing multivariate continuous effects is an essential step towards better understanding of complex phenotypes.

Multivariate adaptive shrinkage (mash) is a Bayesian adaptive shrinkage tool designed to estimate genetic variants associated with multiple phenotypes, thus taking advantage of any correlation in genetic signals that might boost power to detect associations and improve precision of effect size estimates[4,5]. Mash uses empirical estimates of overall covariance structure to model the genetic effect at any given single nucleotide polymorphism (SNP) as a mixture of multivariate normal distributions, where each component of the mixture defines a 'pattern of sharing' of effects among conditions from which the SNP might arise **(Fig 1)** and is readily available using the accompanying software mashR[4]. Using empirical Bayes methods, the patterns of covariance are estimated from the strongest signals in the dataset, scaled to a unit variance, and 'stretched' by a grid of magnitudes to reflect an abundance of shape-scale combinations. The relative frequency of each shape-scale combination is then estimated from a sampling of the overall dataset. Armed with this prior information, the likelihood for each SNP is then calculated for each mixture component. If the effects are strongly correlated, such estimates will receive a boost in precision and power by incorporation of additional data. Importantly, if there is no sharing, multivariate estimates have been shown to do no worse than univariate estimates[4,6].

Genome-wide association studies for blood lipid concentrations have advanced our understanding of causal relationships for diverse cardio-metabolic conditions, including coronary artery disease (CAD), the leading cause of death worldwide[7]. Since currently identified variants explain only a small fraction of the estimated overall heritabilities of plasma lipids[8,9], improved efficiency for lipid genetic associations may yield new insights using existing genetic association data.

Results

Multivariate genetic discovery for plasma lipids. Using summary-level lipid GWAS data generated with conventional methods from the Million Veterans Project (MVP) across 291,746 individuals (210,967; 72.3% European ancestry[8])*,* we sought first to jointly estimate effect sizes across the four plasma lipid traits (i.e., total cholesterol [TC], low-density lipoprotein [LDL-C] cholesterol, high-density lipoprotein [HDL-C] cholesterol, and triglycerides [TG]) for approximately 28 million SNPs using mash. We identify 5,007 500-kb linkage disequilibrium

(LD) blocks (from 1000 Genomes European Samples[10]) in the MVP data set containing at least one variant with a local false sign rate less than 0.05 for at least one lipid trait. Here, the local false sign rate *(lfsr)* is defined as the posterior probability of incorrectly identifying the sign of the effect, and has been used to help provide a bridge between FDR (false discovery rates) and effect size estimation[5].

Omnigenic model. We next used mashR to analyze the distribution of regression coefficients from the set of all SNPs[5]. Mash models the GWAS results as a mixture of SNPs that have a true effect size of exactly zero, with SNPs that have a true effect size that is not zero, across multiple traits. The additional multivariate layer captures information about patterns of sharing across traits when compared to univariate shrinkage approaches.[5] Using this approach, we estimated that 84% of all SNPs present in the MVP data set are associated with non-zero effects on LDL cholesterol, including both causal SNPs as well as nearby SNPs in linkage disequilibrium (LD). Critically, this does not imply that the majority of variants are causal: given that the typical extent of LD is around 10kb–100kb[11], this implies that most 100kb windows in the genome include variants that affect lipid levels. Stratifying by the LD Score for each SNP[12], we see a clear effect that SNPs with more LD partners are more likely to be associated with each lipid trait (**Fig 2A)**. Briefly, this was LD Scores from the European ancestry samples in the 1000 Genomes Project (EUR) using an unbiased estimator of r^2 with 1 centiMorgan (cM) windows.[12] This is consistent with the conclusions of Boyle et al[13] who introduced the idea of an 'omnigenic' hypothesis in which many quantitative traits are influenced by the majority of genomic SNPs, each contribution and that there is an extremely large number of causal variants with tiny effect sizes on quantitative traits.

Boyle et al[13] found clear enrichment of shared directional signal for most SNPs, even for SNPs with p-values as large as 0.5, leading us to consider the information contained in those with non-null (but potentially not genome wide significant effects). After applying mashR, we find that the median absolute effect size for a SNP satisfying an *lfsr* threshold of 0.05 is 0.1533 while the median absolute effect size for a SNP satisfying a local false discovery rate (*lfdr)* of 0.05 is 0.1528 and one satisfying a genome-wide significance threshold of 5×10^{-8} is 6.54, consistent with the results of Boyle et al[13] who found that the median effect of non-null SNPs was approximately 10% that of genome-wide significant SNPs (**Fig 2B-D)** Importantly, both the resolution and stringency using *lfsr* thresholds is greater than *lfdr* for a given level of evidence **(Fig 2E).**

Under this observation, we hypothesized that a method which is capable of both incorporating refined joint posterior effect size estimates and quantifying posterior probabilities of being non-zero (or, even more stringently, incorrectly signed) might add to the ability to explain the heritability of complex disease in a polygenic risk score. First, we sought to compare directly to a univariate approach for adaptive Bayesian shrinkage[5] as published GWAS have acknowledged that using an FDR threshold as previous approaches have validated the use of a 5% FDR threshold to replicate GWAS targets[14]. We show that a univariate shrinkage approach which controls for local false discovery (ash) replicates everything in previously published MVP data[8] and results in a consistent power increase across phenotypes **(Fig 3A, Fig 3C**, green to

blue**)** owing to the adaptive control of local false discovery when compared with Bonferroni-corrected GWAS. We depict the number of LD blocks containing an effective variant in at least one block, defined by p-value $<5x10^{-8}$ in traditional analyses or *lfsr* $<5x10^{-2}$ in adaptive shrinkage analyses.

We can then ascertain the improvement in multivariate control of false discovery when comparing multivariate adaptive shrinkage to univariate adaptive shrinkage (**Fig 3-C**, blue to red**).** Furthermore, most effects are shared across conditions, owing to the assessment of effects jointly, in comparison to univariate shrinkage effects (**Fig 3B).**

Improved polygenic risk scoring across ancestries. When compared to weights derived from traditional univariate maximum likelihood association mapping, mash-derived weights show improvements up to 58% in the proportion of variance explained by the polygenic score. Furthermore, these improvements hold when subdividing the testing population into European and Non-European individuals on all traits excluding triglycerides, a phenomenon which has been observed elsewhere[15](**Fig 4A-C, Table S1**). Of note, the baseline performance of models excluding genetic variables from association with LDL-C is quite poor owing to the phenotypic heterogeneity inherent in statin dosing and duration present in the UKBB population, and yet mash inputs still impove the proportion of variation explained by a considerable (53%). Importantly*,* we shrink the input 'weights' derived from our multivariate discovery set using LD score regression from the 1000 Genomes reference [10]. We then compute scores on a completely independent set of individuals in the UKBB, without the need for parameter tuning in an additional set of data. This is due to the analytic solution innate in the infinitesimal model introduced by Privé[16] (details in Materials and Methods).

In simulation in which we simulate true effects arising from the empirical patterns of sharing identified by mash in the MVP data, we find that mash estimates are superior in terms of precision (relative root mean squared error, RRMSE) and in power compared with typical univariate shrinkage approaches and with canonical multivariate methods that attempt to model the set of multivariate effects as arising from a mixture of distributions constrained by a rigid configuration approach[6] (**Extended Data Figure 1)**. Furthermore, when all associations are simulated to be truly null, our mash method detects 0 false positive associations at a *lfsr* threshold of up to 0.54 (and 0 associations are detected at an *lfsr* of 0.05), indicating the accuracy of the false sign rate for thresholding purposes. Furthermore, mash allows us to consider which hierarchical patterns receive the most `weight' in the mixture model and as expected the matrices which received most of the hierarchical weight showed effects that were shared in sign and magnitude among LDL-C, TC and TG and strongly inversely correlated with strong effects in HDL-C (**Extended Data Figure 2**).

Sources of improved heritability explained. To understand from where the additional heritability explained arises, we sought to evaluate extent of cross-replication between UK Biobank (UKBB) and MVP results. After computing mash posterior means for all overlapping

SNPs from models fit separately using MVP and UKBB (**Extended Data Figure 3**), we aimed to identify replication on a per-trait and across trait basis. Using UKBB lipid genetic summary statistics for 315,133 individuals (100 % European ancestry)[17] overlapping the same boundaries used above, we identify 3,935 500 kb-LD blocks containing at least one *lfsr* < 0.05 across traits. 3761 of these 3935 of these identified UKBB blocks are replicated in the MVP discovery. Therefore, 95.6% and 75.1% of the discoveries in UKBB and MVP, respectively, cross-replicate (**Fig 5A; Extended Data Figure 1)**. Additionally, we reproduce all associations captured by the prevailing multivariate GWAS package MTAG[18] and we replicate 12-fold (**Fig 5A; Extended Data Figure 3**) more LD blocks when comparing non-zero blocks containing at least one significant effect across traits between mtag[18] and mash-posterior results without incurring additional false positives (see also simulations, **Extended Data Figure 1**).

In a per-trait analysis, 68% of MVP discoveries and 88% of UKBB cross-replicate for HDL-C, 42% and 82% of MVP and UKBB discoveries cross-replicate for LDL-C, 63% MVP and 91% UKBB for triglycerides, and 50% MVP and 81% UKBB for total cholesterol. Again, when comparing to existing multivariate approaches for common disease using the software MTAG[18], this boost holds across traits. Taken with the result that comparing to traditional univariate assessments of the MVP data set at a genome-wide family wise error-rate of $5x10^{-8}$, we identify 15-fold more blocks containing a variant significant in at least one trait (**Fig 3A**), we see that our Bayesian multivariate effect size estimate using mash improves our sensitivity to both detect and replicate associations between datasets.

We find that this 15-fold improvement over univariate association is driven largely by improvement in precision, and also by control of false discovery versus family-wise error rate, a concept shown to safely increase the power to detect associated non-zero effects [19], [14]. The effective sample size can be determined by considering the ratio of the original standard error to the posterior standard error for an individual trait (**Equation 1):** We find that effects with small sample sizes (and accordingly large standard errors) benefit from the greatest effective sample boost (that is, increase in posterior sample size over initial sample size).

$$\mathrm{n}_{\mathrm{effective}} = \frac{\hat{s}^2_{\mathrm{original}}}{\tilde{s}^2_{\mathrm{posterior}}} \mathrm{n}_{\mathrm{original}} \quad (1)$$

We find a median effective sample boost of 4.4-fold (IQR 3.54-5.64) using the relationship between the original standard error and the posterior marginal variance, consistent with the strong sharing among lipid subgroups (**Extended Data Figure 4**).

Improved causal gene prioritization. We next sought to investigate how mash would prioritize known Mendelian lipid targets in comparison to univariate methods. We used the Polygenic Priority Score (PoPS), which is a gene prioritization method [20] that leverages genome-wide signal from GWAS summary statistics and incorporates data from an extensive set of public bulk and single-cell expression datasets, curated biological pathways, and predicted protein-

protein interactions. Compared to univariate summary statistics, we found that marginal estimates from posterior means supplied by mash better prioritizes known lipid candidate genes and does so more consistently between traits. Among the top 100 prioritized genes for each trait, there are 47 intersections among HDL-C, LDL-C, and TG using mash and only 27 using raw univariate estimates: of these 47, 23 are shared with those prioritized by the univariate approach **(Fig 5B, Table S2**). Consequently, there are 24 genes that are found consistently by mash across all 4 lipid traits but not by univariate methods that include *LPL* (lipoprotein lipase), *CETP* (cholesterol ester transfer protein), *APOA4* (apolipoprotein A4), *LIPG* (endothelial lipase) and *ADIPOQ* (adiponectin precursor) which have established relevance to lipids in model systems and human studies. However, the 4 genes that are found consistently by univariate methods but not by multivariate methods were *VEGFA*, *TP53*, *ESR1*, and *ENG,* which are not currently known to robustly influence lipids in model systems or human studies. When comparing a list of known Mendelian lipid genes[21], the median rank assigned using mash results is consistently higher than when using univariate association summary statistics (**Tables S3,S4.**).

Improved enrichment for functional annotations. We next sought to determine whether mash discoveries better prioritize biologically relevant associations compared to those observed from conventional univariate methods. Using TORUS and conservatively defined LD blocks [22,23], we show that mash versus univariate statistics for this analysis improves estimation of expected enrichment parameters and depletion accordingly using a list of previously described annotations[12]. For example, we demonstrate that areas of the genome known to be enriched for transcriptional activity (i.e., super enhancers) and promoters show stronger log odds ratio (base 2) of enrichment using mash-derived summary statistics compared to the estimation from using univariate association statistics to assess annotation enrichment parameters (**Fig 5C; Extended Data Figures 5-7, Table S5, S6)**. Similarly, annotations with prior evidence for depletion (i.e., repressors, background selection) show a greater degree of depletion using multivariate summary statistics, as evidenced by background selection and repressors which are repressed to a greater degree using mash (p=0.001) and coding regions which are enriched to a greater degree using mash (p=6e-5;1e-4, see **Fig 5C**) compared with using univariate association statistics. The direction of enrichment or depletion is preserved between univariate and mash multivariate estimates.

Discussion

Understanding the genetic basis of common disease is a key paradigm toward modern genomic medicine goals. The principal strategy for improving power for discovery is increasing sample size to improve precision and including diverse ancestries to analyze new alleles. Here, we introduce a multivariate approach for enabling both improved discovery and effect size estimation of genomic variants without increasing the sample size. Not only do such joint

approaches share information across traits to improve power, but critically, we show that improved robust effect size estimates enable more precise prioritization of genomic targets, enhance assessment of enrichment parameters, and finally, show an improvement in prediction on a polygenic basis when coupled with methods to shrink across LD blocks.

False discovery rates offer a flexible way of capturing inherent differences in the relative signal between populations while controlling the proportion of discoveries that are false. Local false discovery rates are thus a compromise between arbitrarily stringent p-values that allow one to include the prior probability of absence of signal (often termed π_0) in computing the posterior probability of being null and is widely accepted in the genomics community[14,24,25]. The local false sign rate is analogous to the "local false discovery rate" (*lfdr*), but measures confidence in the *sign* of each in addition to confidence in each effect being non-zero, and is thus more stringent.[4]

Importantly, we also apply the use of the local false sign rate to characterize effects, a convenient method of controlling for multiple hypotheses while also incorporating the consistency in sign and magnitude of effects. It is more stringent than the *lfdr* because it requires true discoveries to be not only nonzero, but also correctly signed, and the local false discovery rate is well known to conservatively control for false discoveries in genomic applications[19,26,5]. Furthermore, in some settings with many discoveries, the *lfsr* and *lfdr* can be quite different and emphasize benefits of the *lfsr,* particularly its increased robustness to modeling assumptions [5]. This allows for a multiple hypothesis correction that incorporates the precision of the effect size estimate[5] beyond confidence in the binary classification of a variant as associated or unassociated.

Putting the various lines of evidence together, considering the non-null distribution of SNPs provides a much broader understanding of the collective contribution of genetic variation to quantitative phenotype mapping, and the inclusion of local false sign rates broadens this resolution as a non-null SNP can have varying effect sizes (**Figure 2C**). Our multivariate mapping furthers this resolution as two SNPs with the same local false sign rate can have different posterior effect estimates arising from the mixture of multivariate normal distributions that depend on the 'boost' the SNP receives from the effects in correlated phenotypes. More generally, the heritability of complex traits and diseases is spread broadly across the genome [27] implying that a large fraction of all genes contribute to variation in quantitative traits. We then combine this increased power from false discovery control with the precision and adaptive nature of estimates effects *across* conditions, presumably adding to the 15-fold increase in power. We show that these effects are biologically believable. In summary, we improve our power to both predict and to detect through control of false discovery instead of family wise error rate and through incorporation of additional information captured by between trait sharing of effect-size information.

While well-suited to genetic settings in which effects are additive, multivariate normal methods are limited by settings in which the effects are roughly normally distributed in each individual trait. Furthermore, the benefit of such a method is stronger when the effects are more strongly

correlated than the errors. Perhaps most notably, after estimating effects, correction still must be done for linkage disequilibrium as such a method does not currently take correlation between SNPs into account. However, such work must also be done after univariate GWAS estimation, and here we provide a framework to do so with available linkage disequilibrium tools.[16]

In conclusion, these methods show the significant promise of multivariate approaches for genome-wide association studies of complex traits, and the improvement gained in both prediction and prioritization through control of false discovery of multivariate adaptively shrunk effect size estimates.

Acknowledgments: The authors would like to thank the participants and staff of the Million Veterans Program and the UK Biobank. The UK Biobank analyses were supported by application 7089.

Funding: This work was primarily supported by grants to GMP and PN from the National Heart, Lung, and Blood Institute (R01HL142711, R01HL127564). PN is also supported by additional grants from the National Heart, Lung, and Blood Institute (R01HL148050, R01HL151283, R01HL148565, R01HL151152), National Institute of Diabetes and Digestive and Kidney Diseases (R01DK125782), Foundation Leducq (TNE-18CVD04), and Massachusetts General Hospital (Fireman Endowed Chair in Vascular Medicine).

Author contributions:

Conceptualization: SMU, PN, GMP, CJO

Methodology: SMU, GMP, PN

Investigation: SMU, GMP, PN

Visualization: SMU, SK, PN

Funding acquisition: PN, CJO

Project administration: PN, MS

Supervision: PN, CJO, GMP, BN

Writing – original draft: SMU, PN

Writing – review & editing: SMU, MS, PN, CJO, GMP, BN

Competing interests: CJO is an employee of Novartis. PN reports investigator-initiated grants from Apple, Amgen, AstraZeneca, Boston Scientific, and Novartis, personal fees from Apple, AstraZeneca, Genentech, Novartis, Foresite Labs, TenSixteen Bio, and Blackstone Life Sciences, honorarium from Allelica, is a scientific advisory board member and shareholder of TenSixteen Bio and geneXwell, and spousal employment at

Vertex. All disclosures are unrelated to the present work. The other co-authors do not report any disclosures.

Data and materials availability: All data are available in the main text or the supplementary materials and links to the code to reproduce these analyses are available at the links included in the supplementary information first page. UK Biobank individual-level data are available upon application to the UK Biobank. MVP univariate lipid GWAS summary statistics were previously made available in dbGaP (phs001672).

Works Cited

1. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci*. 2012;109(4):1193-1198. Doi:10.1073/pnas.1119675109

2. Manolio T, Collins F, Cox N, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-753.

3. Zhu X, Stephens M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *bioRxiv*. Published online March 4, 2016:042457. Doi:10.1101/042457

4. Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet*. 2019;51(1):187-195. Doi:10.1038/s41588-018-0268-8

5. Stephens M. False discovery rates: a new deal. *Biostatistics*. Published online October 17, 2016:kxw041. Doi:10.1093/biostatistics/kxw041

6. Flutre T, Wen X, Pritchard J, Stephens M. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PloS Genet*. 2013;9(5):e1003486. Doi:10.1371/journal.pgen.1003486

7. Nowbar AN, Gitto M, Howard JP, Francis DP, Al-Lamee R. Mortality From Ischemic Heart Disease: Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration. *Circ Cardiovasc Qual Outcomes*. 2019;12(6). Doi:10.1161/CIRCOUTCOMES.118.005375

8. Klarin D, Damrauer SM, Cho K, et al. Genetics of Blood Lipids Among ~300,000 Multi-Ethnic Participants of the Million Veteran Program. *Nat Genet*. 2018;50(11):1514-1523. Doi:10.1038/s41588-018-0222-9

9. Peloso GM, Natarajan P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr Opin Genet Dev*. 2018;50:1-6. Doi:10.1016/j.gde.2018.01.003

10. Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J Proteome Res*. 2008;7(1):29-34. Doi:10.1021/pr700600n

11. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100(16):9440-9445. Doi:10.1073/pnas.1530509100

12. Käll L, Storey JD, MacCoss MJ, Noble WS. Posterior Error Probabilities and False

Discovery Rates: Two Sides of the Same Coin. *J Proteome Res*. 2008;7(1):40-44. Doi:doi: 10.1021/pr700739d

13. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Stat Soc Ser B Stat Methodol*. 2004;66(1):187-205. Doi:10.1111/j.1467-9868.2004.00439.x

14. Efron B. Size, power and false discovery rates. *Ann Stat*. 2007;35(4):1351-1377. Doi:10.1214/009053606000001460

15. Mei S, Karimnezhad A, Forest M, Bickel DR, Greenwood CMT. The performance of a new local false discovery rate method on tests of association between coronary artery disease (CAD) and genome-wide genetic variants. Li Q, ed. *PLOS ONE*. 2017;12(9):e0185174. Doi:10.1371/journal.pone.0185174

16. Ye Y, Chen X, Han J, Jiang W, Natarajan P, Zhao H. Interactions Between Enhanced Polygenic Risk Scores and Lifestyle for Cardiovascular Disease, Diabetes, and Lipid Levels. *Circ Genomic Precis Med*. 2021;14(1). Doi:10.1161/CIRCGEN.120.003128

17. 23andMe Research Team, Social Science Genetic Association Consortium, Turley P, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50(2):229-237. Doi:10.1038/s41588-017-0009-4

18. Weeks EM, Ulirsch JC, Cheng NY, et al. *Leveraging Polygenic Enrichments of Gene Features to Predict Genes Underlying Complex Traits and Diseases*. Genetic and Genomic Medicine; 2020. Doi:10.1101/2020.09.08.20190561

19. NHLBI TOPMed Lipids Working Group, Natarajan P, Peloso GM, et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat Commun*. 2018;9(1):3391. Doi:10.1038/s41467-018-05747-8

20. Wen X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann Appl Stat*. 2016;10(3). Doi:10.1214/16-AOAS952

21. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*. Published online September 22, 2015:btv546. Doi:10.1093/bioinformatics/btv546

22. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. Doi:10.1038/ng.3211

23. ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.

Doi:10.1038/ng.3404

24. Privé F, Arbel J, Vilhjálmsson BJ. Ldpred2: better, faster, stronger. Schwartz R, ed. *Bioinformatics*. 2021;36(22-23):5424-5431. Doi:10.1093/bioinformatics/btaa1029

25. Clarke L, Fairley S, Zheng-Bradley X, et al. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res*. 2017;45(D1):D854-D859. Doi:10.1093/nar/gkw829

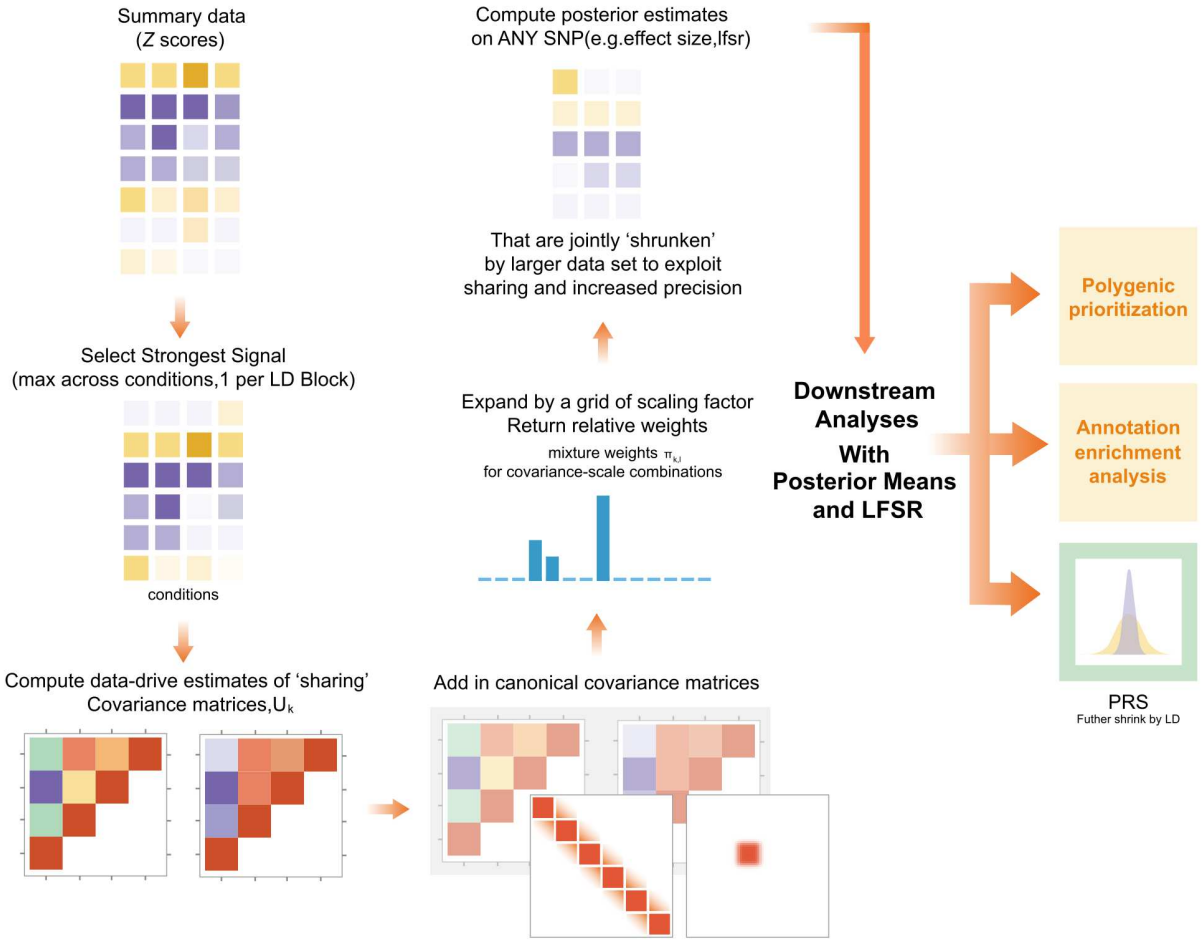
26. Bovy J, Hogg DW, Roweis ST. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Ann Appl Stat*. 2011;5(2B):1657-1677. Doi:10.1214/10-AOAS439

27. Wen X, Stephens M. Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene–environment interactions. *Ann Appl Stat*. 2014;8(1):176-203. Doi:10.1214/13-AOAS695

28. Qiao, Min, Wang, Gao. *TORUS Workflow*. https://github.com/cumc/bioworkflows/blob/master/fine-mapping/gwas_enrichment.ipynb

29. Weeks EM. *PoPs Software*. https://github.com/FinucaneLab/pops/blob/master/README.md

30. Zhang, Mengyu, Wang, Gao. *CUMC Bioworkflows*. https://cumc.github.io/bioworkflows/ldpred/



<u>Fig 1. Mash estimates data-drive covariance patterns of true genetic effects as the multivariate prior to improve posterior estimates for downstream analyses</u>. Mash[4] estimates the covariance of the effects in an empirical bayes fashion, thus estimating patterns of sharing among conditions (here, lipid traits) from the strongest signals in the data, and estimating the relative abundance of such patterns from a random set of all data. This allows us to provide the posterior estimate of the effect and its associated local false sign rate, or posterior probability of incorrectly identifying the sign of the effect, for each SNP and use these posterior estimates to improve performance in polygenic prioritization, enrichment analyses, on polygenic risk scoring. **Mash=multivariate adaptive shrinkage; *SNP=Single Nucleotide Polymorphism; LFSR=Local False Sign Rate; PRS=Polygenic Risk Score; LD=Linkage Disequilibrium.***

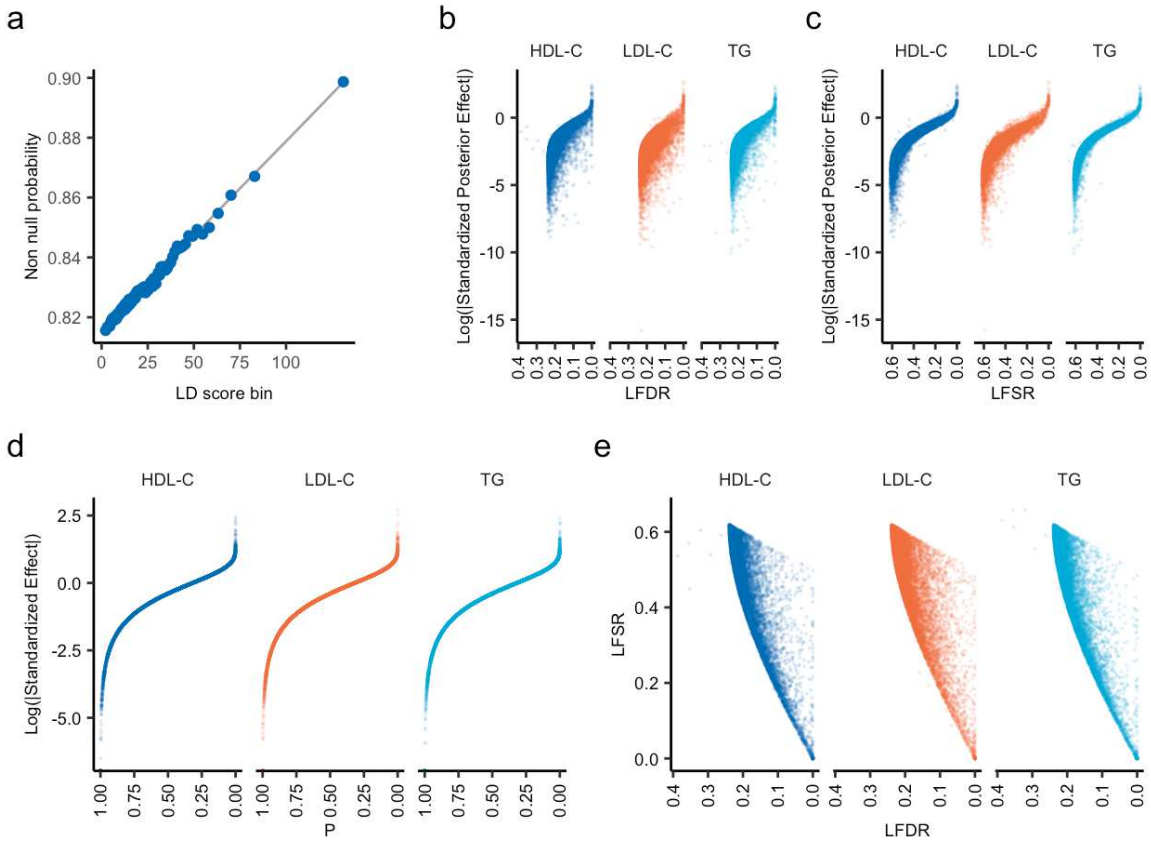


Figure 2. The Utility of Controlling for False Discovery. A) We demonstrate the relationship between effect size and P value and in **B)** A multivariate approach allows that for a given probability of being null (*lfdr*)[28] or for a given local false sign rate (lfsr) **C)** there can be a variety of effect sizes depending on the relative strength of evidence in alternative subgroups. Furthermore, a given non-null rate can lead to greater resolutions in the range of possible local false sign rates **(D)** as reflected in a variety of Local false sign rates for a given non-null rate. **E**) Finally, as the number of SNPs tagged by a given variant increases, the non-null probability increases. **HDL-C=High Density Lipoprotein Cholesterol, LFSR= Local False Sign Rate, LFDR=Local False Discovery Rate, LDL-C= Low-Density Lipoprotein Cholesterol, LDSC=Linkage Disequilibrium Score, TG=Triglycerides.**

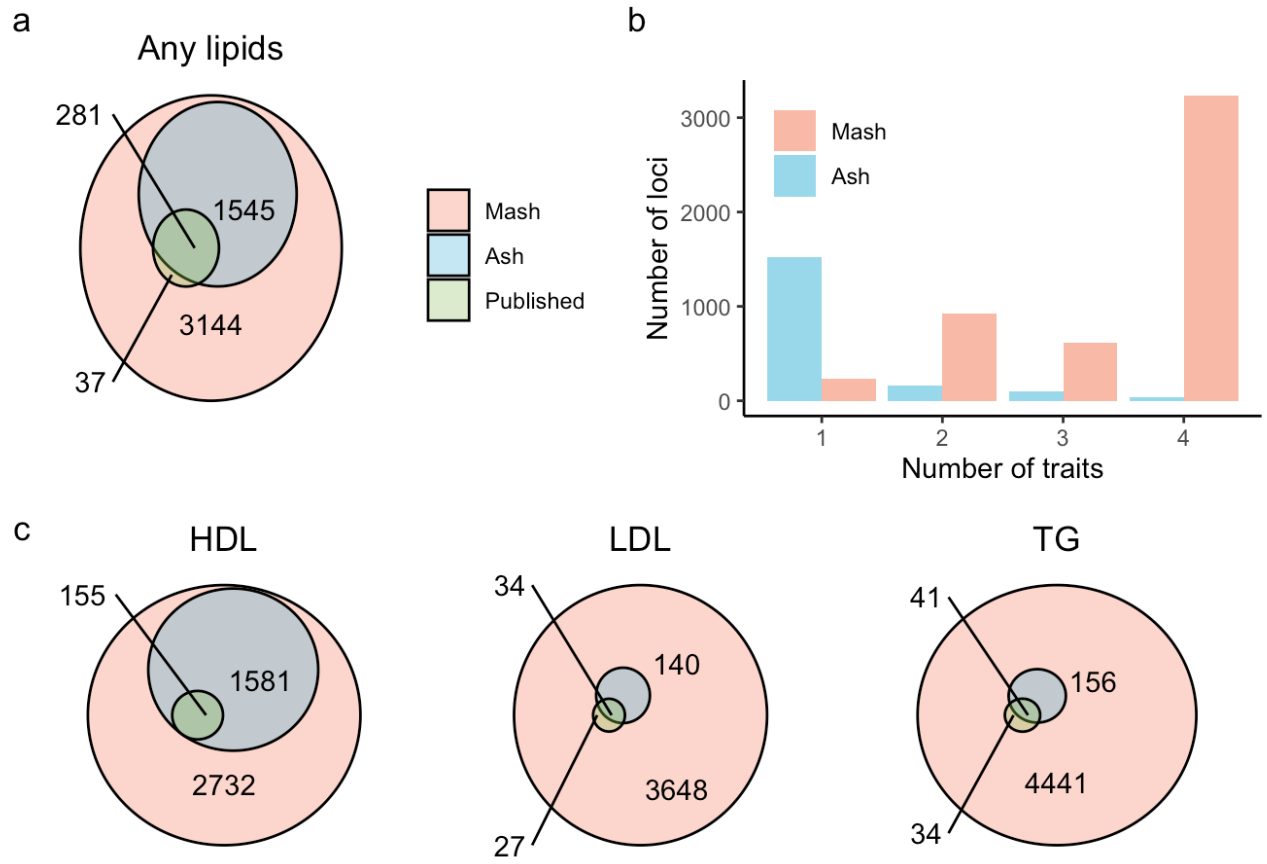


Fig 3. Control of False discovery improves power to detect over control of Family-Wise Error Rate A). A joint approach results in most significant associations being shared in at least 2 subgroups, whereas a univariate approach does not capture the tendency to share effects across conditions. B). Univariate measure of local false sign rate control using ash[5] replicates essentially all existing associations and dramatically increases power to detect. Multivariate adaptive shrinkage adds an additional layer of local false sign rate control by incorporating information across phenotypes. We plot the number of LD blocks containing at least one significant variant across traits in C). HDL-C, LDL-C, and TG. ***Ash= univariate adaptive shrinkage, mash=multivariate adaptive shrinkage, HDL-C=High-Density Lipoprotein Cholesterol, mash=multivariate adaptive shrinkage, LDL-C=Low-Density Lipoprotein Cholesterol, TG=Triglycerides.***

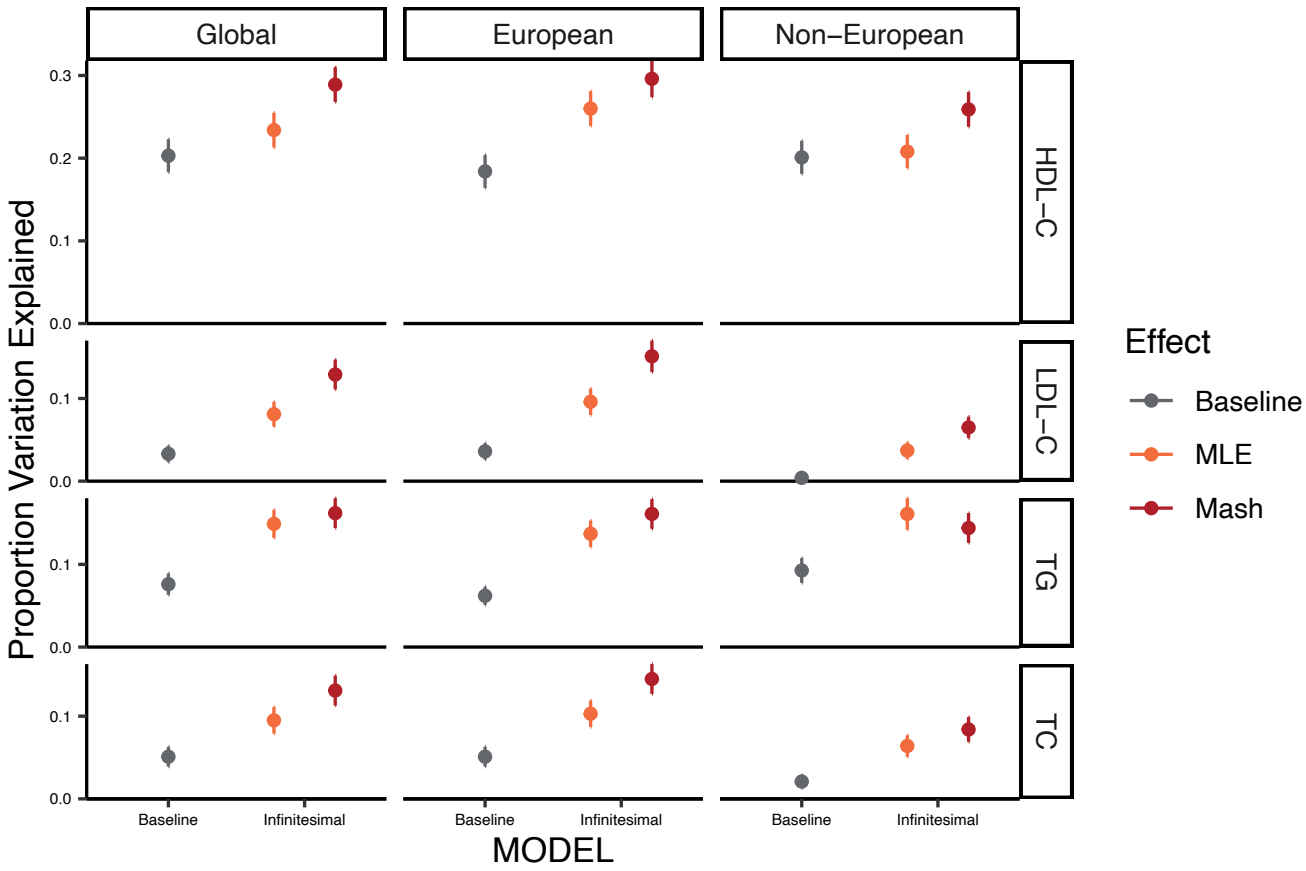


Figure 4. Mash Improves Polygenic Prediction. We consider the improvement in proportion of variation explained by LDpred2[16] on prediction of lipid traits across ethnicities using mash derived posteriors and univariate GWAS estimates as weight inputs over a model including only baseline covariates. We compare the performance of the infinitesimal model using univariate shrinkage (ash) or multivariate (mash) output for all (Global), European ancestry, or non-European ancestry (See Methods for details; **Table S4** for results in tabular form) to a baseline model using only baseline covariates of age and sex in each model. ***GWAS = genome-wide association study. ash = univariate adaptive shrinkage, mash=multivariate adaptive shrinkage, HDL-C=High-Density Lipoprotein Cholesterol, mash=multivariate adaptive shrinkage, LDL-C=Low-Density Lipoprotein Cholesterol, TG=Triglycerides.***

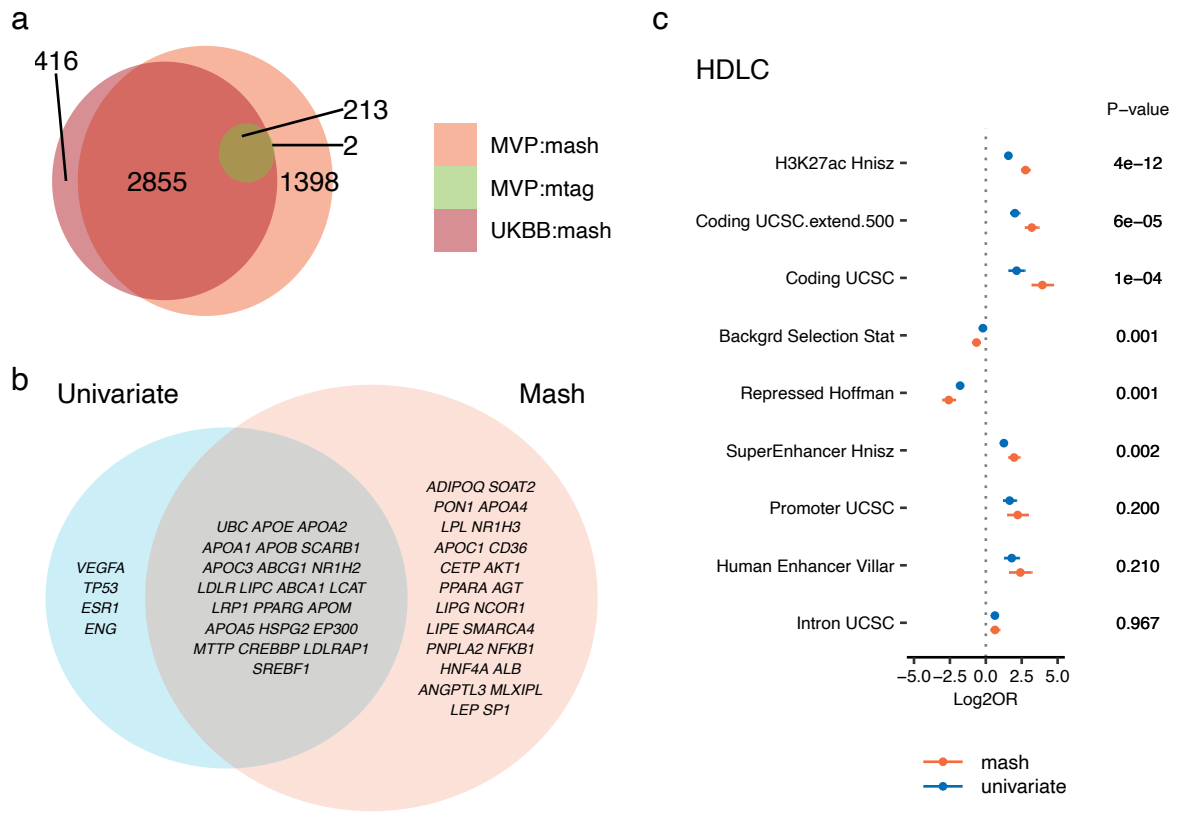


Figure 5: Bayesian multivariate method improves discovery by 15-fold and improves polygenic prioritization consistency of known lipid targets while enhancing known annotation estimation ***A).*** MVP and UKB were fit using mash separately and 500-kb LD blocks were compared containing a significant variant at an *lfsr* of 0.05. We consider all blocks previously identified in Klarin et al.[8] Mash consistently prioritizes 47 genes among LDL-C, HDL-C, and TG, while univariate methods prioritize 23. Of these, 24 are found consistently by mash but not by univariate (MLE) approach, while only 4 are found consistently by univariate approach but not mash. We use polygenic prioritization framework detailed in [20]. ***C)*** Using TORUS[22] we consider enrichment in 27 of the 52 classes examined by Finucane *et al.*[12] and see that mash versus univariate estimates tend to increase features known to be enriched in GWAS hits and decrease those known to be depleted (p-values for difference in the plot). We display for HDL-Cholesterol (***LDL-C, TG,*** and ***TC*** in **Extended Data Figure 5-7; Table S5**). ***GWAS = genome-wide association study; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; mash = multivariate adaptive shrinkage; TG = triglycerides, TC = total cholesterol; MVP:mash = Million Veterans Program data analyzed using mash; MVP:uni =Million Veterans Program Data analyzed using traditional GWAS univariate analysis; UKB:mash = UK Biobank data analyzed using mash; UKBB:uni =UK Biobank data analyzed using traditional GWAS univariate analysis; MVP:mtag: Million Veterans Program Data analyzed using MTAG[18].***

Works Cited

1. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

2. Efron, B. Size, power and false discovery rates. *Ann. Stat.* **35**, 1351–1377 (2007).

3. Stephens, M. False discovery rates: a new deal. *Biostatistics* kxw041 (2016) doi:10.1093/biostatistics/kxw041.

4. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).

5. Klarin, D. *et al.* Genetics of Blood Lipids Among ~300,000 Multi-Ethnic Participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).

6. Weeks, E. M. *et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases*. http://medrxiv.org/lookup/doi/10.1101/2020.09.08.20190561 (2020) doi:10.1101/2020.09.08.20190561.

7. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* **10**, (2016).

8. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

9. 23andMe Research Team *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).

Methods and Materials for: Bayesian Multivariate Genetic Analysis Boosts Translational Insights

Other Methods and Materials for this manuscript include the following:

1. Online methods for rerunning mash with available summary statistics are available at https://broadinstitute.github.io/natarajanlab_wiki/MVP-mfit.html

2. Software for the mashR package: https://github.com/stephenslab/mashr

3. Software for the Torus package: https://github.com/xqwen/torus

4. Online methods for producing Torus metaplots: https://broadinstitute.github.io/natarajanlab_wiki/metaplot_strat_torus.html

5. Online methods for reproducing venn diagrams: https://broadinstitute.github.io/natarajanlab_wiki/venn_diagrams.html

6. Software for the PoPs package: https://github.com/FinucaneLab/pops

7. Software for LDpred2: https://privefl.github.io/bigsnpr/

8. Worklow for running LDpred2 with MVP summary statistics using SoS pipeline for simple univariate case (template for adding analogous mash summary stats for additional traits) https://broadinstitute.github.io/natarajanlab_wiki/hdl_univariate.html

Materials and Methods

Study Participants

Association testing was performed in up to 297,626 white (European ancestry), black (African ancestry), and Hispanic Million Veterans Program (MVP) participants with blood lipids stratified by ethnicity followed by a meta-analysis of results across all three groups as previously described[8]. Samples were imputed to the 1000 Genomes project p3v5 reference panel (b37), and ancestry specific Hardy-Weinberg equilibrium P <1x10-20, posterior call probability < 0.9, imputation quality/INFO <0.3, minor allele frequency (MAF) < 0.0003, call rate < 97.5% for common variants (MAF > 1%), and call rate < 99% for rare variants (MAF < 1%) were used for QC. Variants were also excluded if they deviated > 10% from their expected allele frequency based on reference data from the 1000 Genomes Project. Trans-ethnic meta-analysis of white (European ancestry), black (African ancestry), and Hispanic MVP participants for 291,933 and 297,626 people was produced for Inverse normal transformed HDL cholesterol, Inverse normal transformed LDL cholesterol, Inverse normal transformed triglyceride levels and inverse normal transformed Total Cholesterol.[29]

UK Biobank samples[30] were genotyped on either the UK BiLEVE or UK Biobank Axiom arrays and imputed into the Haplotype Reference Consortium panel and the UK10K+1000 Genomes panel. Variant positions were keyed to the GRCh37 human genome reference. Genotyped variants with genotyping call rate < 0.95 and imputed variants with INFO score < 0.3 or minor allele frequency <= 0.005 in the analyzed samples were excluded. After variant-level quality control, 11,622,901 imputed variants remained for analysis Lipid Levels were collected on the Beckman Coulter AU5800 Platform and were adjusted for Cholesterol medication[31]. Participants without imputed genetic data, or with a genotyping call rate < 0.98, mismatch between self-reported sex and sex chromosome count, sex chromosome aneuploidy, excessive third-degree relatives, or outliers for heterozygosity were excluded from genetic analysis.[30]

Model and Fitting of mash model

Using the procedure outlined in Urbut et al[4], we summarize here for the purpose of completeness: Let b_{jr} (j = 1,…, J; r = 1,…, R) denote the true value of effect j in condition r. Further, let $\hat{b}_{jr}$ denote the "observed" estimate of this effect, and let $\hat{s}_{jr}$ be the standard error of this estimate, so $z_{jr} \coloneqq \hat{b}_{jr}/\hat{s}_{jr}$ is the standard Z statistic used to test whether b_{jr} is zero. Let B, $\hat{B}$, S and Z denote the corresponding J × R matrices, and let $\boldsymbol{b_j}$ (respectively, $\widehat{\boldsymbol{b_j}}, \boldsymbol{z_j}$) denote the jth row of B (respectively, $\hat{B}$, Z).

We assume $\widehat{\boldsymbol{b}}_{\boldsymbol{j}}$ is normally distributed about $\boldsymbol{b}_{\boldsymbol{j}}$, with variance-covariance matrix $\boldsymbol{V}_{\boldsymbol{j}}$ (defined below), and that the $\boldsymbol{b}_{\boldsymbol{j}}$ follows eq. 1:

$$p(\boldsymbol{b}; \boldsymbol{\pi}, \boldsymbol{U}) = \sum_{k=1}^{K} \sum_{l=1}^{L} \pi_{k,l} N_R(\boldsymbol{b}; \mathbf{0}, \omega_l U_k), \tag{1}$$

$$p\left(\widehat{\boldsymbol{b}}_j \mid \boldsymbol{b}_j, V_j\right) = N_R\left(\widehat{\boldsymbol{b}}_j; \boldsymbol{b}_j, V_j\right) \tag{2}$$

$$\boldsymbol{p}\left(\boldsymbol{b}_{\boldsymbol{j}} \mid \boldsymbol{\pi}, \boldsymbol{U}\right) = \sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{K}} \sum_{\boldsymbol{l}=\mathbf{1}}^{\boldsymbol{L}} \boldsymbol{\pi}_{\boldsymbol{k},\boldsymbol{l}} \boldsymbol{N}_{\boldsymbol{R}}(\boldsymbol{b}_{\boldsymbol{j}}; \mathbf{0}, \boldsymbol{\omega}_{\boldsymbol{l}} \boldsymbol{U}_{\boldsymbol{k}}). \tag{3}$$

Combining these two implies that the marginal distribution of $\widehat{\boldsymbol{b}}_{\boldsymbol{j}}$, integrating out $\boldsymbol{b}_{\boldsymbol{j}}$, is

$$p\left(\widehat{\boldsymbol{b}}_j \mid \boldsymbol{\pi}, \boldsymbol{U}, V_j\right) = \sum_{k=1}^{K} \sum_{l=1}^{L} \pi_{k,l} N_R(\widehat{\boldsymbol{b}}_j; \mathbf{0}, \omega_l U_k + V_j). \tag{4}$$

Each covariance matrix $\boldsymbol{V}_{\boldsymbol{j}}$ is specified as $\boldsymbol{V}_{\boldsymbol{j}} = \boldsymbol{S}_{\boldsymbol{j}} \boldsymbol{C} \boldsymbol{S}_{\boldsymbol{j}}$, where C is a correlation matrix that accounts for correlations among the measurements in the R conditions, and $\boldsymbol{S}_{\boldsymbol{j}}$ is the R × R

diagonal matrix with diagonal elements $\hat{s}_{j1}, \ldots, \hat{s}_{jR}$. We estimate $\widehat{C}$ as the estimated covariance matrix of the weakest ($|\widehat{Z}| < 2$) effects across condition[4].

The two steps of mash are:

i. Estimate U and π. This involves two sub-steps:

   a. Define a set of (normalized) data-driven and canonical covariance matrices, $\widehat{U}$.

   b. Given $\widehat{U}$, estimate π by maximum likelihood. (A key idea is that if some matrices generated in Step i–a do not help capture patterns in the data, then they will receive little weight.) Let $\widehat{\pi}$ denote this estimate.

ii. Compute, for each j, the posterior distribution $p(b_j \mid \hat{b}_j, \widehat{\pi}, \widehat{U}, V_j)$.

These steps are detailed in ([4]) and described briefly below.

Generate data-driven covariance matrices U_k. We first identify rows j of matrix $\widehat{B}$ that likely have an effect in at least one condition. In the MVP data we chose rows corresponding to the "top" SNP for each LD block, which we define to be the SNP with the highest value of $Z_j^{\max} := \max_r \hat{b}_{jr}/\hat{s}_{jr}$. (We use the maximum, rather than the sum because we would like to include effects that are strong in a single lipid trait, not just effects that are shared among all lipids.)

Next, we fit a mixture of MVN distributions to these strongest effects using methods from Bovy et al[32]. Results in Bovy et al describe an EM algorithm for fitting a model similar to eqs. 2 & 3, with the crucial difference that there are no scaling parameters on the covariances;

$$p(b_j \mid \pi, U) = \sum_{k=1}^{K} \pi_k N_R(b_j; 0, U_k). \quad (5)$$

We used a list of 16 matrices (i.e., K=16) in this application that incorporated 7 data driven matrices as well as the identity and 4 canonical 'trait specific' matrices as well as matrix for equal effects and three matrices which varying levels of fixed heterogeneity[4,6,33]. We then expanded by a grid of 21 scaling factors (i.e, L = 21) ranging from $\omega^2 = 0.07$ to $\omega^2 = 72.43$ as specified by the range of observed `noisy` effects in a training set of 40,000 SNPS.

To compute the data-driven covariance matrices U, the strongest SNP per each of the 1703 conservatively defined LD blocks specified in[23] were chosen as the SNP with the strongest absolute observed Z statistic across traits from the matrix of Z statistics. We then repeated this model fitting procedure with the UKBB data, choosing a new set of `maxes' and computing weights on a training set from UKBB.

We computed posterior means and local false sign rates for 28,686,877 SNPs in the MVP databases and 13,788,619 million SNPs in the UKBB which overlapped at 11,1577,790 loci.

Posterior calculations. According to the standard Bayesian result for a multivariate normal inference and detailed in[5], if $\boldsymbol{b} \sim \boldsymbol{N_R}(\boldsymbol{0}, \boldsymbol{U})$ and $\widehat{\boldsymbol{b}} \mid \boldsymbol{b} \sim \boldsymbol{N_R}(\boldsymbol{b}, \boldsymbol{V})$, then

$$\widehat{\boldsymbol{b}}|\boldsymbol{b} \sim \boldsymbol{N_R}\left(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{U}}\right),$$

where

$$\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{U}}(\boldsymbol{U}, \boldsymbol{V}) \coloneqq \left(\boldsymbol{U^{-1}} + \boldsymbol{V^{-1}}\right)^{-1} \quad (6)$$

$$\widetilde{\boldsymbol{\mu}} = \widetilde{\boldsymbol{\mu}}\left(\boldsymbol{U}, \boldsymbol{V}, \widehat{\boldsymbol{b}}\right) \coloneqq \widetilde{\boldsymbol{U}}(\boldsymbol{U}, \boldsymbol{V})\boldsymbol{V^{-1}}\widehat{\boldsymbol{b}} \quad (7)$$

(6)

This result is easily extended to the case where the prior on b is a mixture of MVNs (eq. 1). In this case, the posterior distribution is a mixture of MVNs, where here again as in[4] we have collapsed all shape scale combinations into P components.

$$p(b_j \mid \hat{b}_j, V_j, \pi) = \sum_{p=1}^{P} \tilde{\pi}_{jp} N_R(b_j;\, \tilde{\mu}_{jp}, \tilde{U}_{jp}),$$

where $\tilde{U}_{jp} := \tilde{U}(\Sigma_p, V_j)$ (eq. 6) and $\tilde{\mu}_{jp} := \tilde{\mu}(\Sigma_p, V_j, \hat{b}_j)$ (eq. 7), and

$$\tilde{\pi}_{jp} := \frac{\pi_p N_R(\hat{b}_j; 0, \Sigma_p + V_j)}{\sum_{p'=1}^{P} \pi_{p'} N_R(\hat{b}_j; 0, \Sigma_{p'} + V_j)}.$$

From these results, it is straightforward to compute the posterior mean

$$\tilde{b}_j := E(b_j \mid \hat{b}_j, V_j, \pi) = \sum_{p=1}^{P} \tilde{\pi}_{jp} \tilde{\mu}_{jp} \quad (7)$$

and posterior variance

$$\mathbf{Var}\left(\boldsymbol{b_{jr}} \mid \widehat{\boldsymbol{b}}_{\boldsymbol{j}}, \boldsymbol{V_j}, \boldsymbol{\pi}\right) = \sum_{\boldsymbol{p=1}}^{\boldsymbol{P}} \widetilde{\boldsymbol{\pi}}_{\boldsymbol{jp}}\left(\widetilde{\boldsymbol{U}}_{\boldsymbol{jp,rr}} + \widetilde{\boldsymbol{\mu}}^{\boldsymbol{2}}_{\boldsymbol{jp,r}}\right) - \left(\sum\nolimits_{\boldsymbol{p'=1}}^{\boldsymbol{P}} \widetilde{\boldsymbol{\pi}}_{\boldsymbol{jp'}} \widetilde{\boldsymbol{\mu}}_{\boldsymbol{jp',r}}\right)^{\boldsymbol{2}},$$

as well as the local false sign rate. We use the posterior mean in (8) as the input to our downstream analyses.

Local false sign rate. To measure "significance" of an effect $\boldsymbol{b_{jr}}$, we use the local false sign rate[15],

$$\boldsymbol{lfsr_{jr}} := \mathbf{min}\left\{\mathbf{Pr}\left(\boldsymbol{b_{jr}} \geq \mathbf{0} \mid \boldsymbol{D}\right), \mathbf{Pr}\left(\boldsymbol{b_{jr}} \leq \mathbf{0} \mid \boldsymbol{D}\right)\right\}, \qquad (8)$$

where D denotes the available data. $\boldsymbol{lfsr_{jr}}$ is the probability that we would incorrectly predict the sign of the effect if we were to call an effect positive or negative based on the sign of the posterior mean. A small lfsr indicates high confidence in the sign of an effect. The lfsr is more conservative than its analogue, the local false discovery rate (lfdr)[28] because requiring high confidence in the sign is more stringent than requiring high confidence that the effect be non-zero. More importantly, the lfsr is more robust to modelling assumptions than the lfdr[5], a particularly important issue in multivariate analyses where modelling assumptions inevitably play a larger role[4].

Gene Enrichment

TORUS[22] uses a hierarchical Bayes model to estimate the enrichment of a variety of genomic annotations in QTL and in turn assigns a posterior probability of inclusion to the 'causal' nature of the corresponding loci. We used the assigned posterior mean expected Z scores from mash and corresponding univariate Z statistics as inputs to the TORUS model and observed the estimation of log OR of an annotation parameter being enriched (depleted) in QTL. This package considers the strongest loci per block, and we used the 1703 blocks as conservatively defined by[23]. Furthermore, we used the set of annotation parameters available through the Price

lab and described in[34]. We provide a link to the commands to run appropriate pipeline[35] and corresponding annotation file.

Gene Prioritization

In order to rank genes based on their strength of effects as demonstrated by associated SNPs and consequent feature enrichment, The Polygenic Gene Prioritization (PoPs) algorithm works in 3 steps[20].

1. Use magma to assign gene association statistics based on significance of associated SNPs as detailed in[20]
2. Select marginally associated features by performing enrichment analysis for each gene feature separately.
3. Estimate Polygenic Priority Scores (PoPs) by fitting a joint model for the enrichment of all selected features using 'leave one out' regression.

Here, we used the local false sign rate (lfsr) and p-value for all SNPs considered by the MVP project for our mash and univariate analyses, respectively. We used the 1000 Genomes human reference to specify LD among variants per the PoPS protocol, the MAGMA list of genetic annotations to map genes to respective loci, and control and PoPS features from the available packages as detailed within the PoPS package and accompanying software[36]. We then ranked genes in order of prioritization score and compared mash estimates to univariate assessments.

Polygenic Score Prediction

We used the LDpred2 ([16]) infinitesimal sites model to control for this genetic correlation. In this model, all markers are causal (p=1), and effects are drawn from a Gaussian distribution, i.e.,

$$B_{jr} \sim N(0, \frac{{h_r}^2}{M}) \quad (9)$$

where h^2 represents the heritability of the trait and M the number of markers. The posterior mean can be derived analytically

$$E\left(B_{jr} \middle| \hat{B}\right\}_{jr}, D \approx \left(\frac{M}{Nh^2} I + D\right)^{-1} \hat{B}_{jr} \quad (10)$$

ordinary least square estimate from a univariate GWAS analysis. Here, D denotes the LD matrix between the markers obtained from an outside reference panel, in our case the 1000 Genomes EUR data set.

In a typical analysis, one uses $\hat{b}_{jr}$ that are estimated from a univariate GWAS analysis in each trait separately. However, we replace this with the posterior means on a per trait basis from

mash. The β_{jr} are each the marginal output of the posterior mean arising from the mixture normal cited in[4] [eq. 7].

Because we use the exchangeable Z- statistic model where we assume the standard errors $\widehat{s^2}$ are known, these estimates are then rescaled by their standard error to compute the input summary statistic weights for rescaling in LDpred2[16].

General workflow as follows:

1) We perform QC on reference panel (1000 genomes[10])

2) We intersect SNPs common to the reference panel (1000 genomes[10]), discovery data set (MVP) and scoring data set (UKBB). This left us with approximately 400K SNPs.

3) Harmonize alleles for shared direction

4) We calculate the LD matrix and fit the LDSC per LDpred2[16].

5) We use this LD matrix and to compute the posterior weights from initial summary statistics (either arising from GWAS summary statistics or from the posterior means of mash output).

6) We compute the PRS for approximately 4,700 randomly chosen scoring individuals in UKBB.

7) We associated these scores with the phenotype of interest in a linear model that includes age and sex as additional (baseline) covariates.

8) We then divide into EUR and non-EUR individuals to assess population specific performance. A sample vignette running it on HDL univariate summary statistics is available at[37].

Works Cited

1. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.* **109**, 1193–1198 (2012).

2. Manolio, T. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

3. Zhu, X. & Stephens, M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *bioRxiv* 042457 (2016) doi:10.1101/042457.

4. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

5. Stephens, M. False discovery rates: a new deal. *Biostatistics* kxw041 (2016) doi:10.1093/biostatistics/kxw041.

6. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLoS Genet* **9**, e1003486 (2013).

7. Nowbar, A. N., Gitto, M., Howard, J. P., Francis, D. P. & Al-Lamee, R. Mortality From Ischemic Heart Disease: Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration. *Circ. Cardiovasc. Qual. Outcomes* **12**, (2019).

8. Klarin, D. *et al.* Genetics of Blood Lipids Among ~300,000 Multi-Ethnic Participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).

9. Peloso, G. M. & Natarajan, P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr. Opin. Genet. Dev.* **50**, 1–6 (2018).

10. Clarke, L. *et al.* The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* **45**, D854–D859 (2017).

11. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).

12. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

13. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

14. EPIC-CVD Consortium *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).

15. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).

16. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).

17. Ye, Y. *et al.* Interactions Between Enhanced Polygenic Risk Scores and Lifestyle for Cardiovascular Disease, Diabetes, and Lipid Levels. *Circ. Genomic Precis. Med.* **14**, (2021).

18. 23andMe Research Team *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).

19. Brzyski, D. *et al.* Controlling the Rate of GWAS False Discoveries. *Genetics* **205**, 61–75 (2017).

20. Weeks, E. M. *et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases*. http://medrxiv.org/lookup/doi/10.1101/2020.09.08.20190561 (2020) doi:10.1101/2020.09.08.20190561.

21. NHLBI TOPMed Lipids Working Group *et al.* Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat. Commun.* **9**, 3391 (2018).

22. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* **10**, (2016).

23. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* btv546 (2015) doi:10.1093/bioinformatics/btv546.

24. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).

25. Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Stat.* **31**, 2013–2035 (2003).

26. Sesia, M., Bates, S., Candès, E., Marchini, J. & Sabatti, C. False discovery rate control in genome-wide association studies with population structure. *Proc. Natl. Acad. Sci.* **118**, e2105841118 (2021).

27. Schizophrenia Working Group of the Psychiatric Genomics Consortium *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).

28. Efron, B. Size, power and false discovery rates. *Ann. Stat.* **35**, 1351–1377 (2007).

29. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).

30. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

31. Peloso, G. M. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).

32. Bovy, J., Hogg, D. W. & Roweis, S. T. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Ann. Appl. Stat.* **5**, 1657–1677 (2011).

33. Wen, X. & Stephens, M. Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene–environment interactions. *Ann. Appl. Stat.* **8**, 176–203 (2014).

34. ReproGen Consortium *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

35. Qiao, Min & Wang, Gao. *TORUS workflow*.

36. Weeks, E. M. *PoPs software*.

37. Zhang, Mengyu & Wang, Gao. *CUMC bioworkflows*.

38. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

39. Stephens, M. False Discovery Rates: A New Deal. *bioRxiv* 038216 (2016) doi:10.1101/038216.

Extended Data Figures

This PDF file includes:

Extended Data Figures 1-7

Other Supplementary Materials for this manuscript include the following:

1. Supplementary Tables S1 to S6 (included in an attached .xlsx file)

2. Online methods for rerunning mash with available summary statistics are available at https://broadinstitute.github.io/natarajanlab_wiki/MVP-mfit.html

3. *Software for the mashR package: https://github.com/stephenslab/mashr*

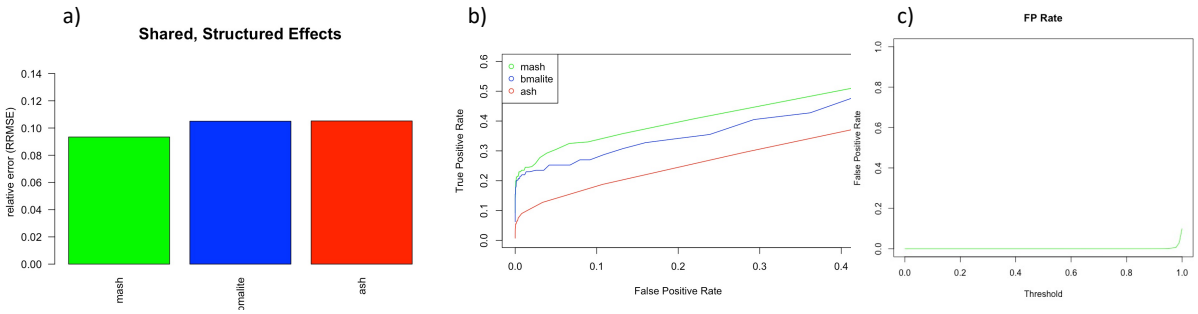
4. *Software for the Torus package:* https://github.com/xqwen/torus

5. Online methods for producing Torus metaplots: https://broadinstitute.github.io/natarajanlab_wiki/metaplot_strat_torus.html
6. Online methods for reproducing venn diagrams: *https://broadinstitute.github.io/natarajanlab_wiki/venn_diagrams.html*

7. *Software for the PoPs package:* https://github.com/FinucaneLab/popste

8. *Software for LDpred2:* https://privefl.github.io/bigsnpr/

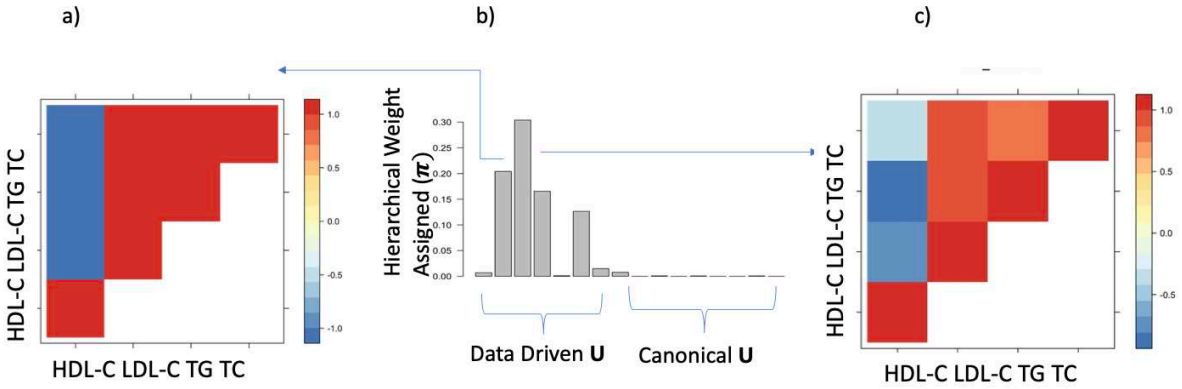
9. *Worklow for running LDpred2 with MVP summary statistics using SoS pipeline for simple univariate case (template for adding analogous mash summary stats for additional traits)* https://broadinstitute.github.io/natarajanlab_wiki/hdl_univariate.html



Extended Data Figure 1: mash estimates are superior in terms of precision and in power.

(a) mash[38] estimates improve the relative root mean squared error, indicating improved precision, when comparing with canonical multivariate (bmalite)[6] and univariate (ash)[39] methods. Mash also offers improved power for a given level of specificity (b) when compared to multivariate canonical and univariate approach. Furthermore, when simulating the instance in which all associations are truly null, mash detects 0 false positive associations at an *lfsr* threshold of up to 0.54 (c), indicating the accuracy of the false sign rate for thresholding purposes. This scenario was based on the fit of the mash model (equation (*1*)) to the MVP data (see Methods). When simulating an instance in which 0 of the effects are real (c), 0 effects are identified at an *lfsr* of up to 0.54.

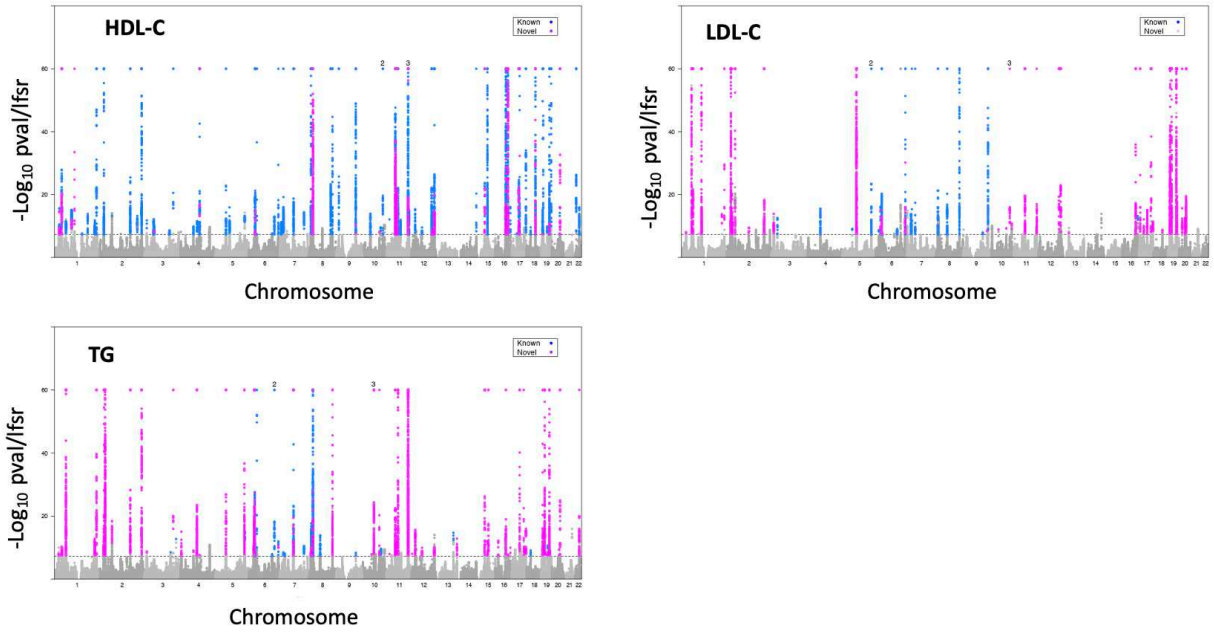
ash = adaptive shrinkage, lfsr = local false sign rate, mash = multivariate adaptive shrinkage, MVP = Million Veterans Program.



Extended Data Figure 2: hierarchical patterns of sharing learned by mash.

Mash allows us to consider the hierarchical patterns which receive the most `weight' in the mixture model and as expected the matrices which received the majority of the hierarchical weight showed effects that were shared in sign and magnitude among LDL-C, TC and TG that were strongly inversely correlated with strong effects in HDL-C. In (**a)** and (**c),** these are the two matrices to receive the predominant weight as estimated by the likelihood step[38], and the histogram demonstrating the hierarchical weight in **(b).**

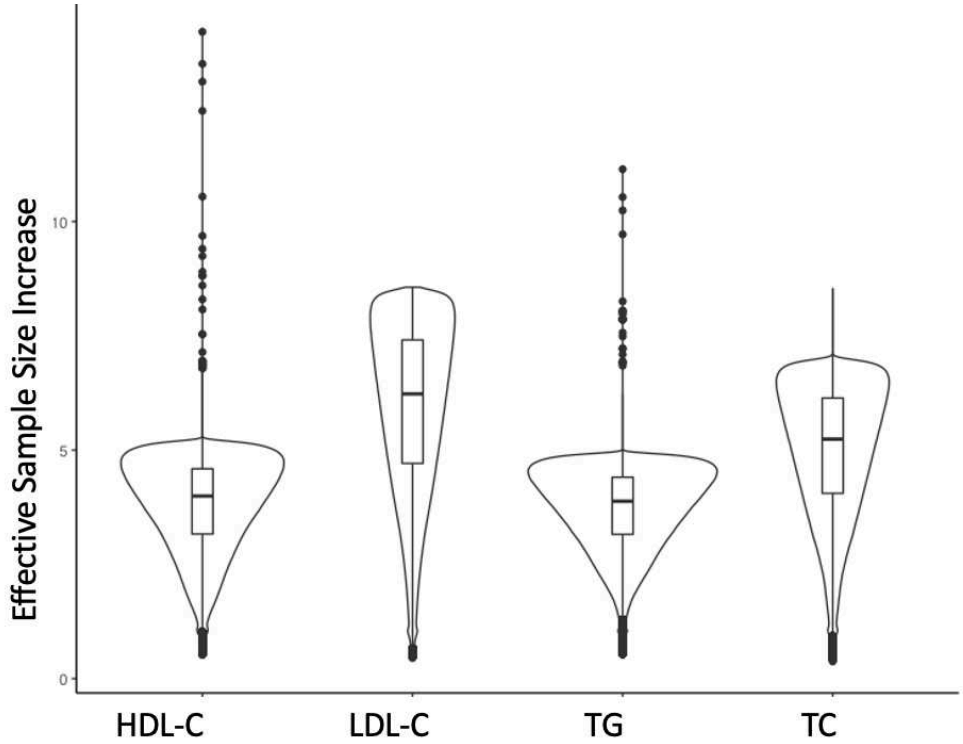
HDL-C = High-density lipoprotein cholesterol, LDL-C = Low-density lipoprotein cholesterol, mash = multivariate adaptive shrinkage, TC = Total cholesterol, TG = Triglycerides.



Extended Data Figure 3: Manhattan plot of all variants deemed significant at a genome-wide *lfsr* 0.05 in a mash analysis of MVP.

Multivariate shrinkage analysis was performed using mashR[38] on 28 million genetic variants in MVP and cross-referenced with publication on existing data set (8). Variants were colored in purple if they were deemed significant in mash analyses of both MVP and UKBB[30] datasets, and replaced with blue if they were within 500 kb of a published[8] result.15-fold increase in identified associations, all replicated in UKBB, as defined by those SNPs which contained at least one significant association across lipid traits.

HDL-C = HDL-cholesterol, LDL-C = LDL cholesterol, mash = multivariate adaptive shrinkage, MVP = Million Veterans Program, TG = Triglyceride, UKBB = UK Biobank.



476

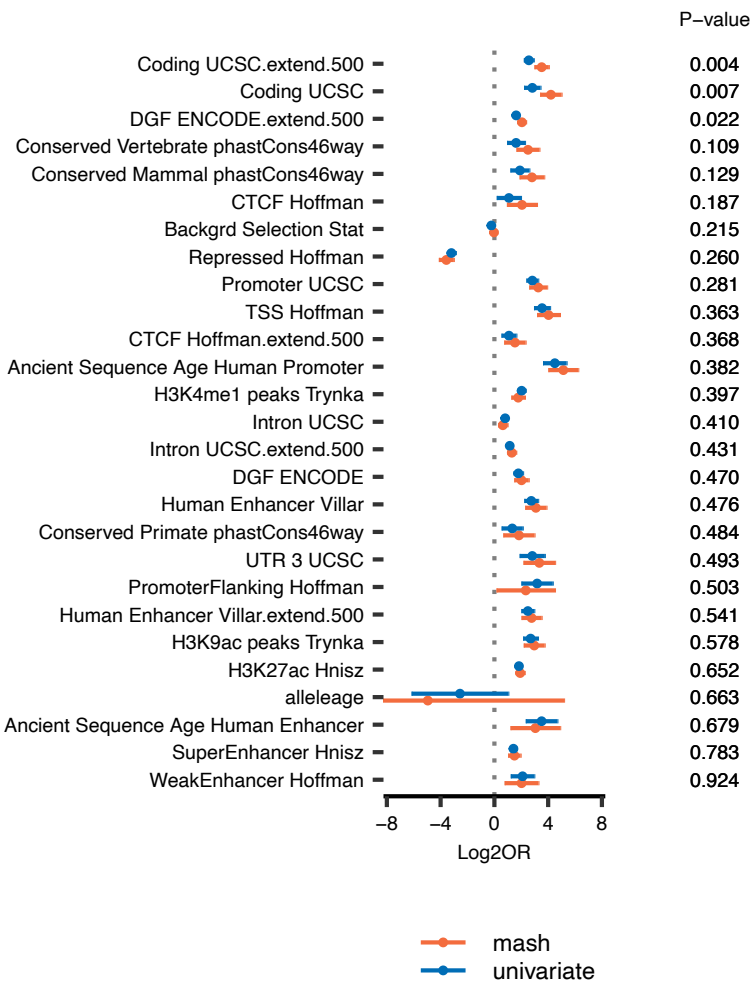
477 **Extended Data Figure 4: mash increases effective sample size by 4-6-fold.**

478 Here, we plot the median increase in effective sample size between multivariate and
479 univariate marginal estimates per trait for all interrogated SNPs. Values computed using
480 **Equation 1** in text. Point represents median and upper, lower whiskers 25th and 75th
481 percentile respectively.

482 ***HDL-C = high-density cholesterol, LDL-C = low-density cholesterol, mash =***
483 ***multivariate adaptive shrinkage, SNP = single nucleotide polymorphism, TC =***
484 ***total cholesterol, TG = triglycerides.***

485

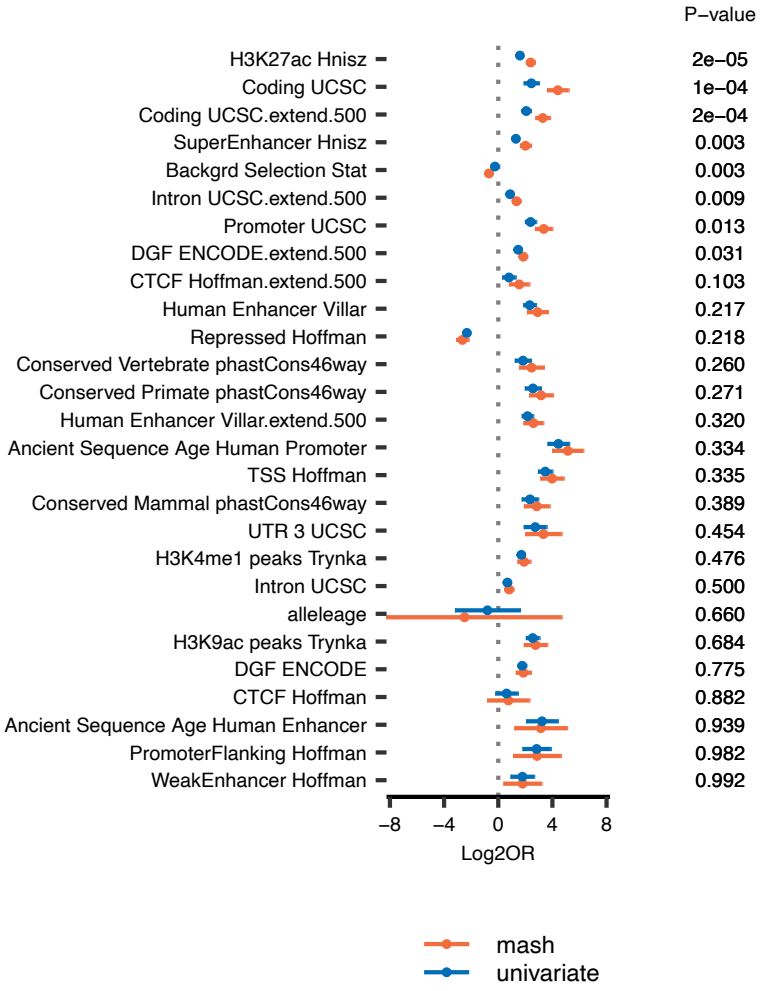
486



Extended Data Figure 5: Log 2 Fold Odds Ratio using TORUS: LDL-Cholesterol.

We fit the annotation enrichment tool TORUS[22] using summary statistics from univariate (GWAS) and multivariate (mash analyses) in LDL cholesterol. We plot the log 2-fold Odds Ratio (OR) for each annotation parameter and compare the difference between estimates using raw univariate and multivariate (mash) estimates with corresponding p-values in the figure.

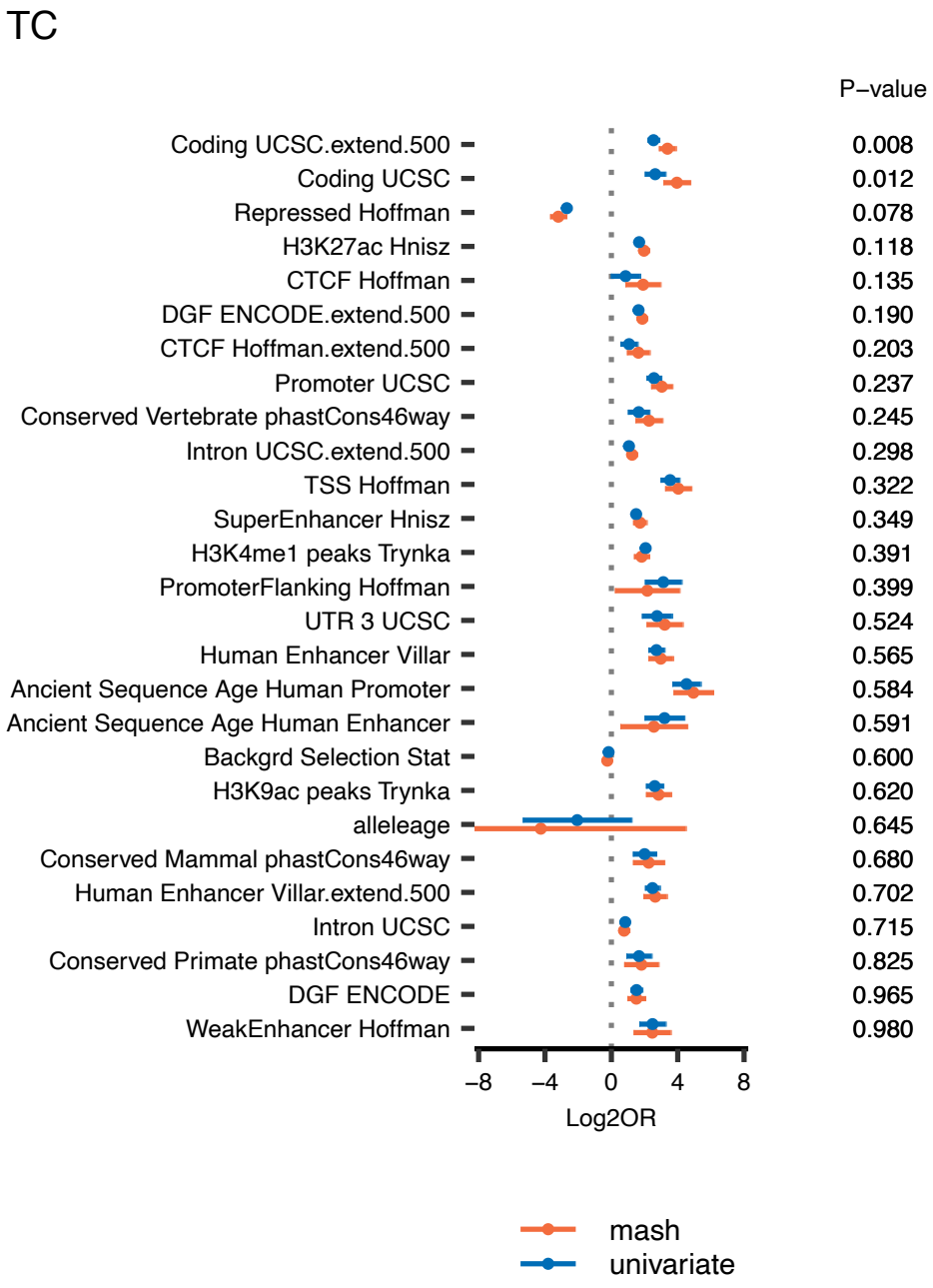
GWAS = genome-wide associate study, LDL-C = low-density lipoprotein cholesterol, mash = multivariate adaptive shrinkage, OR = odds ratio.



Extended Data Figure 6. Log 2 Fold Odds Ratio using TORUS: Triglycerides.

We fit the annotation enrichment tool TORUS[22] using summary statistics from univariate (GWAS) and multivariate (mash analyses) in triglycerides. We plot the log 2-fold Odds Ratio (OR) for each annotation parameter and compare the difference between estimates using raw univariate and multivariate (mash) estimates with corresponding p-values in the figure.

GWAS = genome-wide associate study, mash = multivariate adaptive shrinkage, OR = odds ratio, TG = triglycerides.



509

510 **Extended Data Figure 7: Log2 Fold Odds-Ratio using TORUS: Total Cholesterol**

511 We fit the annotation enrichment tool TORUS[22] using summary statistics from univariate
512 (GWAS) and multivariate (mash analyses) in total cholesterol. We plot the log 2-fold
513 Odds Ratio (OR) for each annotation parameter and compare the difference between

514 estimates using raw univariate and multivariate (mash) estimates with corresponding p-
515 values in the figure.

516 ***GWAS = genome-wide associate study, mash = multivariate adaptive shrinkage,***
517 ***OR = odds ratio, TC = total cholesterol.***
518

Works Cited

1. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

2. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLoS Genet* **9**, e1003486 (2013).

3. Stephens, M. False Discovery Rates: A New Deal. *bioRxiv* 038216 (2016) doi:10.1101/038216.

4. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

5. Klarin, D. *et al.* Genetics of Blood Lipids Among ~300,000 Multi-Ethnic Participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).

6. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* **10**, (2016).

538

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarytablesurbut.xlsx