Extending the Coding-variant Allelic Series Test to Summary Statistics

Zachary R McCaw,[1,*] Jianhui Gao,[1,2] Rounak Dey,[1] insitro Research Team,[1]
Emily Fox,[1] Colm O'Dushlaine,[1] Thomas W Soare[1,*]

October 31, 2024

Abstract

The coding-variant allelic series test (COAST) was introduced to identify genes harboring rare variant allelic series. Specifically, COAST targets genes where a dose-response relationship exists between mutational severity and phenotypic impact. Such genes are of therapeutic interest because the effect of pharmacological modulation can be predicted from the natural variation present in a population. The original COAST required access to individual-level data. However, such data are often unavailable due to privacy concerns or logistical challenges. Meanwhile, single-variant summary statistics of the type produced by genome-wide association studies are plentiful. Here we introduce COAST-SS, an extension of COAST that accepts summary statistics as input, namely the per-variant effect sizes and standard errors, along with estimates of the minor allele frequency and local linkage disequilibrium (LD). As a running example, we consider identifying allelic series for circulating lipid traits in the UK Biobank. Through extensive analyses of real and simulated data, we demonstrate that COAST-SS provides p-values effectively equivalent to those from the original COAST. Interestingly, we find that when LD is low, as is expected among rare variants, COAST-SS is robust to misspecification of the LD matrix, providing valid inference even when the LD matrix is set to the identity. We explore several strategies for annotating the pathogenicity of variants supplied to COAST-SS, finding that they often yield similar power for detecting associations. Lastly, we employ COAST-SS to screen for lipid trait allelic series in an expanded cohort of 350K subjects. COAST-SS has been incorporated into the publicly available `AllelicSeries R` package.

Keywords: Allelic series, Rare variant association testing, Variant Pathogenicity, Whole exome sequencing

[1] Insitro, South San Francisco, CA, USA.

[2] Department of Statistics, University of Toronto, Ontario, CA.

[*] Correspondence: ZRM (zmccaw@insitro.com), TWS (tsoare@insitro.com).

Introduction

An allelic series is a collection of mutations in a gene or pathway that leads to a spectrum of possible phenotypes [1]. Allelic series that modulate gene activity or functionality provide the opportunity to characterize a gene's dose-response behavior from the natural genetic variation in a population [2]. Such genes make appealing therapeutic targets because the impact of pharmacological modulation can be predicted from the impacts of observed genetic alterations [2]. The natural range of functionality observed in the population can also shed light on a candidate modulator's therapeutic window and safety profile [3]. Moreover, with the rise of gene editing technologies, understanding allelic series could guide genetic interventions aimed at achieving beneficial phenotypic changes [4]. Among mutations that affect gene functionality, rare variants can have larger effect sizes than common variants, as purifying selection prevents large-effect deleterious variants from becoming common [5]. Thus, rare variants are particularly informative for characterizing allelic series.

The coding-variant allelic series test (COAST) is a rare-variant association test tailored for the identification of genes harboring allelic series [6]. COAST operates on three classes of coding variants, as annotated by the Ensembl Variant Effect Predictor (VEP) [7]: benign missense variants (BMVs), deleterious missense variants (DMVs), and protein truncating variants (PTVs). By means of an adjustable sequence of weights, which specifies the pattern of effect size sought, COAST targets genes where the impact on the phenotype increases monotonically with mutation severity. COAST was the first rare-variant association test specifically designed not simply to determine whether a gene is associated with a phenotype, but whether that association constitutes a dose-response relationship. However, the original implementation of COAST requires individual-level data [6], which is restrictive insofar as access to individual-level data is often limited by privacy concerns or logistical challenges. In contrast, single-variant summary statistics from genome-wide association studies (GWAS) are often publicly available, even to relatively low minor allele frequencies. To facilitate their application, many methods originally developed in the context of individual-level data have been adapted to accept GWAS summary statistics as input, including heritability estimation, fine-mapping, Mendelian Randomization, and polygenic risk prediction [8, 9]. Doing so not only democratizes research, expanding the group of researchers who can analyze and derive insights from a data set, but also generally accelerates computation, benefiting even those who have access to the original data [10].

Here we extend COAST to enable estimation and inference starting from commonly available GWAS summary statistics. Specifically, COAST-SS performs allelic series testing starting from per-variant effect sizes and standard errors, along with estimates of the minor allele frequency (MAF) and local linkage disequi-

librium (LD). We first validate that COAST-SS, working from summary statistics, can recover essentially the same p-value as the original COAST, working from individual-level data, within the cohort of 150K UK Biobank subjects considered in the original work [6]. Our example phenotypes throughout are a collection of 4 commonly-studied lipid traits: total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides. We proceed to evaluate the operating characteristics of COAST-SS, including type I error and power, as applied to both real and simulated data. Of the inputs required by COAST-SS, the rare variant LD matrix is the least likely to be available. We examine the performance of COAST-SS when the provided LD matrix is either approximate or absent altogether. We find that COAST-SS is robust to the LD matrix, providing valid inference even when supplied with the identity matrix. Compared with other rare-variant association tests, a unique aspect of COAST(-SS) is the classification of variants into 3 distinct categories (BMVs, DMVs, PTVs), across which the effect on the phenotype is expected to increase. We explore the impact on power of alternative schemes for partitioning missense variants into pathogenicity categories based on PolyPhen2 [11] or AlphaMissense [12] scores. Among those explored thus far, we find that results are highly consistent across annotation schemes. Lastly, we expand our initial analysis of lipid trait allelic series in the UKB from 150K subjects with COAST to 350K subjects with COAST-SS, more than doubling the number of candidate allelic series.

Methods

Recap of COAST

Let Y denote a quantitative phenotype (e.g. circulating lipid trait), and X a set of covariates (e.g. age, sex, ancestry). Let $G = (G_1, \ldots, G_J)$, where $G_j \in \{0, 1, 2\}$, denote the genotype at the J rare coding variants within a given gene. We restrict to variants belonging to one of the following annotation categories: benign missense variants (BMVs), deleterious missense variants (DMVs), and protein truncating variants (PTVs). Let A_j denote whether the jth variant is a BMV ($A_j = 1$), DMV ($A_j = 2$), or PTV ($A_j = 3$), and denote by (N_1, N_2, N_3) the total counts of BMVs, DMVs, and PTVs in the gene, respectively. The **baseline allelic series model** is:

$$\mathbb{E}(Y|N, X) = \alpha + N_1\beta_1 + N_2\beta_2 + N_3\beta_3 + X\gamma. \tag{1}$$

Given a set of **allelic series weights** $w = (w_1, w_2, w_3)$, the **allelic series sum model** is:

$$\mathbb{E}(Y|N, X) = \alpha + (w_1N_1 + w_2N_2 + w_3N_3)\beta_S + X\gamma.$$

Note that the absolute scale of the allelic series weights is arbitrary because scaling all the weights up by a factor of ω while scaling β_S by a factor of ω^{-1} would result in the same model. Without loss of generality, in this work, we scale the weight such that $\sum_{a=1}^{3} w_a = 1$. Other normalizations such as fixing $w_1 = 1$ or $||w|| = 1$ are equally valid. Building on the sequence-kernel association test (SKAT) [13], the **allelic series SKAT model** is:

$$\mathbb{E}(Y|N, X) = \alpha + \sum_{j=1}^{J} G_j \beta_j + X\gamma, \qquad \beta_j \sim (0, w_j \tau^2), \qquad w_j = \frac{w_{A_j}}{\mathrm{MAF}_j(1 - \mathrm{MAF}_j)}, \tag{2}$$

where w_{A_j} is the allelic series weight corresponding to the jth variant's annotation, and MAF_j is the variant's minor allele frequency. The overall **omnibus test statistic** is calculated by aggregating the p-values of the $M = 3$ component tests using the Cauchy combination method [14]:

$$Q_{\mathrm{omni}} = \sum_{m=1}^{M} w_{\mathrm{omni},m} \tan\{\pi(0.5 - p_m)\},$$

where the omnibus weights ($w_{\mathrm{omni},m}$) are selected to give burden- and SKAT-type tests equal influence. The final COAST p-value is given by:

$$p_{\mathrm{omni}} = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{Q_{\mathrm{omni}}}{\sum_{m=1}^{M} w_{\mathrm{omni},m}}\right).$$

COAST from Summary Statistics

Baseline model from summary statistics

For exposition, the models in the previous section were subject-level. Hereafter, we adopt matrix notation, with subjects as rows. Suppose we have n independent subjects. The baseline allelic series model is expressible as:

$$Y = \tilde{G}\beta + X\gamma + \epsilon, \tag{3}$$

where $\tilde{G} = (N_1, N_2, N_3)$ is an $n \times 3$ matrix, $\beta = (\beta_1, \beta_2, \beta_3)$ is a 3×1 coefficient vector, and $\epsilon \sim N(0, \sigma^2 I)$ is an $n \times 1$ vector of residuals. We assume the intercept α from the previous section is a column of the covariate matrix X. Note that the normality assumption on the residuals can be relaxed to the requirements that $\mathbb{E}(\epsilon|\tilde{G}, X) = 0$ and $\mathbb{V}(\epsilon|\tilde{G}, X) = \sigma^2 I$. Define the $n \times n$ orthogonal projection matrix: $Q_X = I - X(X'X)^{-1}X'$. The 3×1 score equation for β is:

$$\mathcal{U} = \sigma^{-2}\tilde{G}' Q_X (Y - \tilde{G}\beta)$$

Under the $H_0 : \beta = 0$, the score reduces to $\mathcal{U}_0 = \sigma^{-2}\tilde{G}'Q_X Y$, whose 3×3 variance is $\mathbb{V}(\mathcal{U}_0) = \sigma^{-2}\tilde{G}'Q_X\tilde{G}$. The $H_0 : \beta = 0$ could be evaluated via:

$$T = \mathcal{U}_0'\{\mathbb{V}(\mathcal{U}_0)\}^{-1}\mathcal{U}_0 \overset{H_0}{\sim} \chi^2_3(0),$$

however this test requires access to the individual level data. We next develop an approximate test that can be evaluated using only summary statistics.

Suppose we have access to summary statistics from the single variant association models:

$$Y = G_j\alpha_j + X'\gamma + \epsilon_j, \tag{4}$$

specifically $\hat{\alpha}_j = (G_j'Q_X G_j)^{-1}G_j'Q_X Y$ and $\hat{\mathbb{V}}(\hat{\alpha}_j) = \hat{\sigma}_j^2(G_jQ_XG_j)^{-1}$. Note that $\hat{\alpha}_j$ and $\hat{\mathbb{V}}(\hat{\alpha}_j)$ are scalars. Since under H_0, $\hat{\sigma_j}^2 \xrightarrow{p} \sigma^2$ for $\forall j$, the sampling variance estimate $\hat{\mathbb{V}}(\hat{\alpha}_j) \approx \mathbb{V}(\hat{\alpha}_j) = \sigma^2(G_jQ_XG_j)^{-1}$. Let $\hat{\eta}$ denote the vector of per-variant score statistics:

$$\hat{\eta}_j = \frac{\hat{\alpha}_j}{\hat{\mathbb{V}}(\hat{\alpha}_j)} \approx \sigma^{-2}G_j'Q_XY.$$

Define the $J \times 3$ variant category indicator matrix D such that:

$$D_{jk} = \begin{cases} 1, & \text{if } A_j = k \\ 0, & \text{otherwise.} \end{cases}$$

Then, $\tilde{G}_l = GD_l$ where D_l is the l-th column of D, and thus $\tilde{G} = GD$. That is:

$$\begin{pmatrix} | & | & | \\ N_1 & N_2 & N_3 \\ | & | & | \end{pmatrix}_{n\times 3} = \begin{pmatrix} | & \cdots & | \\ G_1 & \dots & G_J \\ | & \cdots & | \end{pmatrix}_{n\times J} \begin{pmatrix} — & D_1 & — \\ \vdots & \vdots & \vdots \\ — & D_J & — \end{pmatrix}_{J\times 3}$$

Now, we can express the score statistic above and its variance by:

$$\begin{aligned} \mathcal{U}_0 &= \sigma^{-2}\tilde{G}'Q_XY = \sigma^{-2}D'G'Q_XY \approx D'\hat{\eta} \\ \mathbb{V}(\mathcal{U}_0) &= \sigma^{-2}\tilde{G}'Q_X\tilde{G} = D'\left(\sigma^{-2}G'Q_XG\right)D \\ &= D'\left(V^{-1/2}RV^{-1/2}\right)D, \end{aligned}$$

where the $V = \text{diag}(\{\mathbb{V}(\hat{\alpha}_j)\})$ is the $J \times J$ matrix of per-variant sampling variances, which can be approxi-

mated by $\hat{V} = \text{diag}(\{\hat{\mathbb{V}}(\hat{\alpha}_j)\})$, and the j,k-th element of the $J \times J$ correlation matrix R is:

$$R_{jk} = (G_j' Q_X G_j)^{-1/2} (G_j' Q_X G_k)(G_k' Q_X G_k)^{-1/2}.$$

If we assume that the covariates are at most mildly correlated with the genotypes (which is reasonable for rare variants in an ancestrally homogeneous population) and are included in the model mainly for improving precision, R_{jk} can be approximated by the LD between the variants j and k. Hence, we can use the estimated LD matrix $\hat{R}$ to replace R in the above equations to arrive at $\mathbb{V}(\mathcal{U}_0) = D'\hat{V}^{-1/2}\hat{R}\hat{V}^{-1/2}D$. Overall, the test statistic is:

$$T = \mathcal{U}_0' \{\mathbb{V}(\mathcal{U}_0)\}^{-1} \mathcal{U}_0 = \hat{\eta}' D \left(D'\hat{V}^{-1/2}\hat{R}\hat{V}^{-1/2}D \right)^{-1} D'\hat{\eta} \overset{H_0}{\sim} \chi_3^2(0),$$

which can be calculated entirely from summary statistics.

Sum model from summary statistics

The allelic series sum model is a constrained case of the baseline model where $\beta_l = w_l \beta_S$, w_l is the allelic series weight for variant category l, and β_S is a common scalar. We can algebraically express this model as:

$$Y = \tilde{G} w \beta_S + X\gamma + \epsilon, \tag{5}$$

where w is the 3×1 allelic series vector (w_1, w_2, w_3) and β_S is a scalar. Hence, the score statistic and its variance (now scalars) are given by:

$$\begin{aligned} \mathcal{U}_0 &= \sigma^{-2} w' \tilde{G}' Q_X Y = \sigma^{-2} w' D' G' Q_X Y \approx w' D' \hat{\eta} \\ \mathbb{V}(\mathcal{U}_0) &= \sigma^{-2} w' \tilde{G}' Q_X \tilde{G} w = w' D' \left(\sigma^{-2} G' Q_X G\right) D w \\ &= w' D' \left(V^{-1/2} R V^{-1/2} \right) D w. \end{aligned}$$

Therefore, the test statistic for the allelic series sum model is:

$$T = \mathcal{U}_0' \{\mathbb{V}(\mathcal{U}_0)\}^{-1} \mathcal{U}_0 = \hat{\eta}' D w \left(w' D' \hat{V}^{-1/2} \hat{R} \hat{V}^{-1/2} D w \right)^{-1} w' D' \hat{\eta} \overset{H_0}{\sim} \chi_1^2(0).$$

SKAT model from summary statistics

Recall that, under $H_0 : \beta = 0$, the score statistic from the joint model (3) is $\mathcal{U}_0 = \sigma^{-2} \tilde{G}' Q_X Y$. The score statistic for the SKAT model in (2) takes a similar form, $\mathcal{V} = \sigma^{-2} G' Q_X Y$, except $G = (G_1, \ldots, G_J)$ is now the individual-variant genotype matrix. The SKAT statistic [13] is a variance component test of $H_0 : \tau^2 = 0$ in (2)

based on the quadratic form $T = \mathcal{V}'K\mathcal{V}$, where K is a symmetric, positive definite kernel matrix. The kernel matrix K is calculated as a known function of the MAFs of the individual variants. The distribution of T is a mixture of $\chi^2_1(0)$s, whose mixing proportions λ_j for $j \in \{1, \ldots, J\}$ are the eigenvalues of $\mathbb{V}^{1/2}(\mathcal{V})K\mathbb{V}^{1/2}(\mathcal{V})$. From the variant-level summary statistics $\hat{\alpha}_j = (G_j'Q_XG_j)^{-1}G_j'Q_XY$ and $\mathbb{V}(\hat{\alpha}_j) = \sigma_j^2(G_j'Q_XG_j)^{-1}$, we calculate the entries of the score vector by $\mathcal{V}_j = \hat{\alpha}_j/\mathbb{V}(\hat{\alpha}_j) = \sigma_j^{-2}G_j'Q_XY$. The covariance of the per-variant scores is $\Gamma_{ab} = \mathbb{C}(\mathcal{V}_a, \mathcal{V}_b) = \sigma^2\sigma_a^{-2}\sigma_b^{-2}G_a'Q_XG_b$. Writing $\sigma_a^{-2} = \mathbb{V}^{-1}(\hat{\alpha}_a)(G_a'Q_XG_a)^{-1}$, the covariance becomes:

$$\begin{aligned}\Gamma_{ab} &= \sigma^2\mathbb{V}^{-1}(\hat{\alpha}_a)(G_a'Q_XG_a)^{-1}G_a'Q_XG_b(G_b'Q_XG_b)^{-1}\mathbb{V}^{-1}(\hat{\alpha}_a)\\ &= \sigma^2\mathbb{V}^{-1}(\hat{\alpha}_a)(G_a'Q_XG_a)^{-1/2}\big\{(G_a'Q_XG_a)^{-1/2}G_a'Q_XG_b(G_b'Q_XG_b)^{-1/2}\big\}(G_b'Q_XG_b)^{-1/2}\mathbb{V}^{-1}(\hat{\alpha}_a),\end{aligned}$$

which we approximate by $\tilde{\Gamma}_{ab} = \mathbb{V}^{-1/2}(\hat{\alpha}_a)\{\mathrm{LD}_{ab}\}\mathbb{V}^{-1/2}(\hat{\alpha}_b)$. Having calculated $\mathcal{V}$ and estimated its covariance matrix Γ, we calculate the Cholesky decomposition of $\Gamma = LL'$, then take the spectral decomposition of $L'KL$ to obtain the set of eigenvalues $\lambda_j > \epsilon = 10^{-8}$. The $H_0 : \tau^2 = 0$ is evaluated by comparing the test statistic $T = \mathcal{V}'K\mathcal{V}$ to the reference distribution $\sum_j \lambda_j\chi^2_{1,j}(0)$. The final p-value is calculated via Davies' algorithm, as implemented by the `CompQuadForm` package (v1.4.3) [15].

Simulation Studies

Type I error

Type I error studies were performed using real genotypes and phenotypes from 150K unrelated subjects of homogeneous ancestry in the UK Biobank. For the phenotype, LDL was permuted to break its association with genotype, then standardized via the rank-based inverse normal transform [16]. Genotype preparation is described in detail below. In total, more than 500 genome-wide rare-variant association studies (RVAS), including around 17K genes each, were performed, resulting in more than 10M realizations of the COAST-SS p-value under the null.

Power

In order to have fine control over the genetic architecture, power studies were based on simulated data, using the framework described previously [6]. Each simulated gene contained 10^2 variants. A random subset of variants, between 25% and 100%, were selected as causal. The effect sizes of non-causal variants were set to zero. Burden phenotypes were simulated from model (1) with various settings for $(\beta_1, \beta_2, \beta_3)$. For SKAT phenotypes, per-variant effect sizes were generated as: $\beta_j = r_j\gamma_j\sum_{l=1}^3 \mathbb{I}(A_j = l)\beta_l$, where r_j is a random sign, $\gamma_j \sim \Gamma(1, 1)$ is a positive random scalar with $\mathbb{E}(\gamma_j) = 1$ and $\mathbb{V}(\gamma_j) = 1$, and the term $\sum_{l=1}^3 \mathbb{I}(A_j = l)\beta_l$

incorporates an additional fixed weight depending on the variant's annotation.

Preparation of UK Biobank Data

Whole exome sequencing data

Variants in the full UK Biobank WES variant callset (N = 469K, [17]) were filtered out if they were located > 100 bp outside the exome capture regions, within the ENCODE blacklist [18], or within low-complexity regions [19]. We further filtered to autosomal, biallelic variants meeting the following criteria: allele balance fractions of ≤ 0.1 (homozygous) or ≥ 0.2 (heterozygous), call rate ≥ 0.95, mean depth (DP) $12 \leq \text{DP} < 200$, mean genotype quality GQ ≥ 20, Hardy-Weinberg equilibrium $p > 1 \times 10^{-10}$, and inbreeding coefficient $F > -0.3$. Variants were annotated with VEP v109.3 [7] using the following transcript consequences (for any transcript of a gene): protein-truncating variants (PTV): `splice acceptor variant`, `splice donor variant`, `stop gained`, or `frameshift variant`; missense: `missense variant`, `inframe deletion`, `inframe insertion`, `stop lost`, `start lost`, or `protein altering variant`. Missense variants were annotated to be damaging (DMV) if predicted to be `probably damaging` or `possibly damaging` by PolyPhen [20] and `deleterious` or `deleterious low confidence` by SIFT [21]. Missense variants were annotated to be benign (BMV) if predicted to be `benign` by PolyPhen and `tolerated` or `tolerated low confidence` by SIFT [22, 23]. PTVs were further filtered to remove low confidence loss-of-function by LOFTEE [24]. The most severe consequence per variant-gene pair was retained.

Quality control

Sample-level quality control (QC) was performed using 88,813 high-confidence variants passing the above filtering criteria, which were LD-pruned to $r^2 < 0.05$. To mitigate confounding due to population structure, samples were first filtered to unrelated individuals without sex chromosome aneuploidy, within ± 7 standard deviations of the mean of the first six genotype principal components [25], self-reported ethnic background of `White British`, `Irish`, or `White`, had call rate ≥ 0.95, mean DP ≥ 19, and mean GQ ≥ 47.5. We regressed out the top 20 genetic principal components from the following sample QC metrics and removed samples that were > 6 median absolute deviations from the median for any of the following metrics: number of SNPs, number of heterozygous variants, number of homozygous alternate variants, number of insertions, number of deletions, number of transitions, number of transversions, the ratio between the number of transitions and transversions, the ratio between the number of heterozygous and homozygous alternate variants, and the ratio between the number of insertions and deletions [24]. The sample size after filtering was 342,547. All data preprocessing was performed in `Hail v0.2.116` [26].

Analysis cohorts

For analyses where comparison with the original COAST was desired, we used the same subset of 145,735 subjects as in [6]. We refer to this subset as the 150K cohort. Simulation studies were also performed within the 150K cohort because the operating characteristics of statistical methods typically improve with increasing sample size, making the evaluating within the smaller cohort more challenging. The remaining analyses were performed within the superset of 342,547 subjects, referred to as the 350K cohort.

Association analysis

Processed WES data were restricted to BMVs, DMVs, and PTVs with a sample minor allele frequency ≤ 0.01. Following [14, 27], ultra-rare variants, those with a sample MAC ≤ 10, were collapsed into a single pseudo marker separately for each gene $\times$ variant category. Genes were required to have at least 3 distinct rare variants (including the pseudo markers) for inclusion in the genome-wide screen: 17,295 genes passed this threshold. Per-variant effect sizes and standard errors were calculated among 390,865 individuals adjusting for a sparse genetic relatedness matrix in `Saige` (`v1.3.0`) [28]. MAFs and per-gene LD matrices were calculated in `numpy` (`v1.23.5`). COAST-SS was performed using the `AllelicSeries R` package (`v1.0.2`).

Overlap Analysis

For each lipid trait, common-variant associations were downloaded from the NHGRI-EBI GWAS Portal [29] on 2024-10-21 and filtered to those with $p < 5 \times 10^{-8}$. Rare-variant associations from both putative loss of function (pLoF) variants and missense variants, including low-confidence pLoF variants and in-frame insertions or deletions, were obtained from the Genebass Portal [30]. Associations were filtered to those Bonferroni significant based on the number of genes with non-missing p-values, and the union of significant associations from the pLoF and missense analyses was taken.

Results

COAST-SS runs COAST starting from summary statistics

As detailed in the Methods, COAST-SS performs a rare coding-variant allelic series test starting from summary statistics. The inputs to COAST-SS are the per-variant effect sizes (i.e. $\hat{\beta}$s) and standard errors (SEs), together with per-variant MAFs and the local variant-by-variant LD matrix. We first validated that COAST-SS can recover essentially the same p-value as the original COAST when starting from summary statistics. The probability-probability plots in **Figure 1** compare the p-values from COAST-SS with those

from COAST for 4 lipid traits (cholesterol, HDL, LDL, triglycerides) using genotypes from 17K genes and phenotypes from 150K subjects in the UK Biobank. Exact agreement is not expected because COAST includes component tests that could not be computed from standard summary statistics. Nonetheless, the R^2 is at least 0.999 in all cases, and the estimated slopes are within 2% of the identity. Although its close agreement with the original COAST suggests that COAST-SS likewise provides a valid test of association, we confirmed this by performing extensive type I error studies. Specifically, RVAS with real genotypes were performed on random permutations of LDL cholesterol, generating 10M p-values under the null. The quantile-quantile plots in **Figure S1** demonstrate that p-values of COAST-SS, as well as its 3 component tests (the allelic series baseline, sum, and SKAT models), are uniformly distributed in the absence of association. **Table 1** provides estimates of the type I error at $\alpha = 2.5 \times 10^{-6}$, the Bonferroni threshold when adjusting for 20K genes, as well as the expected χ^2 statistic and genome-inflation factor λ_p at the center ($p = 0.5$) and in the lower tail ($p = 0.001$) of the p-value distribution. In all cases, the type I error is at or below the nominal threshold, while the expected χ^2 statistics and genomic-inflation factors are near the expected value of 1. COAST-SS exhibits some deflation in the center of its p-value distribution, an expected consequence of employing the Cauchy combination [31], however this is immaterial insofar as only p-values in the lower tail are of interest. Overall, we find the COAST-SS provides a valid test of association.

COAST-SS remains valid in the absence of perfect LD information

Among the summary statistics required by COAST-SS, the rare-variant LD matrix is the one least readily available. We evaluated several methods of calculating or approximating the LD matrix supplied to COAST-SS. The gold-standard approach utilizes the full sample in which the association statistics were generated to calculate the LD. The sub-sample approach utilizes a random sample of 50K subjects drawn with replacement, emulating a case where the LD matrix is calculated via a different sample from the same population. As rare variants are, by definition, present in only a small proportion of subjects, some variants present in the full sample will not appear in any given sub-sample. In these cases, the corresponding row and column of the sub-sample LD matrix were taken from the identity matrix. The final approximation, emulating a case where LD information is unavailable, simply supplies an identity matrix to COAST-SS. Although ignoring correlations between variants would likely compromise a test's operating characteristics in the common variant setting, in the rare variant setting the LD between variants is often negligible [32]. Indeed, across all 17K genes in our RVAS studies, the mean between variant R^2 in the full sample LD matrices was $\overline{R^2} = 3.4 \times 10^{-4}$, and for half of all genes the mean $\overline{R^2} \leq 1.5 \times 10^{-6}$. As a comparator for COAST-SS performed with the full-sample, sub-sample, and identity LD matrices, we present the calculation performed with the original COAST, using

individual-level data. Interestingly, the type I error studies in **Table 1** and **Figure S1** indicate that all LD methods, including the identity matrix, maintain control of the type I error. Note that these experiments utilized real genotypes, with their inherent LD structure. Moreover, the PP-plots in **Figure S2** exhibit tight concordance between applying COAST-SS with full LD versus the identity matrix to real-data analyses of 4 lipid traits in an expanded cohort of 350K subjects from the UK Biobank. Across phenotypes simulated from a variety of genetic architectures, all methods supplied nearly identical power (**Figures 2, S3**). These results suggest that COAST-SS is robust to the LD matrix, and may even be run with an identity matrix, provided it is reasonable to assume the LD among the supplied variants is small. Out of an abundance of caution, subsequent sections continue to run COAST-SS with the full-sample LD matrix unless otherwise stated.

COAST-SS substantially accelerates genome-wide analyses

A central motivation for the development of COAST-SS was to enable allelic series testing when individual-level data are unavailable. As shown in **Table 2**, a secondary advantage is that COAST-SS dramatically reduces the time required to perform a genome-wide analysis. Compared with the original COAST, COAST-SS reduces the runtime by approximately 100×. With COAST-SS, the expected time to complete a 20K genes scan is 113 ms × 20000 = 38 mins, which we further reduced by running batches of genes in parallel. There was not an appreciable difference in runtime between COAST-SS with a dense LD matrix versus the identity matrix. Of course, the reason COAST-SS is substantially faster is that much of the computation has been performed prior to allelic series testing. The computation of per-gene LD matrices, per-variant MAFs, and per-variant annotations is a one-time cost, and the resulting variant-level information can be used for analyzing any subsequent phenotype. We performed the LD and MAF calculations with `Hail` [26] and `Python`, and annotation using Ensembl's `VEP` [7]. The other prerequisite to running COAST-SS is calculating the per-variant effect sizes and SEs. Many highly optimized tools are available for this task, including `PLINK` [33], `GCTA` [34], `BOLT-LMM` [35], `Saige` [28], and `Regenie` [36].

Alternate annotation schemes provides concordant results

As described in the whole exome sequencing methods, a baseline annotation of BMV, DMV, or PTV was assigned to each variant on the basis of transcript consequence information from VEP [7] and qualitative missense pathogenicity classifications from SIFT [21] and PolyPhen2 [11]. To explore whether the baseline annotations could be improved, quantitative PolyPhen2 (S_{PP}) and AlphaMissense (S_{AM}) [12] scores were queried for each variant. Both scores are continuous values between 0 and 1, with a larger value indicating

a greater likelihood of pathogenicity. PTVs, which are not annotated by PolyPhen2 or AlphaMissense, were assigned a score of 1. Following each method's documentation, missense variants were re-categorized on the basis of their score. For PolyPhen2, variants with $S_{PP} \leq 0.15$ were assigned to the lowest category, those with $0.15 < S_{PP} \leq 0.85$ to the intermediate category ($A = 2$), and those with $S_{PP} > 0.85$ to the highest category ($A = 3$). For AlphaMissense, the cutoffs were $S_{AM} \leq 0.34$ for $A = 1$, $0.34 < S_{AM} \leq 0.564$ for $A = 2$, and $S_{AM} > 0.564$ for $A = 3$. Although annotations defined by means of categorizing pathogenicity scores may not correspond precisely to "benign", "damaging", and "protein-truncating", these categories nonetheless define a discrete, monotonic scale, across which the phenotypic impact of mutations is expected to increase.

Table 3 presents the allocation of variants to pathogenicity categories stratified by annotation scheme. Across the 17,295 genes in the analysis, the mean number of rare coding variants per gene was 30.5. The baseline annotation scheme placed the fewest variants in the high pathogenicity category, and the greatest number in the intermediate category. For both AlphaMissense and PolyPhen2, the intermediate pathogenicity category was depleted, with most variants being assigned to either the low or high categories. AlphaMissense placed the greatest number of variants in the low pathgenicity category, while PolyPhen2 placed the greatest number of variants in the high category. Across all 529K rare variants included in the analysis, 208K (39.2%) changed categories under the AlphaMissense scheme, while 184K (34.6%) changed categories under the PolyPhen2 scheme. Most shifts were between adjacent categories. Only 3.0K (0.6%) variants with AlphaMissense and 2.2K (0.4%) with PolyPhen2 changed from high to low or vice versa.

To compare the annotation schemes with respect to genetic discovery, we applied each to run COAST-SS on summary statistics for 4 lipid traits (cholesterol, HDL, LDL, triglycerides) from 350K subjects in the UK Biobank. **Figure 3(a)** depicts the number of Bonferroni significant associations with each trait under each annotation scheme, and **Figure 3(b)** reports the mean χ^2 statistic across Bonferroni significant genes, a proxy for the power to detect the association. In general, all methods provided highly concordant results. Across gene-trait pairs that were significant according to at least one annotation scheme, the Pearson R^2 between the negative $\log_{10}$ p-values was 0.990 for baseline and AlphaMissense, 0.989 for baseline and PolyPhen2, and 0.986 for AlphaMissense and PolyPhen2.

COAST-SS identifies new candidate allelic series for lipid traits

Candidate allelic series for the 4 lipid traits were previously reported based on applying COAST to individual-level data from 150K subjects in UK Biobank [6]. We compared the Bonferroni significant associations identified by COAST-SS applied to summary statistics from the expanded 350K cohort with those from the initial 150K analysis. For simplicity and comparability with the previous findings, the results in this section utilize the baseline annotations. The Venn diagrams in **Figure S4** depict the overlaps between the Bonferroni significant genes from each analysis. On average, 92.3% of associations identified in the smaller cohort and tested in the larger cohort were Bonferroni significant in the latter, while between 0 and 2 reaching Bonferroni significance in the initial analysis were not tested or were no longer significant in the expanded analysis. Two genes (*AC126283.2* and *APOA1*) comprising 3 of the 6 associations not observed in the larger cohort failed QC in the latter, and so were not tested. As expected, among associations significant in both analyses, the mean χ^2 increased substantially between the 150K and 350K analyses (**Figure S5**), by 2.2× on average. For each trait, COAST-SS in the expanded cohort identified between 9 and 12 additional associations, increasing the number of candidate allelic series by 118% on average. **Figure 5** presents the pattern of effect sizes across variant annotation categories among all Bonferroni significant associations from the 350K analysis. As previously observed, although genes identified by COAST-SS do not universally follow an allelic series pattern, as indicated by a monotonic increase in effect size magnitude from BMVs to DMVs to PTVs, a majority of the genes identified do follow the intended pattern. We also explored the pattern of effect sizes among the 3 genes that reached significance in the 150K analysis but not the 350K analysis. As seen in **Figure S6**, in 2 cases the pattern of effect sizes was not monotone increasing, while in the third (*ALB* with total cholesterol) the pattern was consistent with an allelic series and the p-value was suggestive.

For each lipid trait, we overlapped the genes identified in the expanded 350K analysis but not the original 150K analysis with previously reported associations from both Genebass [30], to assess rare-variant support, and from the GWAS Catalog [29], to assess common-variant support. The results are presented in **Table 4**. Among the 43 total newly identified allelic series candidates, in all cases the same gene had previously been associated with the same trait using either common or rare variants; in 30 cases (69.8%), the association had prior support from both common and rare variant genetics. Note that while the existence of an association had previously been reported, here we are reporting the presence of a potential dose-response relationship.

Discussion

In this work, we described COAST-SS, an extension of the coding-variant allelic series test that accepts commonly available GWAS-type summary statistics as input. In two respects, COAST-SS is not identical to the original COAST. First, reconstructing the allelic series max test from standard summary statistics was not possible. Rather than requiring non-standard summary statistics as input, we elected to remove the max test from COAST-SS. Second, the distinction between the count and indicator models was dropped because investigators working from summary statistics often will not have control over how the genotypes are encoded (i.e. as indicators or counts). Nonetheless, analyses of lipid traits among the subset of 150K UK Biobank subjects utilized for developing the original COAST demonstrate that COAST-SS is able to recover practically equivalent p-values. Moreover, extensive type I error studies based on real data (with a permuted phenotype) indicate that COAST-SS provides a valid test of association.

Although our experiments suggest that COAST-SS remains valid in the absence of LD information, we recommend providing an in-sample LD matrix whenever possible. While minimal LD is expected among rare variants, the extent to which LD can be safely ignored likely depends on accuracy with which the full sample LD matrix can be approximated by the identity matrix. In our analyses of the UK Biobank, the LD among rare variants in the full sample LD matrix was negligible, making approximation by the identity reasonable. A limitation of the present study is that it is unclear how broadly the conclusion that LD can be safely ignored generalizes. Important directions for future research are to characterize the extent of rare variant LD in cohorts beyond the UK Biobank, and to quantify the sensitivity of COAST-SS to unaccounted-for LD. That is, if the LD matrix is approximated by the identity out of convenience or necessity, but the true LD is non-negligible, to what extent are the operating characteristics of COAST-SS disrupted?

Contrary to expectations, utilizing pathogenicity scores from AlphaMissense and PolyPhen2 to reclassify missense variants had only a minor impact on the power to identify associations with lipid traits. This is likely because the annotation categories of most variants ($> 60\%$) were unchanged under the alternate classification schemes, and among variants that did change categories, most ($> 98\%$) shifted by only a single category, predominantly from intermediate (under baseline) to either low or high. In principle, reclassifying variants is only expected to materially affect power when enough variants are reclassified, or those that are reclassified have sufficient effect sizes, that the mean effect sizes across the 3 annotation categories change appreciably. Future work will explore whether other emerging techniques for quantifying variant pathogenicity can increase power for detecting allelic series.

In an expanded analysis of lipid allelic series among 350K subjects from the UK Biobank, 92.3% of the 39 associations identified in the 150K analysis and tested in the 350K analysis remained significant in the latter, with the strength of association increasing by 2.2× on average. Meanwhile, the number of associations more than doubled, with 43 new associations achieving Bonferroni significance in the 350K analysis. While the pattern of effect sizes across variant classes was not always monotone increasing, most of the newly identified associations did exhibit the target pattern of effects. Collectively, these findings suggest that the discovery of allelic series for lipid traits has not yet reached saturation.

In conclusion, we believe that COAST-SS will provide researchers with a valuable and easy-to-use tool for identifying rare-variant allelic series starting from commonly available summary statistics, and in doing so, will significantly expand the set of traits on which well-powered allelic series testing can be performed. COAST-SS has been incorporated into the `AllelicSeries R` package, which is available on `CRAN`.

References

[1] B McClintock. The relation of homozygous deficiencies to mutations and allelic series in maize. *Genetics*, 29(5):478 – 502, 1944.

[2] PM Plenge, EM Scolnick, and D Altshuler. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*, 12(8):581 – 594, 2013.

[3] K Estrada, S Froelich, A Wuster, et al. Identifying therapeutic drug targets using bidirectional effect genes. *Nat Commun*, 12:2224, 2021.

[4] T Li, Y Yang, H Qi, et al. Crispr/cas9 therapeutics: progress and prospects. *Signal Transduct Target Ther*, 8(1):36, 2023.

[5] IP Gorlov, OY Gorlova, ML Frazier, MR Spitz, and CI Amos. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet*, 79(3):199 – 206, 2011.

[6] ZR McCaw, C O'Dushlaine, H Somineni, M Bereket, C Klein, T Karaletsos, FP Casale, D Koller, and TW Soare. An allelic-series rare-variant association test for candidate-gene discovery. *Am J Hum Genet*, 110(8):1330 – 1342, 2023.

[7] W McLaren, L Gil, SE Hunt, et al. The ensembl variant effect predictor. *Genome Biol*, 17(1):122, 2016.

[8] W Chen, Y Wu, Z Zheng, et al. Improved analyses of gwas summary statistics by reducing data heterogeneity and errors. *Nature Communications*, 12(1):7117, 2021.

[9] PI Kontou and PG Bagos. The goldmine of gwas summary statistics: a systematic review of methods and tools. *BioData Mining*, 17(1):31, 2024.

[10] X Li, C Quick, H Zhou, et al. Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nat Genet*, 55(1):154 – 164, 2023.

[11] IA Adzhubei, DM Jordan, and SR Sunyaev. Predicting functional effect of human missense mutations using polyphen-2. *Curr Protoc Hum Genet*, 7(20), 2013.

[12] J Cheng, G Novati, J Pan, et al. Accurate proteome-wide missense variant effect prediction with alphamissense. *Science*, 381(6664):eadg7492, 2023.

[13] MC Wu, S Lee, T Cai, Y Li, M Boehnke, and X Lin. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*, 89(1):82 – 93, 2011.

[14] Y Liu and J Xie. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402, 2020.

[15] P. Lafaye de Micheaux. *CompQuadForm: Distribution Function of Quadratic Forms in Normal Variables*, 2017.

[16] ZR McCaw, JM Lane, R Saxena, S Redline, and X Lin. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*, 76(4):1262 – 1272, 2020.

[17] JD Backman, AH Li, A Marcketta, et al. Exome sequencing and analysis of 454,787 uk biobank participants. *Nature*, 599(7886):628 – 634, 2021.

[18] HM Amemiya, A Kundaje, and AP Boyle. The encode blacklist: Identification of problematic regions of the genome. *Sci Reports*, 9:9354, 2019.

[19] H Li. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20):2843 – 2851, 2014.

[20] Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4):248 – 249, 2010.

[21] Ng PC and Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*, 11(5):863 – 874, 2001.

[22] ET Cirulli, S White, RW Read, et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun*, 11:542, 2020.

[23] CV Van Hout, I Tachmazidou, JD Backman, et al. Exome sequencing and characterization of 49,960 individuals in the uk biobank. *Nature*, 586:749 – 756, 2020.

[24] KJ Karczewski, LC Francioli, G Tiao, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581:434 – 443, 2020.

[25] C Bycroft, C Freeman, D Petkova, et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562:203 – 209, 2018.

[26] Hail Team. *Hail*.

[27] W Zhou, W Bi, Z Zhao, et al. Saige-gene+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet*, 54(10):1466 – 1469, 2022.

[28] W Zhou, JB Nielsen, LG Fritsche, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, 50(9):1335 – 1341, 2018.

[29] A Buniello, JAL MacArthur, M Cerezo, et al. The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1):D1005–D1012, 2019.

[30] KJ Karczewski, M Solomonson, KR Chao, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 uk biobank exomes. *Cell Genomics*, 2(9):100168, 2022.

[31] DJ Liu, GM Peloso, H Yu, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*, 49(12):1758 – 1766, 2017.

[32] B Li, DJ Liu, and SM Leal. Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Gene*, 1(26), 2013.

[33] S Purcell, B Neale, K Todd-Brown, et al. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559 – 575, 2007.

[34] J Yang, SH Lee, ME Goddard, and PM Visscher. Gcta: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, 88(1):76 – 82, 2011.

[35] PR Loh, G Tucker, BK Bulik-Sullivan, et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*, 47(3):284 – 290, 2015.

[36] J Mbatchou, L Barnard, J Backman, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*, 53(7):1097 – 1103, 2021.

Figures

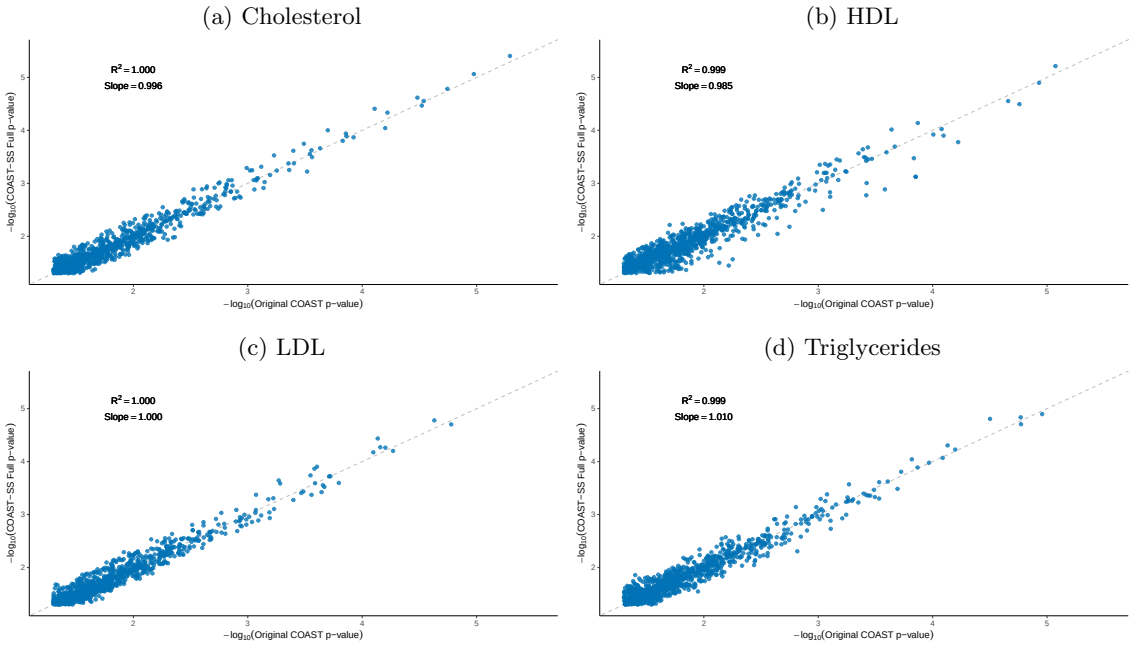


Figure 1: **COAST from summary statistics (COAST-SS) recovers essentially the same p-value as the original COAST.** Note that perfect agreement is not expected because COAST-SS includes a different set of component tests than COAST. Shown are genes with a $P \leq 0.05$ for 4 lipid traits from the UK Biobank. The sample size was $N = 150$K. The dotted gray line is the identity. The ideal value for both the R^2 and the estimated slope is 1.

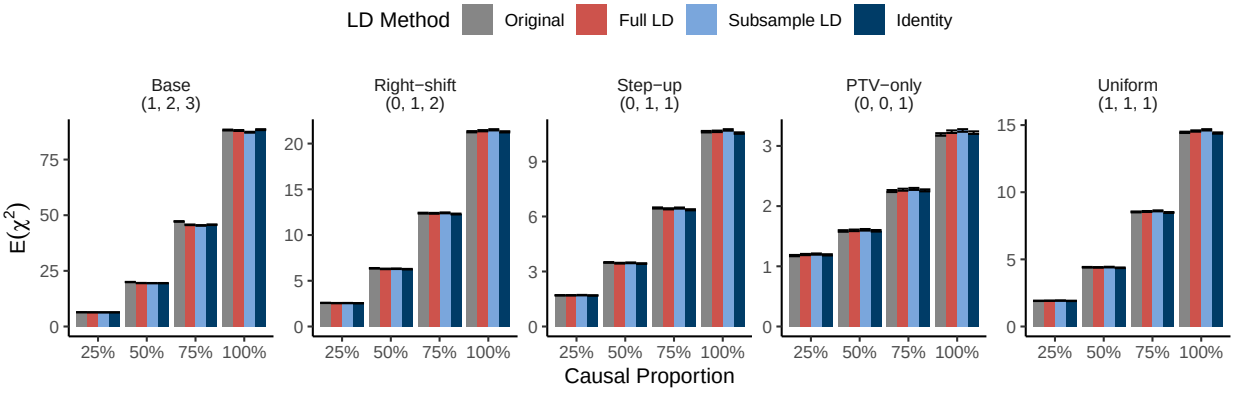
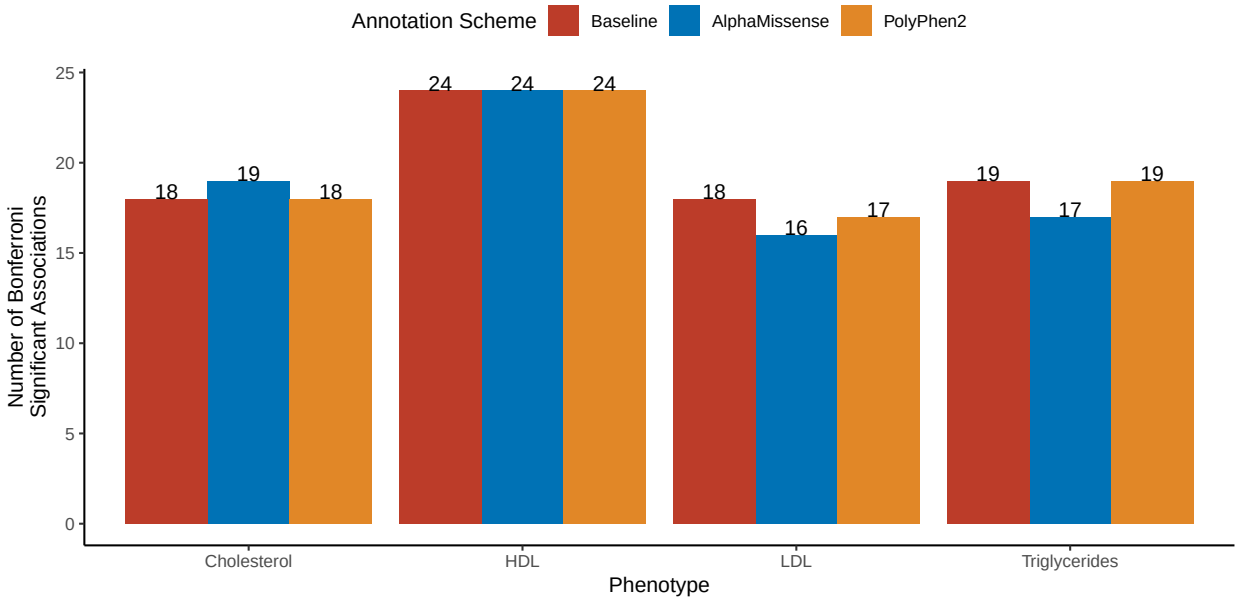


Figure 2: **COAST-SS has equivalent power to COAST, and is robust to the provided LD matrix.** Original refers to COAST run with individual-level data; full LD was calculated using all 150K subjects; sub-sampled LD was calculated using a random sample of 50K subjects, drawn with replacement; identity ignores LD. Results are presented under 5 different generative models. The tuple above each set of results indicates the true effect sizes of BMVs, DMVs, and PTVs in the generative model. Power is quantified by the expected χ^2 statistic. The simulation sample size was $N = 10^4$. Error bars are 95% confidence intervals for the mean across $R = 10^5$ simulation replicates.

(a) Count of Bonferroni Significant Associations



(b) Mean χ^2 at Bonferroni Significant Associations

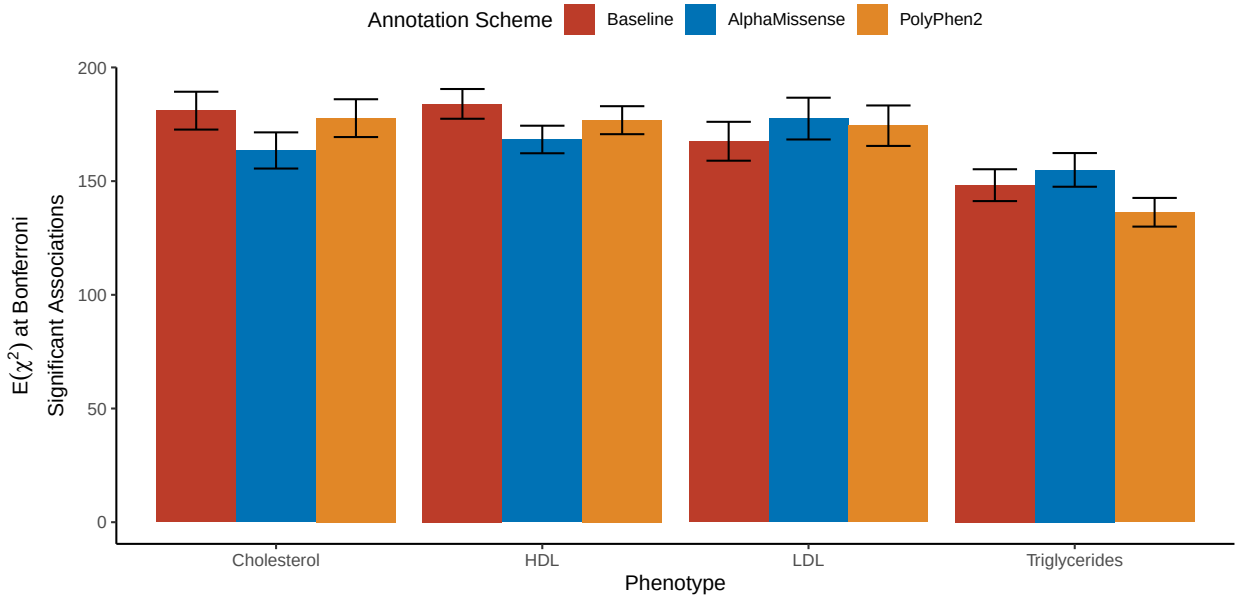


Figure 3: **Different missense variant annotation schemes provided similar power for detecting associations with lipid traits.** COAST-SS with full LD was run on summary statistics for lipid traits from $N = 350$K subjects in the UK Biobank. Baseline is the original annotation scheme, which classifies variants into BMVs, DMVs, and PTVs. AlphaMissense and PolyPhen2 are annotation schemes in which missense variants were reclassified based on their quantitative pathogenicity score. (a) depicts the number of Bonferroni significant associations, adjusting for 17,295 tests of association, and (b) presents the mean χ^2 statistics across the Bonferroni significant associations, a proxy for power. Error bars are 95% confidence intervals.

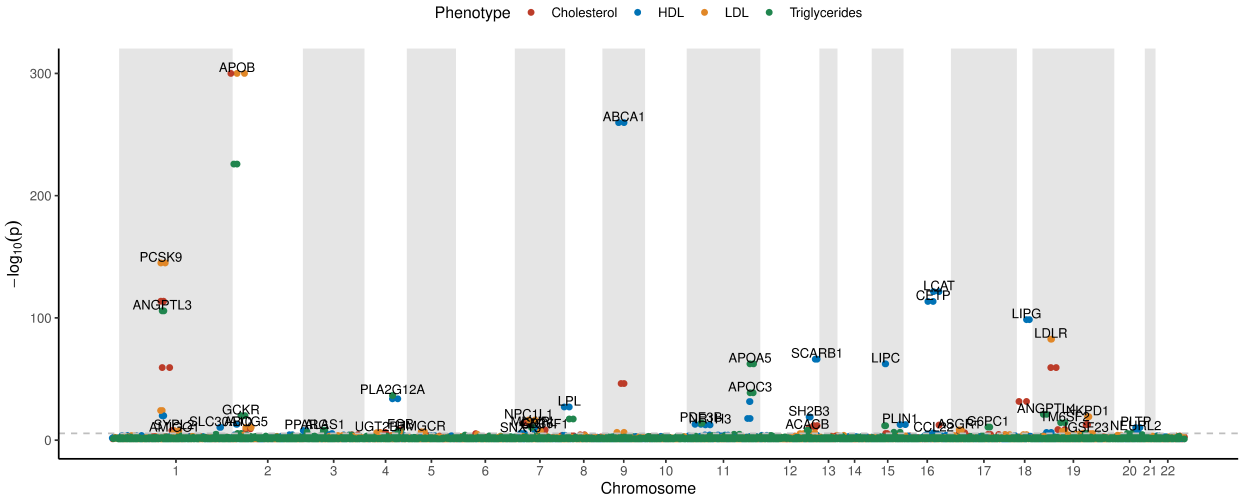


Figure 4: **Manhattan plots for allelic series analysis of lipid trait summary statistics generated from 350K subjects in the UK Biobank.** The analysis includes 17,295 genes. Note that modest horizontal jitter has been applied to facilitate visualization of super-imposed points. Gene symbols are shown for Bonferroni significant associations.

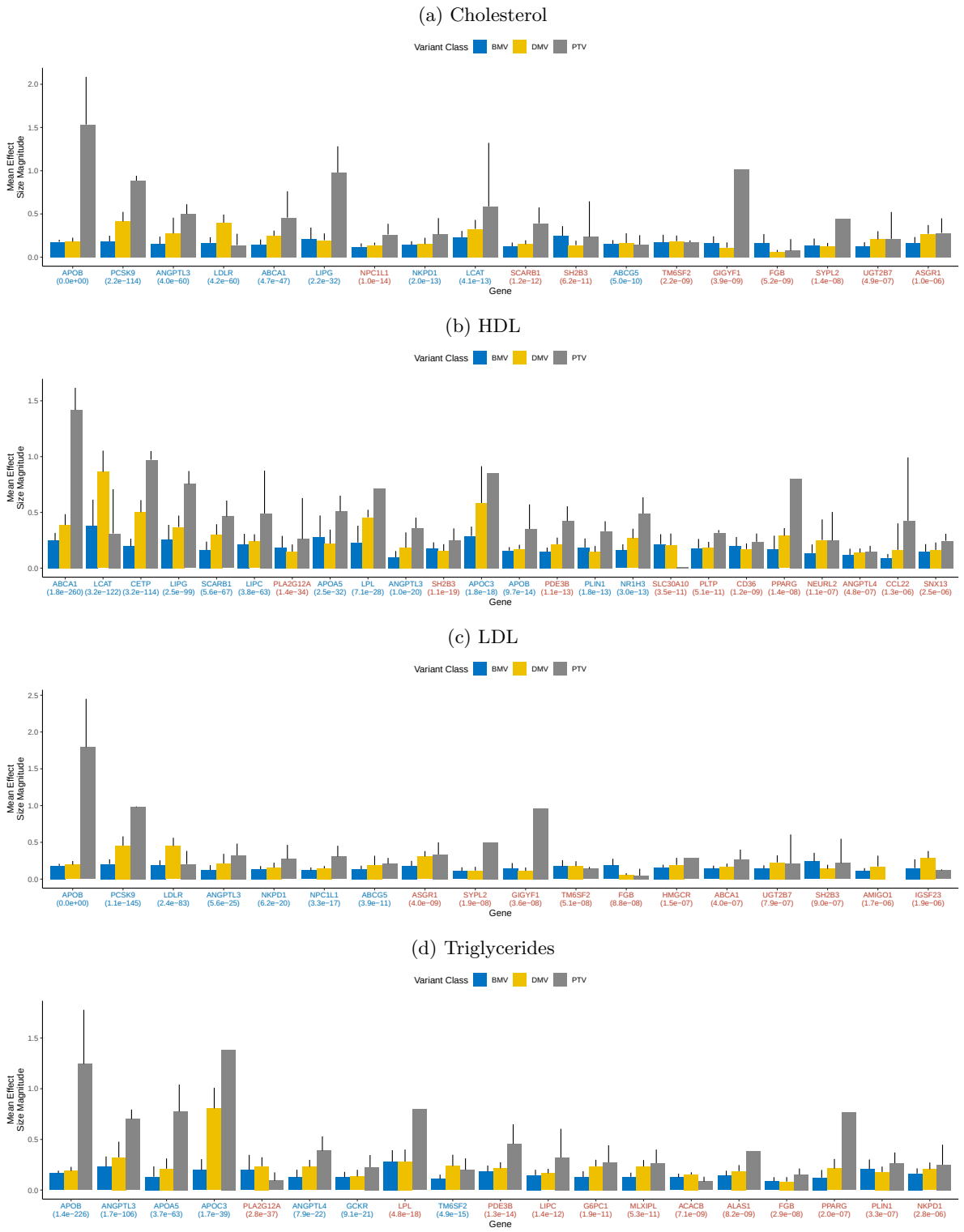


Figure 5: **Effect size patterns among candidate lipid trait allelic series identified by COAST-SS using summary statistics from 350K subjects in the UK Biobank.** COAST-SS was run with the full LD and baseline annotation scheme. Shown are the mean absolute effect sizes across the BMVs, DMVs, and PTVs in a given gene. Error bars are 95% confidence intervals (CIs). The absence of a CI indicates insufficiently many variants to estimate a standard error. The gene symbol is colored blue if the gene was already reported in the 150K analysis, and in red if the gene is newly reported in the 350K analysis. The number in parentheses below the gene symbol is the COAST-SS p-value.

Tables

Test	LD Method	T1E ($\times 10^6$)	$\mathbb{E}(\chi^2)$	$\lambda_{0.5}$	$\lambda_{0.001}$
Baseline	Original	2.10	1.00	1.00	1.00
Baseline	Full LD	2.10	1.00	1.00	1.00
Baseline	Sub-sample LD	2.40	1.00	1.00	1.00
Baseline	Identity	2.30	1.00	1.00	1.00
Allelic Sum	Original	1.80	1.00	1.00	1.00
Allelic Sum	Full LD	1.80	1.00	1.00	1.00
Allelic Sum	Sub-sample LD	2.00	1.00	1.00	1.00
Allelic Sum	Identity	2.40	1.01	1.01	1.01
Allelic SKAT	Original	1.50	1.00	0.98	1.00
Allelic SKAT	Full LD	2.00	1.00	0.99	1.00
Allelic SKAT	Sub-sample LD	2.20	1.00	0.99	1.00
Allelic SKAT	Identity	2.40	1.01	0.99	1.01
COAST-SS	Original	2.40	1.00	0.89	1.00
COAST-SS	Full LD	2.00	1.00	0.90	1.00
COAST-SS	Sub-sample LD	2.20	1.00	0.90	1.01
COAST-SS	Identity	2.50	1.01	0.90	1.01

Table 1: **COAST-SS controls type I error under the null.** Results are based on 10M p-values generated using real data on 150K subjects from the UK Biobank. The phenotype was LDL, which was permuted to break any association with genotype. Type I error ($\times 10^6$) is reported at $\alpha = 2.5 \times 10^{-6}$, which is the Bonferroni threshold when adjusting for 20K genes; the expected value is 2.5. Also reported are the expected χ^2, the median genomic inflation factor $\lambda_{0.5}$, and the tail genomic inflation factor $\lambda_{0.001}$. LD method refers to how the LD matrix was estimated; original means the test was performed with the original COAST (i.e. using individual-level data).

Test	LD Method	Min	p_{25}	p_{50}	μ	p_{75}	Max
COAST	Original	7007.00	9822.00	10309.00	10653.14	11062.00	62635.00
COAST-SS	Full LD	3.00	68.00	79.00	113.21	99.00	2319.00
COAST-SS	Identity	2.00	69.00	80.00	113.88	102.00	1834.00

Table 2: **COAST-SS substantially accelerates genome-wide analyses.** 5 genome-wide RVAS were conducted using real data on 150K subjects from the UK Biobank. Results refer to per-mean runtimes and are measured in milisections (ms). Min refers to the minimum, p_{25} the 25th percentile, p_{50} the median, μ the mean, p_{75} the 75th percentile, Max the maximum.

Annotation	Lowest	Intermediate	Highest
Baseline	13.4	14.7	2.5
AlphaMissense	20.3	3.4	6.8
PolyPhen2	14.1	4.2	12.2

Table 3: **Mean number of variants in each annotation category (predicted pathogenicity) by annotation scheme.** For baseline, the three categories correspond to BMVs, DMVs, and PTVs, respectively. For AlphaMissense and PolyPhen2, variants were reclassified on the basis of their quantitative pathogenicity score. The mean is calculated across the 17,295 genes included in the UK Biobank 350K analysis. Marginally, the mean number of variants per gene is 30.5.

Trait	N Genes	N Catalog	N Genebass	N Both	N Neither
Cholesterol	9	9	8	8	0
HDL	11	10	9	8	0
LDL	11	11	7	7	0
Triglycerides	12	9	10	7	0

Table 4: **Most genes newly identified by COAST-SS in the expanded cohort have both common and rare variant support for association with the corresponding lipid trait.** For each trait, the GWAS Catalog was queried for genome-wide significant common-variant associations, and Genebass was queried for Bonferroni significant rare-variant associations based on both missense and putative loss of function variants. N Genes is the total number of Bonferroni significant associations identified in the expanded 350K cohort but not the original 150K cohort. N Catalog and N Genebass are the counts of the newly significant associations that overlap with the GWAS Catalog and Genebass, respectively. N Both and N Neither are the counts of newly significant associations in both or neither the GWAS Catalog and Genebass. While the existence of an association had previously been reported in all cases, here we are highlighting the presence of a potential dose-response relationship.

Declaration of Interests

ZRM, RD, the insitro Research Team, EF, CO, TS are employees and shareholder of Insitro. JG completed an internship at Insitro.

Acknowledgements

The authors are grateful to the participants of the UK Biobank, whose data were used with permission.

Author Contributions

ZRM, CO, and TWS conceived of the project. TWS curated genetic and phenotypic data. ZRM, JG, and RD developed and implemented the method. ZRM, JG, and TWS performed analyses. All authors provided scientific input. ZRM wrote the first draft of the manuscript. All authors contributed to critical revision of the final manuscript.

Data and Code Availability

This work uses genotypes and phenotypes from the UK Biobank (https://www.ukbiobank.ac.uk/) accessed under approved application number **51766**. The coding-variant allelic series test has been implemented in the `AllelicSeries R` package, which is available on CRAN.