Functional gene embeddings improve rare variant polygenic risk scores

Shubhankar Londhe[1,2,+], Jonas Lindner[1,+], Zhifen Chen[6,7], Eva Holtkamp[1,2], Florian R. Hölzlwimmer[1], Francesco Paolo Casale[1,4,5], Felix Brechtmann[1,*], Julien Gagneur[1,2,3,4,*]

[1] School of Computation, Information and Technology, Technical University of Munich, Garching, Germany
[2] Helmholtz Association – Munich School for Data Science (MUDS), Munich, Germany
[3] Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany
[4] Computational Health Center, Helmholtz Center Munich, Neuherberg, Germany
[5] Helmholtz Pioneer Campus, Helmholtz Center Munich, Neuherberg, Germany
[6] Department of Cardiology, Deutsches Herzzentrum München, Technical University Munich, 80636 Munich, Germany
[7] Deutsches Zentrum für Herz- und Kreislaufforschung (DZHK), partner site Munich Heart Alliance

+: These authors contributed equally.

*: To whom correspondence should be addressed. Email: gagneur@in.tum.de. Correspondence may also be addressed to Felix Brechtmann. Email: felix.brechtmann@in.tum.de

Abstract

Rare variant association testing is a promising approach to identify effector genes for common traits. However, ensuring sensitive and robust rare variant association testing is challenging due to the scarcity of high-impact rare-allele carriers. Here we introduce FuncRVP, a Bayesian rare variant association framework that addresses this issue by leveraging functional gene embeddings, i.e. multidimensional representations of gene function. FuncRVP models the accumulated effects of rare variants on traits as a weighted sum of rare-variant gene impairment scores. A prior, learnt from data, regularizes the weight of each gene depending on the location of the gene in a functional gene embedding. We investigated this approach on 41 quantitative traits from unrelated UK Biobank exome samples. For the gene impairment score, we found that DeepRVAT scores strongly outperformed loss-of-function variant counts. Moreover, effective weight regularizations were obtained when using a prior on gene weight variance, but not on gene weight expectation. Compared to linear regressions on significantly associated genes, FuncRVP typically improved trait prediction, and did so more effectively for traits with higher burden heritability. Moreover, FuncRVP led to more robust gene effect estimates and to increased gene discoveries notably among genes that are more genetically constrained. Altogether, these results demonstrate that the integration of functional information across genes improves rare-variant phenotype prediction and gene discovery.

Introduction

Population-scale whole-exome and whole-genome sequencing vastly expand the potential to understand the genetic underpinnings of human traits and diseases. Such large-scale sequencing studies have unveiled that an average individual carries dozens of potentially deleterious rare variants resulting from germline mutations [1]. Rare variants can exhibit large effects, aiding the discovery of effector genes with major roles in the molecular mechanisms underlying diseases and therefore prioritizing potential drug targets. For instance, the discovery of protective loss-of-function (LoF) variants in proprotein convertase subtilisin kexin 9 (*PCSK9*) led to the development of inhibitors for coronary heart disease (CAD) [2,3]. Furthermore, rare variant derived polygenic risk scores (PRS) not only allow identifying individuals at high disease risk but also better generalize across populations with different ethnicity, benefitting from a milder dependence on linkage disequilibrium (LD) structure compared to common variants [4].

While genome-wide association studies have been instrumental in identifying common genetic variants associated with complex traits, they face challenges when analyzing rare variants due to the scarce number of allele carriers and increased multiple testing burden. To address this limitation, rare variant association tests (RVATs), including burden tests and variance-component tests such as sequence kernel association tests, aggregate rare variants within a gene or a genomic region [5–10]. In this context, we have recently developed DeepRVAT, a model that integrates annotations of all rare variants into a trait-agnostic and individual-specific gene impairment score [11]. DeepRVAT borrows information across genes and traits to learn a trait-agnostic and individual-specific gene impairment score that integrates annotations of rare variants. DeepRVAT models the phenotype as a weighted sum of gene impairment scores of significantly associated genes. Despite these strong assumptions, DeepRVAT outperformed previous rare variant association testing approaches for gene discovery and phenotype prediction [11].

Rare variant association testing often reveals functionally related genes being associated with the same traits, for instance genes associated with cholesterol found by Fiziev et al. [4,12]. Leveraging this observation, gene-set RVATs perform burden tests or variance-component tests on a-priori-defined groups of genes such as pathways [13–16]. Gene-set RVAT can show increased sensitivity in capturing pathways associated with traits [17], however, this approach neither allows pinpointing individual gene associations nor deriving phenotype predictors.

Recently, functional gene embeddings, numerical vectors capturing gene function such that genes with similar functions are close in the embedding space, have been proposed as an alternative to gene sets to represent gene functions [18–20]. Since they are vector representations rather than sets, functional gene embeddings can retain some quantitative information about functional similarities and are more straightforwardly integrated into machine learning algorithms. Recently, the development of Polygenic Priority Score (PoPS) has shown the utility of functional gene embedding to prioritize genes at associated loci to be helpful for genome-wide association studies [21]. However, their utility for rare variant association testing has not yet been explored.

Here we propose to leverage functional gene embedding to increase the power of RVAT while maintaining the capability of developing phenotype predictors and individual gene discovery. To this end, we developed a genotype-to-phenotype model we named FuncRVP (Functional Rare Variant Polygenic Risk Score). In the following, we present the architecture of the model, explore the modeling assumptions using 41 quantitative traits across 161,850 unrelated individuals of the UK Biobank, and assess the model robustness on an evaluation cohort with a similar number of unrelated UK Biobank individuals. We end by highlighting some novel associations with cardiovascular traits.

Methods

Datasets

Phenotype processing and filtering samples

We used 200,643 whole-exome sequencing samples with corresponding blood and urine measurements from the UK Biobank [22] for training and benchmarking FuncRVP phenotype prediction performance against other approaches. We used the *ukb_gen_samples_to_remove* function of the ukbtools R-package (v0.11.3, [23]) together with pre-computed relatedness scores (see UK Biobank Resource 668) to remove closely related individuals, keeping only one representative of groups that are related to the 3rd degree or less. Individuals with European ancestry were identified using UK Biobank data field 22006 (termed 'Caucasian'). This filtering resulted in a dataset (called the *discovery cohort*) of 161,850 individuals.

To evaluate the generalizability of gene effects estimated by our model we used the remaining 307,929 whole-exome sequencing samples from the UK Biobank. These samples were filtered identically to the previous ones, resulting in 209,492 further unrelated individuals with European ancestry (called the *evaluation cohort*).

A detailed list of phenotypes used can be found in Table S1. All measurements were inverse rank normal transformed, corresponding to the data preprocessing in the Genebass study [24]. For each trait, the dataset of genotypes and corresponding trait measurements was split randomly into training, validation, and test data. The split was done such that the splits contain 60%, 20%, and 20% of the data, respectively. The training and validation data were used to optimize the model.

Covariates and Polygenic Risk Scores (PRS)

For each sample, the covariates age (UKBB data field 21003), age^2, sex (UKBB data field 31), age:sex, age^2:sex, and the first 20 genetic principal components (PCs) (UKBB data field 22009_0.[1-20]) were obtained directly from the UK Biobank. All of these were used as covariates for the burden test and for FuncRVP. Polygenic risk scores were obtained from the PGS catalog [25] using the study by Privé et al. [26], and applied to the genotyping array data using PLINK 2.0 [27].

DeepRVAT gene impairment scores

We use the DeepRVAT gene impairment scores as the gene burdens for all models [11]. DeepRVAT summarizes multiple variant effect annotations into one aggregated score per gene. The DeepRVAT gene impairment scores were obtained for 469,382 individuals from the UK Biobank for 17,984 protein-coding genes. These scores were mean shifted to center the mean score of each gene to zero. This ensures that no gene impairment results in no effect of the gene on the phenotype when burden testing.

Gene Embedding

We utilize gene function representations from Brechtmann et al. [18]. In these functional gene embeddings, each protein-coding gene is encoded as a 256-dimension real vector. All the embeddings benchmarked in this work are considered, and their performance is compared in our task of phenotype prediction using FuncRVP. The PoPS feature matrix had 57743 dimensions, which were projected to 256 dimensions using PCA.

Across all 41 traits used in the study, we found that a combination of the “Omics” and “PoPS” embeddings (concatenating the two embeddings per gene) performed the best overall (Figure S1). This combined “Omics + PoPS” gene embedding has a 512-dimensional vector for each gene, as it is a concatenation of two embeddings. We use these embeddings as input to FuncRVP to generate the functional prior. Upon intersection with the genotype matrix from DeepRVAT, we obtain 17,517 coding genes.

Discoveries from other RVAS

We included gene-trait associations found by the genebass study [24] and the study by Backman et al. [28]. These RVAS were performed on 394,841 and 454,787 individuals from the UK Biobank. Gene-trait associations for all 41 traits that we use were available in both studies.

GWAS index variants

On the Pan-UK Biobank [29] GWAS summary statistics we identified independently associated variants through LD-based clumping using PLINK v1.9 [27] with default parameters. These variants were filtered at the conventional GWAS p-value threshold of 5×10^{-8} to obtain the index variants for each phenotype.

Burden heritability

Burden heritability scores were obtained from supplementary table 8 of the work by Weiner et al. [30]. The column “aggregated_h2” from this table was used as the burden heritability score for each trait. The traits studied here overlapped with only 16 out of the 41 traits that we benchmark on.

LOEUF scores

The loss-of-function observed/expected upper bound fraction (LOEUF) were obtained from the download portal of gnomad v2.1.1 [1].

OMIM genes

OMIM genes associated with UK biobank phenotypes were obtained from supplementary table 7 of the work by Fiziev et al. [4]. For genes associated with common phenotypes found through their analysis and a previous RVAS [28], they selected OMIM disease phenotypes associated with these genes. This resulted in triplets of gene, UK Biobank and OMIM phenotypes. They curated

these triplets so that the UK Biobank phenotype was clinically related to the OMIM phenotype. We filter this table to only include the 41 phenotypes on which we have performed our analysis.

Gene investigation

Single-cell expression profiles for the gene *ARHGAP15* were retrieved from the Human cell Atlas Portal on 4th July 2024 under: https://www.proteinatlas.org/ENSG00000075884-ARHGAP15/single+cell+type

Single-cell expression profiles for the gene *RHOF* were retrieved from the Human cell Atlas Portal on 4th July 2024 under: https://www.proteinatlas.org/ENSG00000139725-RHOF/single+cell+type/PBMC

GWAS correction

For associations we investigated, we conducted conditional association tests using GWAS summary statistics from the Pan-UK Biobank [29]. Independently associated variants were identified from GWAS summary statistics through LD-based clumping using PLINK v1.9 [27] with default parameters, restricting to associations with a p-value $< 10^{-7}$ and MAF > 1%. P-values from the GWAS on Europeans only were used from the Pan-UK Biobank as our model is also trained on this data.

FuncRVP Model

For each individual *i*, the trait measurement y_i was modeled as:

$$y_i \sim N(\boldsymbol{g}_i^\top \boldsymbol{\beta} + \boldsymbol{c}_i^\top \boldsymbol{\gamma} + b, \sigma^2) \quad \text{(Eq. 1)}$$

where $\boldsymbol{g}_i$ the vector of gene impairment score, $\boldsymbol{\beta}$ the vector of gene effects, $\mathbf{c}_i$ is the vector of covariates, which includes the polygenic risk score and the first 20 genetic principal components, $\boldsymbol{\gamma}$ is the vector of covariate effects, and σ^2 the trait variance.

We assumed a prior on the effect of the jth gene to be normally distributed around zero with a gene-specific variance $\sigma^2_{\beta_j}$ (index of index), i.e. we assumed:

$$\beta_j \sim N(0, \sigma^2_{\beta_j}) \quad \text{(Eq. 2)}$$

$$\sigma^2_{\beta_j} = f(\boldsymbol{e}_j; \boldsymbol{\theta}) + \sigma_0^2 \quad \text{(Eq. 3)}$$

where $\mathbf{e}_j$ is the functional embedding of gene *j*. The function f is modeled using a neural network with parameters $\boldsymbol{\theta}$. The gene-specific variance $\sigma^2_{\beta_j}$ is given as the sum of a gene-specific term modeled by f on the embedding $\mathbf{e}_j$ and a constant shared variance σ_0^2.

To simplify notation, we use Σ_θ to denote the diagonal matrix whose diagonal elements are the $\sigma^2_{\beta_j}$ across all genes.

This Bayesian model was optimized in a two-step process. Initially, the parameterization of the neural network f was optimized by maximizing the likelihood of $\boldsymbol{y}$ after marginalizing out the gene effects $\boldsymbol{\beta}$. The MAP of $\boldsymbol{y}$ is given by:

$$\max_{\boldsymbol{\theta},\boldsymbol{\gamma},b,\sigma_0^2} N(\boldsymbol{y}|\boldsymbol{C}^\top\boldsymbol{\gamma} + b\mathbb{1}, \sigma^2 + \boldsymbol{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{G}^\top) \quad \text{(Eq. 4)}$$

We learn the optimal parameters by minimizing the negative log-likelihood with an L1-penalty on the parameters of f, denoted by $\boldsymbol{\theta}$, and minimized the resulting quantity using the Adam optimizer [31]:

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}, \hat{b}, \hat{\sigma_0^2}) = \operatorname{argmin}_{\boldsymbol{\theta},\boldsymbol{\gamma},b,\sigma_0^2} N(\boldsymbol{y}|\boldsymbol{C}^\top\boldsymbol{\gamma} + b\mathbb{1}, \sigma^2 + \boldsymbol{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{G}^\top) + \alpha||\boldsymbol{\theta}||_1 \quad \text{(Eq. 5)}$$

The learned gene effects after training were obtained from the posterior distribution of $\boldsymbol{\beta}$, given by:

$$\boldsymbol{\beta}|\boldsymbol{y} \sim N(\boldsymbol{V}\boldsymbol{G}^\top(\hat{\sigma^2})^{-1}(\boldsymbol{y} - (\boldsymbol{C}^\top\hat{\boldsymbol{\gamma}} + \hat{b}\mathbb{1})), \boldsymbol{V}) \quad \text{(Eq. 6)}$$

where the posterior variance $\mathbf{V}$ is given by:

$$\boldsymbol{V} = [\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1} + (\hat{\sigma^2})^{-1}\boldsymbol{G}^\top\boldsymbol{G}]^{-1} \quad \text{(Eq. 7)}$$

Train-validation-test split

The 161,850 samples from the UK Biobank were split into a 60:20:20 training, validation, and testing split. The training data was utilized to learn the model parameters, the validation data was employed to optimize the hyperparameters (see below), and the test data was used to evaluate the phenotype prediction performance of the model. The same train-validation-test split was used for all models.

Parameter optimization

The neural network f consisted of a variable number of layers. Softplus activation functions were applied between all layers to introduce non-linearity and after the last layer to ensure positive output values. We optimized the model using the Adam optimizer [31], with a learning rate of 10^{-3}. Training was performed for 25 epochs, with a batch size of 16,384 samples. The model performance was monitored on a validation dataset, and the parameters at the epoch with the highest validation R^2 were saved.

Hyperparameter optimization

The best model architecture for each trait was determined through hyperparameter optimization. Five trials were conducted for each combination of hyperparameters, and the second lowest validation R^2 was selected as a metric for the performance of that particular combination. We found selected the second lowest validation R^2 led to more robust results than selecting the highest R^2. Hyperparameter tuning was accomplished using a TPE (Tree-structured Parzen Estimator) algorithm [32], implemented via the Optuna Python package [33], to optimize the number of hidden layers : [1, 4], size of hidden layers : [2, 256], L_1 regularization strength, α : [10^{-10}, 10], activation function used : {Softplus, ReLU}, and the initialization of the bias term of the last hidden layer : [-12,-4].

Last-layer bias

The bias term of the last hidden layer acts as the variance of the prior. A larger last-layer bias would allow more genes to have non-zero prior variances and, therefore a higher variance of the prior. While the last layer bias is a gradient-optimized parameter, we observed that the initialization of this parameter is crucial for model convergence. Hence, we included the initialization of this parameter as a hyperparameter.

Phenotype prediction

The phenotype predicted by the model was obtained from Eq. 1. Given the genotype matrix $\mathbf{G}$, the gene effect vector $\boldsymbol{\beta}$, the covariate matrix $\mathbf{C}$, and the coefficient vector $\boldsymbol{\gamma}$, we can compute the predicted phenotype. $\mathbf{G}$ and $\mathbf{C}$ are obtained from the data, while $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are learnt during model training. We used the mean of the posterior distribution $\boldsymbol{\beta}|\boldsymbol{y}$, computed on the training and validation data, as the $\boldsymbol{\beta}$ in Eq.1, to obtain the model predictions on the test data.

Posterior of gene effects

Following optimization of the prior, the posterior is computed by incorporating both training and validation data using Eq. 6. This approach maximizes the number of samples utilized in the computation, ensuring the robustness of the estimates. Consequently, the direction of the gene effect is inferred from the data, with the prior serving as a regularizer. Moreover, the mean posterior gene effect for each gene is determined, providing a summary of the estimated effects.

To rank gene associations, we compute the probability of direction (p_d) [34] on the posterior distribution of a gene effect. The probability of direction indicates the certainty in the sign of an effect and lies between 0.5 and 1. Given a posterior probability p_{post} on a variable $\boldsymbol{\beta}$, p_d is given by:

$$p_d = \max(p_{\text{post}}(\beta > 0), p_{\text{post}}(\beta < 0))$$

Alternative models

We benchmark FuncRVP against a burden test and a Bayesian model with constant regularization for gene effect estimation, and a linear model using burden test significant genes for phenotype prediction.

Burden test

Burden tests aggregate variant effects in a region, usually a gene. These aggregated effects are known as the gene burdens and these are used to test association of the gene with a trait [8]. For an individual *i* and gene *j*, we perform the burden test using a linear regression with the covariates:

$$y_i = g_{i,j}\beta_j + \boldsymbol{c}_i^\top \boldsymbol{\gamma} + b$$

Where $g_{i,j}$ is the DeepRVAT gene burden score of gene *j* of individual *i*, $\mathbf{c}_i$ is the vector of covariates (including the PRS) for individual *i*, and b is the intercept.

Phenotype prediction linear model (LM)

The LM uses genes found to be significantly associated with the phenotype using a burden test as features for phenotype prediction. After running the burden test across all genes, we select the genes significantly associated with the phenotype using Bonferroni significant p-values at 0.05 (multiple testing correction performed per trait). We then use only these significant genes to train a linear model for phenotype prediction:

$$y_i = \sum_{j \epsilon s_g} (g_{i,j}\beta_j) + \boldsymbol{c}_i^\top \boldsymbol{\gamma} + b$$

where s_g is the set of significant genes for the phenotype.

Constant regularization

To delineate the effect of the embedding and the model strategy, we train a Bayesian model in the same way as FuncRVP, with the exception that the neural network f is replaced by a constant. The result is a constant regularization applied to all genes instead of the gene-specific regularization of FuncRVP. The training and evaluation procedure is identical to FuncRVP.

Model evaluation

Comparison of phenotype prediction

FuncRVP predicts the phenotype given an individual's genotype. We trained independent models for each phenotype and evaluated them on a held-out test set by comparing phenotype prediction R^2 against the baseline models. Predictions on the test set were computed using the posterior of the gene effects obtained after the training procedure. To evaluate the phenotype prediction improvements compared to the common variant baseline phenotype predictor, we use the relative improvement in explained variance as our metric. The common variant baseline phenotype predictor is given by:

$$y_i = \boldsymbol{c}_i \boldsymbol{\gamma} + b$$

where $\mathbf{c}_i$ is the vector of covariates (including the PRS) for individual *i*, and b is the intercept. The R^2 of the predictions of this model on the test set are referred to as R^2_{common}. The final metric $relative\ \Delta R^2$ is given by:,

$$relative\ \Delta R^2 = \frac{R^2_{model} - R^2_{common}}{R^2_{common}}$$

where R^2_{model} is the prediction R-squared of FuncRVP or the LM.

We conducted a bootstrapping to identify traits where the two models perform significantly differently. Specifically, we drew 2,000 bootstraps from the test set and computed the phenotype prediction R^2. The compute the p-value of these comparisons as follows [35]:

$$P = 2\min(\frac{N_{\text{greater}}+1}{N_{\text{total}}+1}, \frac{N_{\text{lesser}}+1}{N_{\text{total}}+1})$$

where N_{greater} is the number of bootstraps where FuncRVP performed better, N_{lesser} is the number of bootstraps where FuncRVP performed worse, and N_{total} is the total number of bootstraps drawn which is always 2,000. We consider $P < 0.05$ as the significance threshold.

Comparison of gene effects

We evaluate the predicted gene effects by comparing them against the gene effects observed on the 209,492 samples of the UK Biobank that our model has not seen during training, validation, or testing. We call this the validation cohort. This provides an opportunity to benchmark the results on a larger and independent cohort. We perform a burden test on 209,492 samples using DeepRVAT burdens to obtain the gene effects. We then compare these to the gene effects obtained by FuncRVP and the linear regression on the 161,850 samples. The *MAD* is computed for associations obtained from all models up to rank 1,000 .

Trait-wise jackknife

At rank N, the MAD is computed for the top N association excluding one trait. The procedure is repeated for all traits, resulting in 41 MADs per rank. The 95% interval is obtained as the [2.5th percentile, 97.5th percentile].

Computing resources

The models were trained on NVIDIA A40 GPUs. Model training was capped at 25 epochs as more epochs led to minimal performance improvements on the validation set. The training process took a little more than 6 minutes per trait on one NVIDIA A40.

Results

The FuncRVP prior regularizes gene effects using functional gene embeddings

FuncRVP is a Bayesian model that predicts quantitative phenotypes from genotype and leverages a functional gene embedding to borrow information across genes (Figure 1). Building on DeepRVAT [11], FuncRVP models phenotype as a weighted sum of precomputed gene impairment scores and covariates including common-variant PRS. In contrast to DeepRVAT, no statistical testing is applied to select associated genes. Instead, the weighted sum is computed across all genes. This is robustly achieved by regularizing the weights through a trait-specific prior. The prior depends nonlinearly on a functional gene embedding. We made Gaussian assumptions to model the prior and the phenotype leading to a closed-form solution of the posterior of the phenotype. The model was implemented in the deep learning framework Pytorch and fitted by maximizing the posterior of the phenotype. Hyperparameter selection for the neural network used to model the prior and model evaluation were achieved through a train-validation-test split including 60%, 20%, and 20% of the individuals respectively (Methods).

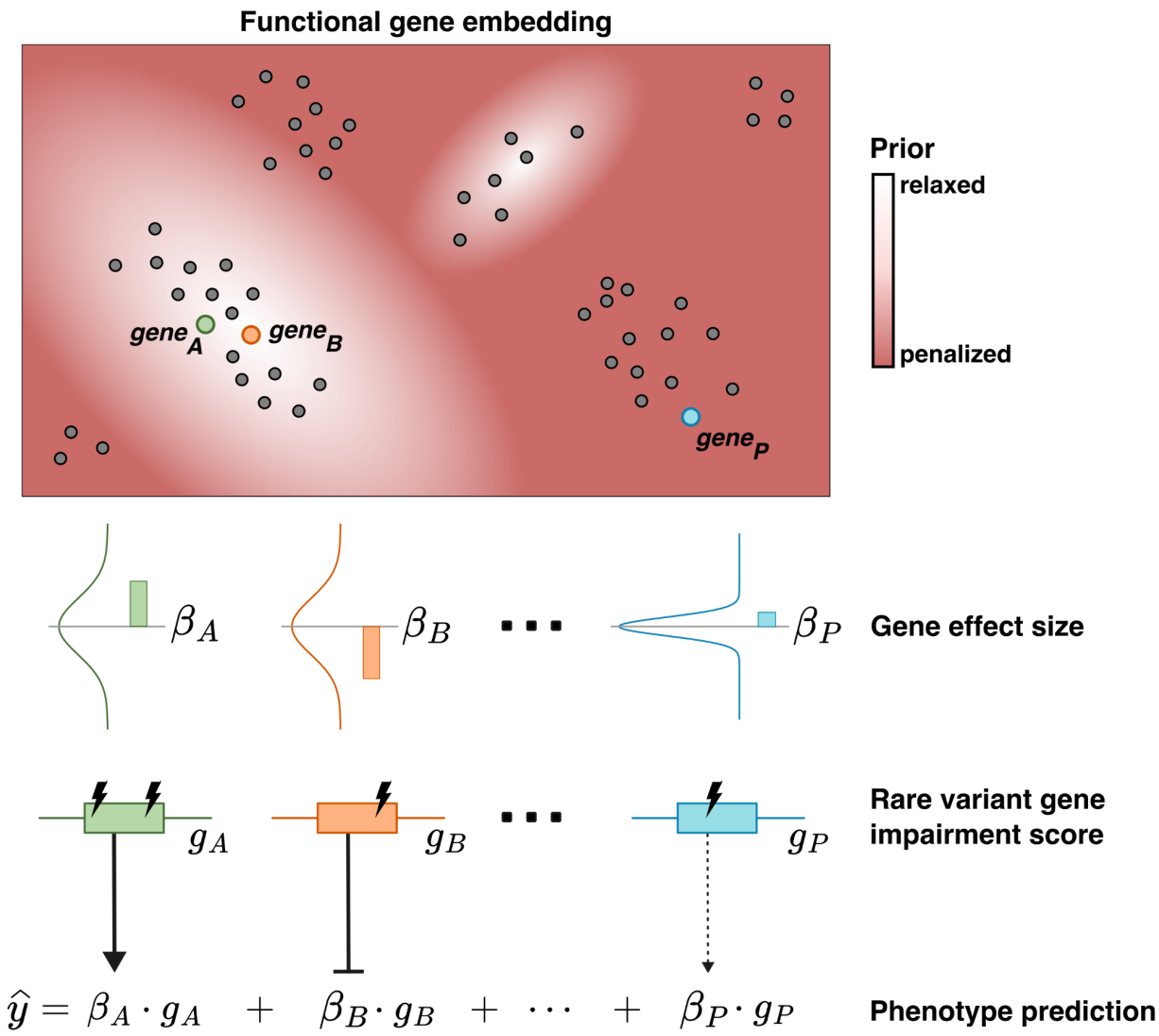

Figure 1. FuncRVP model overview. FuncRVP is a Bayesian model that predicts phenotypes from rare variants. An individual's phenotype is modeled as a weighted sum of their rare variant gene impairment scores (bottom). A prior, learnt from data, constrains the magnitude of gene effects depending on their location in a functional gene embedding, resulting in relaxed regularizations on areas densely populated with trait-associated genes (e.g. genes A-B).

We trained the model on 41 quantitative phenotypes for 161,850 unrelated individuals with European ancestry in the UK Biobank who underwent whole-exome sequencing (Bycroft et al. 2018). We used the DeepRVAT gene impairment scores precomputed on the UK Biobank [11]. As functional gene embeddings we considered the concatenation of two best-performing embeddings investigated by Brechtmann and colleagues [18]. The so-called “omics” embedding is derived from gene expression, genome-wide CRISPR screens and a protein language model. The other embedding is a low-dimensional representation of the gene feature matrix from the

PoPS study [21], which retained the same performance as the original one across a range of benchmarks [18].

We assessed two alternative structures for the prior. On the one hand, we assumed the expectation of the weights to depend on the embedding. On the other hand, we assumed the variance of the weights to depend on the embedding. Across all traits, we found that modeling the variance consistently led to substantially more accurate phenotype predictors on held-out data (Figure 2A). These results indicate that the position in the embedding was predictive of whether genes affect a phenotype but not of the value of the effect. One example illustrating this phenomenon is shown in Figure 2B, which shows the distribution of HDL cholesterol levels among individuals with a putative loss of function variant for the Cholesteryl ester transfer protein *CETP* and its 24 nearest genes in the functional embedding. Like *CETP*, 4 out of the 24 nearest neighbors have been reported to be among the 42 significantly associated genes with the phenotype [24] (Enrichment 94-fold, $P < 2.7 \times 10^{-7}$, Fisher test). However, the effects of these genes vary widely in amplitude and direction.

Moreover, we evaluated multiple ablations and combinations of gene embeddings proposed by Brechtmann and colleagues [18], and found the combination of Omics and PoPS to have the best phenotype prediction performance (Figure S1). In particular, using a regularization independent of any gene embedding, equivalent to an L_2 regularization on the gene effects, performed poorly, showing that it is the information on how genes functionally relate rather than a mere regularization that explains the model performance (Figure S2A). Furthermore, we found that using the number of putative loss of function of variants as gene impairment score as opposed to the DeepRVAT score, led to drastic loss of performance (Figure 2C). Consequently, we went on with DeepRVAT as the gene impairment score and modeling the variance with the omics-PoPS embeddings in all further analyses.

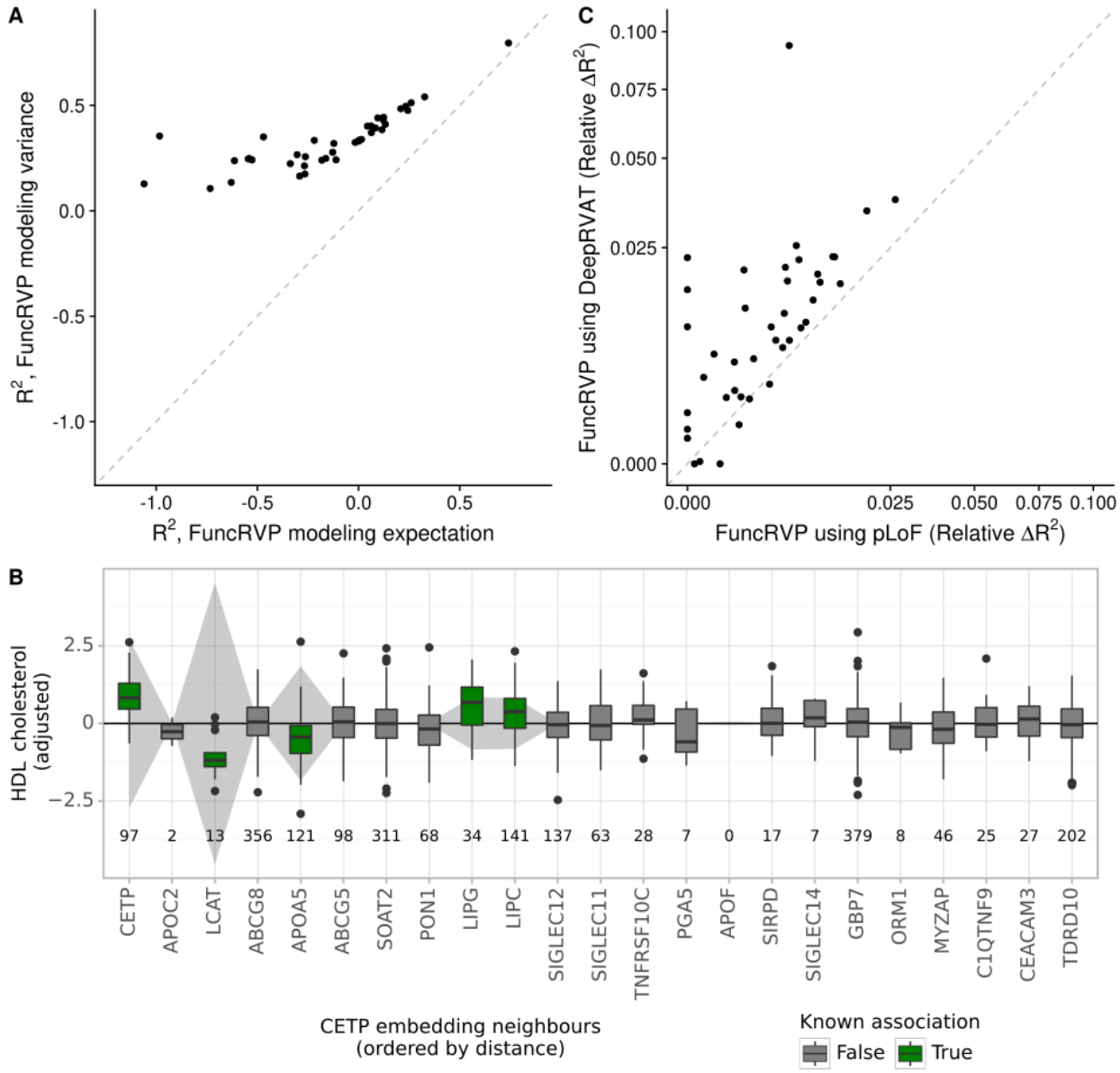

Figure 2. Assessing FuncRVP model assumptions A) Explained variance on held-out data using a prior on the gene effect variance (y-axis) against a prior on gene effect expectations (x-axis) across 41 traits. **B)** Distribution of HDL cholesterol values (adjusted for common variants and covariates) among individuals with a putative loss-of-function (pLoF) variant in a given gene. The gene *CETP* and its nearest neighbors in the embedding space are ordered by increasing distance. Center line, median; box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles. **C)** Similar to A, comparing FuncRVP using DeepRVAT gene impairment scores (y-axis) against FuncRVP using the number of putative loss of function variants [1] (x-axis).

FuncRVP improves phenotype prediction

Having established the model features and structure, we next assessed the added value of FuncRVP for phenotype prediction. As a state-of-the-art comparison point, we used linear models on DeepRVAT gene impairment scores of significantly associated genes, as described earlier [11]. We found significant improvements (bootstrapping: $P < 0.05$, Methods) in phenotype prediction for 11 out of 41 traits and significantly worse on 2 traits (Figure 3A). The traits for which FuncRVP improved also showed greater burden heritability, which is an estimated fraction of phenotypic variance explained by minor allele burden [30], a larger number of genes previously found to be associated by RVATs [24,28], and larger number of significant GWAS loci (Methods) (Figure 3B-D). Collectively these results show that FuncRVP can provide improved phenotypic predictors. However, the model best improves for traits for which the genetic basis is stronger, presumably because this helps learning the prior function.

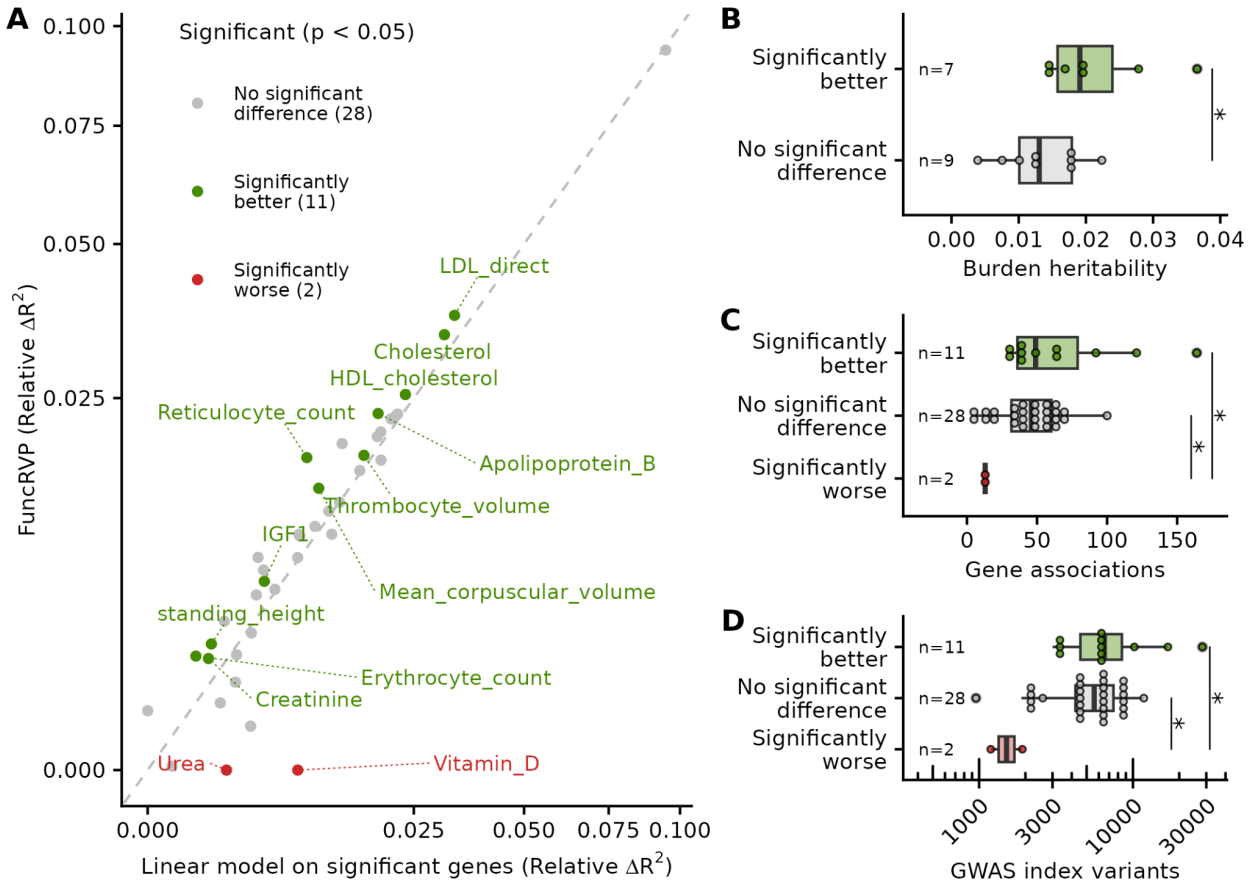

Figure 3. FuncRVP improves phenotype prediction, particularly for phenotypes with a stronger genetic basis. A) Relative improvement in phenotype prediction of FuncRVP (y-axis) compared to a linear regression on gene impairment scores of significant genes (x-axis) for 41 phenotypes. The relative fit improvement was computed as the improvement in R^2 relative to a baseline linear model using only the PRS and covariates (Methods). Significance was determined by case bootstrapping (Methods). **B)** Burden heritability scores for 16 traits [30] between traits that performed significantly better with FuncRVP and those

where FuncRVP showed no significant improvement. **C)** Comparison of the number of gene-trait associations found for each trait by previous RVASs [24,28] across three categories: traits where FuncRVP performed significantly better, traits where there was no significant difference, and traits where FuncRVP performed significantly worse. **D)** As in C using the number of GWAS index variants associated with each trait across the three categories. For B-D) Center line, median; box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles. Significance was determined with the one-sided Wilcoxon test.

Better estimation of gene effects using FuncRVP

Having shown that FuncRVP improves phenotype predictions we next asked whether it helped provide robust estimates of gene effects on phenotypes compared to the burden test approach as implemented by DeepRVAT. This is difficult because the two approaches use different statistical paradigms: Bayesian for FuncRVP and frequentist for burden testing. We defined as our discovery cohort the train and validation splits form the 161,850 unrelated individuals we considered so far. Moreover, we defined as our evaluation cohort a set of 209,492 further unrelated samples from the UK Biobank on which we estimated gene effects by standard, non-regularized, linear regression. The models were benchmarked for giving accurate estimations of those evaluation cohort effects when fitted on the discovery cohort (Figure 4A). Burden-test associations were ranked by increasing p-values whereas FuncRVP associations were ranked by decreasing probability of direction, a Bayesian metric that quantifies the certainty about the sign of an effect (reviewed in [36]) (Methods). We emphasize that here the effects are not estimated the same way for FuncRVP in the discovery cohort and in the evaluation cohort, unlike in classical replication analyses. This is intended to not bias the benchmark toward FuncRVP and to assess the benefit of its functional embedding based regularization.

Firstly, we assessed the agreement in effect direction estimated by FuncRVP in the discovery cohort compared to those obtained by linear regression in the evaluation cohort. Generally, the probability of direction overestimated the empirical sign agreements. This overestimation could be either due to model miscalibration, or to noise in the empirical effect estimations in the evaluation cohort, or a combination of both. Nonetheless, the agreement in direction increased as the probability of direction increased, showing that ranking based on the probability of direction is informative about the robustness of the estimates. Moreover, the directions of effects estimated by FuncRVP with a probability of direction larger than 0.999 perfectly agreed with those in the evaluation cohort for almost all traits (Figure 4B). Consequently, we considered a probability of direction of 0.999 as the minimum threshold for a FuncRVP association to be identified as discovered. Across the 41 traits, this cutoff yielded on average 12.2 associations per trait (+/- 7.8 standard deviation).

Next, we evaluated the generalizability of two models by comparing the gene effects estimated in the discovery cohort with those in the evaluation cohort. Linear regression identified 258 trait-wise significant associations in the discovery cohort (Bonferroni-adjusted $P < 0.05$), whereas FuncRVP identified 486 associations (probability of direction > 0.999). The deviation in gene effect estimates for the top 258 and 486 associations of both models was compared using the median absolute deviation (MAD). In both cases, FuncRVP gene effect estimates had a significantly lower MAD compared to those from linear regression (Figure 4C), and the improvement was significant independently of association rank cutoff (Figure 4D). Altogether, these results demonstrate that FuncRVP improves effect estimations over burden-test based approaches.

Genes with low tolerance to loss function variants [1] are an interesting class of genes because they are more likely to have a functional importance. However, as they are, by definition, depleted for high-impact variants, these genes particularly lack power for rare variant association testing. We found that FuncRVP-specific associations involved genes less tolerant to loss of function

variants (median=0.8 versus 0.91 genome-wide, Figure 4E). For comparison, Medelian genes reported for the same traits [4] had a median LOEUF score of 0.6. In contrast, the LOEUF scores of the discoveries shared between FuncRVP and the burden test as well as of the discoveries exclusive to the burden test were not significantly lower. These results indicate that the Bayesian regularization implemented by FuncRVP is particularly helpful at finding association with constrained genes.

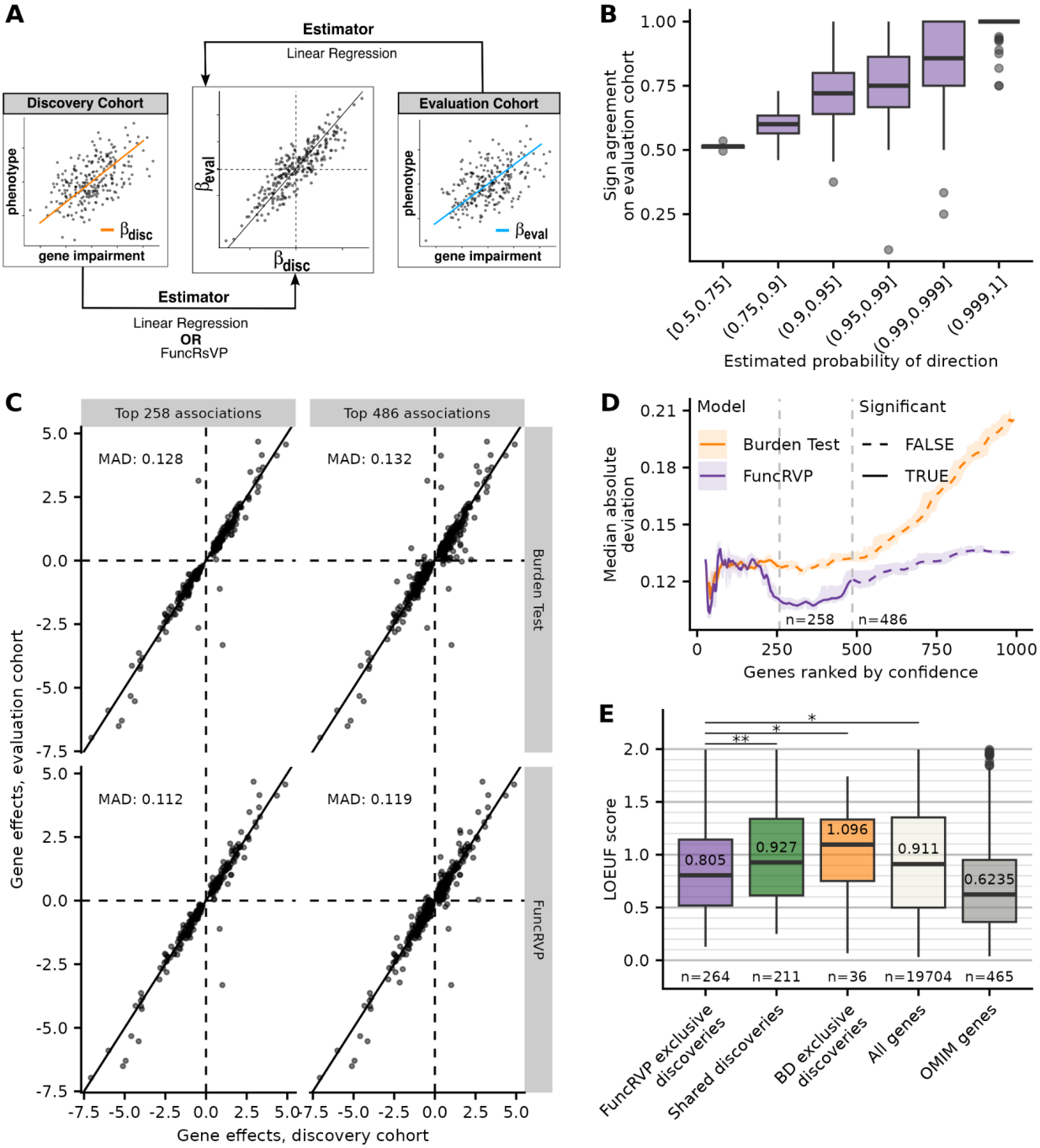

Figure 4. Comparing gene effect estimates on a validation cohort shows the robustness of FuncRVP. A) Schematic representing the strategy used for assessing the gene effect estimates of FuncRVP and a linear regression. Gene effect estimates of both models on the discovery cohort are compared against those obtained on the evaluation cohort using linear regression. **B)** Distribution across 41 traits of the proportion of sign agreements between FuncRVP gene effects estimated on the discovery cohort and gene effects estimated by linear regression on the evaluation cohort. **C)** Gene effect estimates on the evaluation cohort (y-axis) against discovery cohort (x-axis). Associations up to the discovery thresholds of both methods were compared (258 trait-wise Bonferroni significant associations, $P < 0.05$, found by linear regression, and 486 associations with probability of direction > 0.999 found by FuncRVP). The median absolute deviation is indicated for each scatter plot. **D)** Median absolute deviation of predicted gene effects (y-axis) up to a given rank (x-axis) across all 41 traits. Genes are ranked by p-value for the burden test and probability of direction for FuncRVP. The ribbon represents a 95% confidence interval created by jackknife across traits. **E)** Distribution of LOEUF score for five gene sets: exclusive to FuncRVP, exclusive to burden tests performed on DeepRVAT burdens (BD), shared discoveries, all coding genes, and trait-relevant OMIM genes (Methods). One-sided Wilcoxon tests were performed across these categories; the boxes significantly higher than FuncRVP exclusive discoveries are indicated. For B,E): center line, median; box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles.

Novel associations identified by FuncRVP

Out of the FuncRVP gene-trait associations, 88 had not been reported by published RVAS performed on the complete UK Biobank with about ~500,000 samples [4,24,28] (Table S4).

We found the association between *ARHGAP15* (Rho GTPase Activating Protein 15) and lymphocyte percentage particularly interesting. It was reported as the top novel discovery associated with the trait by FuncRVP, although no direct lymphocyte-related function of *ARHGAP15* has been reported, studies on *ARHGAP15* knockout mice demonstrated its role as a temporal signaling regulator of macrophage-1 antigen (Mac-1) affinity during sterile inflammation [37]. Single-cell data from the Human Protein Atlas [38] indicate the highest expression of *ARHGAP15* in T-lymphocytes and, to a lesser extent, B-lymphocytes, in the majority of the tissue or organ types (Methods). Additionally, *ARHGAP15* was found to be associated with lymphocyte count and percentage by GWAS [21,39,40]. FuncRVP only controls for common variants via the genetic principal components and the polygenic risk scores, leaving the possibility that this gene discovery is confounded by local common variants. However, the FuncRVP effect estimate replicated with effects estimated on the evaluation cohort while conditioning on nearby trait-associated common variants (Figure 5A, Methods).

Another interesting novel association was *RHOF* (Rho Family GTPase Rif) with mean platelet volume. The small GTPase Rif is dispensable for platelet filopodia generation in mice [41]. In addition, *RHOF* has the highest expression in platelets among blood cells (Methods). *RHOF* was also identified as the strongest association for mean platelet volume by GWAS [40,42]. Here too the FuncRVP's predicted gene effect was confirmed on evaluation data correcting for associated local common variants (Figure 5B).

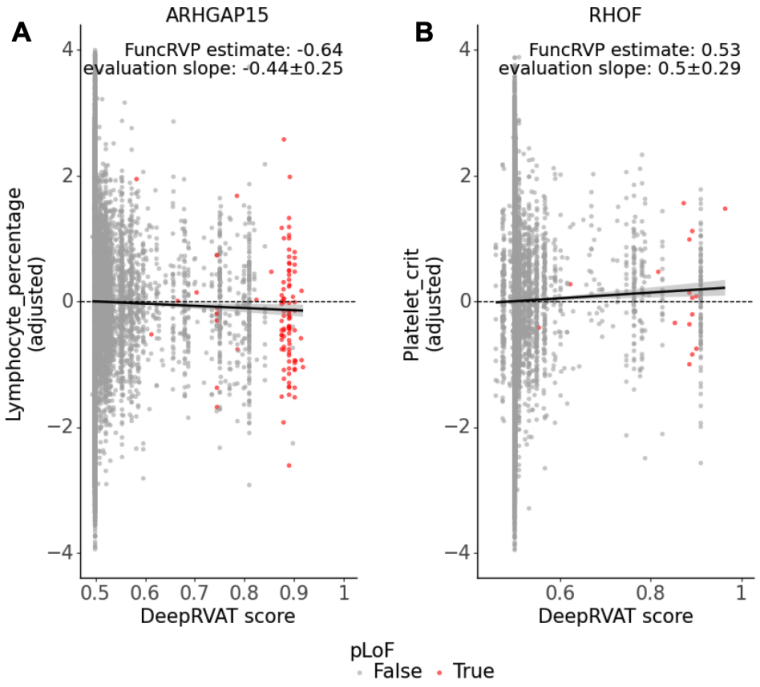

Figure 5. FuncRVP captures associations missed by other methods. A) Lymphocyte percentage z-score (adjusting for common variants and covariates) of individuals of the UK Biobank in the evaluation cohort (y-axis) against the DeepRVAT gene impairment score of *ARHGAP15* (x-axis). Individuals with pLoF variants in the gene are highlighted in red. The trend line is plotted using a linear regression on the evaluation cohort. The "FuncRVP estimate" indicates the gene effect estimated by FuncRVP and the "evaluation slope" is the slope of the linear regression with a 95% confidence interval computed on the evaluation cohort. The p-values for the slope is 5.9×10^{-4}. **B)** same as A) for the gene *RHOF* and the adjusted mean platelet volume (Platelet_crit). The p-values for the slope is 8.2×10^{-4}. The p-values were computed using the Wald Test with t-distribution of the test statistic.

Discussion

We have introduced FuncRVP, a novel rare variant model for phenotype prediction and discovering gene-trait associations. By incorporating gene function information through embeddings, FuncRVP improves upon state-of-the-art burden tests (DeepRVAT) in phenotype prediction and yields more robust gene effect estimates. Investigations of modeling strategies showed that DeepRVAT gene impairment scores strongly outperformed loss-of-function variant counts. Moreover, effective weight regularizations were obtained when using a prior on gene weight variance, but not on gene weight expectation. Our analysis of 41 quantitative traits revealed 88 novel associations including *ARHGAP15* for lymphocyte percentage and *RHOF* for mean platelet volume.

While FuncRVP outperforms the conventional linear regression-based methods, it has its limitations. First, it currently models both the prior and the phenotype using a normal distribution. This approach is restrictive when dealing with disease traits, which are typically categorical and cannot be modeled as a normal distribution. These non-conjugate distributions do not have a closed-form posterior and extending FuncRVP to handle such traits will require the development of numerical or iterative approaches. Although FuncRVP incorporates common variant contributions through principal components and polygenic risk scores, it cannot directly control for the independently associated common variants surrounding individual genes. LD-based clumping to obtain the most significant variants in a gene's vicinity and controlling for these variants has to be performed post-facto. Interestingly, in our investigation of FuncRVP's novel associations, controlling for these variants did not invalidate any of the findings. Additionally, while the overall improvements in phenotype prediction were modest, this aligns with the understanding that the phenotype prediction R^2 is largely driven by common variants [43,44].

FuncRVP's ability to exploit information from all genes is limited when using sparse gene burden scores, such as binary annotations of loss-of-function (LoF) variants. In these cases, FuncRVP performs on par with current models. The development of new impairment score models like DeepRVAT, AlphaMissense, and PrimateAI-3D underscores the increasing availability of informative scores, making sparse burden scores a less favorable option in most cases [4,11,45]. With the growth in the number and size of large-scale genome sequencing studies, along with improved rare variant annotations, FuncRVP's ability to detect subtle genetic effects will be further enhanced, leading to the identification of a broader spectrum of genes contributing to human diseases.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We thank Brian Clarke, Marcel Mück, and Magnus Wahlberg for the computation of DeepRVAT gene impairment scores. S.L. thanks Vicente Yépez for feedback and advice.

This research has been conducted using data from UK Biobank, a major biomedical database (project IDs 25214 and 81358). This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via the project NFDI 1/1 "GHGA - German Human Genome-Phenome Archive" (#441914366 to S.L., E.H. and J.G.), by the German Bundesministerium für Bildung und Forschung (BMBF) through the ERA PerMed project PerMiM (01KU2016B to S.L., J.L., F.B. and J.G.) and through the Model Exchange for Regulatory Genomics project MERGE (031L0174A to F.R.H. and J.G.). This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the TRR267 (403584255, sub-projects, Z03 to J.G. and B05 to Z.C.). This work was funded via the EVUK programme ("Next-generation AI for Integrated Diagnostics") of the Free State of Bavaria (to F.B. and J.G.). This work was funded by the Initiative and Networking Fund of the Helmholtz Association in the framework of the Helmholtz AI project call (project DeepVar, ZT-I-PF-5-156). S.L. and E.H. are supported by the Helmholtz Association under the joint research school 'Munich School for Data Science – MUDS'. F.P.C. was funded by the Free State of Bavaria's Hightech Agenda through the Institute of AI for Health (AIH).

Statement: During the preparation of this work the author(s) used ChatGPT in order to polish the manuscript and to fill this statement. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Contributions

F.B. and J.G. conceptualized the project. F.B., F.P.C., and J.G. designed the methodology. J.L. and S.L. implemented the methods and analyzed the data. F.R.H., E.H., J.L., and Z.C. curated the data. S.L. and J.G. wrote the original draft of the manuscript. S.L. performed visualizations. J.G. supervised the project.

Corresponding author

Correspondence to Julien Gagneur and Felix Brechtmann.

Data and code availability

The code is made available on github: https://github.com/gagneurlab/funcRVP

References

1. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
2. Abifadel, M., Varret, M., Rabès, J.-P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat. Genet. *34*, 154–156. https://doi.org/10.1038/ng1161.
3. Cohen Jonathan C., Boerwinkle Eric, Mosley Thomas H., and Hobbs Helen H. (2006). Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease. N. Engl. J. Med. *354*, 1264–1272. https://doi.org/10.1056/NEJMoa054013.
4. Fiziev, P.P., McRae, J., Ulirsch, J.C., Dron, J.S., Hamp, T., Yang, Y., Wainschtein, P., Ni, Z., Schraiber, J.G., Gao, H., et al. (2023). Rare penetrant mutations confer severe risk of common diseases. Science *380*, eabo1131. https://doi.org/10.1126/science.abo1131.
5. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am. J. Hum. Genet. *95*, 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009.
6. Li, B., and Leal, S.M. (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. Am. J. Hum. Genet. *83*, 311–321. https://doi.org/10.1016/j.ajhg.2008.06.024.
7. Madsen, B.E., and Browning, S.R. (2009). A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. PLOS Genet. *5*, e1000384. https://doi.org/10.1371/journal.pgen.1000384.
8. Morgenthaler, S., and Thilly, W.G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). Mutat. Res. Mol. Mech. Mutagen. *615*, 28–56. https://doi.org/10.1016/j.mrfmmm.2006.09.003.
9. Povysil, G., Petrovski, S., Hostyk, J., Aggarwal, V., Allen, A.S., and Goldstein, D.B. (2019). Rare-variant collapsing analyses for complex traits: guidelines and applications. Nat. Rev. Genet. *20*, 747–759. https://doi.org/10.1038/s41576-019-0177-4.
10. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. Am. J. Hum. Genet. *89*, 82–93. https://doi.org/10.1016/j.ajhg.2011.05.029.
11. Clarke, B., Holtkamp, E., Öztürk, H., Mück, M., Wahlberg, M., Meyer, K., Munzlinger, F., Brechtmann, F., Hölzlwimmer, F.R., Gagneur, J., et al. (2023). Integration of variant annotations using deep set networks boosts rare variant association genetics. Preprint at bioRxiv, https://doi.org/10.1101/2023.07.12.548506 https://doi.org/10.1101/2023.07.12.548506.
12. Li, X., Li, Z., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett, D.K., Aslibekyan, S., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. Nat. Genet. *52*, 969–983. https://doi.org/10.1038/s41588-020-0676-4.
13. Elmarakeby, H.A., Hwang, J., Arafeh, R., Crowdis, J., Gang, S., Liu, D., AlDubayan, S.H., Salari, K., Kregel, S., Richter, C., et al. (2021). Biologically informed deep neural network for prostate cancer discovery. Nature *598*, 348–352. https://doi.org/10.1038/s41586-021-03922-4.
14. van Hilten, A., Kushner, S.A., Kayser, M., Ikram, M.A., Adams, H.H.H., Klaver, C.C.W., Niessen, W.J., and Roshchupkin, G.V. (2021). GenNet framework: interpretable deep learning for predicting phenotypes from genetic data. Commun. Biol. *4*, 1–9. https://doi.org/10.1038/s42003-021-02622-z.

15. Schlüter, A., Vélez-Santamaría, V., Verdura, E., Rodríguez-Palmero, A., Ruiz, M., Fourcade, S., Planas-Serra, L., Launay, N., Guilera, C., Martínez, J.J., et al. (2023). ClinPrior: an algorithm for diagnosis and novel gene discovery by network-based prioritization. Genome Med. *15*, 68. https://doi.org/10.1186/s13073-023-01214-2.
16. Zheng, Z., Liu, S., Sidorenko, J., Wang, Y., Lin, T., Yengo, L., Turley, P., Ani, A., Wang, R., Nolte, I.M., et al. (2024). Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. Nat. Genet. *56*, 767–777. https://doi.org/10.1038/s41588-024-01704-y.
17. Richardson, T.G., Timpson, N.J., Campbell, C., and Gaunt, T.R. (2017). A pathway-centric approach to rare variant association analysis. Eur. J. Hum. Genet. *25*, 123–129. https://doi.org/10.1038/ejhg.2016.113.
18. Brechtmann, F., Bechtler, T., Londhe, S., Mertes, C., and Gagneur, J. (2023). Evaluation of input data modality choices on functional gene embeddings. NAR Genomics Bioinforma. *5*, lqad095. https://doi.org/10.1093/nargab/lqad095.
19. Du, J., Jia, P., Dai, Y., Tao, C., Zhao, Z., and Zhi, D. (2019). Gene2vec: distributed representation of genes based on co-expression. BMC Genomics *20*, 82. https://doi.org/10.1186/s12864-018-5370-x.
20. Theodoris, C.V., Xiao, L., Chopra, A., Chaffin, M.D., Al Sayed, Z.R., Hill, M.C., Mantineo, H., Brydon, E.M., Zeng, Z., Liu, X.S., et al. (2023). Transfer learning enables predictions in network biology. Nature *618*, 616–624. https://doi.org/10.1038/s41586-023-06139-9.
21. Weeks, E.M., Ulirsch, J.C., Cheng, N.Y., Trippe, B.L., Fine, R.S., Miao, J., Patwardhan, T.A., Kanai, M., Nasser, J., Fulco, C.P., et al. (2023). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. Nat. Genet. *55*, 1267–1276. https://doi.org/10.1038/s41588-023-01443-6.
22. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.
23. Hanscombe, K.B., Coleman, J.R.I., Traylor, M., and Lewis, C.M. (2019). ukbtools: An R package to manage and query UK Biobank data. PLOS ONE *14*, e0214311. https://doi.org/10.1371/journal.pone.0214311.
24. Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G., Lu, W., Riley-Gillis, B.M., Tsai, E.A., Kim, H.I., Zheng, X., et al. (2022). Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. Cell Genomics *2*, 100168. https://doi.org/10.1016/j.xgen.2022.100168.
25. Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat. Genet. *53*, 420–425. https://doi.org/10.1038/s41588-021-00783-5.
26. Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O'Reilly, P.F., and Vilhjálmsson, B.J. (2022). Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am. J. Hum. Genet. *109*, 12–23. https://doi.org/10.1016/j.ajhg.2021.11.008.
27. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8.
28. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. Nature, 1–7. https://doi.org/10.1038/s41586-021-04103-z.
29. Karczewski, K.J., Gupta, R., Kanai, M., Lu, W., Tsuo, K., Wang, Y., Walters, R.K., Turley, P., Callier, S., Baya, N., et al. (2024). Pan-UK Biobank GWAS improves discovery, analysis

of genetic architecture, and resolution into ancestry-enriched effects. Preprint at medRxiv, https://doi.org/10.1101/2024.03.13.24303864 https://doi.org/10.1101/2024.03.13.24303864.
30. Weiner, D.J., Nadig, A., Jagadeesh, K.A., Dey, K.K., Neale, B.M., Robinson, E.B., Karczewski, K.J., and O'Connor, L.J. (2023). Polygenic architecture of rare coding variation across 394,783 exomes. Nature *614*, 492–499. https://doi.org/10.1038/s41586-022-05684-z.
31. Kingma, D.P., and Ba, J. (2017). Adam: A Method for Stochastic Optimization. Preprint at arXiv, https://doi.org/10.48550/arXiv.1412.6980 https://doi.org/10.48550/arXiv.1412.6980.
32. Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. In Advances in Neural Information Processing Systems, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, eds. (Curran Associates, Inc.).
33. Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining KDD '19. (Association for Computing Machinery), pp. 2623–2631. https://doi.org/10.1145/3292500.3330701.
34. Makowski, D., Ben-Shachar, M.S., and Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. J. Open Source Softw. *4*, 1541. https://doi.org/10.21105/joss.01541.
35. Davison, A.C., and Hinkley, D.V. (1997). Bootstrap Methods and their Application (Cambridge University Press) https://doi.org/10.1017/CBO9780511802843.
36. Makowski, D., Ben-Shachar, M.S., Chen, S.H.A., and Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. Front. Psychol. *10*. https://doi.org/10.3389/fpsyg.2019.02767.
37. Margraf, A., Cappenberg, A., Vadillo, E., Ludwig, N., Thomas, K., Körner, K., Zondler, L., Rossaint, J., Germena, G., Hirsch, E., et al. (2020). ArhGAP15, a RacGAP, Acts as a Temporal Signaling Regulator of Mac-1 Affinity in Sterile Inflammation. J. Immunol. *205*, 1365–1375. https://doi.org/10.4049/jimmunol.2000047.
38. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single–cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169. https://doi.org/10.1126/sciadv.abh2169.
39. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D., et al. (2020). Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell *182*, 1198-1213.e14. https://doi.org/10.1016/j.cell.2020.06.045.
40. Vuckovic, D., Bao, E.L., Akbari, P., Lareau, C.A., Mousas, A., Jiang, T., Chen, M.-H., Raffield, L.M., Tardaguila, M., Huffman, J.E., et al. (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell *182*, 1214-1231.e11. https://doi.org/10.1016/j.cell.2020.08.008.
41. Goggs, R., Savage, J.S., Mellor, H., and Poole, A.W. (2013). The Small GTPase Rif Is Dispensable for Platelet Filopodia Generation in Mice. PLOS ONE *8*, e54663. https://doi.org/10.1371/journal.pone.0054663.
42. Jurgens, S.J., Pirruccello, J.P., Choi, S.H., Morrill, V.N., Chaffin, M., Lubitz, S.A., Lunetta, K.L., and Ellinor, P.T. (2023). Adjusting for common variant polygenic scores improves yield in rare variant association analyses. Nat. Genet. *55*, 544–548. https://doi.org/10.1038/s41588-023-01342-w.
43. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., Purcell (Leader), S.M., Stone, J.L., Sullivan, P.F., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752. https://doi.org/10.1038/nature08185.
44. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden,

P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569. https://doi.org/10.1038/ng.608.

45. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. https://doi.org/10.1126/science.adg7492.