SR-TWAS: Leveraging Multiple Reference Panels to Improve TWAS Power by Ensemble Machine Learning

Randy L. Parrish[1,2], Aron S. Buchman[3], Shinya Tasaki[3], Yanling Wang[3], Denis Avey[3], Jishu Xu[3], Philip L. De Jager[4], David A. Bennett[3], Michael P. Epstein[1], Jingjing Yang[1*]

1. Center for Computational and Quantitative Genetics, Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, 30322, USA
2. Department of Biostatistics, Emory University School of Public Health, Atlanta, GA 30322, USA
3. Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612, USA
4. Center for Translational and Computational Neuroimmunology, Department of Neurology and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY10032, USA

[*] Correspondence Author: jingjing.yang@emory.edu

Abstract

Multiple reference panels of a given tissue or multiple tissues often exist, and multiple regression methods could be used for training gene expression imputation models for TWAS. To leverage expression imputation models (i.e., base models) trained with multiple reference panels, regression methods, and tissues, we develop a Stacked Regression based TWAS (SR-TWAS) tool which can obtain optimal linear combinations of base models for a given validation transcriptomic dataset. Both simulation and real studies showed that SR-TWAS improved power, due to increased effective training sample sizes and borrowed strength across multiple regression methods and tissues. Leveraging base models across multiple reference panels, tissues, and regression methods, our real application studies identified 6 independent significant risk genes for Alzheimer's disease (AD) dementia for supplementary motor area tissue and 9 independent significant risk genes for Parkinson's disease (PD) for substantia nigra tissue. Relevant biological interpretations were found for these significant risk genes.

Introduction

Two-stage transcriptome-wide association studies (TWAS) have been widely used in genetic studies of complex traits due to the convenience of using publicly-available transcriptomic reference panels and summary-level genome-wide association study (GWAS) datasets[1–5]. The standard two-stage TWAS method[6,7] first trains gene expression imputation models (per gene per tissue) using a transcriptomic reference panel in Stage I, taking quantitative gene expression traits as response variables and nearby (cis-) or genome-wide (cis- and trans-) genetic variants as predictors. The non-zero genetic effect sizes estimated in the gene expression imputation models are considered effect sizes of a broad sense of expression quantitative trait loci (eQTL), which are taken as variant weights to conduct gene-based association tests with GWAS data (individual-level or summary-level) in Stage II.

Various TWAS techniques have been developed, employing diverse regression methods to train models for imputing gene expression. Additionally, multiple transcriptomic reference panels are made available to the public and could be used in TWAS. Consequently, it is possible to train multiple gene expression imputation models by employing distinct regression methods, employing multiple transcriptomic reference panels of the same tissue type, or utilizing transcriptomic data from multiple tissues within a given reference panel. For example, multiple regression methods, such as penalized regression with Elastic-Net penalty (used by PrediXcan[7,8]) and nonparametric Bayesian Dirichlet process regression (DPR) model (used by TIGAR[9]), have trained gene expression imputation models using the same Genotype-Tissue Expression (GTEx)[10] V8 reference data of 48 human tissue types. The Religious Orders Study (ROS)[11], Rush Memory and Aging Project (MAP)[11], and the GTEx[10] V8 project all profile transcriptomic data of prefrontal cortex (PFC) brain tissue and genome-wide genetic data of the same samples,

53 providing multiple reference panels of PFC tissue for TWAS. Thus, leveraging multiple trained
54 gene expression imputation models of the same target gene across multiple regression methods,
55 multiple reference panels, and multiple tissue types is expected to improve TWAS power, for
56 more robustly modeling the unknown genetic architecture of the target gene expression by
57 multiple regression models, having an increased training sample size with multiple reference
58 panels, or borrowing strength across multiple tissue types with correlated gene expression.

59 Multiple approaches that can take advantage of transcriptomic reference data for multiple tissues
60 and/or multiple reference panels have been developed. For example, UTMOST uses group
61 LASSO-penalized multivariate regression to impute cross-tissue expression[12]. SWAM estimates
62 a vector of weights for input expression imputation models such that the weighted average of the
63 input models will give the lowest mean squared error with respect to individual-level reference
64 expression of the target tissue[13]. However, these approaches have drawbacks such as requiring
65 individual-level reference data, being computationally expensive, and user-unfriendly. For
66 example, UTMOST requires individual-level reference data for all tissues to train gene
67 expression imputation models. In order to control for multicollinearity, SWAM considers a
68 regularization parameter which requires fine-tuning based on the covariance structure of
69 Genetically Regulated gene eXpression (GReX) of all considered tissues which must be derived
70 using individual-level reference transcriptomic data[13]. Additionally, SWAM requires that the
71 input of trained gene expression imputation models in the same SQL database format as used for
72 PrediXcan output[13].

73 To fill this gap, we develop a novel TWAS method to leverage multiple summary-level gene
74 expression imputation models (i.e., base models) trained for the same target gene by the
75 ensemble machine learning technique of stacked regression[14,15]. We refer to this novel TWAS

method as Stacked Regression-based TWAS (i.e., SR-TWAS). SR-TWAS first uses a validation transcriptomic dataset of the target tissue type to optimally train a set of weights for the multiple expression imputation base models per target gene (Stage I), by optimizing the gene expression prediction R^2 (i.e., the squared correlation between observed and predicted gene expression levels) in the validation dataset. Then SR-TWAS takes the weighted average eQTL effect sizes as the corresponding variant weights for gene-based association tests in Stage II. The trained expression imputation models by SR-TWAS are specific for the tissue type of the validation data, and the identified TWAS risk genes are interpreted with potential genetic effects mediated through their gene expression of the validation tissue type.

In the following sections, we first briefly describe the stacked regression method used by SR-TWAS and the GTEx V8 and ROS/MAP reference transcriptomic datasets used in this study. Then we describe the results of our simulation studies, validation studies using the real reference data, as well as application TWAS of AD dementia and PD. Last, we end with a discussion.

Results

Overview of SR-TWAS

In the framework of TWAS[7,9,16], a multiple linear regression model is assumed for training gene expression imputation models, taking quantitative gene expression levels $\boldsymbol{E_g}$ of the target gene and tissue as the response variable and cis-acting genetic variants nearby the target gene region (genotype matrix $\boldsymbol{G}$) as predictors, as shown in the following formula:

$$\boldsymbol{E_g} = \boldsymbol{G}\boldsymbol{w} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(0, \boldsymbol{I}).$$

The eQTL effect sizes $\boldsymbol{w}$ could be trained by different regression methods and/or using different reference panels with matched expression and genotype data ($\boldsymbol{E_g}, \boldsymbol{G}$).

Assume there are a total of K base gene expression imputation models trained for the same target gene, with $\widehat{\boldsymbol{w}_k}, k = 1, \dots, K$. Let $\mathbf{E}_{vg}$ denote the gene expression levels of the target gene g for the target tissue type in the validation data, and $\mathbf{G}_v$ denote the genotype matrix of genetic predictors in the validation data. Then the predicted GReX of the validation samples by the kth base model are given by $\mathbf{G}_v\widehat{\mathbf{w}}_k$. The stacked regression method[14,15] will solve for a set of optimal base model weights $\zeta_1, \dots, \zeta_K$, by maximizing the regression R^2 between the profiled gene expression $\mathbf{E}_{vg}$ and the weighted average GReX, $\sum_{k=1}^{K} \zeta_k\, \mathbf{G}_v\widehat{\mathbf{w}}_k$, of K base models, i.e., minimizing the following loss function of $1 - R^2$:

$$\text{minimize}_{(\zeta_k;\, k=1,\dots,K)} \frac{\left\|\mathbf{E}_{vg} - \sum_{k=1}^{K} \zeta_k\, \mathbf{G}_v\widehat{\mathbf{w}}_k\right\|^2}{\left\|\mathbf{E}_{vg} - \bar{E}_{vg}\right\|^2}, \qquad \text{s.t.} \sum_{k=1}^{K} \zeta_k = 1, \quad \zeta_k \in [0,1].$$

As a result, we will obtain a set of model weights ζ_k for $k = 1, \dots, K$ base models, and a set of eQTL effect sizes $\widetilde{\mathbf{w}}$ given by the weighted average of the eQTL effect sizes of K base models, $\widetilde{\mathbf{w}} = \sum_{k=1}^{K} \zeta_k\, \widehat{\mathbf{w}}_k$ (Stage I). Then the final predicted GReX for test genotype data $\mathbf{G}_t$ is given by $\widehat{\text{GReX}}_g = \mathbf{G}_t\widetilde{\mathbf{w}}$, and $\widetilde{\mathbf{w}}$ will be taken as variant weights in the gene-based association tests by SR-TWAS in Stage II. Genes with 5-fold cross validation (CV) $R^2 > 0.5\%$ in the validation dataset by SR-TWAS are considered as having a valid imputation model and will be tested in Stage II[9]. Here, $\widetilde{\mathbf{w}}$ is the trained eQTL effect sizes by SR-TWAS (Stage I) for the target gene of the tissue of the validation data and identified significant genes from Stage II have potential genetic effects mediated through the transcriptome of the tissue of the validation data.

117 **Reference Transcriptomic Data**

118 *GTEx V8.* The Genotype-Tissue Expression project version 8 (GTEx V8) release provides a
119 comprehensive reference dataset of Whole Genome Sequencing (WGS) genotype data with
120 matched RNA-seq transcriptomic data from 54 non-diseased tissues of 838 post-mortem donors
121 of European ancestry[10]. For our real data analysis, we used publicly-available GReX imputation
122 models trained by TIGAR-V2[9] and PrediXcan[7,8,17] with GTEx V8 reference data for European
123 subjects as base models. Supplementary Table 1 shows the disease status of GTEx subjects for
124 tissues used in the AD or PD TWAS. AD or other dementia was reported for only 4% of
125 subjects, with missing data for ~1% subjects. The majority of GTEx subjects used in PD TWAS
126 were reported to be without PD, with <1% reported with PD and with missing data for 13% of
127 training and 6% of validation subjects.

128 *ROS/MAP.* The Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP) are
129 two longitudinal clinical-pathologic cohort studies of aging and AD, which are collectively
130 referred to as ROS/MAP[11]. ROS recruits participants from religious orders across the U.S. while
131 MAP recruits lay persons in the northeastern Illinois area[11]. All participants in both studies are
132 without known dementia and agree to annual clinical evaluations and brain donation upon
133 death[11]. Only subjects of European ancestry were used in this study. Disease status of ROS/MAP
134 participants are shown in Supplementary Table 1. Of the ROS/MAP subjects used in the training
135 data of AD TWAS, 57% had normal cognitive function, 40% had AD or other dementia, and 4%
136 had missing disease status. Of the ROS/MAP subjects used in the validation data of AD TWAS,
137 75% had normal cognitive function and 25% had AD or other types of dementia.

138

139 **Simulation Studies**

140 *Simulation study design*

141 We used the real genotype data of gene *ABCA7* from ROS/MAP and GTEx V8 to simulate gene
142 expression and phenotypes, and considered multiple scenarios with varying proportions of causal
143 SNPs ($p_{causal} = (0.001, 0.01, 0.05, 0.1)$) and gene expression heritability (i.e., the proportion of
144 gene expression variation due to genetics, $h_e^2 = (0.1, 0.2, 0.5)$). We randomly selected n=465
145 training samples with Whole Genome Sequencing (WGS) genotype data from the ROS/MAP
146 cohort and GTEx V8 cohort, respectively. We randomly selected n=400 and n=800 samples with
147 WGS genotype data from ROS/MAP as our validation and test cohorts, respectively. ROS/MAP
148 training, validation, and test samples were simulated with the same causal SNPs (i.e., eQTL),
149 while training samples from the GTEx V8 cohort were simulated with true causal SNPs that
150 were 50% overlapped with the ones for ROS/MAP samples. The simulated expression
151 heritability was the same for both ROS/MAP and GTEx V8 samples.

152 We compared the performance of SR-TWAS with a Naïve approach which takes the average of
153 base models as the trained gene expression imputation model, that is, taking $\zeta_k = \frac{1}{K}, k =$
154 $1, \ldots, K$. Two base models per gene were respectively trained by PrediXcan (penalized
155 regression with Elastic-Net penalty; PrediXcan-GTEx) with the GTEx training samples (n=465),
156 and by TIGAR (nonparametric Bayesian Dirichlet process regression; TIGAR-ROSMAP) with
157 the ROS/MAP training samples (n=465). SR-TWAS and Naïve models were then obtained by
158 using these trained base models. Validation data ($n = 400$) were used to train SR-TWAS models
159 and an additional TIGAR base model (TIGAR_ROSMAP_valid). The SR-TWAS and
160 TIGAR_ROSMAP_valid models were then averaged to create a new model (Avg-valid+SR).
161 Gene expression imputation models (by SR-TWAS and Naïve methods) with 5-fold cross-
162 validation $R^2 > 0.5\%$ in the validation cohort were considered valid models, which were used to

163 produce the Avg-valid-SR models and used in the follow-up gene-based association tests. Test
164 data (n=800) were used for assessing GReX prediction performance and TWAS power, with
165 1,000 repeated simulations per scenario. We compared the performance of three base models of
166 the training and validation data (PrediXcan-GTEx, TIGAR-ROSMAP, and TIGAR-
167 ROSMAP_valid), Naïve method, SR-TWAS, and Avg-valid+SR models.

168 *Simulation study results*

169 As shown in **Fig. 1**, we showed that Avg-valid+SR obtained the highest test R^2 for gene
170 expression imputation across 11 of 12 scenarios, and the SR-TWAS models had the second-best
171 performance. SR-TWAS obtained the highest test R^2 for gene expression imputation in the
172 $p_{\text{causal}} = 0.01$ and $h_e^2 = 0.5$ scenario, where it slightly outperformed the averaged models. The
173 base models trained by TIGAR with ROS/MAP training and validation samples (TIGAR-
174 ROSMAP, TIGAR-ROSMAP_valid) performed similarly well, because the test data were
175 generated under the same model assumptions as the ROS/MAP training and validation data.
176 Here, the Avg-valid+SR models performed best for leveraging the predictive information
177 provided by all three base models. The Naïve method and PrediXcan-GTEx base models did not
178 perform well because the PrediXcan-GTEx base models were trained using GTEx training data
179 which only shared half of the true causal SNPs as the validation and test data. The Naïve
180 approach of taking averages of the PrediXcan-GTEx and TIGAR-ROSMAP base models had
181 poor performance because of the heterogeneous genetic architecture between the GTEx training
182 cohort and test cohort.

183 As expected, model performance improved with increasing true expression heritability h_e^2 with
184 the same training sample size. For all considered scenarios, the highest test R^2 were obtained

185 under a sparse causality model with $p_{\text{causal}} = 0.001$, where true causal SNP effect sizes would be
186 relatively larger given the same h_e^2. The comparison of CV R^2 and training R^2 for Naïve and
187 SR-TWAS approaches (Supplementary Figures 1-2) also showed that SR-TWAS outperformed
188 the Naïve approach under all scenarios. Because the averaging step used to obtain Avg-valid+SR
189 models does not include training and cross validation steps, no training R^2 or CV R^2 are
190 obtained for comparison.

191 In order to assess TWAS power, phenotypes were simulated with a certain proportion of
192 variance due to simulated gene expression (h_p^2). We considered a series of h_p^2 values in the range
193 of $(0.05, 0.875)$. The TWAS power comparison with three base models of the training and
194 validation data (PrediXcan-GTEx, TIGAR-ROSMAP, and TIGAR-ROSMAP_valid), Naïve
195 method, SR-TWAS, and Avg-valid+SR models were shown in **Fig. 2**, where the results were
196 consistent with the test R^2 comparison as in **Fig. 1**. The Avg-valid+SR approach performed best,
197 followed by SR-TWAS, TIGAR-ROSMAP training base models, and TIGAR-ROSMAP_valid
198 validation base models, while the Naïve method and the PrediXcan_GTEx training base models
199 performed poorly in comparison. In the $p_{\text{causal}} = 0.01$ and $h_e^2 = 0.5$ scenario, SR-TWAS slightly
200 outperformed Avg-valid+SR and had a noticeable advantage over the TIGAR-ROSMAP training
201 base models and TIGAR-ROSMAP_valid validation models. The results showed the SR-TWAS
202 approach indeed gained power by leveraging base models trained with multiple reference panels
203 and by multiple statistical methods.

204 Although desirable TWAS power ~80% was only obtained in simulation scenarios with a
205 relatively high h_p^2 that might be higher than the value in real studies, simulation power would
206 increase along with increased test sample sizes. Because real GWAS test data would have a

207 larger sample size than the 800 considered in our simulations, we expect desirable power for our
208 SR-TWAS method in real studies.

209 *Additional simulation studies*

210 Additionally, we conducted similar simulation studies for two other settings, where samples from
211 ROS/MAP and GTEx cohorts have the same set of true causal SNPs (i.e., the same genetic
212 architecture), and (i) the expression heritability was the same for both ROS/MAP and GTEx V8
213 cohorts, or (ii) the expression heritability for GTEx V8 cohort was half that of the ROS/MAP
214 cohort. The results of these settings were similar to that of the previously described setting. Avg-
215 valid+SR still had the best performance, while SR-TWAS, TIGAR-ROSMAP training base
216 models, and TIGAR-ROSMAP_valid validation base models outperformed the PrediXcan-GTEx
217 training base models and Naïve models.

218 Comparisons of CV R^2 and training R^2 for Naïve and SR-TWAS approaches for these scenarios
219 (Supplementary Figures 3, 4, 7, and 8) showed that SR-TWAS outperformed the Naïve approach
220 under all scenarios. For all considered scenarios, again the highest test R^2 was obtained under a
221 sparse causality model with high expression heritability (Supplementary Figures 5 and 9). Power
222 comparison results show that the Avg-valid+SR models obtained the highest power in most
223 scenarios, while SR-TWAS, TIGAR-ROSMAP training base models, and TIGAR-
224 ROSMAP_valid validation base models generally outperformed the PrediXcan-GTEx training
225 base models and the Naïve models (Supplementary Figures 6 and 10). The SR-TWAS approach
226 gained more power in the setting in which the expression heritability for GTEx V8 cohort was
227 only half that of ROS/MAP (Supplementary Figure 10). SR-TWAS once again had the best
228 performance in the $p_{\text{causal}} = 0.01$ and $h_e^2 = 0.5$ scenario under both of these additional settings.

229 *Model weights estimated by SR-TWAS*

230 Plots of the weights (zeta values) of base models that were estimated by SR-TWAS in all three
231 simulation settings and 12 scenarios (Supplementary Figure 11) showed that the SR-TWAS
232 training consistently estimated higher weights for the TIGAR-ROSMAP training base models
233 compared to the PrediXcan-GTEx training base models, with many models selecting only the
234 TIGAR-ROSMAP training base model (ie, models in which the zeta value estimate for the
235 TIGAR-ROSMAP training base model is 1 and the zeta value estimate for the PrediXcan-GTEx
236 model is 0). This makes sense because the validation data were generated under the same model
237 assumptions as the ROSMAP training data. In particular, when the GTEx training data were also
238 generated under the same model as the validation data (orange bars in Supplemental Figure 11),
239 zeta value estimates for the TIGAR-ROSMAP models are more evenly distributed in [0, 1] for
240 scenarios with sparse causality model (p_causal = 0.001) and high expression heritability (he2 =
241 0.2, 0.5). When the GTEx training data were generated under a different setting where half of the
242 true causal eQTL in the validation data were also causal in the ROSMAP training data (black
243 bars in Supplemental Figure 11), the zeta value for the TIGAR-ROSMAP base model is more
244 frequently estimated to be 1.

245 Even when both ROSMAP and GTEx training data were generated under the same model
246 assumptions, the SR-TWAS method is still shown with gained power because the two base
247 models trained by TIGAR and PrediXcan have complementary properties. In particular,
248 PrediXcan uses a parametric penalized regression model with Elastic-Net penalty which is
249 preferred for sparse genetic architecture of gene expression quantitative traits. Whereas TIGAR
250 uses a nonparametric Bayesian Dirichlet process regression model which assumes an
251 infinitesimal model for the underlying genetic architecture of gene expression quantitative traits.

252 As shown in previous studies[9], PrediXcan will perform better when the true causal eQTL is
253 sparse, while TIGAR will perform better as the true causal eQTL proportion increases. Our
254 simulation studies showed that SR-TWAS had improved performance across various scenarios
255 with (p_{causal}= 0.001, 0.01, 0.05, 0.1; h_e^2 = 0.1, 0.2, 0.5; Supplementary Figure 5).

256 *Type I error assessment*

257 We also assessed type I error under the example scenario with $p_{\text{causal}} = 0.1$, $h_e^2 = 0.1$. Base
258 model weights for TIGAR-ROSMAP, PrediXcan-GTEx, and TIGAR-ROSMAP-valid were
259 permuted 10^6 times. The TIGAR-ROSMAP and PrediXcan-GTEx training base models with
260 permuted weights were then used to obtain SR-TWAS and Naïve models. The Avg-valid+SR
261 was obtained by averaging the SR-TWAS models and TIGAR-ROSMAP_valid training base
262 models with permuted weights. All models were then used to conduct gene-based association
263 tests with a phenotype generated randomly from $N(0, 1)$. As shown in Supplementary Table 2,
264 all models controlled well for type I errors for significance thresholds $(10^{-4}, 10^{-5}, 2.5 \times$
265 $10^{-6}, 10^{-6})$.

266

267 **Real Validation Studies**

268 To compare the GReX prediction accuracy with real gene expression data, we considered three
269 sets of base models that were trained by TIGAR with ROS samples (n=237,
270 TIGAR_ROS_DLPFC) of dorsolateral prefrontal cortex (DLPFC) tissue, trained by TIGAR with
271 GTEx V8 data of brain frontal cortex tissue (n=157, TIGAR_GTEx_BRNCTXB)[9], and trained
272 by PrediXcan with the same GTEx reference data of brain frontal cortex tissue (n=157,
273 PrediXcan_GTEx_BRNCTXB)[7,8,17]. SR-TWAS (SR-TWAS_MAP_DLPFC) and Naïve

274 (Naive_MAP_DLPFC) models were trained from these three sets of base models with respect to
275 a validation dataset with half of the MAP samples (n=114, randomly selected) of DLPFC tissue.
276 Valid gene expression imputation models trained by SR-TWAS and Naïve methods with 5-fold
277 CV $R^2 > 0.5\%$ in validation data were tested using the other half of the MAP samples (n=114)
278 of DLPFC tissue.

279 By comparing test R^2 obtained by SR-TWAS, Naïve, and three sets of base models in the test
280 MAP samples (Supplementary Table 3), we showed that PrediXcan_GTEx_BRNCTXB[7,8,17] had
281 the highest median (0.070) and mean (0.113) test R^2 but only for 867 valid gene expression
282 imputation models, SR-TWAS had the second highest median (0.026) and mean (0.068) test R^2
283 for 8425 valid genes expression imputation models, and Naïve model performed similarly to SR-
284 TWAS but with a slightly lower median (0.025) and mean (0.065) test R^2 and fewer valid genes
285 expression imputation models (8360). By pair-wise comparison of test R^2 for all genes with
286 valid expression imputation models (Supplementary Figure 12), SR-TWAS (y-axis) performed
287 noticeably better than the Naïve and three sets of base models (x-axis).

288

289 **Application TWAS of AD Dementia**

290 *Training expression imputation models of SMA tissue by SR-TWAS*

291 We considered four sets of base models — TIGAR and PrediXcan models trained with 465
292 ROS/MAP samples of DLPFC tissue (TIGAR_ROSMAP_DLPFC,
293 PrediXcan_ROSMAP_DLPFC), TIGAR and PrediXcan models trained with 157 GTEx V8
294 samples of prefrontal cortex tissue (TIGAR_GTEx_BRNCTXB[9], PrediXcan_GTEx_BRNCTXB
295 [7,8,17]). An additional 76 ROS/MAP samples of the supplementary motor area (SMA) brain tissue

296 were used as the validation dataset to train SR-TWAS models and to calculate the 5-fold CV R^2
297 that was used to select genes with valid imputation models. Plots of zeta weights estimated by
298 SR-TWAS for each set of training base models were presented in Supplementary Figure 13. We
299 observed that the weights of the TIGAR_GTEx_BRNCTXB[9] training base models was estimated
300 to be 1 more often than the other training base models. These results showed that the base
301 models trained by TIGAR with the GTEx data of BRNCTXB tissue
302 (TIGAR_GTEx_BRNCTXB[9]) contributed solely to SR-TWAS models for almost half of the
303 genes, which was consistent with the numbers of valid gene expression imputation models as
304 shown in Table 1.

305 PrediXcan and TIGAR models trained using the 76 ROS/MAP samples of SMA validation
306 dataset (PrediXcan_ROSMAP_SMA, TIGAR_ROSMAP_SMA) were also included for
307 comparison and were averaged with SR-TWAS models to obtain an additional set of models
308 (Avg-valid+SR). Here, we compared Avg-valid+SR and SR-TWAS models to training base
309 models, as well as validation base models trained by PrediXcan and TIGAR using the validation
310 data of the target SMA brain tissue, to show the advantages of SR-TWAS for leveraging multiple
311 regression models, multiple reference panels, and multiple tissues.

312 By comparing the CV R^2 and numbers of genes with valid expression imputation models (**Table**
313 **1**), we found that gene expression imputation models trained by SR-TWAS for the SMA tissue
314 (SR-TWAS_ROSMAP_SMA) had the highest median CV R^2 (0.072) and second highest mean
315 CV R^2 (0.09) for ~20K genes with valid expression imputation models. Although the
316 PrediXcan_GTEx_BRNCTXB[7,8,17] training base models had the third highest median CV R^2
317 (0.061) and highest mean CV R^2 (0.10), only 4563 genes had valid expression imputation
318 models. These results showed that SR-TWAS obtained improved CV R^2 in a real validation

319 cohort of SMA tissue by leveraging multiple regression methods from two reference panels of
320 multiple relevant tissues.

321 *TWAS results of AD dementia*

322 By using the eQTL weights obtained by training base models, validation base models, SR-
323 TWAS models, and Avg-valid+SR models, we conducted TWAS with the summary-level data
324 of the recent GWAS of AD dementia (n=~762K)[18]. Since Avg-valid+SR models were shown to
325 have the best performance in our simulation studies, here we focused on summarizing the results
326 by Avg-valid+SR here. In particular, Avg-valid+SR identified a total of 89 significant TWAS
327 risk genes of AD dementia with p-values $< 2.5 \times 10^{-6}$. Of these, 19 are known GWAS risk
328 genes, 70 are within 1Mb of a known GWAS risk gene, and 14 have been previously observed as
329 TWAS risk genes of AD dementia[19–24] (Supplementary Table 4).

330 *Validation of Significant TWAS Risk Genes of AD by PMR-Egger*

331 In order to account for potential horizontal pleiotropy effects (genetic effects on phenotype that
332 are not mediated by the considered GReX), we applied the PMR-Egger tool to the 89 significant
333 TWAS risk genes of AD obtained by the Avg-valid+SR models. This analysis was performed
334 using the same validation transcriptomic dataset and GWAS summary data as in the application
335 TWAS of AD dementia. Of these 89 analyzed genes, 61 (68.5%) had a significant causal p-value
336 by PMR-Egger after Bonferroni correction for multiple testing.

337 *Independent significant TWAS risk genes of AD dementia*

338 Because TWAS considers genotype data within a $\pm$1Mb region of the test gene, nearby
339 significant TWAS genes with overlapping test regions often have correlated GReX values and
340 might not represent independent associations. We curated 6 independent TWAS risk genes of

341 AD from the 61 significant genes that were validated by PMR-Egger (**Table 2, Fig. 3**) by
342 selecting the most significant gene in a cluster of significant genes with overlapped test regions
343 as the independent risk gene. We found that 1 of these independent risk genes, *HLA-DRA,* was a
344 known GWAS[21,25] risk gene and also previously observed as a TWAS risk gene[20]. The other 5
345 independent risk genes were near known GWAS risk genes[18,20,21,25–27] and near previously
346 observed TWAS risk genes[19,21]. Compared to the TWAS results using training base models
347 (Supplementary Figure 14) and validation base models (**Fig. 3**), Avg-valid+SR models identified
348 the greatest number of independent risk genes.

349

350 **Application TWAS of PD**

351 *Training expression imputation models of brain substantia nigra tissue by SR-TWAS*

352 We considered six sets of base models trained by TIGAR on six different tissues from GTEx V8
353 — brain anterior cingulate cortex BA24 (BRNACC) (n=136), brain caudate basal ganglia
354 (BRNCDT) (n=173), brain cortex (BRNCTXA) (n=184), brain nucleus accumbens basal ganglia
355 (BRNNCC) (n=182), brain putamen basal ganglia (BRNPTM) (n=154), and whole blood
356 (BLOOD) (n=574). With these six sets of training base models, an additional 101 GTEx samples
357 of brain substantia nigra (BRNSNG) tissue were used as the validation data to train SR-TWAS
358 models. We presented the plots of zeta weights of training base models that were estimated by
359 SR-TWAS in Supplementary Figure 15. For all 6 sets of considered training base models, zeta
360 weights were similarly distributed with 0's for most genes and other values distributed over (0,
361 1). 5-fold CV R^2s were calculated and used to select genes with valid expression imputation
362 models for TWAS. PrediXcan and TIGAR models trained on the validation data of brain
363 substantia nigra tissue (PrediXcan_GTEx_BRNSNG[7,8,17], TIGAR_GTEx_BRNSNG[9]) were used

364 to obtain average models (Avg-valid+SR) of these two sets of validation base models and SR-
365 TWAS models.

366 By comparing the CV R^2 and number of genes with valid expression imputation models (**Table**
367 **3**), we found that gene expression imputation models trained by SR-TWAS for the brain
368 substantia nigra tissue (SR-TWAS_GTEx-BRNSNG[9]) had the highest median CV R^2 (0.068)
369 and highest mean CV R^2 (0.094) for ~23K genes with valid expression imputation models.
370 These results showed that SR-TWAS obtained improved CV R^2 in a real validation cohort of
371 substantia nigra tissue by leveraging multiple regression methods from two reference panels of
372 multiple relevant tissues.

373 *TWAS results of PD*

374 We conducted TWAS using GWAS summary statistics by the recent GWAS of PD (n=~33K
375 cases, ~18k UK Biobank proxi-cases, and ~828K controls)[28], using eQTL weights estimated by
376 the above 6 sets of training base models of multiple tissues, validation base models of the
377 BRNSNG tissue, SR-TWAS models, and Avg-valid+SR models (Supplementary Figure 16).
378 Here, we also focused on the results by using the Avg-valid+SR models (**Fig. 4; Table 4;**
379 **Supplementary Table 5**), including a total of 60 significant TWAS risk genes of PD. Of these,
380 11 are known GWAS risk genes, 47 are within 1Mb of a known GWAS risk gene, and 11 have
381 been previously observed as TWAS risk genes of PD.

382 *Validation of Significant TWAS Risk Genes of PD by PMR-Egger*

383 We applied the PMR-Egger tool to the 60 significant TWAS risk genes of PD obtained by the
384 Avg-valid+SR models. This analysis was performed using the same validation transcriptomic

385 dataset and GWAS summary data as in the application TWAS of PD. Of these genes, 46 (76.6%)
386 had a significant causal p-value by PMR-Egger after Bonferroni correction for multiple testing.

387 *Independent Significant PD TWAS Risk Genes*

388 Similarly, from these 46 replicated risk genes with significant causal p-values by PMR-Egger,
389 we curated 9 independent TWAS risk genes of PD (**Fig. 4; Table 4**), including 6 novel TWAS
390 risk genes (*LA16c-431H6.7*, *ADORA2B*, *AC005082.12*, *MAPK8IP1P2*, *MYLPF*, and *PARL*). Of
391 these novel TWAS risk genes, four (*AC005082.12*, *MAPK8IP1P2*, *MYLPF*, and *PARL*) are near
392 known GWAS risk genes (*GPNMB*[28,29], *MAPT*[28] , *MCCC1*[28], *SETD1A*[28], *ZSWIM7*[30]). The other
393 3 previously observed TWAS risk genes were also known GWAS risk genes (*CD38*[29,30],
394 *MMRN1*[28,30], and *NDUFAF2*[28]). Compared to the TWAS results using these six training base
395 models (Supplementary Figure 16) and validation base models (**Fig. 4**), Avg-valid+SR models
396 still identified the greatest number of independent risk genes that were validated by PMR-Egger.

397

398 ## Discussion

399 We present a novel TWAS tool (SR-TWAS) using the ensemble machine learning technique of
400 stacked regression[14,15,31] for leveraging multiple gene expression imputation models (i.e., base
401 models) trained by different regression methods and/or using different transcriptomic reference
402 panels of different tissue types. We also constructed a set of average models (Avg-valid+SR) of
403 the SR-TWAS models and validation base models that were trained using the validation data.
404 Different from existing methods such as UTMOST[12] and SWAM[13], SR-TWAS requires only
405 summary-level base models, providing the flexibility of using publicly available base models.

406 With comprehensive simulation studies, we compared the Avg-valid+SR models to SR-TWAS
407 models, the naïve approach of averaging all training base models, training base models, and
408 validation base models. We showed that the Avg-valid+SR expression imputation models had
409 the best prediction accuracy and led to the best TWAS power across 11 out of 12 considered
410 scenarios, and that SR-TWAS models performed the best in the remaining scenario.

411 In the real data validation and application studies using ROS/MAP and GTEx V8 reference
412 panels and GWAS summary data of Alzheimer's disease (AD) dementia and Parkinson's disease
413 (PD), Avg-valid+SR models also outperformed base models trained using single reference panels
414 and tissue types. Avg-valid+SR models identified a greater number of total independent risk
415 genes that were replicated by PMR-Egger than any of the base training or validation models.
416 Besides known GWAS/TWAS risk genes that were identified by Avg-valid+SR models, we also
417 found 5 novel independent TWAS risk genes for AD dementia and 6 novel independent TWAS
418 risk genes for PD with known functions in respective disease pathology (Table 2; Table 4). Most
419 of these novel findings are located within 1Mb region of previously known GWAS and TWAS
420 risk genes. Additionally, we found interesting biological interpretations relevant to AD dementia
421 and PD for our identified TWAS risk genes.

422 *Important findings by application TWAS of AD dementia*

423 The results of the TWAS of AD include genes with known associations with AD dementia, or
424 with relevant biological processes like immune response and regulation of AD-associated genes.
425 Among these 6 curated independent significant genes by Avg-valid+SR models (**Table 2**), *HLA-*
426 *DRA* is located in the major histocompatibility complex region that is expressed in glial cells[32]
427 and which has also been previously identified by eQTL analysis[21]. The other 5 independent
428 significant TWAS AD risk genes of AD (*AC092849.1*, *AL110118.2*, *FOSB*, *RN7SL225P*, and

429 *SRD5A3P1*) are within 1Mb of known GWAS risk genes of AD[18,20,21,25–27] and previously
430 observed TWAS risk genes[19,20]. *FOSB* was the most significant of a cluster of 38 identified
431 significant TWAS risk genes with test region overlapped with the well-known GWAS risk gene
432 *APOE*[18,20,21,27,33,34] (**Fig. 3**). An alternatively-spliced product of the *FOSB* gene has been
433 implicated in the regulation of gene expression and cognitive dysfunction in mouse models of
434 AD[35]. *SRD5A3P* is located in the AD-associated *MS4A* gene cluster[26], which contains multiple
435 known GWAS risk genes[18,20–22] as well as TWAS risk gene *MS4A2*[19] of AD. The *MS4A* gene
436 cluster is notable due its role in the regulation of soluble TREM2, which is encoded by the
437 known AD risk gene *TREM2*, in cerebrospinal fluid in AD[26].

438 *Important findings by application TWAS of PD*

439 Similarly, results of the TWAS of PD include genes with known associations with PD, with
440 related conditions, and with relevant biological processes like inflammation. Among these 9
441 curated independent significant genes (**Table 4**), *PARL* plays a role in regulating cellular
442 processing of the mitochondrial kinase protein encoded by *PINK1*, mutations in which are a
443 known cause recessively-inherited, early-onset PD[36]. *NDUFAF2* encodes for a component of
444 mitochondrial complex I and loss of its functionality results in a rare mitochondrial
445 encephalopathy with frequent substantia nigra pathology and motor symptoms[37]. *NDUFAF2* was
446 also identified as a potential drug target in a Mendelian randomization study of potential drug
447 targets for PD treatment[38]. A study of PD-associated *GPNMB* found that it is upregulated with
448 lncRNA gene *AC005082.12*[39].

449 Additionally, *ADORA2B* encodes adenosine receptor A2B which is an important cell receptor
450 involved in numerous pathways and implicated in a broad variety of diseases including asthma,
451 sepsis, inflammatory bowel disease, cancer, renal disease, diabetes, vascular diseases, and lung

452 disease[40,41]. The immunomodulatory effects and role in inflammatory processes have made A2B
453 a target for pharmacological therapeutics[40]and antagonists the similar adenosine receptor A2A
454 were the first non-dopaminergic drug therapy for PD[42]. *MAPK8IP1P2* is a pseudogene near a
455 known TWAS PD risk gene *LRRC37A2*[30] (also identified by Avg-valid+SR models). *CD38* is
456 involved in neurodegeneration, neuroinflammation, and aging[30,43]. *MMRN1* is a carrier protein
457 for platelet factor V and lies ~84KB downstream of a well-established GWAS PD risk locus
458 found in multiple populations[28].

459 *Tool for implementing SR-TWAS*

460 The SR-TWAS tool, including options for constructing SR-TWAS models, models by the Naïve
461 method, and Avg-valid+SR models, is publicly-available on GitHub. The SR-TWAS tool
462 implements user-friendly features, including accepting genotype data in standard VCF-format as
463 input, enabling parallel computation, and using efficient computation strategies to reduce time
464 and memory usage. The most computationally expensive part is to train all base models with
465 different reference panels, which is subject to the regression method. For example, with training
466 sample size n=465, PrediXcan (Elastic-Net) costs ~1 CPU minute and TIGAR (DPR) costs ~3
467 CPU minutes on average per gene. Publicly-available trained models can also be used as base
468 models by the SR-TWAS tool. The process of training SR-TWAS models from base models and
469 validation data is quite computational efficient. For example, with the ROS/MAP SMA tissue
470 validation dataset (n=76) and four base models in our real studies SR-TWAS model training
471 costs ~15 CPU seconds per gene. With the GTEx substantia nigra tissue validation dataset
472 (n=101) and six base models in our real studies SR-TWAS model training costs ~103 CPU
473 seconds per gene.

474 *Limitations*

475 SR-TWAS still has its limitations. For example, SR-TWAS only considers cis-eQTL during
476 model training, uses the standard two-stage TWAS, requires an additional validation dataset of
477 the target tissue independent of those used for base model training[15], and assumes samples of the
478 reference panels and validation dataset are of the same ancestry[44].

479 First, previous studies have illustrated the importance of considering both cis- and trans- eQTL
480 in TWAS[45], and a joint modeling of the gene expression imputation and the gene-based
481 association test[46 47]. The stacked regression technique used by SR-TWAS also applies to
482 scenarios considering both cis- and trans- eQTL, when base models trained with both cis- and
483 trans- eQTL are available.

484 Second, because the standard two-stage TWAS was implemented by SR-TWAS, the potential
485 horizontal pleiotropy (i.e., genetic effects on the phenotype of interest that are not mediated by
486 GReX) was not considered. However, the recently published PMR-Egger[48] tool that tests the
487 genetic effects mediated through the GReX term while controlling for horizontal pleiotropy
488 could be used to validate the findings by SR-TWAS.

489 Third, in this study we only evaluated scenarios where reference samples used to train base
490 models are of the same ancestry as the validation data. Although the method can be generalized
491 to scenarios where base and validation data are of different ancestries, having at least one set of
492 training base models with the same ancestry as the validation data would be a requirement for
493 promising TWAS results. Evaluating the performance of SR-TWAS with base models of
494 multiple ancestries is beyond the scope of this study, and is part of our ongoing study.

495 *Summary*

496 Overall, the SR-TWAS tool provides a useful resource for researchers to take advantages of the
497 publicly-available gene expression imputation models by using multiple regression methods
498 (e.g., PrediXcan[7,8], FUSION[16], TIGAR[9]) and different reference panels of multiple tissue types
499 (e.g., ROS/MAP[11], GTEx V8[10]). In particular, the final trained gene expression imputation model
500 by SR-TWAS will be with respect to the same tissue type as the validation data set. Because
501 multiple base models would not only increase the robustness of the gene expression imputation
502 model but also increase the total effective training sample size, SR-TWAS is expected to further
503 increase TWAS power for studying complex human diseases. The approach of constructing
504 average models of the SR-TWAS models and validation base models (Avg-valid+SR) provides a
505 set of optimal gene expression imputation models that can leverage both training base models
506 and validation base models to achieve the best TWAS performance.

507

508 **Acknowledgements**

509 RP and JY are supported by National Institutes of Health (NIH/NIGMS) grant award
510 R35GM138313. MPE was supported by NIH/NIGMS grant award R01GM117946 and NIH/NIA
511 grant award RF1AG071170. ROS/MAP study data were provided by the Rush Alzheimer's
512 Disease Center, Rush University Medical Center, Chicago, IL. Data collection was supported
513 through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146,
514 R01AG36836, R01AG56352, U01AG32984, U01AG46152, U01AG61356, the Illinois
515 Department of Public Health, and the Translational Genomics Research Institute.

516 **Table 1. Comparison of CV R^2 of SMA tissue for valid gene expression imputation models**
517 **trained by training base models with ROS/MAP and GTEx V8 reference panels of DLPFC**
518 **tissue, validation base models, SR-TWAS models, and Avg-valid+SR models.**

	Sample Size	Median CV R^2	Mean CV R^2	N_{genes}
PrediXcan_GTEx_BRNCTXB[7,8,17a]	157	0.061	0.101	4563
PrediXcan_ROSMAP_DLPFC[a]	465	-	-	6532
TIGAR_GTEx_BRNCTXB[9a]	157	0.038	0.065	21921
TIGAR_ROSMAP_DLPFC[a]	465	0.017	0.049	11981
PrediXcan_ROSMAP_SMA[b]	76	0.031	0.053	23888
TIGAR_ROSMAP_SMA	76	0.064	0.077	32350
SR-TWAS_ROSMAP_SMA	-	0.072	0.090	20216
Avg-valid+SR	-	-	-	32238

519 *a.* Base training model used by SR-TWAS;
520 *b*. Base validation model.
521
522

523 **Table 2. Independent TWAS risk genes of AD dementia identified by Avg-valid+SR models**
524 **of SMA tissue.** TWAS Zscores and p-values were presented here for the models of Avg-
525 valid+SR, SR-TWAS, PrediXcan validation base models, and TIGAR validation base models.
526 The signs of Zscores show the directions of the mediated genetic effects on AD dementia.
527

		Avg-valid+SR ROSMAP SMA		SR-TWAS ROSMAP SMA		PrediXcan ROSMAP SMA		TIGAR ROSMAP SMA	
Gene	**CHR**	**Zscore**	**P-value**	**Zscore**	**P-value**	**Zscore**	**P-value**	**Zscore**	**P-value**
HLA-DRA[ac]	6	-6.81	1.00e-11	-6.88	5.93e-12	-	-	-	-
AC092849.1[bd]	7	6.54	6.23e-11	-	-	6.48	9.02e-11	5.98	2.21e-09
SRD5A3P1[bd]	11	8.73	2.49e-18	-	-	-	-	8.73	2.49e-18
RN7SL225P[bd]	11	5.05	4.33e-07	-	-	5.03	5.02e-07	4.93	8.04e-07
AL110118.2[bd]	14	-5.32	1.01e-07	-	-	-5.25	1.52e-07	-5.63	1.77e-08
FOSB[bd]	19	29.19	2.44e-187	-	-	29.20	1.87e-187	13.79	2.98e-43

528 *a.* Known GWAS risk gene of AD.
529 *b.* Gene within 1Mb of known GWAS risk gene of AD.
530 *c.* Previously observed TWAS risk gene of AD.
531 *d.* Gene within 1Mb of previously observed TWAS risk gene of AD.
532

533

534 **Table 3. Comparison of CV R^2 of BRNSNG tissue for valid gene expression imputation**
535 **models trained by training base models with GTEx V8 reference panel of multiple tissues,**
536 **validation base models, SR-TWAS models, and Avg-valid+SR models.**

	Sample Size	Median CV R^2	Mean CV R^2	N_{genes}
TIGAR_GTEx_BRNACC[9a]	136	0.040	0.062	23068
TIGAR_GTEx_BRNCDT[9a]	173	0.034	0.063	23362
TIGAR_GTEx_BRNCTXA[9a]	184	0.033	0.066	23379
TIGAR_GTEx_BRNNCC[9a]	182	0.033	0.060	23382
TIGAR_GTEx_BRNPTM[9a]	154	0.037	0.062	22481
TIGAR_GTEx_BLOOD[9a]	574	0.020	0.071	16682
TIGAR_GTEx_BRNSNG[9b]	101	0.050	0.068	23051
PrediXcan_GTEx_BRNSNG[7,8,17b]	101	0.052	0.080	2559
SR-TWAS_GTEx_BRNSNG	-	0.068	0.094	22913
Avg-valid+SR	-	-	-	21683

537 *a.* Base training model used by SR-TWAS;
538 *b*. Base validation model.
539

540 **Table 4. Independent TWAS risk genes of Parkinson's disease identified by Avg-valid+SR**
541 **models of SMA tissue.** TWAS Zscores and p-values were presented here for the models of Avg-
542 valid+SR, SR-TWAS, PrediXcan validation base models, and TIGAR validation base models.
543 The signs of Zscores show the directions of the mediated genetic effects on AD dementia.
544

		Avg-valid+SR GTEx BRNSNG		SR-TWAS GTEx BRNSNG		PrediXcan GTEx BRNSNG		TIGAR GTEx BRNSNG	
Gene	**CHR**	**Zscore**	**P-value**	**Zscore**	**P-value**	**Zscore**	**P-value**	**Zscore**	**P-value**
PARL[bd]	3	-4.75	2.02e-06	-	-	-	-	-	-
CD38[ac]	4	7.59	3.31e-14	7.68	1.57e-14	-	-	-	-
MMRN1[ac]	4	-7.99	1.34e-15	-7.91	2.55e-15	-7.90	2.80e-15	-8.50	1.95e-17
NDUFAF2[ac]	5	-5.33	9.80e-08	-5.45	4.91e-08	-	-	-	-
AC005082.12[bd]	7	5.94	2.83e-09	5.47	4.44e-08	-	-	5.35	8.66e-08
LA16c-431H6.7	16	-4.95	7.60e-07	-	-	-	-	-	-
MYLPF[bd]	16	6.34	2.31e-10	-	-	-	-	6.31	2.78e-10
ADORA2B[d]	17	-5.02	5.16e-07	-5.20	2.04e-07	-4.94	7.96e-07	-5.29	1.21e-07
MAPK8IP1P2[bd]	17	9.31	1.30e-20	9.05	1.39e-19	-	-	9.42	4.38e-21

545 *a.* Known GWAS risk gene of PD.
546 *b.* Gene within 1Mb of known GWAS risk gene of PD.
547 *c.* Previously observed TWAS risk gene of PD.
548 *d.* Gene within 1Mb of previously observed TWAS risk gene of PD.

549 ## Figures legends

550 **Fig. 1. Boxplots of gene expression prediction R^2 for simulations with varying proportion**
551 **of true causal SNPs $p_{\text{causal}} = (0.001, 0.01, 0.05, 0.1)$ and true expression heritability**
552 **$h_e^2 = (0.1, 0.2, 0.5)$.** Avg-valid+SR models obtained the highest test R^2 for gene expression
553 imputation across all 11 of 12 scenarios, while SR-TWAS performed best in the scenario with
554 $p_{\text{causal}} = 0.01$ and $h_e^2 = 0.5$. This is because test samples are simulated under the same genetic
555 architecture as the ROSMAP training cohort and the validation cohort, which only have ~50%
556 overlapped true causal SNPs as the GTEx training cohort.

557

558 **Fig. 2. Power comparison for simulations with varying proportion of true causal SNPs**
559 $p_{\text{causal}} = (0.001, 0.01, 0.05, 0.1)$**, true expression heritability** $h_e^2 = (0.1, 0.2, 0.5)$**, and**
560 **phenotype heritability** $h_p^2 \in (0.05, 0.875)$**.** Similar patterns were observed as shown in Fig. 1,
561 where Avg-valid+SR models obtained the highest test R^2 for gene expression imputation across
562 all 11 of 12 scenarios and SR-TWAS models performed best in the scenario with $p_{\text{causal}} = 0.01$
563 and $h_e^2 = 0.5$.

564

565 **Fig. 3. Manhattan plots of TWAS results by validation base models, SR-TWAS models, and**
566 **Avg-valid+SR models of SMA tissue for studying AD dementia.** Validation base models were
567 trained by PrediXcan and TIGAR using the ROS/MAP validation data of SMA tissue. Avg-
568 valid+SR models were obtained by averaging the SR-TWAS and these two sets of validation
569 base models. A total of 89 (6 independent and significantly causal by PMR-Egger) TWAS risk
570 genes were identified by the Avg-valid+SR models. Significant genes are shown in orange and
571 significant genes discussed in the main text are labeled and shown in red.

572

573 **Fig. 4. Manhattan plots of TWAS results by validation base models, SR-TWAS models, and**
574 **Avg-valid+SR models of substantia nigra tissue for studying Parkinson's Disease.**
575 Validation base models were trained by PrediXcan and TIGAR using the GTEx validation data
576 of substantia nigra tissue. Avg-valid+SR models were obtained by averaging the SR-TWAS and
577 these two sets of validation base models. A total of 60 (9 independent and significantly causal by
578 PMR-Egger) significant TWAS risk genes were identified by the Avg-valid+SR models.
579 Significant genes are shown in orange and significant genes discussed in the text are labeled and
580 shown in red.

581

582

583 **Methods**

584 **SR-TWAS using Stacked Regression**

585 Stacked regression is a machine learning method for forming optimal linear combinations of
586 different predictors to improve prediction accuracy[15]. The theoretical background for combining
587 predictors rather than selecting a single best predictor is well-established and has been developed
588 since the 1970s[15,49,50]. The "stacking" method of combining predictors originated in a 1992
589 paper[14] by Wolpert, who described the concept as any scheme for feeding information from a set
590 of cross-validated models to another before forming the final prediction in order to reduce
591 prediction error[14]. The idea is further expanded with stacked regression, a specific framework for
592 combining the initial predictors by weighted average with coefficient constraints to control for
593 multicolinearity[15].

594 In standard two-stage TWAS, we need to first fit a gene expression imputation model, which is
595 assumed as a multivariable linear regression model, with quantitative gene expression levels E_g
596 for the target gene and tissue type as the response variable, and genotype matrix G of
597 nearby/genome-wide SNPs as predictors,

598 $$\boldsymbol{E_g} = \boldsymbol{Gw} + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim N(0, 1).$$

599 This gene expression imputation model can be trained per gene per tissue type, using a
600 transcriptomic reference panel which profiles both transcriptomic and genetic data of the same
601 training cohort. SNPs with non-zero effect sizes $\boldsymbol{w}$ are referred to as a broad sense of eQTL. The
602 eQTL effect sizes $\boldsymbol{w}$ will be estimated from each trained model by different regression methods
603 and/or using different reference data of multiple tissue types.

604 Assume there are a total of K base gene expression imputation models that are trained for the
605 same target gene and tissue type, with $\widehat{\boldsymbol{w_k}}, k = 1, \ldots, K$, as the trained eQTL effect sizes per base

606 model. Let $\mathbf{E}_{vg}$ denote the gene expression levels of the same target gene g and tissue type in the
607 validation data, and $\mathbf{G}_v$ denote the genotype matrix of the same genetic predictors in the
608 validation data. Then the predicted Genetically Regulated gene eXpression (GReX) of the
609 validation samples are given by $\mathbf{G}_v\widehat{\mathbf{w}}_k$, by the kth base model. The stacked regression
610 method[14,15] will solve for a set of optimal model weights $\zeta_1, \dots, \zeta_K$, by maximizing the regression
611 R^2 between the profiled gene expression $\mathbf{E}_{vg}$ and the weighted average GReX, $\sum_{k=1}^{K} \zeta_k\, \mathbf{G}_v\widehat{\mathbf{w}}_k$, of
612 K base models, ie, minimizing the following loss function of $1 - R^2$:

613 $$\text{minimize}_{(\zeta_k;\, k=1,\dots,K)} \frac{\left\|\mathbf{E}_{vg} - \sum_{k=1}^{K} \zeta_k\, \mathbf{G}_v\widehat{\mathbf{w}}_k\right\|^2}{\left\|\mathbf{E}_{vg} - \bar{E}_{vg}\right\|^2}, \qquad \text{s.t.} \sum_{k=1}^{K} \zeta_k = 1, \quad \zeta_k \in [0,1].$$

614 As a result, we will obtain a set of model weights ζ_k for $k = 1, \dots, K$ base models, and a set of
615 eQTL effect sizes $\widetilde{\mathbf{w}}$ given by the weighted average of the eQTL effect sizes of K base models,
616 $\widetilde{\mathbf{w}} = \sum_{k=1}^{K} \zeta_k\, \widehat{\mathbf{w}}_k$ (Stage I). Then the final predicted GReX for test genotype data $\mathbf{G}_t$ is given by
617 $\widehat{\text{GReX}}_g = \mathbf{G}_t\widetilde{\mathbf{w}}$, and $\widetilde{\mathbf{w}}$ will be taken as variant weights in the gene-based association tests by
618 SR-TWAS in Stage II.

619 Genes with 5-fold CV $R^2 > 0.5\%$ in the validation dataset by SR-TWAS are considered as
620 having a valid imputation model and will be tested in Stage II. That is, the validation dataset will
621 be randomly split into 5 folds. For each fold of data, SR-TWAS model will be trained using the
622 other 4-fold data and then use to calculate prediction R^2 with the current fold. The average
623 prediction R^2 across all 5 folds of data is considered as the 5-fold CV R^2. Here, we use a more
624 liberal threshold (0.005) than the threshold 0.01 used by previous studies [16,51,52] to allow more
625 genes to be tested in follow-up TWAS. Because the follow-up gene-based association Z-score
626 test statistic is essentially a weighted average of single variant GWAS Z-score statistics with

627 variant weights provided by the eQTL effect sizes[9], poorly estimated eQTL weights would only
628 reduce power but will not increase false positive rate under null hypothesis.
629
630 **Naïve Method**
631 In this paper, we compared SR-TWAS to a Naïve approach which just takes the average of base
632 models as the trained gene expression imputation model, that is, takes $\zeta_k = \frac{1}{K}, k = 1, \ldots, K$.
633 Using a validation dataset, we can still evaluate the validation R^2 which can be used to select
634 valid genes with validation $R^2 > 0.5\%$.
635
636 **Avg-valid+SR Models**
637 We further constructed average models of the SR-TWAS models and validation base models
638 trained using the validation dataset, which are referred to as Avg-valid+SR models. Because SR-
639 TWAS and validation base models are averaged directly, training R^2 and CV R^2 are not obtained
640 for Avg-valid+SR models. We compared the Avg-valid+SR models to the SR-TWAS models
641 and validation base models in both simulation and real studies.
642
643 **SR-TWAS Tool Framework**
644 SR-TWAS tool was designed to be compatible with the TIGAR-V2 tool framework[9]; it accepts
645 models trained by TIGAR-V2 as input, imports utility functions from TIGAR-V2, and outputs
646 model files which can be used as input for TIGAR-V2 GReX prediction and summary-level
647 TWAS. Much of the structure of the SR-TWAS code was derived from existing TIGAR-V2
648 scripts and it shares dependencies on TABIX[53] and the Python libraries of numpy[54,55], pandas[54],
649 scipy[56], statsmodels[57], and scikit-learn[58,59].

650 The SR-TWAS script utilizes scikit-learn's consistent, extensible interfaces for defining
651 estimators and predictors and for initializing objects[59]. The script trains a stacked regression
652 model using a modified version of scikit-learn's StackingRegressor class, which trains a final
653 estimator from cross-validated predictions from base estimators fitted on the full design matrix.
654 The script defines two custom classes to be used as input for the stacking regressor object: a base
655 estimator class (WeightEstimator) which converts trained GReX prediction models into scikit-
656 learn-compatible estimator objects and a final estimator class (ZetasEstimator) which obtains the
657 values of $\zeta_1, \ldots, \zeta_K$ that minimize the loss function under the constraints $\zeta_k \geq 0$ and $\sum_{k=1}^{K} \zeta_k =$
658 1[15].
659 During the stacked regression, SNP minor allele frequencies and effect sizes for the specified
660 target are first read from each of the K user-specified weight files. The SNPs are then matched to
661 SNPs in the validation genotype data and filtered to exclude effect sizes of SNPs for which the
662 difference between the MAF of the genotype data and the MAF from the corresponding weight
663 file exceeds a user-specified MAF difference threshold. The effect sizes from each weight file
664 are used to initialize K separate instances of the WeightEstimator class. These K
665 WeightEstimator objects are used as base estimators and fit on genotype and expression data
666 from the validation data.
667 A separate script allows users to specify one or more base models trained on the validation data
668 to average with the SR-TWAS model produced in the previous step. The tool will read SNP
669 effect sizes for the specified target from each of these models and output an Avg-valid+SR
670 model averaged from the validation data base model(s) and the model obtained by stacked
671 regression.
672

673 Only SR-TWAS models trained from $K = 2, 4, 6$ base models are presented in this paper. The
674 code was designed to accept any $K \geq 2$, and while the stacked regression script has been
675 primarily tested using $K = 2, 4, 6$ base models, preliminary testing with dummy weight files
676 confirms it can train stacked regression models from $K > 6$ base models.
677

678 **ROS/MAP Reference Panel**

679 The Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP) are two
680 ongoing longitudinal, epidemiologic clinical-pathologic cohort studies of aging and Alzheimer's
681 disease collectively referred to as ROS/MAP[11]. ROS enrolls Catholic nuns, priests, and brothers
682 from religious groups across the United States, primarily from communal living settings[11]. While
683 the similar adult lifestyle of participants allows for more control of potential confounders such as
684 education and socioeconomic status, it simultaneously limits the ability to study such variables[11].
685 MAP was designed to complement and extend studies like ROS by including subjects from a
686 wider range of life experiences, socioeconomic status, and educational attainment and recruits
687 participants primarily from retirement communities in the Chicago area, but also subsidized
688 housing, retirement homes, and through organizations serving minorities and low-income
689 elderly[11]. All participants in both studies are without known dementia and agree to annual
690 clinical evaluations and brain donation upon death[11]. Similarity in study design and data
691 collection procedures allows the ROS and MAP datasets to be merged for use in joint
692 analyses[11,60].
693 Quality-controlled ROS/MAP WGS data for European subjects[60] were used for both the real data
694 application and simulation studies. Transcriptomic data of ROS/MAP samples of brain PFC were
695 profiled by RNA-sequencing (RNA-seq). Gene expression data of Transcripts Per Million (TPM)

696 per sample were provided by Rush Alzheimer's Disease Center. Genes with > 0.1 TPM in ≥ 10
697 samples were considered. Raw gene expression data (TPM) were then log2 transformed and
698 adjusted for age at death, sex, postmortem interval, study (ROS or MAP), batch effects, RNA
699 integrity number scores, cell type proportions (with respect to oligodendrocytes, astrocytes,
700 microglia, neurons), top five genotype principal components, and top probabilistic estimation of
701 expression residuals (PEER) factors[61] by linear regression models. SNPs with minor allele
702 frequency (MAF) > 1%, Hardy-Weinberg p-value $> 10^{-5}$ were analyzed. For each gene, cis-
703 SNPs within the 1Mb of the flanking 5' and 3' ends were used in the imputation models as
704 predictors.
705
706 **GTEx V8 Reference Panel**
707 The Genotype-Tissue Expression (GTEx) project V8 profiles both whole genome sequencing
708 (WGS) genotype data and RNA-seq transcriptomic data for Brain PFC tissue type of n=157
709 donors[10]. Gene expression data of Transcripts Per Million (TPM) per sample per tissue were
710 downloaded from the GTEx portal. Genes with > 0.1 TPM in ≥ 10 samples were considered.
711 Raw gene expression data (TPM) were then log2 transformed and adjusted for age, body mass
712 index (BMI), top five genotype principal components, and top probabilistic estimation of
713 expression residuals (PEER) factors[61]. SNPs with minor allele frequency (MAF) > 1%, Hardy-
714 Weinberg p-value $> 10^{-5}$ were analyzed. For each gene, cis- SNPs within the 1Mb of the
715 flanking 5' and 3' ends were used in the imputation models as predictors.
716
717 **Simulation Study Design**

718 We conducted in depth simulation studies under various scenarios to assess the performance of
719 SR-TWAS, Avg-valid+SR, a Naïve method, and training base models by PrediXcan and
720 TIGAR. We used the real genotype data of gene *ABCA7* from ROS/MAP and GTEx V8 to
721 simulate gene expression and phenotypes. We considered three different settings: (i) Samples
722 from ROS/MAP and GTEx cohorts have the same set of true causal SNPs (i.e., the same genetic
723 architecture). The expression heritability was the same for both ROS/MAP and GTEx V8
724 cohorts. (ii) Samples from ROS/MAP and GTEx cohorts have the same set of true causal SNPs
725 (i.e., the same genetic architecture). The expression heritability for GTEx V8 cohort is only half
726 of the one for ROS/MAP. (iii) Samples from the ROS/MAP cohort were simulated with the same
727 causal SNPs (i.e., eQTL), while samples from the GTEx V8 cohort were simulated with true
728 causal SNPs that were 50% overlapped with the ones for ROS/MAP. The expression heritability
729 was the same for both ROS/MAP and GTEx V8 cohorts.
730 Under each setting, we considered multiple scenarios with varying proportions of causal SNPs
731 ($p_{causal} = (0.001, 0.01, 0.05, 0.1)$) and gene expression heritability (i.e., the proportion of gene
732 expression variation due to genetics, $h_e^2 = (0.1, 0.2, 0.5)$). We randomly selected n=465 training
733 samples with WGS genotype data from ROS/MAP and GTEx V8, respectively. We randomly
734 selected n=400 and n=800 samples with WGS genotype data from ROS/MAP as our validation
735 and test cohorts, respectively. We considered a series of h_p^2 values, the proportion of phenotype
736 variance due to simulated gene expression, in the range of $(0.05, 0.875)$.
737 For each scenario, gene expression $\mathbf{E}_i$ for the ith simulation iteration is generated using the
738 following formula

739 $$\mathbf{E}_i = \gamma_i \mathbf{G}^* \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad \gamma_i = \sqrt{\frac{h_e^2}{\mathrm{Var}(\mathbf{G}^* \boldsymbol{\beta}_{\bullet i})}}, \qquad \boldsymbol{\varepsilon}_i \sim \mathrm{N}\left(0, \sqrt{1 - h_e^2}\right),$$

740 where $\mathbf{G}^*$ denotes the genotype matrix of N_{causal} randomly chosen true causal SNPs for all

741 samples, effect size vector $\boldsymbol{\beta}_{\text{i}}$ was generated from $N(0, I)$, and γ_i is a scale factor chosen to

742 ensure the targeted h_e^2 value. The phenotype vector $\mathbf{Y}_i$ for the ith simulation iteration was

743 generated using the following formula

744 $$\mathbf{Y}_i = \varphi_i \mathbf{E}_i + \boldsymbol{\varepsilon}_i, \qquad \varphi_i = \sqrt{\frac{h_p^2}{\mathrm{Var}(\mathbf{E}_i)}}, \qquad \boldsymbol{\varepsilon}_i \sim \mathrm{N}\left(0, \sqrt{1 - h_p^2}\right)$$

745 where $\mathbf{E}_i$ is the simulated gene expression, and φ_i is a scale factor to ensure the targeted h_p^2

746 value.

747 Two base models per gene were trained by PrediXcan with the GTEx training samples (n=465)

748 (PrediXcan-GTEx), and by TIGAR with the ROS/MAP training samples (n=465) (TIGAR-

749 ROSMAP). SR-TWAS and Naïve models were then obtained by using these trained base

750 models. Validation data ($n = 400$) were used to train SR-TWAS models and filter out gene

751 expression imputation models with 5-fold cross-validation $R^2 < 0.5\%$ in the validation cohort

752 for both SR-TWAS and Naïve models. A validation data base model was trained by TIGAR on

753 the validation data (TIGAR-ROSMAP_valid) to compare results with that of the ensemble

754 models and to obtain a model from the average of the the validation data base models and SR-

755 TWAS (Avg-valid+SR). Test data (n=800) were used for assessing GReX prediction

756 performance and TWAS power. Each causal simulation scenario was repeated for 1,000 times.

757 We compared the performance by SR-TWAS, Avg-valid+SR, the Naïve method, training base

758 models, and validation base models with respect to prediction imputation R^2 in the test data and

759 the power of TWASs.

760 The predicted $\widehat{\mathrm{GReX}}_i$ by each trained gene expression imputation model was used to calculate

761 expression prediction R^2, which is equivalent to the regression R^2 between profiled and

762 predicted gene expression, given by

763 $$R_{E_i}^2 = \mathrm{Cor}\left(\mathbf{E}_i, \widehat{\mathrm{GReX}}_i\right)^2.$$

764 The power will be given by the proportion of simulation iterations that have TWAS p-value $<$

765 2.5×10^6 out of a total of 1,000 simulation iterations.

766

767 **Data availability**

768 All ROS/MAP data analyzed in this study are de-identified and available to any qualified

769 investigator with application through the Rush Alzheimer's Disease Center Research Resource

770 Sharing Hub, https://www.radc.rush.edu, which has descriptions of the studies and available

771 data. GTEx V8 data are available from dbGaP with accession phs000424.v8.p2. TIGAR base

772 models trained from GTEx V8 are available at Synapse

773 https://www.synapse.org/TIGAR_V2_Resource_GTExV8. PrediXcan base models trained from

774 GTEx V8 are available from https://predictdb.org/. GWAS summary data of AD are available

775 from https://ctg.cncr.nl/software/summary_statistics, and GWAS summary data of PD are

776 available from https://bit.ly/2ofzGrk. TIGAR and PrediXcan base models, Avg-valid+SR, SR-

777 TWAS, and Naïve models trained from ROS/MAP brain PFC tissue in this study, SR-TWAS

778 and Avg-valid+SR models trained from GTEx brain substantia nigra tissue in this study, and all

779 TWAS summary statistics will be made available at SYNAPSE when the manuscript is accepted

780 for publication.

781

782 **Code availability**

783 The SR-TWAS tool, including the Naïve and Avg-valid+SR methods, is publicly-available on
784 GitHub, https://github.com/yanglab-emory/SR-TWAS. Code for replicating analyses described
785 in this paper is available at https://github.com/rndparr/SR-TWAS_analysis.

786

787 **References**

788 1. Feng, H. *et al.* Transcriptome-wide association study of breast cancer risk by estrogen-
789 receptor status. *Genet. Epidemiol.* **44**, 442–468 (2020).
790 2. Kar, S. *et al.* Pleiotropy-guided transcriptome imputation from normal and tumor tissues
791 identifies new candidate susceptibility genes for breast and ovarian cancer. *bioRxiv* (2020)
792 doi:10.1101/2020.04.23.043653.
793 3. Strunz, T., Lauwen, S., Kiel, C., Hollander, A. den & Weber, B. H. F. A transcriptome-
794 wide association study based on 27 tissues identifies 106 genes potentially relevant for
795 disease pathology in age-related macular degeneration. *Scientific Reports* **10**, 1584 (2020).
796 4. Wu, C. *et al.* Transcriptome-wide association study identifies susceptibility genes for
797 rheumatoid arthritis. *Arthritis Res Ther* **23**, (2021).
798 5. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association
799 studies. *Nat Genet* **51**, 592–599 (2019).
800 6. Nagpal, S. *et al.* TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation
801 Enhances Gene Mapping of Complex Traits. *The American Journal of Human Genetics*
802 **105**, 258–266 (2019).
803 7. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene
804 expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1–20 (2018).
805 8. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
806 transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
807 9. Parrish, R. L., Gibson, G. C., Epstein, M. P. & Yang, J. TIGAR-V2: Efficient TWAS tool
808 with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8. *Human*
809 *Genetics and Genomics Advances* **3**, 100068 (2022).

10. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

11. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis* **64**, S161–S189 (2018).

12. Hu, Y. *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat Genet* **51**, 568–576 (2019).

13. Liu, A. E. & Kang, H. M. Meta-imputation of transcriptome from genotypes across multiple datasets by leveraging publicly available summary-level data. *PLOS Genetics* **18**, e1009571 (2022).

14. Wolpert, D. H. Stacked generalization. *Neural Networks* **5**, 241–259 (1992).

15. Breiman, L. Stacked regressions. *Mach Learn* **24**, 49–64 (1996).

16. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).

17. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biology* **22**, 49 (2021).

18. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).

19. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *The American Journal of Human Genetics* **100**, 473–487 (2017).

20. Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl Psychiatry* **8**, 99 (2018).

832 21. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways
833 influencing Alzheimer's disease risk. *Nat Genet* **51**, 404–413 (2019).
834 22. Nazarian, A., Yashin, A. I. & Kulminski, A. M. Genome-wide analysis of genetic
835 predisposition to Alzheimer's disease and related sex disparities. *Alz Res Therapy* **11**, 1–21
836 (2019).
837 23. Gockley, J. *et al.* Multi-tissue neocortical transcriptome-wide association study implicates 8
838 genes across 6 genomic loci in Alzheimer's disease. *Genome Medicine* **13**, 76 (2021).
839 24. Jing, Q. *et al.* A Comprehensive Analysis Identified Hub Genes and Associated Drugs in
840 Alzheimer's Disease. *BioMed Research International* **2021**, e8893553 (2021).
841 25. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping, and integrative
842 prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* **53**, 392–402 (2021).
843 26. Deming, Y. *et al.* The MS4A gene cluster is a key modulator of soluble TREM2 and
844 Alzheimer's disease risk. *Science Translational Medicine* **11**, eaau2291 (2019).
845 27. Shigemizu, D. *et al.* Ethnic and trans-ethnic genome-wide association studies identify new
846 loci influencing Japanese Alzheimer's disease risk. *Transl Psychiatry* **11**, 1–10 (2021).
847 28. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for
848 Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet*
849 *Neurology* **18**, 1091–1102 (2019).
850 29. Kia, D. A. *et al.* Identification of Candidate Parkinson Disease Genes by Integrating
851 Genome-Wide Association Study, Expression, and Epigenetic Data Sets. *JAMA Neurology*
852 **78**, 464–472 (2021).
853 30. Yao, S. *et al.* A transcriptome-wide association study identifies susceptibility genes for
854 Parkinson's disease. *npj Parkinsons Dis.* **7**, 1–8 (2021).

855 31. Sagi, O. & Rokach, L. Ensemble learning: A survey. *WIREs Data Mining and Knowledge*
856 *Discovery* **8**, e1249 (2018).
857 32. Tang, H. & Harte, M. Investigating Markers of the NLRP3 Inflammasome Pathway in
858 Alzheimer's Disease: A Human Post-Mortem Study. *Genes (Basel)* **12**, 1753 (2021).
859 33. Lambert, J.-C. *et al.* Genome-wide association study identifies variants at CLU and CR1
860 associated with Alzheimer's disease. *Nat Genet* **41**, 1094–1099 (2009).
861 34. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new
862 risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet* **51**, 414–430
863 (2019).
864 35. Corbett, B. F. *et al.* ΔFosB regulates gene expression and cognitive dysfunction in a mouse
865 model of Alzheimer's disease. *Cell Rep* **20**, 344–355 (2017).
866 36. Shi, G. *et al.* Functional alteration of PARL contributes to mitochondrial dysregulation in
867 Parkinson's disease. *Human Molecular Genetics* **20**, 1966–1974 (2011).
868 37. Subrahmanian, N. & LaVoie, M. J. Is there a special relationship between complex I
869 activity and nigral neuronal loss in Parkinson's disease? A critical reappraisal. *Brain*
870 *Research* **1767**, 147434 (2021).
871 38. Storm, C. S. *et al.* Finding genetically-supported drug targets for Parkinson's disease using
872 Mendelian randomization of the druggable genome. *Nat Commun* **12**, 7342 (2021).
873 39. Murthy, M. N. *et al.* Increased brain expression of GPNMB is associated with genome wide
874 significant risk for Parkinson's disease on chromosome 7p15.3. *Neurogenetics* **18**, 121–133
875 (2017).
876 40. Haskó, G., Linden, J., Cronstein, B. & Pacher, P. Adenosine receptors: therapeutic aspects
877 for inflammatory and immune diseases. *Nat Rev Drug Discov* **7**, 759–770 (2008).

878 41. Sun, Y. & Huang, P. Adenosine A2B Receptor: From Cell Biology to Human Diseases.
879 *Frontiers in Chemistry* **4**, (2016).
880 42. Jenner, P. Chapter Three - An Overview of Adenosine A2A Receptor Antagonists in
881 Parkinson's Disease. in *International Review of Neurobiology* (ed. Mori, A.) vol. 119 71–
882 86 (Academic Press, 2014).
883 43. Guerreiro, S., Privat, A.-L., Bressac, L. & Toulorge, D. CD38 in Neurodegeneration and
884 Neuroinflammation. *Cells* **9**, 471 (2020).
885 44. Mogil, L. S. *et al.* Genetic architecture of gene expression traits across diverse populations.
886 *PLOS Genetics* **14**, e1007586 (2018).
887 45. Luningham, J. M. *et al.* Bayesian Genome-wide TWAS Method to Leverage both cis- and
888 trans-eQTL Information through Summary Statistics. *The American Journal of Human*
889 *Genetics* **107**, 714–726 (2020).
890 46. Yang, C. *et al.* CoMM: a collaborative mixed model to dissecting genetic contributions to
891 complex traits by leveraging regulatory information. *Bioinformatics* **35**, 1644–1652 (2019).
892 47. Yuan, Z. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic
893 Mendelian randomization in transcriptome-wide association studies. *Nature*
894 *Communications* **11**, 3861 (2020).
895 48. Yuan, Z. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic
896 Mendelian randomization in transcriptome-wide association studies. *Nat Commun* **11**, 3861
897 (2020).
898 49. Rao, J. N. K. & Subrahmaniam, K. Combining Independent Estimators and Estimation in
899 Linear Regression with Unequal Variances. *Biometrics* **27**, 971–990 (1971).

50. Efron, B. & Morris, C. Combining Possibly Related Estimation Problems. *Journal of the Royal Statistical Society* **35**, 379–421 (1973).

51. Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat Genet* **50**, 968–978 (2018).

52. Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. *PLoS Genet* **17**, e1009398 (2021).

53. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718–719 (2011).

54. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference (SciPy 2010)* 56–61 (2010). doi:10.25080/Majora-92bf1922-00a.

55. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

56. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272 (2020).

57. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in *Proceedings of the 9th Python in Science Conference (SciPy 2010)* 92–96 (2010). doi:10.25080/Majora-92bf1922-011.

58. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

59. Buitinck, L. *et al.* API design for machine learning software: experiences from the scikit-learn project. in *Proceedings of the European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases (ECMPKDD'13)* 108--122 (2013).

923 60. De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and
924 Alzheimer's disease research. *Sci Data* **5**, (2018).
925 61. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of
926 expression residuals (PEER) to obtain increased power and interpretability of gene
927 expression analyses. *Nat Protoc* **7**, 500–507 (2012).
928