Supplementary Materials for

A deep generative framework for nonlinear association analysis between brain imaging and genetic data of Alzheimer's diseases

Jia-Ni Li[1], Shao-Wu Zhang[1]*, Qing-Qing Zhang[1], Yan-Rui Qiang[1], Ting-He Zhang[2]*

1. MOE Key Laboratory of Information Fusion Technology, School of Automation, Northwestern Polytechnical University, Xi'an 710072, China.
2. Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA,15261, USA.

* Corresponding authors: zhangsw@nwpu.edu.cn, Tinghe.Zhang@pitt.edu

Appendix A: Related works

A.1 Brain imaging genetics association analysis with nonlinear models

The association between brain imaging QTs and SNPs is highly complex, rendering simple linear models inadequate for its analysis [1]. Consequently, some nonlinear imaging genetic methods have been developed to analyze the associations between brain imaging QTs and SNPs. For instance, Ge et al. [2] proposed a Kernel machine-based Canonical Correlation Analysis (KCCA) method to model the joint and epistatic effect of a collection of SNPs, testing the nonlinear interactions between sets of variables in a flexible framework to detect the effects of the interactions between candidate AD risk genes and a collection of cardiovascular disease (CVD) risk factors on hippocampal volume measurements derived from structural brain magnetic resonance imaging (MRI) scans, finding that CR1 and EPHA1 genes influence AD-related neurodegeneration in the presence of CVD risks. However, due to the curse of dimensionality and limitations in the dynamic selection of kernel functions and parameters, KCCA is mainly applied in feature fusion and classification. Huang et al. [3] presented a nonlinear association model (i.e., T-GSRAM) to identify nonlinear associations between brain imaging QTs and SNPs for detecting the potential AD biomarkers. Assuming that the effects of each SNP on QT are regarded as a smooth function (it can be a nonlinear transformation) of time, T-GSRAM first constructed the temporal group sparsity regression model and the additive model to identify associations between longitudinal imaging QTs and SNPs, then used longitudinal QTs

to identify the trajectory of imaging genetic patterns over time. Similarly, Wang et al. [4] first built a connection matrix, then used PageRank to estimate the importance of different brain regions as the brain region features. Finally, they developed a deep self-reconstruction sparse canonical correlation analysis (DS-SCCA) method to identify the associations between SNPs and brain network QTs. However, existing these methods fail to yield a direct association matrix. To address this gap, we designed a nonlinear association analysis (SQAA) module in our SQVE framework to directly identify SNPs associated with specific QT, facilitating the discovery of AD biomarkers.

A.2 Deep generative-based models in brain imaging genomics

The deep generative models in brain imaging genomics mainly include generative adversarial network (GAN) [5], auto encoder (AE) [6], variational autoencoder (VAE) [7], and their variants. For instance, Ko et al. [8] proposed a deep generative and discriminative learning framework to learn the distributional properties of the QTs and SNPs, explore the nonlinear relation between QTs and SNPs, facilitate the identification of Alzheimer's Disease (AD) biomarkers, diagnosis, and cognitive score prediction. Similarly, Ghosal et al. [9] proposed a coupled generative-discriminative framework to explore the interconnectedness of different brain regions and identify AD biomarkers from imaging and genetics data. In Ghosal's framework, the generative module is used to capture the brain activations and the genetic features with dictionary learning, and the discriminative module is used to track the AD disease status with logistic regression, while the classification module guides the generative process to find a low dimensional space for diagnosing AD patients. Dolci et al. [10] first incorporated two pretrained cycle generative adversarial networks (GANs) to perform the knowledge transfer to fill in missing data, and then developed a deep generative imaging genomics framework for AD prediction. However, GAN-based methods have a complex generation process, and provide poor biological explanations for the QTs and SNPs.

Compared to the GAN, the AE/VAE uses neural networks as the encoder and decoder architectures to obtain the shared latent space for learning the underlying

relationships/dependencies among different modalities, and integrating the muti-modality data without any prior knowledge or statistical model [11]. Therefore, AE has been applied in brain imaging genomics. For instance, Li et al. [12] developed deep principal correlated auto-encoders (DPCAE) to integrate fMRI imaging and genomics data in correlation analysis, classification, and clustering tasks. DPCAE consists of two back propagation neural networks and two multilayer belief networks, where the principal component analysis is used to learn the multi-modality linear features and the multilayer belief networks are used to learn the multi-modality nonlinear features. However, existing AE-based imaging genetics association methods often overlook modality differences, whereas VAE takes these differences into account to effectively integrate multi-modal data.

Considering that VAEs can effectively learn the structure/characteristics of modal data and the association among different modalities from the learned shared latent distribution, we designed a dual-encoder VAE module to integrate multimodal information, learning the underlying relationships between SNPs and QTs. In addition, in view of that the perturbation-based method can identify the effects of input data on the outputs of VAE model by perturbing the input data, we built the SNPs-QTs association analysis (SQAA) module on the trained dual-encoder VAE to analyze the nonlinear association between SNPs and QTs by perturbing the input SNPs.

References

[1] X. Wang *et al.*, "Quantitative trait loci identification for brain endophenotypes via new additive model with random networks," *Bioinformatics,* vol. 34, no. 17, pp. i866-i874, Sep 1, 2018.

[2] T. Ge *et al.*, "A kernel machine method for detecting effects of interaction between multidimensional variable sets: an imaging genetics application," *Neuroimage,* vol. 109, pp. 505-514, Apr 1, 2015.

[3] M. Huang, X. Chen, Y. Yu, H. Lai, and Q. Feng, "Imaging Genetics Study Based on a Temporal Group Sparse Regression and Additive Model for Biomarker Detection of Alzheimer's Disease," *IEEE Trans Med Imaging,* vol. 40, no. 5, pp. 1461-1473, May, 2021.

[4] M. Wang, W. Shao, X. Hao, S. Huang, and D. Zhang, "Identify connectome between genotypes and brain network phenotypes via deep self-reconstruction sparse canonical correlation analysis," *Bioinformatics,* vol. 38, no. 8, pp. 2323-2332, 2022.

[5] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems,* vol. 27, 2014.

[6] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Autoencoders," *Machine learning*, pp. 193-208: Elsevier, 2020.

[7] D. P. Kingma, and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[8] W. Ko, W. Jung, E. Jeon, and H. I. Suk, “A Deep Generative-Discriminative Learning for Multimodal Representation in Imaging Genetics,” *IEEE Trans Med Imaging,* vol. 41, no. 9, pp. 2348-2359, Sep, 2022.

[9] S. Ghosal *et al.*, “A generative-discriminative framework that integrates imaging, genetic, and diagnosis into coupled low dimensional space,” *Neuroimage,* vol. 238, pp. 118200, Sep, 2021.

[10] G. Dolci, M. A. Rahaman, J. Chen, K. Duan, Z. Fu, A. Abrol, G. Menegaz, and V. D. Calhoun, “A deep generative multimodal imaging genomics framework for Alzheimer's disease prediction,” in 2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE), 2022, pp. 41-44.

[11] J. N. Nissen *et al.*, “Improved metagenome binning and assembly using deep variational autoencoders,” *Nat Biotechnol,* vol. 39, no. 5, pp. 555-560, May, 2021.

[12] G. Li *et al.*, “Deep Principal Correlated Auto-Encoders With Application to Imaging and Genomics Data Integration,” *IEEE Access,* vol. 8, pp. 20093-20107, 2020.

Appendix B: Methods and Results

B.1 Generating synthetic data

The synthetic datasets include the paired synthetic SNPs data $\mathcal{T}$ and synthetic QTs data $\mathcal{H}$ for subjects. The details of generating synthetic data are described as follows.

B.1.1 Generating the synthetic SNPs data $\mathcal{T}$

The synthetic SNPs data is defined as $\mathcal{T}=[\mathcal{T}_{\text{normal}},\mathcal{T}_{\text{disease}}]$, where $\mathcal{T}_{\text{normal}}$ is the synthetic SNPs data from the normal control (NC) subjects and $\mathcal{T}_{\text{disease}}$ is the synthetic SNPs data from the Alzheimer’s disease (AD) subjects.

For the NC subjects, we first extract the n_{ts} risk SNPs from the ADNI dataset, then calculate the mutation rate $\varphi_n(g_n)$ of each $(g_n)^{\text{th}}$ subject by adopting the following formula, $\varphi_n(g_n)=\{\frac{1}{n_{ts}}\sum_{ts=1}^{n_{ts}}\sum_{i=1}^{2}t_{g_n}^{(ts)_i} \mid t_{g_n}^{(ts)_1}=1,\ t_{g_n}^{(ts)_2}=2,\ g_n=1,2,...,G_n\}$, where $t_{g_n}^{ts}$ is the value of $(ts)^{\text{th}}$ SNP in $(g_n)^{\text{th}}$ subject, G_n is the number of NC subjects. We assume the mutation rate φ_n following a normal distribution, *i.e.* $\varphi_n \sim \mathcal{N}(\mu_n,\sigma_n^2)$ (p-value = 2.2E-16) as shown in Fig. S2 (a), where

$\mu_n = \frac{1}{G_n}\sum_{g_n=1}^{G_n}\varphi_n(g_n)$ and $\sigma_n^2 = \frac{1}{G_n}[\varphi_n(g_n)-\mu_n]^2$. Using φ_n, we synthetic SNPs data $t_{m_1,v}$ of m_1 SNP in subject v following a Bernoulli distribution, *i.e.* $t_{m_1,v} \sim Bernoulli(\varphi_n)$. In this work, we generate 500 NC subjects, each with 1000 SNPs.

For the AD subjects, we first extract n_{ts} risk SNPs from ADNI dataset, and then calculate the mutation rate $\varphi_d(g_d)$ of each $(g_d)^{\text{th}}$ subject by adopting the following formula, $\varphi_d(g_d) = \{\frac{1}{n_{ts}}\sum_{ts=1}^{n_{ts}}\sum_{i=1}^{2} t_{g_d}^{(ts)_i} \mid t_{g_d}^{(ts)_1}=1,\ t_{g_d}^{(ts)_2}=2,\ g_d=1,2,...,G_d\}$, where $t_{g_d}^{ts}$ is the value of $(ts)^{\text{th}}$ SNP in $(g_d)^{\text{th}}$ subject, G_d is the number of AD subjects. We assume the mutation rate φ_n following a normal distribution, *i.e.* $\varphi_d \sim \mathcal{N}(\mu_d, \sigma_d^2)$ (p-value = 2.7E-18) as shown in Fig. S2 (b). Using φ_d, we synthetic SNPs data $t_{m_2,d}$ of m_2 SNP in subject d following a Bernoulli distribution, *i.e.* $t_{m_2,d} \sim Bernoulli(\varphi_d)$. In this work, we generate 500 AD subjects, each with 1000 SNPs.

In this work, $n_{ts}=2000$, $G_n=223$, $\mu_n=0.146$, $\sigma_n^2=0.0292$, $G_d=277$, $\mu_d=0.481$, $\sigma_d^2=0.0444$.

B.1.2 Calculating the related mean μ_1 and nonrelated mean μ_2

First, based on the empirical distribution as shown in Fig. S2 (c), we obtain the related SNPs-QTs distribution $\beta_1 \sim \mathcal{N}(-0.3, 0.1)$ from AD subjects, the unrelated SNPs-QTs distribution $\beta_2 \sim \mathcal{N}(0, 0.02)$ from NC subjects, respectively.

Second, we assume the random effect variable $b_{qt,v}$ of $(qt)^{\text{th}}$ QT in subject v following a Gaussian distribution, *i.e.* $b_{qt,v} \sim \mathcal{N}(\mu_{b'}, \sigma_{b'}^2)$, and then use the generalized linear mixed model (glmmTMB) to estimate the parameters $\mu_{b'}$ and $\sigma_{b'}^2$ for each subject based on the matrix composed of QT n_r $(r=1,2,...,119)$ and the top selected n_{ts} risk SNPs (Fig. S2 d-e). In this work, $\mu_{b'}=3.07$ and $\sigma_{b'}^2=1.48$.

Third, after generating the synthetic NC SNPs data $t_{m_1,v}$, the synthetic AD SNPs data $t_{m_2,d}$, and obtaining the related regression coefficients β_1, the unrelated

regression coefficients β_2, and the random effect variable $b_{qt,v}$, we calculate the normalized expected related mean $\mu_{1,d} = e^{\beta_1 t_{m_2,d} + b_{qt,v}}$, and the unrelated mean $\mu_{2,v} = e^{\beta_2 t_{m_1,v} + b_{qt,v}}$, $\mu_{2,d} = e^{\beta_2 t_{m_2,d} + b_{qt,v}}$.

B.1.3 Calculating random variance σ^2

After getting the QT matrix $\mathrm{M_R} \in \mathbb{R}^{G_n \times 119}$ for NC subjects, where G_n is the number of NC subjects, we generate the scatter plot of the mean μ and the variance σ^2 based on the $\mathrm{M_R}$ as shown in Fig. S2 (f). From Fig. S2 (f), we can see that the variance σ^2 follows a binomial distribution, *i.e.* $\sigma^2 \sim binomial\ distribution(\mu, 2),\ i = 1, 2$. Therefore, the random variance corresponding to μ_1 is $\sigma_1^2 \sim binomial\ distribution(\mu_1, 2)$ and the random variance corresponding to μ_2 is $\sigma_2^2 \sim binomial\ distribution(\mu_2, 2)$.

B.1.4 Generating the synthetic QTs data

The synthetic QTs data is defined as $\mathcal{H} = [\mathcal{H}_{\text{normal}}, \mathcal{H}_{\text{disease}}]$. $\mathcal{H}_{\text{normal}}$ is generated by $h_{normal} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{H}_{\text{disease}}$ is generated by $h_{disease} \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

For the AD subjects, let two sparse vectors: $\boldsymbol{Q} \in \mathbb{R}^{s_a \times 1}$ and $\boldsymbol{T} \in \mathbb{R}^{s_b \times 1}$ represent AD-related QTs and SNPs, respectively, s_a denotes the dimension number for QTs randomly selected from 30% of the 119 QTs, and s_b denotes the number of SNPs related to QTs of AD subjects. In this work, we set $s_a = 36$, and s_b to 10, 20, 25. For example, we set $s_a = 36, s_b = 10$, and $QT_{18}, QT_{19}, ..., QT_{53}$ are the QTs related with AD, i.e., $\underbrace{QT_1, QT_2, ..., QT_{17}}_{\text{unrelated QTs}}, \underbrace{QT_{18}, QT_{19}, ..., QT_{53}}_{\text{related QTs}}, \underbrace{QT_{54}, QT_{55}, ..., QT_{119}}_{\text{unrelated QTs}}$

where, $QT_{18} : \underbrace{\beta_1^{18}, \beta_2^{18}, ..., \beta_{10}^{18}}_{\text{10 related SNPs}}, \underbrace{..., \beta_{1000}^{18}}_{\text{unrelated SNPs}}$,

$QT_{19} : \underbrace{\beta_1^{19}, \beta_2^{19}, ..., \beta_{10}^{19}}_{\text{unrelated SNPs}}, \underbrace{\beta_{11}^{19}, \beta_{12}^{19}, ..., \beta_{20}^{19}}_{\text{10 related SNPs}}, \underbrace{..., \beta_{1000}^{19}}_{\text{unrelated SNPs}}$, and so on.

B.2 Effects of different neural network architectures and parameters on SQVE

B.2.1 Effect of different neural network architectures on SQVE

To investigate the performance of SQVE with different neural network architectures, we compared CNN used in SQVE with other four neural network architectures on synthetic dataset25, such as Attention-Net (AN), Recurrent Neural Network (RNN), Deep Neural Network (DNN), and Long Short-Term Memory (LSTM). Table S4 shows the results of CNN used in SQVE with different neural network architectures on synthetic dataset25.

The results from clearly indicate that CNN outperform the other four neural network architectures of AN, RNN, DNN and LSTM. For example, the CBR(QT) of CNN is 3.83%, 3.48%, and 0.77% higher than that of the AN, RNN, and DNN, respectively. The CBR(SNP) of CNN is 19.64%, 19.38%, 13.64%, and 5.24% higher than that of the AN, RNN, DNN, and LSTM, respectively. These results demonstrate that CNN is powerful in processing high-dimensional genome data. The reason may be that CNN can effectively extract the genetic features from the high-dimensional genome data, thus improving the classification performance. Given these findings, the CNN architecture has been chosen for the SDRS module.

B.2.2 Effect of fragment size on SQVE

To investigate the performance of SQVE with different fragment sizes, we set different fragment sizes (e.g.,10, 20, 30, **...** , 140 SNPs) on synthetic dataset25. The results of SQVE with different fragment sizes are shown in Fig. S2. As shown in Fig. S2, we can see that the CBR(QT) of SQVE gradually increases with the enlargement of fragment size in the range of 10 to 40 SNPs, then slowly decreases in the range of 40 to 140 SNPs. The CBR(SNP) of SQVE gradually increases with the enlargement of fragment size in the range of 10 to 40 SNPs, then gradually decreases in the scope of 40 to 140 SNPs. These results suggest that the selection of fragment size has a certain impact on SQVE.

Consequently, in this work, we segmented the entire genome into 3,397 fragments, with each fragment containing 40 SNPs.

B.2.3 Effect of top SNPs number n_{ts} on SQVE

To explore the performance of SQVE with different top SNPs number n_{ts}, we set different numbers of top SNPs on synthetic dataset25. Fig. S3 shows the CBR(QT) and CBR(SNP) of SQVE with different numbers of top SNPs on synthetic dataset25. As shown in Fig. S3, we can see that when the number of top SNPs is 100, the CBR(QT) and CBR(SNP) of SQVE reach their highest values. Specifically, CBR(SNP) sharply increases in the range of 20 to 100 SNPs, then gradually decreases in the range of 120 to 500 SNPs. CBR(QT) gradually increases in the range of 20 to 50 SNPs, then slowly decreases in the range of 50 to 500 SNPs. These results show that the number of top SNPs has an impact to the performance of SQVE, especially on CBR(SNP) with the greatest influence. Thus, in this work, we set the number of top SNPs to 100.

B.3 Evaluation Metrics

In this work, we use two metrics: the Causal Biomarker Rate (CBR) and Rank Biased Overlap (RBO), to evaluate the performance of SQVE. The CBR(QT) and CBR(SNP) are defined as:

$$\mathrm{CBR(QT)} = n_a^{qt} / s_a \tag{1}$$

$$\mathrm{CBR(SNP)} = n_b^{snp} / s_b \tag{2}$$

where, n_a^{qt} is the number of the detected QTs from sMRIs, s_a is the QT number from sMRIs; n_b^{snp} the number of the detected SNPs from genetic data, s_b is the SNP number from genetic data.

RBO is typically utilized to measure the similarity between two ranking lists. $RBO(QT)$ and $RBO(SNP)$ are defined as:

$$RBO(QT) = RBO(M_{QT}^{G}, M_{QT}^{P}, \psi_{QT}) = (1-\psi_{QT}) \sum_{d_1=1}^{\infty} \psi_{QT}^{d_1-1} \cdot A_{d_1} \quad (3)$$

$$RBO(SNP) = RBO(M_{SNP}^{G}, M_{SNP}^{P}, \psi_{SNP}) = (1-\psi_{SNP}) \sum_{d_2=1}^{\infty} \psi_{SNP}^{d_2-1} \cdot A_{d_2} \quad (4)$$

where M_{QT}^{G} and M_{QT}^{P} denote the ground truth QT list and the predicted QT list, respectively; M_{SNP}^{G} and M_{SNP}^{P} denote the ground truth SNP list and the predicted SNP list, respectively; ψ is the weighted parameter in (0, 1), here we set $\psi_{QT} = \psi_{SNP} = 0.1$ (Fig. S6), which implies the strong emphasis on the top-ranked elements; $A_{d_1} = |M_{1:d_1}^{G} \cap M_{1:d_1}^{P}| / d_1$ denotes the agreement between M_{QT}^{G} and M_{QT}^{P}, and $A_{d_2} = |M_{1:d_2}^{G} \cap M_{1:d_2}^{P}| / d_2$ denotes the agreement between M_{SNP}^{G} and M_{SNP}^{P}; d is the depth of the ranking list, here we set $d_1 = 50$ and $d_2 = 100$ for QT list and SNP list, respectively. The RBO values range between 0 and 1, where a higher value indicates a better model performance.

Appendix C: Tables and Figures

Tables S1, S4-S6

Table S1 Statistics of demographic, neuropsychological and cognitive assessment of 500 subjects. MMSE (Mini-Mental State Examination), CDR (Clinical Dementia Rating Scale), ADAS11 (Alzheimer's Disease Assessment Scale-11 items score), FAQ (functional activities questionnaire).

Characteristic	Number	Gender(M/F)	Education	Age	MMSE	CDR	ADAS11	FAQ
NC	223	105/118	16.61±2.63	74.19±6.42	29.07±1.29	0.01±0.09	0.02±0.08	0.19±1.05
AD	277	167/110	16.02±2.65	73.08±7.67	24.76±2.33	1.82±1.03	14.72±7.96	7.92±6.52

Table S4 The results of CNN and other four neural network architectures on synthetic dataset25 in terms of CBR(QT), CBR(SNP).

Neural network architecture	CBR(QT) (%)	CBR(SNP) (%)
AN	96.17±0.02	74.06±2.15
RNN	96.52±0.01	74.32±2.08
DNN	99.23±0.01	80.06±2.01

LSTM	100.00	88.46±1.19
CNN	100.00	93.70±1.04

Table S5 The optimal value and the selection range of hyperparameters on the synthetic datasets

Module	Name	The range of hyperparameters
ITSS	The number of fragments of SNP	10; 20; 30; 40*; 60; 80; 100; 120; 140
	Batch size	1*; 2
	Learning rate	1e-2; 1e-3; 1e-4*; 1e-5
	Epoch	200; 300; 400; 500*; 600; 700
Dual-Encoder VAE	The number of hidden layers	1*; 2; 3
	The sizes of hidden layers	1000; 2000; 3000*; 4000
	The sizes of latent space	20; 50; 100; 200; 300; 500*
	Dropout rate	0.1*; 0.2; 0.3
	Batch size	2; 5*; 10
	Learning rate	1e-2; 1e-3; 1e-4*; 1e-5
	Epoch	500; 700; 1000*; 1500; 2000
	The size of top SNPs	20; 50; 100*; 120; 150; 200; 220; 250; 300; 400; 500

"*" denotes that this hyperparameter is selected as the hyperparameter of our SQVE model.

Table S6 The optimal value and the selection range of hyperparameters on the real dataset

Module	Name	The range of hyperparameters
ITSS	The number of fragments of SNP	10; 20; 30; 40*; 60; 80; 100; 120; 140
	Batch size	1*; 2
	Learning rate	1e-2; 1e-3; 1e-4*; 1e-5
	Epoch	200; 300; 400; 500*; 600; 700
Dual-Encoder VAE	The number of hidden layers	1*; 2; 3
	The sizes of hidden layers	1000; 2000; 3000*; 4000
	The sizes of latent space	20; 50; 100*; 200; 300; 500
	Dropout rate	0.1*; 0.2; 0.3
	Batch size	2; 5*; 10
	Learning rate	1e-2; 1e-3; 1e-4*; 1e-5
	Epoch	500; 700; 1000*; 1500; 2000
	The size of top SNPs	100; 200; 300; 400; 500*; 600; 700; 800; 1000; 3000; 5000; 7000; 10000

"*" denotes that this hyperparameter is selected as the hyperparameter of our SQVE model

Figures S1-S5

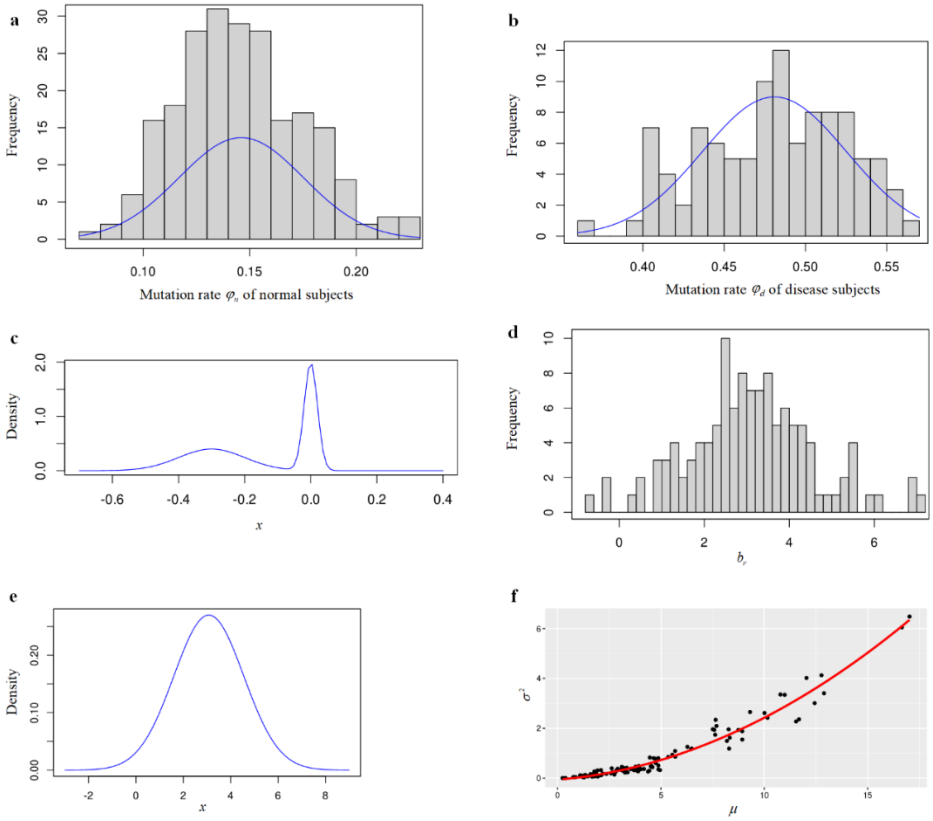

Fig. S1. a. The mutation rate φ_n of NC subjects from ADNI dataset and its gaussian distribution map (blue line). b. The mutation rate φ_d of AD subjects from ADNI dataset and its gaussian distribution map (blue line). c. The gaussian distribution map of SNPs related to QT ($\beta_1 \sim \mathcal{N}(-0.3, 0.1)$) and SNP unrelated to QT ($\beta_2 \sim \mathcal{N}(0, 0.02)$). d. The histogram of the random effect variable b_r. e. The gaussian distribution map based on b_r. f. The binomial distribution map of random variance σ^2 based on the μ.

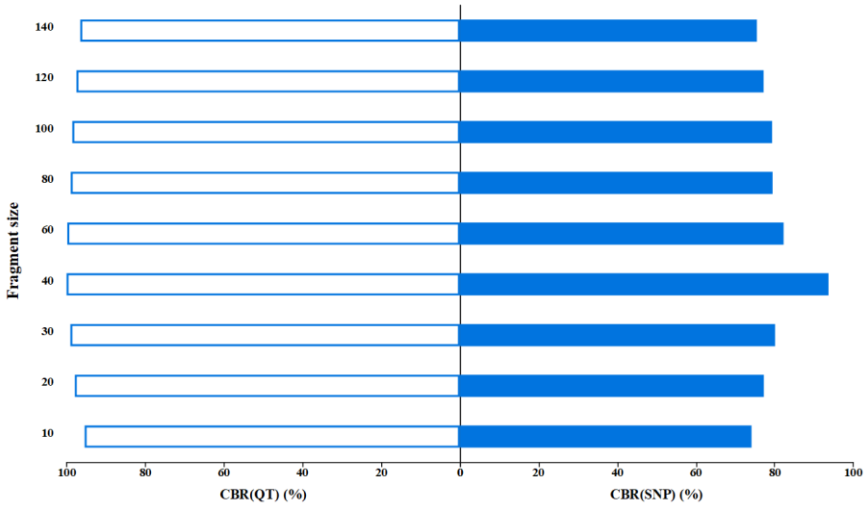

Fig. S2. Results of SQVE with different fragment sizes on synthetic dataset25

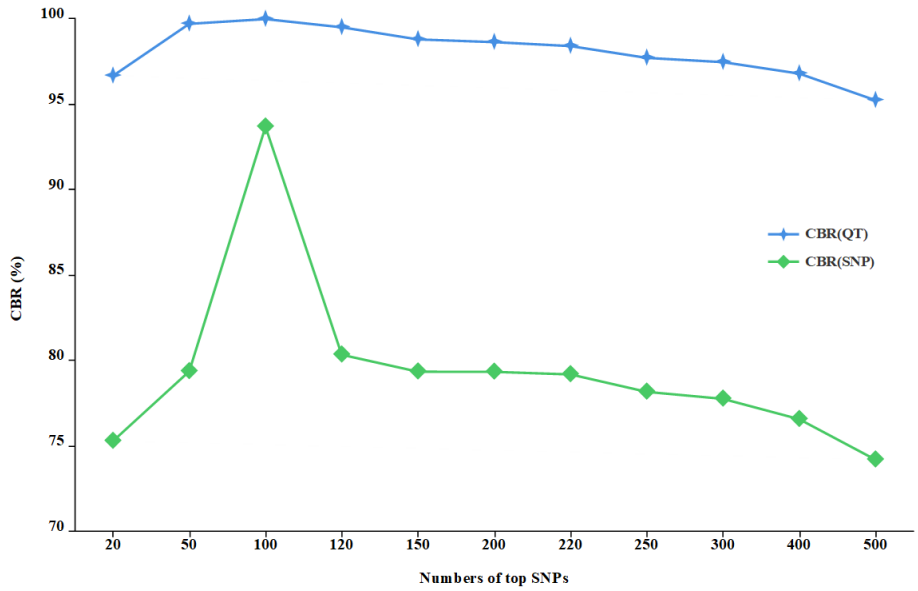

Fig. S3. Results of SQVE with different numbers of top SNPs on synthetic dataset25

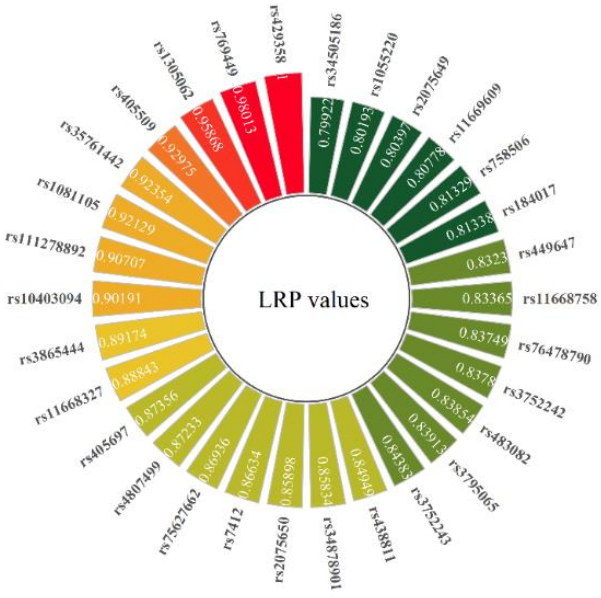

Fig. S4. Visualization of TOP 30 AD-related SNPs and their LRP values on SDRS module.

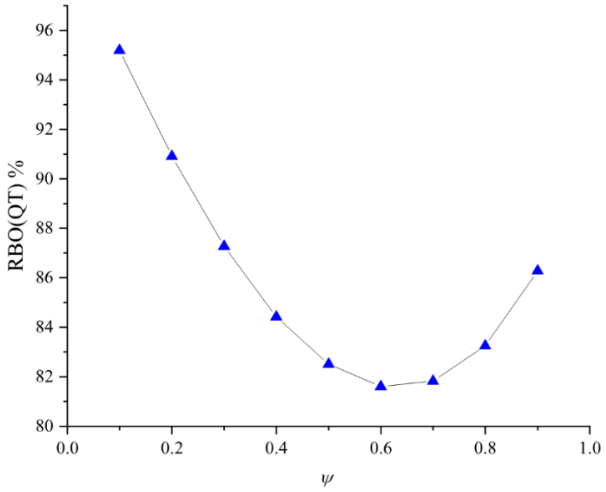

Fig. S5. Results of RBO(QT) with different ψ on the synthetic dataset25.