Supplemental file for "A high-dimensional omnibus test for set-based association analysis"

Haitao Yang[1,2,3], Xin Wang[1], Zechen Zhang[1,2], Fuzhao Chen[1], Hongyan Cao[4], Lina Yan[1,2], Xia Gao[1,2], Hui Dong[5], and Yuehua Cui[6*]

[1]*Division of Health Statistics, School of Public Health, Hebei Medical University, Shijiazhuang, PR China*

[2]*Hebei Key Laboratory of Environment and Human Health, Hebei Province, Shijiazhuang 050017, P.R. China*

[3]*Hebei Key Laboratory of Forensic Medicine, Hebei Province, Shijiazhuang 050017, P.R. China*

[4]*Division of Mathematics and Health Statistics, Shanxi Medical University, Taiyuan, PR China.*

[5]*Department of Neurology, Second Hospital of Hebei Medical University, Shijiazhuang, P.R. China*

[6]*Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA*

*Corresponding author: cuiy@msu.edu

This supplemental file includes details about the method (Section I) along with additional simulation and real data analysis results (Section II).

Section I

A brief summary of the p-value combination methods

Fisher's product test may lose power when there are correlated or large p-values. Zaykin et al. (2002) developed the TPM method, which combines p-values from all tests whose significance exceeds some specified threshold τ, commonly set as 0.05 (Zaykin et al. 2002), thus provides more power. However, choosing an appropriate threshold for the TPM is challenging due to its sensitivity to sample size and the number of tests. As a complementary method to TPM, RTP is calculated as the product of the k most significant p-values from all L hypothesis tests and is suitable for detecting a small set of fixed effects among many nulls. This makes RTP particularly useful for genome-wide association scans, as it often has greater power than the TPM by focusing on the combined evidence from a fixed-size subset of hypotheses. Additionally, RTP is suitable for detecting the combined weak effects of minor loci (Dudbridge et al. 2003); but analytic expressions for combined top p-values is usually cumbersome, which makes interpretation and practical implementation complicated and limits its further development. The ART method retains main characteristics of the RTP and is at least as powerful as RTP. ART is substantially simpler to implement than RTP, but requires to choose k beforehand. ART-A is the adaptive version of ART, which searches through a number of candidate values of truncation points and finds an optimal number in terms of combined p-value (Vsevolozhskaya et al., 2019). However, the number of candidate values also needs to be prespecified. Our simulation found that different numbers of

candidate values of truncation points have substantial effect on type I error and power. All these motivate us to propose an improved version of ART-A, termed iART-A, by using the Cauchy combination test (Liu and Xie, 2019).

A brief review of the omnibus test application

Barnett et al. (2017) proposed an omnibus test that integrates the MinP, SKAT, and generalized higher criticism (GHC) methods, each of which works well under certain disease assumptions. By aggregating the three testing results, the omnibus test gives robust results without knowing the true gene action mode. Similar omnibus tests have also been proposed by Liu and Lin (2019) and Cai et al. (2012). Note that these candidate tests are correlated as they are calculated using the same data, making the p-values correlated. Thus, calculating the p-value of an omnibus test statistic must take the correlations into account (Liu and Lin, 2019). Obtaining the exact or asymptotic distribution of an omnibus test statistic is difficult in practice, so permutation or bootstrap methods are typically used to simulate the null distribution from which an empirical p-value can be obtained. However, obtaining the correlation matrix of candidate test statistics under the null distribution can be challenging, and resampling methods can be time-consuming. To overcome these challenges, Liu and Xie (2019) proposed a powerful *p*-value combination test, termed Cauchy combination test, which works well under any arbitrary dependency structures. The method was later applied to combine p-values in rare-variant analysis in sequencing studies and was termed as ACAT (Liu et al. 2019).

A brief introduction of the HOLP procedure

Here we briefly introduce the HOLP procedure (Wang and Leng, 2016). Let Y_i be a disease response and X_i be a p dimension vector of SNP variables for individual $i (i = 1, \cdots, n)$. For a regression model $Y = X\beta + \epsilon$, one can compute the regression coefficient estimates by $\hat{\beta} = AY$. This gives $\hat{\beta} = AY = A(X\beta + \epsilon) = (AX)\beta + A\epsilon$, where X is an $n \times p$ SNP matrix. It is clear that $\hat{\beta}$ can be decomposed into two parts, the signal part $(AX)\beta$ and the noise part $A\epsilon$. In order to preserve the rank order of β, one would like to have $AX = I$ or $AX \approx I$, where I is the identity matrix. To find A, an intuitive way is to find some inverse of X. For the $p < n$ case, $A = (X^T X)^{-1} X^T$ gives the OLS estimator. For the $p > n$ case, the Moore-Penrose inverse of X gives $A = X^T (XX^T)^{-1}$ which is unique to high-dimensional data. This leads to the high-dimensional OLS projection (HOLP) estimator as $\hat{\beta} = X^T (XX^T)^{-1} Y$. Write,

$$\hat{\beta} = X^T (XX^T)^{-1} Y = X^T (XX^T)^{-1} X\beta + X^T (XX^T)^{-1} \epsilon, \qquad (1)$$

where $X^T (XX^T)^{-1} X\beta$ can be seen as a projection of β in which HOLP projects β onto the row space of X (in contrast, OLS projects β onto the column space of X). Although AX is not an identity matrix (for OLS estimate, $AX = (X^T X)^{-1} X^T X = I$), AX is diagonally dominant. Thus, $\hat{\beta}_i (i \in S,$ where $S = \{j: \beta_j \neq 0, j = 1, \ldots, p\})$ can take advantage of the large diagonal terms of AX to dominate those zero $\hat{\beta}_i's (i \notin S)$. Therefore, HOLP gives a diagonally dominant projection matrix, such that the product of this matrix and β would be more likely to preserve the rank order of the entries in β. Moreover, XX^T is of full rank for $p > n$. Thus, HOLP estimator is unique to high-dimensional data from this viewpoint.

A brief introduction of the de-sparsified LASSO procedure

In a low-dimensional case (i.e., $d \ll n$), the jth ordinary least square estimator $\hat{\beta}_j$ can be obtained from the projection of Y onto the residuals $Z_j = X_j - X_{-j}\hat{\gamma}_{OLS}^{(j)}$, where $\hat{\gamma}_{OLS}^{(j)}$ is the coefficients of X_{-j} by regression X_j on the rest of the X variables, X_{-j}. Multiplying Z_j^T on both sides of equation (1) shown in the main context, we get

$$Z_j^T Y = Z_j^T X_j \beta_j^0 + \sum_{k \neq j} Z_j^T X_k \beta_k^0 + Z_j^T \epsilon. \tag{2}$$

where β_j^0 refers to the true parameter. Dividing by $Z_j^T X_j$ in both sides of equation (2), we get,

$$\frac{Z_j^T Y}{Z_j^T X_j} = \beta_j^0 + \sum_{k \neq j} \frac{Z_j^T X_k}{Z_j^T X_j} \beta_k^0 + \frac{Z_j^T \epsilon}{Z_j^T X_j}, \tag{3}$$

$$\hat{\beta}_j^0 = \frac{Z_j^T Y}{Z_j^T X_j} - \sum_{k \neq j} \frac{Z_j^T X_k}{Z_j^T X_j} \beta_k^0. \tag{4}$$

In equation (4), the estimator $\hat{\beta}_j^0$ is called the low dimensional projection estimator (LDPE) (Zhang and Zhang, 2014). In a low-dimensional case, the residual Z_j is obtained from the OLS estimator, thus $\sum_{k \neq j} \frac{Z_j^T X_k}{Z_j^T X_j} = 0$ due to orthogonality, hence the estimator β_j^0 has the form,

$$\hat{\beta}_j = \frac{Z_j^T Y}{Z_j^T X_j} \tag{5}$$

In equation (5), $Z_j^T Y$ and $Z_j^T X_j$ are the projection of Y and X_j onto Z_j, respectively. Thus, $\hat{\beta}_j$ is the regression coefficient (partial effect) of X_j after eliminating the effects of X_{-j}.

In a high-dimensional situation (i.e., $d > n$), we can use the LASSO residuals ($X_j - X_{-j}\hat{\gamma}_{LASSO}^{(j)}$) as Z_j. When using the LASSO residuals, we do not have exact orthogonality and $\sum_{k \neq j} \frac{Z_j^T X_k}{Z_j^T X_j} \neq 0$, hence a bias arises. Thus $\hat{\beta}_j$ needs a bias correction by plugging in the LASSO estimator $\hat{\beta}_{LASSO}$. This gives the de-sparsified LASSO estimator as,

$$\hat{\beta}_j = \frac{Z_j^T Y}{Z_j^T X_j} - \sum_{k \neq j} \frac{Z_j^T X_k}{Z_j^T X_j} \hat{\beta}_{LASSO,k} \tag{6}$$

Subtracting equation (6) from equation (3) and multiplying $\sqrt{n}$ on both sides, we can get,

$$\sqrt{n}(\hat{\beta}_j - \beta_j^0) = \frac{n^{-1/2}\epsilon^T Z_j}{n^{-1/2}X^T Z_j} + \sum_{k \neq j} \sqrt{n} \frac{Z_j^T X_k}{Z_j^T X_j}(\beta_k^0 - \hat{\beta}_{LASSO,k}) \tag{7}$$

In equation (7), the first term on the right-hand side follows a Gaussian distribution and the second item is negligible (Dezeure et al. 2015). Asymptotic normality was established (Zhang and Zhang, 2014; Van de Geer et al. 2014) as,

$$\frac{\sqrt{n}(\hat{\beta}_j - \beta_j^0)}{\sigma_\epsilon \sqrt{\Omega_{jj}}} \to N(0,1) \text{ as } d \geq n \to \infty, \tag{8}$$

where $\Omega_{jj} = \frac{n Z_j^T Z_j}{(X_j^T Z_j)(X_j^T Z_j)}$, which is exploited in order to address the potential loss of power from avoiding conservative adjustment due to the dependence between the variables (Bühlmann et al., 2014), and Ω_{jj} can be computed from the data. From (8), we can easily conduct hypothesis testing by plugging in an estimator σ_ϵ which can be obtained based on the scaled LASSO (Van de Geer et al. 2014). In short, the aforementioned de-sparsifying LASSO estimator is based on regular

LASSO, and yields a non-sparse estimator which follows a Gaussian distribution. The asymptotic normality distribution allows us to assess the significance of each coefficient β_j, and compute p-values for testing the null, i.e., $H_0: \beta_j = 0$ in a high-dimensional regression setup(Javanmard and Montanari, 2014).

Reference

1. Barnett I, Mukherjee R, Lin X. The generalized higher criticism for testing SNP-set effects in genetic association studies, Journal of the American Statistical Association 2017;112:64-76.
2. Bühlmann P, Kalisch M, Meier L. High-dimensional statistics with a view toward applications in biology 2014.
3. Cai T, Lin X, Carroll RJ. Identifying genetic marker sets associated with phenotypes via an efficient adaptive score test, Biostatistics 2012;13:776-790.
4. Dezeure R, Bühlmann P, Meier L et al. High-dimensional inference: Confidence intervals, p-values and r-software hdi, Statistical science 2015:533-558.
5. Dudbridge F, Koeleman BP. Rank truncated product of P-values, with application to genomewide association scans, Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society 2003;25:360-366.
6. Javanmard A, Montanari A. Confidence intervals and hypothesis testing for high-dimensional regression, The Journal of Machine Learning Research 2014;15:2869-2909.
7. Liu Y, Xie J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures, Journal of the American Statistical Association 2019:1-18.
8. Liu Z, Lin X. A Geometric Perspective on the Power of Principal Component Association Tests in Multiple Phenotype Studies, Journal of the American Statistical Association 2019:1-32.
9. Van de Geer S, Bühlmann P, Ritov Ya et al. On asymptotically optimal confidence regions and tests for high-dimensional models, The Annals of Statistics 2014;42:1166-1202.
10. Vsevolozhskaya OA, Hu F, Zaykin DV. Detecting weak signals by combining small P-values in genetic association studies, Frontiers in Genetics 2019;10:1051.
11. Wang X, Leng C. High dimensional ordinary least squares projection for screening variables, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2016;78:589-611.
12. Zaykin DV, Zhivotovsky LA, Westfall PH et al. Truncated product method for combining P-values, Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society 2002;22:170-185.
13. Zhang CH, Zhang SS. Confidence intervals for low dimensional parameters in high dimensional linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2014;76:217-242.

Section II

Simulation studies

We followed the ADEMP scheme (Morris et al. 2019) for the simulation scenarios. The following three steps are common to all scenarios,

(1) Target of analysis: testing the null hypothesis that the given group or gene is not associated with the phenotype, and assessing the power of testing under the alternative that a gene or a group is associated with a disease outcome.

(2) Method implemented: the proposed high-dimensional statistical inference framework.

(3) Performance measures: type I error and power of the given group or gene in different scenarios, such as different sample sizes, different within-group correlations, different

numbers of predictors in a group, and different disease models (CWSM and DSSM).

Case I: simulation for the small-scale discrete predictors
Data-generating mechanisms

In this simulation, we evaluated the performance of statistical inference for groups consisting of discrete predictors (e.g., SNP genotypes). To borrow the linkage disequilibrium (LD) information from real data, real SNP genotype data from Alzheimer's disease was used to assess the inference performance of the methods at the gene level. The original genotype data was obtained from the ADNI project for the Alzheimer's Disease Neuroimaging Initiative[1] and can be accessed through their website https://adni.loni.usc.edu/.

In the ADNI data, SNP genotypes are coded as 0, 1, or 2 based on minor allele frequency. In this simulation study, we selected SNP data from two genes, *CAMTA1* and *CSMD1*, which have similar sizes to those in the previous simulation. *CAMTA1* has 168 SNPs, and *CSMD1* contains 823 SNPs. *CAMTA1* is located from 6785454 to 7769706 on chromosome 1, while *CSMD1* is located from 2935353 to 4994972 on chromosome 8. We obtained a population of 1,511 individuals with genotype data for both *CAMTA1* and *CSMD1* (human genome assembly-GRCh37). We then randomly sampled $n = 600$ and 800 individuals out of the 1,511 samples without replacement to form the simulation sample. The empirical type I error rate was calculated using the model $Y=\mu+\varepsilon$, where $\varepsilon \sim N(0,1)$. The powers of *CAMTA1* and *CSMD1* were calculated from the model $Y = \mu + X_g \beta_g + \varepsilon$. The details of the two scenarios, CWSM and DSSM, are as follows:

(a) In CWSM, we set $\boldsymbol{\beta}_{\text{CAMTA1}} = \left(\underbrace{c, \ldots, c}_{15}, \underbrace{0, \ldots, 0}_{153} \right)$ and $\boldsymbol{\beta}_{\text{CSMD1}} = \left(\underbrace{c, \ldots, c}_{50}, \underbrace{0, \ldots, 0}_{773} \right)$, where c=(0.1, 0.15, 0.20).

(b) In DSSM, we set $\boldsymbol{\beta}_{\text{CAMTA1}} = \left(c, \underbrace{0, \ldots, 0}_{167} \right)$ and $\boldsymbol{\beta}_{\text{CSMD1}} = \left(c, \underbrace{0, \ldots, 0}_{822} \right)$, where c=(0.5, 0.6, 0.7).

We set the number of variables selected in the HOLP procedure as $d = \frac{n/2}{\log(n/2)}$.

Case II: simulation with the large-scale genome-wide SNP data
Data-generating mechanisms

In this simulation, we evaluated the performance of statistical inference for groups consisting of discrete predictors (e.g., SNP genotypes) on a genome-wide scale. To borrow the LD information from real data, real SNP genotype data from Alzheimer's disease were used to assess the inference performance of the methods at the gene level. The original genotype data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and can be accessed through their website https://adni.loni.usc.edu/.

---

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at:
http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

In the ADNI data, SNP genotypes are coded as 0, 1, or 2 based on minor allele frequency. In this simulation study, we selected a total of 17,477 genes which include 196,998 SNPs. We focused on evaluating the power of two genes, namely *SGCZ* and *CSMD1*. Gene *SGCZ* has 254 SNPs, and *CSMD1* contains 895 SNPs, both are located on chromosome 8. Similar to what we did before with only two genes, we randomly sampled *n*=600 individuals without replacement to form the simulation sample. The empirical power of *CSMD1* and *SGCZ* was calculated using the model,

$$Y = X_{CSMD1}\boldsymbol{\beta}_{CSMD1} + X_{SGCZ}\boldsymbol{\beta}_{SGCZ} + X_{covariates}\boldsymbol{\beta}_{covariates} + \epsilon,$$

controlling for covariates, where $\boldsymbol{\beta}_{CSMD1} = \left(c_1, \underbrace{0, \dots, 0}_{894}\right)$ assuming a DSSM model and $\boldsymbol{\beta}_{SGCZ} =$

$\left(\underbrace{c_2, \dots, c_2}_{15}, \underbrace{0, \dots, 0}_{239}\right)$ assuming a CWSM model and $\boldsymbol{\beta}_{covariates} = (c_3, c_4, c_5, c_6)$ and $\epsilon \sim N(0,1)$.

we set $c_1 = (0.7, 1.0, 1.5)$, $c_2 = (0.2, 0.3, 0.5)$ and we selected four covariates including age (c_3), gender(c_4), education(c_5), and APOE4(c_6) ($c_3 = c_4 = c_5 = c_6 = 0.1$). The remaining genes serve as null genes (i.e., no association with the response). All the 17,477 genes were then subjected to genomewide testing.

We set the number of variables selected in the HOLP procedure as $d = n/2$. We conducted 500 replications and the power is calculated based on the FDR-corrected p-values after HOLP screening. The power for SKAT was based on the FDR-corrected p-value of all 17,477 genes since no screening was applied. For SKAT, we used two kernel functions, namely the linear and the IBS kernel. We used the default parameter settings in the SKAT function as follows,

- `kernel="linear.weighted"`
- `method="davies"`
- `weights.beta=c(1,25)`
- `weights=NULL`
- `impute.method="fixed"`
- `r.corr=0`
- `is_check_genotype=TRUE`
- `is_dosage = FALSE`
- `missing_cutoff=0.15`
- `max_maf=1`
- `estimate_MAF=1`

Case III: simulation with quantitative predictors

Data-generating mechanisms

Data were generated from a linear mode, i.e., $Y = X\beta + \epsilon$, where $X = (X_{G1}, X_{G2})$ represent predictors in two groups (G1 with 100 predictors and G2 with 700 predictors), and each follows a multivariate normal distribution with mean 0 and covariance $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = \rho^{|i-j|}$. $\epsilon \sim N(0, 1)$. In each group, there are signals and noises. Let $\beta_G = (c, \dots, c, 0, \dots, 0)$ be the effect coefficient vector in group G, and c represents the non-zero coefficient.

In Scenario 1, we assume the CWSM model and let $c = 0.1$, i.e. $\beta_{\text{G1}} = \left(\underbrace{0.1,\cdots,0.1}_{15} , \underbrace{0,\cdots,0}_{85} \right)$ and

$\beta_{\text{G2}} = \left(\underbrace{0.1,...,0.1}_{50} , \underbrace{0,...,0}_{650} \right)$. In Scenario 2, we assume the DSSM model and further assume only one

dominated signal ($c = 0.25$) in both G1 and G2, that is $\beta_{\text{G1}} = \left(\underbrace{0.25}_{1} , \underbrace{0,...,0}_{99} \right)$ and $\beta_{\text{G2}} = \left(\underbrace{0.25}_{1} , \underbrace{0,...,0}_{699} \right)$.

In all the scenarios, we varied the sample size from n=600 to 800 and the correlations within the predictors (ρ=0.3, 0.6 and 0.9). In addition, the number of variables to keep after screening via HOLP was set as $d = \dfrac{n}{\log(n)}$ (Wang and Leng, 2016).

Performance measure

***Empirical type I error and power comparison between iART-A and ART-A* in Case I**

Figure S1(A) shows the type I error control of iART-A and ART-A. The two methods can reasonably control the type I error under different sample sizes. The power results under the DSSM and CWSM model are shown in Figures S1(B) and (C), respectively. We can see that the power of *i*ART-A is always higher than that of ART-A when *k*=*L*. The power also increases as the effect size increases. In summary, our proposed *i*ART-A controls the type I error well and has higher power than ART-A under different scenarios and disease model assumptions.
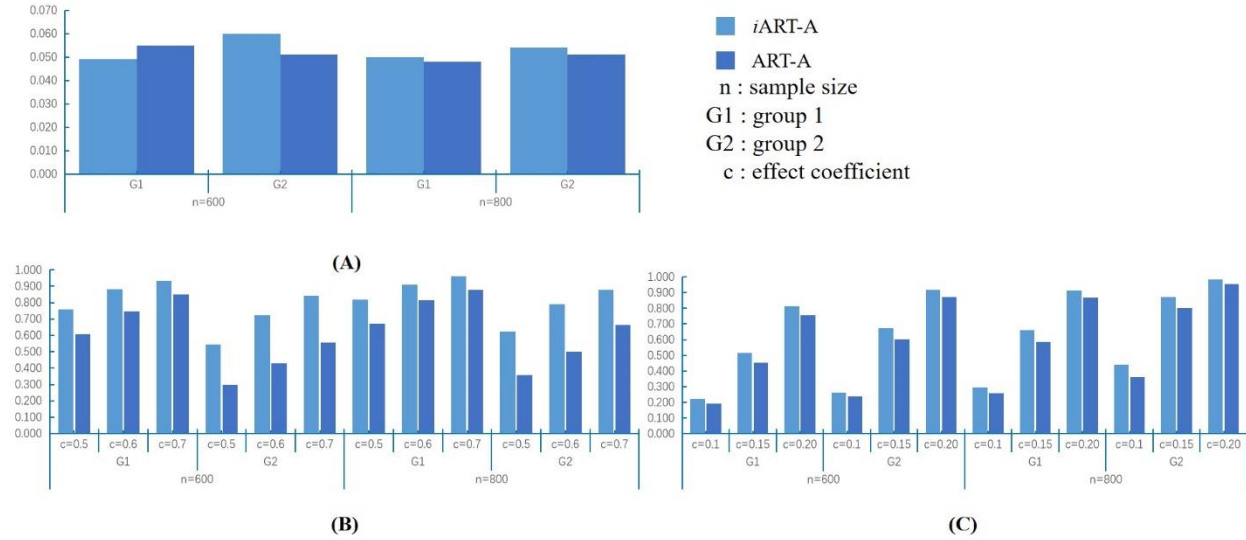

Figure S1. Case I: Empirical type I error and power comparison between *i*ART-A and ART-A. (A) the type I error comparison between *i*ART-A and ART-A; (B) the power comparison between *i*ART-A and ART-A under the DSSM model; (C) the power comparison between *i*ART-A and ART-A under the CWSM model.

***Empirical type I error and power comparison between iART-A and ART-A* in Case III**

We also evaluated the type I error control of *i*ART-A and ART-A (Figure S2(A)). The two methods can reasonably control the type I error under different sample sizes and different correlations. When the sample size increases from 600 to 800, ART-A is a little conservative compared to *i*ART-A, especially under high correlation (ρ=0.9).
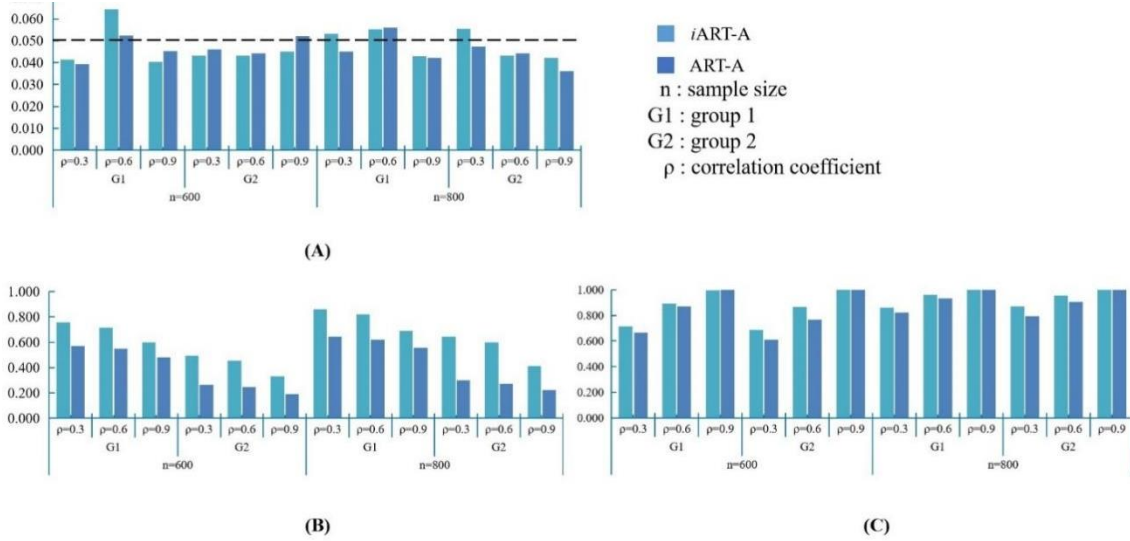
Figure S2. Case III: Empirical type I error and power comparison between *i*ART-A and ART-A. (A) the type I error comparison between *i*ART-A and ART-A; (B) the power comparison between *i*ART-A and ART-A under the DSSM model; (C) the power comparison between *i*ART-A and ART-A under the CWSM model.

Figure S2 (B) shows the power results under the DSSM model. We can see that the power of *i*ART-A is much higher than that of ART-A when *k=L*. In addition, the power of G1 is higher than that of G2 under the same sample size and correlation, which could be due to the high percentage of noise variables in G2. Within each group, the power decreases as the within-group correlation increases. This could be due to the fact that more noise variables were left after screening due to high correlation, leading to diluted signal within each group and hence low power.

The power results under the CWSM model are shown in Figure S2 (C). The proposed *i*ART-A performs similarly or better than the ART-A method does, especially when the correlation is low. In contrast to the DSSM model results, the power increases as the within-group correlation increases under the CWSM model, indicating the effect of correlations on the power under different model assumptions. This further highlights the importance of proposing different models under different model assumptions and further integrating different results to obtain a more robust conclusion.

In summary, our proposed *i*ART-A can control the type I error well and has higher power than ART-A under different scenarios and disease model assumptions. It is safe to apply *i*ART-A in real application and next, we evaluate the performance of the omnibus test approach.

Empirical type I error and power comparison between MinP, iART-A and Min-O in Case III
Figure S3 (A) shows the type I error control under the three methods MinP, *i*ART-A and Min-O. Overall, the type I error can be reasonably controlled under different scenarios in different groups for the three methods.
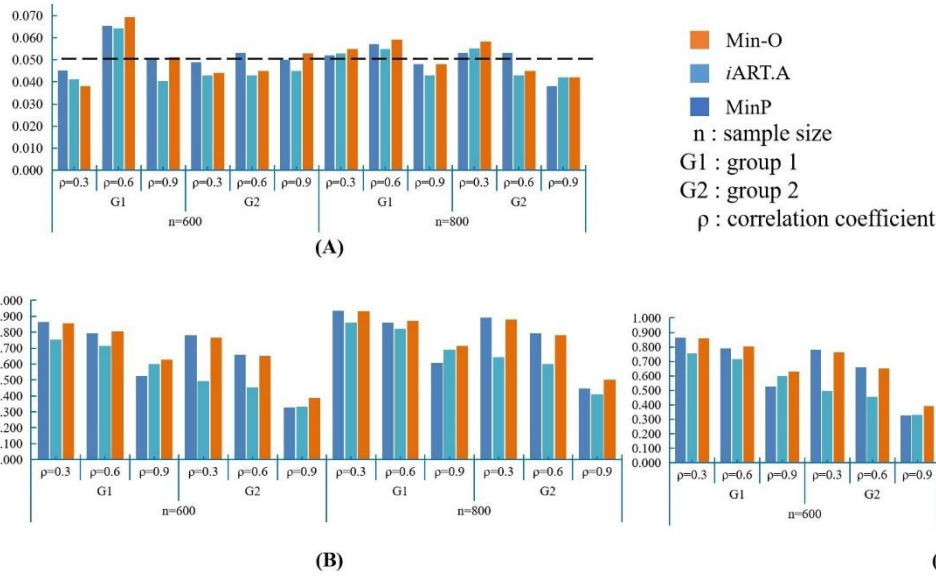
Figure S3. Case III: empirical type I error and power comparison between *i*ART-A, ART-A and Min-O (omnibus test based on ART.A and MinP). (A) the type I error comparison between *i*ART-A, ART-A and Min-O; (B) the power comparison between *i*ART-A, ART-A and Min-O under the DSSM model; (C) the power comparison between *i*ART-A, ART-A and Min-O under the CWSM model.

The power comparison results are presented in Figure S3(B) and Figure S3(C), corresponding to the DSSM and CWSM models, respectively. It is expected that MinP and *i*ART-A would achieve the highest power under the DSSW and CWSM models, respectively. Remarkably, the omnibus test consistently demonstrates comparable or even superior power to the best-performing individual method under the corresponding model (especially when the within-group correlation is high, e.g., $\rho=0.9$). These findings from the power simulation suggest that the omnibus approach can be safely applied in practice, irrespective of the underlying disease model.

The numerical results shown in the figures of Case I and Case III in both the main text and the supplemental file

Table S1. The type I error comparison between different methods in Case I.

Methods	n=600		n=800	
	G1	G2	G1	G2
MinP	0.045	0.044	0.052	0.053
iART.A	0.049	0.060	0.050	0.054
Min-O	0.047	0.057	0.049	0.056
ART.A	0.055	0.051	0.048	0.051

Table S2. The power comparison between different methods under the DSSM model in Case I.

Methods	n=600						n=800					
	G1			G2			G1			G2		
	c=0.5	c=0.6	c=0.7	c=0.5	c=0.6	c=0.7	c=0.5	c=0.6	c=0.7	c=0.5	c=0.6	c=0.7
MinP	0.759	0.877	0.940	0.615	0.783	0.882	0.820	0.915	0.970	0.679	0.835	0.907
iART.A	0.758	0.882	0.932	0.544	0.723	0.841	0.818	0.908	0.960	0.622	0.789	0.878
Min-O	0.794	0.902	0.958	0.647	0.793	0.897	0.845	0.924	0.973	0.696	0.847	0.916

| ART.A | 0.607 | 0.747 | 0.851 | | 0.297 | 0.428 | 0.557 | | 0.673 | 0.812 | 0.876 | | 0.358 | 0.499 | 0.664 |

Table S3. The power comparison between different methods under the CWSM model in Case I.

Methods	n=600						n=800					
	G1			G2			G1			G2		
	c=0.1	c=0.15	c=0.2	c=0.1	c=0.15	c=0.2	c=0.1	c=0.15	c=0.2	c=0.1	c=0.15	c=0.2
MinP	0.173	0.370	0.638	0.161	0.408	0.671	0.225	0.500	0.761	0.236	0.536	0.815
iART.A	0.220	0.514	0.813	0.262	0.675	0.916	0.295	0.662	0.915	0.441	0.873	0.984
Min-O	0.225	0.506	0.796	0.242	0.645	0.891	0.284	0.639	0.903	0.402	0.825	0.972
ART.A	0.191	0.451	0.755	0.236	0.603	0.870	0.258	0.585	0.867	0.363	0.801	0.956

Table S4. The type I error comparison between different methods in Case III.

Methods	n=600						n=800					
	G1			G2			G1			G2		
	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9
MinP	0.045	0.065	0.050	0.049	0.053	0.050	0.052	0.057	0.048	0.053	0.053	0.038
iART.A	0.041	0.064	0.040	0.043	0.043	0.045	0.053	0.055	0.043	0.055	0.043	0.042
Min-O	0.038	0.069	0.051	0.044	0.045	0.053	0.055	0.059	0.048	0.058	0.045	0.042
ART.A	0.039	0.052	0.045	0.046	0.044	0.052	0.045	0.056	0.042	0.047	0.044	0.036

Table S5. The power comparison between different methods under the DSSM model in Case III.

Methods	n=600						n=800					
	G1			G2			G1			G2		
	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9
MinP	0.862	0.791	0.526	0.780	0.658	0.328	0.933	0.860	0.606	0.890	0.792	0.447
iART.A	0.755	0.716	0.600	0.494	0.455	0.330	0.861	0.819	0.689	0.643	0.599	0.412
Min-O	0.857	0.803	0.629	0.765	0.651	0.389	0.929	0.870	0.715	0.880	0.783	0.502
ART.A	0.571	0.550	0.481	0.264	0.247	0.191	0.644	0.620	0.557	0.298	0.275	0.221

Table S6. The power comparison between different methods under the CWSM model in Case III.

Methods	n=600						n=800					
	G1			G2			G1			G2		
	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9	ρ=0.3	ρ=0.6	ρ=0.9
MinP	0.544	0.714	0.987	0.412	0.528	0.938	0.694	0.852	0.995	0.536	0.701	0.984
iART.A	0.716	0.894	0.996	0.688	0.864	1.000	0.861	0.958	1.000	0.868	0.954	1.000
Min-O	0.697	0.875	0.997	0.635	0.818	0.999	0.843	0.948	1.000	0.829	0.934	1.000
ART.A	0.665	0.872	0.998	0.611	0.768	1.000	0.821	0.933	1.000	0.792	0.906	1.000

Table S7. Properties of the cohort used for method validation.

Covariates	Statistical Description
age (years; mean ± SD)	74.6 ± 6.8
gender (male/female (%))	34.5/65.5(%)
education (Year; min/median (IQR)/max)	6/16 (4)/20
APOE4 （copy number, 1/2/3 (%))	50.7/37.9/11.4 (%)

The analysis results of the other three brain volume traits

This supplemental file contains the analysis of the other three brain volume traits. For the volume of Fusiform, we identified two genes (see Table S8). MinP cannot detect *LSAMP*, while iART-A and Min-O can identify the two.

Table S8. List of covariates, genes, and SNPs associated with the volume of Fusiform.

Gene (Ensemble ID)	SNP ID	$P_{\text{DS-LASSO}}$	P_{DOT}
	rs1401069	0.1956	0.2430
	rs1518343	0.5008	0.4292
	rs1523262	0.7308	0.2885
	rs1619470	0.5584	0.6085
	rs2030230	0.3965	0.7935
	rs1910044	0.0039	0.0213
	rs1914812	0.2375	0.4584
	rs1920363	0.0039	0.0213
	rs1980081	0.9103	1e-16
	rs1712999	0.1879	1e-16
	rs2869501	0.0253	1e-16
	rs2869787	0.5437	1e-16
	rs2927271	0.3452	0.2600
LSAMP	rs4305418	0.1825	0.2451
(ENSG00000185565)	rs4405915	0.5204	0.3688
	rs4855941	0.6247	0.5774
	rs6771286	0.1511	0.0820
	rs6797385	0.5133	0.5422
	rs6805241	0.0705	0.0593
	rs6806238	0.9769	0.9095
	rs6782605	0.3392	0.4172
	rs7612435	0.2101	0.3129
	rs7621196	0.4869	0.5903
	rs9842731	0.4715	0.3863
	rs10934370	0.9707	0.9952
	rs11921574	0.5733	0.4812
	rs11929035	0.7978	0.8490
	rs13077308	0.2577	0.2411
	rs2744943	0.6517	8.57e-01
ILRUN	rs2744949	0.4089	2.89e-13
(ENSG00000196821)	rs2744974	0.9273	0.1755
	rs2814945	0.3618	0.6219
	rs2814992	0.4166	1.24e-07

$P_{\text{DS-LASSO}}$: p-values of SNPs inferred via the desparsified-LASSO.
P_{DOT}: p-values of SNPs after decorrelation with the orthogonal transformation.
The significant covariates include age (p=2.18e-05), gender (p=5.64e-12), and education (p=4.76e-05).

For the volume of Entorhinal, we identified four genes (see Table S9). MinP cannot detect *AMPH*, *KCNQ3,* and *ABCG1* and iART-A cannot identify *BCL2*, while Min-O can identify the four.

Table S9. List of covariates, genes, and SNPs associated with the volume of Entorhinal.

Gene (Ensemble ID)	SNP ID	$P_{\text{DS-LASSO}}$	P_{DOT}
	rs2028209	0.2139	0.2308
AMPH	rs2028210	0.4009	1e-16
(ENSG00000078053)	rs2043260	0.6826	1e-16
	rs2043786	0.2822	1e-16

Gene	SNP		
	rs2043787	0.8150	1e-16
	rs2267809	0.1974	1e-16
	rs2284248	0.3774	1e-16
	rs2215955	0.6406	1e-16
	rs4720279	0.3951	1e-16
	rs4427071	0.2205	0.2583
	rs3778878	0.6328	1e-16
	rs3778880	0.0235	1e-16
	rs3800795	0.2363	1e-16
	rs10248732	0.6472	1e-16
	rs10233502	0.2606	1e-16
	rs10263645	0.7663	0.0076
	rs10265223	0.4632	0.0490
	rs10464364	0.5259	0.5237
	rs10499602	0.5335	0.5327
	rs17500182	0.3524	0.2308
	rs973527	0.0026	0.0031
	rs1457785	0.2454	0.2704
	rs1457788	0.2876	1e-16
	rs2469510	0.4568	0.3495
	rs2469520	0.0225	1e-16
	rs2469522	0.9268	0.4332
	rs2198985	0.8888	1e-16
	rs2673567	0.1346	1e-16
	rs2721905	0.1079	0.0561
	rs1864772	0.9460	0.7063
KCNQ3	rs1864773	0.6614	0.4905
(ENSG00000184156)	rs6471065	0.5121	0.4440
	rs6471070	0.8647	0.6380
	rs6984395	0.0752	0.0142
	rs6993047	0.9744	0.4235
	rs7007544	0.0496	0.0405
	rs9643288	0.1562	0.2977
	rs9693071	0.7077	0.7240
	rs10094856	0.4872	0.7574
	rs10956658	0.7210	0.9687
	rs17659499	0.6875	0.7438
	rs915842	0.4455	1e-16
	rs915843	0.4455	1e-16
	rs3787986	0.9041	1e-16
ABCG1	rs3788005	0.5771	1e-16
(ENSG00000160179)	rs4148120	0.5201	1e-16
	rs4148125	0.4028	0.0270
	rs4148134	0.9236	1e-16
	rs2839482	0.1254	1e-16
	rs7281345	0.3480	2.62E-06
	rs1982673	0.9672	0.8046
BCL2	rs4987852	0.9912	0.8560
(ENSG00000171791)	rs7236090	0.2534	0.3195
	rs12958785	1.64E-05	1.70E-05

The significant covariates include age (p=7.05e-06), gender (p=2.34e-04), education (p=1.90e-06), and APOE4 (p=3.30e-06).

For the volume of the Middle Temporal Gyrus, we identified three genes (see Table S10). MinP cannot detect any, while iART-A and Min-O can identify all.

Table S10. List of covariates, genes, and SNPs associated with the volume of Middle Temporal Gyrus.

Gene (Ensemble ID)	SNP ID	$P_{\text{DS-LASSO}}$	P_{DOT}
SLCO1B1 (ENSG00000134538)	rs981262	0.1016	0.1897
	rs987839	0.1603	0.2425
	rs2100996	0.9981	0.9372
	rs2169970	0.7133	0.9195
	rs2291076	0.1619	0.4756
	rs2306283	0.1021	1e-16
	rs4149026	0.0481	1e-16
	rs4149029	0.0279	1e-16
	rs4149030	0.7168	0.8959
	rs4149036	0.0951	0.0094
	rs4149038	0.0771	1e-16
	rs4149042	0.0157	0.1943
	rs4149046	0.5843	0.3780
	rs4149050	0.0738	1e-16
	rs4149056	0.5528	0.0349
	rs4149058	0.0746	0.3231
	rs4149059	0.0738	1e-16
	rs4149062	0.5499	0.6336
	rs4149076	0.0861	0.0403
	rs4149078	0.1813	0.1108
	rs6487213	0.6969	0.6748
	rs7966613	0.3943	0.0597
	rs11045812	0.7740	0.9373
	rs10841763	0.8120	0.6378
	rs12317268	0.8120	0.6378
KLC1 (ENSG00000126214)	rs861547	0.4803	0.5480
	rs1078756	0.9474	0.5881
	rs729438	0.8076	0.3863
	rs2296482	0.8830	0.8607
	rs2896489	0.2098	1.11E-16
	rs11850979	0.2893	0.4975
	rs11160755	0.1178	0.1521
	rs11160756	0.2101	6.52E-13
GRIK1 (ENSG00000171189)	rs363512	0.7537	6.98E-09
	rs363514	0.0282	1e-16
	rs363518	0.0411	1e-16
	rs363522	0.0551	1e-16
	rs363582	0.0556	1e-16
	rs363598	0.3872	3.33E-16
	rs408302	0.5345	0.4365
	rs459617	0.1445	0.0037
	rs2251036	0.7530	0.9307
	rs2832438	0.0199	0.0672
	rs16985084	0.0133	0.0566

The significant covariates include age (p=6.46e-03), gender 7.10e-13) and education (p=1.30e-03)

Q-Q plots of the gene-level p-values using the three methods (Min-O, iART.A, and MinP)
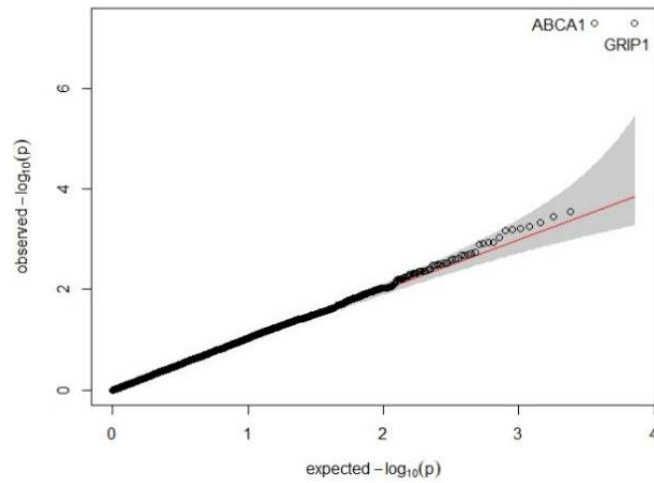

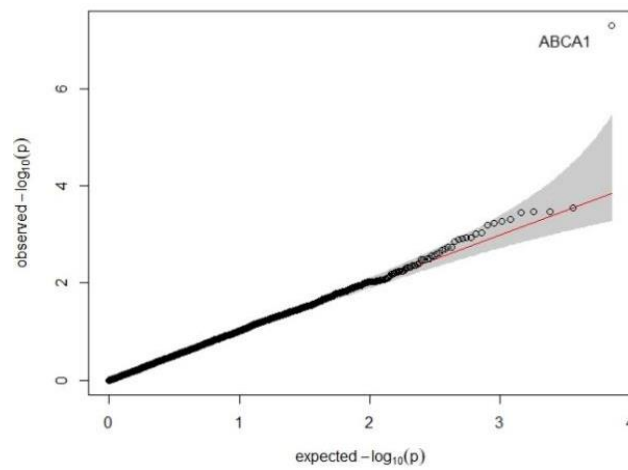Figure S4. The Q-Q plot of gene p-values with the Min-O method.

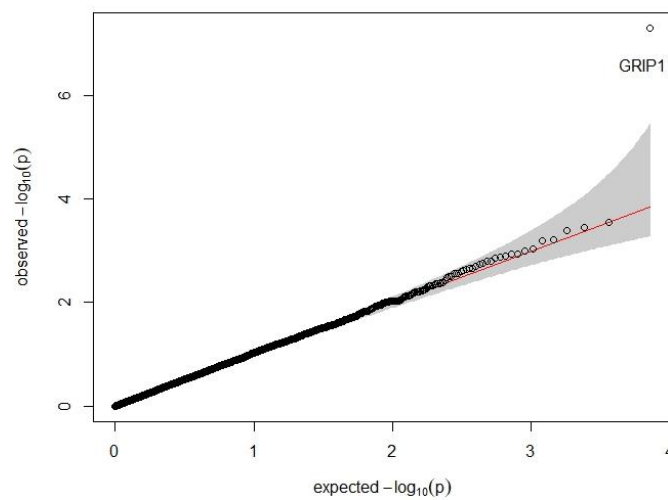Figure S5. The Q-Q plot of gene p-values with the iART-A method.

Figure S6. The Q-Q plot of gene p-values with the MinP method.

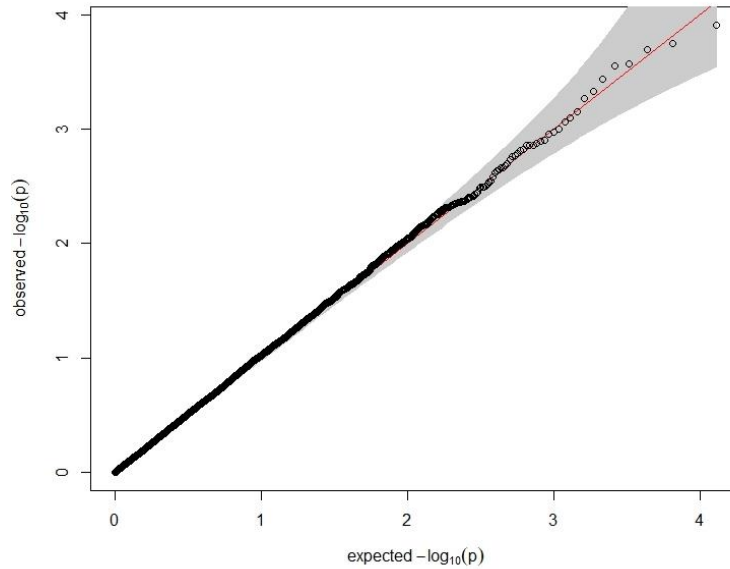# Q-Q plots of the gene-level p-values of ADNI data using the MAGMA software

Figure S7. The Q-Q plot of gene-level p-values of ADNI data with the MAGMA software.

Real data analysis II: gene-level analysis of the human birth weight dataset

The birth weight data

We compared the performance of the omnibus test with the SKAT method by analyzing a human birth weight dataset in the Thai population from the Gene Environment Association Studies initiative GENEVA founded by the trans-NIH (National Institute of Health) Genes, Environment, and Health Initiative (GEI). The dataset was obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession number phs000096.v4.p1.

The pre-processing of the data

There are a total of 590,913 SNPs after removing SNPs with minor allele frequency (MAF)<0.05, missing rate <0.05, and those deviating from Hardy-Weinberg equilibrium (p-value < 0.001). We only focused on genes containing three or more SNPs in our analysis and the missing SNP genotypes were imputed by using the *KNNcatImpute R* package. This ended up with 251,816 SNPs mapped to 12,001 genes based on the human genome assembly 37 (GRCh37). The final sample size is 1,090. Gene-level FDR-adjusted p-values were reported. We also included the covariates in the proposed high-dimensional inference framework. The covariates considered in the analysis are listed in Table S11.

Table S11. Covariates included in the analysis.

Names of covariates	Description
m_Age_OGTT	Mother's age at OGTT
b_CordCP_ug	Baby's cord C-peptide
b_CordPGC_mg	Baby's cord glucose
m_DBPM_OGTT	Mother's mean OGTT diastolic blood pressure

m_FCP_ug	Mother's fasting C-peptide at OGTT
m_FH_DM_Any	Mother's family history of diabetes
m_FH_HBP_Any	Mother's family history of hypertension
m_FPG_CLC_mg	Mother's fasting glucose at OGTT
b_Gestage	Baby's gestational age at delivery
m_HbA1c_percent	Mother's HbA1c percent
m_HtM_OGTT	Mother's mean OGTT height
b_NN_Gender	Baby's gender
m_OneHrPG_CLC_mg	Mother's one hour OGTT glucose
m_SBPM_OGTT	Mother's mean OGTT systolic blood pressure
m_TwoHrPG_CLC_mg	Mother's two hour OGTT glucose
m_WtM_OGTT	Mother's pre-pregnancy weight
m_BMI_OGTT	Mother's BMI

Results of real data analysis II

In our study, we took the baby's birth weight as the response variable to demonstrate the implementation of our method. Firstly, we conducted the HOLP screening procedure to keep the top 20,000 most significant SNPs, where 9 covariates and 19, 991 SNPs mapping to 4,730 genes were included.

(i) Results obtained using the proposed method: We identified four covariates (b_CordCP_ug, m_WtM_OGTT, b_NN_Gender, and b_Gestage) and five genes (*ANTXR1, KCTD16, C4orf54, PPARG,* and *RNGTT*) after the FDR correction. The detailed results are summarized in Tables S12. The Q-Q plots of the gene-level p-values for the three methods are given in Figures S7-9. The MinP method is very conservative and the results of the omnibus test are mainly contributed by the iART-A method.

Table S12. List of covariates, genes, and SNPs associated with baby's birth weight.

Covariates (p-value)	Gene Name (Gene ID)	SNP ID	$P_{\text{DS-LASSO}}$	P_{DOT}	False Discovery Rate		
					P_{MinP}	$P_{i\text{ART.A}}$	P_{Omnibus}
		rs10164798	0.4817	**1.20e-05**			
		rs10167510	0.0919	0.9707			
		rs10176087	0.5570	0.9684			
		rs10185336	0.1293	**3.00e-15**			
		rs11126223	0.7071	0.9937			
b_CordCP_ug		rs11126231	0.0065	0.9313			
(1.6676e-04)	*ANTXR1*	rs11683511	0.8129	0.9900			
	(Gene ID: **84168**)	rs11687792	0.2469	0.9767	0.9739	**9.99e-17**	**2.00e-16**
		rs13422656	0.7505	**2.40e-04**			
		rs2667	0.3661	0.9447			
b_Gestage		rs3923626	0.4952	0.9562			
(3.0429e-03)		rs4416261	0.2246	0.9709			
		rs5017686	0.8823	0.9909			
		rs6546497	0.2244	0.3824			
		rs6546503	0.1539	**3.37e-04**			
		rs6709954	0.3707	0.9668			
m_WtM_OGTT		**rs6722492**	0.1539	**3.37e-04**			
(7.8093e-07)		rs6732795	0.0396	0.9421			
		rs7422275	0.1858	0.9732			
		rs7564412	0.2687	0.9657			
		rs7565538	0.2713	0.9649			
		rs7585658	0.6558	0.9912			
b_NN_Gender		**rs1154456**	0.3528	**1.00e-16**			
(1.9258e-02)		**rs1154457**	0.8883	**1.00e-16**			

Gene	SNP	$P_{\text{DS-lasso}}$	P_{DOT}	P_{MinP}	$P_{i\text{ART.A}}$	P_{Omnibus}
C4orf54 (Gene ID: **285556**)	**rs1154459**	0.4116	**1.00e-16**			
	rs1154460	0.7751	**1.00e-16**			
	rs1154469	0.1901	0.2751	0.9893	**1.00e-16**	**1.00e-16**
	rs12504511	0.5914	**1.00e-16**			
	rs4147549	0.4921	**1.00e-16**			
	rs930478	0.5133	**1.00e-16**			
	rs971074	0.3627	**1.00e-16**			
KCTD16 (Gene ID: **57528**)	rs10037551	0.8337	0.1814			
	rs10515538	0.3250	**0.0019**			
	rs10515542	0.4000	**1.11e-16**			
	rs13182980	0.7708	0.9666			
	rs13357117	0.9582	0.9555			
	rs1422708	0.9725	0.8464			
	rs2080991	0.5361	**0.0269**			
	rs2112157	0.4978	0.8932			
	rs244524	0.7715	**3.55e-06**	0.9966	**1.00e-16**	**1.00e-16**
	rs244537	0.7315	**1.00e-16**			
	rs29893	0.9387	**0.0387**			
	rs29894	0.3479	**6.36e-11**			
	rs29900	0.4000	**1.11e-16**			
	rs3095960	0.3251	**1.00e-16**			
	rs4912969	0.8337	0.1814			
	rs831419	0.3399	**1.00e-16**			
PPARG **GeneID:5468**	rs1152002	0.0305	0.9529			
	rs1175541	0.6650	0.9552			
	rs1175542	0.7371	0.9441			
	rs1175543	0.1421	0.9537			
	rs1175544	0.5257	0.9806			
	rs12497191	0.0307	0.9490			
	rs1875796	0.6108	0.7302			
	rs2028759	0.1182	**3.77e-05**	0.9739	**3.89e-24**	**7.77e-24**
	rs2921190	0.2105	**2.22e-16**			
	rs2921193	0.4413	**5.86e-14**			
	rs2938387	0.5261	**0.0161**			
	rs2938392	0.4086	**1.49e-11**			
	rs2938394	0.1884	**6.82e-07**			
	rs2938395	0.1884	**6.82e-07**			
	rs2938397	0.8446	0.3972			
	rs2959268	0.5012	**1.22e-15**			
	rs2972162	0.8701	**7.52e-11**			
	rs709158	0.3658	0.9600			
	rs796313	0.6248	0.8572			
	rs9809905	0.3771	0.9914			
	rs9855622	0.1212	0.9424			
RNGTT **GeneID:8732**	rs1040676	0.6396	0.5620			
	rs12174206	0.3316	**9.30e-05**			
	rs2757728	0.5694	0.9872			
	rs6922340	0.9963	0.7010			
	rs7747355	0.7697	0.1807			
	rs9342152	0.5471	**1.35e-04**	0.9966	**2.45e-06**	**4.90e-06**
	rs9344870	0.4936	**7.53e-05**			
	rs9351179	0.7079	0.5421			
	rs9353592	0.7337	0.6272			
	rs9362566	0.4885	**1.00e-16**			
	rs9362583	0.5006	0.7799			

The significant SNPs are shown in bold font.

$P_{\text{DS-lasso}}$: p-values of SNPs inferred via desparsifying-lasso, a high-dimensional inference approach.

P_{DOT}: decorrelated $P_{\text{DS-lasso}}$ by orthogonal transformation.

P_{MinP}: p-values of SNPs inferred via minimum p-value distribution.

$P_{i\text{ART.A}}$: p-values of SNPs inferred via our proposed improved ART.A approach.

P_{Omnibus}: p-values of SNPs inferred via omnibus test, Min-O, based on MinP and iART.A.
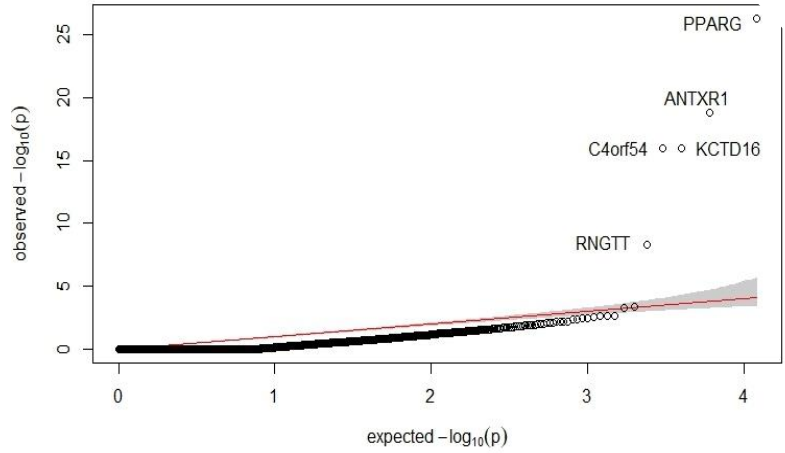
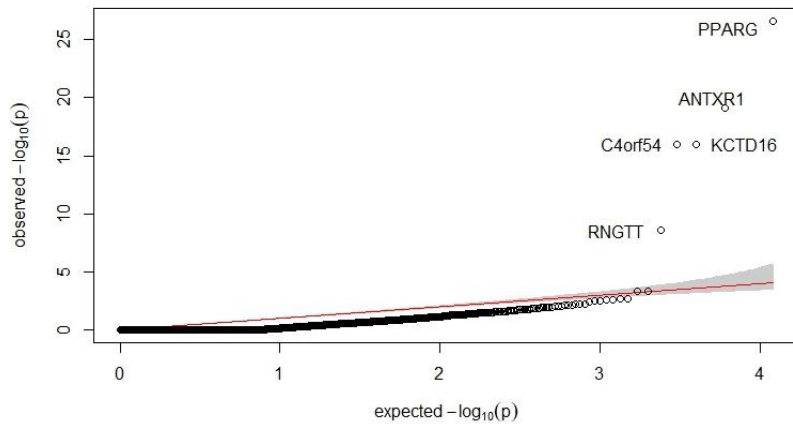Figure S8. The Q-Q plot of gene-level p-values with the Min-O method.

Figure S9. The Q-Q plot of gene-level p-values with the iART.A method.
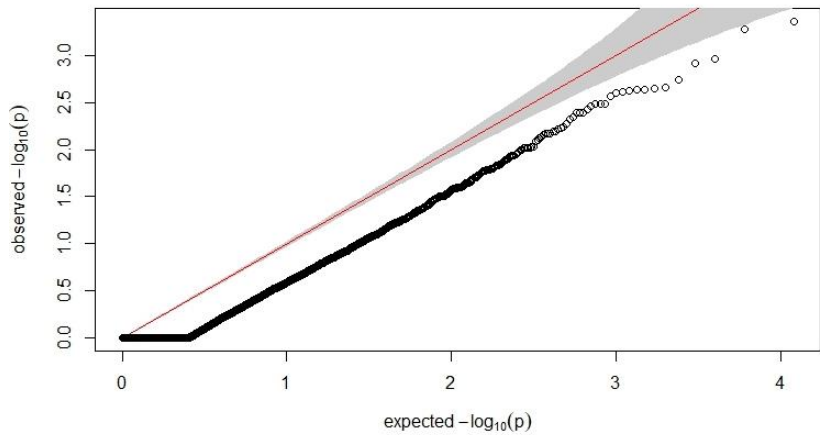

Figure S10. The Q-Q plot of gene-level p-values with the MinP method.

(ii) Results obtained using other methods (e.g., SKAT and MAGMA): The SKAT method did not identify any significant genes after the FDR control after adjusting for the effects of the covariates. For the MAGMA method using summary statistics, we obtained the gene-level p-

values of the 11,114 genes mapped by 251,816 SNPs (following the same procedure as described in the main text for the ADNI data anlaysis). After the gene-level FDR control, no significant gene was identified by the MAGMA package. The Q-Q plot of the gene-level p-values by the MAGMA software can be found in Figure S11.
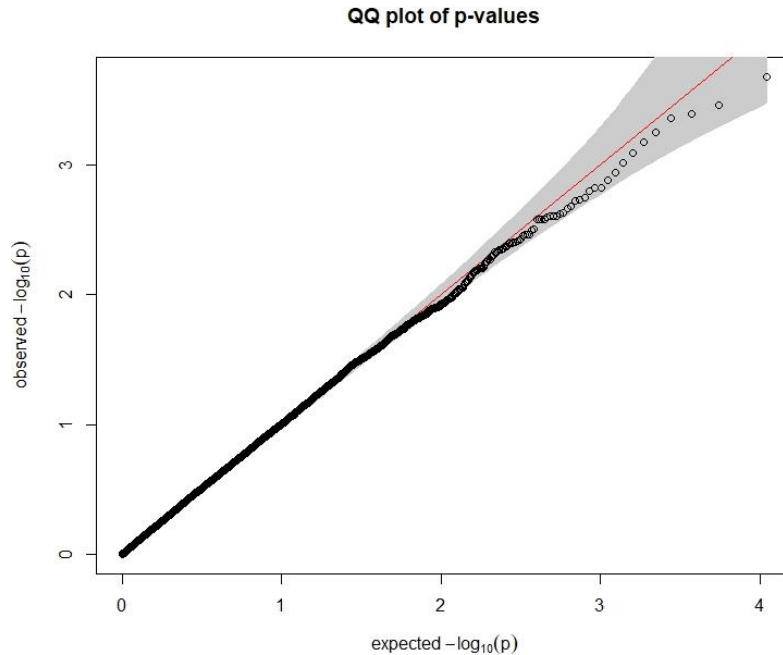
Figure S11. The Q-Q plot of gene-level p-values of birthweight data with the MAGMA software.

(iii) A brief description of genes identified: It is well known that the baby's gender, the mother's pre-pregnancy weight, and the baby's gestational age at delivery were associated with the baby's birthweight[1-5]. Elevated cord C-peptide is significantly associated with increasing birthweight and Cord blood C-peptide is an important mediator of the association between maternal and infant adiposity, across the spectrum of maternal glucose tolerance[6, 7].

We found literature reports for four of the five significant genes, indicating the clinical significance of the method. Mutations in *ANTXR1* can cause GAPO syndrome, a condition characterized by growth retardation, alopecia, pseudoanodontia, and progressive visual impairment[8]. Many studies have shown that children born with low birth weight are associated with an increased risk of developing type 2 diabetes (T2D) in adulthood [9]. Several GWAS studies have identified genetic factors associated with T2D and birth weight (e.g., [10]). *PPARG* is a member of the nuclear receptor superfamily of transcription factors and regulates the gene expression involved in lipid metabolism within adipocytes. Its pivotal role in adipose tissue differentiation and susceptibility to type 2 diabetes[4] in humans has been well-established. Numerous studies have substantiated a significant association between *PPARG* and birth weight[11-13]. The proximal small intestine *GABABR* signaling is disrupted in prediabetes independent of age, with possible involvement of decreased *KCTD16*, which implies *KCTD16* is a potential pharmacological target to study prediabetes prevention[14]. *C4orf54*, also known as

Chromosome 4 Open Reading Frame 54, is a gene located on chromosome 4. It is a relatively less studied gene, and its exact function and role in biological processes are still not fully understood. Currently, proteins with *C4orf54* have defective phosphosites, representing opportunities to further expand our knowledge of the insulin signaling network and how it goes awry in insulin resistance[15]. *RNGTT* is also a less studied gene and the direct relationship between the gene and birth weight or T2D has not been reported.

We also examined the LD structure of the SNPs in the five significant genes, *ANTXR1*, *KCTD16*, *C4orf54*, *PPARG*, and *RNGTT*. As shown in Figure S12-S16, all the LD plots revealed strong correlations between SNPs in the corresponding gene. In addition, all significant SNPs in each gene belong to the same block. The strong LD between the SNPs may explain why SNPs are significant after applying decorrelation by the orthogonal transformation.
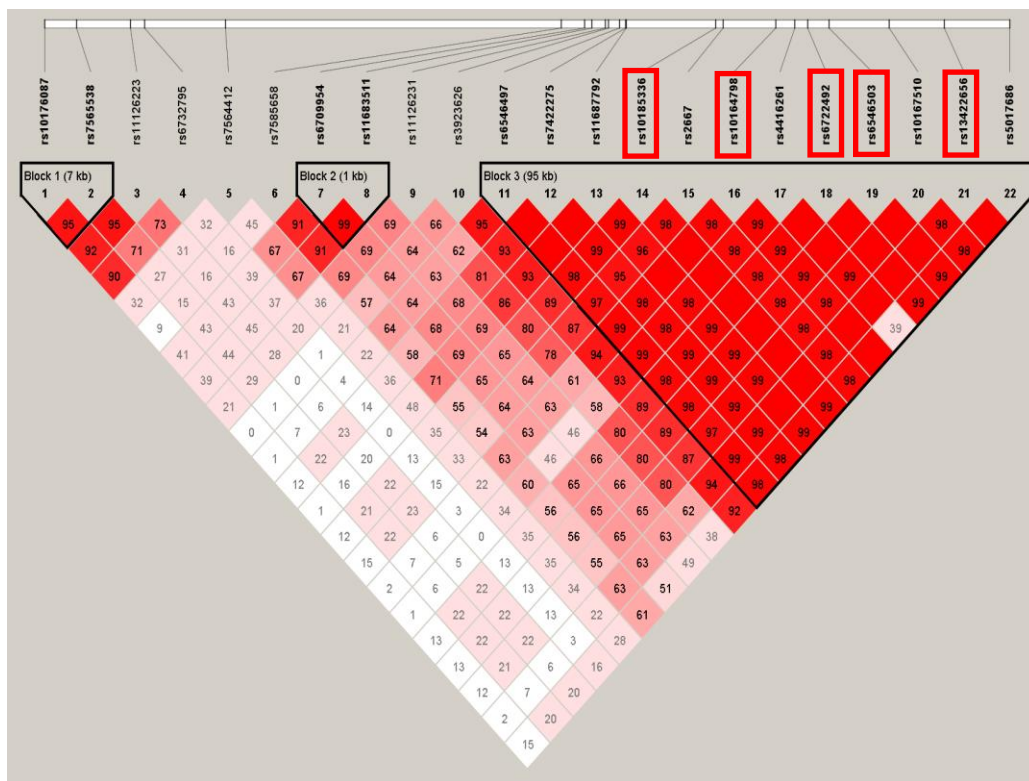
Figure S12. The LD plot of SNPs in gene *ANTXR1*. SNPs showing significance are highlighted with red rectangles.
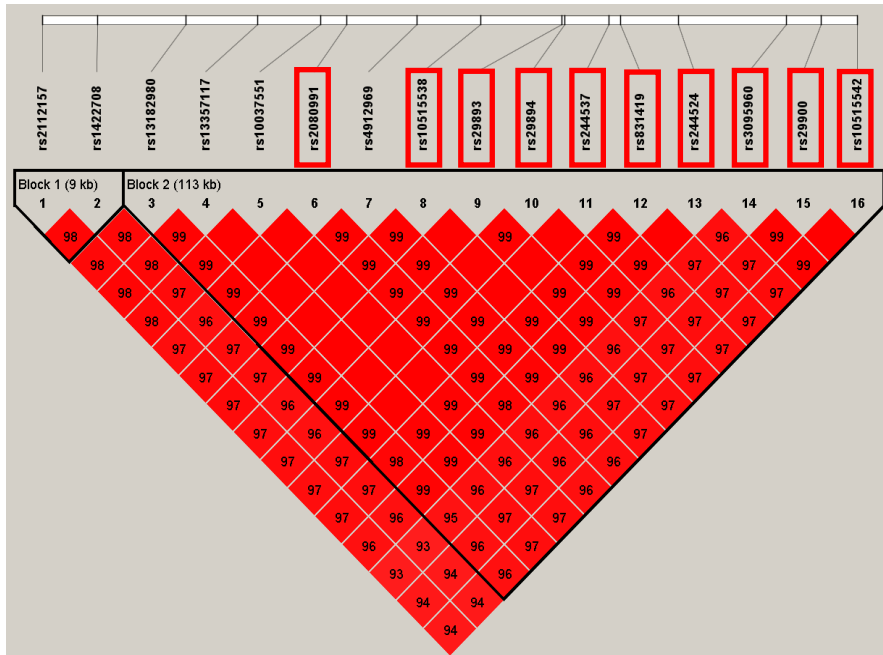
Figure S13. The LD plot of SNPs in gene *KCTD16*. SNPs showing significance are highlighted with red rectangles.
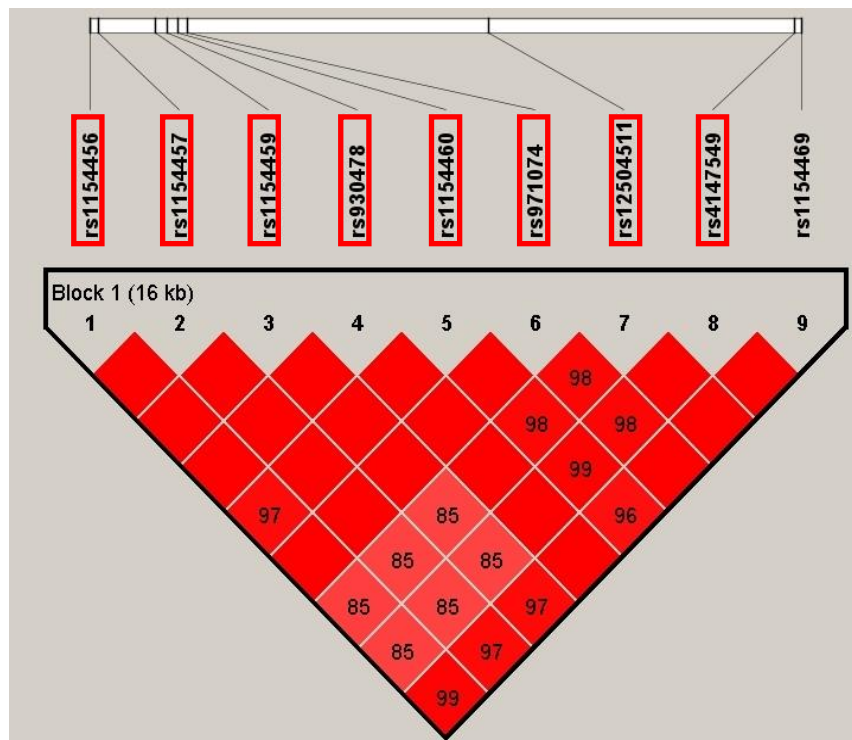

Figure S14. The LD plot of SNPs in gene *C4orf54*. SNPs showing significance are highlighted with red rectangles.
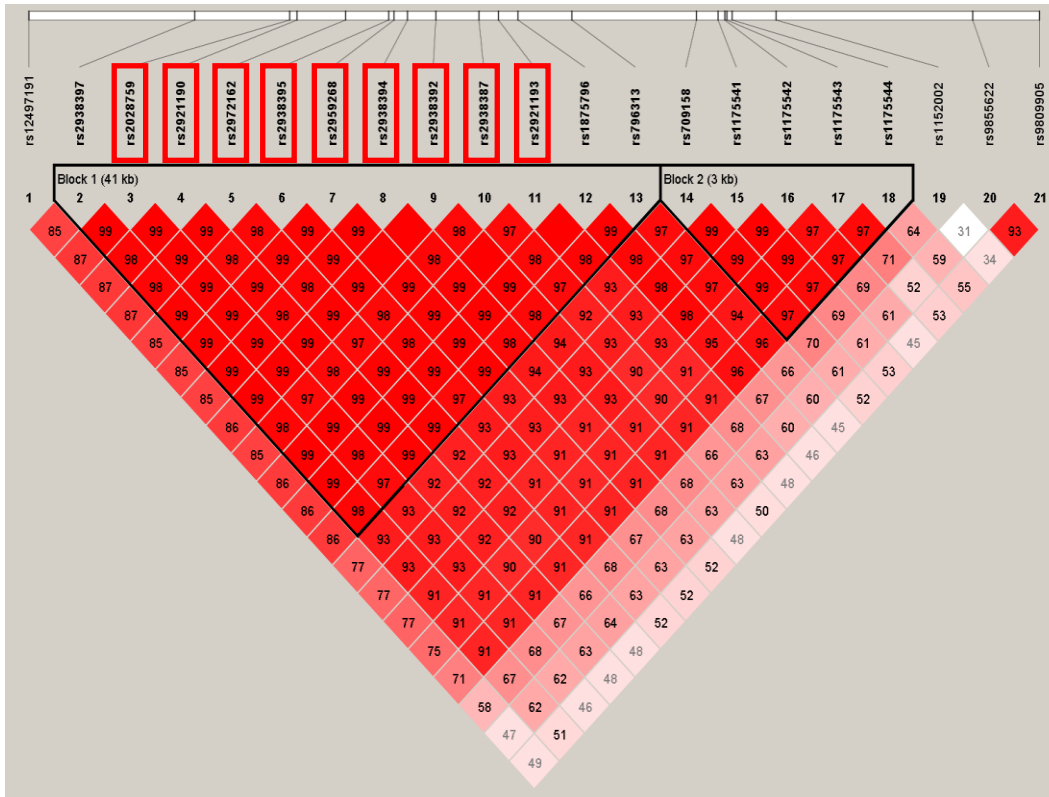
Figure S15. The LD plot of SNPs in gene *PPARG*. SNPs showing significance are highlighted with red rectangles.
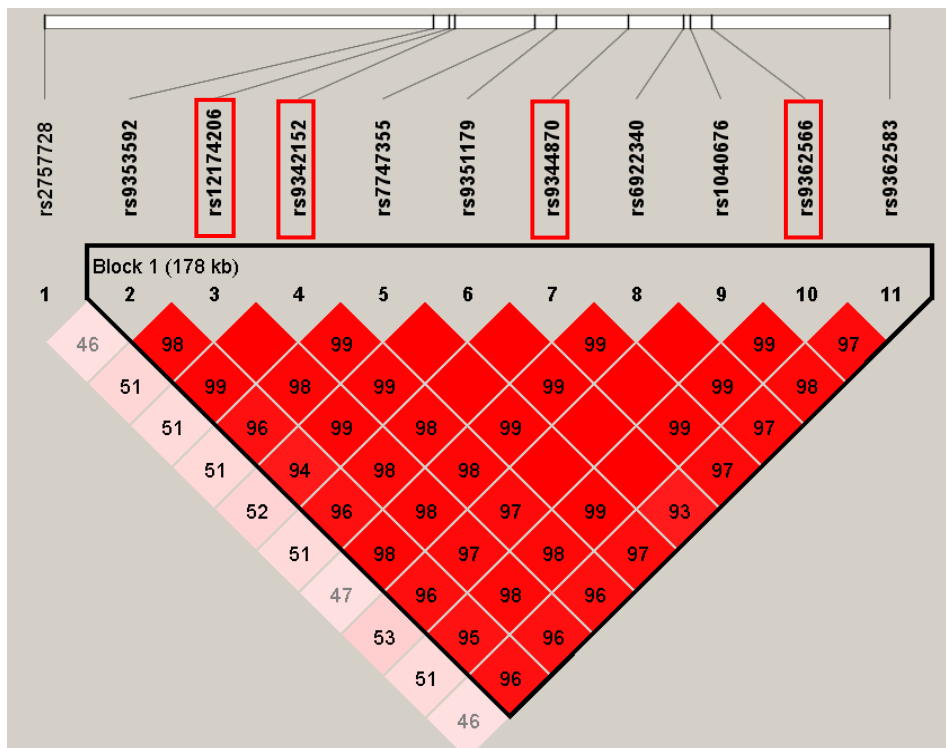

Figure S16. The LD plot of SNPs in gene *RNGTT*. SNPs showing significance are highlighted with red rectangles.

Reference

1. Zhao R, Xu L, Wu M et al. Maternal pre-pregnancy body mass index, gestational weight gain influence birth weight, Women and Birth 2018;31:e20-e25.

2. Sun Y, Shen Z, Zhan Y et al. Effects of pre-pregnancy body mass index and gestational weight gain on maternal and infant complications, BMC Pregnancy and Childbirth 2020;20:1-13.

3. Lumme R, Rantakallio P, Hartikainen A-L et al. Pre-pregnancy weight and its relation to pregnancy outcome, Journal of Obstetrics and Gynaecology 1995;15:69-75.

4. Onyekwelu J, Nwankwo CH, Oyeka I. Statistical analysis of birth weight and gender of newborn infants, Birth 2018;17:18.

5. Oldroyd J, Renzaho A, Skouteris H. Low and high birth weight as risk factors for obesity among 4 to 5-year-old Australian children: does gender matter? European Journal of Pediatrics 2011;170:899-906.

6. Lee I-L, Barr EL, Longmore D et al. Cord blood metabolic markers are strong mediators of the effect of maternal adiposity on fetal growth in pregnancies across the glucose tolerance spectrum: the PANDORA study, Diabetologia 2020;63:497-507.

7. Abo Amer A, Asar EH, Rashwan MT et al. Measurement of C-Peptide Level in The Umbilical Cord of Infants of Diabetic Mothers and Its Relationship to The Risk of Hypoglycemia, Benha Medical Journal 2024.

8. Stránecký V, Hoischen A, Hartmannová H et al. Mutations in ANTXR1 cause GAPO syndrome, The American Journal of Human Genetics 2013;92:792-799.

9. Johansson S, Iliadou A, Bergvall N et al. The association between low birth weight and type 2 diabetes: contribution of genetic factors, Epidemiology 2008;19:659-665.

10. Horikoshi M, Yaghootkar H, Mook-Kanamori DO et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism, Nature Genetics 2013;45:76-82.

11. Meirhaeghe A, Boreham CA, Murray LJ et al. A possible role for the PPARG Pro12Ala polymorphism in preterm birth, Diabetes 2007;56:494-498.

12. Verier C, Meirhaeghe A, Bokor S et al. Breast-Feeding Modulates the Influence of the Peroxisome Proliferator–Activated Receptor-γ (PPARG2) Pro12Ala Polymorphism on Adiposity in Adolescents: The Healthy Lifestyle in Europe by Nutrition in Adolescence (HELENA) cross-sectional study, Diabetes Care 2010;33:190-196.

13. Broholm C, Olsson AH, Perfilyev A et al. Epigenetic programming of adipose-derived stem cells in low birthweight individuals, Diabetologia 2016;59:2664-2673.

14. Mira MÂdF. Investigating cellular and molecular neuroendocrine signalling in the diabetic gut. 2021.

15. Needham EJ, Hingst JR, Onslev JD et al. Personalized phosphoproteomics of skeletal muscle insulin resistance and exercise links MINDY1 to insulin action, medRxiv 2024:2024.2003. 2011.24304084.

16. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods, Statistics in Medicine 2019;38:2074-2102.