

Supplementary Information

A Top 10 influential genes for a subset of AD patients

Table S1: Top 10 influential genes for a subset of AD patients (example rows).

Patient	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Gene 6	Gene 7	Gene 8	Gene 9	Gene 10
1	RRAS	FAM92A1P2	SUSD2	IGF1R	RIC8A	TAMM41	HS3ST3B1	ITPR3	MTTP	AC009495.2
2	RRAS	FAM92A1P2	SUSD2	TAMM41	ITPR3	MTTP	RIC8A	AC009495.2	HS3ST3B1	TMEM72
3	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	AC009495.2	TMEM72	RP11-312O7.2
4	RRAS	FAM92A1P2	TAMM41	ITPR3	SUSD2	RIC8A	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
5	RRAS	FAM92A1P2	TAMM41	ITPR3	SUSD2	RIC8A	HS3ST3B1	AC009495.2	TMEM72	RP11-312O7.2
6	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	RP11-312O7.2	TMEM72	AC009495.2
7	RRAS	FAM92A1P2	TAMM41	SUSD2	ITPR3	RIC8A	HS3ST3B1	RP11-312O7.2	AC009495.2	TMEM72
8	RRAS	FAM92A1P2	HS3ST3B1	TAMM41	SUSD2	ITPR3	RIC8A	AC009495.2	TMEM72	RP11-312O7.2
9	RRAS	FAM92A1P2	TAMM41	ITPR3	RIC8A	SUSD2	HS3ST3B1	AC009495.2	TMEM72	IGF1R
10	RRAS	FAM92A1P2	TAMM41	RIC8A	SUSD2	HS3ST3B1	AC009495.2	ITPR3	TMEM72	RP11-312O7.2
11	RRAS	FAM92A1P2	TAMM41	RIC8A	SUSD2	ITPR3	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
12	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	AC009495.2	TMEM72	RP11-312O7.2
13	RRAS	FAM92A1P2	TAMM41	ITPR3	SUSD2	RIC8A	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
14	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
15	RRAS	FAM92A1P2	TAMM41	SUSD2	RIC8A	HS3ST3B1	ITPR3	AC009495.2	TMEM72	RP11-312O7.2
16	RRAS	FAM92A1P2	TAMM41	HS3ST3B1	RIC8A	ITPR3	SUSD2	RP11-312O7.2	AC009495.2	TMEM72
17	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	AC009495.2	TMEM72	RP11-312O7.2
18	RRAS	FAM92A1P2	TAMM41	ITPR3	RIC8A	SUSD2	HS3ST3B1	TMEM72	AC009495.2	RP11-312O7.2
19	RRAS	FAM92A1P2	TAMM41	ITPR3	SUSD2	RIC8A	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
20	RRAS	FAM92A1P2	TAMM41	RIC8A	SUSD2	HS3ST3B1	AC009495.2	ITPR3	IGF1R	RP11-312O7.2
21	RRAS	FAM92A1P2	TAMM41	SUSD2	RIC8A	HS3ST3B1	AC009495.2	ITPR3	RP11-312O7.2	TMEM72
22	RRAS	FAM92A1P2	SUSD2	ITPR3	TAMM41	RIC8A	HS3ST3B1	AC009495.2	TMEM72	IGF1R
23	RRAS	FAM92A1P2	TAMM41	ITPR3	SUSD2	RIC8A	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72
24	RRAS	FAM92A1P2	TAMM41	SUSD2	RIC8A	HS3ST3B1	RP11-312O7.2	AC009495.2	ITPR3	TMEM72
25	RRAS	FAM92A1P2	TAMM41	SUSD2	ITPR3	RIC8A	HS3ST3B1	AC009495.2	TMEM72	RP11-312O7.2
26	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	RP11-312O7.2	AC009495.2	TMEM72
27	RRAS	FAM92A1P2	TAMM41	SUSD2	ITPR3	RIC8A	HS3ST3B1	AC009495.2	IGF1R	TMEM72
28	RRAS	FAM92A1P2	TAMM41	RIC8A	ITPR3	SUSD2	HS3ST3B1	AC009495.2	RP11-312O7.2	TMEM72

B ADNI preprocessing and genotype quality control

The primary ADNI analysis used baseline visits only. CN and AD subjects were retained according to the ADNI diagnosis table, with `DIAGNOSIS = 1` coded as CN and `DIAGNOSIS = 3` coded as AD. MCI subjects were excluded from the primary AD-versus-CN analysis. After PTID normalization, duplicate-subject removal, and matching to genotype files, the final analysis included 815 subjects with matched genotype and phenotype data.

Genotype data were assembled from multiple ADNI PLINK releases by retaining the common SNP intersection, aligning all release-specific genotype matrices to the same SNP order, and concatenating subjects across releases. Invalid loci were removed, and missing genotypes were mean-imputed by SNP. The common genotype intersection contained 75,339 SNPs, of which 68,059 passed all three SNP-level QC filters: call rate ≥ 0.95 , $MAF \geq 0.01$, and $HWE p \geq 10^{-6}$ in CN controls. Sample-level QC and relatedness summaries are shown in Figure S2, Figure S3, Table S4, and Table S5.

C Robustness checks for the smaller AD subgroup

Because the AD-only clustering produced an imbalanced two-cluster solution, with 232 patients in Cluster 1 and 19 patients in Cluster 2, we performed additional checks to determine whether the smaller subgroup could be explained by obvious non-biological outlier structure. We focused on three classes of potential explanations: ancestry stratification, baseline demographic or clinical imbalance, and genotype quality. All analyses were performed among the 251 AD patients assigned to the two Het-PR clusters.

Genetic ancestry was evaluated using genotype principal components computed from common variants. PC1–PC10 were compared between clusters using Mann–Whitney U tests, with Bonferroni correction across the ten tested PCs. None of the ten PCs differed significantly after correction (Table S2), indicating that

the smaller subgroup was not driven by gross ancestry stratification. Consistent with this result, the PC1–PC2 visualization shows that Cluster 2 lies within the main ancestry cloud rather than forming a separate ancestry outlier group (Figure S1).

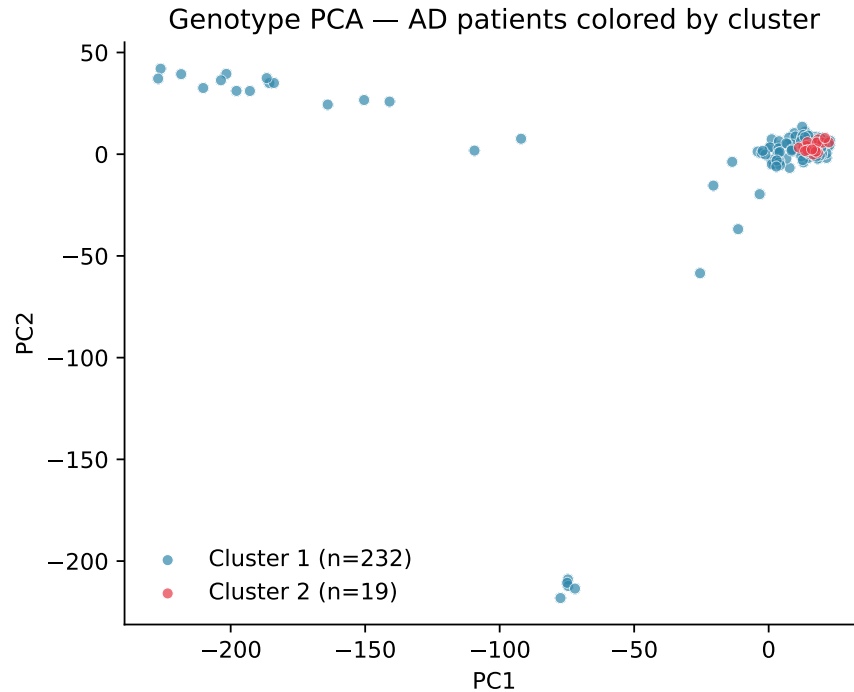


Figure S1: Genotype principal-component visualization of AD patients colored by Het-PR cluster assignment. Cluster 2 does not form a distinct ancestry-separated group in the PC1–PC2 space.

Table S2: Comparison of genotype principal components between the two AD clusters. PC1–PC10 were compared using Mann–Whitney U tests, with Bonferroni correction across ten PCs.

PC	Mean (Cluster 1)	Mean (Cluster 2)	U statistic	p-value	Bonferroni p-value
PC1	-1.3526	16.5156	1768.0	0.1523	1.0000
PC2	-0.2996	3.6586	2091.0	0.7116	1.0000
PC3	-0.6633	8.0991	1851.0	0.2466	1.0000
PC4	-0.1793	2.1892	2018.0	0.5421	1.0000
PC5	-0.0202	0.2463	2322.0	0.6994	1.0000
PC6	-0.5531	6.7541	1781.0	0.1649	1.0000
PC7	0.2865	-3.4984	2382.0	0.5596	1.0000
PC8	-0.3928	4.7969	1907.0	0.3298	1.0000
PC9	0.2224	-2.7157	2384.0	0.5552	1.0000
PC10	-0.4712	5.7531	1872.0	0.2759	1.0000

We next compared baseline demographic and clinical variables between the two clusters. Age at baseline, years of education, sex distribution, and baseline MMSE did not significantly differ between Cluster 1 and Cluster 2 (Table S3). These results argue against the interpretation that the smaller subgroup is explained by baseline demographic or clinical imbalance.

Finally, we evaluated sample-level genotype quality metrics, including genotype call rate, missingness rate, and heterozygosity F-statistic. Both clusters had high median call rates exceeding 99.7%. Cluster 2 showed a higher median call rate and lower missingness than Cluster 1, while heterozygosity did not significantly differ between clusters (Table S4; Figure S2). Thus, the smaller subgroup does not show evidence of being a low-quality genotype outlier group.

Table S3: Baseline demographic and clinical comparison between the two AD clusters. Continuous variables are reported as mean $\pm$ standard deviation. Sex was compared using Fisher’s exact test; continuous variables were compared using Mann–Whitney U tests.

Variable	Cluster 1 (n = 232)	Cluster 2 (n = 19)	p-value
Age at baseline (years)	75.2 $\pm$ 7.4	73.8 $\pm$ 10.8	0.717
Education (years)	14.8 $\pm$ 3.0	16.2 $\pm$ 2.8	0.090
Female sex, n (%)	129/232 (55.6%)	13/19 (68.4%)	0.340
Baseline MMSE	23.2 $\pm$ 2.3	22.7 $\pm$ 2.3	0.223

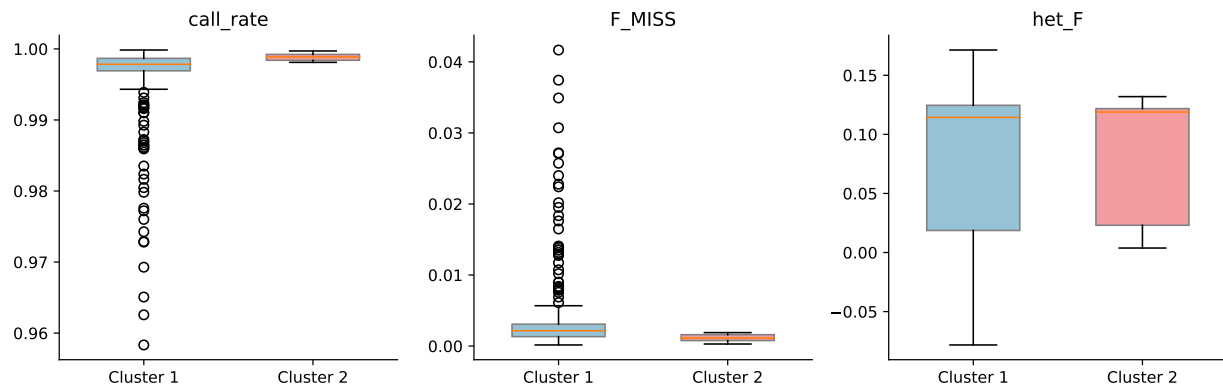


Figure S2: Sample-level genotype quality metrics by AD cluster. Cluster 2 does not show poorer genotype quality: both clusters have high call rates, Cluster 2 has lower missingness, and heterozygosity F-statistics do not differ significantly.

Although call rate and missingness differed between clusters, the direction of the difference does not support a poor-quality outlier interpretation: Cluster 2 showed higher call rate and lower missingness rather than poorer genotype quality. Heterozygosity F-statistics were also comparable between clusters.

We also summarized SNP-level QC and sample relatedness for the genotype data used in the AD analysis. SNP-level filters included per-SNP call rate in AD patients, minor allele frequency in AD patients, and Hardy–Weinberg equilibrium in CN controls. Of 75,339 SNPs in the common genotype intersection, 68,059 passed all three SNP-level QC filters. Hardy–Weinberg equilibrium was evaluated in CN controls rather than AD cases to avoid removing variants that may deviate from equilibrium because of disease association. Pairwise IBS analysis within the AD cohort found no likely duplicate pairs and no likely first-degree relatives (Table S5; Figure S3).

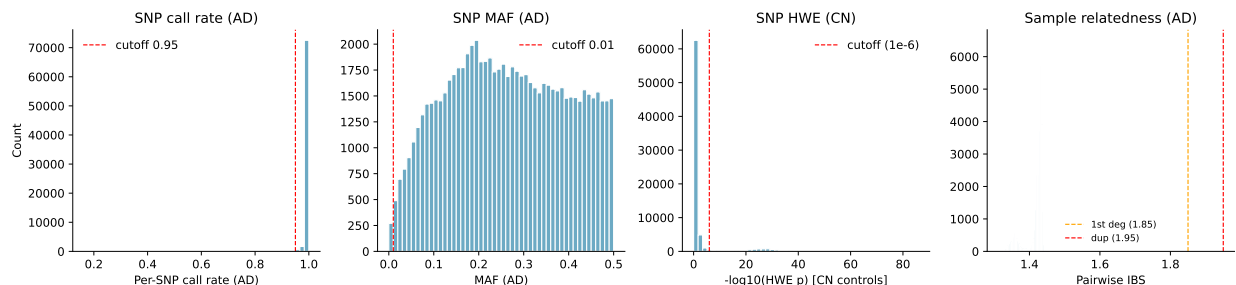


Figure S3: SNP-level QC and sample relatedness summaries. The panels show per-SNP call rate in AD patients, minor allele frequency in AD patients, Hardy–Weinberg equilibrium in CN controls, and pairwise IBS among AD patients.

Together, these analyses provide supporting robustness evidence for the smaller AD subgroup by exclud-

Table S4: Sample-level genotype quality comparison between the two AD clusters. Values are reported as median and interquartile range (IQR).

Metric	Cluster 1 median (IQR)	Cluster 2 median (IQR)	p-value
Call rate	0.99784 (0.00176)	0.99886 (0.00083)	8.46×10^{-5}
F_MISS	0.00216 (0.00176)	0.00114 (0.00083)	8.46×10^{-5}
Heterozygosity F	0.11434 (0.10594)	0.11890 (0.09874)	0.873

Table S5: SNP-level QC and AD sample relatedness summary.

QC item	Count	Percentage / threshold
Total SNPs in common genotype intersection	75,339	100.0%
Pass per-SNP call rate ≥ 0.95 in AD	74,827	99.3%
Pass MAF ≥ 0.01 in AD	75,068	99.6%
Pass HWE p-value $\geq 10^{-6}$ in CN controls	68,658	91.1%
Pass all three SNP-level QC filters	68,059	90.3%
AD pairwise IBS comparisons checked	31,375	–
Likely duplicate pairs, IBS ≥ 1.95	0	none detected
Likely first-degree relative pairs, IBS ≥ 1.85	0	none detected

ing several straightforward outlier explanations, including gross ancestry separation, baseline demographic imbalance, poor sample-level genotype quality, and duplicated or closely related AD samples.

C.1 Cluster robustness and sensitivity analysis

To evaluate whether the AD-only two-cluster solution was sensitive to patient composition, SNP composition, score representation, or nominal thresholding, we repeated the clustering under multiple perturbation schemes and compared each resulting partition with the reference AD-only clustering. Cluster agreement was quantified using adjusted Rand index (ARI), normalized mutual information (NMI), best label-matched accuracy, and minimum cluster-wise Jaccard similarity.

The reference clustering was exactly reproduced by the primary rerun of the analysis. In 500 patient-subsampling runs using 80% of AD patients, the clustering was recovered exactly on the overlapping samples. SNP-subsampling analyses also showed complete stability across all tested SNP-retention fractions. Alternative score representations, including raw score values, one-minus-score values, $-\log_{10}(p)$ -transformed values, and within-patient rank-transformed score values, produced the same two-cluster partition. Binary thresholding at $p \leq 0.10$ and $p \leq 0.05$ exactly reproduced the reference clustering, while $p \leq 0.01$ yielded a highly concordant solution (Table S6).

C.2 Dominant-SNP removal sensitivity analysis

We further evaluated whether the AD-only two-cluster solution was driven by a small number of dominant SNPs with unusually strong personalized association signals. SNPs were ranked according to three dominance criteria: mean $-\log_{10}(p)$ across AD patients, frequency of nominal threshold passing at $p \leq 0.05$, and variance of $-\log_{10}(p)$ across AD patients. For each ranking criterion, we removed the top 1, 5, 10, 25, 50, 100, 200, or 500 SNPs and repeated the same AD-only hierarchical clustering procedure.

The AD subgroup assignment was unchanged under all dominant-SNP removal settings. Even after removing the top 500 SNPs under each ranking criterion, the clustering retained the original 232:19 cluster-size split and achieved ARI = 1.000 and NMI = 1.000 relative to the original clustering. These results are summarized in Table S7 and Figure S4, indicating that the AD subgrouping was not driven by a small number of dominant variants, but instead reflected a distributed personalized SNP-profile structure.

Table S6: Robustness and sensitivity analysis of the AD-only clustering solution. Agreement with the reference clustering was quantified using adjusted Rand index (ARI), normalized mutual information (NMI), best label-matched accuracy, and Jaccard similarity.

Analysis	Setting	ARI	NMI	Accuracy	Min Jaccard
Primary reproduction	Reference rerun	1.000	1.000	1.000	1.000
Patient subsampling	500 runs, 80% AD patients	1.000	1.000	1.000	1.000
SNP subsampling	50% SNPs retained	1.000	1.000	1.000	1.000
SNP subsampling	70% SNPs retained	1.000	1.000	1.000	1.000
SNP subsampling	80% SNPs retained	1.000	1.000	1.000	1.000
SNP subsampling	90% SNPs retained	1.000	1.000	1.000	1.000
Score representation	Raw p-values	1.000	1.000	1.000	1.000
Score representation	One-minus-p values	1.000	1.000	1.000	1.000
Score representation	$-\log_{10}(p)$ values	1.000	1.000	1.000	1.000
Score representation	Within-patient ranks	1.000	1.000	1.000	1.000
Binary thresholding	$p \leq 0.10$	1.000	1.000	1.000	1.000
Binary thresholding	$p \leq 0.05$	1.000	1.000	1.000	1.000
Binary thresholding	$p \leq 0.01$	0.966	0.922	0.996	0.947

Table S7: Dominant-SNP removal sensitivity analysis for the AD-only clustering solution. SNPs were ranked by mean $-\log_{10}(p)$, nominal threshold-passing frequency, or variance of $-\log_{10}(p)$ across AD patients. The table reports the result after removing up to the top 500 SNPs under each ranking criterion.

Removal ranking	Max removed	Features kept	Final cluster size	Min ARI	Min NMI
Mean $-\log_{10}(p)$	500	1184	232 vs 19	1.000	1.000
Threshold-passing frequency	500	1184	232 vs 19	1.000	1.000
Variance of $-\log_{10}(p)$	500	1184	232 vs 19	1.000	1.000

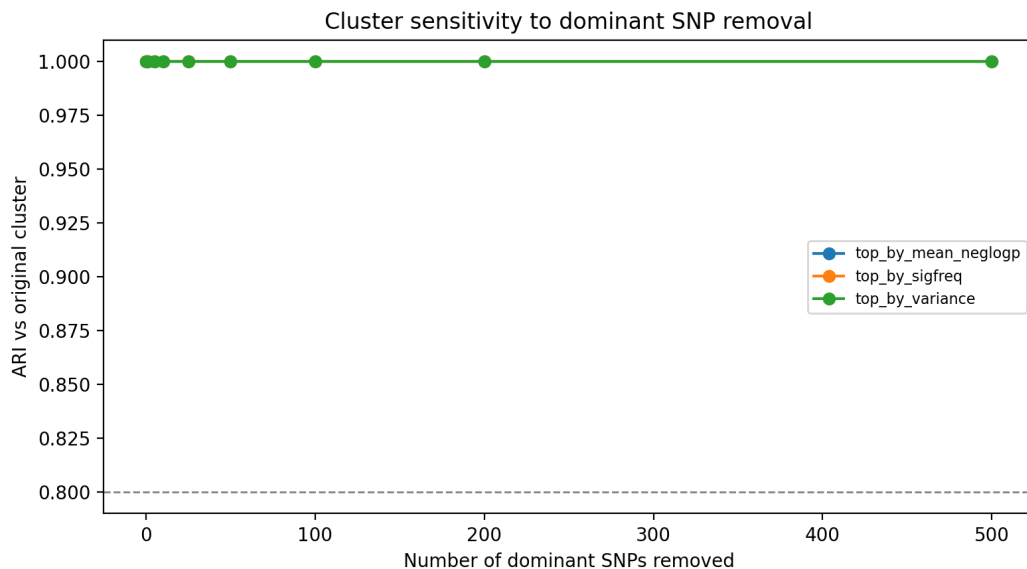


Figure S4: Dominant-SNP removal sensitivity analysis. The AD-only clustering remained identical to the original clustering after removing up to the top 500 SNPs ranked by mean $-\log_{10}(p)$, nominal significance frequency, or variance across AD patients.

D Simulation Results

D.1 Simulation Data Generation

We first validate our method in comparison to several others using simulated data. To begin, we consider a simplified scenario: instead of giving each individual their own distinct oracle effect size vector, we partition the individuals into g cohorts, such that all individuals within the same cohort share similar parameter values. We generate the simulation data as follows. First, we draw n samples, each with p features sampled from a uniform distribution, forming the feature matrix X . We then evenly split these n samples into g cohorts, and assign to each cohort a unique effect size vector β of length p with exactly k non-zero entries (i.e., a sparsity level of k). To construct each of the k non-zero values in an effect size vector, we draw a scalar $a \sim \mathcal{N}(0, 1)$ and set

$$\beta^{(i)} = \begin{cases} a \cdot s + s, & \text{if } a > 0, \\ a \cdot s - s, & \text{otherwise,} \end{cases} \quad (15)$$

where s is a scaling factor controlling the magnitude of the effect size. All samples within a cohort share the same effect size vector β for that cohort. Consequently, for the i th sample, we generate its response (label) Y_i as

$$Y_i = X_i B_i + \varepsilon, \quad (16)$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. In our notation, B denotes the matrix of all effect size vectors β (one for each cohort), and σ_ε is a hyperparameter specifying the noise level in the simulation. Finally, to evaluate our model’s behavior in this setting, we must provide covariate information (analogous to “demographics” for each sample). In the original idealized benchmark, we used the oracle cohort assignment information (the true effect size pattern B for each sample) as the covariate matrix C . This benchmark evaluates whether Het-PR can recover heterogeneous effects when the sample-similarity structure is correctly specified. Because this setting may favor the proposed method by providing oracle subgroup information, we additionally performed sensitivity analyses in which C was replaced by noisy group covariates, feature-derived principal components, or random uninformative covariates. We also evaluated model-misspecification settings involving nonlinear interaction effects and a homogeneous no-subgroup structure, as described in Section D.3.

D.2 Simulation Result

Our primary interest is the ability of each method to recover the underlying “genetic” pattern of the disease, which in the simulation corresponds to identifying the non-zero coefficient entries. We therefore assess each method by how well it detects the true non-zero features, using the Area Under the ROC Curve (AUC-ROC) as the evaluation metric. This metric compares the ranked feature “importance” scores produced by a method against the ground-truth sparsity pattern of the oracle coefficients. In particular, depending on the nature of a given method, we define a feature’s “importance” either as its p -value (when the method provides significance measures) or as the absolute value of its estimated coefficient (when p -values are not available). As discussed earlier, our Het-PR framework can incorporate different base models. We evaluate two versions in our experiments: “**p1 pvalue**,” which uses a variant of a linear mixed model as the base (capable of reporting p -values even in high-dimensional settings), and “**p1 beta**,” which uses a standard linear regression as the base model (reporting raw coefficient estimates). We compare these against several baseline methods: **dmr**, a personalized regression approach based on distance matching; **par**, a penalized angular regression method for personalized modeling; a clustering-based regression baseline (**cr**); and a traditional population-level regression model (**pr**). The simulation results show that our Het-PR method consistently outperforms all of these competing models in accurately recovering the true sparse coefficient pattern across a wide range of parameter settings. In particular, we conducted five replicate simulation runs (with different random seeds) for each combination of parameters, varying the sample size (n), number of features (p), sparsity level (k), and number of cohorts (denoted as “Block Num”). In total, we evaluated **27** distinct parameter configurations, each repeated **5** times. The overall performance distribution across the methods is summarized in Figure S5, and the detailed performance for each specific parameter setting is shown in Figure S6. In nearly all cases, our Het-PR approach achieves higher AUC-ROC scores than the other methods except telling cr true clusters, indicating better recovery of the non-zero features.

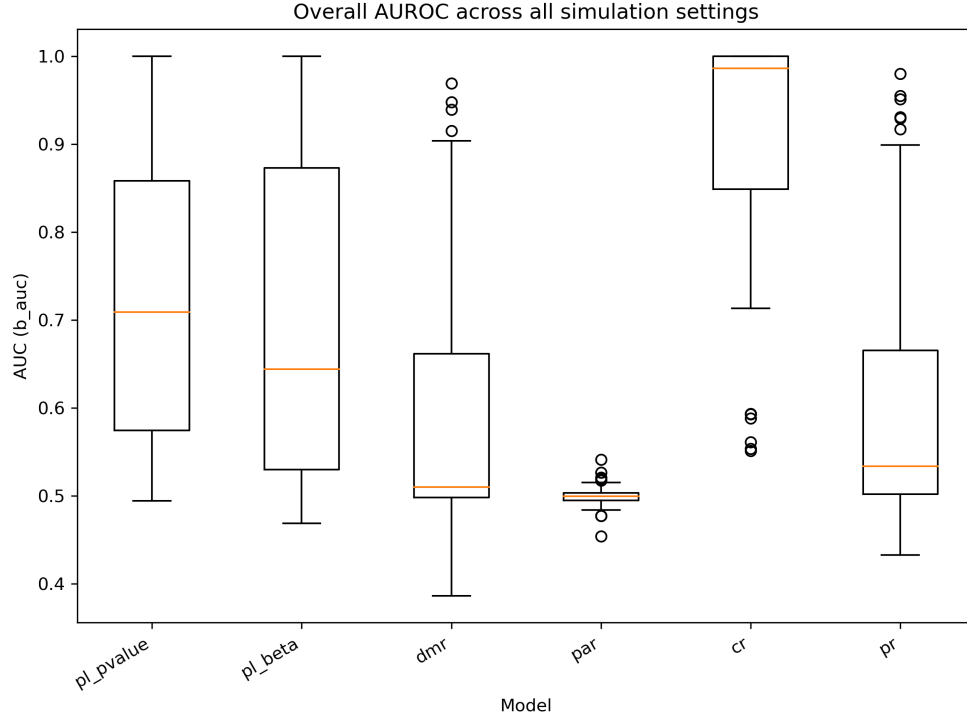


Figure S5: Overall performance distribution of the simulation tests across six different modeling approaches. **pl pvalue** refers to the personalized regression with a linear mixed model base using model-reported score quantities to calculate the AUC-ROC; **pl beta** is the personalized regression with a linear regression base (using estimated coefficients to calculate the AUC-ROC); **dmr** denotes the personalized regression with distance matching; **par** denotes the penalized angular regression method; **cr** is a clustering regression baseline; and **pr** is the standard population-level regression.

b_auc by model for each simulation setting

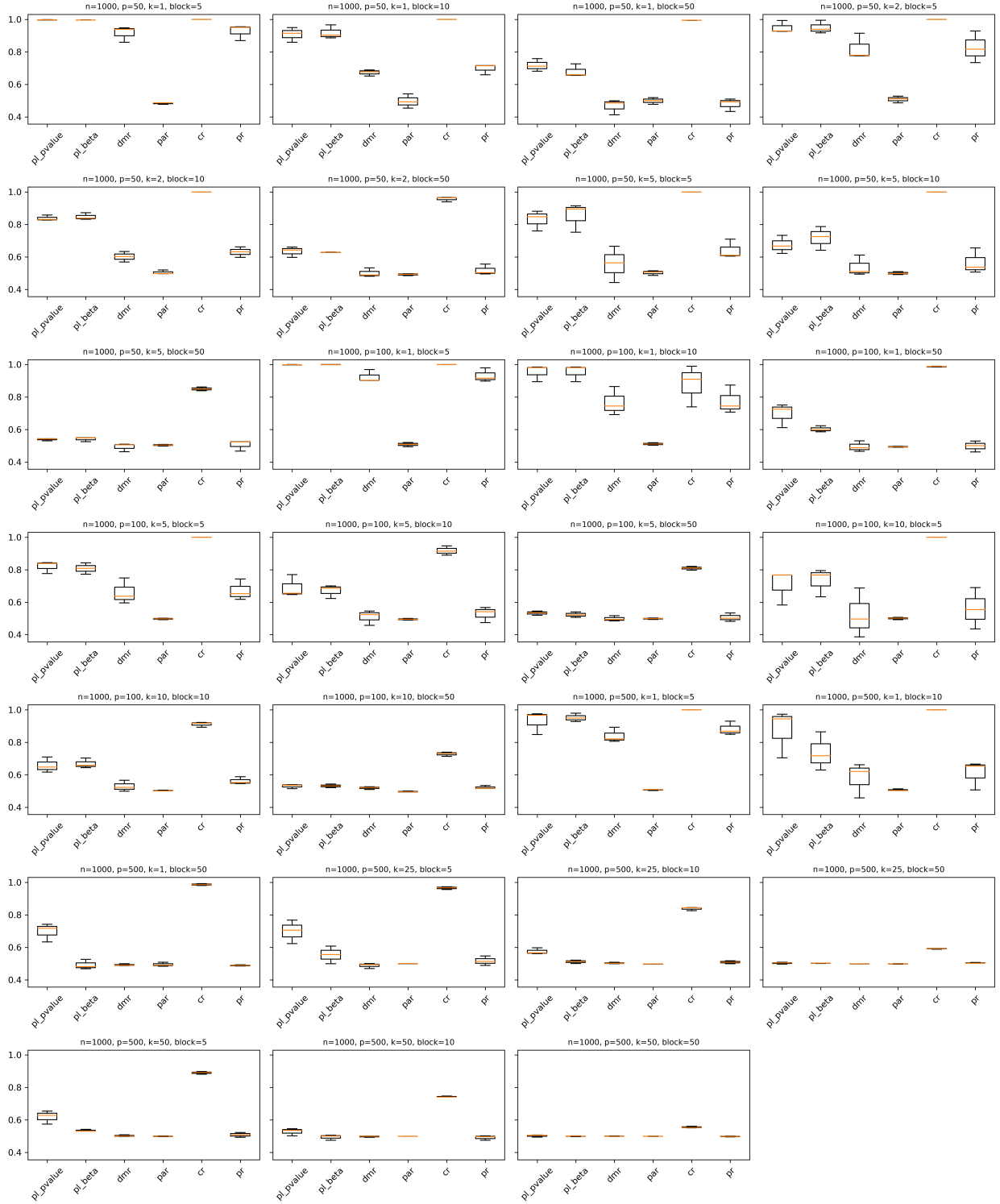


Figure S6: Model performance under different simulation parameter configurations. There are 27 parameter sets in total, each described in the title of its corresponding sub-figure. We conducted five simulation runs for each parameter set using different random seeds.

D.3 Sensitivity to non-oracle auxiliary covariates and model misspecification

Because the original structured-subgroup simulation used oracle cohort information as the auxiliary covariate matrix C , we performed additional sensitivity analyses to evaluate whether the performance of Het-PR depended on direct access to the true grouping structure. We compared the oracle- C benchmark with three non-oracle alternatives: noisy group covariates, feature-derived principal components, and random uninformative covariates. We also evaluated two additional model-misspecification settings: a nonlinear-interaction data-generating mechanism and a homogeneous no-subgroup setting.

Across structured-subgroup simulations, Het-PR retained favorable sample-level causal-SNP recovery under non-oracle C settings (Table S8). With uniformly generated features, Het-PR achieved mean AUROC values of 0.855 using noisy group C with noise level 0.10, 0.843 using noisy group C with noise level 0.30, 0.843 using feature-derived C , and 0.841 using random C , compared with 0.806 for the population-level baseline. The corresponding oracle- C benchmark achieved 0.864. Thus, although oracle C gave the strongest benchmark performance, Het-PR did not collapse when C was noisy, feature-derived, or uninformative.

We also evaluated structured-subgroup simulations using real-genotype features sampled from the ADNI genotype-derived feature matrix. In this more genetic-like feature setting, Het-PR achieved mean AUROC values of 0.924 using noisy group C , 0.922 using feature-derived C , and 0.914 using random C , compared with 0.901 for the population-level baseline. The corresponding oracle- C benchmark achieved 0.945. These results indicate that the simulation advantage of Het-PR was not solely driven by oracle access to true subgroup labels and was also observed under real-genotype feature structure.

To evaluate robustness to nonlinear model misspecification, we further considered a nonlinear-interaction data-generating mechanism. Under this setting, Het-PR achieved mean AUROC values of 0.816 using feature-derived C and 0.819 using random C , compared with 0.785 for the population-level baseline. These results suggest that Het-PR retained favorable causal-SNP recovery under nonlinear effect misspecification.

Finally, to evaluate behavior when there was no clear subgroup structure, we considered a homogeneous no-subgroup setting. In this setting, Het-PR achieved mean AUROC values of 1.000 using either feature-derived C or random C , matching the population-level baseline. Thus, when subgroup heterogeneity was absent, the personalized framework did not degrade causal-SNP recovery relative to the population-level model.

Table S8: Sensitivity of Het-PR simulation performance to non-oracle auxiliary covariates and model-misspecification settings. Values are mean sample-level causal-SNP AUROC across three random seeds.

Scenario	Feature source	C setting	C noise	Het-PR AUC	Population AUC
Structured subgroups	Uniform	Oracle group	0.00	0.864	0.806
Structured subgroups	Uniform	Noisy group	0.10	0.855	0.806
Structured subgroups	Uniform	Noisy group	0.30	0.843	0.806
Structured subgroups	Uniform	Feature PCA	0.00	0.843	0.806
Structured subgroups	Uniform	Random	0.00	0.841	0.806
Structured subgroups	Real genotype	Oracle group	0.00	0.945	0.901
Structured subgroups	Real genotype	Noisy group	0.30	0.924	0.901
Structured subgroups	Real genotype	Feature PCA	0.00	0.922	0.901
Structured subgroups	Real genotype	Random	0.00	0.914	0.901
Nonlinear interaction	Uniform	Feature PCA	0.00	0.816	0.785
Nonlinear interaction	Uniform	Random	0.00	0.819	0.785
Homogeneous no subgroup	Uniform	Feature PCA	0.00	1.000	1.000
Homogeneous no subgroup	Uniform	Random	0.00	1.000	1.000

E Y-permutation negative-control analysis

To provide an empirical null reference for the Het-PR individualized score summaries, we performed 50 full-data Y -permutation refits. In each refit, the AD/CN outcome labels were randomly permuted, the same

covariate residualization procedure was repeated, and Het-PR was refitted using the same X and C matrices. We then recomputed the fraction of AD patient-SNP scores passing the nominal reference threshold $p \leq 0.05$.

The observed fraction of AD patient-SNP scores below 0.05 was 5.66%, compared with a Y -permutation null median of 4.13% and range of 2.49%–10.83% (Table S9). This result provides an empirical null reference for the score-threshold summaries used in the main analysis and shows that the observed threshold-passing rate was close to the Y -permutation null distribution at the nominal $p \leq 0.05$ reference level.

Table S9: Y -permutation negative-control analysis of Het-PR individualized score summaries. Fifty full-data Y -permutation refits were performed. The observed refit used the original AD/CN labels.

Metric	Observed	Null median	Null range	Empirical $p_{\geq obs}$
AD score fraction $p \leq 0.05$	0.0566	0.0413	0.0249–0.1083	0.2549

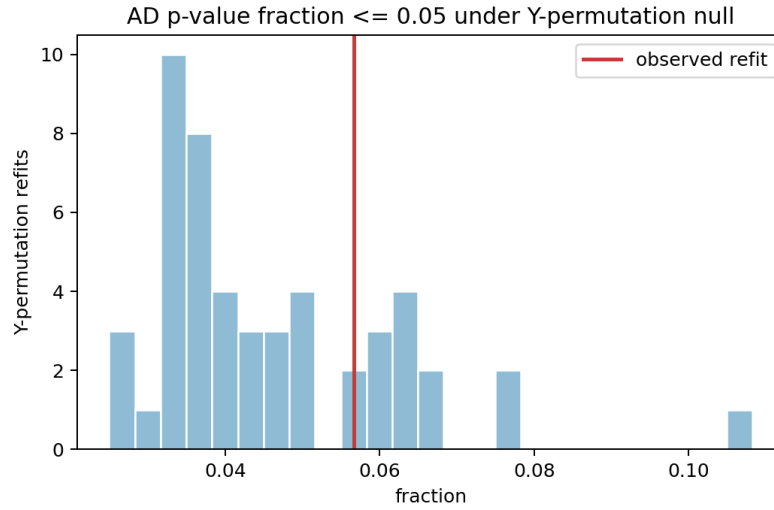


Figure S7: Fraction of AD patient-SNP Het-PR individualized scores passing the nominal threshold $p \leq 0.05$ under the Y -permutation null.