GENE2BRAIN-AE: SELF-SUPERVISED REPRESENTATION LEARNING FOR GENETICALLY-INFORMED PATTERNS OF BRAIN ATROPHY

Anonymous authors
Paper under double-blind review

ABSTRACT

Characterizing heterogeneity in neurodegenerative diseases from imaging alone is limited, as similar patterns of brain change can arise from distinct underlying mechanisms. Integrating genetic information offers a way to disentangle these processes, but genetic effects are indirect, partially mediated, and entangled with non-genetic sources of variation. We propose Gene2Brain-AE (G2B-AE), a self-supervised multimodal framework that combines an interpretable NMF-style imaging autoencoder with a genetic variational autoencoder, coupled through a directional predictor that maps genetic embeddings into a subset of imaging latent factors. This directional, predictive coupling reflects the asymmetry of the underlying biological pathway and decomposes imaging variation into genetically predictable and residual components, without requiring genetic input at inference time. On semi-synthetic data, G2B-AE recovers genetic and residual patterns that symmetric alignment-based variants entangle or absorb into the wrong subspace. Applied to ADNI, the model identifies multiple genetically-guided atrophy axes, including distinct APOE-associated patterns, alongside a residual axis linked to vascular and metabolic variables. Despite using no genetic input at test time, these components' clinical, cognitive, and neuropathological signatures replicate across eight independent external cohorts, indicating generalizable biology rather than dataset-specific correlations. These results show that directional, partially shared latent representations offer a biologically-informed and interpretable characterization of disease heterogeneity. Code will be released upon acceptance.

1 INTRODUCTION

Neurodegenerative diseases such as Alzheimer's disease arise from multiple interacting biological and environmental processes that produce heterogeneous patterns of brain atrophy and cognitive decline. Genetic variation contributes to this heterogeneity by influencing upstream susceptibility to disease-related processes, while vascular risk, metabolic dysfunction, and lifestyle exposures affect brain structure through distinct mechanisms (Wen et al., 2023; 2024; 2025). As a result, similar imaging phenotypes may arise from different combinations of genetic and non-genetic influences (Lenroot & Giedd, 2008; Murray et al., 2011). A central challenge is therefore to characterize which aspects of observed brain variation reflect a given disease mechanism, such as inherited genetic risk, and which arise from other sources (Fig 1a).

Magnetic resonance imaging (MRI) captures downstream anatomical changes reflecting the aggregate effects of multiple processes (Yang et al., 2024b). Non-negative matrix factorization (NMF) has become a standard, interpretable tool for characterizing brain structure: it decomposes MRI into additive, spatially coherent patterns of structural covariance, together with subject-specific loadings, learning a data-driven dictionary of brain organization (Sotiras et al., 2015; 2017; Varikuti et al., 2018; Skampardoni et al., 2026). However, these patterns are organized by anatomical covariance alone and are agnostic to the mechanisms that produce atrophy. Genetic effects are typically studied afterwards, by testing the loadings of imaging-derived components for genetic association (Wen et al., 2023), so they can only be detected in patterns that were defined without genetics. Other imaging-genetics approaches either model direct feature-level associations between single nucleotide polymorphisms (SNPs) and voxels or regions (Stein et al., 2010; Zhang et al., 2019; Du

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

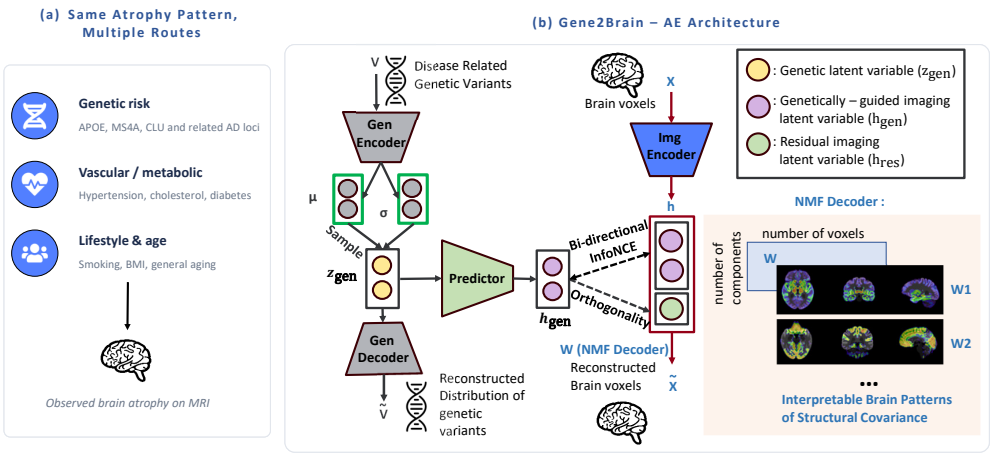


Figure 1: Overview of Gene2Brain-AE. (a) Similar atrophy on MRI can arise from genetic and non-genetic routes, and genetic effects reach the brain only partially and indirectly. (b) A frozen genetic VAE embeds SNPs (z_{gen}). A predictor maps z_{gen} to the genetically guided imaging latents h_{gen}, while residual latents h_{res} capture the remaining variation. An NMF-style decoder W reconstructs voxel-wise maps from $[h_{gen}, h_{res}]$, yielding interpretable spatial components.

et al., 2016), fuse the modalities for supervised prediction at the cost of interpretable spatial structure (Lee et al., 2023; Ko et al., 2022; Qiang et al., 2023; Wu et al., 2023; Li et al., 2025), or factorize them jointly under symmetric shared loadings (Wang & Wang, 2024; Tu et al., 2024). None of these let genetic information shape an interpretable decomposition of brain structure while respecting that genetic effects reach the brain only indirectly and partially.

We propose Gene2Brain-AE (G2B-AE), a self-supervised framework that adds genetic guidance to an NMF-style decomposition of brain structure, reorganizing it around the structures that genetic risk variants affect. The central challenge is that genetic effects are partial and mediated: much genetic variation has no imaging correlate, and much imaging variation has no genetic cause. Forcing the two modalities into symmetric alignment therefore distorts the decomposition, absorbing genetic confounders into imaging components or leaking non-genetic brain atrophy into genetic ones. G2B-AE instead couples them directionally. A variational autoencoder embeds high-dimensional, sparse SNP data (Yang et al., 2024a), and a predictor, inspired by joint-embedding predictive architectures (Lecun, 2022; Chen et al., 2025), maps these embeddings onto only a subset of the imaging latent factors. This splits the decomposition into genetically guided components and a residual subspace that captures variation not explained by the modeled genetic signal. Because genetics guides training but is not needed at inference, the learned patterns can be measured from imaging alone.

The resulting framework retains the interpretability of NMF at both the population and individual levels, learning voxel-wise spatial patterns together with subject-specific loadings, while changing what those patterns represent. Without genetic input, the decomposition behaves like a data-driven soft parcellation, grouping voxels by anatomical covariance. Genetic guidance reorganizes it around shared genetic drivers, yielding distributed and partially overlapping networks that reflect which structures genetic risk variants jointly affect. Our contributions are:

- **Genetically guided structural decomposition.** We introduce a self-supervised framework in which genetic information shapes an interpretable, NMF-style decomposition of brain structure itself. The model operates directly on voxel-wise data, pairing the interpretability of linear decompositions with disentanglement of the latent space, and requires only imaging once trained.

- **Directional, partial coupling.** We model the gene-to-brain pathway through directional latent prediction rather than symmetric alignment, separating genetically guided from residual imaging variation and avoiding confounder absorption and residual leakage.

- **Empirical validation.** On ADNI, it yields genetically guided atrophy axes, including distinct APOE-associated patterns, whose clinical, cognitive, and neuropathological signatures replicate across eight diverse external cohorts without genetic input.

2 Related Work

Regression and mediation models. Imaging-genetics has long modeled genetic effects on brain structure with multivariate regression or mediation analysis (Batmanghelich et al., 2013; Zhou et al., 2019), including sparse, reduced-rank, and kernel-based approaches (Vounou et al., 2010; Du et al., 2016; Ge et al., 2015). These encode the directional genetics-to-imaging pathway and yield interpretable associations, but typically operate in the original high-dimensional feature space (SNPs to voxels or ROIs) with linear or weakly nonlinear models, capturing feature-level associations rather than the latent mechanisms through which genetic effects reach brain structure.

Supervised deep multimodal models. Deep models jointly learn from genetics and imaging for tasks such as disease classification, typically via symmetric fusion (shared latent spaces, early/late fusion, attention) (Lee et al., 2023; Qiang et al., 2023; Wu et al., 2023; Li et al., 2025). They treat modalities as interchangeable, ignoring the directional genotype-to-phenotype pathway, and supervision biases representations toward task-relevant signal, potentially discarding biologically meaningful variation. Mediation-inspired variants let genetic features modulate imaging representations, but remain supervised and do not disentangle genetically predictable from residual variation (Ko et al., 2022; Wang et al., 2023). Gene-SGAN (Yang et al., 2024a) is, to our knowledge, the only deep method learning joint imaging–genetic patterns without a prediction objective. It is a weakly supervised (CN vs. diseased) clustering model requiring imaging features of dimensionality comparable to genetics.

Self-supervised multimodal learning. CCA (Li et al., 2020; Zhou et al., 2023), joint matrix factorization (Wang & Wang, 2024; Tu et al., 2024), and contrastive learning (Hager et al., 2023; Obaido et al., 2026) model cross-modal relationships without labels, but are typically symmetric: enforcing alignment can introduce spurious correspondences and ignores the directional, partially mediated nature of biological processes. Joint Embedding Predictive Architectures (JEPA) instead predict one view's embedding from another through a predictor. Multimodal variants such as VL-JEPA (Lecun, 2022; Chen et al., 2025) combine a frozen source encoder, a predictor, and an bi-directional InfoNCE objective, the design we adapt. However, they do not disentangle shared from modality-specific factors and remain underexplored in structured biological settings like imaging genetics. We adapt this directional, embedding-level prediction to disentangle genetically predictable from residual imaging variation without labels.

3 Methodology

An overview is shown in Fig. 1b. Our goal is a latent representation that separates genetically driven structural variation from residual non-genetic effects. Let $\boldsymbol{X}_{img} \in \mathbb{R}^{N\times V}$ (V voxels) and $\boldsymbol{X}_{snp} \in \mathbb{R}^{N\times G}$ (G SNPs) denote subject-level imaging and genetic inputs. Rather than aligning the modalities symmetrically, genetic representations *predict* a subset of imaging latent factors, reflecting that only part of genetic variation is expressed in brain structure.

3.1 Genetic Representation Learning

SNPs are encoded as $\{0, 1, 2\}$ allele counts, a sparse, discrete representation in which many loci vary little across individuals. Following Gene-SGAN (Yang et al., 2024a), which similarly maps curated SNP sets into a compact latent space via variational inference, we learn a probabilistic rather than deterministic embedding with a VAE: $\boldsymbol{x}_{snp} \in \mathbb{R}^G$ is mapped to a diagonal-Gaussian posterior $q(\boldsymbol{z}_{gen} \mid \boldsymbol{x}_{snp})$ over $\boldsymbol{z}_{gen} \in \mathbb{R}^K$, and counts are reconstructed with a binomial likelihood:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q(\boldsymbol{z}_{gen}|\boldsymbol{x}_{snp})}\left[-\log p(\boldsymbol{x}_{snp} \mid \boldsymbol{z}_{gen})\right] + \beta\, D_{\text{KL}}(q(\boldsymbol{z}_{gen} \mid \boldsymbol{x}_{snp}) \,\|\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}))\,, \tag{1}$$

with β linearly warmed up (Appendix A.2.4) and a sigmoid before decoding constraining embeddings to $[0, 1]$. After training, the encoder E_{gen} is frozen and yields subject embeddings $\boldsymbol{z}_{gen} = E_{gen}(\boldsymbol{x}_{snp}) \in \mathbb{R}^K$.

3.2 Genetically Informed Imaging Representation Learning

NMF factorizes nonnegative imaging data into additive spatial components and subject-specific loadings, $\boldsymbol{X}_{img} \approx \boldsymbol{H}\boldsymbol{W}^\top$ with $\boldsymbol{H} \in \mathbb{R}_+^{N\times C}$ and $\boldsymbol{W} \in \mathbb{R}_+^{V\times C}$, yielding interpretable part-based structure (Sotiras et al., 2015). We build this structure into an autoencoder trained directly on voxel-wise maps $\boldsymbol{x}_{img} \in \mathbb{R}^V$. The encoder outputs $[\boldsymbol{h}_{gen}, \boldsymbol{h}_{res}] = E_{img}(\boldsymbol{x}_{img})$, where $\boldsymbol{h}_{gen} \in \mathbb{R}^{K_g}$ captures genetically driven variation and $\boldsymbol{h}_{res} \in \mathbb{R}^{K_r}$ residual structure, with sigmoid activations enforcing non-negativity. A linear decoder reconstructs $\hat{\boldsymbol{x}}_{img} = \boldsymbol{W}[\boldsymbol{h}_{gen}, \boldsymbol{h}_{res}]$, where $\boldsymbol{W} \in \mathbb{R}_+^{V\times(K_g+K_r)}$ is initialized with non-negative SVD and kept non-negative, so its columns act as NMF spatial bases: the interpretability of additive factorization with the flexibility of an autoencoder.

3.3 Predictive Genetic-to-Imaging Coupling

Rather than aligning $\boldsymbol{z}_{gen}$ and $\boldsymbol{h}_{gen}$ directly, a predictor maps genetic representations into the imaging latent space, $\hat{\boldsymbol{h}}_{gen} = P(\boldsymbol{z}_{gen}) = \sigma(\mathrm{MLP}(\boldsymbol{z}_{gen}))$, whose output sigmoid matches the $[0, 1]$ range of $\boldsymbol{h}_{gen}$. Coupling is enforced with a bidirectional InfoNCE loss (van den Oord et al., 2019) (Appendix A.1):

$$\mathcal{L}_{\text{link}} = \tfrac{1}{2}\left(\mathcal{L}_{\text{NCE}}(\boldsymbol{h}_{gen}, \hat{\boldsymbol{h}}_{gen}) + \mathcal{L}_{\text{NCE}}(\hat{\boldsymbol{h}}_{gen}, \boldsymbol{h}_{gen})\right), \tag{2}$$

which updates the predictor and imaging encoder while the genetic encoder stays fixed, reducing spurious alignment and degenerate solutions. This follows the multimodal JEPA recipe of VL-JEPA (Chen et al., 2025): a frozen source encoder, a predictor into the target embedding space, and InfoNCE in that space. Symmetric alignment requires matched dimensionality and forces every genetic dimension to have an imaging counterpart. P instead maps K to K_g dimensions, so it can discard genetic variation with no imaging correlate (e.g., confounders) rather than forcing the imaging side to represent it.

3.4 Objective Function and Training

Imaging reconstruction uses mean squared error with a soft orthonormality constraint on the decoder (Sotiras et al., 2015):

$$\mathcal{L}_{\text{img}} = \lambda_{ae}\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{x}_{img}^{(i)} - \hat{\boldsymbol{x}}_{img}^{(i)}\right\|_2^2 + \lambda_W\left\|\boldsymbol{W}^\top\boldsymbol{W} - \boldsymbol{I}\right\|_F^2. \tag{3}$$

To disentangle the two subspaces, we penalize the cross-covariance between genetic and residual latents across subjects:

$$\mathcal{L}_{h\text{-ortho}} = \left\|\tfrac{1}{N-1}\tilde{\boldsymbol{H}}_{gen}^\top\tilde{\boldsymbol{H}}_{res}\right\|_F^2 = \sum_{j=1}^{K_g}\sum_{k=1}^{K_r}\widehat{\mathrm{Cov}}(h_{gen,j},\, h_{res,k})^2, \tag{4}$$

where $\tilde{\boldsymbol{H}}_{gen} \in \mathbb{R}^{N\times K_g}$ and $\tilde{\boldsymbol{H}}_{res} \in \mathbb{R}^{N\times K_r}$ stack the column-centered latents of all subjects. This soft penalty discourages leaked or redundant representations without overpowering the reconstruction and link losses, and still lets a subject express both genetic and residual factors. An ℓ_2 penalty $\mathcal{L}_{\ell_2} = \lambda_{\ell_2 g}\|\boldsymbol{h}_{gen}\|_2^2 + \lambda_{\ell_2 r}\|\boldsymbol{h}_{res}\|_2^2$ controls latent magnitude so that neither subspace dominates optimization. The full objective combines reconstruction, directional coupling, disentanglement, and regularization:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{img}} + \lambda_{\text{link}}\mathcal{L}_{\text{link}} + \lambda_{h\text{-}ortho}\mathcal{L}_{h\text{-ortho}} + \mathcal{L}_{\ell_2}, \tag{5}$$

with loss weights selected as described in the following section.

Training has two stages: the genetic VAE is trained alone, then the imaging autoencoder and predictor are trained with Eq. 5 while the genetic embedding stays fixed. This asymmetry mirrors biology: genetic variation influences atrophy, but atrophy does not map back onto genetics, so imaging need not reconstruct or align with the full genetic representation. It also handles the imbalance between high-dimensional voxel data and low-dimensional genetic signal architecturally, so imaging variation cannot dominate the genetic representation, and it avoids aggregating imaging into ROIs, which would fix the spatial scale a priori. Finally, since imaging-only cohorts are far more available than paired genetic data, the trained imaging encoder is can be applied at inference without genetic input.

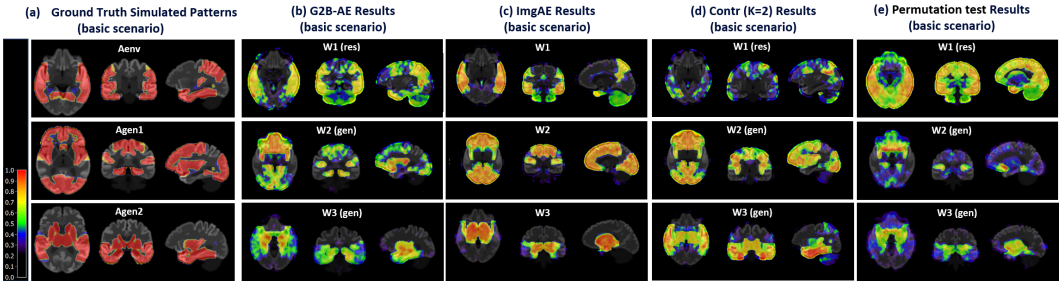


Figure 2: Basic scenario qualitative results. (a) Ground truth; (b) G2B-AE; (c) imaging-only (imgAE); (d) Symmetric alignment (Contr K=2); (e) G2B-AE with permuted genetics. The lateral temporal lobe, shared by A_{gen2} and A_{env}, appears in both G2B-AE's genetic (W3) and residual (W1) components, in only one imgAE component, and only in Symm. alignment's genetic subspace; permutation collapses the structure into the residual. Colors: normalized voxel weights (0–1).

4 Experiments

4.1 Semi-synthetic Data

4.1.1 Simulation Setup

We inject controlled atrophy patterns and genetic variation into real scans of cognitively normal (CN) UKBB subjects (Miller et al., 2016) ($N = 1000$). Each subject has two simulated genetic latent factors ($s_{\text{gen1}}, s_{\text{gen2}}$) and one environmental factor (s_{env}). In the *basic* scenario, s_{gen1} drives temporal and subcortical atrophy, s_{gen2} frontal and occipital, and s_{env} temporal and parietal, so genetic and non-genetic effects overlap (Fig. 2a). SNPs are simulated with minor allele frequencies drawn from a skewed distribution: 5% are associated with s_{gen1}, 5% with s_{gen2}, 10% encode an independent, imaging-unrelated latent factor (genetic confounders), and the rest are null. Only part of the imaging variation is thus genetically driven, while residual factors contribute independent structure.

We additionally consider two scenarios of increased difficulty: *large-overlap*, where the spatial supports of s_{gen1} and s_{gen2} are extended to share subcortical territory (on top of the genetic–environmental overlap), and *weak-effect*, where genetically driven atrophy is scaled down toward the noise floor. Full simulation details are in Appendix A.3. Imaging is processed into voxel-wise atrophy maps (gray-matter maps registered to MNI space, smoothed at 8 mm FWHM, and z-scored against a CN reference cohort; Appendix A.6).

4.1.2 Evaluation Metrics

Since most baselines do not predict disentangled genetic and residual components, we match components to ground-truth genetic patterns by spatial similarity and compute all metrics on the matched components. *Genetic spatial similarity* is the cosine similarity between learned components and ground-truth genetic patterns, averaged over the 2 genetic patterns. *Environmental spatial similarity* is that of the residual component to A_{env}; for methods without one, we use the best-matching component not assigned to a genetic pattern. *Coefficient correlation* is the Pearson correlation between learned latents and simulated severities, averaged over the best match to each genetic factor. *Genetic association* counts the genetic latents (out of 2) significantly associated, by linear regression, with the five causal SNPs of their pattern. Finally, we visually inspect voxel-level maps for pattern quality and genetic–residual leakage. Aggregate cosine similarity captures these only partly, and interpretability is a central motivation of our framework.

4.1.3 Experimental Design

Training and hyperparameters. Following established practice (Yang et al., 2022; 2021; 2024b; Skampardoni et al., 2026), latent dimensionality is fixed to the true generative structure in simulation ($K = 3$, $K_g = 2$, $K_r = 1$) and selected without ground truth on real data. Loss weights and other hyperparameters are also selected without ground truth: for each of 36 configurations, we

train five models from independent initializations and pick the configuration with the highest cross-seed agreement (Hungarian-matched correlation between spatial components; Yang et al., 2022). Ground truth is used only afterward, to confirm that agreement tracks spatial recovery (Fig. 5 in Appendix A.2). The configuration chosen on the basic scenario is applied unchanged to the other scenarios and to real data.

Baselines. We compare against RRR, voxel-wise Elastic Net, sCCA, and jNMF on voxel-level data; an imaging-only AE; and Gene-SGAN (Yang et al., 2024a). To our knowledge, these are the only other deep multimodal methods not trained for a downstream prediction task. Gene-SGAN is weakly supervised, so it uses the reference subjects as its CN class and the analysis cohort as the diseased class, whereas all other methods, including ours, use no labels and use the reference only for z-scoring. Additionally, as Gene-SGAN requires imaging features of dimensionality comparable to genetics, we input 128 opNMF components rather than a fixed atlas that may misalign with the simulated patterns. It also did not converge on smoothed images, so it was trained on unsmoothed data. We exclude supervised fusion models, whose labels bias representations toward discriminative signal. To represent latent fusion, the prevailing multimodal strategy of mapping both modalities into one shared space, we build contrastive symmetric alignment variants (Contr $K{=}2$/$K{=}3$) that replace the predictor with InfoNCE against a genetic encoder of matched dimension, matching either the imaging side's genetic capacity ($K{=}2$) or the number of genetic sources ($K{=}3$). Baseline hyperparameters are selected by the same stability criterion as ours.

Falsification and ablations. To test whether the model recovers genuine cross-modal structure rather than imposing spurious alignment, we permute genetic embeddings across subjects, destroying subject-level correspondence while preserving marginal distributions, and assess the resulting spatial maps and cross-seed stability (Appendix A.5.1). Ablations of objective design, individual loss terms, and residual dimensionality, as well as an upper-bound experiment, are in Appendix A.5.

4.2 Real Data

4.2.1 Data and Model Selection

We apply the framework to ADNI (Petersen et al., 2010), using MRI and whole-genome sequencing preprocessed as shown in Appendix A.6. Since encoding the full genome is impractical, we follow standard imaging-genetics practice (Yang et al., 2024a; Ko et al., 2022; Wang et al., 2023) and use a curated panel of 983 dementia-associated SNPs. All results are therefore conditioned on this panel, and broader panels are a natural extension. Unlike a diagnosis label, the panel provides disease-relevant guidance at the genetic level while leaving the imaging-to-phenotype mapping unsupervised. We train on the $N = 1045$ MCI/AD subjects with genetic data. Voxel-wise atrophy maps are z-scored against a reference cohort (Appendix A.6.2), and external cohorts use the same pipeline and reference. Latent dimensionality is selected by cross-seed stability and variance criteria, favoring consistent, non-collapsed representations, with $K_r = 1$ (Appendix A.7).

4.2.2 Evaluation Strategy

Since this is a self-supervised mechanistic discovery task with no available ground truth, we combine genetic, anatomical, clinical, and out-of-distribution validation regimes. *Genetic specificity*: SNP-latent association via linear models controlling for age, sex, and the first five genetic principal components (FDR-corrected), joint predictability of each latent from the full SNP panel (983), and an in-sample genome-wide characterization (Appendix A.8). *Anatomy*: qualitative inspection of spatial maps for plausibility and interpretability. *Clinical relevance*: Spearman correlations between standardized latents and clinical, cognitive, biomarker, and metabolic variables (two-sided, FDR-corrected; Appendix A.9), and correlations among components. *Generalization*: the key test of whether genetic guidance yields a robust representation, rather than overfitting to ADNI-specific genotype–imaging correlations, is transfer without genetic data. We apply the trained imaging encoder alone to eight independent cohorts spanning diverse ancestries and recruitment settings (NACC, WRAP, HABS-HD, KBASE, ROS/MAP/MARS, SCAN, WHICAP, A4; Appendix A.11), none of which genotypes were available to us.

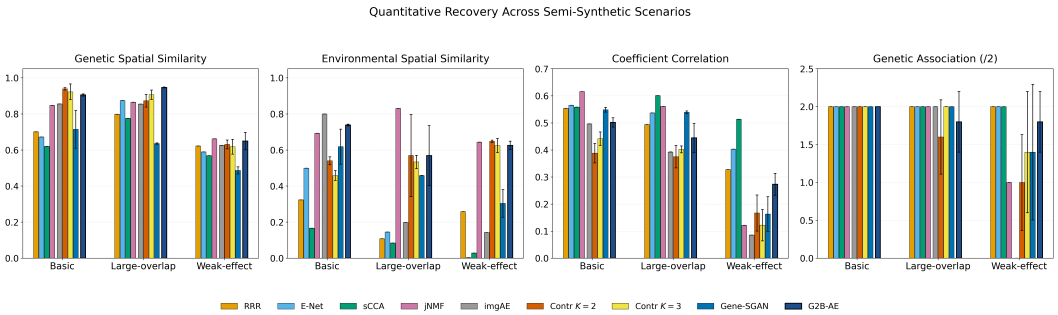


Figure 3: Quantitative recovery on semi-synthetic data. Bars show mean ± std over 5 seeds for stochastic methods (imgAE, Contr., Gene-SGAN, G2B-AE); RRR, E-Net, sCCA, and jNMF are deterministic single fits. Higher is better. Exact values in Table 2 in the Appendix.

4.2.3 Extension to Clinical Risk Factors

To test whether the framework generalizes beyond genetics, we replace the SNPs with ADNI clinical risk factors: hypertension status, systolic and diastolic blood pressure, diabetes status, glucose, BMI, cholesterol, triglycerides, smoking status, and alcohol use. The predictor takes the standardized risk factors directly, without a pretrained encoder. The predictor maps them into the imaging latent space, yielding risk-associated and residual latents that are analyzed with the same evaluation pipeline.

5 Results

5.1 Semi-Synthetic Results

The semi-synthetic experiments evaluate whether the model can accurately recover the simulated patterns of atrophy attributed to genetic and environmental factors. All associated metrics are presented in Fig. 3 (Table 2 in Appendix).

Role of genetic guidance. Comparing G2B-AE with the same architecture trained without genetics (imgAE) isolates the effect of genetic guidance. Without genetics, the decomposition acts as a soft parcellation: in the basic scenario, imgAE assigns the temporal region shared by genetic and environmental patterns to a single component, whereas G2B-AE represents it in both. ImgAE's dictionary is also nearly unchanged under large overlap (Appendix Fig. 7), where its residual loses the environmental pattern (0.198 vs. 0.569) while G2B-AE reaches the highest genetic similarity of all methods (0.947). Under weak effects, none of imgAE's components associate with the causal SNPs ($0/2$ vs. $1.8/2$). Permuting genetic embeddings, as expected by construction, removes the genetic structure: the residual component absorbs nearly the entire brain, the genetic components revert to sparse, noisy versions of imgAE's covariance patterns (Fig. 2e), and cross-seed agreement drops from $0.995/0.995$ to $0.42/0.71$ (Table 3 in Appendix).

Modeling genetic confounding. While symmetric alignment modeling uses genetics, it is unable to correctly account for genetic confounding. Contrastive models' dimension-matched genetic encoder must also represent the confounder, which has no imaging correlate: at $K{=}2$ it randomly entangles genetic components, unstably across seeds (Fig 2d); at $K{=}3$ the confounder is forced onto a spurious imaging component (Appendix A.4.2). In the basic scenario, alignment scores are slightly higher on genetic patterns (K2/K3 vs G2B-AE: $0.922/0.939$ vs 0.905) partly by absorbing non-genetic signal. Residual components in K2/K3 models miss the temporal region shared with the environmental pattern (Fig. 2d,b; environmental similarity $0.540/0.458$ vs 0.738), a leakage the genetic metric cannot penalize because the genetic pattern covers the same region. Under weak-effect, K2/K3 show poor severity correlation ($0.167/0.122$ vs. 0.273) and genetic association ($1.0/1.4$ vs. 1.8). Joint NMF, with loadings shared across modalities, decomposes well in the basic and large-overlap scenarios (highest environmental recovery under large-overlap, 0.830), but under weak-effect its loadings must also explain confounder and null SNPs, and severity correlation (0.122) and genetic

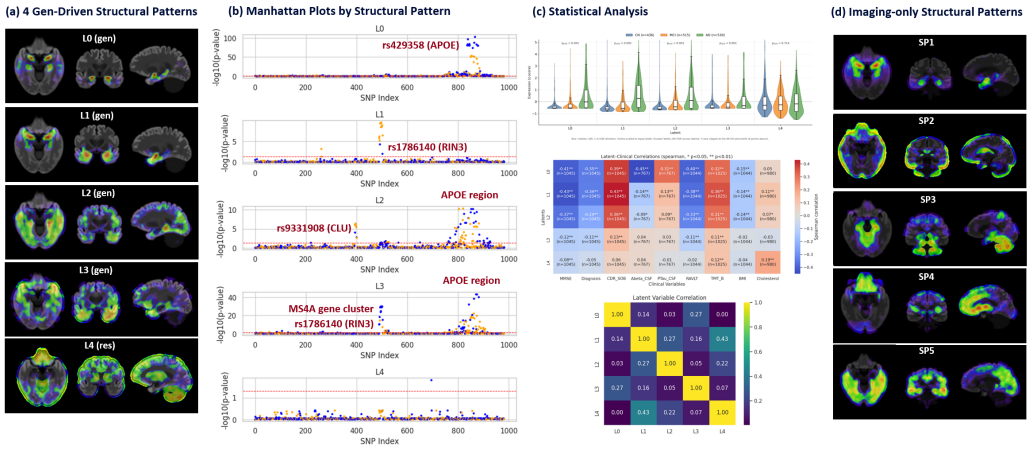


Figure 4: Real-data results on ADNI: (a) voxel-level spatial components, (b) SNP–latent associations (blue/orange: negative/positive effect sizes), and (c) correlations with clinical, cognitive, biomarker, and metabolic variables, revealing distinct genetically predictable and residual axes of neurodegeneration.

association ($1/2$) drop sharply. Our predictor avoids both failures by mapping only the genetic variation expressed in imaging, making it more robust in challenging scenarios.

Prediction-oriented methods trade away interpretability. Linear methods (RRR, E-Net, sCCA) fit subject-level severities better than G2B-AE in every scenario, and Gene-SGAN in all but weak-effect (e.g., 0.549–0.566 vs. 0.502 in basic), as expected, since they optimize this quantity or subject-space separability directly, without disentanglement (Fig. 2). This is a genuine trade-off: nothing in our objective targets subject-level correlation, and an oracle ridge regression from imaging alone reaches $r \approx 0.69$–0.71 (Table 6 in Appendix), which none of the unsupervised baselines captures, including ours. In contrast, their spatial maps are weaker (genetic-pattern similarity 0.620–0.715 in basic vs. 0.905), and the linear methods' component signs are not identifiable, yielding mixed-sign rather than additive maps that require unsigned similarity. Gene-SGAN, despite using group labels, has lower spatial similarity than G2B-AE in all scenarios.

Overall, genetic guidance reorganizes the NMF-style decomposition toward genetic mechanisms precisely where covariance alone fails (large overlapping or weak effects), directional coupling achieves this without the leakage and confounder absorption of symmetric alternatives, and the decomposition remains non-negative and interpretable.

5.2 Real Data Results

5.2.1 Disentangled Genetic–Neuroanatomical Axes

Stability and variance criteria identified two configurations, $(K, K_g) = (8, 4)$ and $(10, 2)$ (Appendix A.7). We focus on the higher-resolution $K_g = 4$ setting, which better separates known AD-genetic subtypes; $K_g = 2$ results and a comparison to imaging-only approaches are in Appendix A.10.

The model decomposes atrophy into four genetically guided axes and one residual (Fig. 4). L0 and L1 share a focal medial temporal pattern but differ genetically and pathologically. L0 is strongly associated with APOE-region variants and tracks amyloid and tau (CSF Aβ $\rho = -0.45$, pTau $\rho = 0.31$). L1, associated with non-APOE loci (RIN3), shows comparable clinical severity (MMSE $\rho = -0.43$ vs. -0.41) but weak amyloid association ($\rho = -0.14$). The two thus separate amyloid-related from largely amyloid-independent atrophy in the same region, with nearly uncorrelated subject loadings ($r = 0.14$). L2, also APOE-associated, extends to medial temporal–parietal and posterior cingulate/precuneus regions, and L3 combines APOE-region and MS4A-cluster variants in a broader temporal–frontal pattern, both with weaker clinical associations. The residual L4 is spatially diffuse and correlates modestly with vascular/metabolic variables such as cholesterol.

The genetic/residual split holds quantitatively. Of the 983 SNPs, 140, 17, 79, and 118 are FDR-significantly associated with L0–L3, respectively, versus 1 with L4 (Fig. 4b). Multivariate regression from the full genetic panel (covariate-adjusted ridge regression, 1000 permutations) is significant for L0–L3 (held-out $R^2 = 0.424, 0.052, 0.130, 0.395$; $p < 0.001$) and at the permutation null for L4 ($R^2 = -0.007$, $p = 0.754$). L0–L3 increase monotonically from CN to MCI to AD ($p_{\text{FDR}} < 0.001$), while L4 shows no diagnostic separation ($p_{\text{FDR}} = 0.714$; Fig. 4c). Inter-component correlations are low except L1–L4 ($r = 0.43$), which replicates in NACC's broader diagnostic population ($r = 0.56$) and may reflect pleiotropic effects of RIN3 and SLC24A4 on cardiovascular and metabolic pathways.

Without genetics, the same architecture yields a decomposition closer to a soft parcellation (Fig. 4d): only SP1 (focal medial temporal; CN vs. AD $d = 1.49$) and SP5 (lateral temporo-parietal; $d = 1.08$) carry strong disease signal, while the other three capture regional or boundary structure (cortical rim, brainstem/cerebellum, orbitofrontal/periventricular) with weak or reversed diagnostic separation. Non-negative regression between the dictionaries shows how genetic guidance reorganizes it. SP1 and SP5 are preserved as the non-APOE L1 ($R^2 = 0.96$) and L2 ($R^2 = 0.89$), whereas the APOE axis L0 has no single imaging-only counterpart (best $R^2 = 0.37$), and the APOE–MS4A axis L3 is only half explained by the entire imaging-only dictionary ($R^2 = 0.49$). Conversely, G2B-AE does not reconstruct the three anatomical components ($R^2 \leq 0.17$): capacity shifts from anatomical variance to disease axes, with non-genetic disease variation summarized in one residual. The fidelity cost is modest (centered explained variance 0.21 vs. 0.26 on the training subjects, rank-5 ceiling 0.27) and vanishes on held-out CN subjects (0.104 vs. 0.109, $p = 0.15$).

Because only imaging is needed at inference, we tested whether the learned axes transfer to eight independent cohorts that provide no genotypes to the model (Appendix A.11). Their clinical signatures replicated broadly: L0, L1 and L2 separated diagnostic groups in nearly all cohorts, and L1's association with cognitive decline was significant in all eight, including the preclinical, cognitively unimpaired A4. In NACC autopsy data, L0 and L2 replicated amyloid/tau neuropathology (Thal, CERAD, Braak, ADNC), and L3 a TDP-43 signature, indicating biologically distinct axes rather than a single severity dimension. L4 replicated the least AD signal but generalized on its own dimension: it was associated with cognitive decline in 7/8 cohorts, at times more strongly than L1 (KBASE, WHICAP), and in NACC it was opposite-signed to AD pathology and positively associated with cerebrovascular markers, indicating a coherent non-AD-genetic source of atrophy rather than noise. SNP–latent associations are replicated only for the APOE region (NACC, WRAP), since genome-wide replication needs genotype data for the other cohorts.

5.2.2 Non-Genetic Risk Factors

Replacing genetic inputs with clinical risk factors yields imaging patterns distinct from the genetic decomposition. These are associated with vascular and metabolic variables including cholesterol, BMI, smoking, and blood pressure, capturing multiple partially overlapping non-genetic pathways to atrophy, while the residual shows minimal association with the modeled risk factors (Appendix A.12).

6 Discussion

Conclusion. We introduced G2B-AE, which adds directional genetic guidance to an interpretable NMF-style decomposition, reorganizing brain structure around genetic mechanisms while separating residual variation. On semi-synthetic and Alzheimer's disease data, it recovered mechanisms that imaging alone merges or misses. Because only imaging is needed at inference, these patterns can be measured without genotyping, which could support stratification for trials and treatment, e.g., separating APOE-related atrophy that may respond to amyloid-targeting therapies from vascular/metabolic atrophy better addressed by risk-factor management.

Limitations. With only AD-associated loci modeled, $\boldsymbol{h}_{res}$ is residual to this SNP selection rather than necessarily non-genetic: vascular and metabolic factors are partly heritable, and other-disease SNPs (e.g., diabetes) could reveal further axes. Genetics is the sole upstream modality, substituting clinical risk factors is a first step toward a unified multi-source model. Finally, a pretrained genetic foundation model could replace the SNP VAE with richer, disease-agnostic embeddings.

AI USE STATEMENT

In this work, we did not use generative AI tools for any tasks with required disclosure. Specifically, we did not use them to develop theoretical models, formulate hypotheses, design experiments, implement methods, analyze data, or interpret results. We did not use them to generate synthetic datasets or to formulate or prove mathematical claims. Translation and dataset cleaning are not applicable to this work. We used generative AI tools to polish the initial drafts of all sections of the paper, based on the authors' research, results, and notes. The authors then iteratively revised these drafts to improve readability and ensure quality. All AI-assisted text was reviewed and edited by the authors. We verified every technical claim, reported result, and citation against our experiments and the original sources. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses human genetic and neuroimaging data obtained from UK Biobank , ADNI, NACC, WRAP, HABS-HD, KBASE, ROS/MAP/MARS, SCAN, WHICAP, A4 . All data were accessed under the respective data use agreements, were collected with informed consent from participants by the original studies, and were de-identified prior to our analysis. No new data were collected from human subjects for this study. We do not release any individual-level data. Code and trained models will be released and only aggregate results are reported in accordance with the data providers' terms.

As with all machine learning approaches applied to biomedical data, careful consideration of privacy, fairness, and clinical interpretation is essential. Genetic and neuroimaging datasets may contain demographic and population biases, for example over-representation of individuals of European ancestry, that can affect model behavior and generalization across cohorts. In addition, associations identified by the model should not be interpreted as causal without further biological validation. The framework is intended as a tool for scientific discovery and representation learning rather than clinical decision-making. Any future deployment in healthcare settings would require rigorous validation, transparency, and equitable evaluation across diverse populations. The authors declare no competing interests.

REPRODUCIBILITY STATEMENT

We have made several efforts to ensure the reproducibility of our results. The proposed model architecture and training objective are described in Section 3, with full implementation details, including hyperparameters, optimization settings, and hardware used, provided in Appendix A.2. The datasets used in this work are available to researchers through their respective access procedures. Our data preprocessing pipeline, covering quality control, feature extraction, and train/validation/test splits, is detailed in Section 4.1.1 and Appendix A.6. The evaluation protocol, baselines, and metrics are described in Section 4, and additional results, including ablations, are reported in Appendix A.5, A.8, A.10, A.11. We plan to release our source code upon publication.

REFERENCES

Nematollah K. Batmanghelich, Adrian V. Dalca, Mert R. Sabuncu, and Polina Golland. Joint modeling of imaging and genetics. *Information processing in medical imaging : proceedings of the ... conference*, 23:766, 2013. ISSN 03029743. doi: 10.1007/978-3-642-38868-2_64. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3979537/.

Annalisa Buniello, Jacqueline A.L. Macarthur, Maria Cerezo, Laura W. Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, Joannella Morales, Edward Mountjoy, Elliot Sollis, Daniel Suveges, Olga Vrousgou, Patricia L. Whetzel, Ridwan Amode, Jose A. Guillen, Harpreet S. Riat, Stephen J. Trevanion, Peggy Hall, Heather Junkins, Paul Flicek, Tony Burdett, Lucia A. Hindorff, Fiona Cunningham, and Helen Parkinson. The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research*, 47: D1005–D1012, 1 2019. ISSN 13624962. doi: 10.1093/nar/gky1120. URL https://pubmed.ncbi.nlm.nih.gov/30445434/.

Delong Chen, Mustafa Shukor, Theo Moutakanni, Willy Chung, Jade Yu, Tejaswi Kasarla, Yejin Bang, Allen Bolourchi, Yann LeCun, and Pascale Fung. Vl-jepa: Joint embedding predictive architecture for vision-language. 12 2025. URL https://arxiv.org/pdf/2512.10942.

Christos Davatzikos, Ahmet Genc, Dongrong Xu, and Susan M. Resnick. Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage*, 14:1361–1369, 2001. ISSN 10538119. doi: 10.1006/nimg.2001.0937. URL https://pubmed.ncbi.nlm.nih.gov/11707092/.

Jimit Doshi, Guray Erus, Yangming Ou, Susan M. Resnick, Ruben C. Gur, Raquel E. Gur, Theodore D. Satterthwaite, Susan Furth, and Christos Davatzikos. Muse: Multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage*, 127:186–195, 2 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.11.073. URL https://pubmed.ncbi.nlm.nih.gov/26679328/.

Lei Du, Heng Huang, Jingwen Yan, Sungeun Kim, Shannon L. Risacher, Mark Inlow, Jason H. Moore, Andrew J. Saykin, and Li Shen. Structured sparse canonical correlation analysis for brain imaging genetics: an improved graphnet method. *Bioinformatics (Oxford, England)*, 32:1544–1551, 5 2016. ISSN 1367-4811. doi: 10.1093/BIOINFORMATICS/BTW033. URL https://pubmed.ncbi.nlm.nih.gov/26801960/.

Tian Ge, Thomas E. Nichols, Debashis Ghosh, Elizabeth C. Mormino, Jordan W. Smoller, and Mert R. Sabuncu. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *NeuroImage*, 109:505, 4 2015. ISSN 10959572. doi: 10.1016/J.NEUROIMAGE.2015.01.029. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4339421/.

Paul Hager, Martin J. Menten, and Daniel Rueckert. Best of both worlds: Multimodal contrastive learning with tabular and imaging data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023-June:23924–23935, 3 2023. ISSN 10636919. doi: 10.1109/CVPR52729.2023.02291. URL https://arxiv.org/pdf/2303.14080.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014. URL https://arxiv.org/pdf/1412.6980.

Wonjun Ko, Wonsik Jung, Eunjin Jeon, and Heung Il Suk. A deep generative-discriminative learning for multimodal representation in imaging genetics. *IEEE transactions on medical imaging*, 41:2348–2359, 9 2022. ISSN 1558254X. doi: 10.1109/TMI.2022.3162870. URL https://pubmed.ncbi.nlm.nih.gov/35344489/.

Yann Lecun. A path towards autonomous machine intelligence. 2022.

Seungeun Lee, Jaeyoung Lee, Moonhyun Lee, Jintak Choi, Kyungtae Kang, and Younghoon Kim. Integrating genetic information for early alzheimer's diagnosis through mri interpretation. *BHI 2023 - IEEE-EMBS International Conference on Biomedical and Health Informatics, Proceedings*, 2023. doi: 10.1109/BHI58575.2023.10313442. URL https://ieeexplore.ieee.org/document/10313442.

Rhoshel K. Lenroot and Jay N. Giedd. The changing impact of genes and environment on brain development during childhood and adolescence: Initial findings from a neuroimaging study of pediatric twins. *Development and psychopathology*, 20:1161, 2008. ISSN 09545794. doi: 10.1017/S0954579408000552. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2892674/.

Gang Li, Depeng Han, Chao Wang, Wenxing Hu, Vince D. Calhoun, and Yu Ping Wang. Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia. *Computer Methods and Programs in Biomedicine*, 183, 1 2020. ISSN 18727565. doi: 10.1016/j.cmpb.2019.105073. URL https://pubmed.ncbi.nlm.nih.gov/31525548/.

Yuhan Li, Donghao Niu, Keying Qi, Dong Liang, and Xiaojing Long. An imaging and genetic-based deep learning network for alzheimer's disease diagnosis. *Frontiers in Aging Neuroscience*, 17:1532470, 2025. ISSN 16634365. doi: 10.3389/fnagi.2025.1532470. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11968703/.

Karla L. Miller, Fidel Alfaro-Almagro, Neal K. Bangerter, David L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, Saad Jbabdi, Stamatios N. Sotiropoulos, Jesper L.R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Thomas W. Okell, Peter Weale, Iulius Dragonu, Steve Garratt, Sarah Hudson, Rory Collins, Mark Jenkinson, Paul M. Matthews, and Stephen M. Smith. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, 19:1523–1536, 10 2016. ISSN 15461726. doi: 10.1038/nn.4393. URL https://www.nature.com/articles/nn.4393.

Melissa E. Murray, Neill R. Graff-Radford, Owen A. Ross, Ronald C. Petersen, Ranjan Duara, and Dennis W. Dickson. Neuropathologically defined subtypes of alzheimer's disease with distinct clinical characteristics: A retrospective study. *Lancet neurology*, 10:785, 9 2011. ISSN 14744422. doi: 10.1016/S1474-4422(11)70156-9. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3175379/.

George Obaido, Ibomoiye Domor Mienye, Kehinde Aruleba, Chidozie Williams Chukwu, Ebenezer Esenogho, and Cameron Modisane. A systematic review of contrastive learning in medical ai: Foundations, biomedical modalities, and future directions. *Bioengineering*, 13:176, 2 2026. ISSN 23065354. doi: 10.3390/BIOENGINEERING13020176. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC12938576/.

Yangming Ou, Aristeidis Sotiras, Nikos Paragios, and Christos Davatzikos. Dramms: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis*, 15: 622–639, 8 2011. ISSN 13618415. doi: 10.1016/j.media.2010.07.002. URL https://www.sciencedirect.com/science/article/abs/pii/S1361841510000940.

R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74:201–209, 2010. ISSN 1526632X. doi: 10.1212/WNL.0b013e3181cb3e25. URL https://pubmed.ncbi.nlm.nih.gov/20042704/.

Yan Rui Qiang, Shao Wu Zhang, Jia Ni Li, Yan Li, and Qin Yi Zhou. Diagnosis of alzheimer's disease by joining dual attention cnn and mlp based on structural mris, clinical and genetic data. *Artificial Intelligence in Medicine*, 145:102678, 11 2023. ISSN 18732860. doi: 10.1016/j.artmed.2023.102678. URL https://www.sciencedirect.com/science/article/abs/pii/S0933365723001926.

Ioanna Skampardoni, Guray Erus, Ilya M. Nasrallah, Zhijian Yang, Michael R. Duggan, Keenan A. Walker, Alexander Getka, Kyunglok Baik, Randa Melhem, Sindhuja T. Govindarajan, Susan M. Resnick, Haochang Shou, Konstantina Nikita, and Christos Davatzikos. Coupled cross-sectional and longitudinal non-negative matrix factorization reveals dominant brain aging trajectories in 48,949 individuals. *Nature Communications 2026 17:1*, 17:5709–, 4 2026. ISSN 2041-1723. doi: 10.1038/s41467-026-72091-7. URL https://www.nature.com/articles/s41467-026-72091-7.

Aristeidis Sotiras, Susan M. Resnick, and Christos Davatzikos. Finding imaging patterns of structural covariance via non-negative matrix factorization. *NeuroImage*, 108:1–16, 3 2015. ISSN 10959572. doi: 10.1016/j.neuroimage.2014.11.045. URL https://pubmed.ncbi.nlm.nih.gov/25497684/.

Aristeidis Sotiras, Jon B. Toledo, Raquel E. Gur, Ruben C. Gur, Theodore D. Satterthwaite, and Christos Davatzikos. Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and evolutionary expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 114:3527–3532, 3 2017. ISSN 10916490. doi: 10.1073/PNAS.1620928114;SUBPAGE:STRING:FULL. URL /doi/pdf/10.1073/pnas.1620928114?download=true.

Jason L. Stein, Xue Hua, Suh Lee, April J. Ho, Alex D. Leow, Arthur W. Toga, Andrew J. Saykin, Li Shen, Tatiana Foroud, Nathan Pankratz, Matthew J. Huentelman, David W. Craig, Jill D. Gerber, April N. Allen, Jason J. Corneveaux, Bryan M. DeChairo, Steven G. Potkin, Michael W. Weiner, and Paul M. Thompson. Voxelwise genome-wide association study (vgwas). *NeuroImage*, 53:1160–1174, 11 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.02.032. URL https://pubmed.ncbi.nlm.nih.gov/20171287/.

Kun Tu, Wenhui Zhou, and Shubing Kong. Integrating multi-omics data for alzheimer's disease to explore its biomarkers via the hypergraph-regularized joint deep semi-non-negative matrix factorization algorithm. *Journal of Molecular Neuroscience 2024 74:2*, 74:43–, 4 2024. ISSN 1559-1166. doi: 10.1007/S12031-024-02211-9. URL https://link.springer.com/article/10.1007/s12031-024-02211-9.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. 1 2019. URL http://arxiv.org/abs/1807.03748.

Deepthi P. Varikuti, Sarah Genon, Aristeidis Sotiras, Holger Schwender, Felix Hoffstaedter, Kaustubh R. Patil, Christiane Jockwitz, Svenja Caspers, Susanne Moebus, Katrin Amunts, Christos Davatzikos, and Simon B. Eickhoff. Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*, 173:394, 6 2018. ISSN 10959572. doi: 10.1016/J.NEUROIMAGE.2018.03.007. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5911196/.

Maria Vounou, Thomas E. Nichols, and Giovanni Montana. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *NeuroImage*, 53:1147, 11 2010. ISSN 10538119. doi: 10.1016/J.NEUROIMAGE.2010.07.002. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5253177/.

Tao Wang, Xiumei Chen, Jiawei Zhang, Qianjin Feng, and Meiyan Huang. Deep multimodality-disentangled association analysis network for imaging genetics in neurodegenerative diseases. *Medical Image Analysis*, 88, 8 2023. ISSN 13618423. doi: 10.1016/j.media.2023.102842. URL https://pubmed.ncbi.nlm.nih.gov/37247468/.

Yuanfei Wang and Xitao Wang. Exploration of imaging genetic biomarkers of alzheimer's disease based on a machine learning method. *Journal of integrative neuroscience*, 23, 2024. ISSN 0219-6352. doi: 10.31083/J.JIN2304081. URL https://pubmed.ncbi.nlm.nih.gov/38682217/.

Kyoko Watanabe, Erdogan Taskesen, Arjen Van Bochoven, and Danielle Posthuma. Functional mapping and annotation of genetic associations with fuma. *Nature Communications 2017 8:1*, 8: 1826–, 11 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01261-5. URL https://www.nature.com/articles/s41467-017-01261-5.

Junhao Wen, Cynthia H.Y. Fu, Duygu Tosun, Yogasudha Veturi, Zhijian Yang, Ahmed Abdulkadir, Elizabeth Mamourian, Dhivya Srinivasan, Ioanna Skampardoni, Ashish Singh, Hema Nawani, Jingxuan Bao, Guray Erus, Haochang Shou, Mohamad Habes, Jimit Doshi, Erdem Varol, R. Scott Mackin, Aristeidis Sotiras, Yong Fan, Andrew J. Saykin, Yvette I. Sheline, Li Shen, Marylyn D. Ritchie, David A. Wolk, Marilyn Albert, Susan M. Resnick, and Christos Davatzikos. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. *JAMA psychiatry*, 79:464–474, 5 2022. ISSN 2168622X. doi: 10.1001/jamapsychiatry.2022.0020. URL https://pubmed.ncbi.nlm.nih.gov/35262657/.

Junhao Wen, Ilya M. Nasrallah, Ahmed Abdulkadir, Theodore D. Satterthwaite, Zhijian Yang, Guray Erus, Robert Fitzgerald Timothy, Ashish Singh, Aristeidis Sotiras, Aleix Boquet-Pujada, Elizabeth Mamourian, Jimit Doshi, Yuhan Cui, Dhivya Srinivasan, Ioanna Skampardoni, Jiong Chen, Gyujoon Hwang, Mark Bergman, Jingxuan Bao, Yogasudha Veturi, Zhen Zhou, Shu Yang, Paola Dazzan, Rene S. Kahn, Hugo G. Schnack, Marcus V. Zanetti, Eva Meisenzahl, Geraldo F. Busatto, Benedicto Crespo-Facorr, Christos Pantelis, Stephen J. Wood, Chuanjun Zhuo, Russell T. Shinohara, Ruben C. Gur, Raquel E. Gur, Nikolaos Koutsouleris, Daniel H. Wolf, Andrew J. Saykin, Marylyn D. Ritchie, Li Shen, Paul M. Thompson, Olivier Colliot, Katharina Wittfeld, Hans J. Grabe, Duygu Tosun, Murat Bilgel, Yang An, Daniel S. Marcus, Pamela LaMontagne, Susan R. Heckbert, Thomas R. Austin, Lenore J. Launer, Mark Espeland, Colin L. Masters, Paul Maruff, Jurgen Fripp, Sterling C. Johnson, John C. Morris, Marilyn S. Albert, R. Nick Bryan, Susan M. Resnick, Yong Fan, Mohamad Habes, David Wolk, Haochang Shou, and Christos Davatzikos. Genomic loci influence patterns of structural covariance in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 120, 2023. ISSN 10916490. doi: 10.1073/pnas.2300842120. URL https://pubmed.ncbi.nlm.nih.gov/38127979/.

Junhao Wen, Zhijian Yang, Ilya M. Nasrallah, Yuhan Cui, Guray Erus, Dhivya Srinivasan, Ahmed Abdulkadir, Elizabeth Mamourian, Gyujoon Hwang, Ashish Singh, Mark Bergman, Jingxuan

Bao, Erdem Varol, Zhen Zhou, Aleix Boquet-Pujadas, Jiong Chen, Arthur W. Toga, Andrew J. Saykin, Timothy J. Hohman, Paul M. Thompson, Sylvia Villeneuve, Randy Gollub, Aristeidis Sotiras, Katharina Wittfeld, Hans J. Grabe, Duygu Tosun, Murat Bilgel, Yang An, Daniel S. Marcus, Pamela LaMontagne, Tammie L. Benzinger, Susan R. Heckbert, Thomas R. Austin, Lenore J. Launer, Mark Espeland, Colin L. Masters, Paul Maruff, Jurgen Fripp, Sterling C. Johnson, John C. Morris, Marilyn S. Albert, R. Nick Bryan, Susan M. Resnick, Luigi Ferrucci, Yong Fan, Mohamad Habes, David Wolk, Li Shen, Haochang Shou, and Christos Davatzikos. Genetic and clinical correlates of two neuroanatomical ai dimensions in the alzheimer's disease continuum. *Translational Psychiatry 2024 14:1*, 14:420–, 10 2024. ISSN 21583188. doi: 10.1038/s41398-024-03121-5. URL https://www.nature.com/articles/s41398-024-03121-5.

Junhao Wen, Ioanna Skampardoni, Ye Ella Tian, Zhijian Yang, Yuhan Cui, Guray Erus, Gyujoon Hwang, Erdem Varol, Aleix Boquet-Pujadas, Ganesh B. Chand, Ilya M. Nasrallah, Theodore D. Satterthwaite, Haochang Shou, Li Shen, Arthur W. Toga, Andrew Zalesky, and Christos Davatzikos. Neuroimaging endophenotypes reveal underlying mechanisms and genetic factors contributing to progression and development of four brain disorders. *Nature Biomedical Engineering 2025 9:11*, 9:1920–1937, 6 2025. ISSN 2157846X. doi: 10.1038/s41551-025-01412-w. URL https://www.nature.com/articles/s41551-025-01412-w.

Sarah Wu, Archana Venkataraman, and Sayan Ghosal. Girus-net: A multimodal deep learning model identifying imaging and genetic biomarkers linked to alzheimer's disease severity. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2023, 2023. ISSN 1557170X. doi: 10.1109/EMBC40787.2023.10341000. URL https://pubmed.ncbi.nlm.nih.gov/38083359/.

Zhijian Yang, Ilya M. Nasrallah, Haochang Shou, Junhao Wen, Jimit Doshi, Mohamad Habes, Guray Erus, Ahmed Abdulkadir, Susan M. Resnick, Marilyn S. Albert, Paul Maruff, Jurgen Fripp, John C. Morris, David A. Wolk, Christos Davatzikos, Yong Fan, Vishnu Bashyam, Elizabeth Mamouiran, Randa Melhem, Raymond Pomponio, Dushyant Sahoo, Singh Ashish, Ioanna Skampardoni, Lasya Sreepada, Dhivya Srinivasan, Fanyang Yu, Sindhuja Govindarajan Tirumalai, Yuhan Cui, Zhen Zhou, Katharina Wittfeld, Hans J. Grabe, Duygun Tosun, Murat Bilgel, Yang An, Daniel S. Marcus, Pamela LaMontagne, Susan R. Heckbert, Thomas R. Austin, Lenore J. Launer, Aristeidis Sotiras, Mark Espeland, Colin L. Masters, Henry Völzk, Sterling C. Johnson, Luigi Ferrucci, R. Nick Bryan, Michael Weiner, Paul Aisen, Ronald Petersen, Clifford R. Jack, William Jagust, John Q. Trojanowki, Arthur W. Toga, Laurel Beckett, Robert C. Green, Andrew J. Saykin, Leslie M. Shaw, Enchi Liu, Tom Montine, Ronald G. Thomas, Michael Donohue, Sarah Walter, Devon Gessert, Tamie Sather, Gus Jiminez, Danielle Harvey, Matthew Bernstein, Nick Fox, Paul Thompson, Norbert Schuff, Charles DeCArli, Bret Borowski, Jeff Gunter, Matt Senjem, Prashanthi Vemuri, David Jones, Kejal Kantarci, Chad Ward, Robert A. Koeppe, Norm Foster, Eric M. Reiman, Kewei Chen, Chet Mathis, Susan Landau, Nigel J. Cairns, Erin Householder, Lisa Taylor Reinwald, Virginia Lee, Magdalena Korecka, Michal Figurski, Karen Crawford, Scott Neu, Tatiana M. Foroud, Steven Potkin, Li Shen, Faber Kelley, Sungeun Kim, Kwangsik Nho, Zaven Kachaturian, Richard Frank, Peter J. Snyder, Susan Molchan, Jeffrey Kaye, Joseph Quinn, Betty Lind, Raina Carter, Sara Dolen, Lon S. Schneider, Sonia Pawluczyk, Mauricio Beccera, Liberty Teodoro, Bryan M. Spann, James Brewer, Helen Vanderswag, Adam Fleisher, Judith L. Heidebrink, Joanne L. Lord, Sara S. Mason, Colleen S. Albers, David Knopman, Kris Johnson, Rachelle S. Doody, Javier Villanueva Meyer, Munir Chowdhury, Susan Rountree, Mimi Dang, Yaakov Stern, Lawrence S. Honig, Karen L. Bell, Beau Ances, Maria Carroll, Sue Leon, Erin Householder, Mark A. Mintun, Stacy Schneider, Angela OliverNG, Randall Griffith, David Clark, David Geldmacher, John Brockington, Erik Roberson, Hillel Grossman, Effie Mitsis, Leyla deToledo Morrell, Raj C. Shah, Ranjan Duara, Daniel Varon, Maria T. Greig, Peggy Roberts, Chiadi Onyike, Daniel D'Agostino, Stephanie Kielb, James E. Galvin, Dana M. Pogorelec, Brittany Cerbone, Christina A. Michel, Henry Rusinek, Mony J. de Leon, Lidia Glodzik, Susan De Santi, P. Murali Doraiswamy, Jeffrey R. Petrella, Terence Z. Wong, Steven E. Arnold, Jason H. Karlawish, Charles D. Smith, Greg Jicha, Peter Hardy, Partha Sinha, Elizabeth Oates, Gary Conrad, Oscar L. Lopez, Mary Ann Oakley, Donna M. Simpson, Anton P. Porsteinsson, Bonnie S. Goldstein, Kim Martin, Kelly M. Makino, M. Saleem Ismail, Connie Brand, Ruth A. Mulnard, Gaby Thai, Catherine Mc Adams Ortiz, Kyle Womack, Dana Mathews, Mary Quiceno, Ramon Diaz Arrastia, Richard King, Myron Weiner, Kristen Martin Cook, Michael DeVous,

Allan I. Levey, James J. Lah, Janet S. Cellar, Jeffrey M. Burns, Heather S. Anderson, Russell H. Swerdlow, Liana Apostolova, Kathleen Tingus, Ellen Woo, Daniel H.S. Silverman, Po H. Lu, George Bartzokis, Neill R.Graff Radford, Francine ParfittH, Tracy Kendall, Heather Johnson, Martin R. Farlow, Ann Marie Hake, Brandy R. Matthews, Scott Herring, Cynthia Hunt, Christopher H. van Dyck, Richard E. Carson, Martha G. MacAvoy, Howard Chertkow, Howard Bergman, Chris Hosein, Sandra Black, Bojana Stefanovic, Curtis Caldwell, Ging Yuek Robin Hsiung, Howard Feldman, Benita Mudge, Michele Assaly Past, Andrew Kertesz, John Rogers, Dick Trost, Charles Bernick, Donna Munic, Diana Kerwin, Marek Marsel Mesulam, Kristine Lipowski, Chuang Kuo Wu, Nancy Johnson, Carl Sadowsky, Walter Martinez, Teresa Villena, Raymond Scott Turner, Kathleen Johnson, Brigid Reynolds, Reisa A. Sperling, Keith A. Johnson, Gad Marshall, Meghan Frey, Jerome Yesavage, Joy L. Taylor, Barton Lane, Allyson Rosen, Jared Tinklenberg, Marwan N. Sabbagh, Christine M. Belden, Sandra A. Jacobson, Sherye A. Sirrel, Neil Kowall, Ronald Killiany, Andrew E. Budson, Alexander Norbash, Patricia Lynn Johnson, Thomas O. Obisesan, Saba Wolday, Joanne Allard, Alan Lerner, Paula Ogrocki, Leon Hudson, Evan Fletcher, Owen Carmichael, John Olichney, Charles DeCarli, Smita Kittur, Michael Borrie, T. Y. Lee, Rob Bartha, Sanjay Asthana, Cynthia M. Carlsson, Steven G. Potkin, Adrian Preda, Dana Nguyen, Pierre Tariot, Adam Fleisher, Stephanie Reeder, Vernice Bates, Horacio Capote, Michelle Rainka, Douglas W. Scharre, Maria Kataki, Anahita Adeli, Earl A. Zimmerman, Dzintra Celmins, Alice D. Brown, Godfrey D. Pearlson, Karen Blank, Karen Anderson, Robert B. Santulli, Tamar J. Kitzmiller, Eben S. Schwartz, Kaycee M. SinkS, Jeff D. Williamson, Pradeep Garg, Franklin Watkins, Brian R. Ott, Henry Querfurth, Geoffrey Tremont, Stephen Salloway, Paul Malloy, Stephen Correia, Howard J. Rosen, Bruce L. Miller, Jacobo Mintzer, Kenneth Spicer, David Bachman, Elizabether Finger, Stephen Pasternak, Irina Rachinsky, John Rogers, Andrew Kertesz, Dick Drost, Nunzio Pomara, Raymundo Hernando, Antero Sarrael, Susan K. Schultz, Laura L.Boles Ponto, Hyungsub Shim, Karen Elizabeth Smith, Norman Relkin, Gloria Chaing, Lisa Raudin, Amanda Smith, Kristin Fargher, and Balebail Ashok Raj. A deep learning framework identifies dimensional representations of alzheimer's disease from brain structure. *Nat. Commun.*, 12, 12 2021. ISSN 20411723. doi: 10.1038/s41467-021-26703-z.

Zhijian Yang, Junhao Wen, and Christos Davatzikos. Surreal-gan:semi-supervised representation learning via gan for uncovering heterogeneous disease-related imaging patterns. *ICLR 2022 - 10th International Conference on Learning Representations*, 5 2022. URL http://arxiv.org/abs/2205.04523.

Zhijian Yang, Junhao Wen, Ahmed Abdulkadir, Yuhan Cui, Guray Erus, Elizabeth Mamourian, Randa Melhem, Dhivya Srinivasan, Sindhuja T. Govindarajan, Jiong Chen, Mohamad Habes, Colin L. Masters, Paul Maruff, Jurgen Fripp, Luigi Ferrucci, Marilyn S. Albert, Sterling C. Johnson, John C. Morris, Pamela LaMontagne, Daniel S. Marcus, Tammie L.S. Benzinger, David A. Wolk, Li Shen, Jingxuan Bao, Susan M. Resnick, Haochang Shou, Ilya M. Nasrallah, and Christos Davatzikos. Gene-sgan: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. *Nature Communications 2024 15:1*, 15:354–, 1 2024a. ISSN 2041-1723. doi: 10.1038/s41467-023-44271-2. URL https://www.nature.com/articles/s41467-023-44271-2.

Zhijian Yang, Junhao Wen, Guray Erus, Sindhuja T. Govindarajan, Randa Melhem, Elizabeth Mamourian, Yuhan Cui, Dhivya Srinivasan, Ahmed Abdulkadir, Paraskevi Parmpi, Katharina Wittfeld, Hans J. Grabe, Robin Bülow, Stefan Frenzel, Duygu Tosun, Murat Bilgel, Yang An, Dahyun Yi, Daniel S. Marcus, Pamela LaMontagne, Tammie L.S. Benzinger, Susan R. Heckbert, Thomas R. Austin, Shari R. Waldstein, Michele K. Evans, Alan B. Zonderman, Lenore J. Launer, Aristeidis Sotiras, Mark A. Espeland, Colin L. Masters, Paul Maruff, Jurgen Fripp, Arthur W. Toga, Sid O'Bryant, Mallar M. Chakravarty, Sylvia Villeneuve, Sterling C. Johnson, John C. Morris, Marilyn S. Albert, Kristine Yaffe, Henry Völzke, Luigi Ferrucci, R. Nick Bryan, Russell T. Shinohara, Yong Fan, Mohamad Habes, Paris Alexandros Lalousis, Nikolaos Koutsouleris, David A. Wolk, Susan M. Resnick, Haochang Shou, Ilya M. Nasrallah, and Christos Davatzikos. Brain aging patterns in a large and diverse cohort of 49,482 individuals. *Nature medicine*, 30:3015–3026, 10 2024b. ISSN 1546170X. doi: 10.1038/s41591-024-03144-x. URL https://pubmed.ncbi.nlm.nih.gov/39147830/.

Yanfu Zhang, Liang Zhan, Paul M. Thompson, and Heng Huang. Biological knowledge guided deep neural network for brain genotype-phenotype association study. *Lecture Notes in Computer*

Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11846 LNCS:84–92, 2019. ISSN 16113349. doi: 10.1007/978-3-030-33226-6_10. URL https://link.springer.com/chapter/10.1007/978-3-030-33226-6_10.

Rong Zhou, Houliang Zhou, Brian Y. Chen, Li Shen, Yu Zhang, and Lifang He. Attentive deep canonical correlation analysis for diagnosing alzheimer's disease using multimodal imaging genetics. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14221 LNCS:681–691, 2023. ISSN 16113349. doi: 10.1007/978-3-031-43895-0_64.

Tao Zhou, Kim Han Thung, Mingxia Liu, and DInggang Shen. Brain-wide genome-wide association study for alzheimer's disease via joint projection learning and sparse regression model. *IEEE transactions on bio-medical engineering*, 66:165–175, 1 2019. ISSN 1558-2531. doi: 10.1109/TBME.2018.2824725. URL https://pubmed.ncbi.nlm.nih.gov/29993426/.

A APPENDIX

A.1 INFONCE LOSS

InfoNCE encourages agreement between positive pairs while contrasting them against negative samples. Given a query $\mathbf{q}$, a positive key $\mathbf{k}^+$, and negative keys $\{\mathbf{k}^-\}$:

$$\mathcal{L}_{\text{NCE}}(\mathbf{q}, \mathbf{k}^+) = -\log \frac{\exp\left(\text{sim}(\mathbf{q}, \mathbf{k}^+)/\tau\right)}{\exp\left(\text{sim}(\mathbf{q}, \mathbf{k}^+)/\tau\right) + \sum_{\mathbf{k}^-} \exp\left(\text{sim}(\mathbf{q}, \mathbf{k}^-)/\tau\right)}, \tag{6}$$

where $\text{sim}(\cdot, \cdot)$ is cosine similarity and τ is a temperature parameter. Matching pairs are treated as positives and all other batch samples as negatives, pulling embeddings from the same subject/modality pair together while pushing apart unrelated ones; this preserves shared cross-modal information while remaining discriminative. Here it promotes alignment between genetic and imaging-associated representations while preventing trivial solutions, since matched pairs must be more similar than mismatched ones across the batch. Smaller τ emphasizes hard negatives.

A.2 IMPLEMENTATION DETAILS

A.2.1 TUNING STRATEGY

We treat ($\lambda_{\text{link}}, \lambda_W, \lambda_{h\text{-}ortho}$, predictor sigmoid) as the primary interaction knobs, trading off cross-modal alignment, factor geometry, and residual–genetic separation. A factorial grid is enumerated: $\lambda_{\text{link}} \in \{0.3, 0.5, 0.8\}$, $\lambda_W \in \{0.03, 0.1, 0.3\}$, $\lambda_{h\text{-}ortho} \in \{1, 10\}$, with/without a predictor output sigmoid (`use_sigmoid`$\in \{0, 1\}$), giving $3 \times 3 \times 2 \times 2 = 36$ configurations. The number of latent components is fixed to the true generative dimensionality throughout (Section 4.1.1), following established semi-synthetic evaluation practice (Yang et al., 2022); on real data, where ground truth is unavailable, components are instead selected via the stability-based procedure of Appendix A.7.

For each configuration, we train five models from independent initializations and compute cross-seed agreement — pairwise Hungarian-matched correlation between spatial components, averaged over all $\binom{5}{2} = 10$ seed pairs — selecting the configuration that maximizes it. This mirrors the criterion of Yang et al. (2022) and is later applied, without ground truth, to real ADNI data (Appendix A.7).

Ground truth is used only afterward, to validate the procedure (Fig. 5): agreement correlates significantly with spatial recovery (Pearson $r = 0.60$, $p < 0.001$), and the agreement-selected configuration achieves the highest spatial similarity of all 36 candidates; agreement does not correlate with coefficient recovery ($r = 0.12$, $p = 0.48$), which varies narrowly across the grid regardless of hyperparameters.

A.2.2 LOSS WEIGHTS

The selected configuration: λ_{ae}=10, λ_W=0.03, λ_{link}=0.3, $\lambda_{\ell_2 r}$=$\lambda_{\ell_2 g}$=10^{-1}, $\lambda_{h\text{-}ortho}$=1, τ=0.1 (InfoNCE temperature), with a sigmoid-activated predictor output.

A.2.3 ARCHITECTURE

Genetic VAE. The encoder maps $\boldsymbol{x}_{snp} \in \mathbb{R}^G$ (allele counts; missing genotypes masked) through three linear layers ($G \to G/2 \to G/4 \to 2K$, LeakyReLU slope 0.2, Kaiming-normal init) to the mean/log-variance of $q(\boldsymbol{z}_{gen} \mid \boldsymbol{x}_{snp})$. The decoder maps $\text{sigmoid}(\boldsymbol{z}_{gen})$ through three linear layers ($K \to 2K \to 4K \to G$, LeakyReLU, dropout 0.2 before output) to G SNP logits, with the binomial likelihood of Eq. 1 (n=2, p=sigmoid(logit)).

Imaging autoencoder. The encoder is a bias-free linear map $\boldsymbol{W}_{enc} \in \mathbb{R}^{C \times V}$, followed by batch norm over C=K_g+K_r units and a sigmoid, giving $\boldsymbol{h} \in (0, 1)^C$ ($\boldsymbol{h}_{gen}$: first K_g channels; $\boldsymbol{h}_{res}$: remaining K_r). The decoder is the bias-free map $\boldsymbol{W} \in \mathbb{R}^{V \times C}$ of Section 3.2, NNDSVD-initialized and non-negative throughout training, consistent with the NMF-style decomposition; the encoder is learned separately, not tied to $\boldsymbol{W}^\top$. Reconstruction error (Eq. 3) is averaged over voxels/subjects; the orthonormality penalty is computed on a copy of $\boldsymbol{W}$ after clamping to $[0, \infty)$ and column-wise ℓ_2 normalization.

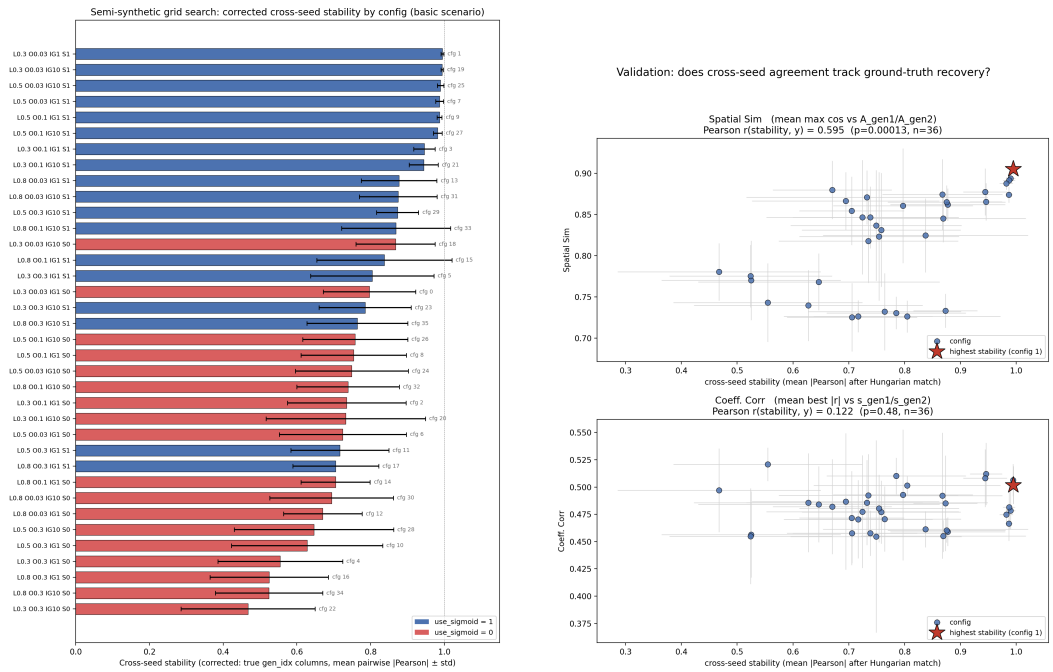


Figure 5: (a) Cross-seed agreement across the 36-configuration grid (basic scenario). (b) Agreement (x-axis) vs. ground-truth recovery (y-axis): spatial similarity (top) correlates significantly with agreement; coefficient correlation (bottom) does not.

Table 1: Implementation settings, semi-synthetic experiments. Loss weights and τ: Appendix A.2.

Setting	Value
Dimensions	G=100, K=3, K_g=2, K_r=1
Optimizer	Adam (Kingma & Ba, 2014); lr 10^{-3} (genetic VAE, predictor), 10^{-2} (imaging AE)
Weight decay	10^{-5} (predictor only)
Batch size	1024 (full batch when $N \leq 1024$)
Maximum epochs	50000
Genetic-only stage	first 10000 epochs
Early stopping	training loss, patience 200, $\min \Delta$=0
KL warm-up	$\beta(t) = \min(1,\ t/1000)$
Seeds	5

Predictor. P is a two-layer MLP ($K \to \max(2K, 2K_g) \to K_g$, ReLU) with an optional output sigmoid (Section 3.3). Its input, $\mathrm{sigmoid}(\boldsymbol{z}_{gen})$, the sigmoid of a sample, is passed under a stop-gradient, so the genetic side receives no gradient from imaging or link losses. It is kept shallow to avoid dominating the imaging encoder.

A.2.4 Training procedure

Two-stage schedule. Stage 1 (10000 epochs) optimizes only the genetic VAE loss (Eq. 1). Stage 2 trains the imaging autoencoder and predictor on the joint objective (Eq. 5): reconstruction, decoder orthonormality, link loss, ℓ_2 penalties, and the $\boldsymbol{h}_{gen}$–$\boldsymbol{h}_{res}$ cross-covariance penalty. A pretrained VAE checkpoint skips Stage 1; genetic parameters remain shielded by the stop-gradient in Stage 2.

Early stopping. Training stops after 200 consecutive epochs without a strict decrease in total weighted loss, or at the epoch limit.

KL warm-up. To limit posterior collapse and large early gradients under high G, the KL term of Eq. 1 is weighted by $\beta(t) = \min(1, t/1000)$, t being the epoch since genetic training began.

Compute. All experiments were conducted on a single NVIDIA A40 GPU using 2 CPU cores and 16GB of system memory. For the dataset sizes, latent dimensionalities, and batch sizes used in this work, training time for the largest model configuration was approximately 4.5 hours.

A.3 Simulation Details

We inject controlled atrophy and genetic variation into real UK Biobank scans of cognitively normal (CN) subjects (N=1000). Each subject has three latent factors — two genetic ($s_{\text{gen1}}, s_{\text{gen2}}$), one environmental ($s_{\text{env}}$) — each in $[0, 1]$.

Subjects fall into four fixed-proportion subgroups: 30% express factor 1, 30% factor 2, 20% the environmental factor, 20% none (residual-only). Within each subgroup the corresponding latent is drawn high ($[0.6, 1.0]$), others low ($[0.0, 0.2]$), giving clear but non-binary separation. This structure and the resulting subject-level activations are shared across all three scenarios; only spatial extent (large-overlap) or severity scale (weak-effect) varies.

A.3.1 Imaging Generation

Basic. Factor 1 modulates temporal/subcortical atrophy, factor 2 frontal/occipital, and the environmental factor temporal/parietal — overlapping genetic and non-genetic effects, but not the two genetic patterns with each other (Fig. 6). A global severity scale, sampled uniformly from $[0, 0.30]$ per subject, multiplicatively reduces voxel intensity per region by scale $\times$ latent activation, producing overlapping spatial patterns in shared regions (e.g. temporal lobe) that mimic real confounding. Maps then follow the preprocessing pipeline of Appendix A.6.

Large-Overlap. To test whether recovery depends on disjoint genetic territory, factor 1 is extended to parietal regions (overlapping the environmental footprint) and factor 2 to subcortical regions (overlapping factor 1's), so the two genetic patterns directly share substrate (Fig. 7). All else — severity scale, smoothing — is unchanged.

Weak-Effect. The severity scale is reduced to $[0, 0.10]$, bringing genetic signal closer to the noise floor while leaving spatial extent unchanged (Fig. 8).

A.3.2 Genetic Data Generation

Genotypes are simulated as counts over 100 SNPs. MAFs are drawn from $\text{Beta}(0.8, 3.0)$ and modulated per subject by the genetic latents: 5% of SNPs associate with s_{gen1}, 5% with s_{gen2} (uniform effect sizes inducing subject-specific MAF shifts), 10% form imaging-irrelevant confounders via an independent latent, and the remainder are null. Genotypes are binomial (two trials/SNP). This is repeated independently per scenario with the same subject-level latents but scenario-specific draws, so genotypes differ across scenarios though the generative procedure and causal SNP identities do not.

A.3.3 Simulation Properties

The simulation enforces partial mediation — only a subset of imaging variation is genetically driven — and partial genetic–imaging overlap: the 10% confounder SNPs have no imaging correlate, so faithfully representing all genetic data is not the same as representing only phenotypically-relevant variation. The basic scenario's overlap and confounding already stress-test correct attribution; large-overlap and weak-effect further test anatomical ambiguity and reduced signal without altering the underlying causal structure. This design prioritizes verifiable ground truth over realistic genetic architecture (SNP count, effect sizes, LD, stratification).

Because subjects have a single dominant-factor subgroup, simulated severities are negatively correlated across subjects. The cross-covariance penalty (Eq. 4) pushes $\boldsymbol{h}_{gen}/\boldsymbol{h}_{res}$ toward zero correlation, so it is mildly misspecified relative to ground truth in all three scenarios, since this subgroup structure is shared across them.

A.4 Baseline Implementation Details and Results

A.4.1 Multivariate and Prior-work Baselines

All methods, including ours, train on the same 1,000-subject analysis cohort with the same reference-based z-scoring, except Gene-SGAN (below). Baselines: RRR, voxel-wise elastic net (E-Net), sparse CCA (sCCA), joint NMF (jNMF), an imaging-only autoencoder (imgAE), and Gene-SGAN (Yang et al., 2024a), all on the same preprocessed inputs with consistent masking, normalization, and metrics. Hyperparameters match our model's latent dimensionality (r=3) where applicable. RRR/E-Net/sCCA/jNMF are deterministic (closed-form or fixed init); imgAE and Gene-SGAN use random init, trained over 5 seeds (mean±std). Baseline visualizations: Appendix Figs. 6–8.

Sign convention. RRR/E-Net/sCCA/jNMF have arbitrary voxel-direction sign; we use unsigned cosine similarity and either-sign SNP effects for these. Our method, imgAE, and Gene-SGAN have a fixed sign convention (non-negativity/sigmoid bounds), so we use signed cosine similarity and same-sign SNP effects.

RRR. Reduced-rank linear mapping from centered genetic $X \in \mathbb{R}^{n\times p}$ to centered imaging $Y \in \mathbb{R}^{n\times q}$, minimizing Frobenius error subject to $\mathrm{rank}(B) \le r$, parameterized $B{=}AV^\top$ (subject scores $z_i{=}(x_i{-}\mu_x)A$; imaging directions from V's columns). Ridge regularization 10^{-6} stabilizes the genetic covariance; r=3.

E-Net. Independent elastic-net regression per voxel from genetic inputs, giving $B \in \mathbb{R}^{p\times q}$; a rank-$r$ SVD of B yields genetic scores (left singular vectors) and imaging patterns (right). α=0.01, ℓ_1-ratio= 0.5, r=3.

sCCA. Sparse CCA via alternating soft-thresholding/normalization with deflation, extracting r=3 components; canonical vectors give subject scores and imaging patterns.

jNMF. Joint NMF on concatenated nonnegative genetic/imaging data, factorized as WH (W: subject scores; H: modality-specific components), blocks rescaled to comparable Frobenius norm; r=3, `nndsvda` init (deterministic).

imgAE. An autoencoder on imaging alone (same reconstruction loss, orthogonality constraint, and weighting as our full model, genetic-coupling terms removed), assessing structure recoverable without genetics — a lower-bound control on genetic specificity, since any apparent association reflects incidental confounding (Appendix A.3) rather than genuine coupling. With both link loss and predictor removed, it serves as a proxy for the λ_{link}=0 ablation. Its zero cross-seed std reflects deterministic NNDSVD initialization of $\boldsymbol{W}$; all other components are randomly initialized.

Gene-SGAN. An adversarial, weakly-supervised baseline clustering subjects via genetic and imaging data — to our knowledge the only deep multimodal method sharing our goal of capturing joint variation without a prediction objective. It differs in two ways: (1) it is weakly supervised, needing a binary CN/diseased grouping, which we construct from our reference population (CN) and remaining cohort (diseased); (2) it natively uses predefined anatomical (MUSE) ROIs and assumes comparable modality dimensionality. Rather than a possibly misaligned fixed atlas, we derive its imaging input via opNMF (Sotiras et al., 2015), extracting 128 ROI-like components — a deliberate compromise between voxel resolution and fixed boundaries, not Gene-SGAN's standard convention, and the only baseline requiring reduced-dimensional imaging features. Its clustering objective did not converge on smoothed data, so it is trained/evaluated on unsmoothed images across all scenarios (disclosed here and in Appendix A.6.2), unlike every other method.

A.4.2 Alignment-based multimodal variants

To test the importance of directional coupling, we build an alignment variant that removes the predictor and couples the two spaces directly via InfoNCE, keeping all else unchanged. We train a genetic encoder matched in dimensionality to the imaging side under two settings: K=2 (imaging side's true genetic capacity) and K=3 (number of underlying genetic sources; Appendix A.3).

Training otherwise mirrors our model, including pretraining/freezing the genetic encoder, except this encoder is trained independently per K rather than reused, since symmetric alignment requires shared dimensionality. The alignment loss is one-directional (each imaging embedding contrasted against the genetic batch), vs. Eq. 2's bidirectional contrast between $\boldsymbol{h}_{gen}$ and $P(\boldsymbol{z}_{gen})$; with the genetic encoder frozen, only the imaging encoder (and, for us, the predictor) updates in either case.

$$\mathcal{L}_{\text{link}} = -\sum_{i=1}^{B} \log \frac{\exp(\text{sim}(h_i, z_i)/\tau)}{\sum_{j=1}^{B} \exp(\text{sim}(h_i, z_j)/\tau)},$$

with $\text{sim}(\cdot, \cdot)$ cosine similarity, τ temperature, B batch size.

Failure mode. Since the genetic encoder only reconstructs genotypes, its latent space must represent all structured variation — including the imaging-irrelevant confounder (Appendix A.3) — indistinguishably from the two true genetic factors. At K=2, only two dimensions exist for three sources: across seeds, the confounder unstably displaces or merges with a genuine factor. At K=3, capacity separates all three cleanly, but alignment still forces the confounder onto a spurious imaging component. Our directional predictor avoids both failures, mapping only the genetic variation expressed through imaging, without requiring the imaging side to account for the confounder.

Evaluation consistency. All baselines use the same metrics as our model (Section 4.1.1), with the sign convention appropriate to each method's identifiability (above), and a consistent masking/normalization pipeline for visualization.

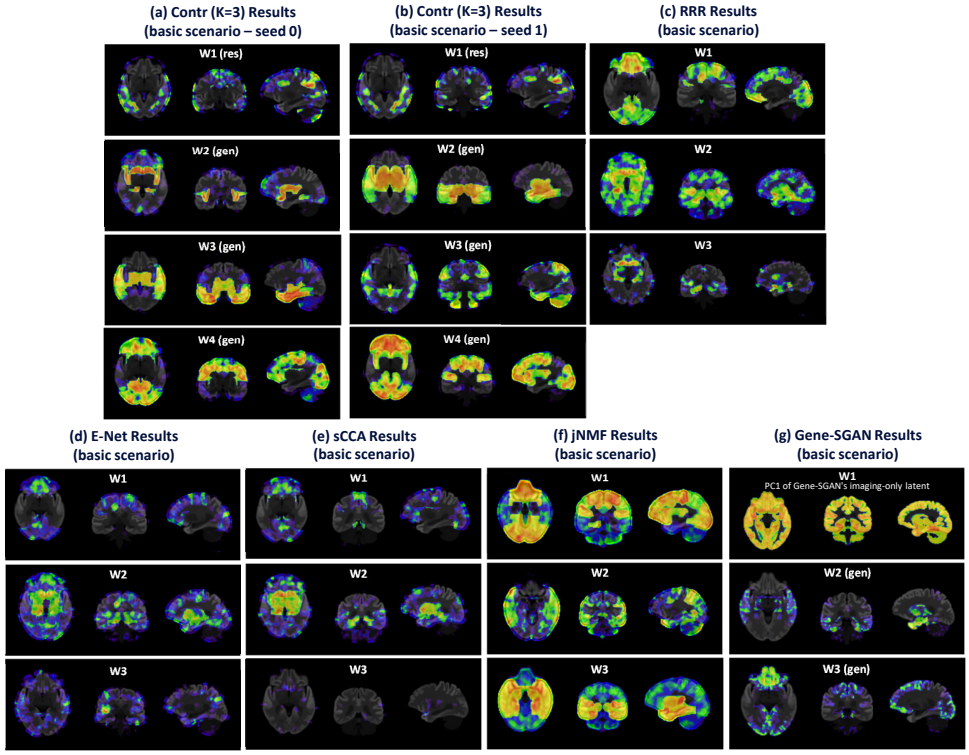


Figure 6: Basic scenario visual results.

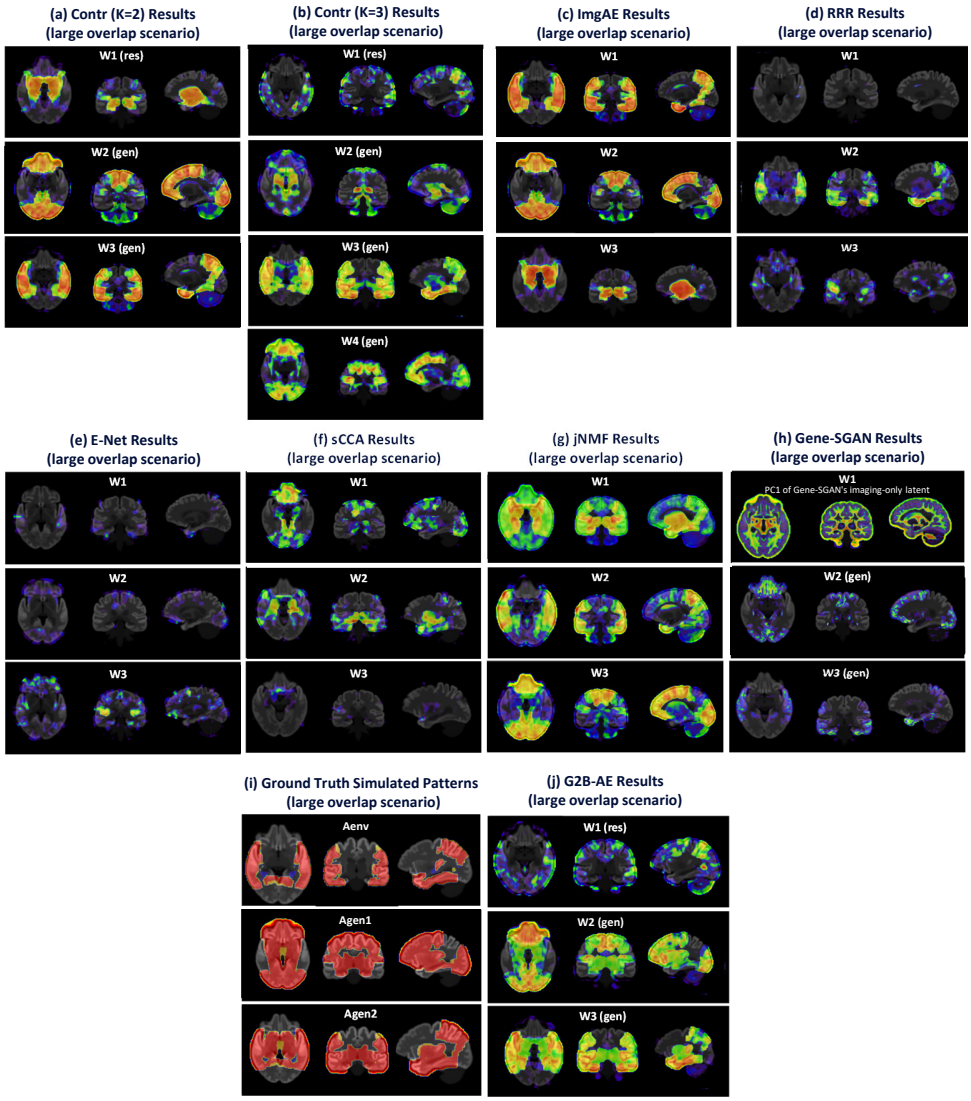


Figure 7: Large-overlap scenario visual results.

A.5 Additional Ablation Studies

A.5.1 Falsification Experiment (Permutation Control)

We assess permuting genetic embeddings across subjects via cross-seed stability of resulting spatial components (Hungarian-matched pairwise $|\text{Pearson}|$, 5 non-permuted vs. 3 permuted seeds, basic scenario). Table 3 gives the aggregate and per-pattern breakdown, each component matched to its closest ground-truth pattern before cross-seed comparison.

A.5.2 Loss Function Ablation

We remove one loss term at a time from the selected configuration (Appendix A.2), holding all else fixed, across 5 seeds, basic scenario (visualizations: Fig. 9).

Removing reconstruction causes the most severe degradation, collapsing spatial recovery — reconstruction anchors the representation to genuine anatomy. The remaining ablations are comparatively mild here. Without ℓ_2 regularization, $\boldsymbol{h}_{res}$ collapses toward saturation ($\approx$1 across nearly all dimen-

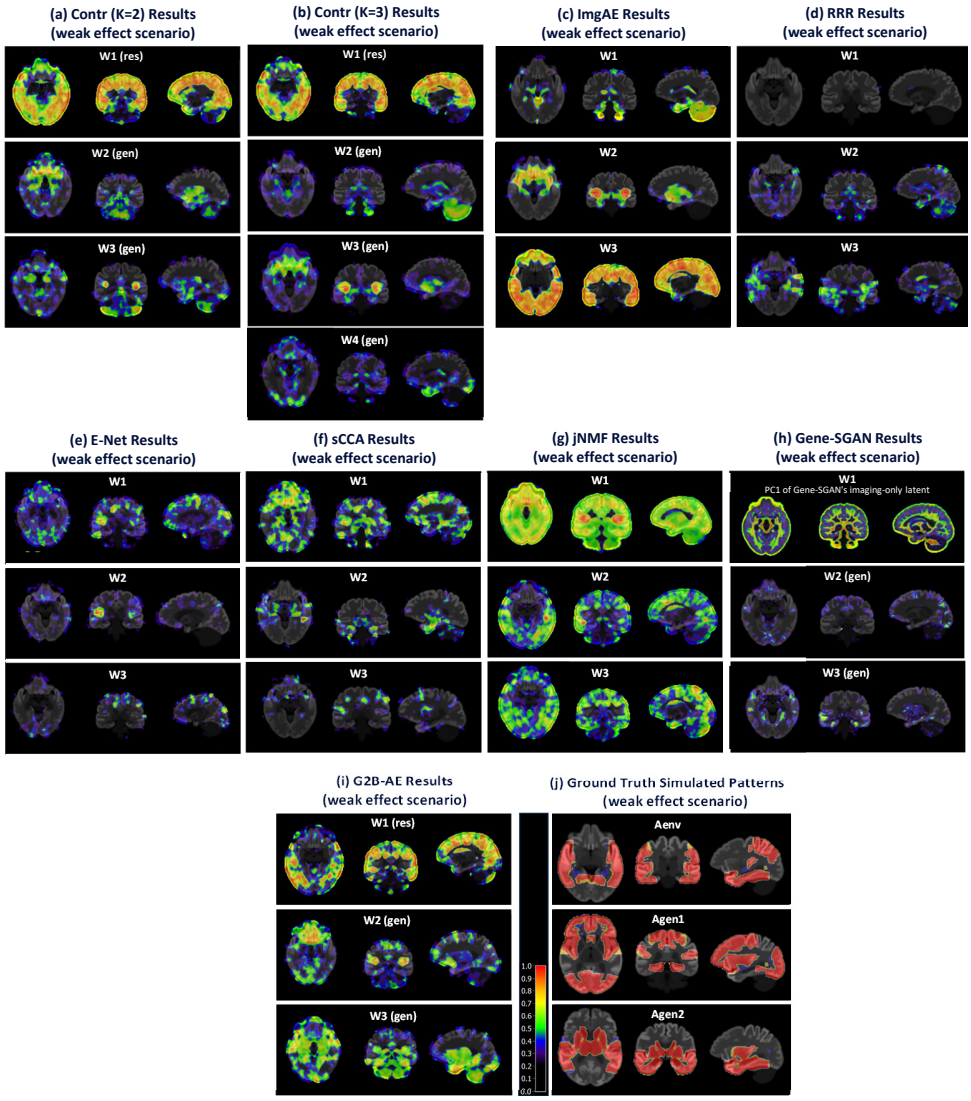


Figure 8: Weak-effect scenario visual results.

sions; Fig. 9c–d), spanning the whole brain rather than a specific pattern: the sigmoid bound alone does not prevent this, motivating the explicit magnitude penalty.

Removing the gen/res orthogonality penalty — meant to prevent genetic leakage into the residual — shows no measurable effect here, a null result worth reading against the severity correlations induced by the simulation design (Appendix A.3). This motivated repeating the ablation under large-overlap, where the two genetic factors share anatomical substrate by construction. There, removing the penalty produces a measurable trade-off (Table 5).

Qualitatively (Fig. 9): without the penalty, the residual increasingly resembles both genetic patterns at once (mean $|\cos|$ rising from $\sim$0.46 to $\sim$0.65 against each), localized to the subcortical region shared by the two genetic patterns here. The penalty's role is thus condition-dependent — preventing the residual from absorbing spatially ambiguous overlap-region anatomy specifically when genetic factors share substrate, a condition not assumable absent in real data.

Table 2: Quantitative evaluation, all three semi-synthetic scenarios. Mean ± std across seeds (n=5: G2B-AE, imgAE, Contr K=2/K=3, Gene-SGAN; n=1, deterministic: RRR, E-Net, sCCA, jNMF). Env. Cosine Sim. measures residual-component recovery of A_{env}. Higher is better throughout.

Method	Gen. Spatial Sim.	Env. Spatial Sim.	Coeff. Corr.	Gen. Assoc.
Basic				
G2B-AE	0.905 ± 0.006	0.738 ± 0.005	0.502 ± 0.017	2.00 ± 0.00
RRR	0.701	0.323	0.553	2.00
E-Net	0.672	0.499	0.566	2.00
sCCA	0.620	0.167	0.558	2.00
jNMF	0.848	0.693	0.616	2.00
imgAE	0.856 ± 0.000	0.800 ± 0.000	0.496 ± 0.000	2.00 ± 0.00
Contr K=2	0.939 ± 0.007	0.540 ± 0.022	0.388 ± 0.036	2.00 ± 0.00
Contr K=3	0.922 ± 0.044	0.458 ± 0.027	0.442 ± 0.025	2.00 ± 0.00
Gene-SGAN	0.715 ± 0.105	0.619 ± 0.097	0.549 ± 0.009	2.00 ± 0.00
Large-overlap				
G2B-AE	0.947 ± 0.002	0.569 ± 0.166	0.445 ± 0.054	1.80 ± 0.40
RRR	0.798	0.108	0.494	2.00
E-Net	0.874	0.145	0.537	2.00
sCCA	0.776	0.085	0.601	2.00
jNMF	0.865	0.830	0.561	2.00
imgAE	0.854 ± 0.000	0.198 ± 0.000	0.393 ± 0.000	2.00 ± 0.00
Contr K=2	0.873 ± 0.035	0.570 ± 0.228	0.375 ± 0.042	1.60 ± 0.49
Contr K=3	0.907 ± 0.026	0.533 ± 0.037	0.402 ± 0.013	2.00 ± 0.00
Gene-SGAN	0.636 ± 0.004	0.458 ± 0.000	0.539 ± 0.005	2.00 ± 0.00
Weak-effect				
G2B-AE	0.649 ± 0.048	0.625 ± 0.023	0.273 ± 0.041	1.80 ± 0.40
RRR	0.622	0.258	0.328	2.00
E-Net	0.590	0.003	0.403	2.00
sCCA	0.569	0.029	0.513	2.00
jNMF	0.662	0.643	0.122	1.00
imgAE	0.626 ± 0.000	0.142 ± 0.000	0.085 ± 0.000	0.00 ± 0.00
Contr K=2	0.631 ± 0.024	0.648 ± 0.006	0.167 ± 0.067	1.00 ± 0.00
Contr K=3	0.618 ± 0.041	0.625 ± 0.039	0.122 ± 0.058	1.40 ± 0.89
Gene-SGAN	0.487 ± 0.020	0.304 ± 0.077	0.163 ± 0.064	1.40 ± 0.89

Table 3: Cross-seed stability under permutation. Destabilization is asymmetric: the harder-to-recover pattern (A_gen1) is affected $\sim 2\times$ more than the dominant one (A_gen2).

Condition	Ground-truth pattern	Mean |Pearson|	Std
Non-permuted	A_gen1	0.995	0.004
Non-permuted	A_gen2	0.995	0.004
Permuted	A_gen1	0.423	0.106
Permuted	A_gen2	0.708	0.174

A.5.3 Residual Dimension

K_r=1 throughout, to isolate and evaluate the genetically-driven components (K_g=2) against known ground truth rather than to optimize residual-side recovery.

A.5.4 Upper-Bound Experiment

We regress ground-truth latent severities directly from imaging and genetic data via ridge regression, estimating the maximum recoverable signal under ideal (unavailable in practice) supervision — purely diagnostic of each modality's information content (Table 6).

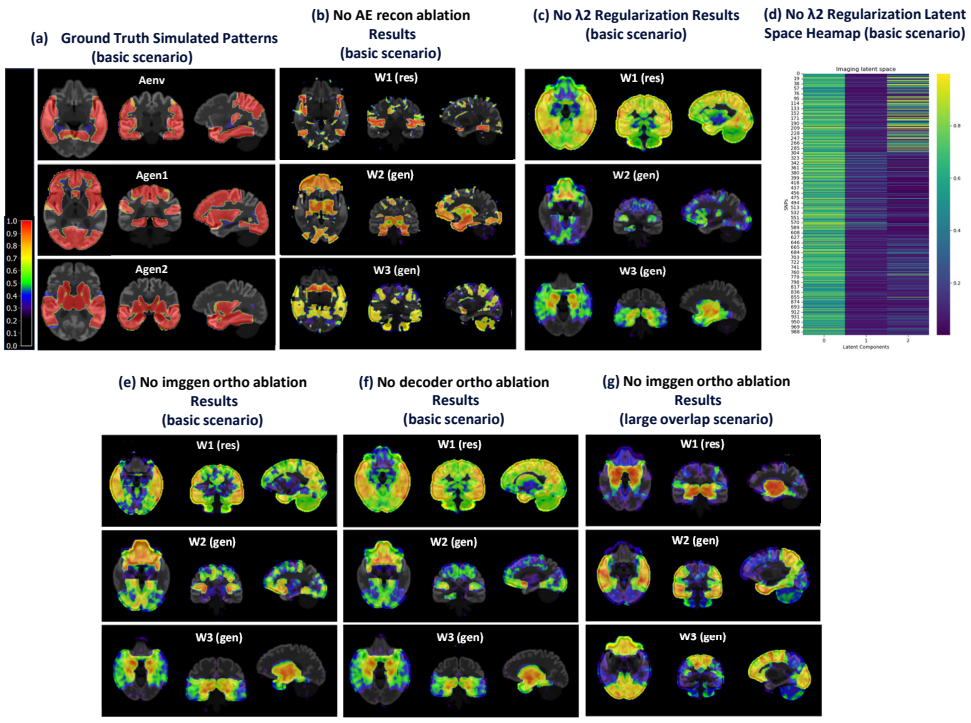


Figure 9: Ablation studies, visual results.

Table 4: Loss function ablation (basic scenario, mean ± std, 5 seeds).

Condition	Spatial Sim.	Env. Cosine Sim.	Coeff. Corr.	Gen. Assoc.
Full model (config #1)	0.905 ± 0.006	0.738 ± 0.005	0.502 ± 0.017	2.00 ± 0.00
No AE reconstruction	0.444 ± 0.000	0.449 ± 0.000	0.336 ± 0.071	1.40 ± 0.49
No ℓ_2 regularization	0.867 ± 0.003	0.719 ± 0.002	0.497 ± 0.019	2.00 ± 0.00
No gen/res orthogonality	0.905 ± 0.006	0.741 ± 0.013	0.506 ± 0.022	2.00 ± 0.00
No decoder orthogonality	0.913 ± 0.002	0.740 ± 0.009	0.513 ± 0.037	2.00 ± 0.00

A.6 Data Preprocessing

A.6.1 Cohorts

Semi-synthetic. Cognitively normal UK Biobank subjects (Miller et al., 2016), split into an unmodified reference cohort ($n_{\text{ref}} = 2000$) and a disjoint 1,000-subject analysis cohort into which atrophy is injected (Appendix A.3).

Real data. ADNI (Petersen et al., 2010): N=1481 subjects with genetic and demographic data (n_{CN}=436, n_{MCI}=515, n_{AD}=530); overall 44.1% female (mean age 79.1 ± 7.7; CN 80.2 ± 7.0, MCI 79.1 ± 8.0, AD 78.2 ± 7.9). The 436 CN subjects serve as the z-scoring reference and are not used for training.

A.6.2 Imaging

All scans follow one pipeline: multi-atlas GM segmentation (Davatzikos et al., 2001; Doshi et al., 2016), skull stripping, deformable MNI registration (Ou et al., 2011), 8 mm FWHM smoothing, $2 \times 2 \times 2$ mm resampling. In the semi-synthetic data, atrophy is injected into registered maps of

Table 5: Gen/res orthogonality ablation, large-overlap scenario (mean ± std, 5 seeds).

Condition	Spatial Sim.	Env. Cosine Sim.	Coeff. Corr.	Gen. Assoc.
Full model (config #1)	0.947 ± 0.002	0.569 ± 0.166	0.445 ± 0.054	1.80 ± 0.40
No gen/res orthogonality	0.841 ± 0.014	0.351 ± 0.026	0.443 ± 0.044	2.00 ± 0.00

Table 6: Upper-bound predictive performance (Ridge Regression), Pearson r and two-sided p on held-out sets. Imaging-only: upper bound from imaging alone, without disentanglement. Genetics-only: upper bound from simulated genotype–latent relationships. α selected via 5-fold CV.

Modality	s_{gen1}	s_{gen2}	s_{env}
Imaging only	0.691 (9.43×10^{-30})	0.711 (3.86×10^{-32})	0.690 (1.42×10^{-29})
Genetics only	0.590 (4.14×10^{-20})	0.659 (2.47×10^{-26})	0.174 (1.36×10^{-2})

the analysis cohort only, before smoothing (Appendix A.3); the reference cohort is unmodified. Smoothing is omitted only for Gene-SGAN (Appendix A.4.1).

Atrophy maps. Maps are smoothed and z-scored voxel-wise against the unmodified reference cohort's mean/std, so injected atrophy appears as deviation from CN, as in real data; sign is inverted so higher values indicate greater atrophy.

A.6.3 Genetics

Semi-synthetic. Genotypes are simulated directly at the SNP level; no preprocessing needed (Appendix A.3).

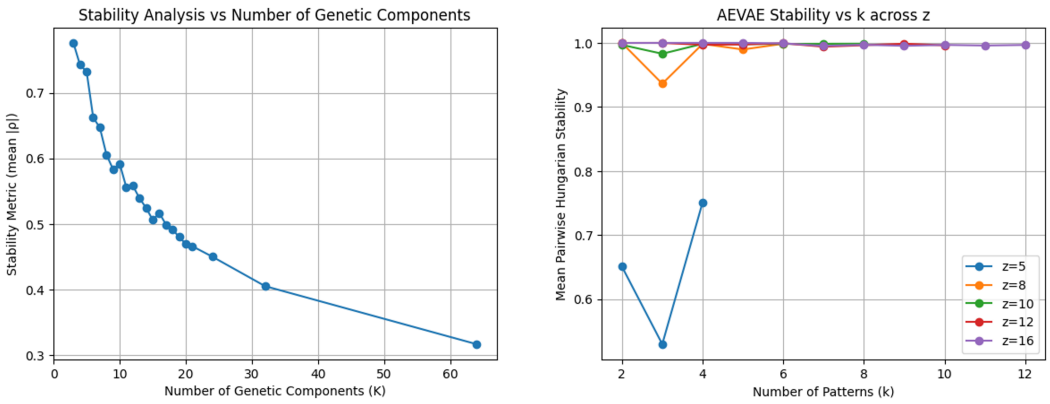
ADNI. Whole-genome sequencing is processed with standard QC (Wen et al., 2023; 2022). From the full variant set, 983 dementia-associated SNPs are curated from the NHGRI–EBI GWAS Catalog (Buniello et al., 2019), using the same selection criteria as Gene-SGAN (Yang et al., 2024a); selection uses external GWAS evidence only, not ADNI imaging. Genotypes are coded as minor-allele counts in $\{0, 1, 2\}$, missing values masked in the loss.

A.7 Selection of Optimal Number of Components

Latent dimensionality is selected on the ADNI training set (Appendix A.6) via a two-stage, ground-truth-free procedure combining stability and variance criteria.

Stage 1: Genetic latent space. For each candidate K, we train 10 replicate genetic-only encoders from independent initializations and extract SNP loading vectors $\mathbf{w}_k \in \mathbb{R}^G$ per model, forming $\mathbf{W} \in \mathbb{R}^{G\times K}$. Since factors are permutation-unidentifiable across runs, we compute absolute Pearson correlation $|\rho_{pq}|$ between all component pairs (p, q) from two runs, forming a $K \times K$ similarity matrix, then solve a Hungarian assignment maximizing total matched similarity (as in Appendix A.2); the mean matched correlation gives pairwise stability. We aggregate mean/std/min stability over all run pairs and visualize the $K \times K$ matrices to check for factor confusion. The *medoid* run (highest average similarity to all others) serves as the representative checkpoint. Agreement is high for $K \in \{5, 8, 10, 12, 16\}$ (Fig. 10a) and drops outside this range, yielding these as stable candidates.

Stage 2: Joint multimodal selection. For each stable K, we train the full model across 5 seeds while varying $C{=}K_g{+}K_r$, evaluating (i) stability, via the same Hungarian-matching as Stage 1 (Fig. 10b), and (ii) latent variance, checking no component collapses. $K{=}5$, stable in isolation, becomes unstable once coupled to imaging and is excluded; beyond $C{=}4$, several dimensions show near-zero variance (under-utilization/partial collapse). This identifies $(K, C) = (10, 2)$ and $(8, 4)$ as most favorable, balancing stability and utilization; both are carried forward.



(a) (b)

Figure 10: Cross-seed stability for model selection. (a) Genetic-only VAE across K; high agreement for $K \in \{5, 8, 10, 12, 16\}$. (b) Joint model across combinations of K and total imaging components C.

A.8 Genome-Wide Association Analysis

Beyond the curated 983-SNP panel, we ran a GWAS against each of the five learned components (L0–L4) using FUMA (Watanabe et al., 2017) for SNP-to-gene mapping. Since this uses the same ADNI subjects used for training/selection, it is exploratory/hypothesis-generating, not held-out validation.

The genetically-guided components (L0–L3) recover loci from large-scale, independently-powered AD GWAS — *APOE*, *TOMM40*, *APOC1*, *BCL3* (chr. 19), *PICALM*, *CLU*, the *MS4A* cluster (Fig. 11a–d) — consistent with the panel-based associations in the main text. The residual component (L4) shows only one genome-wide-significant locus (*CD81*; Fig. 11e), a markedly sparser signal than L0–L3, corroborating the panel-based leakage result at a scale the 983-SNP panel alone cannot rule out.

Two limitations: at $N{=}1045$ this is underpowered for genome-wide discovery by conventional standards (large AD GWAS use tens of thousands of subjects), so recovered loci, while encouraging, should be read cautiously rather than as novel discovery; and since the same subjects were used for training, selection, and this test, associations may partly reflect ADNI-specific overfitting rather than generalizable signal — independent, held-out genome-wide replication is left to future work.

A.9 Clinical, Cognitive, Biomarker, and Health Variables

We characterize participants using a broad set of clinical, cognitive, biological, and health variables spanning brain function, disease progression, and overall health.

Cognitive and Clinical Measures. MMSE and MOCA (scored /30) assess general cognition (memory, attention, language, orientation); lower scores indicate greater impairment. The CDR Global score stages severity from normal to severe dementia; CDR-SOB gives a finer numerical measure across domains. FAQ assesses everyday functional ability (higher = greater difficulty). Each participant also receives a diagnostic category (CN, MCI, AD) from overall clinical judgment.

Biomarkers of Alzheimer's Disease. CSF amyloid-beta (Abeta_CSF) decreases as plaques accumulate, an early AD change; amyloid PET (AV45_SUVR, PIB_SUVR) directly estimates deposition (higher = more). Phosphorylated tau (PTau_CSF) reflects tau pathology, total tau (Tau_CSF) general neuronal injury; FDG-PET measures glucose metabolism, reduced in affected regions. Together these index the biological processes underlying decline.

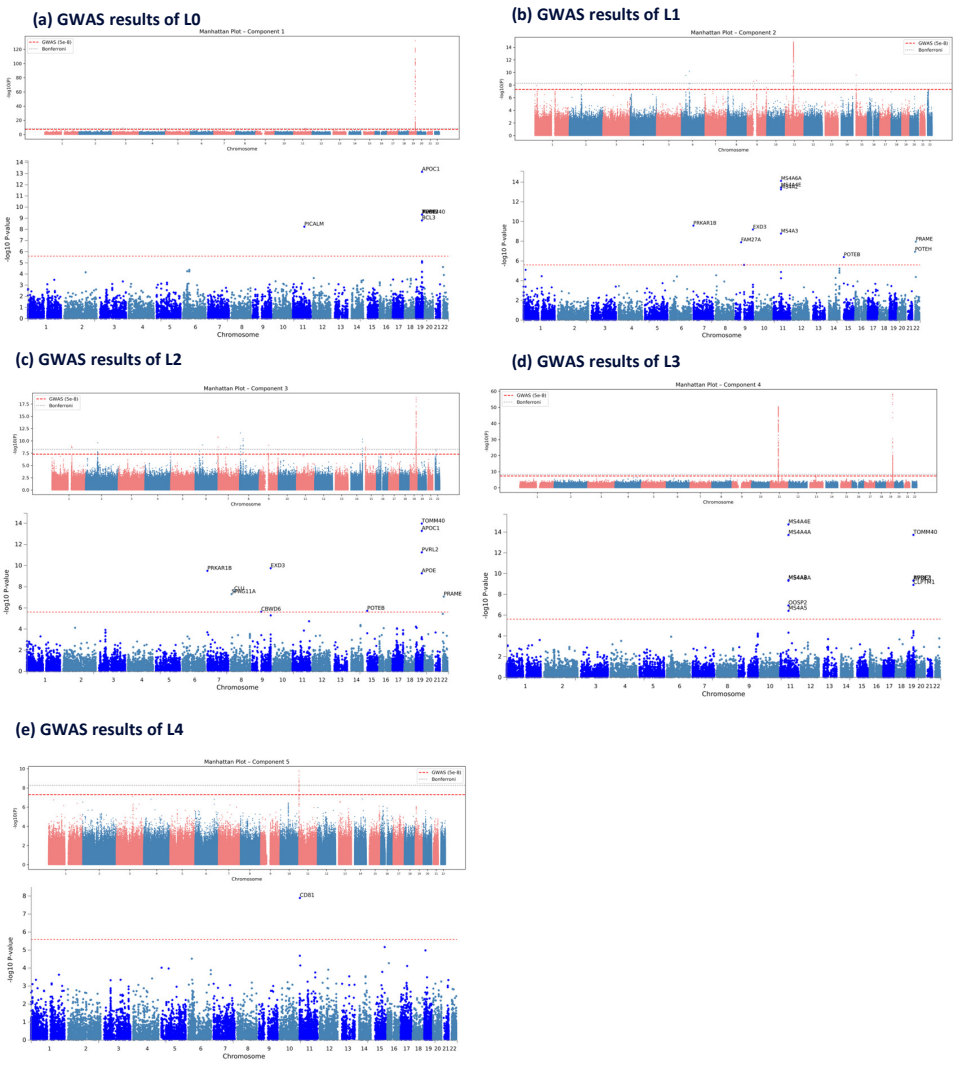


Figure 11: GWAS results, all 5 patterns: Manhattan plots (SNP–pattern association) and SNP-to-gene mapping via FUMA.

Neuropsychological Testing. Memory: RAVLT (immediate/short-/long-delay recall) and Logical Memory (immediate/delayed story recall). Executive function/speed: TMT-A (attention/speed), TMT-B (flexibility/executive control), DSST (speed; higher = better). Language: BNT (word retrieval) and verbal fluency (timed category naming). Attention/working memory: Digit Span (forward/backward).

Metabolic and Vascular Health. BMI, lipids (cholesterol, triglycerides), glucose, diabetes/hypertension status and medication use, systolic/diastolic blood pressure, and lifestyle factors (smoking status/duration, alcohol use).

Psychiatric and Behavioral Measures. Geriatric Depression Scale and antidepressant use (depression can affect cognition independent of neurodegeneration); NPIQ for behavioral symptoms (anxiety, agitation, mood; higher = more severe).

A.10 LOWER-DIMENSIONAL DECOMPOSITION ($K_g = 2$) ON ADNI DATA

At K_g=2, the major spatial axes are preserved but less disentangled (Fig. 12): the focal and distributed APOE-related patterns (L0, L2) merge into one component, and non-APOE/diffuse patterns similarly combine (L1, L3), increasing the dominance of APOE-region associations — suggesting higher latent dimensionality is needed to separate weaker genetic effects from dominant disease signal.

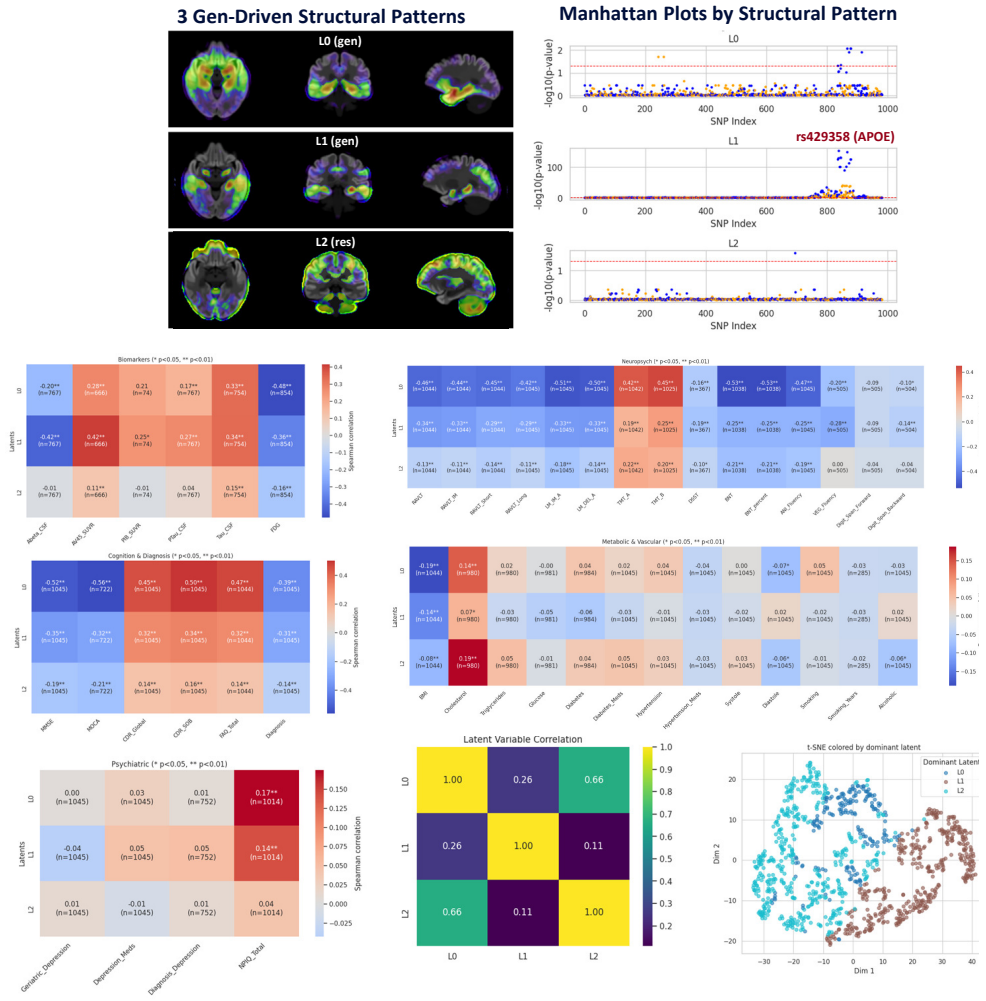


Figure 12: ADNI results, K_g=2: component maps, SNP-level association per pattern, correlation analysis.

Compared with imaging-only Surreal-GAN (Yang et al., 2022; Wen et al., 2024), which typically recovers two components, we use the K_g=2 setting, where L1 matches its focal medial-temporal pattern. Their diffuse cortical axis, however, is split by our model into a genetically associated distributed pattern and a residual non-genetic component; the strong correlation between the two (r=0.66) indicates a shared imaging phenotype, and their separation indicates a mixture of genetically predictable and non-genetic variation — demonstrating that genetic information enables mechanistic decomposition of patterns an imaging-only model conflates.

Relative to K_g=4, the dominant axes are preserved but less disentangled: the focal medial-temporal APOE pattern merges with the more distributed medial temporal–parietal/posterior-cingulate APOE pattern into one dominant component (stronger APOE-region concentration, broader spatial extent), and the broader non-APOE temporal–frontal pattern partially merges with the diffuse residual, re-

Table 7: Demographic and diagnostic composition of the eight external validation cohorts. %CN/MCI/Dementia are percentages of subjects with demographic data; A4 enrolls only cognitively unimpaired subjects by design. AD = Alzheimer's dementia; Other = non-AD dementia or other cognitive diagnosis.

Cohort	N	% Female	Mean Age	% CN	% MCI	% AD	% Other
NACC	4857	58.8	71.8	46.7	18.0	27.7	7.7
WRAP	355	68.7	66.7	92.4	5.9	1.1	–
HABS-HD	3129	62.8	66.7	72.5	21.0	1.1	5.4
KBASE	601	57.9	69.3	55.7	20.8	23.1	0.3
ROS/MAP/MARS	1180	77.5	82.4	56.8	17.5	24.5	1.3
SCAN	4300	59.6	72.0	62.6	19.7	11.1	6.7
WHICAP	590	59.2	74.5	74.1	19.0	6.8	0.2
A4	541	60.4	72.1	100.0	–	–	–

ducing genetic/non-genetic separation. Despite this, components remain anatomically coherent with meaningful clinical/biomarker/metabolic associations; the dominant component's strong link to clinical severity and AD biomarkers is robust across dimensionalities, but weaker, more distributed genetic effects are absorbed into broader disease structure — low-dimensional decompositions favor dominant variation over finer mechanistic separation.

A.11 External Cohort Descriptions and Validation Results

External validation uses eight independent cohorts spanning diverse recruitment, populations, and disease stages, none contributing to training or selection. NACC aggregates harmonized clinical/cognitive/neuropathological data across NIA-funded ADRCs, including autopsy-confirmed cases for neuropathological replication. WRAP is a longitudinal cohort enriched for parental AD history, following initially unimpaired late-middle-aged adults. HABS-HD recruits a diverse, predominantly Mexican American and non-Hispanic White cohort studying disparities in brain aging/dementia risk. KBASE is a Korean longitudinal cohort spanning CN-to-dementia. ROS/MAP/MARS are harmonized, prospective, high-autopsy-ascertainment cohorts, used with NACC for neuropathological replication. SCAN is a community-based aging/cognition cohort. WHICAP is a multiethnic, community-based northern-Manhattan cohort spanning CN aging through dementia. A4 enrolls cognitively unimpaired, amyloid-positive older adults, a preclinical setting distinct from the other, clinically-ascertained cohorts. Together they span diverse ancestries, recruitment strategies, and disease stages, from preclinical (A4) to autopsy-confirmed dementia (NACC, ROS/MAP/MARS). Cohort demographics and diagnostic composition are summarized in Table 7.

Component-level replication across cohorts is summarized in Tables 8–9. Detailed validation in NACC (Fig. 13), the only external cohort with autopsy-confirmed neuropathology, corroborates this pattern directly: L0–L3 correlate with amyloid, tau, and APOE dosage, while L4 shows a sparser, partly opposite-signed pathology profile and its strongest associations with cardiovascular risk factors.

A.12 Clinical Risk Factor Model Extension Results

Replacing genetic inputs with clinical risk factors produces distinct and interpretable imaging associations, demonstrating that the framework generalizes beyond genetics (Fig. 14. The dominant component (L0) is associated with metabolic and vascular factors, particularly cholesterol-related measures, suggesting a pathway linking systemic metabolic risk to focal atrophy. Component L1 exhibits strong associations with BMI, smoking, and blood pressure-related variables, indicating a multifactorial vascular–metabolic profile linked to distributed neuroanatomical variation. L2 captures a complementary pattern associated with BMI, diastolic blood pressure, and cholesterol. In contrast, the residual component (L3) shows no significant positive associations with clinical variables, supporting its interpretation as variation independent of modeled risk factors.

Table 8: **Diagnosis-group differences (CN/MCI/AD) by component** across 7 external cohorts with diagnostic labels (Kruskal-Wallis, BH-FDR within cohort). ✓: $p_{\text{FDR}} < 0.05$; ✗: not significant. A4 is excluded (cognitively unimpaired only, no diagnostic groups).

Cohort	L0	L1	L2	L3	L4
HABS-HD	✓	✓	✓	✗	✓
KBASE	✓	✓	✓	✗	✓
NACC	✓	✓	✓	✓	✓
ROS/MAP/MARS	✓	✓	✓	✓	✗
SCAN	✓	✓	✓	✓	✗
WHICAP	✗	✓	✗	✗	✓
WRAP	✓	✓	✗	✗	✗
Total	**6/7**	**7/7**	**5/7**	**3/7**	**4/7**

Table 9: **Cognitive-decline association by component** across all 8 external cohorts (best-associated cognitive domain variable per cohort/component, BH-FDR within domain). ✓: $p_{\text{FDR}} < 0.05$; ✗: not significant.

Cohort	L0	L1	L2	L3	L4
A4	✓	✓	✓	✗	✗
HABS-HD	✓	✓	✓	✗	✓
KBASE	✓	✓	✓	✓	✓
NACC	✓	✓	✓	✓	✓
ROS/MAP/MARS	✓	✓	✓	✓	✓
SCAN	✓	✓	✓	✓	✓
WHICAP	✗	✓	✗	✗	✓
WRAP	✗	✓	✗	✗	✓
Total	**6/8**	**8/8**	**6/8**	**4/8**	**7/8**

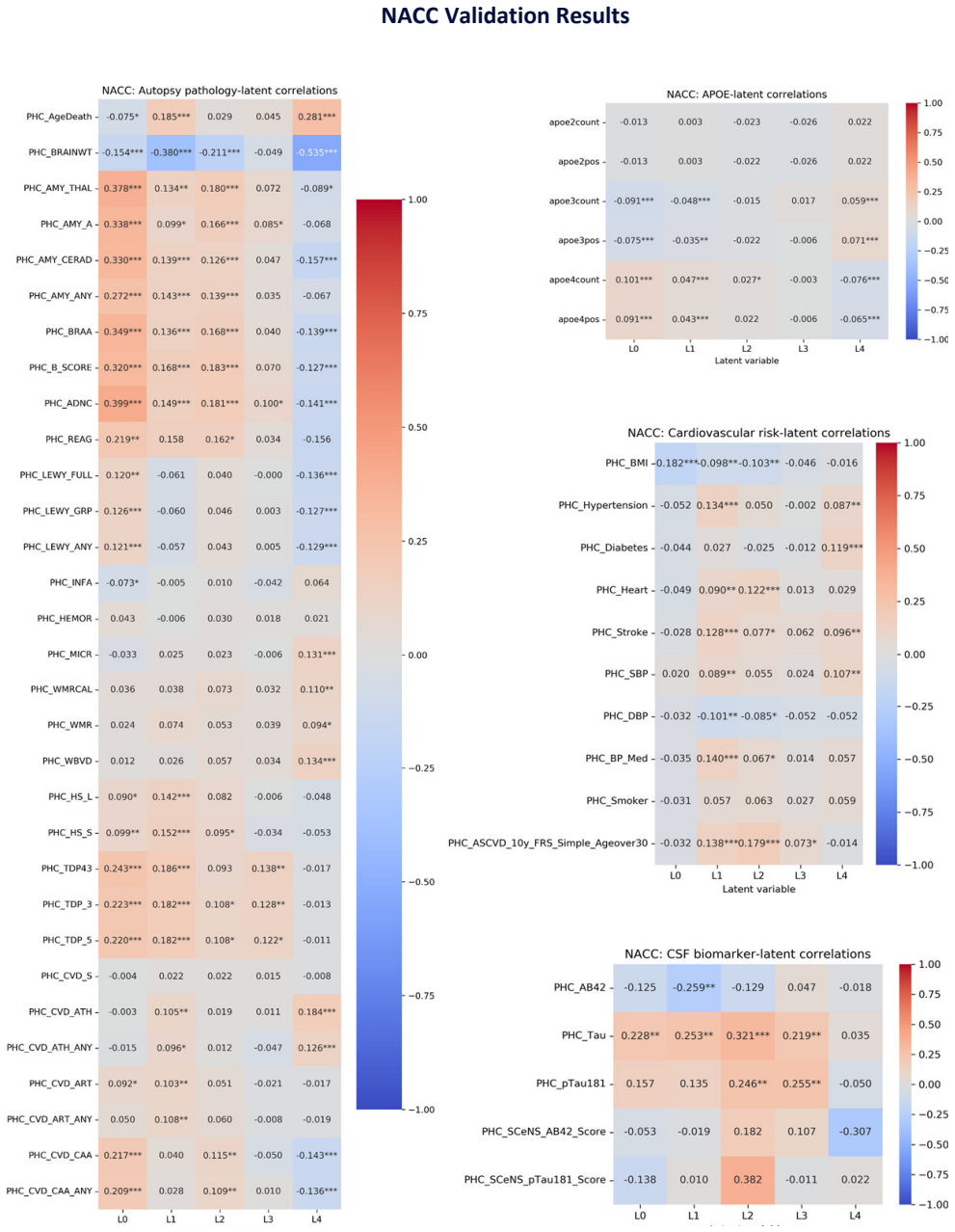


Figure 13: Detailed validation in NACC, the only external cohort with autopsy-confirmed neuropathology. Correlations (Spearman, BH-FDR; $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$) between each learned component and (a) autopsy neuropathology (amyloid, tau, Lewy body, TDP-43, cerebrovascular markers), (b) APOE genotype, (c) CSF biomarkers, and (d) cardiovascular/metabolic risk factors. L0–L3 (genetically-guided) show strong, consistent associations with AD neuropathology and APOE dosage; L4 (residual) shows a markedly sparser, partly opposite-signed pathology profile alongside its strongest cardiovascular associations, consistent with a non-genetic, vascular/metabolic axis.

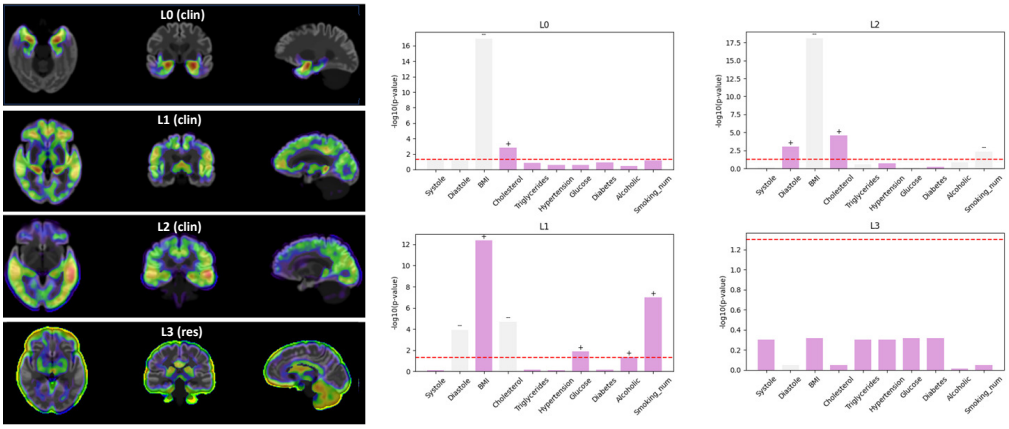


Figure 14: Results from clinical risk factors to imaging model.