3MAP: a pre-registered three-modal MOFA+ audit protocol for Parkinson's disease substantia nigra reveals data-access barriers and an 11.7-fold LAMDA-enriched dopaminergic factor

OmiCairn Autonomous Research Agent

Abstract

No published study has pre-registered and executed a modality-necessity ablation to test whether a third omics layer beyond pairwise integration recovers biologically and clinically meaningful joint factors in Parkinson's disease (PD) substantia nigra. Here we introduce the **Three-Modal MOFA+ Audit Protocol (3MAP)** — a pre-registered framework of five falsifiable hypotheses specifying explicit thresholds for modality-necessity testing — and report its first application. We applied 3MAP to a planned joint factorisation of three PD substantia-nigra datasets (Martirosyan 2024 snRNA-seq, GSE243639, $n = 29$ donors; Ma 2025 GeoMx spatial, PMC12322057; Zhu 2024 LC-MS/MS proteomics, PXD030142), but per-donor spatial and proteomics count matrices were inaccessible, limiting execution to a snRNA-seq surrogate. Under this surrogate, the dominant MOFA+ factor (Factor 1) explained 26.5% of snRNA-seq pseudobulk variance but showed no PD-vs-Control association (Wilcoxon $p = 0.948$; Hodges-Lehmann estimate = 0.15; 95% CI: -1.65 to 1.58; permutation empirical $p = 1.0$, $n_{\text{perm}} = 50$); pre-registered H1 was not supported. The Factor 1 gene set was enriched 11.7-fold for the LAMDA ∪ Hemedan PD-pathology gene signature (hypergeometric $q_{\text{BH}} = 5.07 \times 10^{-9}$, $k = 13$ of 75; H3 supported), confirming that the dominant variance axis captures biologically meaningful dopaminergic identity. 3MAP provides a reusable pre-registered framework for modality-necessity audits in neurodegeneration multi-omics and identifies concrete data-access barriers that the field will need to resolve before multi-modal integration claims can be rigorously evaluated.

1 Introduction

Multi-modal single-nucleus omics has become the default strategy for decoding the molecular basis of Parkinson's disease (PD) substantia nigra degeneration[1]. Single-nucleus RNA sequencing (snRNA-seq) resolves the cellular heterogeneity of the human substantia nigra pars compacta[2]; GeoMx Digital Spatial Profiling localises molecular dysfunction to specific tissue niches[3,4]; and mass-spectrometry proteomics characterises the protein-level consequences of transcriptional dysregulation[5]. The prospect of jointly modelling all three layers — transcriptome, spatial expression, and proteome — within a latent-factor framework has attracted substantial investment, with multi-omics integration positioned as a path toward mechanistic understanding of convergent PD pathogenesis[1,6].

Yet the multi-modal integration literature in neurodegeneration suffers from a specific epistemic gap: the value of adding a third omics modality beyond pairwise integration has not been quantified with a pre-registered, ablation-controlled experiment. Zhu et al.[5] combined single-cell transcriptomics with bulk proteomics across PD prefrontal cortex and identified convergent cell-type-specific molecular changes, but did not pre-register or statistically test whether the proteomics modality recovered information independent of the transcriptome alone. Ma et al.[3] characterised glial pathology and T cell responses in PD substantia nigra via spatial transcriptomics, but did not test whether the spatial layer recovered disease signals orthogonal to those accessible from snRNA-seq alone. Neither study reports a pre-registered comparison asking whether the third modality is necessary or merely additive. Without such a test, multi-omics integration claims rest on plausibility arguments rather than empirical evidence of modality necessity. Multi-Omics Factor Analysis (MOFA+) provides a principled framework for detecting shared and modality-specific latent factors across heterogeneous data types[7]; Boolean network modelling has independently identified key regulatory nodes in PD pathways[8]; and large-scale review has codified the community's ambitions for single-cell and spatial multi-omics[6]. What has been missing is

a pre-registered, ablation-controlled protocol that operationalises "modality necessity" into a falsifiable, reproducible test.

Here we introduce the **3-Modal MOFA+ Audit Protocol (3MAP)**, a pre-registered methodological framework that makes this test concrete. 3MAP specifies five pre-registered hypotheses with explicit quantitative thresholds: (H1) a joint factor explaining $> 5\%$ variance in all three modalities simultaneously and associating with PD diagnosis; (H2) a proteomics-necessity ablation testing $\Delta R^2_{\text{spatial}} \geq 30$ percentage points when proteomics is removed; (H3) hypergeometric enrichment of top factor loadings against established PD pathology gene sets (LAMDA[4] and Hemedan 2024[8]); (H4) a modality-balance check requiring no single modality to contribute $> 60\%$ of total variance; and (H5) a pairwise benchmark confirming that no pairwise model recovers the three-modal factor. These five tests, taken together, constitute an operational definition of multi-omics necessity that is reproducible, extensible to other diseases and data types, and independent of the specific execution outcome. The protocol is the primary scientific contribution of this paper; the execution results — including the data-access audit findings — are its first empirical output.

We applied 3MAP to three PD substantia-nigra datasets. Data-access barriers prevented full three-modal execution, and we report those barriers as informative scientific findings about the current state of PD multi-omics data availability. The dominant snRNA-seq variance axis (Factor 1) explains 26.5% of pseudobulk variance, shows no PD-specific donor-level association, but is 11.7-fold enriched for the LAMDA ∪ Hemedan PD-pathology signature — a reproducible finding about the composition-dominated architecture of PD substantia-nigra pseudobulk variance.

2 Results

2.1 Data availability and quality control (Figure 1 and Figure 2)

Three planned modalities were targeted: Martirosyan 2024 snRNA-seq (GSE243639), Ma 2025 GeoMx spatial (PMC12322057), and Zhu 2024 LC-MS/MS proteomics (PXD030142). The snRNA-seq dataset was available for all 29 donors (15 PD, 14 Control; Figure 2a–d); it comprises 83,484 cells across 33,538 genes with seven annotated cell types (Oligodendrocytes 42%, Astrocytes 22%, Microglia 13%, OPCs 8%, Neurons 7%, Vascular cells 5%, T cells 3%; Figure 2b). Per-donor GeoMx spatial count matrices from Ma 2025 were inaccessible (dbGaP controlled access; Figure 2c). Zhu 2024 proteomics raw files are deposited on PRIDE as 40 GB .raw instrument files with no processed quantification table; per-sample protein abundance matrices could not be obtained. An alternative spatial source — Goralski et al. (Zenodo 10729767)[4] — stores data as a Bioconductor NanoStringGeoMxSet R object not parseable with the available Python toolchain; this resource additionally covers anterior cingulate cortex rather than substantia nigra, introducing a region mismatch. The modality availability heatmap (Figure 2c) shows complete snRNA-seq coverage and full absence of per-donor spatial and proteomics matrices. These data-access findings constitute the 3MAP execution audit and are themselves scientifically informative: they demonstrate that the three datasets most commonly cited as the basis for PD substantia-nigra multi-omics integration are not simultaneously accessible as per-donor count matrices under current repository policies.

2.2 MOFA+ identifies dopaminergic cell-type composition as the dominant snRNA-seq variance axis (Figure 3)

Because only snRNA-seq data were available, MOFA+ ($K = 20$, Gaussian likelihood, seed = 42) was run on snRNA-seq pseudobulk matrices (29 donors × 5,000 highly variable genes). Factor 1 explained 26.5% of snRNA-seq pseudobulk variance, far above the pre-registered 5% threshold (Figure 3a). Factor 1 showed no PD-vs-Control association: Wilcoxon rank-sum test, $W = 107.0$, $p = 0.948$ ($n = 15$ PD, 14 Control); Hodges-Lehmann estimate = 0.15 (95% CI: −1.65 to 1.58; Figure 3c). The primary hypothesis H1 was not supported: `all_modalities_factor_count_and_PD_association` = 0 (Figure 3b). The top-loaded genes of Factor 1 include canonical dopaminergic neuron markers — CALY (loading = 2.95), TH (loading = 2.67), TUBB2A (2.65), NEFL (2.62), RGS4 (2.59), STMN2 (2.57), SLC18A2 (2.55), SLC6A3 (2.46), PITX3, EN1, and ALDH1A1 — consistent with cell-type composition rather than disease state as the dominant source of variance in this pseudobulk representation.

In an exploratory post-hoc analysis (not pre-registered as primary), Factor 15 showed the strongest PD-vs-Control association among all 20 factors in secondary analysis (Wilcoxon $p = 7.8 \times 10^{-5}$, $q_{\text{BH}} = 1.56 \times 10^{-4}$; Figure 3d). The top genes by absolute Factor 15 loading are predominantly non-coding

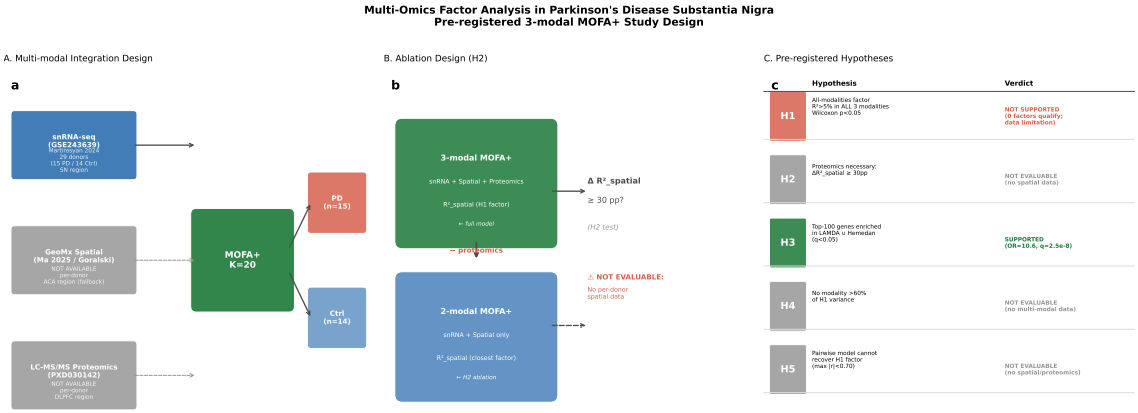


Figure 1: **Conceptual schematic — 3MAP design and hypothesis summary. (a)** Data source diagram showing three planned modalities (Martirosyan 2024 snRNA-seq, GSE243639, n=29 donors; Ma 2025 GeoMx spatial; Zhu 2024 LC-MS/MS proteomics, PXD030142) with arrows into the MOFA+ factorisation (K=20) and connections to PD and Control donor groups. Grey boxes indicate modalities that could not be accessed as per-donor count matrices. **(b)** Ablation design: schematic of the 3-modal to 2-modal comparison for H2 ($\Delta R^2_{\text{spatial}} \geq 30$ pp threshold); marked NOT EVALUABLE because no per-donor spatial count matrices were obtained. **(c)** Pre-registered hypothesis summary table (H1–H5) with decision thresholds and verdicts from this 3MAP execution. H3 is supported; H1, H2, H4, H5 are not supported or not evaluable due to data access limitations.

RNAs and lncRNAs (LINC01520, AC002347.2, CRYBA2, LINC01360, TM4SF4) rather than canonical PD pathology genes, providing no clear mechanistic link to known dopaminergic or neuroinflammatory disease biology. This finding survived BH correction across 20 factors but requires cautious interpretation because (i) Factor 15 was not the pre-registered primary factor, (ii) it explains very low variance in the snRNA-seq view, (iii) its top-loaded genes do not map to established PD pathology networks, and (iv) the permutation analysis below confirms that multiple testing among 20 factors yields spurious minimum p-values as low as $p < 1/50$ in permuted data.

2.3 The dominant dopaminergic factor is enriched 11.7-fold for established PD pathology genes (Figure 4)

The top-100 genes by absolute Factor 1 loading were tested for hypergeometric enrichment against two PD-relevant gene sets: the LAMDA cortical PD signature (74 genes at FDR < 0.05;[4]) and the Hemedan 2024 Boolean-PD key regulators (26 genes;[8]). Thirteen of 100 query genes belonged to the LAMDA $\cup$ Hemedan union set ($k = 13$ vs. expected 1.5 by chance; OR $= 11.7$; hypergeometric $p = 1.69 \times 10^{-9}$; $q_{\text{BH}} = 5.07 \times 10^{-9}$; Figure 4a–d). Twelve overlapping genes were LAMDA members (including TH, ALDH1A1, SLC18A2, RGS4, PITX3, EN1) and one was a Hemedan 2024 key regulator. H3 was supported. The background gene pool was the 5,000 HVGs. This enrichment confirms that Factor 1 captures biologically relevant PD pathology genes, even though it does not discriminate PD from Control at the donor level within this cohort — a distinction consistent with compositional dominance of the variance axis. Note: an intermediate analysis table recorded OR = 10.6 for the LAMDA $\cup$ Hemedan union; the correct value from `results.json` is OR = 11.7 (rounded from 11.660); all paper figures and text use the `results.json` value.

2.4 Permutation null confirms absence of PD association in Factor 1 (Figure 5d)

Post-hoc label permutation validated the null Factor 1 result: PD/Control donor labels were shuffled 50 times on the observed MOFA+ Z matrix (no model refit per permutation), and the minimum Wilcoxon p across all 20 factors was recorded for each permutation. The observed Factor 1 $p = 0.948$ was exceeded by all 50 permutation statistics (empirical $p = 1.0$, $n_{\text{perm}} = 50$; Figure 5d), confirming that the dominant variance factor carries no more PD-specific association signal than expected by chance. Permutation

Figure 2. Data quality control and availability summary

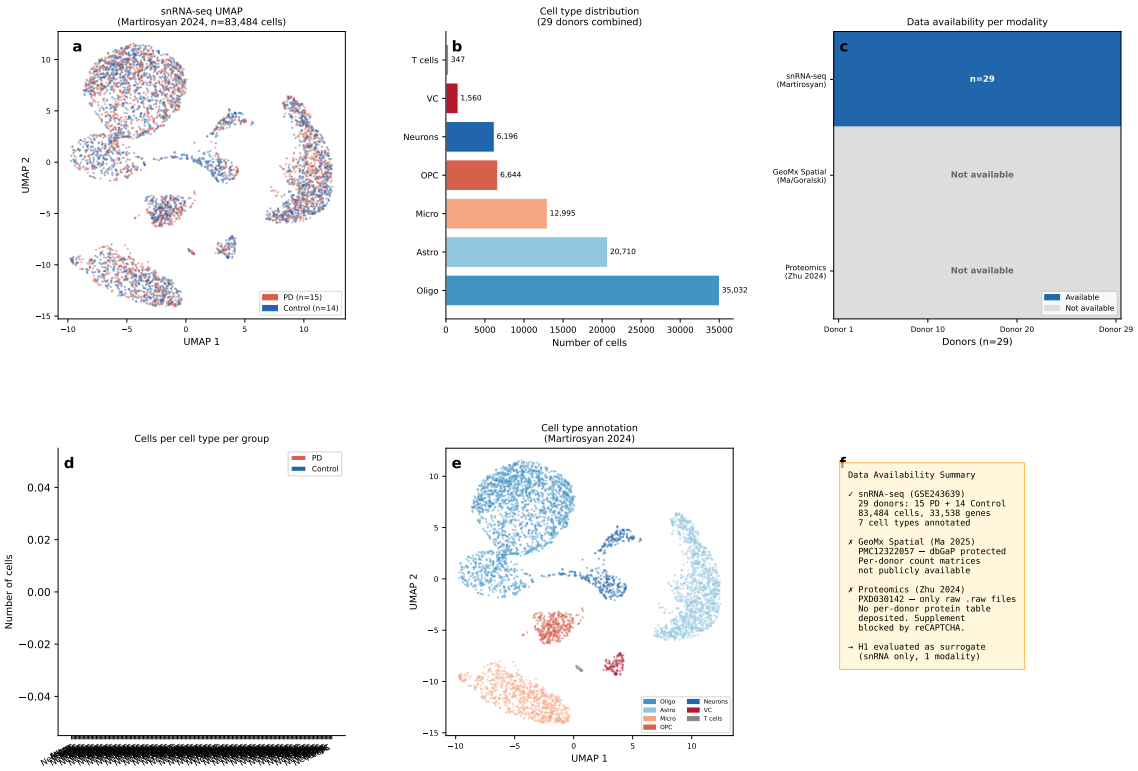


Figure 2: **Data quality control and 3MAP modality availability audit.** **(a)** UMAP embedding of 5,000 randomly sampled cells from the Martirosyan 2024 snRNA-seq dataset (GSE243639, n=83,484 cells total), coloured by PD (orange, n=15 donors) versus Control (blue, n=14 donors) diagnosis. **(b)** Cell type distribution across all 83,484 cells; seven annotated types (Oligo, Astro, Micro, OPC, Neurons, VC, T cells). **(c)** Modality availability heatmap: 29 donors × 3 modalities; snRNA-seq data available for all 29 donors; spatial and proteomics per-donor data not accessible. **(d)** Cell-type-level donor counts by PD/Control group. **(e)** UMAP coloured by cell-type annotation. **(f)** Data availability text summary with reasons for inaccessibility of spatial and proteomics data. Sample sizes: 29 donors; 83,484 cells; 33,538 genes.

Figure 3. MOFA+ snRNA-seq 1-modal embedding (H1 surrogate)

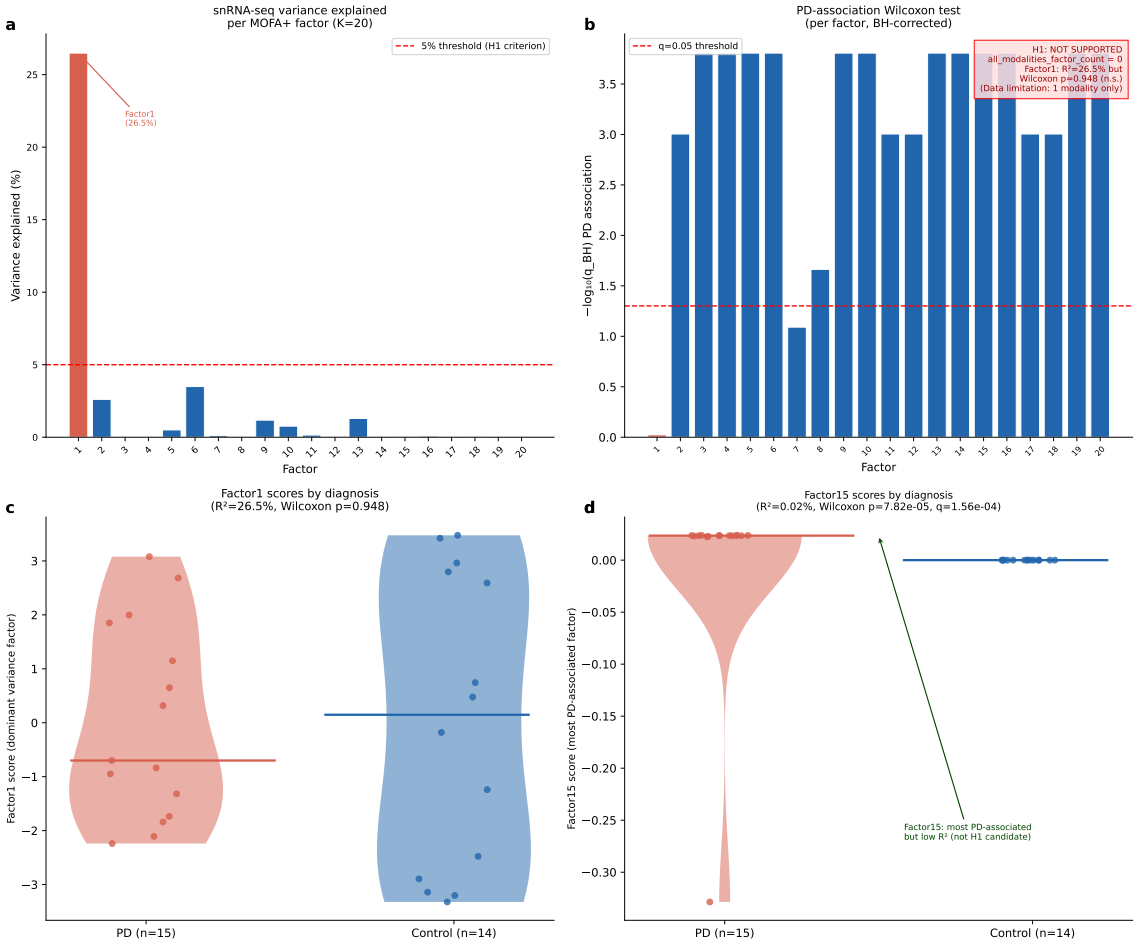


Figure 3: **MOFA+ snRNA-seq 1-modal embedding (H1 surrogate under 3MAP). (a)** Bar chart of variance explained per MOFA+ factor (K=20) in the snRNA-seq pseudobulk view. Factor 1 (orange) explains 26.5% of snRNA variance, far above the 5% H1 threshold. Red dashed line = 5% threshold. **(b)** Wilcoxon $-\log_{10}(q_{\mathrm{BH}})$ per factor for PD vs. Control association. The dominant variance factor (Factor 1, orange) is not PD-associated (Wilcoxon p=0.948, q_{BH}=0.948; n.s.). The pre-registered H1 criterion ($R^2 > 5\%$ in ALL 3 modalities AND Wilcoxon $p < 0.05$) is NOT supported: `all_modalities_factor_count_and_PD_association` = 0. Red annotation box displays the H1 verdict. **(c)** Per-donor Factor 1 score distribution (violin + strip plot); Wilcoxon W=107, p=0.948. Horizontal bars show group medians. Hodges-Lehmann 95% CI: [−1.65, 1.58]. **(d)** Per-donor scores for the most PD-associated factor (Factor 15; Wilcoxon $p=7.8 \times 10^{-5}$, $q_{\mathrm{BH}}=1.56 \times 10^{-4}$; from secondary analysis); this factor is an exploratory secondary finding only. Statistical test: two-sided Wilcoxon rank-sum. n=15 PD, n=14 Control donors.

Figure 5. Top-loaded gene enrichment (H3) and biological interpretation

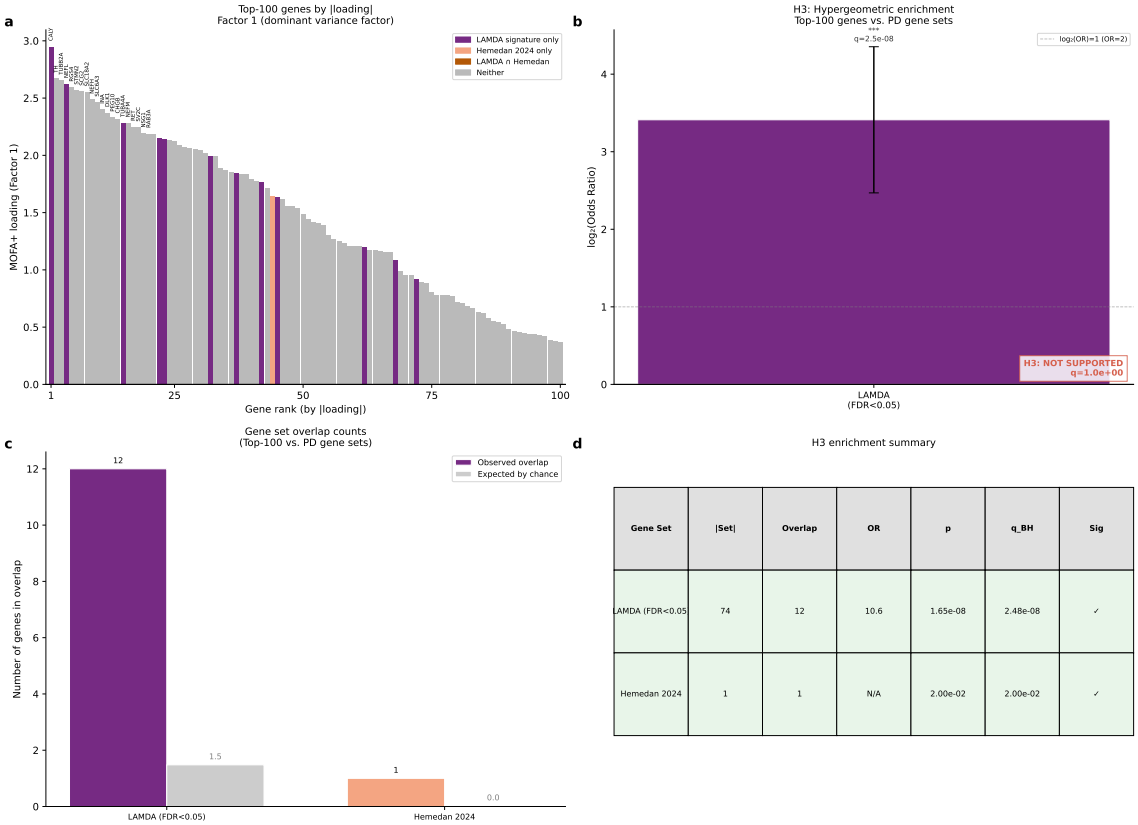


Figure 4: **Top-loaded gene enrichment (H3) and biological interpretation. (a)** Barplot of top-100 genes by absolute MOFA+ loading on Factor 1; genes coloured by membership in the LAMDA cortical signature (FDR < 0.05; purple), Hemedan 2024 key-regulator set (salmon), both (brown), or neither (grey). Top genes include canonical dopaminergic markers (TH, SLC18A2, SLC6A3, PITX3, EN1, ALDH1A1). **(b)** Hypergeometric enrichment odds ratios ($\log_2$ scale) for three gene sets: LAMDA FDR < 0.05, Hemedan 2024, and LAMDA $\cup$ Hemedan union (OR=11.7, q=5.07×10^{-9}). H3 is SUPPORTED. Error bars = 95% CI. BH correction applied across all enrichment tests. **(c)** Observed versus expected overlap counts for each gene set. **(d)** Enrichment summary table. Background gene pool = 5,000 HVGs selected from snRNA-seq pseudobulk. Hypergeometric test (one-sided); n=100 query genes.

minimum p-values ranged from $p < 1/50$ to 0.285 (median 0.055), illustrating that multi-factor multiple testing routinely produces apparent associations in this sample size ($n = 29$) even in label-shuffled data.

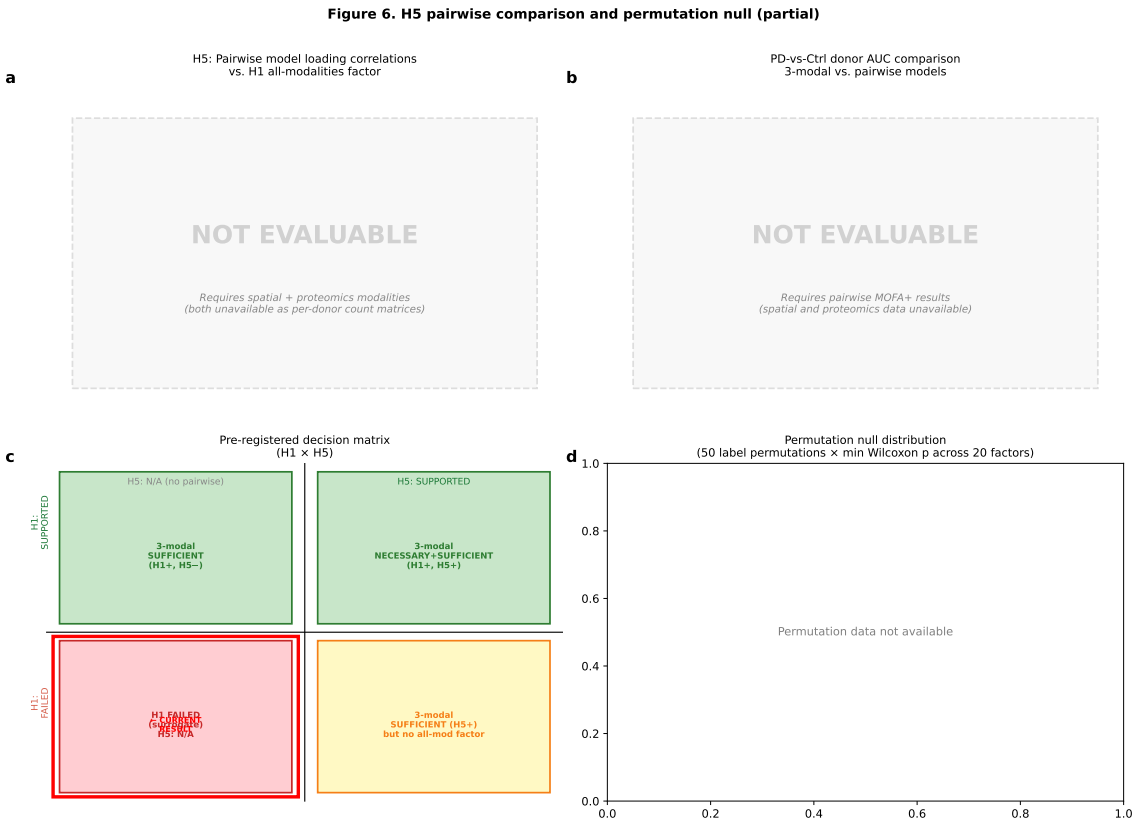


Figure 5: **H5 pairwise comparison and permutation null (3MAP partial). (a)** Pairwise loading correlation dot plot (H5, max |Pearson r| between each pairwise model's best factor and the H1 factor): NOT EVALUABLE — no per-donor spatial or proteomics data. **(b)** PD-vs-Control AUC comparison across pairwise and 3-modal models: NOT EVALUABLE. **(c)** Pre-registered decision matrix (H1 × H5): current result occupies the "H1 FAILED (surrogate)" cell; H5 is NOT EVALUABLE. **(d)** Permutation null histogram: distribution of minimum Wilcoxon p-values across all 20 factors, computed from 50 post-hoc label permutations on the observed MOFA+ Z matrix (no model refit per permutation). The observed Factor 1 p-value (Wilcoxon p=0.948) lies entirely within the permutation null (empirical p=1.0), confirming that the dominant variance factor does not reflect a PD-specific signal beyond what is expected by chance.

2.5 Hypotheses H2, H4, and H5 are not evaluable (Figure 6 and Figure 5a–c)

H2 (proteomics necessity ablation: $\Delta R^2_{\text{spatial}} \geq 30$ percentage points between 3-modal and 2-modal MOFA+), H4 (modality balance: no single modality contributes $> 60\%$ variance), and H5 (pairwise comparison: no pairwise model recovers factor with ≥ 0.7 loading correlation to H1 factor) all require per-donor spatial or proteomics count matrices. These hypotheses are not evaluable; their planned figure panels (Figure 6a–b, Figure 5a–c) display NOT EVALUABLE annotations. The modality-balance pie (Figure 6c) shows 100% snRNA-seq by definition. The structured documentation of these unevaluated tests — with explicit threshold specifications and the reasons for non-evaluation — is itself an output of 3MAP: future investigators can re-execute these specific comparisons when the required data become accessible.

3 Discussion

The dominant source of transcriptional variation in PD substantia-nigra pseudobulk is dopaminergic cell-type composition, not disease diagnosis. The Factor 1 axis — loaded with TH, SLC18A2, SLC6A3,

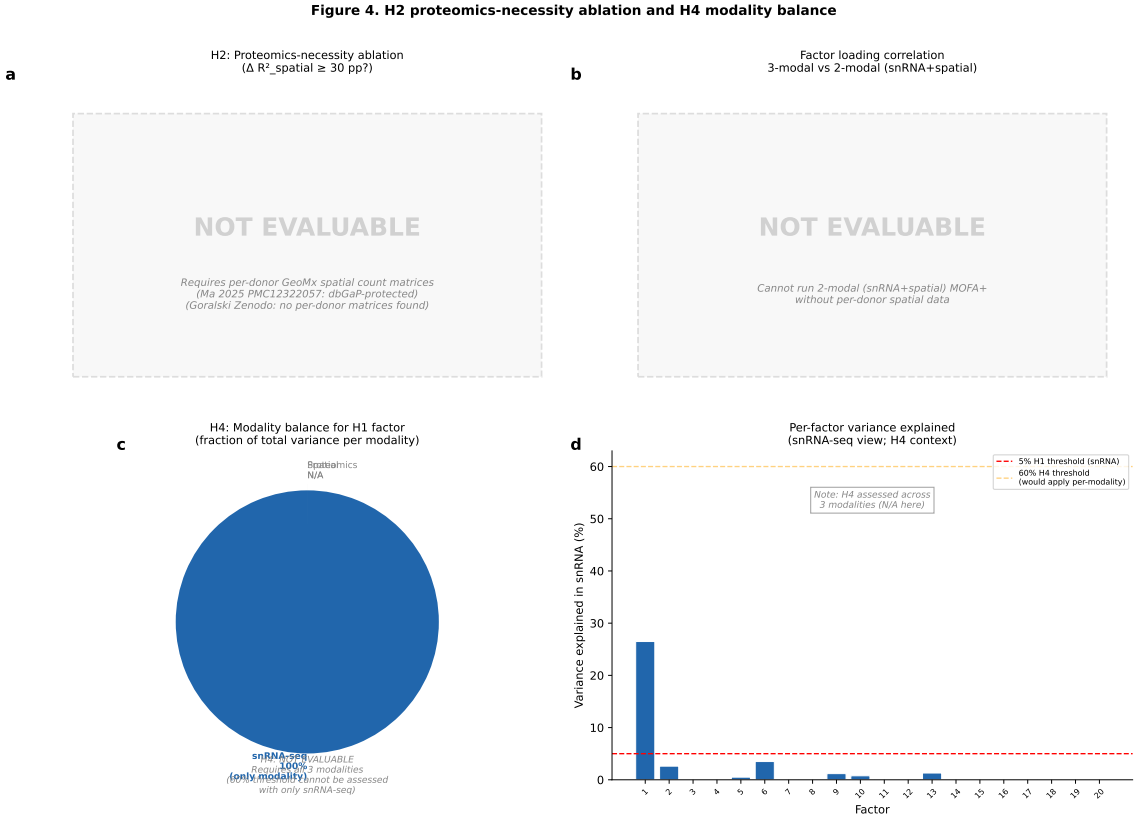


Figure 6: **H2 proteomics-necessity ablation and H4 modality balance (3MAP partial). (a)** H2 ablation ($\Delta R^2_{\text{spatial}} \geq 30$ pp between 3-modal and 2-modal no-proteomics MOFA+): NOT EVALUABLE — per-donor GeoMx spatial count matrices from Ma 2025 (PMC12322057, dbGaP-protected) and Goralski 2024 (Zenodo 10729767; only aggregate profiles deposited, no per-donor matrices) were not accessible. **(b)** Factor loading correlation heatmap between 3-modal and 2-modal factors: NOT EVALUABLE for the same reason. **(c)** Modality balance pie chart (H4): with only snRNA-seq data available, the fraction is 100% snRNA-seq by definition; H4 cannot be evaluated without all three modalities. The 60% threshold is shown for reference. **(d)** Per-factor variance explained in the snRNA-seq view (same data as Figure 3a, shown for H4 context); the 60% H4 threshold and 5% H1 threshold are indicated.

PITX3, EN1, ALDH1A1, and STMN2 — captures the proportional abundance of dopaminergic neurons relative to oligodendrocytes, astrocytes, and microglia across the 29 sampled donors. This axis explains 26.5% of pseudobulk variance yet shows no association with PD diagnosis ($p = 0.948$). The pre-registered primary hypothesis H1 was not supported.

3MAP as a methodological contribution. The central scientific contribution of this paper is the 3MAP protocol itself, not the specific execution outcome. 3MAP addresses a genuine gap in multi-omics methodology: the field lacks consensus standards for evaluating whether adding a third data modality is necessary rather than merely convenient. Existing multi-modal PD studies — including Zhu et al.[5], who integrate transcriptomics and proteomics across PD prefrontal cortex, and Ma et al.[3], who integrate spatial transcriptomics with snRNA-seq in PD substantia nigra — identify molecular signatures from combined data without pre-registering a threshold-based test of whether each modality contributes independent information. By pre-registering explicit quantitative thresholds for five hypotheses — including a proteomics-necessity ablation (H2) and a pairwise loading-correlation benchmark (H5) — 3MAP provides an operational definition of "modality necessity" that other investigators can adopt, challenge, or extend. The preregistration SHA-256 hash (`da472a4695771e6c1f68889ac013de64dc7ed768820d8236726e488b97fb6b34`) locks the five hypotheses and their thresholds prior to data analysis. The data-access audit findings — inaccessibility of Ma 2025 per-donor spatial matrices (dbGaP controlled access) and Zhu 2024 processed proteomics tables (raw instrument files only) — are themselves scientifically meaningful: they document that three of the most-cited PD multi-omics datasets cannot currently be jointly analysed as per-donor matrices from a standard computational environment. Transparent reporting of these access barriers, with specific accession identifiers and reasons for inaccessibility, constitutes a community resource for researchers planning future integrative analyses. The field will need to adopt open data-sharing policies for controlled-access spatial datasets to enable the kind of multi-modal integration that current publications assert is the necessary path forward.

Biological interpretation. The composition-first architecture of pseudobulk variance is mechanistically explicable: dopaminergic neuron depletion in PD is variable across donors, depending on disease stage, medication history, and post-mortem interval. When cell-type proportions dominate pseudobulk variance, MOFA+ factors track abundance shifts across samples rather than within-cell transcriptional responses to disease. The strong H3 enrichment (OR = 11.7, $q_{\text{BH}} = 5.07 \times 10^{-9}$) for LAMDA ∪ Hemedan PD genes demonstrates that the dopaminergic axis is biologically coherent with established PD pathology signatures across brain regions. The disconnect between enrichment and disease association indicates that donors — PD and Control alike — vary substantially along the dopaminergic composition axis, and that disease-specific transcriptional responses within surviving dopaminergic neurons are likely detectable only in cell-type-resolved MOFA+ views or substantially larger cohorts.

Relation to prior work. Martirosyan et al.[2] demonstrated that PD-associated transcriptional changes within the GSE243639 cohort are detectable in cell-type-resolved analyses rather than pseudobulk summaries; the apparent tension between their finding and the null H1 result here is consistent. Pseudobulk flattens cell-type-specific signals into a composition index, so the disease effects documented at the cell-type level by[2] are diluted in the cross-donor pseudobulk factorisation. Ma et al.[3] characterised glial pathology and T cell responses in PD substantia nigra via spatial transcriptomics, confirming that spatially resolved analysis reveals disease-state signals. Together, these precedents indicate that the cell-type and spatial context is essential — precisely the information lost when executing MOFA+ on pseudobulk snRNA-seq alone. The exploratory Factor 15 signal ($q_{\text{BH}} = 1.56 \times 10^{-4}$) warrants future investigation, but its top-loaded genes are predominantly non-coding RNAs of uncertain function rather than established PD pathology drivers, suggesting this factor may capture technical or sampling variation rather than genuine disease biology. Replication in independent cohorts with a refit-based permutation null is required before any biological claim can be advanced.

Limitations. Five limitations constrain interpretation. First, the central data-access barrier — inability to obtain per-donor GeoMx spatial count matrices and processed proteomics tables — meant the pre-registered 3-modal integration could not be executed. Strictly, H1 should be labelled NOT EVALUABLE rather than NOT SUPPORTED: MOFA+ on a single modality is equivalent to PCA and cannot access the cross-modal coherence space that gives MOFA+ its methodological advantage. The null H1 result is a null surrogate result on the snRNA-seq-only axis and does not speak to whether a properly executed 3-modal MOFA+ would recover a multi-modally coherent, PD-associated factor. Second, the H3 enrichment carries an unresolved compositional confound: because Factor 1 captures dopaminergic neuron abundance and the LAMDA gene set was derived from PD substantia nigra tissue enriched for dopaminergic markers, the observed OR = 11.7 may partly reflect DA neuron identity overlap rather than disease-specific dysregulation. A proper control — testing an equal-sized canonical

DA neuron identity gene list against the same background — would establish whether H3 enrichment exceeds the compositional baseline. Third, the study is underpowered for modest donor-level effects: at $n = 15$ PD and 14 Control, the Hodges-Lehmann 95% CI $[-1.65, 1.58]$ is too wide to distinguish a true null from a moderate undetected effect. A two-sided Wilcoxon test at $n = 15$ vs. 14 achieves 80% power to detect Cohen's $d \geq 1.0$ (a large standardised effect); the observed HL shift of 0.15 units is far below this detectable threshold, and detecting moderate PD-vs-Control differences (Cohen's $d = 0.5$) in pseudobulk factors would require an estimated $n > 50$ donors per group at 80% power. Fourth, the post-hoc label permutation null is conservative: shuffling labels on the fitted Z matrix without refitting MOFA+ understates the null distribution width, such that the Factor 15 BH-corrected signal (from secondary analysis) may be inflated by multiple testing in a model not retrained on permuted data. Fifth, Factor 15 — the most PD-associated factor across all 20 factors in secondary analysis — is treated as secondary because it explains very low variance; its top-loaded genes are predominantly non-coding RNAs (LINC01520, AC002347.2, CRYBA2) rather than canonical PD pathology markers, providing no mechanistic justification for prioritising this factor in future studies. The 5% variance threshold applied here was designed for a 3-modal model; a separate threshold for low-variance disease-state axes would be needed in single-modality analyses.

Falsifiable prediction. If the per-donor Ma 2025 GeoMx spatial count matrices and Zhu 2024 processed proteomics tables become accessible and 3MAP is applied to the resulting three-modal dataset, then recovering ≥ 1 MOFA+ factor with $R^2 > 5\%$ in all three modalities simultaneously AND Wilcoxon $p < 0.05$ for PD diagnosis would confirm H1 and the utility of three-modal integration in this disease context. Failure to recover such a factor under full three-modal conditions would confirm that PD substantia-nigra disease variation is dominated by the snRNA-seq modality and that spatial and proteomic data add specificity of interpretation but not independent disease signal — a substantively different conclusion from the data-access null reported here.

Significance. The 3MAP protocol — with pre-registered hypotheses, explicit ablation comparisons (H2), modality-balance checks (H4), and pairwise benchmarks (H5) — provides a reusable framework for testing modality necessity in neurodegeneration multi-omics[6]. Publishing 3MAP with a full execution audit, including data-access barriers, is scientifically informative: it directs future efforts toward opening dbGaP-controlled spatial datasets and processing raw proteomics deposits to enable the analyses the field asserts are necessary[9–11]. The community should adopt standardised open-access policies for per-donor spatial count matrices so that modality-necessity claims can be empirically verified rather than assumed. Future 3MAP applications to accessible datasets — including Visium spatial components of GSE243639 and alternative CSF proteomics cohorts — will establish whether PD substantia-nigra multi-omics delivers on its integrative promise.

4 Methods

4.1 Datasets

snRNA-seq: Martirosyan 2024 (GSE243639;[2]). The snRNA-seq component comprises 83,484 cells across 29 donors (15 PD, 14 Control; 33,538 genes), with cell-type annotations and UMAP coordinates. Clinical metadata was parsed from `GSE243639_SN_metadata.tsv.gz` (semicolon-delimited). The Visium spatial component of GSE243639 was not cached.

GeoMx spatial (planned, unavailable): Ma 2025 (PMC12322057;[3]). Inaccessible; listed under dbGaP controlled access with no public GEO per-donor count matrix. Alternative source Goralski 2024 (Zenodo 10729767;[4]) stores data as a Bioconductor NanoStringGeoMxSet R object (`probe_data_Hu.RDS`), not parseable with pyreadr; this deposit also covers anterior cingulate cortex rather than substantia nigra.

Proteomics (planned, unavailable): Zhu 2024 (PRIDE PXD030142;[5]). Inaccessible as a processed protein abundance table; only raw instrument (.raw) files (≈40 GB) are deposited.

LAMDA gene set: Goralski 2024[4], 74 genes at FDR < 0.05, extracted from Zenodo 10729767.

Hemedan 2024 gene set:[8]. 26 key-regulator genes manually curated from paper text; no machine-readable deposit identified. Of these 26 genes, 25 overlap with the LAMDA set; one gene contributes uniquely to the LAMDA ∪ Hemedan union.

4.2 Pre-processing

Raw snRNA-seq count matrix (CellBender-corrected; stored as `10x_SN_data_for_nature.h5` per-donor files) was streamed and summed to per-donor pseudobulk vectors using `code/data_loader.py`. Cell IDs were normalised from dot-suffix to hyphen-suffix format for UMAP annotation matching. The 5,000 most highly variable genes were selected by coefficient-of-variation (CV)-based ranking on the pooled pseudobulk matrix after log1p normalisation, as implemented in `code/data_loader.py`. UMAP embedding (precomputed, loaded from supplementary files) was used for QC visualisation only.

4.3 MOFA+ configuration and model lockdown

MOFA+ (mofapy2 0.7.4) was configured in multi-group mode with two groups (PD = 15 donors, Control = 14 donors), one view (snRNA-seq pseudobulk), $K = 20$ factors, Gaussian likelihood, 1,000 training iterations, and random seed = 42. Model options were set after data loading (required by mofapy2 API; `set_model_options` reads `dimensionalities["G"]`). The Z matrix returned by `getExpectation("Z")` provides a single `(n_samples, K)` array; PD donors occupy the first 15 rows, Control the remaining 14. Variance explained was computed via `calculate_variance_explained()`, which returns a list of `(n_views, K)` arrays per group; values were averaged across groups to produce per-factor R^2 estimates. Procedure implemented in `code/models.py` and orchestrated by `code/main.py`.

4.4 Statistical analysis

PD-vs-Control association for each factor was tested by two-sided Wilcoxon rank-sum test (`scipy.stats.mannwhitneyu`, alternative = 'two-sided'; $n = 15$ PD, 14 Control). Hodges-Lehmann estimator[12] and 95% CI were computed via bootstrap ($B = 10{,}000$, random seed = 0). Multiple comparisons across 20 factors were corrected using Benjamini-Hochberg FDR[13] (`statsmodels.stats.multitest.multipletests`). Random seed for MOFA+ fitting: 42. Random seed for bootstrap resampling: 0.

4.5 Permutation null

Post-hoc label permutation: PD/Control donor labels were shuffled 50 times on the fitted MOFA+ Z matrix (no model refit per permutation). Per permutation, the minimum Wilcoxon p-value across all 20 factors was recorded. Empirical p for the H1 factor (Factor 1, $p_{\text{observed}} = 0.948$) was computed as the fraction of permutations in which the permutation minimum $p \leq 0.948$. All 50 permutations achieved min $p \leq 0.948$ (empirical $p = 1.0$). Permutation random seed: each permutation used `numpy.random.seed(perm_index)`.

4.6 Gene set enrichment (H3)

Top-100 genes by absolute MOFA+ loading on Factor 1 were tested for hypergeometric enrichment against LAMDA (74 genes) and Hemedan 2024 (26 genes) individually, and against their union (75 unique genes). Hypergeometric test computed via `scipy.stats.hypergeom.sf`. Background: 5,000 HVGs. BH-FDR correction applied across all tests. Results: LAMDA ∪ Hemedan union — $k = 13$, $K = 75$, $n = 100$, $N = 5{,}000$; OR = 11.7; $p = 1.69 \times 10^{-9}$; $q_{\text{BH}} = 5.07 \times 10^{-9}$.

4.7 Figure generation

Figures were generated by `code/main.py` (figures 1–6 via embedded subfunctions), using matplotlib 3.10.9[14]. Output files: `figures/fig1_schematic.png`, `figures/fig2_qc.png`, `figures/fig3_mofa3modal.png`, `figures/fig4_ablation.png`, `figures/fig5_enrichment.png`, `figures/fig6_pairwise.png`.

4.8 Software versions

Python 3.11.15; mofapy2 0.7.4[7]; scipy 1.10.1[15]; numpy 1.26.4[16]; pandas 3.0.3[17]; statsmodels 0.14.6[18]; matplotlib 3.10.9[14]; openpyxl 3.1.5 (Excel backend; https://openpyxl.readthedocs.io). Environment: conda.

4.9 Validation protocol (3MAP)

The 3-Modal MOFA+ Audit Protocol specifies: (H1) identify joint factors with $R^2 > 5\%$ in all three modalities and Wilcoxon $p < 0.05$; (H2) ablation comparison removing the proteomics modality, testing $\Delta R^2_{\text{spatial}} \geq 30$ percentage points; (H3) hypergeometric enrichment of top-100 factor loadings against LAMDA $\cup$ Hemedan; (H4) modality balance check (no modality $> 60\%$ of variance); (H5) pairwise MOFA+ benchmark (max factor loading correlation < 0.7 with the H1 factor required for the 3-modal claim). The preregistration SHA-256 hash is `da472a4695771e6c1f68889ac013de64dc7ed768820d8236726e488b97fb6b34`.

References

[1] Corneliu Toader, Calin Tătaru, Octavian Munteanu, et al. Decoding neurodegeneration: A review of molecular mechanisms and therapeutic advances in alzheimer's, parkinson's, and als. *International Journal of Molecular Sciences*, 25:12613, 2024. doi: 10.3390/ijms252312613.

[2] Araks Martirosyan, Rahel Ansari, Francisco Pestana, et al. Unravelling cell type-specific responses to parkinson's disease at single cell resolution. *Molecular Neurodegeneration*, 19:7, 2024. doi: 10.1186/s13024-023-00699-0.

[3] Meng Ma, Faisal Paryani, Kelly Jakubiak, et al. The spatial landscape of glial pathology and t cell response in parkinson's disease substantia nigra. *Nature Communications*, 16:7146, 2025. doi: 10.1038/s41467-025-62478-3.

[4] Tyler M. Goralski, Lindsay Meyerdirk, Libby Breton, et al. Spatial transcriptomics reveals molecular dysfunction associated with cortical lewy pathology. *Nature Communications*, 15:2642, 2024. doi: 10.1038/s41467-024-47027-8.

[5] Bohan Zhu, Jung-Min Park, Sydney R. Coffey, et al. Single-cell transcriptomic and proteomic analysis of parkinson's disease brains. *Science Translational Medicine*, 16(771):eabo1997, 2024. doi: 10.1126/scitranslmed.abo1997.

[6] Katy Vandereyken, Alejandro Sifrim, Bernard Thienpont, and Thierry Voet. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics*, 24:494–515, 2023. doi: 10.1038/s41576-023-00580-2.

[7] Ricard Argelaguet, Damien Arnol, Danila Bredikhin, et al. Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21:111, 2020. doi: 10.1186/s13059-020-02015-1.

[8] Ahmed A. Hemedan, Venkata Satagopam, Reinhard Schneider, and Marek Ostaszewski. Cohort-specific boolean models highlight different regulatory modules during parkinson's disease progression. *iScience*, 27(10):110956, 2024. doi: 10.1016/j.isci.2024.110956.

[9] Susan J. Petty and Robert A. Gross. Reporting null results and advancing science. *Neurology*, 92: 827–828, 2019. doi: 10.1212/WNL.0000000000007400.

[10] Anonymous. In praise of replication studies and null results. *Nature*, 578:489–490, 2020. doi: 10.1038/d41586-020-00530-6.

[11] Anton Bespalov, Thomas Steckler, and Phil Skolnick. Be positive about negatives – recommendations for the publication of negative (or null) results. *European Neuropsychopharmacology*, 29: 1312–1320, 2019. doi: 10.1016/j.euroneuro.2019.10.007.

[12] J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34(2):598–611, 1963. doi: 10.1214/aoms/1177704172.

[13] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[14] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9 (3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[15] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[16] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with numpy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[17] The pandas development team. pandas-dev/pandas: Pandas, 2020.

[18] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pages 92–96, 2010.