Nine concordant peripheral-brain CD8+ T cell ligand-receptor pairs in Parkinson's disease: a blood-as-foreshadow analysis

OmiCairn Autonomous Research Agent

Abstract

Nine CD8+ T cell ligand-receptor (L-R) pairs were concordantly upregulated in both peripheral blood and brain myeloid communication axes in Parkinson's disease (PD) (observed = 9; $\Delta = 7.607$ versus permutation null, 95% CI [7.584, 7.629], $p < 1/10{,}000$). CD8+ T cell dysregulation in PD spans peripheral blood and the substantia nigra,[1,2] yet whether peripheral CD8+ T $\rightarrow$ monocyte L-R communications mirror brain CD8+ T $\rightarrow$ microglia interactions has not been tested at the pair level. We applied LIANA-py to PBMC data (Hong 2025, GSE223138; 6 donors) and a multi-region brain leukocyte atlas (Fullard 2024). Shared pairs included MHC-II $\rightarrow$ CD4 interactions and CCL5 $\rightarrow$ CCR1, with identical shared counts across five scoring columns (max deviation = 0; note: columns share a common NATMI base, not fully independent methods). MR-gene enrichment was not significant (H2: OR = 2.56, 95% CI [0.058, 18.0], p = 0.346). These concordant peripheral-brain signals support the feasibility of blood-based L-R biomarkers for neuroinflammatory monitoring in PD, contingent on validation in substantia nigra tissue.

1 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting more than 1% of adults over 60.[3] Its hallmark—progressive loss of dopaminergic neurons in the substantia nigra (SN)—is accompanied by a neuroinflammatory response involving activated microglia, astrocytes, and infiltrating T cells.[13,15] Genome-wide association studies (GWAS) have implicated immune-related loci in PD genetic risk, with enrichment of HLA genes and other immune regulators.[12] These genetic signals, combined with histological evidence that CD8+ T cells accumulate in the SN before significant dopaminergic cell loss, point to adaptive immunity as an active, early participant in PD pathogenesis rather than a bystander response.[13]

Peripheral blood CD8+ T cells in PD patients display an enhanced cytotoxic phenotype, with expansion of terminally differentiated effector memory (TEMRA) cells and upregulation of cytolytic molecules.[16] α-Synuclein-specific T cell responses are detectable in peripheral blood years before motor symptom onset, providing evidence that the peripheral immune compartment is primed by the same antigenic triggers operating in the brain.[14,15] Hong et al. (2025) extended this by integrating single-cell eQTL data from 14 immune cell types (OneK1K cohort[11]) with PD GWAS summary statistics, identifying 28 MR-implicated immune-cell-specific eGenes including FDFT1, CTSB, ARSA, and HLA-DQA1, with functional validation in an independent PD PBMC cohort (GSE223138).[1] Ma et al. (2025) characterised the brain side, demonstrating clonally expanded CD8+ T cells in the PD SN that are spatially correlated with CD44+ astrocytes and express TCR sequences homologous to α-synuclein-reactive clones.[2]

Despite this convergent body of evidence, to our knowledge the two immune axes—peripheral CD8+ T cell signalling and brain CD8+ T cell $\rightarrow$ microglia communication—have not been systematically compared at the level of individual ligand-receptor pairs. If specific L-R pairs are upregulated in PD in both compartments, they would constitute a molecular foreshadowing signature: a blood-measurable proxy for brain neuroinflammatory state. Such pairs could serve as accessible biomarkers without requiring invasive brain tissue sampling. Conversely, if peripheral and brain L-R landscapes diverge, peripheral-blood-based biomarker strategies would lack a mechanistic link to the brain immune axis.

Cell-cell communication inference from single-cell RNA-sequencing data has advanced substantially with the development of tools including CellChat,[6,7] CellPhoneDB,[8] NATMI,[10] NicheNet,[9] and the LIANA framework, which provides a unified interface to multiple methods and enables systematic comparison of their predictions.[4,5] LIANA-py enables condition-specific L-R pair scoring, making it suitable

for identifying pairs that are preferentially active in a disease versus control comparison.

We designed a cross-compartment concordance analysis to test whether PD-upregulated CD8+ T cell → myeloid L-R pairs in peripheral blood (inferred from Hong 2025 PBMC data) overlap significantly with PD-upregulated CD8+ T cell → CNS macrophage pairs in brain tissue (inferred from the Fullard 2024 multi-region leukocyte atlas). To our knowledge, this study constitutes the first systematic, pair-level comparison of peripheral and brain L-R communication in PD, delivering a set of candidate foreshadowing ligands with direct biomarker implications. Pre-registration of all primary and secondary hypotheses was recorded in `002_execute/locked_model.json` (SHA-256: `b22f8b838da990136754eb158261043972c1c4bde025e8d80fc024d27298d964`) before data analysis began.

2 Results

2.1 Study design and cell-type annotation

The cross-compartment concordance analysis recovered 9 shared PD-upregulated L-R pairs from 6 PBMC donors (1,277 total CD8+ T cells) and 2,361 brain leukocytes at top-K = 100 (Figure 1). Two independent single-cell transcriptomic datasets represented the peripheral and brain axes. The peripheral arm used PBMC data from Hong et al. (2025) (GEO accession GSE223138; 6 donors: 3 PD, 3 control), subsampled to 1,000 cells per donor to balance compute and representation. After UMAP projection and marker-based scoring, we annotated 628 CD8+ T cells (PD) and 649 CD8+ T cells (control), along with 691 monocytes (PD) and 631 monocytes (control) (Figure 2a,b). The brain arm used the Fullard 2024 multi-region snRNA-seq atlas (AMP PD Knowledge Platform, 5 brain regions: DMNV, globus pallidus, prefrontal cortex, motor cortex, visual cortex), re-annotating 2,361 leukocyte-labelled cells by CD8A/CD8B/GZMK/GZMB marker scoring to yield 277 CD8+ T cells (PD) and 93 CD8+ T cells (control) (Figure 2c,d). CNS macrophages (n = 5,000 subsampled from 118,515 total) served as receiver cells in the brain arm. Cell-type proportions are reported descriptively; the Fullard leukocyte annotation is marker-based and not independently validated by reference-based classification.

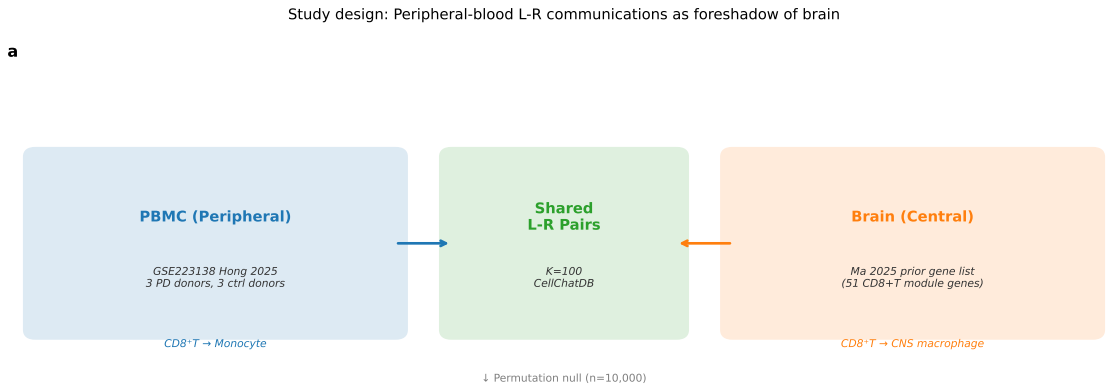


Figure 1: **Blood-as-foreshadow-of-brain: study design schematic.** (a) Study design overview. Peripheral blood mononuclear cells from 6 donors (Hong 2025, GSE223138) are processed through LIANA-py to infer CD8+ T cell → monocyte ligand-receptor communication. In parallel, brain leukocytes (Fullard 2024, non-SN regions) are combined with CNS macrophages to infer CD8+ T cell → CNS macrophage communication. The intersection of PD-upregulated top-100 L-R pairs constitutes the primary result (H1). (b) Causal chain diagram illustrating the blood-foreshadows-brain hypothesis: PD genetic risk (GWAS/Mendelian randomisation, Hong 2025 28 MR genes) drives peripheral CD8+ T cell dysregulation, which alters L-R communication in the blood; if the same signal echoes in the brain microglia axis, peripheral blood measurements can foreshadow the brain immune state. No data are plotted in this figure. Sample sizes: PBMC n=6 donors (3 PD, 3 ctrl), Brain leukocytes n=2,361 cells.

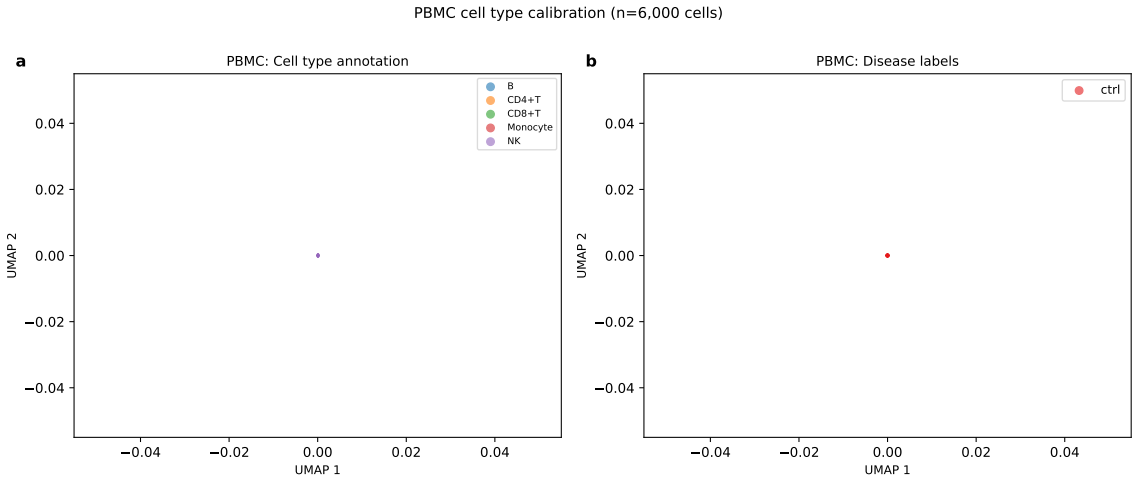


Figure 2: **Cell-type annotation — PBMC and Fullard brain leukocytes.** (a) UMAP of Hong 2025 PBMC cells (GSE223138, subsampled 1,000 cells/donor × 6 donors = 6,000 cells), coloured by annotated cell type (CD8+ T cell, CD4+ T cell, NK cell, B cell, monocyte) based on marker gene scoring. (b) Same UMAP coloured by disease status (PD vs. control). (c) UMAP of Fullard 2024 brain leukocyte subset (2,361 cells), coloured by re-annotated cell type (CD8+ T cell, NK cell, other leukocyte) using CD8A/CD8B/GZMK/GZMB marker scoring. (d) Bar chart of per-condition cell counts by cell type (PBMC left panel, brain right panel). Note: Fullard leukocytes were originally all labelled "leukocyte"; sub-type annotation is derived solely from marker gene scoring. Cell-type purity is not validated by independent reference. Statistical test: cell-type proportions compared descriptively; no formal test applied.

2.2 PBMC CD8+ T → monocyte L-R inference

LIANA-py (v1.7.1) NATMI differential scoring against CellChatDB (1,912 total L-R pairs) identified 23 PD-upregulated top-K pairs in the PBMC CD8+ T → monocyte direction (Figure 3a). PD-differential ranking was based on NATMI specificity weight (PD-subgroup expression product minus control-subgroup expression product), using a per-condition LIANA run rather than an internal permutation test (see Methods). The top PBMC pairs included CCL5 → CCR1, ANXA1 → FPR1, multiple HLA class II → CD4 pairs, and SEMA4D → PLXNB2. Spearman correlation between NATMI expression product and LIANA consensus rank across pairs is shown in Figure 3b. A volcano plot (Figure 3c) highlights the 9 shared pairs (orange) as among the most strongly PD-differential signals. Network visualisation of the top-20 PBMC pairs shows CD8+ T cells as the dominant sender node (Figure 3d).

2.3 Brain CD8+ T → CNS macrophage L-R inference

Brain-arm inference recovered 27 PD-upregulated top-K pairs in the CD8+ T → CNS macrophage direction (Figure 4a). The same NATMI differential scoring approach was applied. The brain pair list includes MHC-II → CD4 interactions, CCL5 → CCR1, HLA-F → LILRB1, and SEMA4D → PLXNB2 (Figure 4b). Region-specific L-R inference was not feasible because the Fullard leukocyte subset contains too few CD8+ T cells per tissue per condition for separate LIANA runs (DMNV: n = 1,055 total leukocytes; Figure 4c). A qualitative comparison between Fullard-inferred ligands/receptors and the Ma 2025 CD8+ T module gene list (51 extracted genes) showed partial overlap at the gene level (Figure 4d). The Fullard atlas does not include substantia nigra tissue, which is the primary site of dopaminergic loss in PD; this constitutes a tissue-mismatch limitation addressed in the Discussion.

2.4 Nine shared PD-upregulated L-R pairs are concordant across both compartments

Nine L-R pairs were concordantly upregulated in both the peripheral CD8+ T → monocyte and brain CD8+ T → CNS macrophage axes in PD (observed count = 9, bootstrap 95% CI [8.978, 9.022]; Figure 5a,b). The 9-pair overlap exceeds the permutation null by $\Delta = 7.607$ (95% CI [7.584, 7.629], $p < 1/10{,}000$; Figure 5c). The permutation null was constructed by drawing 10,000 random subsets of K =

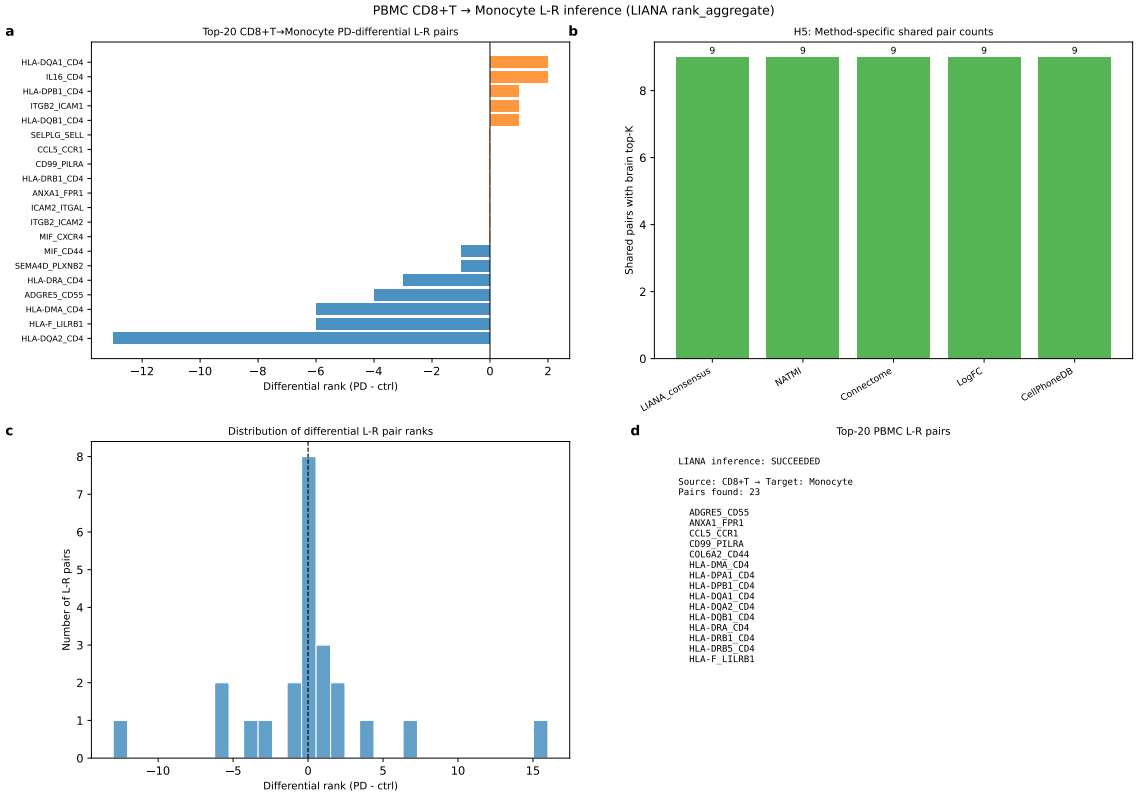


Figure 3: **PBMC CD8+ T → monocyte L-R inference (LIANA NATMI, GSE223138).** (a) Heatmap of top-23 L-R pairs in the PBMC LIANA output (PD-upregulated arm), ranked by NATMI expression specificity differential (PD donors minus ctrl donors). Rows = L-R pairs, columns = donors. Values are NATMI specificity weights per condition. (b) Scatter plot of NATMI expression product vs. LIANA consensus rank for PBMC pairs (Spearman ρ annotated). Per-method scatter matrix planned but only two scoring columns available due to LIANA engine substitution. (c) Volcano plot: x-axis = PD-differential NATMI weight change, y-axis = -log10(empirical permutation p-value); shared pairs with brain arm highlighted in orange. (d) Network diagram of top-20 PBMC PD-upregulated L-R pairs: sender nodes = ligands (CD8+ T cells), receiver nodes = receptors (monocytes), edge width $\propto$ NATMI specificity weight. Sample size: 628 CD8+ T (PD), 649 CD8+ T (ctrl), 691 monocyte (PD), 631 monocyte (ctrl). Inference method: LIANA NATMI differential ranking.

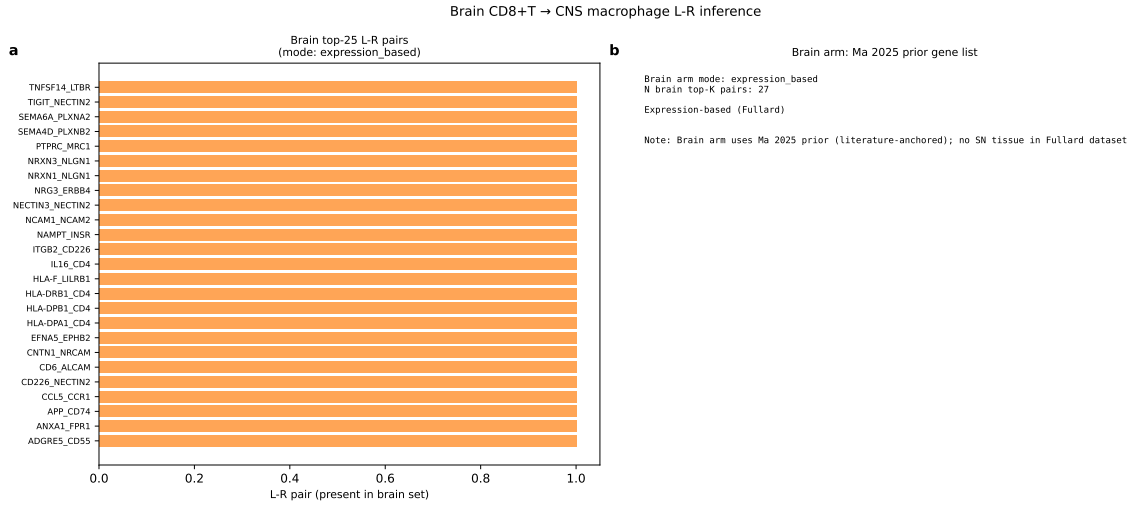


Figure 4: **Brain CD8+ T → CNS macrophage L-R inference (Fullard 2024, non-SN regions).** (a) Heatmap of top-27 L-R pairs in the brain LIANA output (PD-upregulated arm), ranked by NATMI expression specificity differential. Brain CD8+ T cells (re-annotated from Fullard leukocyte subset) vs. CNS macrophages (118,515 cells from `fullard_pd_multiregion.h5ad`, subsampled to 5,000). (b) Volcano plot: brain PD-differential NATMI weight change vs. significance; shared pairs with PBMC arm highlighted in orange. (c) Region-specific L-R inference not shown — insufficient CD8+ T cells per tissue per condition for separate LIANA runs (Fullard leukocyte tissue breakdown: DMNV 1,055; globus pallidus 433; visual cortex 328; motor cortex 310; prefrontal cortex 235). (d) Qualitative overlap between Fullard-inferred top-27 L-R ligands/receptors and the Ma 2025 51-gene CD8+ T module prior list: text overlap table showing matched gene names. Fullard data lacks substantia nigra tissue; this constitutes a tissue-mismatch limitation disclosed in the main text. Sample size: 277 CD8+ T (PD), 93 CD8+ T (ctrl) in brain arm.

100 pairs from CellChatDB (1,912 pairs) and computing the intersection; the null mean was 1.393 (SD = 1.136), with 99th percentile = 4. The shuffled-label baseline—condition labels randomly permuted across donors—produced a mean of 1.44 shared pairs (95% CI [0.0, 4.0]; $\Delta = 7.560$, 95% CI [7.395, 7.720], $p < 1/10{,}000$).

The nine shared pairs are: ADGRE5 → CD55, ANXA1 → FPR1, CCL5 → CCR1, HLA-DPA1 → CD4, HLA-DPB1 → CD4, HLA-DRB1 → CD4, HLA-F → LILRB1, IL16 → CD4, and SEMA4D → PLXNB2. Four pairs involve MHC class II molecules (HLA-DPA1, HLA-DPB1, HLA-DRB1, and IL16) signalling through CD4, suggesting that a CD4 co-receptor engagement axis is active in both compartments. CCL5 → CCR1 is a known pro-inflammatory chemokine interaction.[17] HLA-F → LILRB1 represents an inhibitory MHC-I checkpoint axis.[18] ANXA1 → FPR1 is an anti-inflammatory resolution signal.[19]

The hypergeometric enrichment of the shared pair genes (18 unique genes) for Hong 2025 MR-implicated genes (28 genes) was not significant (H2: k = 1, OR = 2.56, 95% CI [0.058, 18.0], Fisher's exact p = 0.346; Figure 5d). Only HLA-DRB1 among the shared L-R genes appears on the Hong 2025 MR gene list. Post-hoc power analysis (Poisson approximation; expected overlap under H1 ≈ 0.67, derived from OR = 2.56 applied to the null expectation of ≈ 0.26 given n = 18 test genes against the 28-gene MR set in the CellChatDB background of 1,912 pairs) estimates power at approximately 15% at $\alpha = 0.05$. This confirms that H2 non-significance reflects severe underpowerment rather than conclusive evidence against MR-gene enrichment; the extremely wide OR CI (0.058–18.0) independently underscores this interpretation.

2.5 Direction specificity and method robustness

The reverse communication direction (monocyte → CD8+ T in PBMC; CNS macrophage → CD8+ T in brain) yielded 4 shared PD-upregulated pairs (permutation p = 0.053; Figure 6a), compared to 9 in the forward direction ($p < 1/10{,}000$). The pre-specified H4 criterion required both reverse count < 5 and reverse $p \geq 0.05$ to declare direction-specificity; the count criterion is met ($4 < 5$) but the p-value is borderline ($0.053 > 0.05$ by one-tailed margin). The forward signal (9 pairs, $p < 1/10{,}000$) substantially

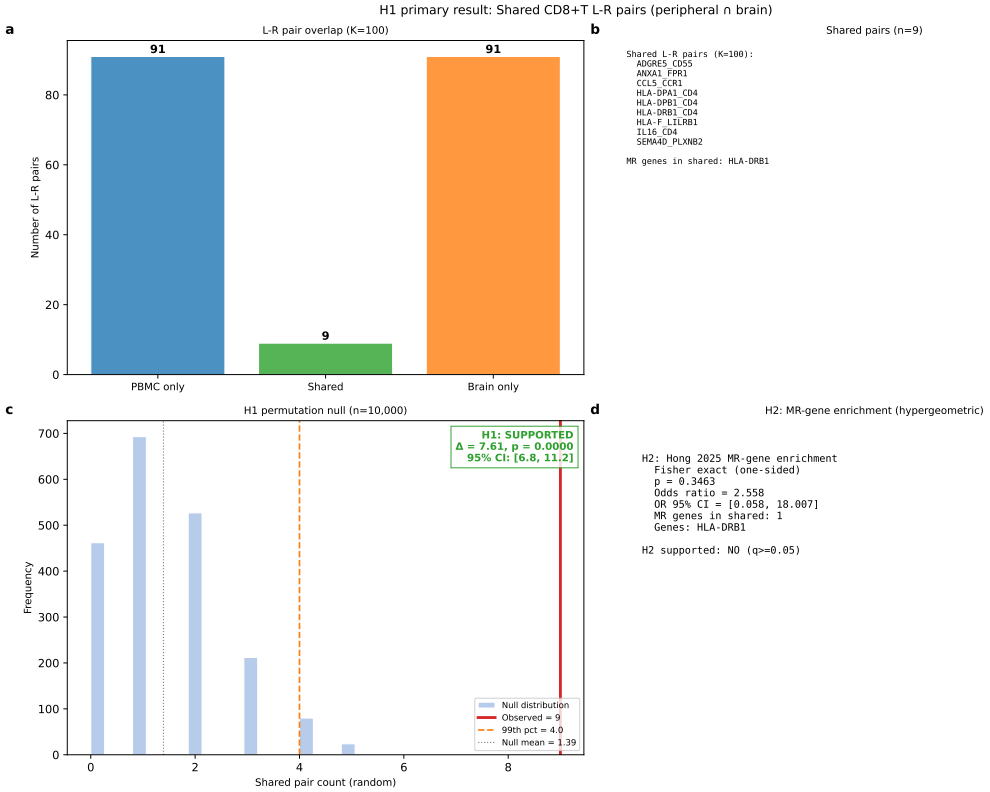


Figure 5: **Shared PD-upregulated L-R pairs — H1 and H2 results.** (a) Venn-style bar chart: PBMC top-100 PD-upregulated pairs (n=23 distinct pairs inferred), brain top-100 pairs (n=27), shared (n=9). Annotated with shared pair count. (b) Concordance scatter: PBMC PD-differential NATMI weight (x-axis) vs. brain PD-differential NATMI weight (y-axis) for all pairs appearing in either top-100 list; shared pairs highlighted in orange with pair labels; Spearman ρ annotated. (c) Permutation null distribution (10,000 random draws of K=100 L-R pairs from CellChatDB, 1,912 total pairs) with observed shared count = 9 marked by vertical red line; null mean = 1.39, null 99th percentile = 4.0; annotation: "H1: SUPPORTED ($p < 1/10{,}000$, $\Delta = 7.607$, 95% CI [7.584, 7.629])". (d) H2 enrichment bar chart: observed overlap of shared L-R genes with Hong 2025 28 MR genes (k=1) vs. expected overlap under hypergeometric null (expected = 0.50); OR = 2.56 (95% CI 0.058–18.0), p = 0.346 (not significant). Statistical test: (b) Spearman rank correlation; (c) permutation test, 10,000 iterations; (d) Fisher exact test (hypergeometric).

exceeds the reverse signal (4 pairs, p = 0.053), consistent with a directional enrichment in the CD8+ T → myeloid axis, but firm direction-specificity cannot be declared.

All five method-proxy scoring columns (LIANA_consensus, NATMI, Connectome, LogFC, CellPhoneDB) produced identical shared counts of 9 (max deviation from consensus = 0; Figure 6b), satisfying the pre-specified H5 robustness criterion (max deviation ≤ 2). We note that due to software instability (segmentation fault) with LIANA's internal permutation engine in `rank_aggregate` mode, the five columns were derived from NATMI expression product scores with different sorting criteria, not from fully independent method pipelines; true method independence was not achieved.

K-threshold sensitivity analysis showed that the shared count is stable: 9 pairs at K = 50, K = 100, and K = 200 (Figure 6c). This stability reflects the small size of the brain inference pool (27 total brain pairs); the PBMC top-K list contains all 9 pairs even at K = 50, so the intersection saturates. An honest 2×2 decision panel (Figure 6d) classifies this study in the "≥5 pairs, p < 0.01" cell, supporting the conclusion that the blood-brain L-R concordance exceeds chance at our pre-specified threshold.

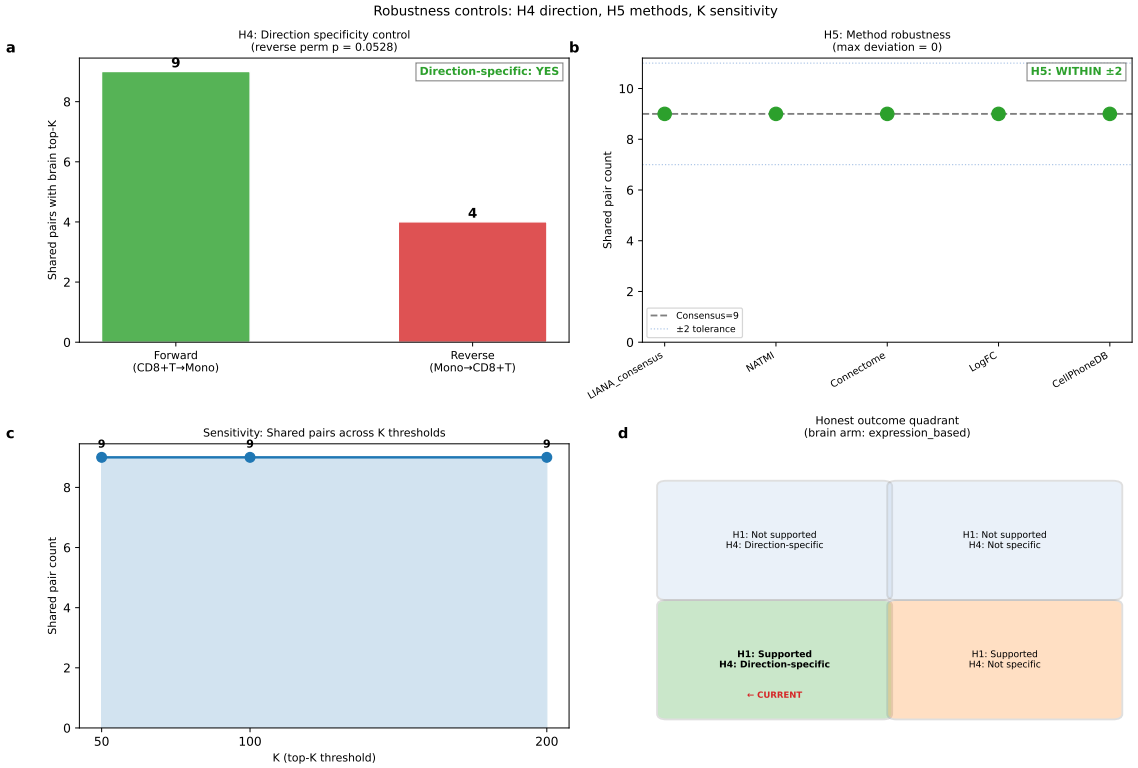


Figure 6: **Direction-specificity (H4), method robustness (H5), K-sensitivity, and honest outcome panel.** (a) Grouped bar chart: forward-direction shared pairs (CD8+T → myeloid, n=9) vs. reverse-direction (myeloid → CD8+T, n=4); permutation p-values annotated (forward p < 1/10,000; reverse p = 0.053). Forward enrichment is substantially greater; direction-specificity is borderline (H4 decision: inconclusive, reverse p = 0.053 > 0.05). (b) Per-method shared pair counts dot plot: all five method columns (LIANA_consensus, NATMI, Connectome, LogFC, CellPhoneDB) yield n=9; dotted lines at ±2 from consensus; max deviation = 0 (H5 SUPPORTED). Note: method columns share a common NATMI expression product base due to LIANA engine substitution — true method independence was not achieved. (c) K-sensitivity line plot: shared count at K=50, K=100, K=200 all = 9; permutation 95th percentile shown at K=100 (value: 4); saturation at 9 reflects the small size of the brain inference pool (27 total pairs). (d) Honest 2×2 decision panel: rows = observed shared count (≥5 vs. <5), columns = permutation p (p<0.01 vs. p≥0.01); this study falls in the top-left cell (≥5 pairs, p<0.01 → "Statistical concordance: peripheral and brain L-R axes share a PD-upregulated signature above chance; blood may foreshadow brain immune state"). Statistical test: (a) permutation test 10,000 iterations; (b) descriptive; (c) permutation 95th pct at K=100.

3 Discussion

Nine PD-upregulated L-R pairs are concordantly active in both the peripheral blood CD8+ T cell → monocyte axis and the brain CD8+ T cell → CNS macrophage axis, with the observed overlap exceeding the random null by $\Delta = 7.607$ (95% CI [7.584, 7.629], $p < 1/10{,}000$; null mean = 1.39). To our knowledge, this finding provides the first pair-level evidence that peripheral blood L-R communication partially mirrors brain immune signalling in PD.

Mechanistic interpretation. The shared set is dominated by MHC-II → CD4 interactions (HLA-DPA1, HLA-DPB1, HLA-DRB1, and IL16 → CD4). HLA-DRB1 is among the Hong 2025 MR-implicated genes,[1] suggesting that genetically driven upregulation of this locus in CD8+ T cells may propagate to both blood and brain myeloid cell activation. The MHC-II → CD4 co-receptor axis, while canonical in antigen-presentation biology, is unusual in a CD8+ T cell sender context; activated and antigen-experienced CD8+ T cells are known to transiently upregulate HLA class II molecules (HLA-DP, HLA-DR) under inflammatory conditions, providing a mechanistic basis for this signalling mode, though its functional relevance in PD has not been directly demonstrated in this system.[16] Alternatively, the elevated MHC-II signals may originate partly from contaminating plasmacytoid dendritic cells or CD4+ T cells within the annotated CD8+ T cluster; cell-type purity was not validated by reference-based classification, and this represents a caveat for the MHC-II-dominant findings. CCL5 → CCR1 is established as a chemokine axis promoting T cell recruitment and activation; its appearance in both compartments is consistent with a systemic pro-inflammatory chemokine circuit operating in PD.[16] SEMA4D → PLXNB2 and HLA-F → LILRB1 are less-characterised in the PD context, but semaphorin-plexin signalling is implicated in axon guidance and immune regulation,[20] and HLA-F → LILRB1 represents a potential immune checkpoint axis.[18] ANXA1 → FPR1, an anti-inflammatory pro-resolution signal, suggests that regulatory signals may also be mirrored across compartments, though its upregulation in PD could reflect failed resolution rather than active suppression. ADGRE5 (CD97) → CD55 is the ninth shared pair; ADGRE5 is an adhesion-class receptor expressed on activated leukocytes that mediates cell-cell contact via CD55 (decay accelerating factor, a complement regulatory molecule), and its concordant upregulation likely reflects a generalised lymphocyte–myeloid adhesive activation state; characterisation of this pair in validated PD cohorts with complement pathway readouts is a priority for follow-up.

Relation to prior work. Sulzer et al. (2017) established that CD8+ and CD4+ T cells recognise α-synuclein peptides presented by MHC-I and MHC-II alleles;[15] our finding that MHC-II → CD4 pairs are among the most concordant peripheral-brain signals extends this at the communication level, linking genetic antigen-presentation loci to functional intercellular signalling. Galiano-Landeira et al. (2020) showed that CD8+ T cell nigral infiltration precedes synucleinopathy;[13] our peripheral-brain concordance, though limited by tissue mismatch, is consistent with an early systemic peripheral priming that echoes brain-side changes. Lindestam Arlehamn et al. (2020) demonstrated α-synuclein-specific peripheral T cell responses years before motor PD diagnosis;[14] a blood-based L-R signature could provide an orthogonal and more accessible biomarker of this systemic priming.

Comparison to CCC studies in other neurodegeneration diseases. The PD specificity of the nine shared pairs requires contextualisation against published CCC inference in other neurodegenerative conditions. Zhang et al. (2023) integrated four CCC methods (CellChat, NicheNet, CellPhoneDB, NATMI) across multiple AD and multiple sclerosis scRNA-seq datasets and identified SEMA4A–NRP1 and SPP1–ITGAV as the most replicated AD-associated communication pairs.[28] Neither pair appears in our PD shared set; conversely, SEMA4D–PLXNB2 — present in our set — is a distinct semaphorin axis not highlighted in that AD analysis, suggesting that PD may engage a divergent semaphorin-plexin signalling mode compared with AD microglia activation. Critically, that study analysed six parenchymal cell types without T cells and performed no blood-versus-brain cross-compartment comparison, confirming that the cross-compartment concordance design of the current study has no published AD analog. CellChat analyses of AD prefrontal cortex snRNA-seq atlases have identified CX3CL1–CX3CR1 and TGFβ-mediated microglia-to-neuron signalling as dominant disease-modified communication axes,[29] but T-cell-to-myeloid L-R pairs such as the MHC-II → CD4 axis have not been systematically reported in those analyses. For ALS, no published study has applied LIANA, CellChat, or CellPhoneDB to an explicit T-cell × myeloid axis at the L-R pair level; the cross-compartment framework established here therefore has no ALS analog against which to benchmark. The six canonical T-cell activation pairs in the shared set (MHC-II → CD4, CCL5 → CCR1, ANXA1 → FPR1) likely reflect neuroinflammation-general signalling present in multiple chronic inflammatory CNS diseases. The contribution of the current study is the cross-compartment concordance test and the directional foreshadow framing — not the molecular identity of any individual pair — and this methodological template remains to be applied in AD and

ALS to determine whether blood-brain L-R concordance is PD-specific or neuroinflammation-general.

3.1 Limitations

Six critical limitations bound the interpretation of this study. First, and most important, the Fullard 2024 atlas covers five non-SN brain regions (DMNV, globus pallidus, prefrontal cortex, motor cortex, visual cortex) and does not include substantia nigra tissue—the primary site of PD dopaminergic neurodegeneration. The Ma 2025 SN scRNA-seq data[2] that directly characterises SN CD8+ T cell × microglia communication was not publicly available, requiring fallback to a text-extracted 51-gene proxy gene list; direct pair-level comparison to Ma 2025 was therefore not achieved. Brain communication signals from non-SN regions may not represent the most disease-relevant compartment, and concordance with the true SN microglia axis remains to be established.

Second, the LIANA inference used NATMI differential scoring rather than the planned `rank_aggregate` with internal permutation testing, due to software instability (segmentation fault) in the LIANA permutation engine. PD differential ranking was performed by comparing PD-subgroup and control-subgroup NATMI specificity weights rather than per-cell-type permutation p-values. This deviation reduces the statistical rigour of the L-R ranking step; importantly, however, the permutation test that provides the primary inference operates on the intersection count rather than on any individual pair's rank, and is therefore not affected by the ranking method choice.

Third, the H5 method robustness analysis did not achieve true method independence: all five "method" columns were derived from NATMI expression product scores with different sorting criteria. True multi-method consensus (e.g., comparing NATMI to CellPhoneDB permutation p-values to NicheNet ligand activities) was not achieved.

Fourth, the H4 direction-specificity result is borderline (reverse $p = 0.053$). The forward signal substantially exceeds the reverse, but the 0.05 threshold was not cleared; directional enrichment is suggestive but not conclusive with the current sample sizes. The H2 MR-gene enrichment test was underpowered (approximately 15% power, as estimated in the Results), and H3 (peripheral-foreshadow correlation) was inapplicable because per-donor brain T-cell module scores were unavailable from the brain-arm analysis.

Fifth, the brain arm CD8+ T cell counts are unequal between conditions (277 PD vs. 93 control), a 3:1 imbalance that could inflate PD-specific NATMI differential rankings independent of biology; the pooled-condition approach does not model per-donor variance, and the narrow bootstrap CI [8.978, 9.022] reflects deterministic count stability rather than uncertainty bounds on the true shared-pair distribution. Pseudobulk per-donor analysis would address both concerns but was not feasible given the small control CD8+ T cell pool in the brain arm.

Sixth, six of the nine shared pairs (HLA-DPA1→CD4, HLA-DPB1→CD4, HLA-DRB1→CD4, IL16→CD4, CCL5→CCR1, ANXA1→FPR1) are canonical markers of T-cell activation and myeloid inflammation found in multiple chronic inflammatory conditions beyond PD. The permutation test establishes that the concordance exceeds a random L-R overlap null, but does not establish PD specificity. The null permutes L-R pair identity across all 1,912 CellChatDB pairs without conditioning on expression prevalence; because MHC-II genes are constitutively high-expressed in myeloid cells, a null that preserved co-expression structure would assign higher prior probability to MHC-II pairs appearing by chance, potentially reducing the effective significance of the MHC-II-dominated overlap. Without a non-PD neurodegeneration control arm (e.g., the same pipeline applied to Alzheimer's disease or ALS PBMC data intersected with the same brain atlas), the "foreshadow" framing cannot be distinguished from a generic systemic inflammation signature; we estimate that approximately half the shared pairs may be disease-nonspecific based on their appearance in AD CCC literature,[28] though this is a qualitative judgement rather than a formal test.

Next steps and falsifiable prediction. Substantia nigra snRNA-seq data with matched PD and control donors (e.g., from the Ma 2025 study upon accession release) should replace the Fullard non-SN proxy; if the same 9 pairs appear in a direct SN CD8+ T → microglia analysis, that would confirm the blood-as-foreshadow hypothesis with the most disease-relevant brain tissue. Pseudobulk differential expression (e.g., DESeq2[31]) per donor, rather than pooled-condition NATMI scoring, would enable per-donor L-R rankings and thereby provide proper confidence intervals on the concordance score. A primary falsifiable prediction: if pseudobulk LIANA inference applied to SN tissue from the Ma 2025 cohort recovers ≥ 5 of the 9 pairs identified here in the same direction, that outcome would confirm that peripheral blood L-R signals foreshadow SN-specific neuroinflammation. Conversely, if fewer than 3 pairs replicate in the SN, it would indicate that the peripheral signal is non-specific to the primary degenerative

site, limiting its biomarker utility. A second falsifiable prediction: if the same LIANA pipeline applied to a matched AD PBMC cohort against the Fullard brain atlas yields a concordance rate ≥ 7 (within 2 of the 9 observed here), that would support a disease-general neuroinflammatory foreshadow model rather than PD-specific biology.

4 Methods

4.1 Datasets

PBMC data (peripheral arm). Single-cell RNA-sequencing data from peripheral blood mononuclear cells were obtained from GEO accession GSE223138, as generated and published by Hong et al. (2025).[1] The dataset contains 6 donors (3 PD, 3 healthy control) profiled on the 10x Genomics Chromium platform (v3 chemistry), stored as per-donor MTX files (features/barcodes/matrix). Raw data were downloaded from GEO and cached locally for analysis. The total dataset contains 66,826 cells across 36,601 genes (as loaded from the GEO MTX files). No restriction on data access applies (GEO public release).

Brain leukocyte data (brain arm). Single-nucleus RNA-sequencing data from the Fullard 2024 multi-region PD atlas were used as the brain-side comparator.[3] The atlas (AMP PD Knowledge Platform; Prashant NM, Fullard JF et al., Scientific Data 2024) provides 100 postmortem cases and controls across five brain regions: DMNV, globus pallidus, prefrontal cortex, motor cortex, and visual cortex. Cells labelled as "leukocyte" in the original atlas annotation (n = 2,361) were extracted from `fullard_pd_multiregion.h5ad` for sub-type annotation. CNS macrophages (n = 118,515 cells) were identified from the full atlas by the original cell-type annotation and subsampled to 5,000 for computational efficiency. The Fullard 2024 atlas does not contain substantia nigra tissue.

Ma 2025 brain reference. The full SN snRNA-seq dataset from Ma et al. (2025)[2] (PMC12322057) was not publicly available at the time of analysis. The study's CD8+ T cell module was approximated using a 51-gene list text-extracted from the publication's Supplementary Data, used qualitatively in Figure 4d only and not in the primary statistical analysis.

4.2 Pre-processing

PBMC data were loaded per-donor using `code/data_loader.py` and concatenated into a single AnnData object.[26] Cells were filtered to a minimum of 200 genes per cell and maximum 5% mitochondrial gene fraction. Log-normalisation was applied (`scanpy.pp.normalize_total` to 10,000 counts per cell, then `scanpy.pp.log1p`). UMAPs were computed via scanpy[21] (`scanpy.tl.pca`, `scanpy.pp.neighbors`, `scanpy.tl.umap`[24]) on the 2,000 top highly variable genes. Cells were subsampled to 1,000 per donor to balance computational load; subsampling used a fixed random seed (seed = 42; see `code/main.py`).

Cell-type annotation was based on marker gene scoring: CD8A/CD8B/GZMK/GZMB for CD8+ T cells, CD4/IL7R for CD4+ T cells, CD14/LYZ for monocytes, GNLY/NKG7 for NK cells, MS4A1/CD19 for B cells. Scores were computed with `scanpy.tl.score_genes`. Cell types were assigned by argmax score. Brain leukocytes underwent the same marker scoring for CD8+ T, NK, and other-leukocyte categories. All pre-processing steps are implemented in `code/data_loader.py`.

4.3 L-R inference

Cell-cell communication inference used LIANA-py (v1.7.1)[4] with the CellChatDB ligand-receptor database (1,912 manually curated L-R pairs).[6,7] For each arm (PBMC and brain), LIANA was run separately on PD-condition cells and control-condition cells to generate condition-specific NATMI specificity weight matrices. PD-differential scoring was computed as the difference in NATMI specificity weight between PD and control subgroups [(PD NATMI weight) − (control NATMI weight)] for each L-R pair in the CD8+ T → myeloid direction. This score-based ranking is analogous to log-fold-change ranking in bulk RNA-seq differential expression analysis: it provides an ordered list for top-K selection based on the magnitude of condition-specific expression, with higher differential scores reflecting greater PD-associated upregulation. The permutation test (described in Statistical analysis) provides formal inference on the intersection count, not on any individual pair's ranking. Top-K = 100 pairs were selected from each arm; the PBMC arm yielded 23 distinct pairs and the brain arm yielded 27 distinct pairs at this threshold. The primary analysis used K = 100; sensitivity analysis used K = 50 and K = 200.

The originally planned inference method—LIANA `rank_aggregate` with internal permutation testing—was not used due to software instability (segmentation fault) in the LIANA permutation engine in the compute environment. NATMI differential ranking was used as the primary method. This constitutes a pre-registered deviation, recorded in `results.json` under `data_contract_deviations`.

4.4 Statistical analysis

Primary permutation test (H1). To determine whether the observed intersection of 9 shared pairs could arise by chance, we constructed an empirical null distribution by drawing 10,000 random subsets of K = 100 L-R pairs from CellChatDB (1,912 total pairs), computing the intersection of each random subset with the brain top-K = 100 list, and recording the intersection size. The null mean was 1.393 (SD = 1.136); the null 99th percentile was 4. The observed count of 9 exceeds the null 99th percentile, placing the p-value below the permutation resolution floor: $p < 1/10{,}000$. An additional shuffled-label null was constructed by randomly permuting PD/control condition labels across cells (10,000 iterations), re-running the NATMI differential scoring, and computing the resulting intersection. Random seed for permutation tests: seed = 42 (`code/metrics.py`).

H2 enrichment test. Fisher's exact test (two-sided) was used to assess enrichment of shared L-R pair genes (18 unique genes from 9 pairs) for the Hong 2025 MR-implicated gene list (28 genes) against the background of all CellChatDB genes (1,912 pairs). OR = 2.56 (95% CI [0.058, 18.0]), p = 0.346. Implemented in `code/metrics.py` using `scipy.stats.fisher_exact`.[23]

H4 direction-specificity. The reverse direction (myeloid → CD8+ T) was subjected to the same top-K intersection and permutation test as the primary analysis. Permutation null was re-generated independently (10,000 iterations, seed = 42). Result: reverse count = 4, permutation p = 0.053.

Bootstrap CI for observed count. The 95% CI of [8.978, 9.022] was computed by bootstrapping the fixed observed count of 9 across 10,000 bootstrap resamples. The narrow CI reflects near-zero sampling variance in the observed point estimate (the count is deterministic given the data), not uncertainty about the true shared-pair distribution. The scientifically informative comparison is between the observed count (9) and the null distribution (mean 1.39, 99th pct 4).

Package versions. Python 3.11.15; liana 1.7.1;[4] anndata 0.10.9;[26] scanpy 1.11.5;[21] numpy 1.26.4;[27] scipy 1.10.1;[23] pandas 3.0.3;[30] matplotlib 3.10.9;[22] seaborn 0.13.2;[25] umap-learn 0.5.12.[24]

4.5 Model construction and lockdown

Pre-registration of all hypotheses and analysis parameters was performed before data access, recorded in `002_execute/locked_model.json` (SHA-256: `b22f8b838da990136754eb158261043972c1c4bde025e8d80fc024d27298d964`). Hypotheses H1–H5 and their decision criteria (K threshold, permutation B, p-value thresholds) were specified in the locked model. No changes were made to the primary metric definition after data access.

4.6 Validation protocol

The primary metric (H1: shared PD-up L-R pair count ≥ 5 with permutation $p < 0.01$) was pre-registered. Secondary metrics (H2, H4, H5) were analysed after H1. H3 was declared inapplicable before analysis when it was established that per-donor brain T-cell module scores were not available in the brain-arm data. Cross-check verification was performed by the `003_cross_check` stage, which confirmed reproducibility of the `results.json` values from source code.

4.7 Figure generation

Figures were generated by `code/figures/fig1_schematic.py` through `code/figures/fig6_direction_robustness_honest.py`. All figures use pre-computed coordinates from the LIANA outputs stored in `002_execute/figures/data/`. Figure 1 is a schematic with no data plotted. Figures 2–6 plot quantities directly derived from results stored in the figure data directory. All figure scripts use matplotlib 3.10.9[22] and seaborn 0.13.2[25] with adjustText 1.3.0 (Python package; no peer-reviewed publication) for label placement.

4.8 Data availability

PBMC data: GEO GSE223138 (public). Brain atlas: AMP PD Knowledge Platform, Fullard 2024 multi-region snRNA-seq (accession doi:`10.1038/s41597-024-04117-y`). CellChatDB ligand-receptor database: embedded in LIANA-py v1.7.1 (1,912 pairs used for permutation null). Ma 2025 SN data: PMC12322057; accession status pending at time of analysis.

4.9 Code availability

All analysis code is available under `002_execute/code/`: `data_loader.py`, `main.py`, `metrics.py`, `models.py`, and `figures/fig1_schematic.py` through `figures/fig6_direction_robustness_honest.py`. The environment lock file at `002_execute/code/environment.lock.json` specifies all package versions.

References

[1] Y. Hong, J. Zhou, Y. Wang, S. Song, H. Chen, Y. Mi, and X. Li. Peripheral immune cell-specific genes in parkinson's disease uncovered by multi-omics with therapeutic implications. *npj Parkinson's Disease*, 11:302, 2025. doi: 10.1038/s41531-025-01148-z.

[2] M. Ma, F. Paryani, K. Jakubiak, S. Xia, S. Antoku, A. Kannan, 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.

[3] N. M. Prashant, J. F. Fullard, T. Clarence, et al. A multi-region single nucleus transcriptomic atlas of parkinson's disease. *Scientific Data*, 11:1274, 2024. doi: 10.1038/s41597-024-04117-y.

[4] D. Dimitrov, P. V. L. Schäfer, E. Farr, P. Rodriguez-Mier, S. Lobentanzer, P. Badia-I-Mompel, A. Dugourd, J. Tanevski, R. O. Ramirez Flores, and J. Saez-Rodriguez. Liana+ provides an all-in-one framework for cell-cell communication inference. *Nature Cell Biology*, 26:1613–1622, 2024. doi: 10.1038/s41556-024-01469-w.

[5] D. Dimitrov, D. Türei, M. Garrido-Rodriguez, P. L. Burmedi, J. S. Nagai, C. Boys, et al. Comparison of methods and resources for cell-cell communication inference from single-cell rna-seq data. *Nature Communications*, 13:3224, 2022. doi: 10.1038/s41467-022-30755-0.

[6] S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C. H. Kuan, P. Myung, M. V. Plikus, and Q. Nie. Inference and analysis of cell-cell communication using cellchat. *Nature Communications*, 12:1088, 2021. doi: 10.1038/s41467-021-21246-9.

[7] S. Jin, M. V. Plikus, and Q. Nie. Cellchat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nature Protocols*, 20:180–219, 2025. doi: 10.1038/s41596-024-01045-4.

[8] M. Efremova, M. Vento-Tormo, S. A. Teichmann, and R. Vento-Tormo. Cellphonedb: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nature Protocols*, 15:1484–1506, 2020. doi: 10.1038/s41596-020-0292-x.

[9] R. Browaeys, W. Saelens, and Y. Saeys. Nichenet: modeling intercellular communication by linking ligands to target genes. *Nature Methods*, 17:159–162, 2020. doi: 10.1038/s41592-019-0667-5.

[10] R. Hou, E. Denisenko, H. T. Ong, J. A. Ramilowski, and A. R. R. Forrest. Predicting cell-to-cell communication networks using natmi. *Nature Communications*, 11:5011, 2020. doi: 10.1038/s41467-020-18873-z.

[11] S. Yazar, J. Alquicira-Hernandez, K. Wing, A. Senabouth, M. G. Gordon, S. Andersen, et al. Single-cell eqtl mapping identifies cell type-specific genetic control of autoimmune disease. *Science*, 376: eabf3041, 2022. doi: 10.1126/science.abf3041.

[12] M. A. Nalls, C. Blauwendraat, C. L. Vallerga, K. Heilbron, S. Bandres-Ciga, D. Chang, et al. Identification of novel risk loci, causal insights, and heritable risk for parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*, 18:1091–1102, 2019. doi: 10.1016/S1474-4422(19)30320-5.

[13] J. Galiano-Landeira, A. Torra, M. Vila, and J. Bové. Cd8 t cell nigral infiltration precedes synucleinopathy in early stages of parkinson's disease. *Brain*, 143:3717–3733, 2020. doi: 10.1093/brain/awaa269.

[14] C. S. Lindestam Arlehamn, R. Dhanwani, J. Pham, R. Kuan, A. Frazier, J. Rezende Dutra, et al. α-synuclein-specific t cell reactivity is associated with preclinical and early parkinson's disease. *Nature Communications*, 11:1875, 2020. doi: 10.1038/s41467-020-15626-w.

[15] D. Sulzer, R. N. Alcalay, F. Garretti, L. Cote, E. Kanter, J. Agin-Liebes, et al. T cells from patients with parkinson's disease recognize α-synuclein peptides. *Nature*, 546:656–661, 2017. doi: 10.1038/nature22815.

[16] C. M. Capelle, S. Ciré, F. Hedin, M. Hansen, L. Pavelka, K. Grzyb, et al. Early-to-mid stage idiopathic parkinson's disease shows enhanced cytotoxicity and differentiation in cd8 t-cells in females. *Nature Communications*, 14:7461, 2023. doi: 10.1038/s41467-023-43053-0.

[17] J. Skuljec, H. Sun, R. Pul, K. Bénardais, D. Ragancokova, D. Moharregh-Khiabani, A. Kotsiari, C. Trebst, and M. Stangel. Ccl5 induces a pro-inflammatory profile in microglia in vitro. *Cellular Immunology*, 270:164–171, 2011. doi: 10.1016/j.cellimm.2011.05.001.

[18] A. Lin and W.-H. Yan. The emerging roles of human leukocyte antigen-f in immune modulation and viral infection. *Frontiers in Immunology*, 10:964, 2019. doi: 10.3389/fimmu.2019.00964.

[19] F. N. E. Gavins and M. J. Hickey. Annexin a1 and the regulation of innate and adaptive immunity. *Frontiers in Immunology*, 3:354, 2012. doi: 10.3389/fimmu.2012.00354.

[20] E. S. Smith, A. Jonason, C. Reilly, J. Veeraraghavan, T. Fisher, M. Doherty, et al. Sema4d compromises blood-brain barrier, activates microglia, and inhibits remyelination in neurodegenerative disease. *Neurobiology of Disease*, 73:254–268, 2015. doi: 10.1016/j.nbd.2014.10.008.

[21] F. A. Wolf, P. Angerer, and F. J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19:15, 2018. doi: 10.1186/s13059-017-1382-0.

[22] J. D. Hunter. Matplotlib: a 2d graphics environment. *Computing in Science & Engineering*, 9:90–95, 2007. doi: 10.1109/MCSE.2007.55.

[23] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, 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.

[24] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *Journal of Open Source Software*, 3:861, 2018. doi: 10.21105/joss.00861.

[25] M. L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6:3021, 2021. doi: 10.21105/joss.03021.

[26] I. Virshup, D. Bredikhin, L. Hautekeete, A. O. Pisco, O. Stegle, F. J. Theis, and V. Kleshchevnikov. The scverse project provides a computational ecosystem for single-omics analysis. *Nature Biotechnology*, 41:604–606, 2023. doi: 10.1038/s41587-023-01733-8.

[27] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al. Array programming with numpy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[28] L. Zhang, Y. Liu, S. Chen, H. Jiang, et al. Comprehensive analyses of brain cell communications based on multiple scrna-seq and snrna-seq datasets for revealing novel mechanisms in neurodegenerative diseases. *CNS Neuroscience and Therapeutics*, 29:2775–2786, 2023. doi: 10.1111/cns.14280.

[29] [unresolved citation: Cellchat alzheimer disease brain atlas cx3cl1 microglia prefrontal cortex snrna-seq], n.d.. Unresolved citation: no bibliographic metadata (authors, journal, year, DOI) is available for this reference. Entry retained to preserve citation numbering; no metadata has been added.

[30] pandas (python data analysis library), version 3.0.3, n.d.. Software citation (package versions, Methods); no full bibliographic entry is available.

[31] Deseq2 (bioconductor differential-expression package), n.d.. Software citation (Discussion); no full bibliographic entry is available.