Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative systems biology identifies shared neuroimmune co-expression programmes across pulmonary tuberculosis and Alzheimer's disease

Oluwaseun Oluwatosin Taofeek

oluwaseun@nextgenbioinstitute.org

NextGenBio Institute

Solomon Adesola Adigun

NextGenBio Institute

Elijah Kolawole Oladipo

Helix Biogen Institute

Aganze Gloire-Aimé Mushebenge

University of South Africa

Afrah Khairallah

University of KwaZulu-Natal

Paul Akinduti

Federal University of Medicine and Medical Sciences

Lateef Adegboyega Sulaimon

University of Abuja

Research Article

Keywords: Pulmonary tuberculosis, Alzheimer's disease, Neuroimmune networks, Systems biology, Transcriptomics, WGCNA, Machine learning.

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11528962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrative systems biology identifies shared neuroimmune co-expression programmes across pulmonary tuberculosis and Alzheimer's disease

Short Title (Running Title)

Shared neuroimmune networks in PTB and Alzheimer's disease

Authors

Oluwaseun Oluwatosin Taofeek[1,2*], Solomon Adesola Adigun[1], Elijah Kolawole Oladipo[2], Aganze Gloire-Aimé Mushebenge[3], Afrah Khairallah[4], Paul Akinduti[5], Lateef Adegboyega Sulaimon[6]

[1]NextGenBio Institute, Abeokuta, Nigeria

[2]Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria

[3]University of South Africa, Pretoria, South Africa

[4]Africa Health Research Institute and University of KwaZulu-Natal, Durban, South Africa

[5]Federal University of Medicine and Medical Sciences, Abeokuta, Nigeria

[6] Department of Biochemistry, University of Abuja, Abuja, Nigeria

* Corresponding author

Email: oluwaseun@nextgenbioinstitute.org

orcid.org/0000-0002-3076-4648

Abstract

Background: There is growing evidence that pulmonary tuberculosis (PTB) exerts long-lasting systemic effects on brain function, whereas Alzheimer's disease (AD) is increasingly recognized as a systemic neuroimmune disorder. BCG vaccination studies suggest that tuberculosis-associated immune pathways influence AD-related biology. However, whether PTB-associated co-expression programmes exhibit concordant activity in AD and display reproducible neuroanatomical organization within the human brain remains unknown. We hypothesized that PTB-associated co-expression programmes would exhibit concordant transcriptional activity in AD, retain aspects of their underlying co-expression architecture across diseases, and identify shared neuroimmune mechanisms relevant to chronic infection and neurodegeneration. We implemented an integrative systems biology framework combining whole-blood transcriptomics, weighted gene co-expression network analysis, functional enrichment, immune-cell deconvolution, and machine learning to identify PTB-associated co-expression programmes and prioritize biomarkers. These programmes were validated in two independent PTB cohorts, evaluated in an AD dataset, and mapped onto the Allen Human Brain Atlas to characterize neuroanatomical organization.

Results: Active PTB was associated with widespread transcriptional remodeling, with 1,389 dysregulated genes organized into 16 PTB-associated co-expression modules. Nested machine learning identified a 12-gene biomarker signature that demonstrated excellent diagnostic performance in the discovery cohort (AUC ≈ 0.99) and retained performance in two independent external PTB datasets without model retraining. Majority of PTB-derived immune co-expression modules exhibited concordant activity in the independent AD cohort. Formal module-preservation analysis demonstrated that several PTB-derived networks retained moderate preservation in AD, supporting partial conservation of co-expression architecture across diseases. Brain transcriptomic mapping demonstrated reproducible region-specific organization across adult brain donors. Integrated neuroanatomical contextualization prioritized a subset of PTB-derived neuroimmune modules using cross-disease transcriptional activity, regional specificity, and inter-donor reproducibility, revealing neuroimmune transcriptional programmes shared across infection and neurodegeneration.

Conclusions: Our findings establish a systems-level framework for investigating shared neuroimmune processes across chronic infection and neurodegeneration while identifying reproducible co-expression modules and candidate biomarkers for future studies.

Keywords: Pulmonary tuberculosis; Alzheimer's disease; Neuroimmune networks; Systems biology; Transcriptomics; WGCNA; Machine learning.

Background

Tuberculosis (TB), caused by *Mycobacterium tuberculosis,* remains one of the leading causes of death from a single infectious agent worldwide despite the availability of effective anti-tuberculosis therapy and widespread Bacillus Calmette–Guérin (BCG) vaccination [1]. According to the World Health Organization Global Tuberculosis Report 2025, an estimated 10.7 million people developed TB and 1.23 million died from the disease in 2024, underscoring its persistent global health burden, particularly in low- and middle-income countries where socioeconomic inequalities and health system constraints continue to impede disease control [1]. Although pulmonary tuberculosis (PTB) has traditionally been regarded as a disease confined to the lungs, accumulating evidence indicates that it is a systemic disease with biological consequences extending well beyond the respiratory tract [2–4]. Even after successful microbiological cure, many individuals continue to experience persistent inflammatory and immune abnormalities, multisystem sequelae, impaired quality of life, and excess long-term mortality [5,6]. More recently, an intriguing body of experimental and clinical evidence has suggested that these systemic effects may also extend to the brain. Patients with PTB, even in the absence of clinically apparent central nervous system infection, have been reported to exhibit cognitive impairment, neuroinflammatory alterations accompanied by evidence of sustained peripheral immune activation [7–10]. These observations extend the current understanding of TB-associated neurological disease by suggesting that, even in the absence of direct central nervous system infection, chronic PTB may influence brain function through sustained systemic immune signalling and lung–brain immune crosstalk, potentially promoting neuroinflammatory responses [7,8]. Yet a fundamental

question remains unanswered: do these persistent systemic immune perturbations simply accompany PTB, or do PTB-associated immune co-expression programmes exhibit concordant transcriptional activity in neurodegenerative disease and retain aspects of their underlying network organization across disease contexts?

The biological plausibility of this hypothesis is supported by growing evidence that the persistent immune remodelling associated with PTB shares important features with the neuroimmune dysregulation that characterizes Alzheimer's disease (AD) [11–13]. Although the innate and adaptive immune responses elicited during active PTB are essential for containing Mycobacterium tuberculosis, persistent antigenic stimulation frequently leads to sustained production of pro-inflammatory cytokines, prolonged activation of monocytes, macrophages, neutrophils, and T cells, and widespread immune remodelling that may persist long after successful antimicrobial treatment [2,3,11,14,15]. These chronic immune alterations are increasingly recognized to extend beyond the circulation, promoting communication between the lung and the central nervous system through circulating cytokines, extracellular vesicles, metabolic reprogramming, immune-cell trafficking, and disruption of blood–brain barrier integrity, thereby facilitating neuroinflammatory responses [7,8]. Notably, many of these same mechanisms—including persistent activation of microglia and astrocytes, inflammasome signalling, complement-mediated synaptic remodelling, Toll-like receptor, NF-κB, NOD-like receptor, interferon, and cytokine signalling pathways—are now recognized as central drivers of AD pathogenesis rather than merely secondary consequences of amyloid-β and tau pathology [12,13]. The growing recognition that Bacillus Calmette–Guérin vaccination induces durable trained immunity through epigenetic and metabolic reprogramming of innate immune cells further supports the existence of long-lasting immune programmes capable of influencing neuroinflammatory processes relevant to neurodegeneration [16,17]. Collectively, these observations suggest that PTB and AD share features of neuroimmune pathway activation, raising the possibility that this convergence extends beyond individual signalling pathways to coordinated transcriptional programmes. Whether such programmes exhibit concordant transcriptional activity across PTB and AD,

how they are organised into coordinated co-expression networks, and whether they exhibit reproducible neuroanatomical organisation within the human brain remain unknown.

Despite this compelling biological rationale, the molecular relationship between PTB and AD remains largely unexplored. While transcriptomic and systems biology studies have substantially advanced our understanding of the molecular architecture underlying both conditions [11–15,18,19], it remains unknown whether PTB-associated immune co-expression programmes exhibit concordant transcriptional activity in AD, whether their underlying co-expression architecture is retained across disease contexts, whether they represent coordinated network-level processes rather than isolated inflammatory signals, or whether they display reproducible neuroanatomical organization within the human brain. Addressing these questions requires an integrative framework capable of identifying PTB-associated transcriptional networks, defining their immunological context, validating them in independent PTB cohorts, evaluating their activity in AD, and characterizing their regional expression patterns in the brain. Without such a systems-level perspective, the potential molecular links between chronic pulmonary infection and neurodegeneration remain poorly understood.

Collectively, these observations led us to hypothesize that PTB-associated immune co-expression programmes exhibit concordant transcriptional activity in AD, retain aspects of their underlying co-expression architecture across diseases, and display reproducible region-specific organization within the human brain, suggesting shared neuroimmune mechanisms relevant to both chronic infection and neurodegeneration. To test this hypothesis, we applied an integrative systems biology framework combining whole-blood transcriptomics, weighted gene co-expression network analysis (WGCNA), functional enrichment, immune-cell deconvolution, machine learning (ML), independent PTB validation, cross-disease evaluation in AD, and neuroanatomical mapping using the Allen Human Brain Atlas. This approach enabled us to identify PTB-associated transcriptional networks, evaluate their reproducibility across independent cohorts, assess their activity and network preservation in AD, and characterize their regional organization within the adult human brain.

Methods

Study Design

This study employed a multi-stage systems biology framework to identify peripheral immune transcriptional programmes associated with active PTB, evaluate their concordant transcriptional activity in an independent AD cohort, and characterize their regional brain expression patterns in the adult human brain. The analytical strategy integrated differential gene expression, WGCNA [24], functional enrichment, immune-cell deconvolution, ML, independent transcriptomic validation, and brain transcriptomic mapping to investigate molecular convergence between chronic infectious disease and neurodegeneration.

The workflow comprised five sequential stages. First, whole-blood RNA-sequencing data from individuals with active PTB and healthy controls were analysed to identify differentially expressed genes and construct weighted gene co-expression networks. Second, PTB-associated co-expression modules were functionally characterized using Gene Ontology (GO) [25], Kyoto Encyclopedia of Genes and Genomes (KEGG) [26], and Disease Ontology (DO) [27] enrichment analyses, followed by immune-cell deconvolution to examine relationships between transcriptional modules and circulating immune-cell populations. Third, network-derived hub genes were prioritized using a ML framework and validated in two independent PTB transcriptomic cohorts. Fourth, PTB-derived transcriptional programmes were subsequently evaluated in an independent AD prefrontal cortex RNA-sequencing cohort using module eigengene analyses, differential biomarker assessment, and formal WGCNA module-preservation analysis to assess both transcriptional activity and conservation of co-expression architecture across diseases. Finally, selected PTB-derived co-expression modules were mapped to the Allen Human Brain Atlas (AHBA) [28], analysed independently in two adult human brain donors, and integrated through a consensus neuroanatomical framework to identify reproducible region-specific molecular signatures shared across PTB and AD.

This integrated workflow enabled the evaluation of PTB-associated transcriptional programmes across multiple analytical layers, from network discovery and biological characterization to external validation, cross-disease assessment, and neuroanatomical mapping. An overview of the analytical workflow is presented in Figure 1.

Study Cohorts and Transcriptomic Datasets

Publicly available RNA-sequencing datasets were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and analyzed using a staged discovery-validation framework (Table 1). The primary discovery cohort was GSE157657 [29], a longitudinal whole-blood RNA-sequencing dataset comprising individuals with active PTB (PTB), healthy controls, household contacts, tuberculosis progressors, and pre-diagnostic tuberculosis. This cohort served as the discovery dataset for differential gene expression analysis, WGCNA, functional enrichment, immune-cell deconvolution, and ML.

Because GSE157657 contains repeated longitudinal samples from individual participants, analyses requiring statistically independent observations were performed using one representative sample per participant. Clinical metadata were harmonized using unique participant identifiers, samples were ordered by collection time, and the earliest available sample from each participant was retained for differential expression, WGCNA, immune-cell deconvolution, and ML analyses. The complete longitudinal metadata were additionally retained to enable a secondary differential expression analysis comparing individuals sampled prior to PTB diagnosis (PTB_PreDiagnosis) with healthy controls.

Two independent whole-blood RNA-sequencing datasets (GSE107991 [30] and GSE279063 [31]) comprising individuals with active PTB and healthy controls were used for external validation. Both cohorts were processed independently using the same preprocessing, quality-control, and variance-stabilizing

transformation procedures as the discovery cohort and were used exclusively to evaluate the reproducibility of the ML-derived TB biomarker signature.

Cross-disease evaluation was performed using the post-mortem prefrontal cortex RNA-sequencing dataset GSE313307 [32], which includes participants classified according to AD diagnosis and respiratory infection status. Samples were stratified into four predefined groups (AD_NoInf, AD_Inf, CTR_NoInf, and CTR_Inf), enabling both one-way comparisons among clinical groups and factorial analyses evaluating the independent and interactive effects of AD and respiratory infection. This dataset was used to determine whether TB-associated transcriptional programmes exhibited concordant activity in a neurodegenerative disease context and to evaluate preservation of PTB-derived co-expression networks in AD. The characteristics of all transcriptomic datasets, including sample size, sequencing platform, biological specimen, clinical groups, and analytical role, are summarized in Table 1.

Differential Gene Expression Analysis

Differential gene expression analysis was performed using the representative-sample discovery cohort derived from GSE157657 to identify transcriptional alterations associated with active PTB. To avoid pseudoreplication arising from repeated longitudinal sampling, only one representative sample per participant was included in the primary differential expression analysis. Raw RNA-sequencing count data were analyzed using DESeq2 [33]. Genes with low expression were excluded by retaining those with a cumulative count ≥ 15 across all samples.

Differential expression between active PTB and healthy controls was assessed using negative binomial generalized linear models implemented in DESeq2. Because complete age, sex, and batch information were unavailable or inconsistent across datasets, disease status was used as the primary design variable. P values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure, and genes with FDR-adjusted $P < 0.05$ and $|\log_2 FC| > 1$ were considered significantly differentially expressed.

Variance-stabilized count data generated using the DESeq2 variance-stabilizing transformation (VST) were used for principal component analysis (PCA). Differential expression results were visualized using volcano plots.

Weighted Gene Co-expression Network Analysis

Weighted gene co-expression network analysis (WGCNA) [24] was performed on variance-stabilized expression data from the representative-sample discovery cohort to identify transcriptional modules associated with active PTB. Sample and gene quality were assessed using the goodSamplesGenes() function, and potential outliers were evaluated by hierarchical clustering and principal component analysis.

A signed co-expression network was constructed using the WGCNA package. The soft-thresholding power was selected as the smallest value achieving a signed scale-free topology fit index (R^2) ≥ 0.85 while maintaining network connectivity; a power of 20 was therefore used for network construction. Modules were identified using blockwiseModules() with dynamic tree cutting (deepSplit = 2) and a minimum module size of 50 genes, and closely related modules were merged using a module eigengene dissimilarity threshold (mergeCutHeight = 0.25).

Module eigengenes were correlated with clinical traits, including disease status, age, and sampling time, using Pearson correlation. Because demographic and technical covariates were unavailable or inconsistently reported across datasets, disease status was treated as the principal analytical variable throughout the integrative analyses. Module–trait relationships were visualized using annotated correlation heatmaps, and eigengene network analysis was performed to examine relationships among modules.

To prioritize biologically relevant genes, module membership (MM), gene significance (GS), and intramodular connectivity (kWithin) were calculated for all genes. Hub genes were defined as the most highly connected genes within each module, and the top 20 hub genes per module were retained for downstream functional enrichment, machine-learning, and validation analyses.

Functional Annotation and Pathway Enrichment Analysis

To characterize the biological functions represented by disease-associated co-expression modules, functional enrichment analyses were performed using the clusterProfiler Bioconductor package [34] together with the org.Hs.eg.db, AnnotationDbi, and DOSE annotation resources [35]. Only modules significantly associated with disease status in the WGCNA analysis were included. Modules significantly associated with disease status ($|r| \geq 0.40$, $P < 0.05$) were selected for enrichment analyses. All genes retained after transcriptomic preprocessing and WGCNA quality control were used as the reference gene universe. GO enrichment [25] was performed using the enrichGO() function for the Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) ontologies. KEGG pathway enrichment [26] was conducted using enrichKEGG() with *Homo sapiens* as the reference organism, and DO enrichment [27] was performed using the enrichDO(). For all enrichment analyses, statistical significance was assessed using Benjamini–Hochberg multiple-testing correction with $P < 0.05$ and $q < 0.20$. Semantically redundant GO terms were reduced using simplify() with a similarity cutoff of 0.70, retaining the most significant representative term from each cluster.

Immune Cell Deconvolution Analysis

To characterize immune-cell populations associated with PTB-related transcriptional programmes, immune-cell composition was estimated using precomputed CIBERSORTx immune-cell fraction estimates derived from the representative-sample discovery cohort [36]. Analyses were restricted to individuals with active PTB. Only PTB-associated co-expression modules identified by WGCNA (FDR < 0.05) were included. Associations between module eigengenes and immune-cell fractions were quantified using Spearman rank correlation, and resulting P values were adjusted using the Benjamini–Hochberg procedure. Associations with an adjusted FDR < 0.05 were considered statistically significant.

Significant module–immune-cell associations were visualized using hierarchically clustered correlation heatmaps, and scatterplots were generated for the strongest associations. Immune-cell populations were further prioritized according to the number and strength of significant module associations.

Machine Learning-Based Biomarker Prioritization

A supervised machine-learning framework integrating penalized regression and ensemble learning was used to prioritize biomarkers capable of discriminating active PTB from healthy controls. Candidate features were derived from PTB-associated WGCNA modules. Specifically, the top 20 intramodular hub genes from modules significantly associated with PTB (FDR < 0.05) were retained as the initial feature set. Analyses were restricted to PTB and healthy control samples from the representative-sample discovery cohort. Model development employed nested cross-validation consisting of an outer 10-fold cross-validation loop for unbiased performance estimation and an inner repeated 10-fold cross-validation loop (five repetitions) for model optimization. Within each outer training fold, LASSO logistic regression implemented in glmnet was used for feature selection, retaining genes with non-zero coefficients at the λ_{1se} penalty. Selected genes were subsequently used to train a Random Forest classifier, with hyperparameters optimized within the inner cross-validation loop. Predictions from held-out outer folds were pooled to generate unbiased performance estimates. Classifier performance was evaluated using ROC analysis with 95% confidence intervals, precision–recall analysis, confusion matrix statistics, calibration metrics (Brier score, calibration intercept, and calibration slope), and Decision Curve Analysis (DCA). Statistical significance was assessed using a permutation test comprising 1,000 random permutations of outcome labels. Following nested cross-validation, the complete discovery cohort was used to train a final locked LASSO-derived Random Forest classifier. The resulting biomarker signature was subsequently evaluated, without retraining, in two independent external validation cohorts (GSE107991 and GSE279063), where discrimination, calibration, and clinical utility were assessed using the same performance framework.

External TB Validation

The robustness and generalizability of the final twelve-gene biomarker signature were evaluated in two independent whole-blood RNA-sequencing cohorts (GSE107991 and GSE279063). Samples were restricted to individuals with active pulmonary tuberculosis (PTB) and healthy controls and were processed using the same preprocessing and variance-stabilizing transformation procedures applied to the discovery cohort. The locked Random Forest classifier developed in the discovery cohort was applied to each external dataset without retraining. Model performance was assessed using receiver operating characteristic (ROC) analysis with corresponding area under the curve (AUC) estimates and 95% confidence intervals. Additional evaluation included confusion matrix statistics, precision–recall analysis, calibration metrics (Brier score, calibration intercept, and calibration slope), and Decision Curve Analysis (DCA) to assess clinical utility across a range of decision thresholds.

Cross-disease Evaluation of PTB-derived Neuroimmune Signatures

Following validation in independent PTB cohorts, the PTB-derived neuroimmune signature was evaluated in the independent post-mortem prefrontal cortex Alzheimer's disease (AD) RNA-sequencing cohort GSE313307 to assess its cross-disease transcriptional activity. The cohort comprised four predefined groups: AD without respiratory infection (AD_NoInf), AD with respiratory infection (AD_Inf), cognitively normal controls without infection (CTR_NoInf), and cognitively normal controls with infection (CTR_Inf).

Raw count data were processed using DESeq2, and genes with low expression were excluded prior to analysis. Because demographic and technical covariates were unavailable or inconsistently reported, disease and infection status were used as the primary design variables. Variance-stabilizing transformation (VST) was applied to generate normalized expression values, and data quality was assessed using principal component analysis.

Differential expression analyses were performed using CTR_NoInf as the reference group to evaluate three predefined contrasts: AD_NoInf versus CTR_NoInf, AD_Inf versus AD_NoInf, and CTR_Inf versus CTR_NoInf. Statistical significance was assessed using Wald tests implemented in DESeq2.

The PTB-derived biomarker signature was subsequently evaluated in GSE313307 by examining expression patterns across the four clinical groups. Group differences were assessed using one-way analysis of variance (ANOVA) with Benjamini–Hochberg false discovery rate correction, and pairwise comparisons between AD_NoInf and CTR_NoInf were performed to identify TB-derived biomarkers exhibiting concordant dysregulation in AD. Expression patterns were visualized using hierarchical clustering heatmaps and boxplots.

Module Preservation Analysis

To determine whether PTB-derived transcriptional programmes retained their underlying co-expression architecture in AD, formal module-preservation analysis was performed using the WGCNA modulePreservation() function. Genes shared between the PTB discovery network and GSE313307 were identified and restricted to those passing quality-control procedures in both datasets. Module assignments derived from the PTB discovery network were used as the reference network, and preservation statistics were calculated in the AD cohort using 500 permutation tests. Preservation strength was assessed using the composite Zsummary statistic, where Zsummary < 2 indicates no evidence of preservation, $2 \leq$ Zsummary ≤ 10 indicates moderate preservation, and Zsummary > 10 indicates strong preservation. Gene-retention statistics were additionally calculated for each module by quantifying the proportion of discovery-network genes represented in the AD dataset.

Allen Human Brain Atlas

Neuroanatomical mapping of PTB-associated co-expression modules was performed using RNA-sequencing data from two neurologically normal adult donors (H0351.2001 and H0351.2002) in the Allen

Human Brain Atlas (AHBA) [28]. Expression matrices, anatomical annotations, and ontology files were imported and matched using sample identifiers. WGCNA module assignments from the PTB discovery cohort were integrated with AHBA data using shared gene symbols.

For each module, mean expression across constituent genes was calculated for every brain sample to generate module-level expression profiles. Module expression values were merged with anatomical metadata and summarized across major brain regions and subregions. To facilitate cross-module and cross-donor comparisons, expression values were standardized using Z-score transformation and visualized using hierarchical clustering heatmaps.

Regional differences in module expression were assessed independently in each donor using Kruskal–Wallis tests. Donor-specific results were subsequently integrated through consensus analyses to identify reproducible neuroanatomical expression patterns associated with PTB-derived transcriptional programmes.

Consensus Neuroanatomical Analysis

Donor-specific AHBA analyses were integrated to identify reproducible neuroanatomical patterns of PTB-associated co-expression modules. Module-level expression profiles generated independently for donors H0351.2001 and H0351.2002 were standardized (Z-scores) and compared across major brain regions.

Inter-donor reproducibility was assessed on a module-by-module basis using Pearson and Spearman correlations, together with root mean square error (RMSE) and mean absolute deviation (MAD). Because multiple anatomical samples originated from the same donor, regional brain samples were not treated as independent biological replicates. Accordingly, analyses were interpreted as describing within-donor regional transcriptional organization, while concordance between donors was used to evaluate reproducibility rather than population-level significance.

Consensus regional expression profiles were generated by averaging standardized module expression values across donors for each major brain region. Consensus matrices were used to identify regions exhibiting the strongest module enrichment and to quantify regional specificity using the variance, expression range, and peak Z-score across regions.

The analysis was subsequently extended to anatomical substructures by averaging standardized expression values across corresponding subregions from both donors. Consensus subregional profiles were used to identify anatomical locations exhibiting the most consistent and pronounced expression of PTB-associated transcriptional modules.

Integrated Neuroanatomical Contextualization

Consensus neuroanatomical metrics derived from the AHBA were integrated with module-level AD validation and module-preservation results to identify PTB-associated co-expression modules exhibiting both disease relevance and reproducible neuroanatomical enrichment. For each validated module, neuroanatomical characteristics, including peak brain region, peak regional enrichment (Z-score), regional specificity metrics, and inter-donor reproducibility statistics (Pearson correlation, Spearman correlation, RMSE, and MAD), were combined with AD validation results.

Disease-associated effects were obtained from the diagnosis term of the two-way ANOVA and included effect size, statistical significance, Benjamini–Hochberg FDR, and generalized eta-squared (η^2G). These measures were integrated with the neuroanatomical metrics to enable joint evaluation of disease association, regional enrichment, and cross-donor reproducibility. Module-preservation statistics were retained as complementary evidence regarding conservation of co-expression architecture but were not incorporated into the integrated ranking framework.

Validated modules were subsequently prioritized according to the strength of AD association, neuroanatomical specificity, and inter-donor reproducibility to identify robust neuroimmune transcriptional programs exhibiting concordant peripheral and neuroanatomical organization.

Statistical Analysis

All analyses were performed in R (version 4.4.1; R Foundation for Statistical Computing, Vienna, Austria) [37] using Bioconductor and CRAN packages appropriate to each stage of the workflow. Statistical methods specific to individual analyses are described in the corresponding subsections. Unless otherwise stated, all tests were two-sided. Differential expression was assessed using Wald tests implemented in DESeq2. Correlation analyses, Student's t-tests, one-way and two-way ANOVA, Kruskal–Wallis tests, and machine-learning performance metrics were applied as appropriate to the analytical design. Network preservation was evaluated using permutation-based WGCNA module-preservation analysis, and preservation strength was quantified using the composite Zsummary statistic. Classifier performance was evaluated using ROC analysis, area under the curve (AUC), confusion-matrix statistics, sensitivity, specificity, balanced accuracy, and Cohen's kappa. Multiple-testing correction was performed using the Benjamini–Hochberg false discovery rate (FDR) procedure where applicable. Unless otherwise specified, an adjusted FDR < 0.05 was considered statistically significant. Effect sizes, including generalized eta-squared (η^2G) for ANOVA, were reported where appropriate.

Results

Overview of the integrative systems biology workflow

A sequential integrative systems biology workflow was implemented to investigate the systems-level molecular relationship between PTB and AD, integrating transcriptomic discovery, network biology, immune characterization, machine learning, independent validation, cross-disease evaluation, and regional brain transcriptomic mapping (Figure 1). The analysis began with differential gene expression and

WGCNA of whole-blood RNA-sequencing data from the discovery PTB cohort (GSE157657), followed by functional enrichment and immune-cell deconvolution to characterize disease-associated transcriptional programmes.

Network-derived biomarkers were subsequently prioritized using machine learning and evaluated in two independent PTB cohorts (GSE107991 and GSE279063). PTB-associated transcriptional programmes were first evaluated for differential activity in an independent AD cohort (GSE313307) and subsequently subjected to formal WGCNA module-preservation analysis to determine whether their underlying network organization was retained across diseases. Finally, selected co-expression modules were mapped to the Allen Human Brain Atlas (AHBA), analysed independently in two neurologically normal adult donors, and integrated through a consensus framework to identify reproducible region-specific expression patterns. The characteristics and analytical roles of all datasets are summarized in Table 1.

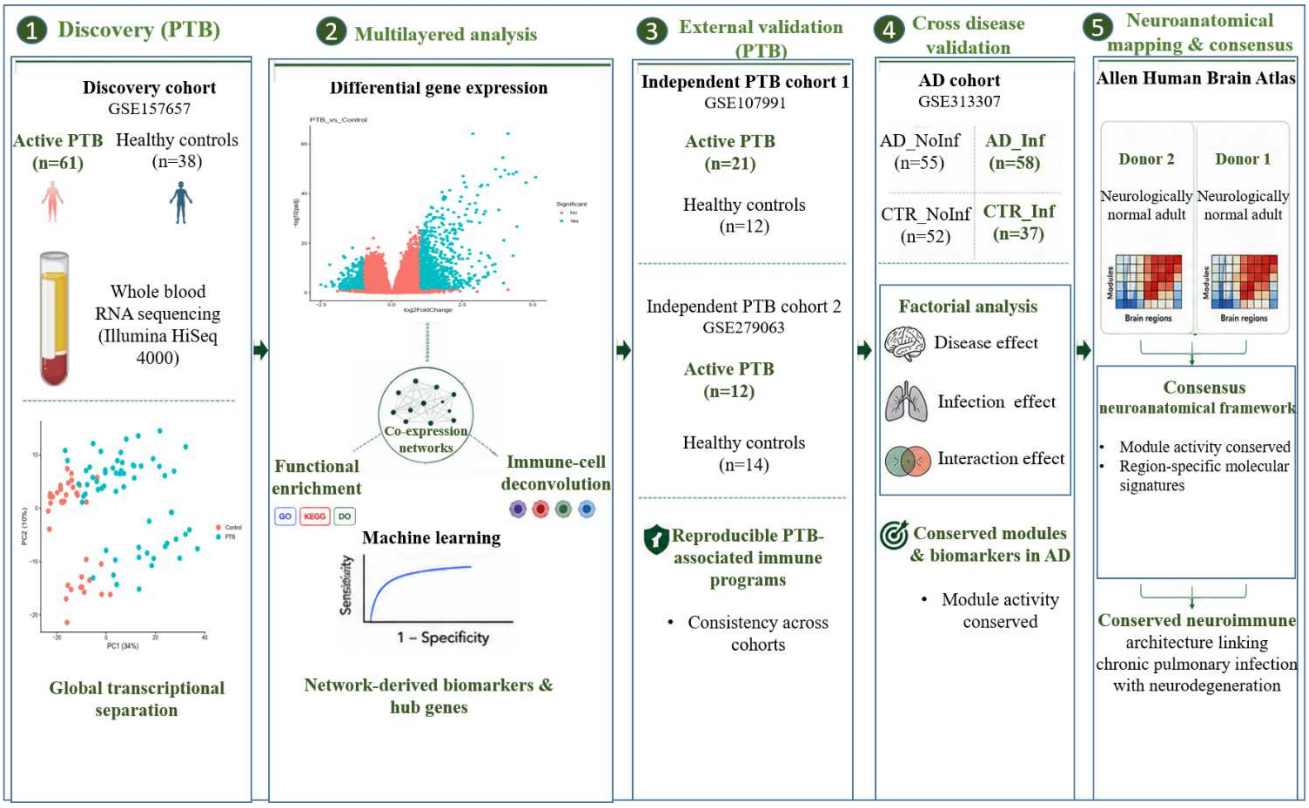


Figure 1. Overview of the integrative systems biology workflow used to identify PTB-associated peripheral immune co-expression programmes, evaluate their activity in an independent AD cohort, and characterize their regional brain expression patterns using the Allen Human Brain Atlas (AHBA).

Table 1. Summary of transcriptomic datasets used for discovery, independent validation, cross-disease evaluation, and neuroanatomical mapping.

Dataset	Disease/Condition	Study Design	Biological Specimen	Sequencing Platform	Clinical Groups Included	Analytical Role
GSE157657	Pulmonary tuberculosis	Longitudinal cohort	Whole blood	Illumina HiSeq 4000	Active pulmonary TB, healthy controls, household contacts, TB progressors, pre-diagnostic TB	Primary discovery cohort for differential expression, WGCNA, functional enrichment, immune-cell deconvolution, and machine learning.
GSE107991	Pulmonary tuberculosis	Independent case-control cohort	Whole blood	Illumina HiSeq 4000	Active pulmonary TB, healthy controls	External validation of the TB-derived gene signature and machine-learning model.
GSE279063	Pulmonary tuberculosis	Independent case-control cohort	Whole blood	Illumina HiSeq 2500	Active pulmonary TB, healthy controls	Independent validation of the TB transcriptional signature and machine-learning model.
GSE313307	Alzheimer's disease	Cross-sectional cohort	Post-mortem prefrontal cortex	Illumina NovaSeq 6000 RNA sequencing	AD_NoInf, AD_Inf, CTR_NoInf, CTR_Inf	Cross-disease evaluation of TB-associated transcriptional modules using one-way ANOVA, two-way ANOVA, and post hoc analyses.
Allen Human Brain Atlas	Neurologically normal adult brain	Reference brain transcriptomic atlas	Multiple postmortem brain regions	Illumina HiSeq 2000	Two neurologically normal adult donors (H0351.2001 and H0351.2002)	Regional brain transcriptomic mapping, donor comparison, and anatomical contextualization, and regional localization of concordant transcriptional programmes.

Active PTB is associated with widespread peripheral transcriptional remodeling

Differential expression analysis of the discovery cohort (GSE157657) was performed to characterize peripheral transcriptional alterations associated with active PTB. Following quality control and selection of one representative sample per participant, the analysis included 99 independent samples comprising 61 individuals with active PTB and 38 healthy controls. Principal component analysis (PCA) demonstrated clear separation between PTB and healthy control samples, indicating distinct global transcriptional profiles. The first two principal components explained 34% and 10% of the total variance, respectively (Figure 2A). Using DESeq2 (adjusted $P < 0.05$ and $|\log_2 \text{fold change}| > 1$), we identified 1,389 differentially expressed genes in PTB relative to healthy controls, including 1,043 upregulated and 346 downregulated genes. Volcano plot visualization revealed extensive transcriptional dysregulation, with a predominance of upregulated genes in PTB (Figure 2B). Complete differential expression results are provided in Table S1, while significantly dysregulated, upregulated, and downregulated genes are listed in Tables S2–S4. Differential expression quality-control diagnostics, including MA and dispersion plots, are presented in Figures S1–S2. Taken together, these findings demonstrate that active PTB is associated with extensive peripheral transcriptional remodeling. However, differential expression analysis alone cannot determine whether these dysregulated genes function independently or are organized into coordinated biological programmes.

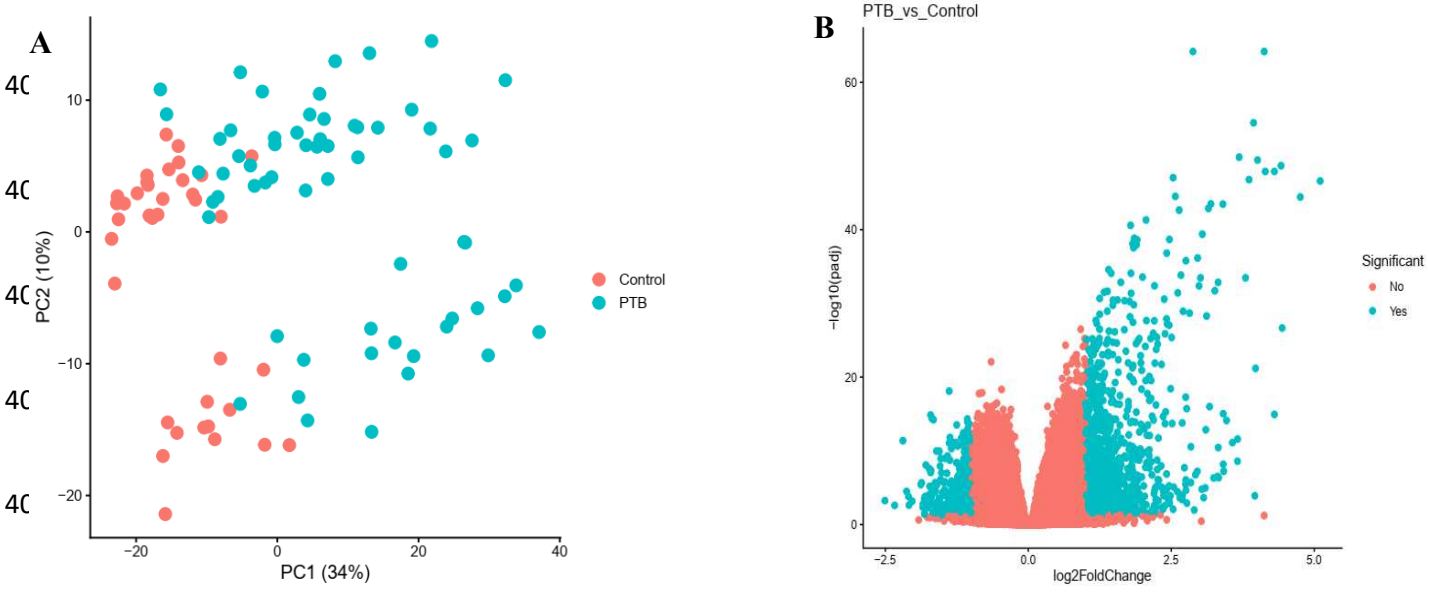


Figure 2. Active PTB is associated with extensive peripheral transcriptional remodeling. **(A)** Principal component analysis (PCA) of whole-blood RNA-sequencing data from the discovery cohort (GSE157657) demonstrates clear separation between active PTB (PTB) and healthy control samples, indicating distinct global transcriptomic profiles. The first two principal components explain 34% and 10% of the total variance, respectively. **(B)** Volcano plot showing differential gene expression between active PTB and healthy controls. Genes meeting the significance criteria (adjusted $P < 0.05$ and $|\log_2 \text{fold change}| > 1$) are highlighted, revealing extensive transcriptional dysregulation with a predominance of upregulated genes in PTB.

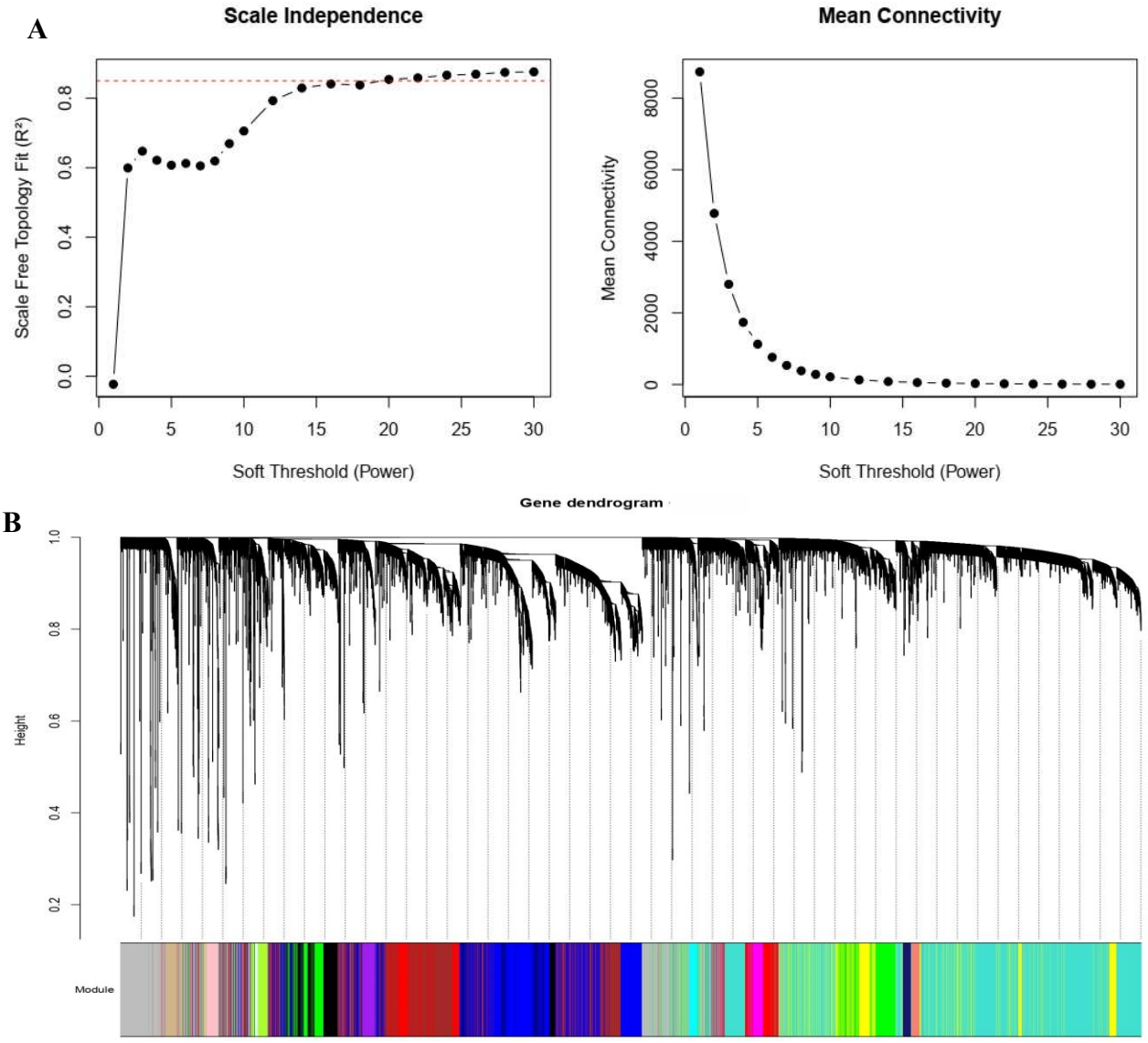
PTB-associated transcriptional alterations **organize into disease-associated co-expression modules**

WGCNA was performed to determine whether PTB-associated transcriptional alterations were organized into coordinated biological networks rather than acting independently. A signed weighted co-expression network was constructed using an automatically selected soft-thresholding power of $\beta = 20$, which satisfied the scale-free topology criterion while preserving network connectivity (Figure 3A). Network construction identified 16 co-expression modules, while 2,031 genes that could not be confidently assigned to a specific module were retained in the grey module (Figure 3B), indicating that PTB-associated transcriptional alterations were organized into discrete network structures rather than isolated gene-expression changes.

Correlation of module eigengenes with clinical traits identified several modules strongly associated with active PTB (Figure 3C). The turquoise module showed the strongest negative association with PTB ($r = -0.588$, $P = 7.57 \times 10^{-12}$), whereas the blue module exhibited the strongest positive association ($r = 0.568$, $P = 5.48 \times 10^{-11}$). Additional PTB-associated modules included the purple ($r = 0.559$), lightcyan ($r = 0.489$), midnightblue ($r = -0.482$), brown ($r = 0.438$), and yellow ($r = -0.420$) modules (all $P < 0.001$). In contrast, PTB pre-diagnosis status and other clinical variables showed comparatively weak associations with most module eigengenes, indicating that active disease status was the principal driver of co-expression variation

within the network. Complete module characteristics and network statistics are provided in Tables S5–S8, with higher-order eigengene relationships shown in Figures S3–S4.

Collectively, these findings demonstrate that PTB-associated transcriptional alterations are organized into coordinated disease-associated co-expression modules, providing a systems-level framework for investigating the biological processes underlying host responses to tuberculosis. Functional enrichment analyses were subsequently performed to characterize the biological functions represented within these modules.



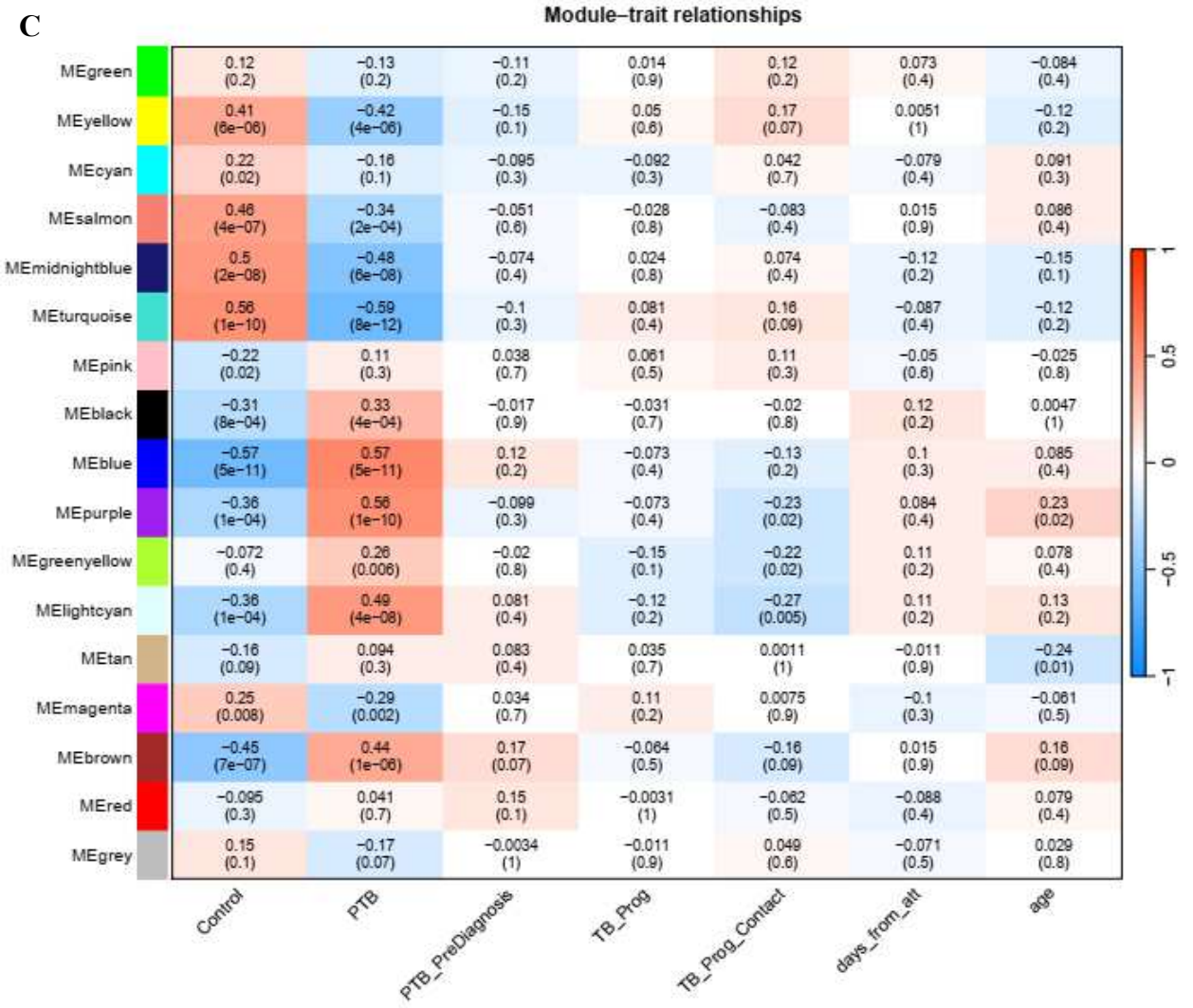


Figure 3. Identification of PTB-associated co-expression modules using weighted gene co-expression network analysis (WGCNA).

(A) Soft-threshold analysis showing scale-free topology fit index (left) and mean connectivity (right), with a soft-thresholding power of β = 20 selected for network construction. **(B)** Hierarchical clustering dendrogram of genes with corresponding module colour assignments generated using dynamic tree cutting. Sixteen co-expression modules were identified, while unassigned genes were grouped into the grey module. **(C)** Module–trait relationship heatmap showing Pearson correlation coefficients (upper value) and corresponding *P* values (lower value) between module eigengenes and clinical traits. Several modules exhibited strong positive or negative associations with active PTB.

PTB-associated co-expression modules reveal coordinated immune, inflammatory, and host defence programmes

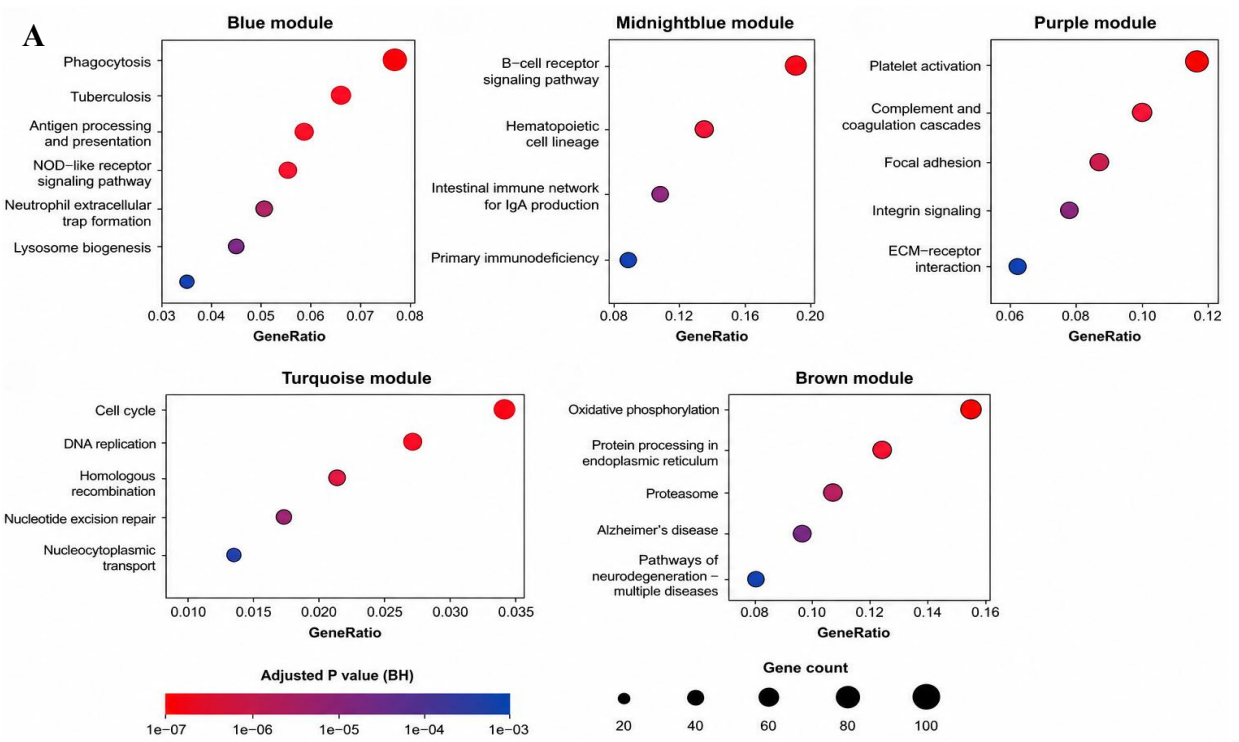
Functional enrichment analysis of the PTB-associated co-expression modules using Gene Ontology Biological Process (GO-BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Disease Ontology (DO) databases revealed the biological functions represented within disease-associated networks, demonstrating that these modules corresponded to coordinated biological programmes associated with active PTB (Figure 4).

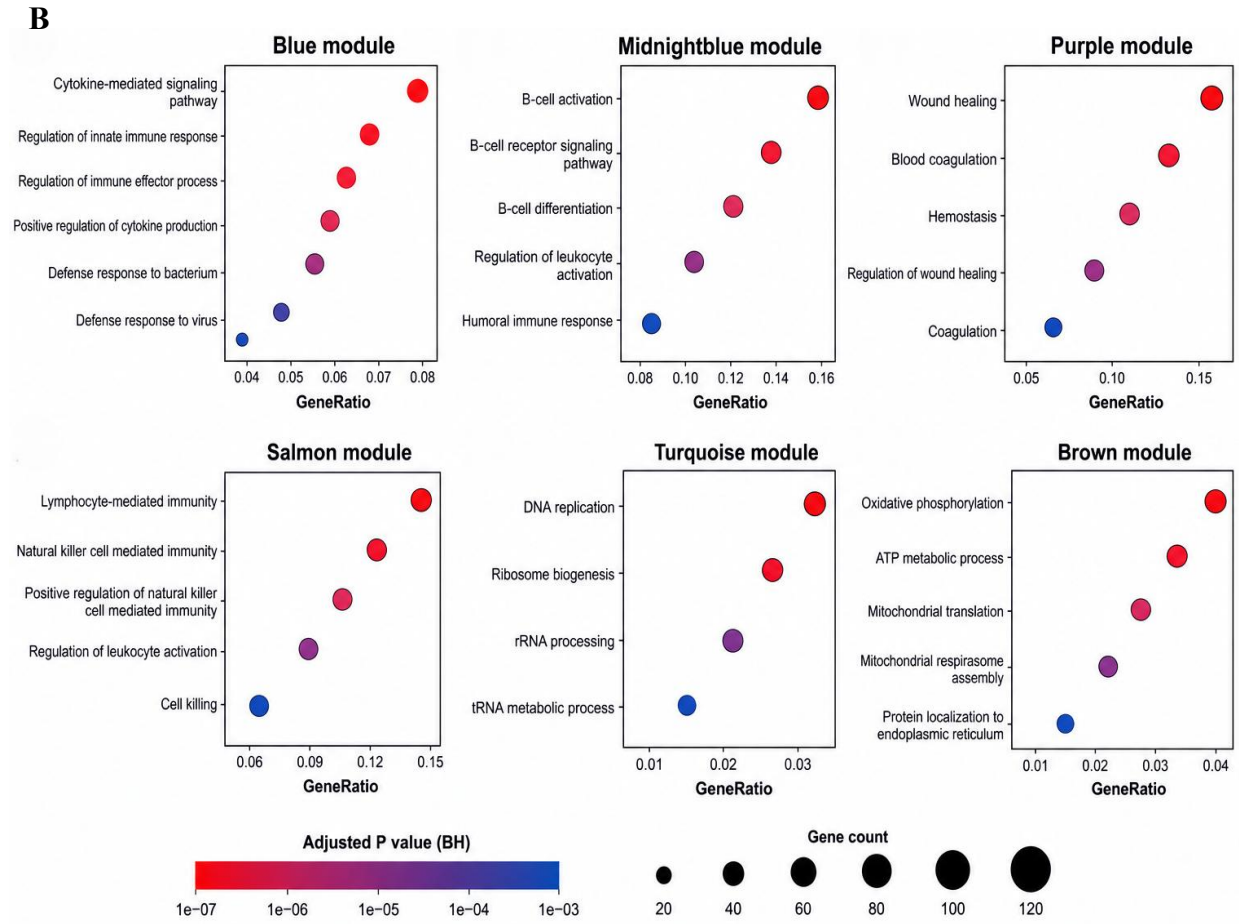
The positively PTB-associated blue module and negatively PTB-associated midnightblue module were enriched for innate and adaptive immune processes, respectively. The blue module was characterized by cytokine-mediated signalling, regulation of innate immune responses, immune effector processes, phagocytosis, antigen processing and presentation, NOD-like receptor signalling, neutrophil extracellular trap formation, lysosome biogenesis, and tuberculosis pathways, indicating activation of host-defence and antimicrobial immune responses. In contrast, the negatively PTB-associated midnightblue module demonstrated enrichment for B-cell activation, humoral immune responses, B-cell receptor signalling, hematopoietic cell lineage, intestinal immune networks for IgA production, and primary immunodeficiency pathways. These findings suggest relative suppression or reduced activity of adaptive B-cell–associated immune programmes during active PTB. (Figure 4A–B).

The purple module was enriched for pathways related to tissue remodelling and vascular responses. Enrichment analyses identified significant associations with platelet activation, complement and coagulation pathways, focal adhesion, integrin signalling, extracellular matrix organization, wound healing, and haemostasis, suggesting coordinated regulation of immune-cell migration, vascular integrity, and tissue repair during active PTB (Figure 4A–B).

In contrast, the negatively PTB-associated turquoise module and positively PTB-associated brown module were enriched for cellular maintenance and metabolic processes. The turquoise module was characterized by enrichment of cell-cycle regulation, DNA replication, homologous recombination, nucleotide excision repair, ribosome biogenesis, and nucleocytoplasmic transport. The brown module showed strong enrichment for mitochondrial respiratory and bioenergetic pathways, including oxidative phosphorylation, aerobic respiration, proton transmembrane transport, ATP metabolic processes, mitochondrial respirasome assembly, mitochondrial translation, proteasome function, and protein processing in the endoplasmic reticulum. Several KEGG neurodegenerative disease pathways, including Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and prion disease, were also enriched; however, these associations were driven predominantly by shared mitochondrial respiratory-chain genes rather than disease-specific neurodegenerative mechanisms. Unlike the turquoise module, which was negatively associated with PTB, the brown module was positively associated with active disease, suggesting enhanced mitochondrial bioenergetic activity and protein homeostasis within PTB-associated transcriptional networks (Figure 4A–B).

Disease Ontology analysis further supported the biological relevance of these transcriptional programmes by demonstrating enrichment for PTB and other infectious diseases within the blue module, primary immunodeficiency disorders within the midnightblue module, hematopoietic and coagulation disorders within the purple module, and mitochondrial metabolic disease categories within the brown and turquoise modules. The concordance across GO-BP, KEGG, and Disease Ontology analyses confirmed that PTB-associated modules captured immune activation, host-defence responses, tissue remodelling, and metabolic dysregulation (Figure 4C). Because these findings suggested coordinated alterations in peripheral immune biology during active PTB, immune-cell deconvolution analysis was subsequently performed to investigate changes in peripheral immune-cell composition.


A
Blue module
Phagocytosis
Tuberculosis
Antigen processing and presentation
NOD−like receptor signaling pathway
Neutrophil extracellular trap formation
Lysosome biogenesis
0.03 0.04 0.05 0.06 0.07 0.08
GeneRatio
Midnightblue module
B−cell receptor signaling pathway
Hematopoietic cell lineage
Intestinal immune network for IgA production
Primary immunodeficiency
0.08 0.12 0.16 0.20
GeneRatio
Purple module
Platelet activation
Complement and coagulation cascades
Focal adhesion
Integrin signaling
ECM−receptor interaction
0.06 0.08 0.10 0.12
GeneRatio
Turquoise module
Cell cycle
DNA replication
Homologous recombination
Nucleotide excision repair
Nucleocytoplasmic transport
0.010 0.015 0.020 0.025 0.030 0.035
GeneRatio
Brown module
Oxidative phosphorylation
Protein processing in endoplasmic reticulum
Proteasome
Alzheimer's disease
Pathways of neurodegeneration − multiple diseases
0.08 0.10 0.12 0.14 0.16
GeneRatio
Adjusted P value (BH)
1e−07 1e−06 1e−05 1e−04 1e−03
Gene count
20 40 60 80 100


B
Blue module
Cytokine-mediated signaling pathway
Regulation of innate immune response
Regulation of immune effector process
Positive regulation of cytokine production
Defense response to bacterium
Defense response to virus
0.04 0.05 0.06 0.07 0.08
GeneRatio
Midnightblue module
B-cell activation
B-cell receptor signaling pathway
B-cell differentiation
Regulation of leukocyte activation
Humoral immune response
0.08 0.10 0.12 0.14 0.16
GeneRatio
Purple module
Wound healing
Blood coagulation
Hemostasis
Regulation of wound healing
Coagulation
0.05 0.10 0.15
GeneRatio
Salmon module
Lymphocyte-mediated immunity
Natural killer cell mediated immunity
Positive regulation of natural killer cell mediated immunity
Regulation of leukocyte activation
Cell killing
0.06 0.09 0.12 0.15
GeneRatio
Turquoise module
DNA replication
Ribosome biogenesis
rRNA processing
tRNA metabolic process
0.01 0.02 0.03
GeneRatio
Brown module
Oxidative phosphorylation
ATP metabolic process
Mitochondrial translation
Mitochondrial respirasome assembly
Protein localization to endoplasmic reticulum
0.01 0.02 0.03 0.04
GeneRatio
Adjusted P value (BH)
1e−07 1e−06 1e−05 1e−04 1e−03
Gene count
20 40 60 80 100 120

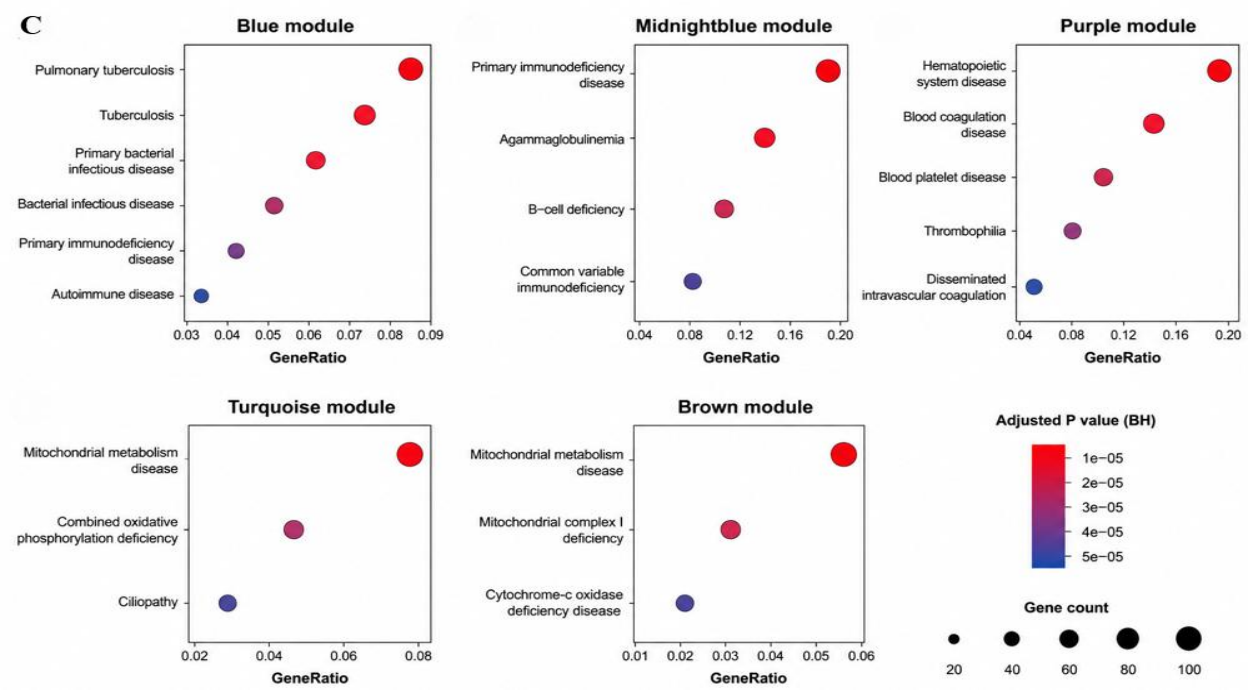


Figure 4. Functional enrichment analysis identifies biologically distinct PTB-associated co-expression modules.

(A) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the disease-associated co-expression modules identified by WGCNA. **(B)** Gene Ontology (GO) Biological Process enrichment analysis highlighting the predominant biological functions represented within each module. **(C)** Disease Ontology (DO) enrichment analysis demonstrating the disease relevance of individual co-expression modules. Dot size represents the number of enriched genes, dot colour indicates the Benjamini–Hochberg (BH)-adjusted *P* value, and the x-axis represents the gene ratio (GeneRatio). In combination, the enrichment analyses demonstrate that PTB-associated co-expression modules are functionally specialized, encompassing innate immune responses, adaptive B-cell immunity, coagulation and extracellular matrix organization, cell-cycle and DNA replication processes, mitochondrial metabolism, and neurodegeneration-related pathways.

Disease-associated co-expression modules exhibit distinct immune-cell associations

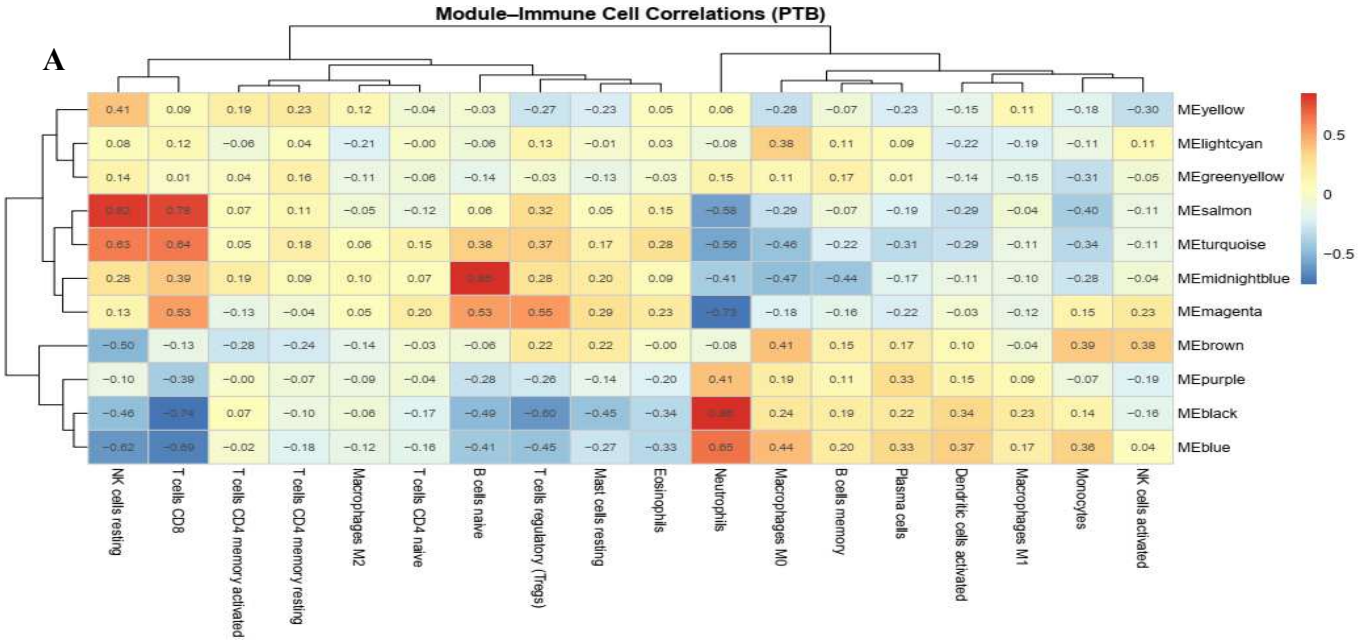
Immune-cell composition was estimated using CIBERSORTx, and module eigengenes were correlated with inferred immune-cell fractions to define the immunological context of PTB-associated co-expression modules. Significant module–immune cell associations were identified using Spearman correlation analysis with FDR correction, revealing extensive interactions between disease-associated transcriptional programmes and both innate and adaptive immune-cell populations (Figure 5A).

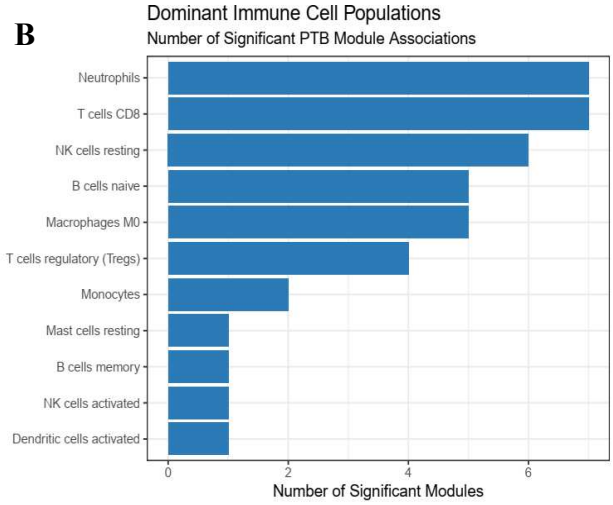
Across the PTB-associated network, neutrophils and $CD8^+$ T cells emerged as the dominant immune populations associated with disease-related transcriptional programmes. Neutrophils were linked to the greatest number of significant co-expression modules and exhibited the highest overall correlation strength, followed by $CD8^+$ T cells, resting natural killer (NK) cells, naive B cells, and macrophages M0 (Figure 5B–C). Additional associations involving regulatory T cells, monocytes, mast cells, activated dendritic cells, activated NK cells, and memory B cells indicated that PTB-associated programmes span multiple innate and adaptive immune compartments rather than a single immune lineage.

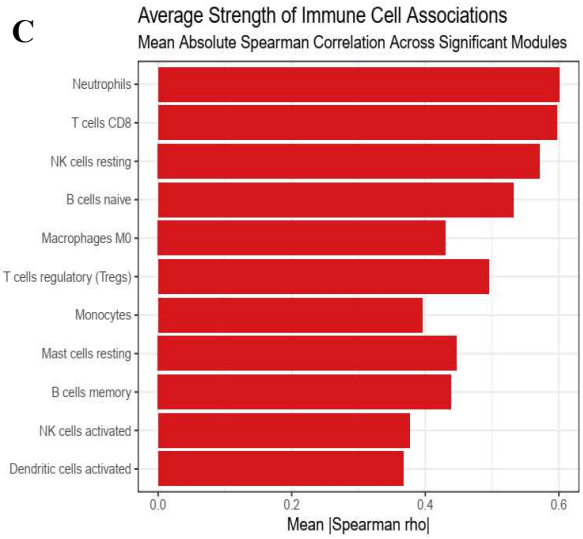
Representative module–immune cell relationships further illustrated the functional specialization of individual co-expression programmes. The MEblack module exhibited a strong positive association with neutrophils ($\rho = 0.86$, FDR $= 2.37 \times 10^{-13}$) together with a strong negative association with $CD8^+$ T cells ($\rho = -0.74$, FDR $= 2.28 \times 10^{-8}$), consistent with a transcriptional programme associated with neutrophil-dominant immune responses accompanied by reduced cytotoxic T-cell representation (Figure 5D). Similarly, the MEblue module demonstrated positive correlations with neutrophils ($\rho = 0.66$, FDR $= 6.24 \times 10^{-6}$) and negative correlations with $CD8^+$ T cells ($\rho = -0.69$, FDR $= 6.49 \times 10^{-7}$), indicating that multiple PTB-associated modules are associated with similar neutrophil-dominated inflammatory profiles (Figure 5E). In contrast, the MEmidnightblue module was most strongly associated with naive B cells ($\rho = 0.85$, FDR $= 4.60 \times 10^{-13}$), highlighting an adaptive immune programme centred on B-cell biology (Figure 5F). Finally, the MEturquoise module showed positive associations with both resting NK cells ($\rho = 0.63$, FDR

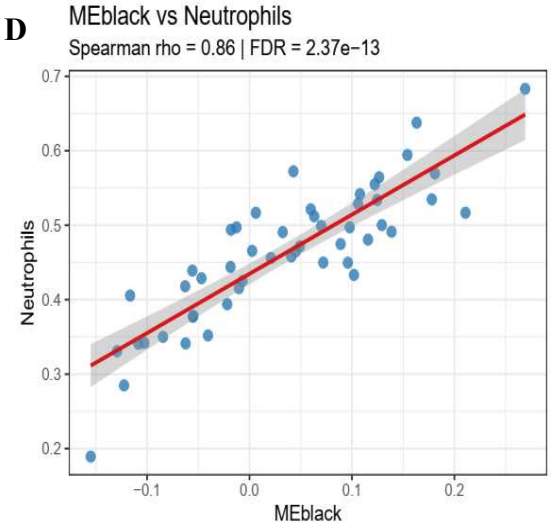
= 2.03 × 10^{-5}) and $CD8^+$ T cells (ρ = 0.64, FDR = 1.12 × 10^{-5}), reflecting coordinated cytotoxic lymphocyte responses within modules negatively associated with active PTB (Figure 5G). Complete module–immune cell correlation coefficients and statistically significant associations following false discovery rate correction are provided in Tables S9–S10.

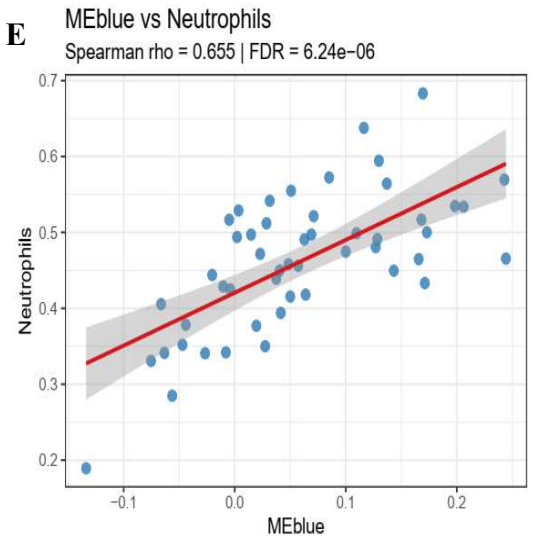
Overall, these findings demonstrate that PTB-associated co-expression modules represent biologically distinct immune programmes linked to specific immune-cell populations rather than generalized inflammatory activation. Modules positively associated with active PTB predominantly correlated with neutrophil-driven innate immune responses, whereas negatively associated modules were more closely related to cytotoxic lymphocytes and adaptive immune populations. These module–immune cell relationships link peripheral transcriptional networks to immune-cell alterations during active tuberculosis. Machine-learning analyses were subsequently performed to identify the most diagnostically informative features within these disease-associated co-expression modules.

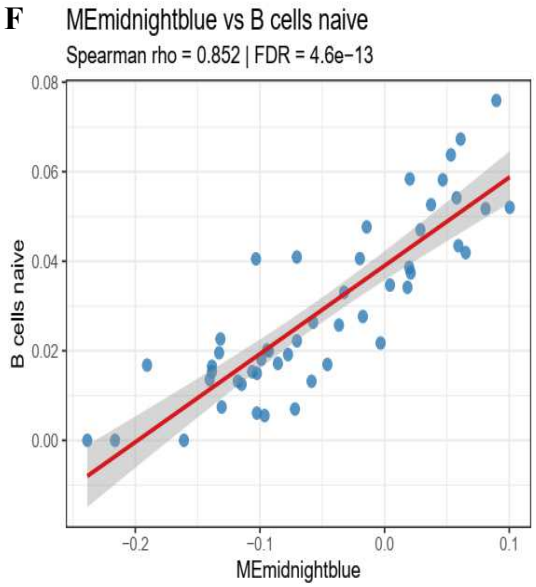



B
Dominant Immune Cell Populations
Number of Significant PTB Module Associations
Neutrophils
T cells CD8
NK cells resting
B cells naive
Macrophages M0
T cells regulatory (Tregs)
Monocytes
Mast cells resting
B cells memory
NK cells activated
Dendritic cells activated
0
2
4
6
Number of Significant Modules


C
Average Strength of Immune Cell Associations
Mean Absolute Spearman Correlation Across Significant Modules
Neutrophils
T cells CD8
NK cells resting
B cells naive
Macrophages M0
T cells regulatory (Tregs)
Monocytes
Mast cells resting
B cells memory
NK cells activated
Dendritic cells activated
0.0
0.2
0.4
0.6
Mean |Spearman rho|


D
MEblack vs Neutrophils
Spearman rho = 0.86 | FDR = 2.37e−13
Neutrophils
MEblack


E
MEblue vs Neutrophils
Spearman rho = 0.655 | FDR = 6.24e−06
Neutrophils
MEblue


F
MEmidnightblue vs B cells naive
Spearman rho = 0.852 | FDR = 4.6e−13
B cells naive
MEmidnightblue

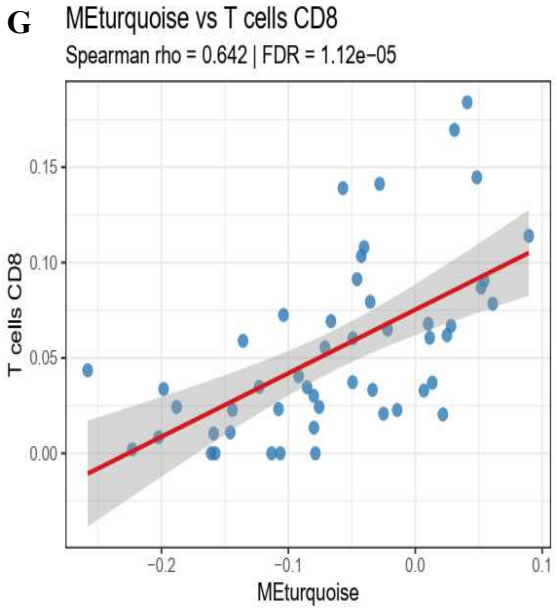

G
MEturquoise vs T cells CD8
Spearman rho = 0.642 | FDR = 1.12e−05
T cells CD8
MEturquoise

Figure 5. Disease-associated co-expression modules exhibit distinct associations with peripheral immune-cell populations. (A) Heatmap showing Spearman correlation coefficients between WGCNA module eigengenes and immune-cell fractions estimated using CIBERSORTx in the PTB discovery cohort. Significant positive and negative module–immune cell associations are represented according to the correlation coefficient. **(B)** Number of significant module associations identified for each immune-cell population, highlighting the immune-cell types most extensively represented across the PTB-associated co-expression network. **(C)** Overall correlation strength of immune-cell populations across significant module–immune associations, indicating the relative contribution of individual immune-cell populations to the transcriptional network. (D–G) Representative scatterplots illustrating the strongest module–immune cell relationships, including MEblack–Neutrophils (D)**,** MEblue–Neutrophils (E)**,** MEmidnightblue–Naive B cells (F)**,** and MEturquoise–$CD8^+$ T cells (G)**.**

Nested machine-learning identifies and externally validates a robust biomarker signature for pulmonary tuberculosis

A two-stage machine-learning framework was applied to prioritize diagnostically informative biomarkers from PTB-associated co-expression modules. Candidate genes were evaluated using nested cross-validation to minimize information leakage and optimism bias. Within each outer training fold, LASSO regression identified informative features, and genes consistently selected across iterations were ranked by selection frequency (Figure 6A). The selected biomarkers were subsequently used to train the final locked Random Forest classifier.

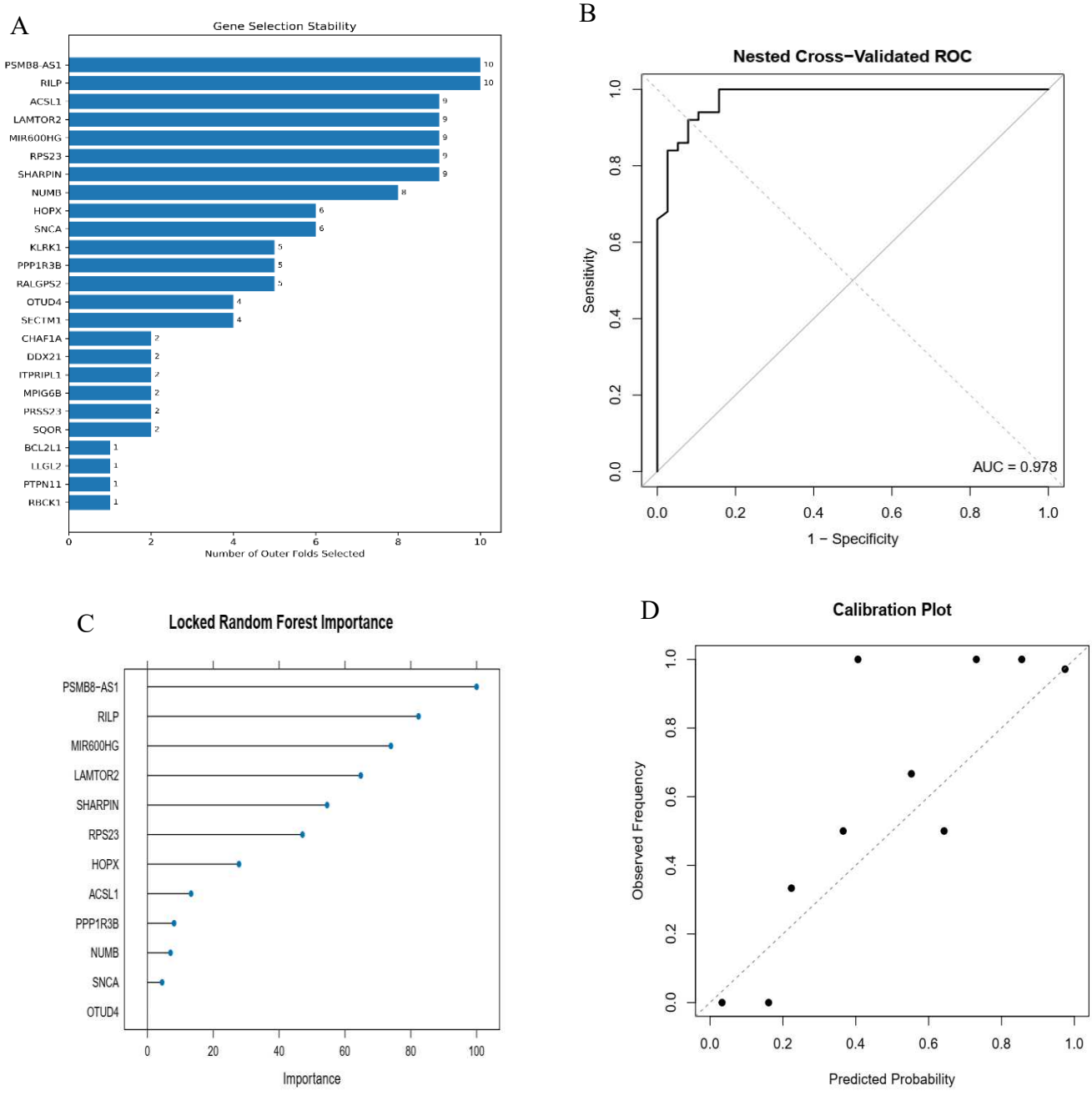
Nested cross-validation consistently identified a compact subset of informative biomarkers. Training the final locked model on the complete discovery cohort produced a 12-gene diagnostic signature comprising NUMB, ACSL1, PPP1R3B, PSMB8-AS1, LAMTOR2, SNCA, SHARPIN, RILP, RPS23, MIR600HG, HOPX, and OTUD4, which constituted the final diagnostic classifier (Table S11). Performance of the classifier was evaluated using pooled out-of-fold predictions generated during the outer cross-validation.

The classifier demonstrated excellent discrimination between PTB and healthy controls, achieving an AUC of approximately 0.99, with correspondingly high sensitivity and specificity (Figure 6B; Table S12). Calibration analysis demonstrated close agreement between predicted and observed outcomes, with a low Brier score, calibration intercept near zero, and calibration slope approaching the ideal value (Figure 6D; Table S13). Decision curve analysis demonstrated positive net benefit across a broad range of threshold probabilities, supporting potential clinical utility (Figure 6E). Furthermore, permutation testing (1,000 permutations) confirmed that classifier performance was significantly greater than expected by chance (Figure 6F).

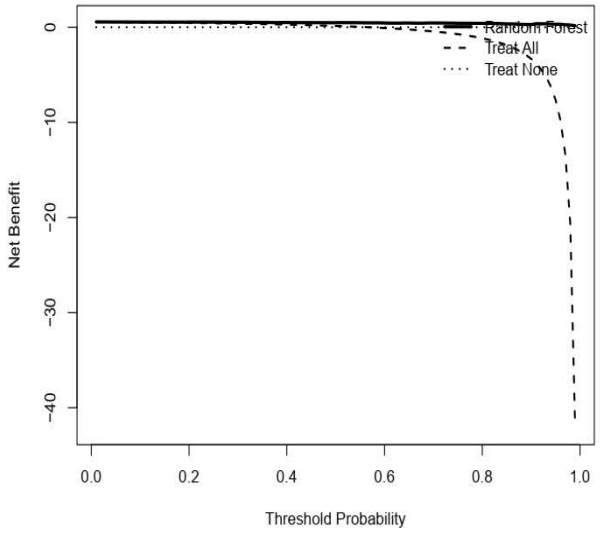
Variable importance analysis identified PSMB8-AS1 as the strongest contributor to classification performance, followed by RILP, MIR600HG, LAMTOR2, SHARPIN, RPS23, HOPX, ACSL1, PPP1R3B, NUMB, SNCA, and OTUD4, indicating that a small subset of network-derived genes accounted for most predictive performance (Figure 6C; Table S14). The locked Random Forest classifier was then applied, without retraining, to two independent external RNA-sequencing cohorts. In GSE107991, the classifier maintained excellent diagnostic performance, achieving an AUC of 0.895 (95% CI: 0.785–1.000), an overall accuracy of 84.8%, sensitivity of 85.7%, specificity of 83.3%, and a PR-AUC of 0.947 (Figure 6G). Calibration analysis demonstrated good agreement between predicted probabilities and observed outcomes (Figure 6I), while decision curve analysis indicated sustained positive net benefit across clinically relevant threshold probabilities (Figure 6K). Detailed prediction results, confusion matrices, calibration statistics, and supporting performance metrics for the independent external validation cohorts are provided in Tables S16–S17.

External validation in GSE279063 demonstrated moderate discriminatory performance, with an AUC of 0.780 (95% CI: 0.584–0.976) and a PR-AUC of 0.806 (Figure 6H). Although calibration remained acceptable (Figure 6J), reduced sensitivity led to misclassification of many PTB samples, lowering overall accuracy despite preservation of ranking performance. Decision curve analysis likewise indicated reduced clinical utility compared with GSE107991 (Figure 6L). Precision–recall curves for the external validation

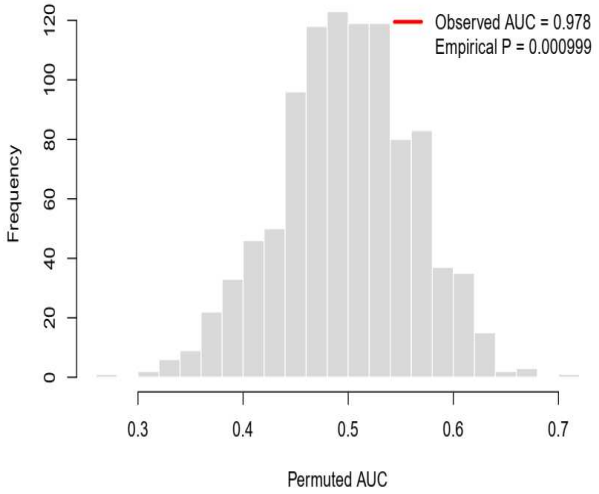
cohorts are presented in Figures S5–S6. These findings as a whole demonstrate that integrating WGCNA with nested machine learning identified a compact and biologically meaningful PTB biomarker signature with excellent internal performance and reproducible external validation, while highlighting the impact of dataset heterogeneity on classifier performance.



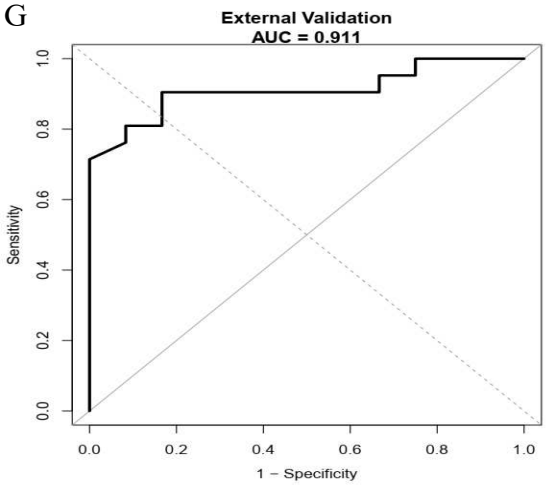
E



F

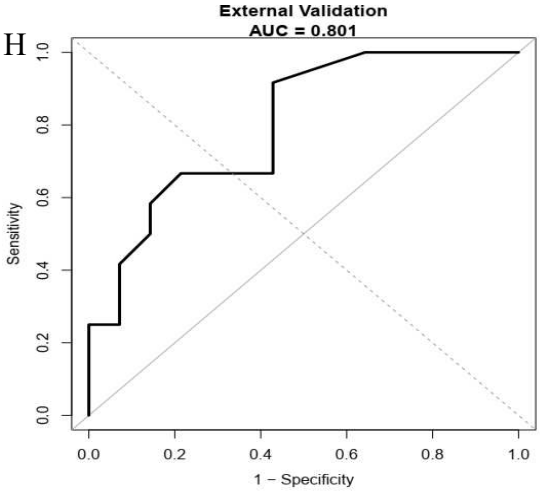


G



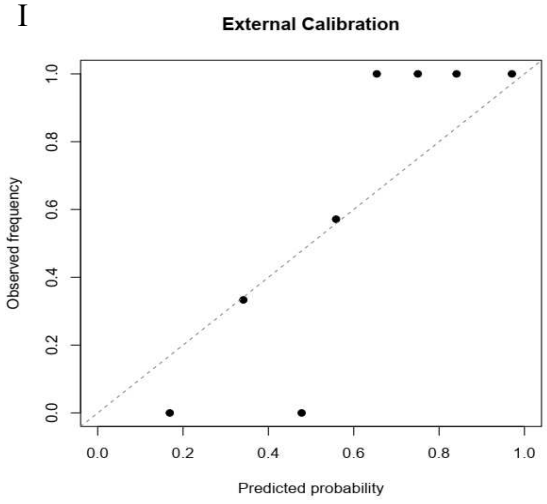
GSE107991

H



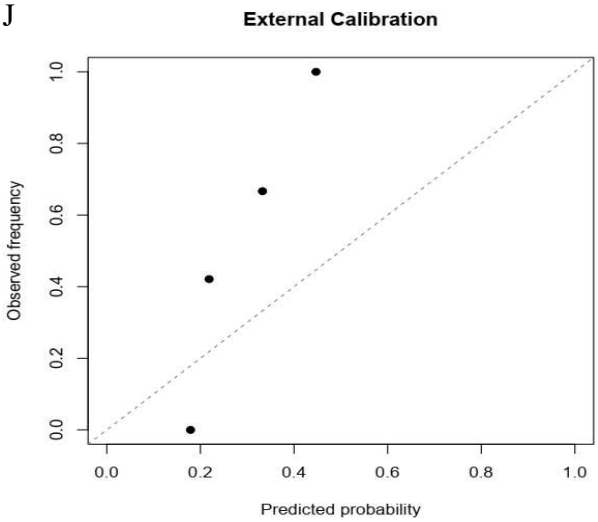
GSE279063

I



GSE107991

J



GSE279063

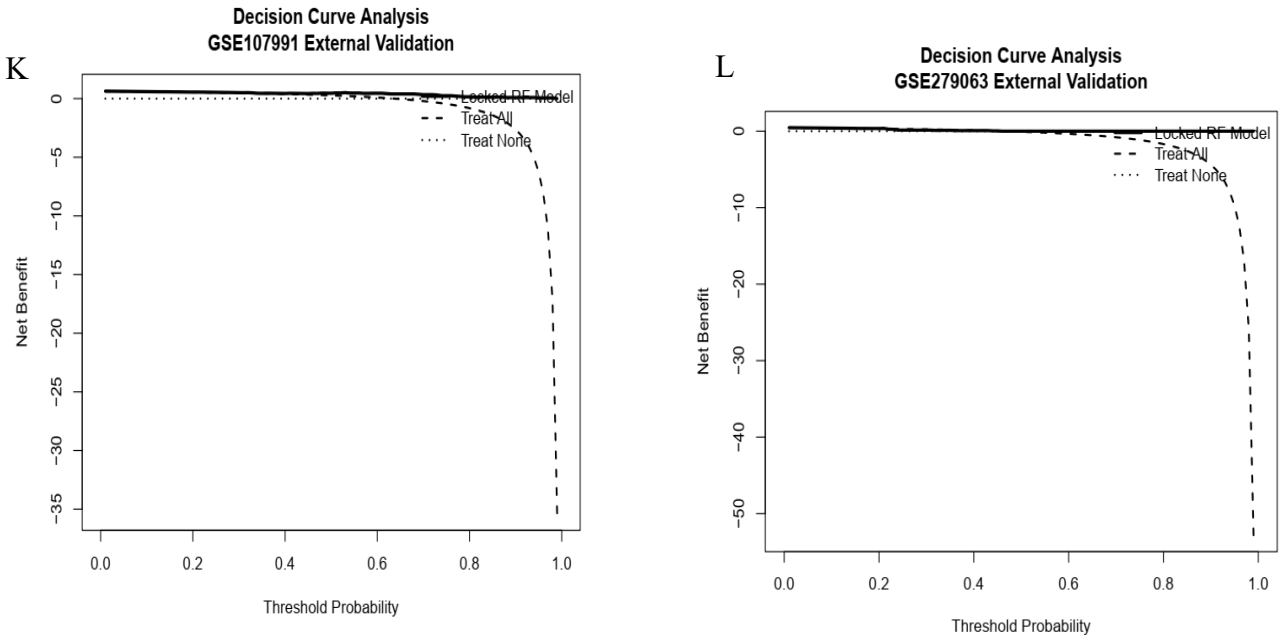


Figure 6. Nested machine-learning framework identifies and externally validates a robust biomarker signature for pulmonary tuberculosis.

(A) Gene selection stability across nested cross-validation showing the frequency with which candidate biomarkers were retained by LASSO across the outer cross-validation folds. (B) Receiver operating characteristic (ROC) curve generated from pooled out-of-fold predictions demonstrating the discrimination performance of the nested Random Forest classifier. (C) Variable importance ranking of the final locked Random Forest model identifying the most influential biomarkers contributing to PTB classification. (D) Calibration plot showing agreement between predicted probabilities and observed outcomes in the discovery cohort. (E) Decision curve analysis demonstrating the net clinical benefit of the locked classifier across a range of threshold probabilities. (F) Permutation test (1,000 permutations) confirming that model performance was significantly greater than expected by chance. (G) Receiver operating characteristic (ROC) curve of the locked classifier in the independent external validation cohort GSE107991. (H) Receiver operating characteristic (ROC) curve of the locked classifier in the independent external validation cohort GSE279063. (I) Calibration plot for GSE107991. (J) Calibration plot for GSE279063. (K) Decision curve analysis for GSE107991. (L) Decision curve analysis for GSE279063.

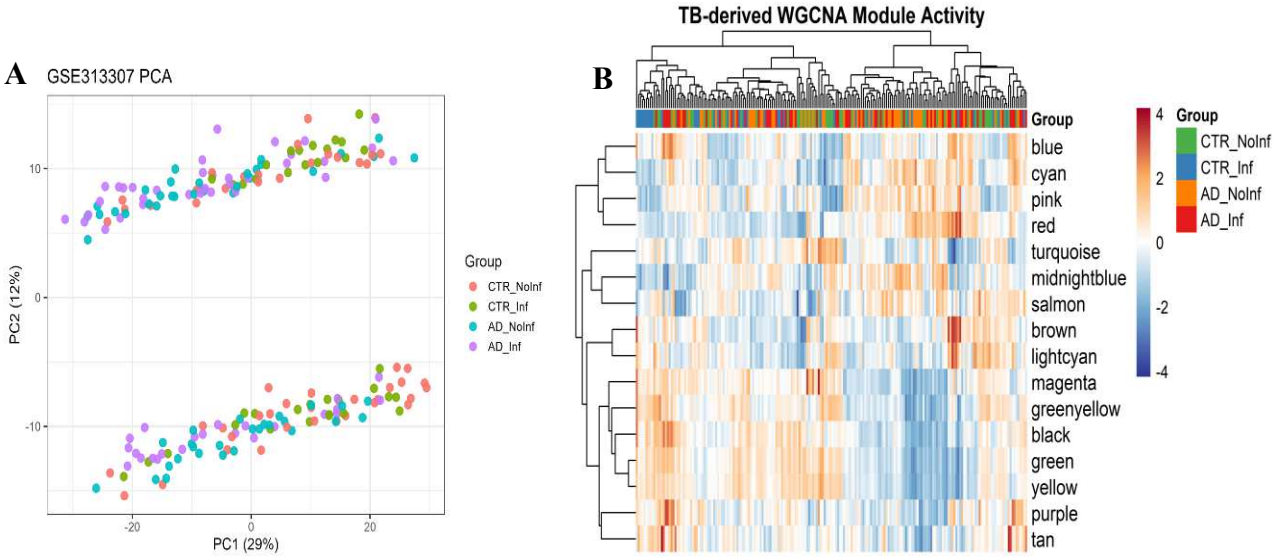
PTB-derived immune co-expression modules exhibit concordant transcriptional activity in an independent Alzheimer's disease cohort

PTB-derived transcriptional programmes were evaluated in an independent AD cohort comprising individuals with and without concurrent respiratory infection (GSE313307). PCA demonstrated substantial overlap among the four clinical groups, indicating broadly similar global transcriptomic profiles while providing a basis for evaluating PTB-derived module activity in an independent disease context (Figure 7A). Module activity heatmaps revealed distinct patterns across the four clinical groups, indicating that PTB-derived transcriptional programmes remained detectable within the AD cohort despite differences in disease phenotype (Figure 7B). Quantitative comparison of module eigengene scores further confirmed this observation, with 11 of the 12 evaluated co-expression modules exhibiting significant differences across the four clinical groups, including cyan, greenyellow, blue, magenta, pink, tan, turquoise, lightcyan, midnightblue, purple, and yellow (FDR < 0.05), whereas the black module did not show significant variation (Figure7C). These findings indicate that the majority of PTB-derived modules exhibited differential activity across the AD cohort.

Two-way ANOVA demonstrated that module activity was influenced by AD diagnosis, respiratory infection, or their interaction depending on the module, indicating that PTB-derived transcriptional programmes respond differentially to these biological factors rather than reflecting a single disease effect. Notably, modules previously associated with innate immune activation, adaptive immune responses, cellular proliferation, and metabolic regulation in the PTB discovery cohort exhibited consistent alterations across the AD comparison groups, indicating that these programmes remained active in the AD cohort rather than representing TB-specific signatures. Complete two-way ANOVA results, module-level statistical summaries, pairwise comparisons, and group mean eigengene values are provided in Tables S18–S21. Similarly, detailed statistical results for the validated neuroimmune biomarkers are presented in Tables S22–S24.

To determine whether the ML-prioritized candidate biomarkers also exhibited differential expression, expression of the validated neuroimmune genes was examined across the four clinical groups. Boxplot analysis demonstrated consistent differential expression of HOPX, SNCA, SHARPIN, PSMB8-AS1, and LAMTOR2, confirming that these biomarkers remained significantly associated with disease status in the independent AD cohort (Figure 7D). The concordant activity of these network-derived biomarkers supports the module-level findings and demonstrates that PTB-derived signatures remain biologically relevant in a distinct neurodegenerative context.

Collectively, these findings demonstrate that PTB-derived immune co-expression modules and their associated neuroimmune biomarkers showed concordant activity in an independent AD cohort, supporting the existence of shared peripheral immune transcriptional programmes across chronic infection and neurodegeneration. These validated transcriptional programmes were subsequently mapped onto the adult human brain using the AHBA to characterize their regional expression patterns.



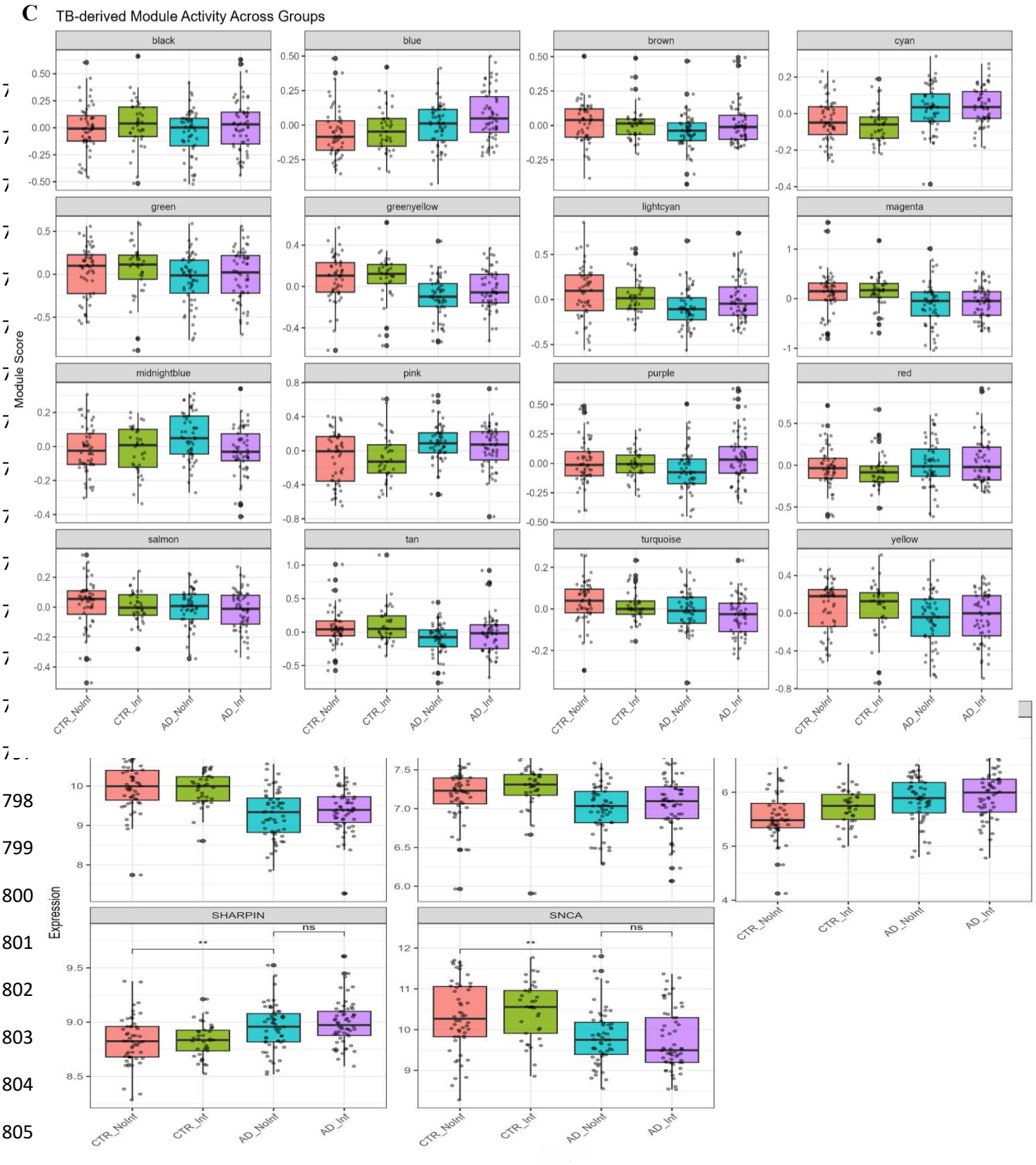


Figure 7. Validation of PTB-derived immune co-expression modules and neuroimmune biomarkers in an independent AD cohort. (A) Principal component analysis (PCA) of the independent validation

cohort (GSE313307), comprising AD and cognitively normal control (CTR) individuals with and without concurrent infection. The PCA demonstrates the overall transcriptomic structure of the validation dataset and the overlap among the four clinical groups. **(B)** Heatmap showing the activity of PTB-derived weighted gene co-expression network (WGCNA) modules across individual samples in the validation cohort. Rows represent module eigengenes, columns represent individual samples, and colors indicate standardized module activity (Z-scores). Hierarchical clustering demonstrates coordinated PTB-derived module activity across the AD cohort. **(C)** Boxplots of PTB-derived module eigengene scores across the four clinical groups. Each panel represents the activity of a single co-expression module, illustrating quantitative differences in module activity among control and AD samples with and without infection. Statistical significance was assessed using one-way ANOVA followed by false discovery rate (FDR) correction. **(D)** Expression profiles of independently validated PTB-derived neuroimmune biomarkers (HOPX, LAMTOR2, PSMB8-AS1, SHARPIN, and SNCA) across the four clinical groups. Boxplots display variance-stabilized (VST) expression values, with statistical significance determined following FDR correction. These validated biomarkers demonstrate differential expression in the independent AD cohort, supporting concordant transcriptional activity of PTB-derived molecular signatures across distinct disease contexts.

PTB-derived co-expression networks demonstrate partial preservation in Alzheimer's disease

Although PTB-derived modules exhibited significant activity within the AD cohort, differential module activity alone does not establish whether the underlying co-expression architecture is retained across diseases. Therefore, we performed formal WGCNA module-preservation analysis using GSE313307 to determine whether PTB-derived transcriptional programmes remained organized as coherent gene networks in AD. Gene overlap between the PTB discovery network and the AD dataset was high, with retention exceeding 75% for most major modules and approaching 90% for the green and magenta modules, enabling robust preservation analysis across all evaluated networks (Table S25). Several modules demonstrated moderate preservation in AD, with green (Zsummary = 8.96), magenta (8.95), and yellow (8.77) exhibiting

the strongest preservation statistics among all evaluated modules (Figure 8; Table S26). Additional modules including black (4.40), blue (3.50), red (2.94), brown (2.87), and turquoise (2.21) also showed evidence of preservation, whereas greenyellow, purple, pink, midnightblue, tan, cyan, salmon, and lightcyan demonstrated weak or absent preservation (Zsummary < 2). The preservation of module topology across independent diseases suggests that these networks capture biologically coherent molecular systems rather than dataset-specific transcriptional signatures, supporting the existence of shared transcriptional organization across PTB and AD.

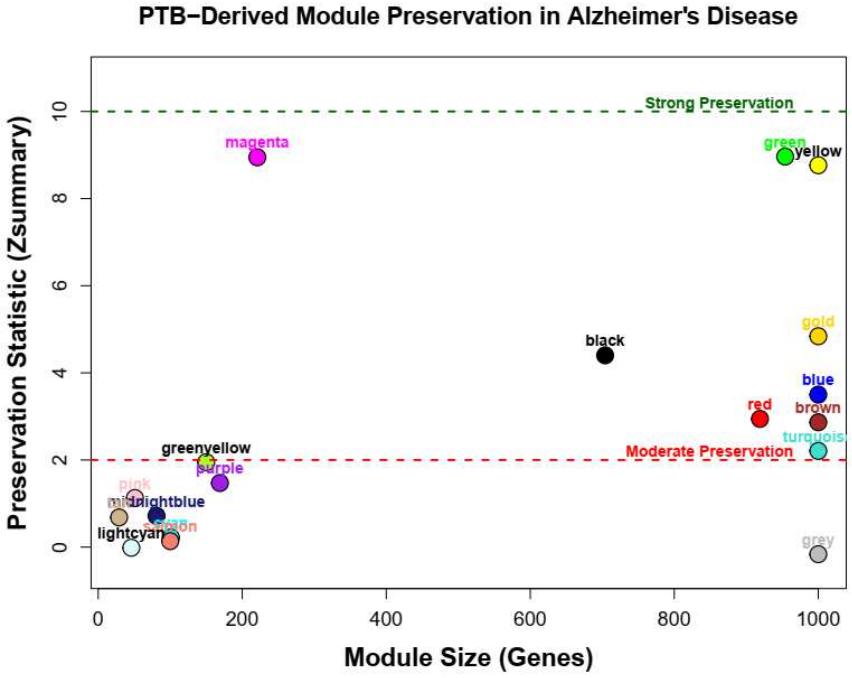


Figure 8. Preservation of PTB-derived co-expression modules in Alzheimer's disease. Module preservation analysis was performed using WGCNA modulePreservation with the PTB discovery network as the reference and GSE313307 as the test dataset. Each point represents a PTB-derived module coloured according to its original WGCNA assignment. The dashed red line indicates the threshold for moderate preservation (Zsummary = 2), whereas the dashed green line indicates strong preservation (Zsummary = 10). Green, magenta, and yellow modules exhibited the strongest preservation, while black, blue, brown,

red, and turquoise demonstrated moderate preservation, indicating partial conservation of PTB-associated co-expression architecture in Alzheimer's disease. Full preservation statistics are provided in Table S25.

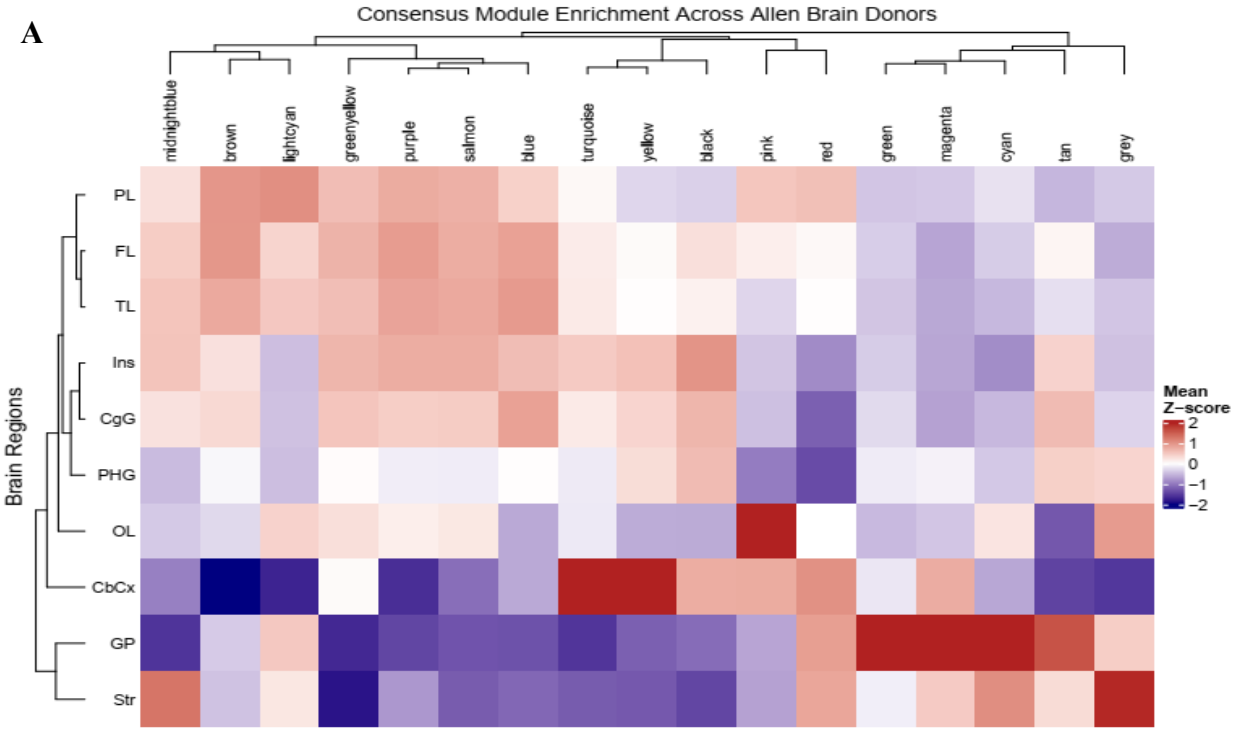
PTB-derived neuroimmune co-expression modules exhibit reproducible regional brain transcriptomic organization in the adult human brain

All PTB-derived co-expression modules were mapped onto AHBA RNA-sequencing data to characterize their regional brain expression patterns. This analysis was performed independently of the AD preservation results to determine whether PTB-associated transcriptional programmes exhibit reproducible neuroanatomical organization within the adult human brain. Consensus analysis across two independent donors revealed reproducible regional expression patterns, with several modules showing preferential enrichment in specific brain regions rather than uniform expression throughout the brain (Figure 9A). Similar spatial expression patterns were observed at the anatomical substructure level, where module-specific expression patterns remained consistent at higher anatomical resolution, further supporting the reproducibility of the observed regional expression patterns. (Figure 9B).

Statistical analysis (Table S27) confirmed significant regional heterogeneity in module expression across major brain regions. Consensus peak-region assignments and standardized regional expression profiles identified distinct anatomical preferences among PTB-derived co-expression modules (Tables S28–S29). Several modules, including green, magenta, cyan, tan, lightcyan, and red, exhibited their highest expression within the globus pallidus, whereas the yellow, turquoise, pink, and black modules were preferentially enriched within the cerebellar cortex. The brown and purple modules showed maximal expression in the frontal lobe, while the salmon, midnightblue, blue, and greenyellow modules demonstrated preferential localization to the insula. In contrast, the grey module exhibited its strongest expression within the striatum, indicating that distinct PTB-derived neuroimmune transcriptional programmes display distinct regional expression preferences rather than converging within a single brain region. Analysis at the anatomical substructure level demonstrated that similar regional expression patterns were observed at higher

anatomical resolution, with individual cortical and subcortical substructures retaining module-specific enrichment profiles. The concordance between regional and subregional expression patterns across both Allen Brain donors supports the reproducibility of the observed regional expression organization and argues against donor-specific effects.

These findings demonstrate that PTB-derived neuroimmune co-expression modules exhibit reproducible and anatomically structured expression patterns within the adult human brain. The preferential localization of distinct transcriptional modules to specific brain regions supports the existence of anatomically organized regional expression profiles that may provide an anatomical context for understanding how chronic peripheral immune activation relates to region-specific neuroimmune organization. Inter-donor reproducibility of the observed neuroanatomical patterns was subsequently evaluated by direct comparison of the two independent AHBA donors.



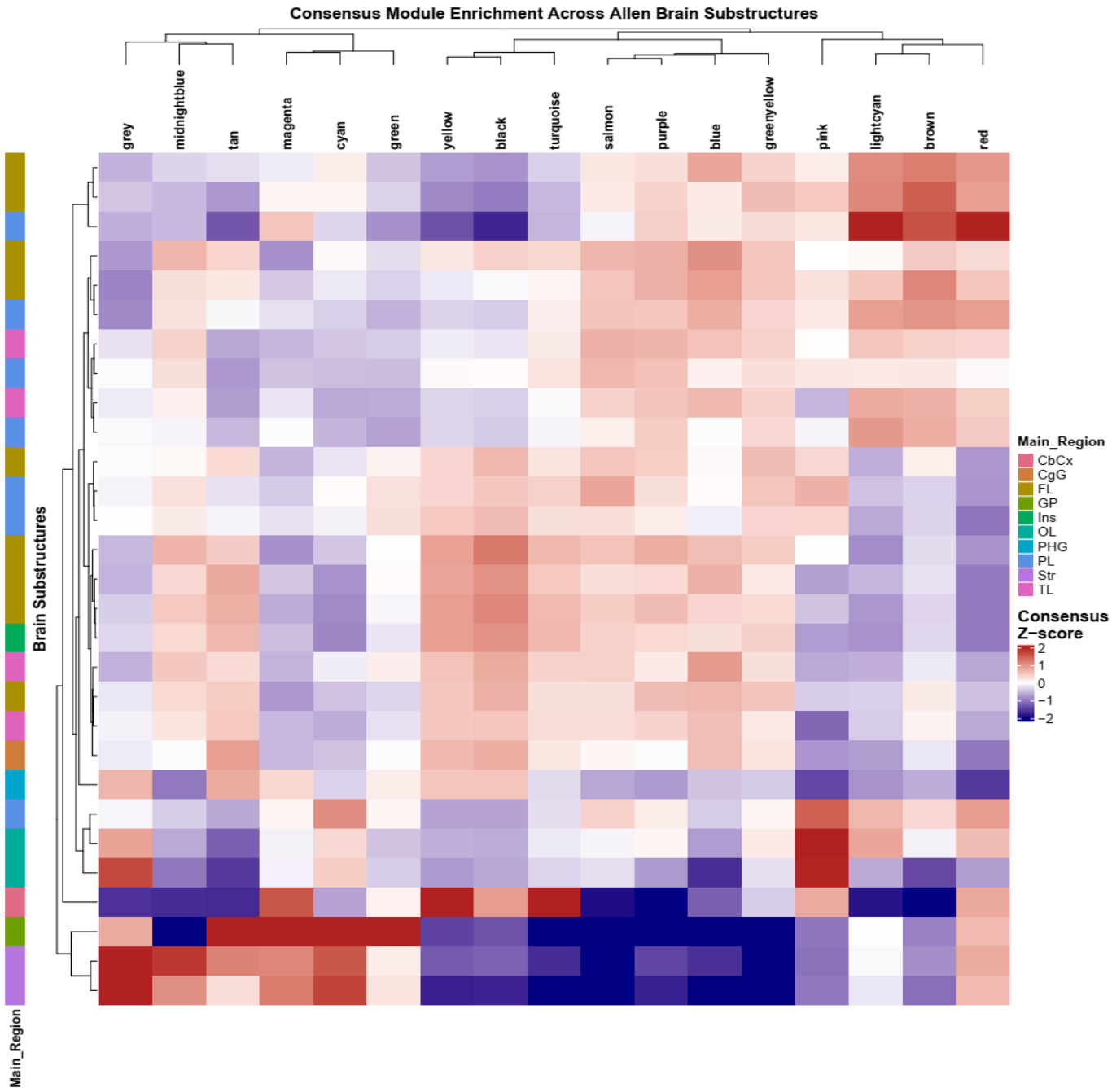


Figure 9. Regional expression patterns of PTB-derived neuroimmune co-expression modules in the adult human brain.

(A) Consensus heatmap showing standardized (Z-score) expression of PTB-derived WGCNA modules across major brain regions in the Allen Human Brain Atlas, generated by integrating two independent adult donors. Rows represent anatomical brain regions and columns represent PTB-derived co-expression modules. Hierarchical clustering was performed using consensus regional expression profiles.

(B) Consensus heatmap showing standardized module expression across anatomical brain substructures in the Allen Human Brain Atlas. Hierarchical clustering illustrates the spatial expression patterns of PTB-derived neuroimmune transcriptional programmes at higher anatomical resolution. Expression values are displayed as standardized Z-scores.

Reproducibility of PTB-derived neuroimmune modules across independent Allen Brain Atlas donors

Regional expression profiles of PTB-derived co-expression modules were compared between the two independent AHBA donors to evaluate the reproducibility of the observed neuroanatomical organization. Most modules demonstrated strong concordance in regional expression, indicating that the neuroanatomical patterns identified in the consensus analysis were largely reproducible across donors rather than donor-specific (Figure 10A). Modules including green**,** turquoise**,** grey**,** purple**,** greenyellow**,** yellow**,** magenta**,** brown**,** and cyan exhibited high inter-donor Pearson correlations, whereas lightcyan**,** black, and midnightblue showed comparatively lower reproducibility (Figure 10A; Table S30).

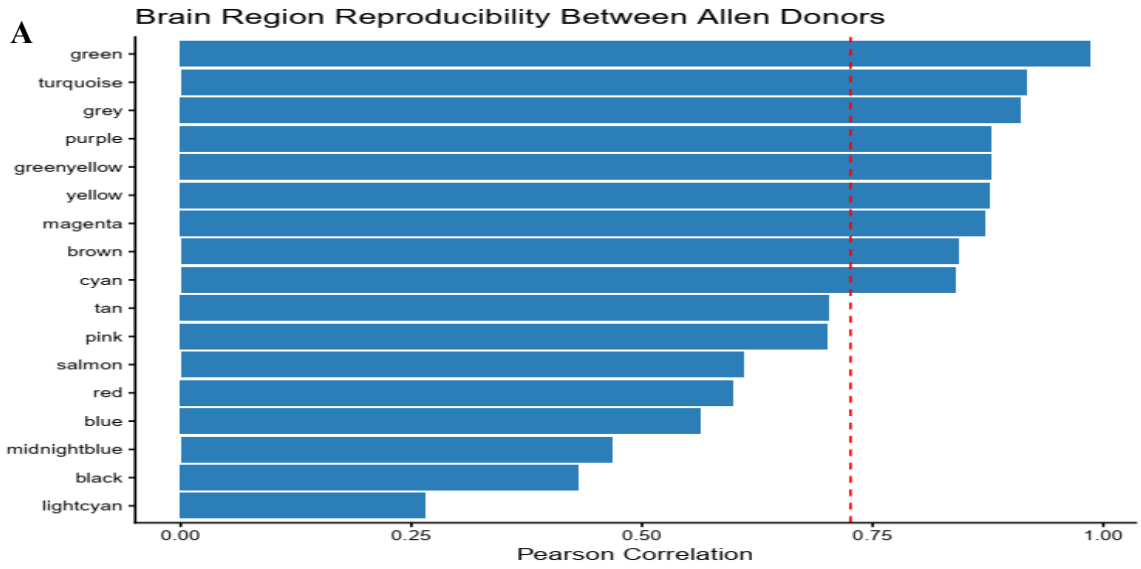
To further characterize the consistency of regional expression patterns, pairwise correlations of regional module expression profiles revealed consistent module clustering across donors, with positively correlated modules grouping together and negatively correlated modules occupying distinct neuroanatomical territories (Figure 10B).

Quantitative reproducibility analyses confirmed moderate-to-high agreement between donors across most modules (Tables S30–S31). Consistent module–brain region associations were further supported by regional enrichment rankings, anatomical substructure analyses, and regional specificity metrics (Tables S32–S36; Figure S7). Notably, modules exhibiting the strongest regional specificity also showed the highest inter-donor reproducibility.

Collectively, these findings demonstrate that the regional expression patterns of PTB-derived immune co-expression modules are highly reproducible across independent Allen Brain Atlas donors. The consistency

of regional enrichment patterns provides independent validation of the consensus neuroanatomical analysis and supports the existence of stable module-specific expression profiles across donors.

Having established that PTB-derived neuroimmune transcriptional programmes are reproducibly localized within specific brain regions across independent donors, we next integrated these findings with the peripheral blood and AD analyses to construct a unified systems-level framework integrating chronic peripheral immune activation with anatomically organized neuroimmune networks.



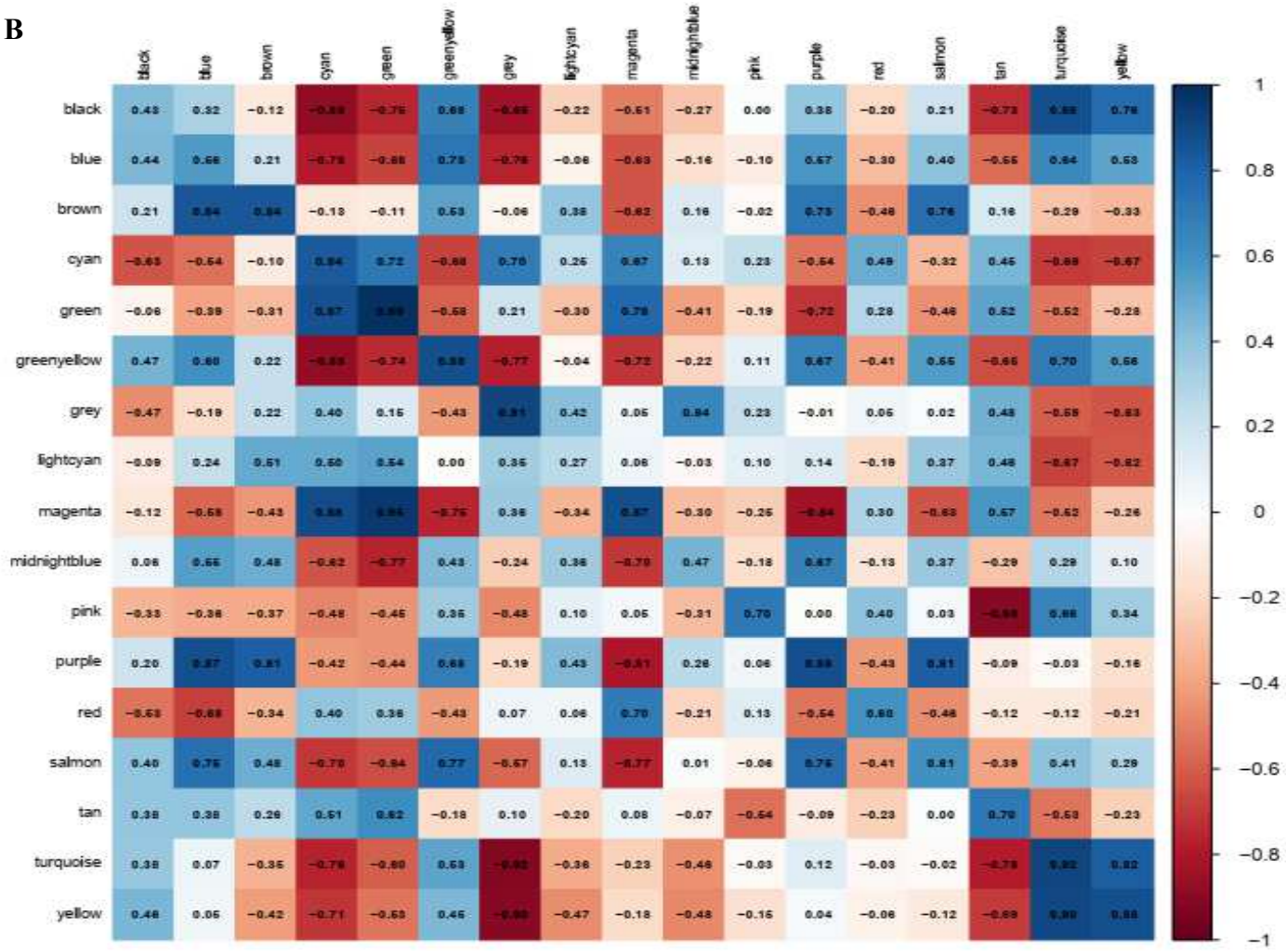


Figure 10. Reproducibility of PTB-derived neuroimmune module expression across independent Allen Human Brain Atlas donors.

(A) Pearson correlation coefficients describing the reproducibility of regional module expression between the two independent Allen Human Brain Atlas donors for each PTB-derived co-expression module. **(B)** Correlation heatmap showing pairwise relationships among regional expression profiles of PTB-derived co-expression modules based on consensus Allen Brain Atlas data. Positive and negative correlations indicate modules exhibiting similar or distinct regional expression patterns across brain regions.

Integrated neuroanatomical contextualization identifies robust PTB-derived neuroimmune modules

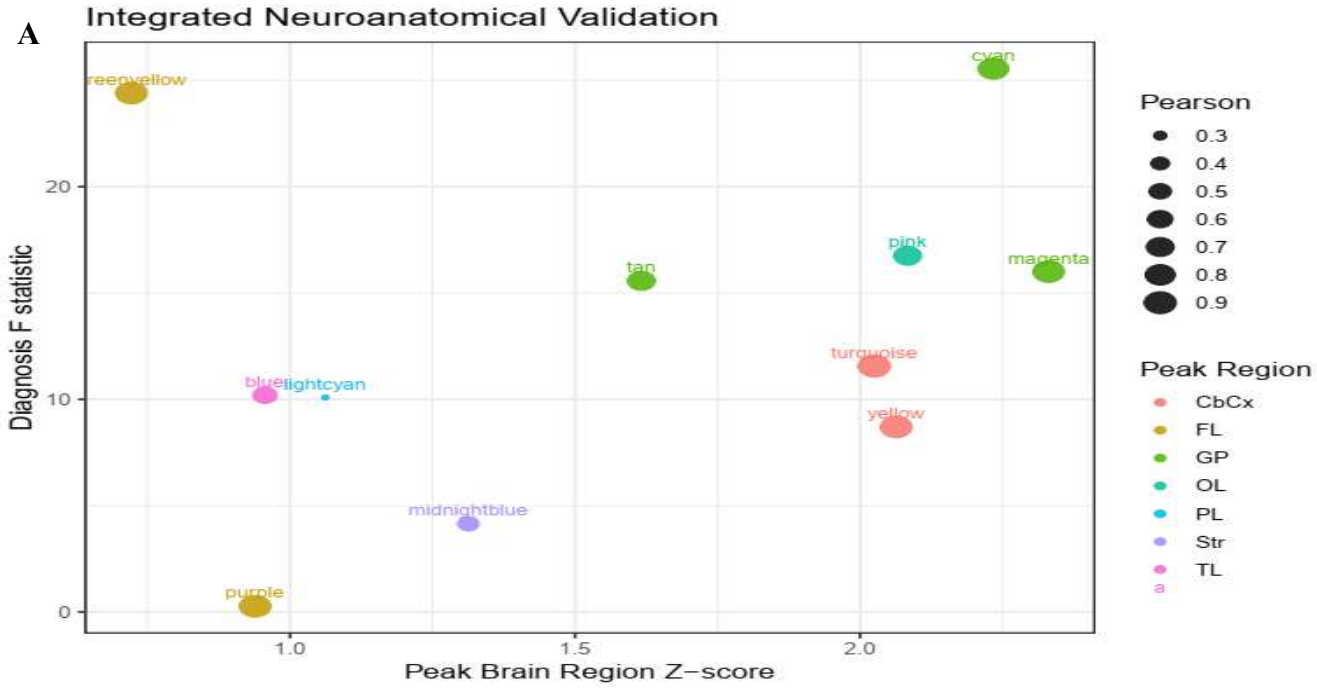
To identify PTB-derived neuroimmune programmes with the strongest combined evidence for cross-disease activity, neuroanatomical organization, and reproducibility, module-level measures of diagnostic significance, regional enrichment, regional specificity, and inter-donor concordance were integrated into a

unified framework (Figure 11A). Module-preservation results were subsequently used to provide additional context regarding the extent to which prioritized modules retained their original co-expression architecture in AD.

Integration of these complementary evaluation criteria demonstrated substantial differences in the overall strength of individual modules. The cyan module exhibited the strongest combined evidence, displaying the highest diagnosis-associated effect together with strong regional enrichment and high reproducibility across Allen Brain Atlas donors. Similarly, the magenta**,** turquoise**,** greenyellow**,** pink, and tan modules consistently ranked among the highest-performing modules, combining strong disease association, regional enrichment, and inter-donor reproducibility. In contrast, modules such as midnightblue**,** lightcyan, and purple exhibited comparatively weaker integrated evidence because of lower diagnosis effects, reduced regional specificity, or diminished agreement between donors (Figure 11A). Notably, preservation analysis revealed that several modules exhibiting strong cross-disease or neuroanatomical evidence, including magenta, yellow, and turquoise, also demonstrated evidence of network preservation in AD, whereas others showed strong disease associations despite limited preservation, suggesting partial rewiring of their underlying co-expression architecture in the AD context.

To further examine the relationship between neuroanatomical organization and reproducibility, regional specificity was compared directly with inter-donor reproducibility across evaluated modules. Modules with greater regional specificity consistently showed higher inter-donor reproducibility, revealing a strong positive relationship between anatomical restriction and donor concordance (Figure 11B). Highly region-specific modules, including turquoise**,** purple**,** greenyellow, yellow**,** magenta**,** and cyan, also displayed the greatest reproducibility, whereas modules with lower anatomical specificity, including lightcyan and midnightblue, exhibited correspondingly lower reproducibility. These findings indicate that spatially restricted neuroimmune transcriptional programmes represent the most stable and reproducible components of the PTB-derived molecular network.

The integrated contextualization framework demonstrates that the principal PTB-derived co-expression modules combine significant disease association, regional specificity, and robust inter-donor reproducibility. The individual diagnosis-associated effects, regional specificity scores, and inter-donor reproducibility metrics contributing to the integrated contextualization framework are presented in Figures S8–S10, while the complete integrated contextualization metrics are provided in Tables S37–S38. Collectively, these findings identify a subset of PTB-derived neuroimmune transcriptional programmes that combine cross-disease concordance in an independent AD cohort with reproducible region-specific organization within the adult human brain, providing the strongest evidence for candidate shared neuroimmune programmes across chronic infection with neurodegeneration.



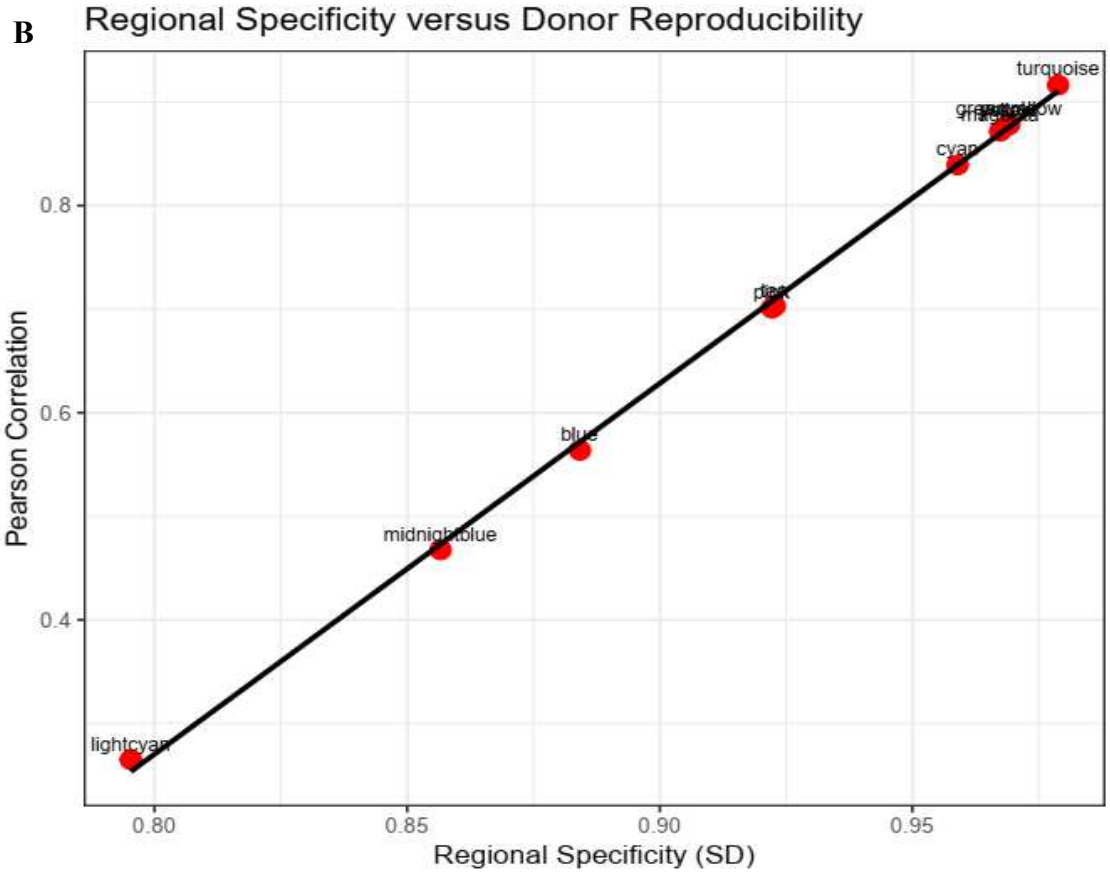


Figure 11. Integrated neuroanatomical contextualization of PTB-derived neuroimmune co-expression modules.

(A) Integrated contextualization framework of PTB-derived neuroimmune modules combining diagnosis-associated effect size (F statistic), peak regional enrichment (maximum regional Z-score), donor reproducibility (Pearson correlation), and peak brain-region assignment. Bubble size represents inter-donor reproducibility, while bubble colour denotes the brain region exhibiting maximal module enrichment. **(B)** Relationship between regional specificity and donor reproducibility of evaluated PTB-derived neuroimmune modules. Each point represents one evaluated module. Regional specificity was quantified as the standard deviation of consensus regional Z-scores, and donor reproducibility was assessed using Pearson correlation between independent Allen Human Brain Atlas donors.

Discussion

The present study provides systems-level evidence that PTB and AD exhibit shared neuroimmune transcriptional programmes that are organized into coordinated co-expression networks and exhibit reproducible neuroanatomical organization within the adult human brain. Rather than representing disease-specific inflammatory responses or isolated transcriptional alterations, the widespread peripheral immune remodelling associated with active PTB defines coherent molecular programmes that exhibited similar activity in the independent AD cohort, supporting the existence of shared neuroimmune transcriptional programmes across otherwise distinct disease contexts. By integrating transcriptomics, network biology, immune-cell deconvolution, machine learning, independent cross-disease evaluation, and brain transcriptomic mapping within a unified analytical framework, this study moves beyond descriptive comparisons to define the systems-level molecular organization linking chronic pulmonary infection with neurodegeneration. Together, these findings provide a framework for investigating how chronic infection-associated immune remodelling may influence neurodegenerative processes and identify candidate molecular systems for future mechanistic and translational studies.

A central biological insight emerging from this study is that the host response to active PTB is organized into coordinated immune transcriptional programmes that define a systems-level architecture of the host immune response. These programmes encompassed complementary components of host immunity, including innate immune activation, adaptive B-cell responses, tissue remodelling, and concurrent suppression of mitochondrial metabolism, oxidative phosphorylation, and fundamental cellular maintenance pathways, reflecting the extensive immunometabolic reprogramming accompanying persistent *Mycobacterium tuberculosis* infection. The strong associations between these transcriptional programmes and neutrophils, cytotoxic lymphocytes, macrophages, and B cells further suggest that sustained immune activation during PTB is coordinated across multiple interacting immune-cell populations rather than occurring within isolated cellular compartments. Collectively, these findings indicate that active PTB is characterized by integrated cellular and molecular remodelling that extends

beyond local pulmonary inflammation, providing a biologically plausible basis through which chronic peripheral infection may influence neuroimmune processes relevant to neurodegeneration.

Our findings extend previous transcriptomic studies of PTB and AD by demonstrating that PTB-associated immune programmes not only exhibited concordant activity in AD but several also retained significant aspects of their original co-expression architecture, indicating conservation of systems-level molecular organization across diseases. Furthermore, these programmes demonstrated reproducible neuroanatomical organization within the adult human brain. Several PTB-derived modules also demonstrated moderate preservation within the independent AD cohort. Whereas differential module activity indicates that transcriptional programmes remain detectable across diseases, module-preservation analysis evaluates whether the underlying relationships among genes are retained. Preservation of the green, magenta, yellow, black, blue, brown, and turquoise modules suggests that these programmes reflect stable biological systems rather than disease-specific gene sets. Conversely, modules exhibiting low preservation despite substantial gene retention may reflect disease-specific network rewiring, highlighting both shared and divergent aspects of PTB- and AD-associated neuroimmune biology.

These observations build upon growing evidence that PTB is a systemic disease characterized by persistent immune dysregulation and multisystem sequelae extending well beyond pulmonary pathology. Recent clinical and mechanistic studies have shown that the long-term consequences of PTB involve sustained inflammatory, metabolic, and immunological perturbations affecting multiple organ systems rather than the lungs alone [2,3,5,6]. Complementing these observations, advances in tuberculosis immunology have demonstrated that chronic *Mycobacterium tuberculosis* infection induces coordinated interactions among macrophages, neutrophils, lymphocytes, and other immune populations, driving persistent immune remodelling instead of transient inflammatory activation [11,14,15]. Consistent with this evolving understanding, our findings demonstrate that PTB-associated immune remodelling is organized into PTB-derived transcriptional programmes that are reproduced in the independent AD cohort, extending current concepts of post-tuberculosis systemic pathology from clinical observations to systems-level molecular

organization. These observations also align with emerging models of AD that recognize the disease as a systemic neuroimmune disorder characterized by dynamic interactions between peripheral and central immune compartments, thereby providing a biological rationale for evaluating PTB-derived transcriptional programmes in peripheral blood [12,13,20]. More broadly, recent advances in neuroimmunology have established that continuous bidirectional communication between the immune and nervous systems is essential for maintaining brain homeostasis but becomes maladaptive during chronic inflammation, ultimately influencing cognition and neurodegenerative processes[21,22]. Experimental evidence demonstrating lung–brain crosstalk, blood–brain barrier modulation, and immune-mediated signalling between the periphery and the central nervous system[7,8] further provides biological plausibility for the shared neuroimmune transcriptional organization identified in this study. Moreover, our findings complement recent transcriptomic meta-analyses demonstrating reproducible immune-related gene networks across multiple AD brain regions and cohorts [18,19,23], while offering a molecular framework that may help explain epidemiological observations reporting an increased risk of dementia among individuals recovering from active tuberculosis [9,10]. These findings, together, bridge previously separate observations from tuberculosis immunology, neuroimmunology, transcriptomic systems biology, and population-based epidemiology into a coherent systems-level model describing shared neuroimmune organization across chronic pulmonary infection and neurodegeneration.

A systems-level interpretation of our findings is that chronic PTB may be associated with coordinated peripheral immune remodelling organized into stable transcriptional networks extending beyond the site of infection. Rather than functioning as isolated inflammatory pathways, these programmes integrate innate and adaptive immunity, immunometabolic reprogramming, and tissue remodelling within a coordinated neuroimmune framework. Within this framework, the concordant activity of these programmes in an independent AD cohort, together with their reproducible region-specific organization within the adult human brain, suggests that persistent peripheral immune activation may contribute to molecular environments associated with region-specific neuroimmune vulnerability rather than diffuse

neuroinflammatory responses. This model shifts attention from individual genes to coordinated biological systems shared across chronic infection and neurodegeneration. Although causal relationships cannot be inferred from the present analyses, recent evidence demonstrating that modulation of systemic immunity through mycobacterial exposure, including BCG vaccination and immunotherapy, reshapes central nervous system immune responses and alters AD-associated biomarkers provides independent biological support for the existence of a functional immune–brain axis linking peripheral immune activity to brain health [16,17]. Collectively, these observations support a systems biology paradigm in which chronic infectious diseases are viewed not solely as localized pathologies but as persistent immune perturbations capable of reshaping shared molecular networks with potential implications for long-term neurological health.

The systems-level framework proposed in this study has several important translational implications. The identification of shared neuroimmune transcriptional programmes linking PTB and AD suggests that persistent peripheral immune remodelling following active tuberculosis may provide an opportunity to investigate whether network-based signatures can identify individuals at increased risk of subsequent neurodegenerative complications before the onset of overt neurological disease. Rather than relying on individual biomarkers, our findings support the development of network-based molecular signatures that capture coordinated immune dysregulation and may provide complementary biological information for patient stratification and longitudinal risk assessment. Beyond biomarker discovery, these PTB-derived transcriptional programmes highlight interconnected immune networks that may represent more informative therapeutic targets than individual molecular components, consistent with emerging systems medicine approaches to complex chronic diseases. In this context, recent evidence demonstrating that BCG vaccination and immunotherapy can remodel central nervous system immunity and alter AD-associated biomarkers further supports the concept that systemic immune pathways are amenable to therapeutic manipulation [16,17]. Although these translational opportunities remain to be validated experimentally and clinically, our findings provide a rational framework for future longitudinal studies integrating molecular profiling, neuroimaging, cognitive assessment, and functional immunological analyses to determine

whether persistent immune remodelling following PTB contributes to subsequent neurodegenerative risk and whether these trajectories can be modified through targeted immune interventions.

Several limitations should be considered when interpreting these findings. First, the analyses were based on publicly available transcriptomic datasets, and therefore inherit the limitations of retrospective studies, including cohort heterogeneity, differences in clinical characteristics, and potential unmeasured confounding variables. Second, the transcriptomic analyses were performed using peripheral whole-blood samples, which may not fully capture the cellular and molecular complexity of immune responses occurring within pulmonary lesions or the central nervous system. Likewise, the neuroanatomical analyses were performed using transcriptomic data from neurologically normal adult brains in the AHBA and therefore define the anatomical organization of PTB-derived transcriptional programmes rather than disease-specific brain expression in PTB or AD. In addition, because the AHBA comprises repeated regional measurements from only two adult donors, these analyses characterize reproducible regional transcriptional organization within the available donors and should not be interpreted as estimates of population-level neuroanatomical variation. Third, although the PTB-derived transcriptional programmes identified in this study were reproducibly validated in an independent AD cohort and exhibited consistent neuroanatomical organization, the observational nature of the analyses precludes inference of causal relationships between PTB and subsequent neurodegenerative processes. Finally, the predicted molecular interactions and biological networks require experimental validation in appropriate cellular and animal models, together with prospective longitudinal studies integrating immunological, neuroimaging, and cognitive assessments. Whether the PTB-derived transcriptional programmes identified here are unique to PTB or represent broader molecular signatures shared across other chronic inflammatory diseases also remains an important question for future comparative investigations. Nevertheless, several features strengthen the robustness of the present study. Rather than relying on a single analytical approach, we integrated complementary systems biology methodologies, including differential transcriptomic analysis, weighted gene co-expression network analysis, immune-cell deconvolution, machine learning, cross-disease evaluation, and

regional brain transcriptomic mapping, allowing convergent biological evidence to emerge across complementary analytical frameworks. The observation of similar transcriptional programmes in the independent AD cohort, together with concordant regional expression patterns across the two Allen Human Brain Atlas donors, increases confidence that the identified transcriptional programmes reflect biologically meaningful organization rather than dataset-specific variation.

Conclusions

In conclusion, we identified PTB-associated neuroimmune transcriptional programmes that were validated in independent PTB cohorts, exhibited concordant activity and partial network preservation in AD, and demonstrated reproducible regional organization within the adult human brain. Integrating transcriptomics, network biology, immune-cell deconvolution, machine learning, and neuroanatomical mapping revealed shared neuroimmune molecular architecture spanning chronic pulmonary infection and neurodegeneration. Although the observational nature of this study precludes causal inference, these findings provide a foundation for future mechanistic, longitudinal, and translational investigations aimed at defining how persistent immune remodelling may contribute to neurodegenerative processes.

List of abbreviations

AD: Alzheimer's disease

AHBA: Allen Human Brain Atlas

ANOVA: Analysis of variance

AUC: Area under the curve

BCG: Bacillus Calmette–Guérin

BH: Benjamini–Hochberg

BP: Biological Process

CC: Cellular Component

CI: Confidence interval

CTR: Control

DCA: Decision curve analysis

DO: Disease Ontology

FDR: False discovery rate

GEO: Gene Expression Omnibus

GO: Gene Ontology

GO-BP: Gene Ontology Biological Process

GS: Gene significance

KEGG: Kyoto Encyclopedia of Genes and Genomes

LASSO: Least absolute shrinkage and selection operator

MAD: Mean absolute deviation

ME: Module eigengene

MF: Molecular Function

ML: Machine learning

MM: Module membership

NCBI: National Center for Biotechnology Information

NK: Natural killer

PCA: Principal component analysis

PR-AUC: Area under the precision–recall curve

PTB: Pulmonary tuberculosis

RMSE: Root mean square error

ROC: Receiver operating characteristic

TB: Tuberculosis

VST: Variance-stabilizing transformation

WGCNA: Weighted gene co-expression network analysis

Declarations

Ethics approval and consent to participate

This study exclusively analyzed publicly available, de-identified transcriptomic datasets obtained from the NCBI Gene Expression Omnibus (GEO) and the Allen Human Brain Atlas. No new human participants were recruited, and no identifiable personal information was accessed. Ethical approval and informed consent were obtained by the original study investigators as described in the corresponding primary

publications. Therefore, no additional institutional ethical approval was required for the present secondary analysis of publicly available data.

Consent for publication

Not applicable

Availability of data and materials

All transcriptomic datasets analysed in this study are publicly available through the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE157657, GSE107991, GSE279063, and GSE313307. Allen Human Brain Atlas transcriptomic data used for neuroanatomical analyses are publicly available from the Allen Institute for Brain Science (https://human.brain-map.org/static/download). All processed results, supplementary tables, and supporting data generated during this study are provided within the manuscript and its Supporting Information files. All code used for data preprocessing, differential expression analysis, weighted gene co-expression network analysis, functional enrichment, immune-cell deconvolution, machine-learning analyses, external validation, and neuroanatomical mapping is available at https://github.com/oluwaseuntosinn/PTB_AD_Neuroimmune_Systems_Biology and permanently archived at https://doi.org/10.5281/zenodo.22257533, enabling full reproduction of the reported findings.

Competing Interests

The authors have declared that no competing interests exist.

Funding

The authors received no specific funding for this work.

Author Contributions

Conceptualization: Oluwaseun Oluwatosin Taofeek. Data curation: Oluwaseun Oluwatosin Taofeek. Formal analysis: Oluwaseun Oluwatosin Taofeek. Investigation: Oluwaseun Oluwatosin Taofeek. Methodology: Oluwaseun Oluwatosin Taofeek and Lateef Adegboyega Sulaimon. Project administration: Oluwaseun Oluwatosin Taofeek and Lateef Adegboyega Sulaimon. Resources: Solomon Adesola Adigun, Elijah Kolawole Oladipo, Aganze Gloire-Aimé Mushebenge, Afrah Khairallah, Paul Akinduti, and Lateef Adegboyega Sulaimon. Software: Oluwaseun Oluwatosin Taofeek. Supervision: Lateef Adegboyega Sulaimon. Validation: Oluwaseun Oluwatosin Taofeek, Solomon Adesola Adigun, Elijah Kolawole Oladipo, Aganze Gloire-Aimé Mushebenge, Afrah Khairallah, Paul Akinduti, and Lateef Adegboyega Sulaimon. Visualization: Oluwaseun Oluwatosin Taofeek. Writing – original draft: Oluwaseun Oluwatosin Taofeek. Writing – review & editing: Oluwaseun Oluwatosin Taofeek, Solomon Adesola Adigun, Elijah Kolawole Oladipo, Aganze Gloire-Aimé Mushebenge, Afrah Khairallah, Paul Akinduti, and Lateef Adegboyega Sulaimon.

Acknowledgments

The authors acknowledge the investigators who generated and made publicly available the GEO datasets analysed in this study and the Allen Institute for Brain Science for providing access to the Allen Human Brain Atlas.

Supporting Information

Supporting Figures

S1 Fig. MA plot showing $\log_2$ fold changes as a function of mean normalized gene expression following differential expression analysis.

S2 Fig. DESeq2 dispersion plot showing gene-wise, fitted, and final dispersion estimates used for differential expression analysis.

S3 Fig. Module eigengene clustering revealing higher-order relationships among PTB-derived co-expression modules.

S4 Fig. Eigengene network illustrating connectivity among PTB-derived co-expression modules.

S5 Fig. Precision–recall curve for external validation cohort GSE107991.

S6 Fig. Precision–recall curve for external validation cohort GSE279063.

S7 Fig. Regional specificity of PTB-derived neuroimmune modules across the Allen Human Brain Atlas. Modules are ranked according to consensus regional specificity scores.

S8 Fig. Diagnosis-associated effects across PTB-derived neuroimmune modules in the Alzheimer's disease validation cohort.

S9 Fig. Regional specificity of validated PTB-derived neuroimmune modules.

S10 Fig. Inter-donor reproducibility of validated PTB-derived neuroimmune modules across independent Allen Human Brain Atlas donors.

Supporting Tables

Differential Expression Analysis

S1 Table. Complete differential expression results for active pulmonary tuberculosis versus healthy controls.

S2 Table. Significantly differentially expressed genes identified in active pulmonary tuberculosis.

S3 Table. Upregulated genes in active pulmonary tuberculosis.

S4 Table. Downregulated genes in active pulmonary tuberculosis.

Weighted Gene Co-expression Network Analysis

S5 Table. Sizes of PTB-derived weighted gene co-expression network modules.

S6 Table. Pearson correlations between module eigengenes and clinical traits.

S7 Table. Statistical significance of module–trait associations.

S8 Table. Network construction parameters used for weighted gene co-expression network analysis.

Immune Cell Deconvolution

S9 Table. Complete correlations between PTB-derived co-expression modules and inferred immune-cell populations.

S10 Table. Significant PTB module–immune-cell associations following false discovery rate correction.

Machine Learning and Biomarker Discovery

S11 Table. Final locked 12-gene PTB biomarker signature identified by nested machine learning.

S12 Table. Performance metrics of the nested cross-validation classifier in the discovery cohort.

S13 Table. Calibration statistics for the locked Random Forest classifier in the discovery cohort.

S14 Table. Variable importance ranking of the final locked Random Forest classifier.

S15 Table. Permutation test results for the nested Random Forest classifier.

S16 Table. External validation results for GSE107991, including predictions, confusion matrix, and calibration statistics.

S17 Table. External validation results for GSE279063, including predictions, confusion matrix, and calibration statistics.

Alzheimer's Disease Validation

S18 Table. Two-way ANOVA results for PTB-derived co-expression modules in the Alzheimer's disease cohort.

S19 Table. One-way ANOVA and false-discovery-rate-adjusted significance of PTB-derived module eigengenes across Alzheimer's disease clinical groups.

S20 Table. Pairwise comparisons of PTB-derived module eigengene activity among Alzheimer's disease clinical groups.

S21 Table. Group mean module eigengene scores across Alzheimer's disease clinical groups.

S22 Table. Summary statistics for validated PTB-derived neuroimmune biomarkers.

S23 Table. Pairwise comparisons for validated PTB-derived neuroimmune biomarkers.

S24 Table. Expression matrix of validated PTB-derived neuroimmune biomarkers in the Alzheimer's disease cohort.

Module Preservation Analysis

S25 Table. Gene retention statistics for PTB-derived co-expression modules in the Alzheimer's disease cohort.

S26 Table. Module preservation statistics for PTB-derived co-expression modules in the Alzheimer's disease cohort.

Neuroanatomical Mapping

S27 Table. Kruskal–Wallis statistics for regional differences in module expression across major brain regions.

S28 Table. Consensus peak brain-region assignments for PTB-derived neuroimmune modules.

S29 Table. Consensus regional Z-score matrix for PTB-derived neuroimmune modules across the Allen Human Brain Atlas.

Reproducibility Analysis

S30 Table. Inter-donor reproducibility of PTB-derived neuroimmune modules across major brain regions.

S31 Table. Summary statistics of regional reproducibility across independent Allen Human Brain Atlas donors.

Integrated Neuroanatomical Prioritization

S32 Table. Consensus ranking of regional enrichment across PTB-derived neuroimmune modules.

S33 Table. Consensus ranking of validated PTB-derived neuroimmune modules.

S34 Table. Absolute ranking of regional enrichment across PTB-derived neuroimmune modules.

S35 Table. Consensus ranking of anatomical substructures showing preferential module enrichment.

S36 Table. Top regionally specific PTB-derived neuroimmune modules identified from consensus Allen **Human Brain Atlas analysis.**

S37 Table. Integrated validation metrics for PTB-derived neuroimmune modules, including diagnosis-associated effects, peak brain region, regional specificity, and inter-donor reproducibility.

S38 Table. Integrated systems-level summary of validated PTB-derived neuroimmune modules combining disease association, neuroanatomical localization, regional specificity, and reproducibility metrics.

References

1. Global gains in tuberculosis response endangered by funding challenges [Internet]. [cited 2026 Jul 28]. Available from: https://www.who.int/news/item/12-11-2025-global-gains-in-tuberculosis-response-endangered-by-funding-challenges

2. Luies L, du Preez I. The Echo of Pulmonary Tuberculosis: Mechanisms of Clinical Symptoms and Other Disease-Induced Systemic Complications. Clin Microbiol Rev. 2020 Jul 1;33(4):e00036-20. doi:10.1128/CMR.00036-20 PubMed PMID: 32611585; PubMed Central PMCID: PMC7331478.

3. Min J, Yoolwon Jeong, Hyung Woo Kim, Choi H, Sohn H, Lipman M, et al. Post-tuberculosis Health: A Holistic Perspective on Multisystem Sequelae and Long-Term Wellbeing. Tuberc Respir Dis (Seoul). 2026 Apr;89(2):202–14. doi:10.4046/trd.2025.0120 PubMed PMID: 41713847; PubMed Central PMCID: PMC13065409.

4. Ravimohan S, Kornfeld H, Weissman D, Bisson GP. Tuberculosis and lung damage: from epidemiology to pathophysiology. European Respiratory Review. 2018 Feb 28;27(147). doi:10.1183/16000617.0077-2017

5. Akshay N Gupte, Boisson-Walsh, Moises A Huaman, Rajita Bhavaraju, Yilan Gao, David Hom, et al. Long-term Sequelae of Pulmonary Tuberculosis: A Narrative Review. Clin Infect Dis. 2026 Apr 15;82(Supplement_2):S60–7. doi:10.1093/cid/ciag123

6. Wang J, Yuan B, Fang Y, Zhao C, Liu J, Xiong K. Post-tuberculosis morbidities and their associated mortality: moving from challenges to solutions. European Respiratory Review. 2025 Oct 22;34(178). doi:10.1183/16000617.0148-2025

7. Villalba N, Beard RS, Yuan SY. Barrier breakdown: lung-brain crosstalk in systemic inflammation. J Neuroinflammation. 2026 Apr 14;23(1):164. doi:10.1186/s12974-026-03807-1 PubMed PMID: 41981636; PubMed Central PMCID: PMC13200398.

8. Amanda S. Latham, Charlize E. Geer, David F. Ackart, Kristin N. Weninger, Chase C. Gross, Brendan K. Podell, et al. Immune cell infiltration and modulation of the blood-brain barrier in a guinea pig model of tuberculosis: Observations without evidence of bacterial dissemination to the brain. PLOS ONE. 2024 Dec 31;19(12):e0307577. doi:10.1371/journal.pone.0307577

9. Yeo Y, Yoo JE, Han K, Jung JH, Choi H, Park J, et al. Risk of dementia in survivors of active tuberculosis in Korea: A nationwide cohort study. Journal of Infection and Public Health. 2024 Feb 1;17(2):286–92. doi:10.1016/j.jiph.2023.12.006

10. Peng YH, Chen CY, Su CH, Muo CH, Chen KF, Liao WC, et al. Increased Risk of Dementia Among Patients With Pulmonary Tuberculosis. Am J Alzheimers Dis Other Demen. 2015 Sep;30(6):629–34. doi:10.1177/1533317515577186 PubMed PMID: 25792663; PubMed Central PMCID: PMC10852775.

11. Russell DG, Simwela NV, Mattila JT, Flynn J, Mwandumba HC, Pisu D. How macrophage heterogeneity affects tuberculosis disease and therapy. Nat Rev Immunol. 2025 May;25(5):370–84. doi:10.1038/s41577-024-01124-3

12. Butovsky O, Rosenzweig N, Kleemann KL, Jorfi M, Kuchroo VK, Tanzi RE, et al. Immune dysfunction in Alzheimer disease. Nat Rev Neurosci. 2026 Mar;27(3):196–218. doi:10.1038/s41583-025-00997-0 PubMed PMID: 41315874; PubMed Central PMCID: PMC13010386.

13. Gopalakrishna PK, Che Mohd Nassir CMN, Anandan S, Hussan F, Dandala KCR, Sura S, et al. The Immuno-Glial Connectome in Alzheimer's Disease: Integrating Central and Peripheral Inflammatory Networks. Cell Mol Neurobiol. 2026 Jan 22;46(1):31. doi:10.1007/s10571-026-01671-4

14. Dupuy P, Monard SC, Neyrolles O, Lugo-Villarino G. From Phagosomes to Niches: Macrophage Biology in Tuberculosis Revisited. Cold Spring Harb Perspect Med. 2026 Apr 2;16(4):a041819. doi:10.1101/cshperspect.a041819 PubMed PMID: 41184126; PubMed Central PMCID: PMC13065343.

15. Prasad GVRK, Philips JA. How Mycobacterium tuberculosis Subverts Innate and Adaptive Immunity and Their Crosstalk: Implications for Vaccine Design. Vaccines. 2026 May;14(5):414. doi:10.3390/vaccines14050414

16. Weinberg MS, Kodali MC, Li Z, Galler JA, Tidball AR, Reynolds WC, et al. Bacillus Calmette-Guérin (BCG) immunotherapy reprograms CNS immunity and alters Alzheimer's biomarkers: results from two open-label clinical trials. Commun Med (Lond). 2026 Jul 2;6(1):358. doi:10.1038/s43856-026-01691-7 PubMed PMID: 42393277; PubMed Central PMCID: PMC13328741.

17. Druszczynska M, Sadowska B, Kulesza J, Kulesza E, Fol M. Bacillus Calmette–Guérin (BCG) Vaccination and the Immune–Brain Axis: Implications for Neuroprotection and Neurodegenerative Disease. Vaccines. 2026 May 2;14(5):412. doi:10.3390/vaccines14050412

18. Lopes KDP, Vialle RA, Green G, Fujita M, Gaiteri C, Menon V, et al. Gene module-trait network analysis uncovers cell type specific systems and genes relevant to Alzheimer's Disease. acta neuropathol commun. 2025 Nov 4;13(1):222. doi:10.1186/s40478-025-02143-4

19. Hou J, Hess JL, Zhang C, Van Rooij JGJ, Hearn GC, Fan CC, et al. Meta-Analysis of Transcriptomic Studies of Blood and Six Brain Regions Identifies a Consensus of 15 Cross-Tissue Mechanisms in Alzheimer's Disease and Suggests an Origin of Cross-Study Heterogeneity. American J of Med Genetics Pt B. 2025 Jun;198(4):e33019. doi:10.1002/ajmg.b.33019

20. Zhang S, Gao Y, Zhao Y, Huang TY, Zheng Q, Wang X. Peripheral and central neuroimmune mechanisms in Alzheimer's disease pathogenesis. Mol Neurodegeneration. 2025 Feb 21;20(1):22. doi:10.1186/s13024-025-00812-5

21. Leunig A, Gianeselli M, Russo SJ, Swirski FK. Connection and communication between the nervous and immune systems. Nat Rev Immunol. 2025 Dec;25(12):912–33. doi:10.1038/s41577-025-01199-6

22. Klein RS. The neuroimmune system and cognition. Nat Immunol. 2026 Jul;27(7):1353–63. doi:10.1038/s41590-026-02560-0

23. Aguzzoli Heberle B, Fox KL, Lobraico Libermann L, Ronchetti Martins Xavier S, Tarnowski Dallarosa G, Carolina Santos R, et al. Systematic review and meta-analysis of bulk RNAseq studies in human Alzheimer's disease brain tissue. Alzheimer's & Dementia. 2025 Mar;21(3):e70025. doi:10.1002/alz.70025

24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec;9(1):559. doi:10.1186/1471-2105-9-559

25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9. doi:10.1038/75556 PubMed PMID: 10802651; PubMed Central PMCID: PMC3037419.

26. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000 Jan 1;28(1):27–30. doi:10.1093/nar/28.1.27

27. Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, et al. The Human Disease Ontology 2022 update. Nucleic Acids Research. 2022 Jan 7;50(D1):D1255–61. doi:10.1093/nar/gkab1063

28. Shen EH, Overly CC, Jones AR. The Allen Human Brain Atlas. Trends in Neurosciences. 2012 Dec;35(12):711–4. doi:10.1016/j.tins.2012.09.005

29. Tabone O, Verma R, Singhania A, Chakravarty P, Branchett WJ, Graham CM, et al. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. Journal of Experimental Medicine. 2021 Oct 4;218(10):e20210915. doi:10.1084/jem.20210915

30. Singhania A, Verma R, Graham CM, Lee J, Tran T, Richardson M, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun. 2018 Jun 19;9(1):2308. doi:10.1038/s41467-018-04579-w

31. Ahn E, Kang Y, Lee H, Kim J. GEO Accession viewer [Internet]. [cited 2026 Sep 1]. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi

32. Pegoraro G, MacBean LF, Smith AR, Soanes DM, Smith RG, Imm J, et al. Systemic infections alter cortical transcriptional signatures in Alzheimer's disease. Neurobiology of Aging. 2026 Nov;167:155–67. doi:10.1016/j.neurobiolaging.2026.06.010

33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550. doi:10.1186/s13059-014-0550-8

34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology. 2012 May 1;16(5):284–7. doi:10.1089/omi.2011.0118

35. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015 Feb 15;31(4):608–9. doi:10.1093/bioinformatics/btu684

36. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 May;12(5):453–7. doi:10.1038/nmeth.3337 PubMed PMID: 25822800; PubMed Central PMCID: PMC4739640.

37. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: https://www.R-project.org/

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf