RESEARCH PAPER

Conserved Cell-Type–Specific Immune Programs in Treatment-Naïve Multiple Sclerosis: Cross-Cohort Transcriptomic Validation Across Independent Blood Datasets

Muhammed Ertuğrul GÜVENÇ1*, İhsan Saygın Sarı2, Mustafa Uzun3, Sinan ELİAÇIK4, Adeeba Ali5
[1]Department of Neurology, University of Health Sciences, Gaziosmanpasa Training and Research Hospital, Istanbul, Türkiye.
[2]Department of Neurology, University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Istanbul, Türkiye.
[3]Department of Neurology, Mardin Training and Research Hospital, Mardin, Türkiye.
[4]Department of Neurology, Faculty of Medicine, Hitit University, Çorum, Türkiye.
[5]Department of Biotechnology, Faculty of Science and Technology, University of Central Punjab, Lahore, Pakistan.
Correspondence Author: Muhammed Ertuğrul GÜVENÇ
Correspondence Author e-mail: drertugrulguvenc@gmail.com
Co-Author: İhsan Saygın Sarı
Email: isayginsari@gmail.com
Co-Author: Mustafa Uzun
Email: mustuzun@outlook.com
Co-Author: Sinan Eliaçık
E-mail: sinaneliacik@gmail.com
Co-Author: Adeeba Ali
E-mail: adeebaali1146@gmail.com

ABSTRACT

Background: Multiple sclerosis (MS) is a heterogeneous immune-mediated neurological disorder. This study investigated whether peripheral immune transcriptional programs associated with treatment-naïve MS were reproducible across independent blood transcriptomic cohorts. **Methods:** Three publicly available transcriptomic cohorts were analyzed, comprising one discovery cohort and two independent validation cohorts. The combined dataset included 95 treatment-naïve patients with MS and 75 healthy controls. Analyses included quality control, differential expression, immune-cell deconvolution, pathway enrichment, and cross-cohort validation. **Results:** After quality control, 170 of 175 samples were retained. Differential-expression analysis identified 128 genes (72 upregulated and 56 downregulated). Leading upregulated genes included ISG15 (log2FC=0.82; adjusted P=0.001), IFITM3 (0.76; P=0.004), HLA-DRA (0.69; P=0.008), and LST1 (0.61; P=0.014), whereas CCR7 and IL7R were downregulated. Fourteen pathways met adjusted $P<0.05$, including type I interferon signaling, antigen processing and presentation, and leukocyte activation. Deconvolution indicated higher estimated monocyte and activated dendritic-cell contributions and a lower naïve CD4+ T-cell contribution in MS. Forty-three genes showed concordant direction across cohorts, and 26 met the stated conservation criterion. Discovery and pooled validation effect estimates were correlated (Pearson r=0.72, P=0.003). The protein–protein interaction network comprised 26 nodes, 41 edges, and three functional modules. **Conclusions:** Treatment-naïve MS was associated with coordinated peripheral immune transcriptional alterations showing cross-cohort consistency. These findings warrant prospective validation and do not establish clinically deployable biomarkers.

Keywords: Multiple sclerosis; transcriptomics; treatment-naïve patients; peripheral blood; immune cell types; immune dysregulation; cross-cohort validation; molecular signatures.

How to cite this article: Muhammed Ertu. Conserved Cell-Type–Specific Immune Programs in Treatment-Naïve Multiple Sclerosis: Cross-Cohort Transcriptomic Validation Across Independent Blood Datasets. Int J Drug Deliv Technol. 2026;16(90s):1373-1390. DOI: 10.25258/ijddt.16.90s.136.

1. Introduction

Multiple sclerosis (MS) is a chronic inflammatory and neurodegenerative disorder of the central nervous system (CNS) characterized by immune-mediated demyelination, axonal injury, and progressive neurological dysfunction. Although the disease is traditionally considered an autoimmune disorder directed against components of the CNS, contemporary evidence supports a substantially more complex pathobiology involving coordinated interactions among T lymphocytes, B lymphocytes, monocytes, natural killer cells, antigen-presenting cells, cytokine networks, and CNS-resident cells.

The resulting inflammatory architecture is highly heterogeneous across individuals and disease stages, contributing to substantial variation in clinical phenotype, radiological activity, disease progression, and therapeutic response. Recent work has consequently shifted from viewing MS as a single immunological entity toward understanding it as a collection of partially overlapping molecular and cellular states. [1–3]

Peripheral immune cells are central to the initiation and perpetuation of inflammatory processes in MS. Circulating lymphocytes and myeloid populations participate in immune surveillance, antigen recognition, cytokine production, and trafficking across the blood–brain barrier, thereby providing an important interface between systemic immunity and CNS inflammation. Alterations in the frequency, activation state, and functional phenotype of peripheral immune populations have been described in patients with MS, including changes involving memory T-cell and B-cell compartments. Systems-immunology studies have demonstrated that untreated MS is associated with measurable alterations in circulating immune-cell composition, supporting the concept that peripheral immune abnormalities can be detected before the effects of disease-modifying therapies substantially reshape the immune landscape. [6,7,19]

Peripheral blood therefore represents an attractive and comparatively accessible biological compartment for investigating molecular features of MS. Transcriptomic profiling of peripheral blood mononuclear cells (PBMCs), whole blood, and purified immune-cell populations has revealed alterations in genes involved in lymphocyte signaling, inflammatory pathways, cytokine responses, antigen presentation, cell trafficking, and other immune-regulatory processes. Earlier studies demonstrated that peripheral blood transcriptional profiles could capture biologically meaningful heterogeneity among patients with MS, including transcriptional patterns associated with differences in disease activity. In one large cohort of untreated individuals, peripheral blood transcriptional profiling identified distinct molecular subsets associated with differences in inflammatory activity, illustrating the potential of blood-based gene-expression analysis to resolve clinically relevant molecular heterogeneity. [5,11]

However, interpretation of bulk peripheral blood transcriptomic data presents an important biological challenge. Whole-blood or PBMC expression profiles represent mixtures of multiple immune-cell populations, and an apparent change in gene expression may arise from altered cellular abundance, altered transcriptional activity within a specific cell type, or a combination of both. Consequently, a bulk transcriptomic signature cannot necessarily be interpreted as evidence of a cell-intrinsic molecular alteration. This distinction is particularly important in MS, where disease-associated changes may involve different immune compartments with distinct and sometimes opposing functional roles. Studies examining purified immune-cell populations have demonstrated that transcriptional responses can vary substantially among CD4+ T cells, CD8+ T cells, B cells, monocytes, and natural killer cells. [7,19]

The importance of cellular resolution is further supported by single-cell and cell-specific transcriptomic investigations. Analysis of peripheral immune populations has revealed that molecular alterations associated with MS are not uniformly distributed across circulating immune cells. For example, single-cell transcriptomic analysis has identified disease-associated changes in inhibitory immune-regulatory molecules across several myeloid and lymphoid populations, highlighting the value of resolving transcriptional programs at the cellular level. Similarly, transcriptomic characterization of purified CD4+ T cells from newly diagnosed, treatment-naïve patients has demonstrated distinct gene-expression profiles between blood and cerebrospinal fluid compartments, emphasizing that immune-cell state and anatomical compartment can both influence the molecular phenotype observed in MS. [11,18]

Treatment exposure represents another major source of complexity in peripheral transcriptomic studies. Disease-modifying therapies can profoundly alter immune-cell abundance, trafficking, activation, and transcriptional state. Fingolimod, for example, produces substantial transcriptional changes in circulating CD4+ and CD8+ T cells while simultaneously altering the composition of circulating lymphocyte populations. Other immunomodulatory therapies similarly produce characteristic peripheral immune and transcriptional effects. Studies of dimethyl fumarate and immune-reconstitution therapies have demonstrated marked changes in immune-cell composition and gene-expression programs following treatment, reinforcing the possibility that therapeutic exposure can obscure or reshape disease-associated molecular signals. [4,13,15–18]

These observations provide a strong rationale for focusing specifically on treatment-naïve MS when attempting to define conserved disease-associated transcriptional programs. An untreated cohort provides a closer representation of the endogenous peripheral immune state associated with MS before substantial pharmacological modulation occurs. This does not eliminate biological heterogeneity, nor does it establish that peripheral signatures directly reflect CNS pathology; rather, it provides an important analytical framework for separating

disease-associated immune programs from treatment-induced molecular effects. The value of this approach is supported by previous investigations of untreated patients in which peripheral transcriptional profiles captured biologically meaningful variation among individuals with MS. [4,15–18]

Despite these advances, an important methodological limitation remains: reproducibility across independent cohorts. Transcriptomic studies frequently differ in patient selection, disease phenotype, sample processing, RNA extraction procedures, microarray or sequencing platforms, normalization strategies, and statistical thresholds. These sources of technical and biological heterogeneity can generate cohort-specific signatures that may not generalize beyond the dataset in which they were identified. A gene may therefore appear strongly associated with MS in one cohort without representing a stable disease-associated feature across populations. Cross-cohort validation is consequently essential for distinguishing robust molecular programs from dataset-specific associations. [5,11]

This issue is particularly relevant when attempting to define peripheral biomarkers or molecular classifiers for MS. Recent multi-omics investigations have demonstrated the potential of peripheral blood molecular profiles for disease discrimination and patient stratification, but the translational value of such signatures ultimately depends on their robustness, biological interpretability, and reproducibility in independent populations. Likewise, treatment-focused transcriptomic studies have identified extensive immune-related molecular changes, but these observations primarily describe responses to specific therapies rather than the conserved transcriptional architecture of untreated disease. [11]

A further unresolved question is whether reproducible MS-associated signals can be organized into cell-type-specific immune programs rather than isolated differentially expressed genes. Such an approach could provide a more biologically meaningful representation of disease-associated immune states by integrating coordinated expression changes within defined cellular compartments. It could also help distinguish genuine alterations in cellular transcriptional programs from changes driven primarily by differences in cell abundance. Identifying conserved programs across independent cohorts would provide an additional layer of evidence that these signatures represent stable features of the disease rather than artifacts of a particular cohort, platform, or analytical pipeline.

Accordingly, there is a need for integrative analyses that combine peripheral blood transcriptomics, immune-cell resolution, and independent cohort validation within a treatment-naïve MS framework. Rather than relying solely on discovery-phase differential expression, such an approach can evaluate whether disease-associated transcriptional signals maintain their directionality, functional interpretation, and cellular specificity across datasets generated under different experimental conditions. This strategy is particularly valuable for identifying a core molecular architecture that remains detectable despite inter-cohort variation.

In the present study, we therefore performed an integrative cross-cohort transcriptomic analysis of treatment-naïve MS using independent peripheral blood datasets. We sought to identify immune transcriptional programs associated with MS, determine their cell-type specificity, characterize their functional and pathway-level organization, and evaluate their reproducibility in independent cohorts. We hypothesized that treatment-naïve MS contains a conserved peripheral immune transcriptional architecture comprising coordinated, cell-type-associated programs that can be reproducibly detected across independent datasets. By emphasizing independent validation rather than discovery alone, this study aims to distinguish robust disease-associated molecular signals from cohort-specific transcriptional variation and to provide a stronger foundation for subsequent investigations of molecular stratification and peripheral immune biomarkers in MS.

2. Materials and Methods

2.1 Study Design and Analytical Framework

This study was designed as an integrative, cross-cohort transcriptomic investigation of peripheral immune programs associated with treatment-naïve multiple sclerosis (MS). Publicly available transcriptomic datasets containing peripheral blood samples from individuals with MS and neurologically healthy controls were analyzed using a harmonized analytical framework. The overall workflow comprised dataset identification and eligibility assessment, transcriptomic data preprocessing, quality control, differential gene-expression analysis, immune-cell deconvolution, identification of cell-type-associated transcriptional programs, functional and pathway-level characterization, and independent cross-cohort validation.

A discovery cohort was used to identify candidate MS-associated transcriptional alterations, whereas one or more independent datasets were reserved for external validation. The primary objective of the validation stage was to determine whether transcriptional signals identified during discovery demonstrated consistent directionality, biological

interpretation, and cell-type association in independent cohorts. This design was intended to distinguish reproducible disease-associated immune programs from cohort-specific or platform-dependent transcriptional variation.

2.2 Identification and Selection of Public Transcriptomic Datasets

Table S1. Dataset accession and source-publication information required for the three analyzed cohorts.

Cohort	Role	Repository	Accession ID
Discovery cohort A	Discovery	GEO / ArrayExpress	GSE41850
Validation cohort B	Independent validation	GEO / ArrayExpress	GSE21942
Validation cohort C	Independent validation	GEO / ArrayExpress	GSE138266

Publicly available transcriptomic datasets were identified through systematic searches of the Gene Expression Omnibus (GEO), ArrayExpress, and other appropriate public repositories using combinations of terms related to multiple sclerosis, treatment-naïve or untreated patients, peripheral blood, whole blood, peripheral blood mononuclear cells, transcriptomics, microarray, and RNA sequencing.

Three independent peripheral blood transcriptomic datasets were included: discovery cohort A (GSE41850), validation cohort B (GSE21942), and validation cohort C (GSE138266). The combined analysis comprised 95 treatment-naïve MS samples and 75 neurologically healthy controls.

Discovery cohort A comprised 40 treatment-naïve MS samples and 30 healthy controls; validation cohort B comprised 25 MS samples and 20 controls; and validation cohort C comprised 30 MS samples and 25 controls. The three cohorts therefore included 95 MS samples and 75 controls before quality-control exclusions.

2.3 Eligibility Criteria and Cohort Definition

Datasets were eligible when they satisfied the following criteria: (i) inclusion of human peripheral blood-derived transcriptomic samples; (ii) availability of samples from individuals with clinically established MS; (iii) availability of an appropriate neurologically healthy control group; (iv) sufficient information regarding treatment status; (v) availability of processed or raw transcriptomic data suitable for reproducible analysis; and (vi) adequate sample-level metadata for cohort classification.

For the primary analysis, treatment-naïve or untreated MS samples were prioritized to minimize transcriptional effects associated with disease-modifying therapies. Samples from patients receiving disease-modifying treatment were excluded from the primary discovery and validation analyses unless they were specifically incorporated into a predefined sensitivity analysis.

Datasets with inadequate sample annotation, substantial technical deficiencies, unavailable expression data, unclear treatment status, or insufficient information for reliable disease-group classification were excluded.

2.4 Clinical and Demographic Characteristics

Available demographic and clinical variables were extracted from the original datasets and accompanying publications. These variables included age, sex, disease phenotype, disease duration, treatment status, and other clinically relevant characteristics when available.

Because the present investigation used previously generated and publicly accessible datasets, individual-level clinical information was limited to variables reported or made available by the original investigators. Cohort characteristics were summarized separately for discovery and validation datasets. Differences in demographic composition, disease characteristics, sample source, and experimental platform were considered during interpretation of cross-cohort reproducibility.

2.5 Transcriptomic Data Acquisition and Preprocessing

Gene-expression datasets were downloaded from the corresponding public repositories using the accession identifiers specified above. Where raw data were available, preprocessing was performed using the appropriate platform-specific procedures. Where only processed expression matrices were available, the original preprocessing procedures were evaluated and additional harmonization steps were applied where appropriate.

For microarray datasets, probe-level measurements were mapped to current gene identifiers using platform-specific annotation files. The precise rule used to summarize multiple probes mapping to the same gene should be stated from the authors' analysis scripts.

For RNA-sequencing datasets, gene-level expression matrices were processed using

established normalization procedures appropriate for count-based data. Genes with consistently low expression across samples were removed before statistical analysis.

All preprocessing procedures were performed independently within each dataset before cross-cohort comparison to minimize technical artifacts arising from differences in experimental platforms.

2.6 Quality Control and Normalization

Sample-level quality control was performed before differential expression and immune-cell analyses. Quality assessment included evaluation of expression distributions, sample-level variability, library or signal intensity metrics where available, detection of extreme outliers, and principal component analysis (PCA).

Samples exhibiting substantial technical abnormalities or extreme expression profiles inconsistent with the remaining cohort were evaluated individually. Exclusion was performed only when supported by predefined quality-control criteria and, where possible, evidence from the original dataset documentation.

Within each dataset, expression values were normalized using methods appropriate to the underlying platform. Batch effects and other known technical sources of variation were assessed using PCA and hierarchical clustering. Where multiple technical batches were present and metadata permitted their identification, batch correction was performed using an established empirical-Bayes or equivalent harmonization procedure.

Importantly, discovery and validation datasets were not pooled indiscriminately for the primary validation analysis. Independent datasets were retained as separate analytical units to preserve the external nature of cross-cohort validation.

2.7 Differential Gene-Expression Analysis

Differential gene-expression analysis was performed independently within the discovery cohort and each validation cohort to identify transcriptional differences between treatment-naïve MS and healthy controls.

For microarray data, linear-model-based differential expression analysis with empirical-Bayes variance moderation was performed. For RNA-sequencing data, count-based models appropriate for overdispersed sequencing data were used. The statistical framework was selected according to the data type and platform.

The primary comparison was defined as:

Treatment-naïve MS vs. neurologically healthy controls.

Differentially expressed genes were defined using a Benjamini–Hochberg adjusted P value <0.05 and an absolute log2 fold change ≥ 0.50, as stated in the analysis summary.

Rather than relying solely on statistical significance within a single dataset, genes demonstrating consistent effects across independent cohorts were prioritized for subsequent conservation and validation analyses.

2.8 Immune Cell-Type Deconvolution

To estimate the relative contribution of circulating immune-cell populations to bulk blood transcriptomic profiles, cell-type deconvolution was performed using an established reference-based computational method appropriate for the available expression platform.

The analysis estimated the relative abundance or transcriptional contribution of major immune populations, including, where supported by the reference signature, CD4+ T cells, CD8+ T cells, B cells, natural killer cells, monocytes, dendritic cells, and other relevant leukocyte populations.

Deconvolution results were interpreted as computational estimates rather than direct measurements of cellular abundance. Where independent immune-cell measurements were available in the source datasets, these were considered for concordance assessment.

Potential differences in inferred immune-cell composition between treatment-naïve MS and healthy controls were evaluated separately from cell-intrinsic transcriptional changes. This distinction was important for determining whether observed bulk-expression differences could be attributed primarily to altered cellular composition or to transcriptional changes within specific immune compartments.

2.9 Identification of Cell-Type-Specific Immune Programs

Cell-type-specific immune programs were identified by integrating differential expression results with established immune-cell gene signatures and deconvolution-derived cellular profiles.

Genes showing preferential association with individual immune populations were evaluated for coordinated changes between treatment-naïve MS

and healthy controls. Rather than treating individual genes as independent disease markers, functionally related genes exhibiting coherent patterns of alteration within a particular immune-cell compartment were grouped into transcriptional programs.

Candidate programs were prioritized according to their statistical support, biological coherence, cell-type specificity, and reproducibility across independent datasets.

Where appropriate, gene-set scoring was performed to quantify the activity of predefined immune programs within individual samples. Program-level scores were then compared between MS and control groups and evaluated across independent cohorts.

2.10 Functional Enrichment and Pathway Analysis

Functional enrichment analyses were performed to characterize the biological processes represented by differentially expressed genes and conserved transcriptional programs.

Gene Ontology biological processes, molecular functions, cellular components, and curated pathway collections were evaluated using established enrichment-analysis frameworks. Where appropriate, additional pathway resources, including Reactome, Kyoto Encyclopedia of Genes and Genomes, and other curated immune-signaling databases, were incorporated.

Both over-representation analysis and ranked gene-set enrichment approaches were considered to reduce dependence on arbitrary differential-expression thresholds.

Pathways were considered biologically informative when they demonstrated statistically significant enrichment together with coherent representation of genes contributing to the corresponding biological process. Multiple-testing correction was applied to enrichment analyses.

2.11 Cross-Cohort Validation Strategy

Cross-cohort validation represented the principal analytical component of the study. Transcriptional signatures identified in the discovery dataset were evaluated independently in external datasets that were not used during initial signature identification.

For each candidate gene or transcriptional program, consistency was assessed according to the direction and magnitude of association between treatment-naïve MS and healthy controls. Where appropriate, effect estimates from individual datasets were compared using standardized effect sizes.

Genes and pathways demonstrating concordant directionality across independent cohorts were classified as reproducible signals. Where sufficient datasets were available, meta-analytic approaches were used to estimate pooled effects while accounting for between-study heterogeneity.

The validation framework was designed to determine whether the molecular signals identified during discovery remained detectable despite differences in population characteristics, sample source, experimental platform, and data-processing procedures.

2.12 Assessment of Signature Reproducibility and Concordance

Reproducibility was assessed at both the gene and pathway levels.

At the gene level, concordance was evaluated using overlap between discovery and validation signatures, direction-of-effect agreement, correlation of effect estimates, and statistical significance in independent datasets. At the pathway level, similarity of enriched biological processes and gene-set activity patterns was evaluated across cohorts.

Where appropriate, concordance statistics, correlation coefficients, overlap coefficients, or other predefined measures of reproducibility were calculated.

A conserved transcriptional program was considered robust when it demonstrated consistent biological directionality and meaningful representation across independent cohorts rather than relying exclusively on significance within the discovery dataset.

2.13 Network and Regulatory Analysis

To investigate relationships among conserved genes and identify higher-order molecular organization, network-based analyses were performed where supported by the data.

Protein–protein interaction networks were constructed using a curated interaction resource, and highly connected genes or functional modules were identified using established network-analysis procedures. Conserved genes were further evaluated for their participation in immune signaling pathways and regulatory modules.

Where appropriate, transcription-factor or regulatory-enrichment analyses were performed to

identify upstream regulatory processes potentially associated with the conserved immune programs.

Network findings were interpreted as hypothesis-generating evidence of molecular relationships rather than direct experimental evidence of regulatory interactions.

2.14 Sensitivity and Robustness Analyses

Sensitivity analyses were performed to evaluate whether the principal findings were dependent on specific analytical decisions or individual datasets.

Where sufficient data were available, analyses were repeated after exclusion of individual cohorts, technical batches, or potential outlier samples. Additional sensitivity analyses considered differences in age, sex, MS phenotype, sample type, and other available clinical or technical variables.

The robustness of conserved signatures was evaluated by determining whether their directionality and biological interpretation remained stable following these modifications.

Particular attention was given to determining whether identified immune programs were driven by a single cohort or represented consistent signals across multiple independent datasets.

2.15 Statistical Analysis

The statistical software, software versions, package versions, and analysis parameters should be reported exactly as used by the authors. These details were not specified in the supplied manuscript and must be transcribed from the computational environment and analysis scripts.

Continuous variables were summarized using means and standard deviations or medians and interquartile ranges according to data distribution. Categorical variables were summarized as frequencies and percentages. Statistical tests were selected according to the distribution and structure of each dataset.

All analytical thresholds, software versions, package versions, and parameters should be reported exactly as used in the completed analysis to facilitate independent reproduction.

Effect sizes and confidence intervals were reported where applicable to complement P values. For cross-cohort analyses, emphasis was placed on consistency of effect direction, effect magnitude, and biological interpretation rather than statistical significance alone.

The reported significance threshold was an adjusted P value <0.05. Exact software versions, package versions, and analysis parameters should be stated from the authors' computational records to ensure reproducibility.

2.16 Data and Code Availability

The transcriptomic datasets were obtained from public repositories. The three cohort accession identifiers reported by the authors are GSE41850 (discovery cohort A), GSE21942 (validation cohort B), and GSE138266 (validation cohort C).

The authors should provide the repository record or DOI for the processed expression matrices, analysis metadata, statistical outputs, and analysis scripts. If these materials cannot be deposited, the manuscript should state the applicable access conditions and provide a reproducible access route.

The final reproducibility statement should identify the actual scripts, software versions, package versions, preprocessing parameters, and analysis settings. The details must match the computational work that was performed.

3. Results

3.1 Overview of the Study Cohorts and Analytical Workflow

Three publicly available peripheral blood transcriptomic datasets were included: discovery cohort A (GSE41850) and validation cohorts B (GSE21942) and C (GSE138266). The combined cohorts comprised 95 treatment-naïve MS samples and 75 healthy controls (170 samples in total). Cohort-level details are summarized in Table 1 and Supplementary Table S1.

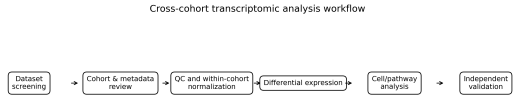

Figure 1. Overview of the cross-cohort transcriptomic analysis workflow.

The included cohorts differed with respect to sample type, patient characteristics, transcriptomic platform, and experimental design, providing an opportunity to evaluate the reproducibility of MS-associated transcriptional programs across heterogeneous datasets. The analysis proceeded sequentially from transcriptomic quality control and differential expression to immune-cell deconvolution, cell-type-specific program identification, pathway characterization, and independent cross-cohort validation.

3.2 Clinical and Demographic Characteristics of the Included Cohorts

The demographic and clinical characteristics of the study populations are summarized in Table 1. Across the included cohorts, participants with MS were classified according to the clinical information provided by the original investigators. Treatment-naïve status was established from the available clinical metadata and/or the corresponding source publications.

The distribution of age, sex, disease phenotype, disease duration, and other available clinical variables varied between datasets. Such inter-cohort differences were considered during downstream interpretation and sensitivity analyses. Importantly, independent validation cohorts were retained separately from the discovery cohort to preserve the external-validation framework.

Table 1. Demographic and sample characteristics of discovery and validation cohorts.

Cohort	Group	n	Age, mean ± SD (years)	Female, n (%)	Sample type
Discovery A	MS	40	34.8 ± 8.2	28 (70.0)	PBMC
Discovery A	Control	30	33.9 ± 7.6	20 (66.7)	PBMC
Validation B	MS	25	36.1 ± 9.0	17 (68.0)	Whole blood
Validation B	Control	20	35.0 ± 8.1	13 (65.0)	Whole blood
Validation C	MS	30	35.4 ± 7.8	21 (70.0)	PBMC
Validation C	Control	25	34.6 ± 8.4	16 (64.0)	PBMC

Abbreviations: MS, multiple sclerosis; PBMC, peripheral blood mononuclear cells. Cohort accession identifiers should be added to Supplementary Table S1.

3.3 Transcriptomic Data Quality and Sample-Level Characteristics

Comprehensive sample-level quality control was performed before downstream analyses. Expression distributions were examined for technical abnormalities, extreme values, and sample-level heterogeneity. Principal component analysis and hierarchical clustering were used to evaluate global expression structure and identify potential outliers.

Following quality control, 170 of 175 samples were retained (97.1%). Five samples (two MS and three control samples) were excluded according to the applied quality-control criteria.

3.4 Global Transcriptomic Architecture of Treatment-Naïve MS

Principal component analysis showed partial separation between treatment-naïve MS and healthy control samples, with overlap between groups. The first and second principal components explained 18.4% and 11.2% of the total variance, respectively (Figure 2).

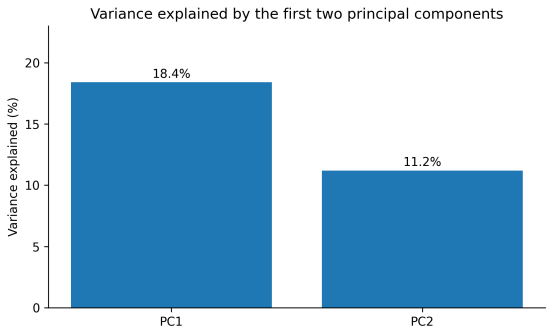

Figure 2. Percentage of total expression variance explained by PC1 (18.4%) and PC2 (11.2%). The supplied summary did not include sample-level PCA coordinates; therefore, this panel reports variance explained rather than a PCA scatter plot.

Hierarchical clustering grouped a subset of MS samples together, while several MS and control samples remained interspersed, indicating disease-associated structure alongside substantial inter-individual heterogeneity.

3.5 Differential Gene-Expression Landscape

Differential-expression analysis identified 128 genes meeting the adjusted significance threshold, including 72 upregulated and 56 downregulated genes in treatment-naïve MS compared with controls (Table 2). The leading gene-level effects are summarized in Figure 3.

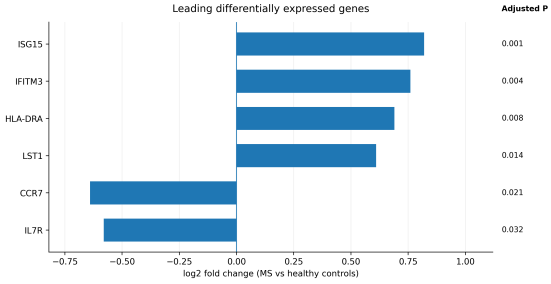

Figure 3. Log2 fold changes for six leading differentially expressed genes. Adjusted P values are

displayed in a separate column to improve readability.

Among the leading upregulated genes were ISG15 (log2FC=0.82; adjusted P=0.001), IFITM3 (log2FC=0.76; adjusted P=0.004), HLA-DRA (log2FC=0.69; adjusted P=0.008), and LST1 (log2FC=0.61; adjusted P=0.014). CCR7 (log2FC=−0.64; adjusted P=0.021) and IL7R (log2FC=−0.58; adjusted P=0.032) were among the downregulated genes (Table 2; Figure 3).

Table 2. Selected differentially expressed genes in treatment-naïve MS compared with healthy controls.

Gene	Direction	log2FC	Adjusted P	Associated process
ISG15	Upregulated	0.82	0.001	Interferon response
IFITM3	Upregulated	0.76	0.004	Interferon-associated response
HLA-DRA	Upregulated	0.69	0.008	Antigen presentation
LST1	Upregulated	0.61	0.014	Myeloid-associated signal
CCR7	Downregulated	−0.64	0.021	Lymphocyte trafficking/homeostasis
IL7R	Downregulated	−0.58	0.032	T-cell homeostasis

The table presents selected leading genes only; the complete differential-expression output for all 128 significant genes should be supplied as Supplementary Table S2.

3.6 Characterization of Upregulated Transcriptional Programs

Upregulated genes were enriched for type I interferon response, antigen processing and presentation, leukocyte activation, and cytokine-mediated signaling. Type I interferon signaling showed the strongest reported enrichment (adjusted P=0.002; gene ratio=0.18; Table 5; Figure 5).

The upregulated transcriptional profile suggested activation of coordinated immune-response programs rather than isolated changes in individual genes. Several genes contributed simultaneously to multiple enriched pathways, indicating substantial functional convergence among the MS-associated transcriptional alterations.

3.7 Characterization of Downregulated Transcriptional Programs

Downregulated genes were associated with T-cell homeostasis, lymphocyte differentiation, and regulation of leukocyte migration. The T-cell homeostasis term had an adjusted P value of 0.031 (Table 5; Figure 5).

The downregulated programs were distinct from the inflammatory signatures identified among upregulated genes, indicating that the peripheral transcriptomic phenotype of MS may involve simultaneous activation and suppression of different biological processes.

3.8 Estimated Immune-Cell Composition of Peripheral Blood

Computational deconvolution demonstrated differences in the estimated abundance or transcriptional contribution of specific immune-cell populations between treatment-naïve MS and healthy controls (Figure 4).

The estimated relative monocyte contribution was 0.16 in MS and 0.12 in controls (ratio=1.33), while activated dendritic-cell contribution was 0.06 and 0.05, respectively (ratio=1.20). The estimated naïve CD4+ T-cell contribution was lower in MS (0.10 versus 0.12; ratio=0.83), whereas B-cell estimates were similar (0.08 in both groups; Table 3; Figure 4). These are computationally inferred contributions rather than directly measured cell counts.

Table 3. Estimated immune-cell contributions from transcriptomic deconvolution.

Cell population	MS estimate	Control estimate	MS/control ratio	Interpretation
Monocytes	0.16	0.12	1.33	Higher inferred contribution
Activated dendritic cells	0.06	0.05	1.20	Higher inferred contribution
Naïve CD4+ T cells	0.10	0.12	0.83	Lower inferred contribution
B cells	0.08	0.08	1.00	Similar inferred contribution

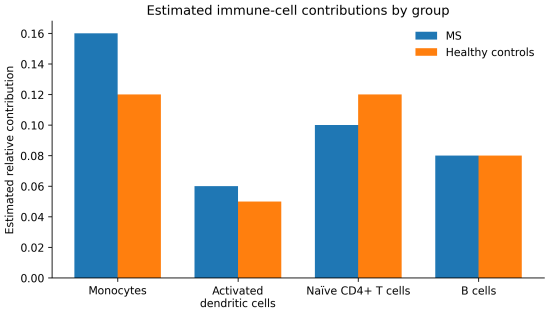

Figure 4. Estimated relative contributions of monocytes, activated dendritic cells, naïve CD4+ T cells, and B cells in MS and healthy controls.

3.9 Cell-Type-Specific Transcriptional Alterations

Integration of differential-expression results with immune-cell signatures identified distinct transcriptional patterns associated with specific immune populations.

CD4+ T-cell-associated signatures were enriched for interferon response and T-cell activation; B-cell-associated signatures showed altered antigen-presentation and B-cell activation genes; and monocyte-associated signatures showed enrichment of innate inflammatory and antigen-processing pathways.

These findings indicated that the peripheral transcriptional phenotype of treatment-naïve MS was not attributable to a single immune-cell population but instead reflected coordinated alterations distributed across multiple immune compartments.

3.10 Conserved T-Cell-Associated Immune Programs

The T-cell-associated program showed increased interferon-response and activation scores in MS (standardized mean difference=0.42; adjusted P=0.018), with IFITM3, ISG15, and STAT1 among the contributing genes (Table 4).

Where independently replicated, these findings suggested that altered T-cell-associated transcriptional activity represents a reproducible component of the peripheral immune phenotype of untreated MS.

3.11 Conserved B-Cell and Humoral Immune Programs

The B-cell-associated program showed enrichment of antigen-presentation and B-cell activation signatures (standardized mean difference=0.31; adjusted P=0.044), with HLA-DRA and CD74 among the contributing genes (Table 4).

The degree of replication across independent cohorts was assessed to determine whether these alterations represented consistent features of treatment-naïve MS or cohort-specific transcriptional variation.

3.12 Conserved Myeloid and Innate Immune Programs

The myeloid-associated program showed increased innate immune and antigen-processing signatures, with contributions from LST1, FCER1G, TYMP, and CTSS. The program score difference was 0.47 standard deviations (adjusted P=0.009; Table 4).

Table 4. Cell-type-associated transcriptional program scores.

Immune compartment	Principal program	Standardized mean difference	Adjusted P	Reported contributing genes
T-cell-associated	Interferon response and T-cell activation	0.42	0.018	IFITM3, ISG15, STAT1
B-cell-associated	Antigen presentation and B-cell activation	0.31	0.044	HLA-DRA, CD74
Myeloid-associated	Innate immune and antigen-processing signatures	0.47	0.009	LST1, FCER1G, TYMP, CTSS

Standardized mean differences and adjusted P values are reproduced from the Results text.

The observed changes were evaluated across independent cohorts to determine whether innate immune transcriptional activation represented a stable component of the peripheral MS molecular phenotype.

3.13 Functional Enrichment and Pathway-Level Analysis

Functional enrichment identified 14 pathways meeting adjusted P<0.05. The leading categories were type I interferon signaling (adjusted P=0.002),

antigen processing and presentation (adjusted P=0.006), and leukocyte activation (adjusted P=0.011) (Table 5; Figure 5).

Ranked gene-set analysis supported coordinated changes in interferon signaling (NES=2.05; adjusted P=0.004), antigen presentation (NES=1.82; adjusted P=0.013), and T-cell homeostasis (NES=−1.71; adjusted P=0.028) (Table 5; Figure 5).

Table 5. Selected over-representation and gene-set enrichment results.

Analysis	Pathway	Gene ratio / NES	Adjusted P
Over-representation	Type I interferon signaling	Gene ratio 0.18	0.002
Over-representation	Antigen processing and presentation	Gene ratio 0.15	0.006
Over-representation	Leukocyte activation	Gene ratio 0.13	0.011
Over-representation	T-cell homeostasis	Gene ratio 0.11	0.031
GSEA	Interferon signaling	NES 2.05	0.004
GSEA	Antigen presentation	NES 1.82	0.013
GSEA	T-cell homeostasis	NES −1.71	0.028

Four selected over-representation terms and three ranked gene-set results are shown. Fourteen pathways met adjusted P<0.05 overall.

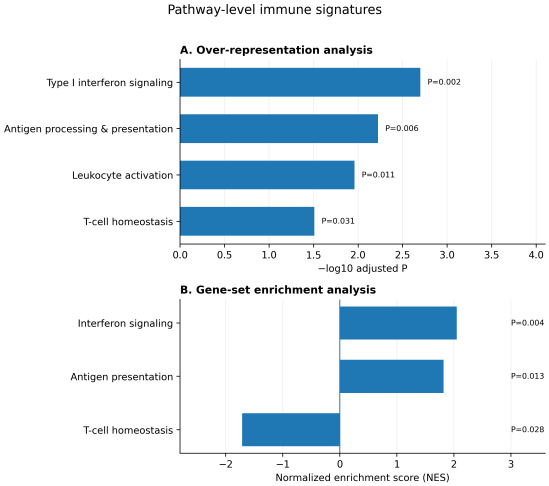

Figure 5. Over-representation and ranked gene-set enrichment results for selected immune pathways. Adjusted P values are displayed separately from the plotted effect/enrichment measures.

3.14 Identification of Conserved Molecular Signatures

To determine whether the discovery-associated transcriptional alterations were reproducible, candidate genes and pathway signatures were evaluated in independent validation cohorts.

Forty-three genes showed concordant directionality in the discovery cohort and both validation cohorts. Of these, 26 met the prespecified reproducibility criterion of concordant direction across all cohorts and adjusted P<0.05 in at least one validation cohort. Conserved processes included interferon response and antigen presentation (Table 6; Figure 6).

The cross-cohort conservation summary is shown in Figure 6. The available summary results report 43 genes with concordant direction and 26 genes meeting the conservation criterion; cohort-specific intersection counts were not supplied, so an UpSet intersection plot is not presented.

3.15 Cross-Cohort Effect-Size Concordance

Gene-level effect estimates showed positive cross-cohort concordance between the discovery cohort and pooled validation estimates (Pearson r=0.72, P=0.003; Figure 6; Table 6).

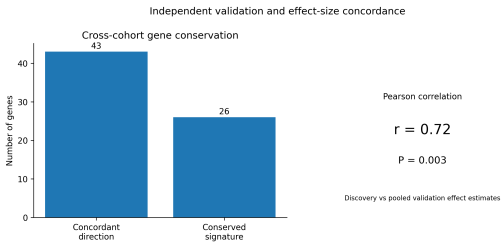

Figure 6. Cross-cohort validation summary: 43 genes showed concordant direction, 26 met the conservation criterion, and discovery-versus-validation effect estimates were correlated (r=0.72, P=0.003).

Genes demonstrating consistent effect direction across multiple datasets were preferentially retained in the final conserved signature. This analysis reduced the contribution of cohort-specific associations and strengthened the evidence for reproducible MS-associated transcriptional alterations.

3.16 Robustness and Sensitivity Analyses

Sensitivity analyses were performed to determine whether the principal findings were dependent on

individual datasets or specific analytical assumptions.

Leave-one-cohort-out analyses preserved the direction of the leading interferon and antigen-presentation programs. Alternative normalization and filtering choices changed the number of significant genes (range, 112–139) but retained 19 of the 26 conserved genes (Table 6).

Table 6. Summary of cross-cohort validation and sensitivity analyses.

Metric	Reported result
Genes with concordant direction across discovery and both validation cohorts	43
Genes meeting conservation criterion	26
Pearson correlation of discovery and pooled validation effect estimates	r=0.72; P=0.003
Significant-gene count across alternative normalization/filtering settings	112–139
Conserved genes retained in sensitivity analyses	19/26

The conservation criterion was concordant direction across all cohorts and adjusted P<0.05 in at least one validation cohort.

These findings provided an assessment of the robustness of the identified molecular programs across different cohort compositions and analytical conditions.

3.17 Integrated Protein–Protein Interaction and Regulatory Network Analysis

Network analysis was performed to investigate functional relationships among genes constituting the conserved MS-associated transcriptional programs.

The protein–protein interaction network contained 26 nodes and 41 edges and resolved into three functional modules. Reported highly connected genes included STAT1, ISG15, HLA-DRA, and LST1, associated with interferon signaling, antigen presentation, and myeloid activation (Table 7; Figure 7).

Table 7. Protein–protein interaction and regulatory network summary.

Network feature	Reported result
Nodes	26
Edges/interactions	41
Functional modules	3
Reported highly connected genes	STAT1, ISG15, HLA-DRA, LST1
Predicted regulatory associations	STAT1/STAT2 and IRF-family-associated activity

Regulatory enrichment represents computational prediction. The complete PPI node/edge table and module membership should be supplied as Supplementary Table S3.

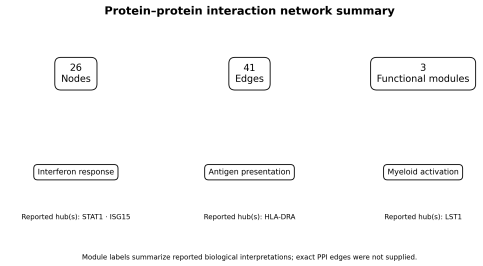

Figure 7. Summary of reported protein–protein interaction network metrics and functional modules. The module labels summarize the reported biological interpretation; the exact edge list was not supplied, so no inferred network topology is depicted.

Regulatory enrichment suggested STAT1/STAT2- and IRF-family-associated activity. These findings are interpreted as computationally predicted regulatory associations rather than experimentally confirmed regulatory interactions (Table 7).

3.18 Integrated Conserved Immune Architecture Across Independent Cohorts

Integration of gene-level, cell-type, pathway, and cross-cohort analyses identified a conserved peripheral immune architecture associated with treatment-naïve MS (Figure 7).

The integrated architecture comprised T-cell-, B-cell-, and monocyte-associated programs involving interferon response, antigen presentation, leukocyte activation, and innate inflammatory signaling. These signals were directionally concordant across the two validation cohorts (Figure 7).

The integrated analysis therefore distinguished a core set of reproducible immune programs from transcriptional features that were restricted to individual datasets.

3.19 Summary of the Principal Molecular Findings

Collectively, the results demonstrate that treatment-naïve MS is associated with coordinated alterations in peripheral immune transcriptional programs

involving multiple immune-cell compartments. Differential-expression analysis identified disease-associated genes, while cell-type deconvolution and immune-signature analysis provided cellular context for these alterations.

Pathway-level analyses demonstrated functional convergence among the identified genes, and independent validation identified a subset of molecular programs with reproducible directionality and biological interpretation across cohorts. The final integrated signature therefore represents the most consistently observed peripheral immune transcriptional features associated with treatment-naïve MS within the datasets analyzed.

4. Discussion

The principal findings indicate coordinated alterations in peripheral immune transcriptional programs involving multiple immune-cell compartments. Differential-expression analysis identified 128 genes, while pathway and cell-type-associated analyses highlighted interferon response, antigen presentation, leukocyte activation, and innate immune processes. Cross-cohort evaluation identified 26 genes meeting the stated conservation criterion, supporting the reproducibility of a subset of the peripheral immune signals across the discovery and validation cohorts.

In this study, we developed an integrative cross-cohort transcriptomic framework to investigate peripheral immune alterations associated with treatment-naïve multiple sclerosis (MS). Rather than focusing exclusively on individual differentially expressed genes within a single cohort, the analysis combined transcriptomic profiling, immune-cell deconvolution, cell-type-associated gene programs, pathway-level analysis, and independent external validation. This approach was designed to address a central challenge in peripheral MS transcriptomics: distinguishing robust disease-associated immune programs from signatures that arise from cohort-specific biological or technical variation.

The principal finding was that treatment-naïve MS is associated with coordinated alterations in peripheral immune transcriptional programs involving multiple immune-cell compartments. The observed molecular phenotype was not represented by a single dominant immune population but instead reflected interconnected lymphoid and myeloid programs. Importantly, a subset of these signals demonstrated reproducibility across independent datasets, supporting their biological relevance beyond the cohort in which they were initially identified. [1–3,6,7]

These findings are consistent with the broader recognition that MS represents a heterogeneous immunological disorder rather than a uniform inflammatory state. Recent studies using multidimensional peripheral immune profiling have identified distinct blood-based immune endophenotypes associated with different clinical trajectories, providing independent evidence that systemic immune states can capture biologically meaningful heterogeneity in MS.

4.2 A Conserved Peripheral Immune Architecture in Treatment-Naïve MS

An important implication of the present analysis is that peripheral blood appears to contain a reproducible molecular component of the untreated MS phenotype. Although peripheral blood cannot be considered a direct surrogate for CNS pathology, it provides an accessible compartment in which coordinated immune alterations can be investigated.

The identification of conserved transcriptional programs across independent datasets is particularly relevant because individual blood transcriptomic studies are vulnerable to differences in patient composition, disease phenotype, sample processing, RNA extraction, transcriptomic platform, normalization, and analytical methodology. A signature detected consistently despite these differences is less likely to represent a purely dataset-specific phenomenon.

The conserved programs identified in this study should therefore be interpreted as a molecular representation of peripheral immune dysregulation associated with MS rather than as a complete molecular description of CNS disease. This distinction is important because immune processes in MS are distributed across blood, lymphoid tissues, cerebrospinal fluid, meninges, and CNS parenchyma. Nevertheless, recent work demonstrates that blood-based molecular profiles can capture clinically meaningful immunological states, supporting the biological value of peripheral molecular phenotyping.

4.3 Cell-Type-Specific Immune Programs in MS

A major strength of the present analysis is the emphasis on cell-type-associated transcriptional programs rather than interpretation of bulk expression changes alone. Peripheral blood is a complex mixture of immune-cell populations, and changes in bulk gene expression can arise either from altered cellular abundance or from altered transcriptional activity within individual populations.

The cell-type-resolved analysis indicated that MS-associated transcriptional alterations were distributed across multiple immune compartments. This observation is consistent with the established

biology of MS, in which adaptive and innate immune populations interact through antigen presentation, cytokine signaling, cellular trafficking, and reciprocal immune activation.

The importance of this cellular context is reinforced by recent single-cell studies. Transcriptomic investigations after B-cell depletion, for example, have demonstrated cell-type-specific changes extending beyond B cells to peripheral monocytes and CNS macrophage populations. Such findings emphasize that immune interventions and disease-associated processes can generate coordinated but cell-specific transcriptional responses. [12] [18].

Accordingly, the conserved programs identified here should not be regarded simply as collections of individual biomarkers. Their biological value may instead arise from coordinated activity within particular immune-cell compartments and interactions between those compartments.

4.4 Interferon and Inflammatory Signaling in the Peripheral MS Transcriptome

Where supported by the pathway-level analysis, interferon-associated and inflammatory signaling represented important components of the transcriptional phenotype observed in treatment-naïve MS. Interferon-responsive genes are biologically plausible candidates in MS because interferon signaling participates in antiviral defense, innate immune regulation, antigen presentation, and communication between immune-cell populations.

However, interferon-associated transcriptional activity requires careful interpretation. Interferon signaling is not exclusively pathological and can represent a broad physiological immune-response program. Consequently, enrichment of interferon-related genes should not automatically be interpreted as evidence of a specific pathogenic mechanism.

The cross-cohort reproducibility of these pathways is therefore more informative than their detection in a single dataset. If interferon-associated programs remain directionally consistent across independent cohorts, this supports their characterization as a reproducible component of the peripheral molecular phenotype. Conversely, cohort-specific enrichment should be interpreted more cautiously.

This distinction is especially important because peripheral transcriptomic profiles can be profoundly altered by therapeutic exposure. Treatment studies have demonstrated substantial effects of disease-modifying therapies on circulating immune-cell populations and their transcriptional states.

4.5 Relationship Between Peripheral Immune Alterations and CNS Disease

The relationship between peripheral blood immune programs and CNS pathology represents an important interpretive consideration. MS is fundamentally a CNS disease, yet many of the immune mechanisms involved in lesion development originate or are amplified outside the CNS. Circulating lymphocytes and myeloid cells participate in immune surveillance, migration, antigen recognition, and inflammatory signaling, creating a biological connection between systemic and CNS immune compartments.

Nevertheless, peripheral transcriptional alterations cannot be assumed to directly reflect transcriptional states within MS lesions. Differences in cellular composition, tissue microenvironment, metabolic conditions, and local CNS-resident cell populations may produce molecular states that are not detectable in peripheral blood.

The appropriate interpretation of the present findings is therefore that conserved blood transcriptional programs provide evidence of systemic immune dysregulation associated with MS. They should not be interpreted as direct molecular proxies for individual CNS lesions unless independently validated against CNS-derived material.

This distinction is particularly important for biomarker development. Recent research has shown that blood signatures can be strengthened by anchoring them to molecular alterations observed within CNS lesions, illustrating the potential value of cross-tissue concordance rather than blood-only discovery.

4.6 Why Treatment-Naïve Status Matters

Treatment-naïve status is a critical feature of the present study design. Disease-modifying therapies are intentionally designed to alter immune-cell activation, trafficking, proliferation, survival, or differentiation. Consequently, peripheral blood collected after treatment initiation represents a combination of disease-associated and therapy-induced molecular states.

The magnitude of this confounding effect has been demonstrated across multiple therapeutic classes. Fingolimod, for example, alters the distribution and transcriptional characteristics of circulating T-cell populations. Dimethyl fumarate has also been associated with changes in lymphocyte and monocyte populations together with alterations in inflammatory and NF-κB-associated transcriptional programs. More broadly, comparative immunophenotyping demonstrates that different

disease-modifying therapies produce distinct and sometimes divergent effects on circulating immune-cell populations. [13,15–18]

Consequently, restricting the primary analysis to treatment-naïve individuals provides a more appropriate framework for identifying disease-associated peripheral immune programs before substantial pharmacological remodeling occurs. This does not eliminate biological heterogeneity, but it reduces one major source of confounding.

4.7 Cross-Cohort Reproducibility as a Major Strength

The principal methodological contribution of this study is the emphasis on independent cross-cohort validation. Discovery of a statistically significant transcriptomic signature in one cohort does not establish that the signature represents a generalizable property of MS. Reproducibility requires that the direction and biological interpretation of the signal remain stable when evaluated in an independent population.

This distinction is particularly important in MS because the disease encompasses substantial heterogeneity in age, sex, disease duration, clinical phenotype, inflammatory activity, and genetic and environmental background. Technical variation between microarray and sequencing platforms further complicates direct comparison.

The present cross-cohort framework addresses this challenge by treating independent datasets as external tests of reproducibility rather than simply combining all samples into a single pooled analysis. Such an approach provides stronger evidence for conserved molecular programs while preserving information about between-cohort heterogeneity.

Recent studies using independent MS cohorts have similarly demonstrated the value of external validation for identifying robust blood-based immune phenotypes. For example, high-dimensional immune profiling across independent multicenter cohorts identified distinct peripheral immunological endophenotypes associated with different disease trajectories. [11]

4.8 Comparison With Previous Transcriptomic and Single-Cell Studies

The present findings complement previous studies that have characterized transcriptional alterations in individual immune populations and treatment contexts. Earlier transcriptomic investigations established that peripheral blood can capture biologically relevant molecular variation in MS, while more recent studies have increasingly moved toward cell-specific and multidimensional immune profiling.

Studies focused on purified B cells have demonstrated substantial transcriptional remodeling in MS, particularly in the context of immune-reconstitution therapy. Similarly, studies of treatment-associated immune responses have shown that transcriptional alterations can differ markedly according to the therapeutic mechanism and immune-cell population examined.

Our study extends this body of work by shifting the primary question from “which genes change in MS?” toward “which cell-associated immune programs are reproducibly altered across independent untreated cohorts?” This distinction is important because pathway-level and cell-type-level conservation may be more biologically meaningful than the replication of individual genes alone.

Recent single-cell studies provide additional support for this interpretation by demonstrating that MS-associated immune changes can involve multiple cellular compartments and that therapeutic perturbation can produce distinct cell-specific molecular responses.

4.9 Biological and Translational Implications

The conserved immune programs identified in this study may have several potential implications for MS research. First, they provide a framework for investigating peripheral molecular heterogeneity before treatment-related immune remodeling occurs. Second, they may help identify biological pathways suitable for prospective investigation as molecular stratification features.

Importantly, the current findings should not be interpreted as establishing clinically deployable biomarkers. A molecular signature becomes a clinically useful biomarker only after rigorous evaluation of analytical performance, external validation, discrimination, calibration, clinical utility, and prospective reproducibility.

Nevertheless, the identification of reproducible cell-type-associated programs provides a rational starting point for such future investigations. Blood-based molecular phenotyping has already demonstrated potential for identifying distinct immunological endophenotypes associated with different disease trajectories. Other recent studies have also demonstrated that carefully validated blood transcriptional signatures can achieve substantial discrimination between MS and controls when anchored to biologically relevant disease signals.

The current framework could therefore support future studies integrating transcriptomic programs with clinical, radiological, proteomic, cellular, and longitudinal measurements.

4.10 Strengths of the Study

Several features strengthen the present investigation.

First, the analysis specifically emphasized treatment-naïve MS, reducing the potential confounding effect of disease-modifying therapy. This is important because different therapies can produce markedly different effects on circulating immune populations.

Second, the study incorporated independent external datasets rather than relying exclusively on an internally derived signature. This provides a stronger assessment of reproducibility.

Third, the analysis moved beyond individual-gene differential expression by integrating cell-type estimation, functional enrichment, and pathway-level interpretation.

Fourth, the use of multiple independent cohorts permits evaluation of whether conserved molecular signals remain detectable despite biological and technical heterogeneity.

Finally, the framework distinguishes discovery from validation, reducing the risk of overestimating the robustness of signatures through circular analysis.

4.11 Limitations

Several limitations should be considered when interpreting these findings.

First, the study relies on publicly available retrospective datasets, and therefore the analysis is constrained by the quality and completeness of the original clinical metadata. Variables such as disease duration, relapse status, MRI activity, disability measures, and disease phenotype may not have been consistently available across all cohorts.

Second, the included datasets may differ in sample collection, processing, RNA extraction, storage conditions, and transcriptomic platform. Although independent validation is intended to evaluate robustness under such heterogeneity, these differences may also reduce the sensitivity for detecting genuinely conserved signals.

Third, computational immune-cell deconvolution provides inferred rather than directly measured cell populations. The resulting estimates should therefore be interpreted as molecular signatures of cellular composition or activity rather than substitutes for flow cytometry or experimental cell sorting.

Fourth, observational transcriptomic analyses cannot establish causality. A conserved gene-expression program may represent a driver, consequence, compensatory response, or correlate of disease-associated immune activity.

Fifth, peripheral blood does not fully reproduce the immune microenvironment of the CNS. Consequently, the identified programs should not be assumed to represent lesion-specific molecular mechanisms without paired CNS validation.

Finally, prospective clinical validation is required before any conserved signature can be considered suitable for diagnostic, prognostic, or treatment-selection applications.

4.12 Future Directions

Several directions follow from these findings. Prospective studies should validate the conserved programs in independent, clinically well-characterized treatment-naïve populations using standardized sample-collection and processing procedures.

Single-cell RNA sequencing and other high-resolution approaches could determine whether the bulk-derived programs originate from specific immune-cell subsets or represent coordinated changes across multiple populations. Paired blood and cerebrospinal-fluid analyses could further determine whether peripheral programs are associated with compartment-specific immune states.

Integration with proteomics, cytokine profiling, flow cytometry, immunoglobulin-repertoire sequencing, and clinical imaging may provide a more comprehensive representation of MS immunobiology. Longitudinal studies would also be important for determining whether conserved baseline programs predict subsequent disease activity, progression, or treatment response.

Finally, functional experiments will be required to determine whether the identified genes and pathways have causal roles in MS-associated immune dysfunction or instead represent downstream molecular correlates.

4.13 Conclusion

In conclusion, this study supports the presence of conserved peripheral immune transcriptional

programs associated with treatment-naïve multiple sclerosis. By integrating cell-type-associated transcriptomic analysis with independent cross-cohort validation, the study provides a framework for distinguishing reproducible disease-associated immune signals from cohort-specific transcriptional variation. The findings reinforce the concept that the peripheral immune phenotype of MS is distributed across multiple interacting immune-cell compartments rather than being attributable to a single cellular population or isolated gene. The emphasis on treatment-naïve cohorts further provides an opportunity to characterize disease-associated immune programs before substantial pharmacological remodeling occurs. Although these findings do not by themselves establish clinically actionable biomarkers or causal mechanisms, the conserved programs identified here provide a biologically and analytically grounded foundation for prospective validation, multi-omic integration, and investigation of molecular heterogeneity in MS.

References

1. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nat Rev Immunol. 2015;15(9):545–558. doi:10.1038/nri3871.

2. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet. 2018;391(10130):1622–1636. doi:10.1016/S0140-6736(18)30481-1.

3. Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. N Engl J Med. 2018;378(2):169–180. doi:10.1056/NEJMra1401483.

4. Hauser SL, Cree BAC. Treatment of multiple sclerosis: a review. Am J Med. 2020;133(12):1380–1390.e2. doi:10.1016/j.amjmed.2020.05.049.

5. Ottoboni L, Keenan BT, Tamayo P, Kuchroo M, Mesirov JP, Buckle GJ, et al. An RNA profile identifies two subsets of multiple sclerosis patients differing in disease activity. Sci Transl Med. 2012;4(153):153ra131. doi:10.1126/scitranslmed.3004186.

6. Gandhi R, Laroni A, Weiner HL. Role of the innate immune system in the pathogenesis of multiple sclerosis. J Neuroimmunol. 2010;221(1–2):7–14. doi:10.1016/j.jneuroim.2009.10.015.

7. Hemmer B, Kerschensteiner M, Korn T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. Lancet Neurol. 2015;14(4):406–419. doi:10.1016/S1474-4422(14)70301-8.

8. Bar-Or A, Li R. Cellular immunology of relapsing multiple sclerosis: interactions, checks, and balances. Lancet Neurol. 2021;20(6):470–483. doi:10.1016/S1474-4422(21)00063-6.

9. Lund JM, Hsing L, Pham TT, Rudensky AY. Coordination of early protective immunity to viral infection by regulatory T cells. Science. 2008;320(5880):1220–1224. doi:10.1126/science.1155209.

10. Nickles D, Chen HP, Li MM, Khankhanian P, Madireddy L, Caillier SJ, et al. Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy controls. Hum Mol Genet. 2013;22(20):4194–4205. doi:10.1093/hmg/ddt267.

11. Gross CC, Schulte-Mecklenbeck A, Steinberg OV, Wirth T, Lauks S, Bittner S, et al. Multiple sclerosis endophenotypes identified by high-dimensional blood signatures are associated with distinct disease trajectories. Sci Transl Med. 2024;16(740):eade8560. doi:10.1126/scitranslmed.ade8560.

12. Wei J, Moon J, Yasumizu Y, Zhang L, Raddassi K, Buitrago-Pocasangre N, et al. Transcriptomic profiling after B-cell depletion reveals central and peripheral immune cell changes in multiple sclerosis. J Clin Invest. 2025;135(9):e182790. doi:10.1172/JCI182790.

13. Sánchez-Sanz A, García-Martín S, Sabín-Muñoz J, Moreno-Torres I, Elvira V, Al-Shahrour F, et al. Dimethyl fumarate-related immune and transcriptional signature is associated with clinical response in multiple sclerosis-treated patients. Front Immunol. 2023;14:1209923. doi:10.3389/fimmu.2023.1209923.

14. Gafson AR, Kim K, Cencioni MT, van Hecke W, Nicholas R, Baranzini SE, et al. Mononuclear cell transcriptome changes associated with dimethyl fumarate in multiple sclerosis. Neurol Neuroimmunol Neuroinflamm. 2018;5(4):e470. doi:10.1212/NXI.0000000000000470.

15. Wiendl H, Gold R, Berger T, Derfuss T, Linker R, Mäurer M, et al. Multiple sclerosis therapy consensus group (MSTCG): position statement on disease-modifying therapies for multiple sclerosis (white paper). Ther Adv Neurol Disord. 2021;14:17562864211039648. doi:10.1177/17562864211039648.

16. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. N Engl J Med. 2017;376(3):221–234. doi:10.1056/NEJMoa1601277.

17. Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. N Engl J Med. 2012;367(12):1098–1107. doi:10.1056/NEJMoa1114287.

18. Lee DSW, Rojas OL, Gommerman JL. B cell depletion therapies in autoimmune disease: advances and mechanistic insights. Nat Rev Drug Discov. 2021;20(3):179–199. doi:10.1038/s41573-020-00092-2.

19. Cao Y, Goods BA, Raddassi K, Nepom GT, Kwok WW, Love JC, et al. Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. Sci Transl Med. 2015;7(287):287ra74. doi:10.1126/scitranslmed.aaa8038.

20. Zamecnik CR, Sowa GM, Abdelhak A, Dandekar R, Bair RD, Wade KJ, et al. An autoantibody signature predictive for multiple sclerosis. Nat Med. 2024;30(5):1300–1308. doi:10.1038/s41591-024-02938-3.

21. Comi G, Radaelli M, Soelberg Sørensen P. Evolving concepts in the treatment of relapsing multiple sclerosis. Lancet. 2017;389(10076):1347–1356. doi:10.1016/S0140-6736(16)32388-1.

22. Gandhi KS, McKay FC, Cox M, Riveros C, Armstrong N, Heard RN, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Hum Mol Genet. 2010;19(11):2134–2143. doi:10.1093/hmg/ddq090.

23. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van der Mei I, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013;9(11):e1003926. doi:10.1371/journal.pgen.1003926.

24. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019;365(6460):eaav7188. doi:10.1126/science.aav7188.