nature portfolio

Corresponding author(s): Tanuja Chitnis

Last updated by author(s): Aug 14, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom software was used for data collection and only commercial instrument acquisition software was used (e.g. BD FACSDiva v9.3 for flow cytometry and on-board Illumina software for sequencing).
Data analysis	All custom analysis code is version-controlled and publicly available at https://github.com/TNRC-MGB/RRMS-Biomarkers (MIT license; Zenodo DOI pending). Cell Ranger v7.2.0/v8.0.1; R v4.5.2 (Seurat 5.4.0, harmony 1.2.4, Azimuth 0.5.0, scater 1.38.0, scDblFinder 1.24.10, scDist 1.1.5, NEBULA, FlowSOM 2.18.0, fgsea 1.36.2, edgeR 4.8.2, lme4 1.1-38, lmerTest 3.2-1, emmeans 2.0.2, RQdeltaCT 1.3.3); Salmon v1.12.0; MAGMA v1.10 and scDRS under Python 3.12.3; FlowJo v10.10.1 was used for flow cytometry compensation (see the flow section below).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data and code availability statements are included in the main text. Raw FASTQ and processed count matrices are in GEO (accessions pending). Flow-cytometry FCS files, qPCR data, Vireo demultiplexing calls and derived analysis objects are in Zenodo (DOI pending). WGS and individual-level clinical data are controlled-access under Mass General Brigham IRB policy. Additional third party data include IMSGC MS GWAS summary statistics (GEO GSE246060, GSE29498, GSE56144; Azimuth PBMC reference).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Both sexes were included. Sex was obtained from clinical documentation in the CLIMB study. The female predominance reflects MS epidemiology. Healthy controls were age- and sex-matched to the extent possible. MS cohort (n=114): 89 F / 25 M [single-cell discovery 11 F / 4 M (n=15), validation 78 F / 21 M (n=99). Healthy controls (n=21): 11 F / 10 M]. Sex-stratified analyses were not performed.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity (self-identified, U.S. Census/OMB-style categories) were recorded from clinical documentation in the CLIMB study and are reported in Table S2. Race/ethnicity was not used for recruitment or group assignment. Whole cohort (n=135): White 122, Black or African American 5, Asian 3, Native Hawaiian or Other Pacific Islander 1, more than one race 2, unknown 2; Hispanic or Latino 9, not Hispanic or Latino 125, unknown 1.
Population characteristics	Adults from the CLIMB MS cohort (n=114: RRMS 106, CIS 4, SPMS 4; 89 F / 25 M; mean age 38.7 y, SD 7.8, range 21 - 57) and healthy controls (n=21; 11 F / 10 M; mean age 36.7 y, SD 10.7, range 23 - 58). Demographics, diagnosis, and treatment status are included in Table S2.
Recruitment	Participants were enrolled in CLIMB, an observational MS cohort at Brigham and Women's Hospital running for ~25 years. CLIMB enrollment includes annual MRI, standardized neurologic exams, and systematic biobanking. Samples were selected retrospectively from the biobank by clinical/radiographic status, as confirmed by independent review from two neurologists.
Ethics oversight	All participants gave informed written consent. The study was approved by the Mass General Brigham (formerly Partners HealthCare) Institutional Review Board and all work complied with the applicable ethical regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size. Pre-relapse specimens are rare and identifiable only retrospectively, so sizes were determined from CLIMB biobank availability: 85 MS patients were screened, 15 selected for single-cell discovery and 99 additional patients for orthogonal validation. We included 21 age- and sex-matched healthy controls based on sample availability.
Data exclusions	Single-cell QC metrics were pre-established. Specifically, cells with >10% mitochondrial reads, <300 genes, or <500 UMIs were excluded, plus per-library adaptive outlier filters (scater, 2 MADs), doublets (union of Vireo and scDblFinder calls), and residual low-quality clusters (representing 2.68% of cells). RT-qPCR: pre-specified reliability filters (Ct>35 or undetermined replicates). 5 donors were excluded for QC failure and LMP-2B was excluded for failure to amplify in any sample. For flow cytometry, one of 16 FlowSOM metaclusters was removed as non-B-cell debris.
Replication	The pre-relapse signature was validated in additional patients using additional cohorts and assays: bulk RNA-seq of sorted CD19+ B cells (n=70 patients, paired when possible), targeted viral RT-qPCR (n=30 patients, paired when possible), and multiparameter flow cytometry (n=23 patients, paired when possible), and reproduced across cohorts and orthogonal methodological approaches. All validation attempts are reported.

Randomization	No randomization was included in this observational study. Samples were assigned to groups by predefined clinical/radiographic status.
Blinding	Blinding of the study team was not possible. The clinical research coordinators needed to know which specimens were relapse-associated and which were remission to select and pair samples. The operator performing the single-cell assays was, however, unaware of how the numbered tubes corresponded to which sample(s) in a pair.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Two panels were used, a FACS sort panel (CD3-FITC, CD19-PE, CD14-PE-Cy7, 7-AAD) to isolate CD19+ B cells for bulk RNA-seq and RT-qPCR, and a 17-reagent B-cell phenotyping panel including EBV gp350. Details (vendor, clone, catalog) are in Supplementary Table 6.
Validation	All antibodies are commercial reagents validated by their manufacturers for flow cytometry on human cells.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	not applicable
Study protocol	No interventional protocol. Observational sampling and follow-up under the CLIMB protocol approved by the Mass General Brigham IRB, which is available from the corresponding author on request.
Data collection	Blood and clinical metadata were collected at CLIMB study visits at Brigham and Women's Hospital between 2001 and 2024.
Outcomes	Disease status definitions are based on neurologist-confirmed relapse with 1 or more new or enlarging gadolinium-enhancing MRI lesion. Remission samples required at least 100 days after any prior relapse and at least 180 days before any subsequent relapse, with clinical and radiographic stability. Reported outcomes include molecular differences between pre-relapse and remission peripheral immune states (single-cell and bulk transcriptomes, viral transcripts, B-cell subset frequencies).

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cryopreserved PBMCs were thawed, washed, surface-stained with a viability dye and the 17-marker B-cell panel, and immediately analyzed on the instrument. CD19+ B cells were FACS-sorted for bulk RNA-seq and RT-qPCR.
Instrument	BD LSRFortessa (5-laser acquisition with BD FACSDiva v9.3).
Software	BD FACSDiva v9.3 was used for acquisition. FlowJo v10.10.1 was used for compensation and export. FlowSOM v2.18.0 in R v4.5.2 was used for metacluster analysis.
Cell population abundance	B cells were gated as CD19+ within viable singlets, and FlowSOM metacluster frequencies were expressed as % of CD19+ B cells. gp350+ frequencies were modeled with a binomial GLMM.
Gating strategy	Sequential gates were applied to isolate lymphocytes (FSC-A vs SSC-A), then singlets (FSC-A vs FSC-H), viable cells, CD3- cells, and finally CD19+ B cells. High confidence CD19+ B cells were used as input to FlowSOM. FlowSOM analysis yielded 16 metaclusters, with one small cluster removed as non-B-cell debris. Ultimately 15 were annotated (based on CD20, CD11c, CD21, CD27, CD18, CD38, CD172a, CD5, CXCR4). gp350 positivity was thresholded per acquisition day from FMO controls at a 1% false-positive rate. We have included a gating figure in Extended Data Fig. 8a.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.