

An EBV-associated atypical B cell signature in clinically isolated syndrome is implicated in progression of multiple sclerosis

Elliott D. SoRelle^{1,2*}, Ellora Haukenfrers³, Vaibhav Jain³, Karen Abramson³, Emily Hocke³, Laura A. Cooney^{5,6}, Kristina M. Harris⁵, Scott S. Zamvil^{5,7}, Simon G. Gregory^{1,3,8^*}, and Micah A. Luftig^{1^*}

¹ Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, NC 27710, USA

² Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA

³ Duke Molecular Physiology Institute, Duke University, Durham, NC 27701, USA

⁴ Department of Immunology, Duke University School of Medicine, Durham, NC 27710, USA

⁵ Immune Tolerance Network, Bethesda, MD 20814, USA

⁶ Department of Internal Medicine, Division of Rheumatology, University of Michigan, Ann Arbor MI 48109, USA

⁷ Department of Neurology, University of California at San Francisco, San Francisco, CA 94185, USA

⁸ Department of Neurology, Duke University School of Medicine, Durham, NC 27710, USA

[^] Equal contributors

^{*} Corresponding authors

KEYWORDS: clinically isolated syndrome, multiple sclerosis, atypical B cell, T-bet, EBV, autoimmunity, chronic infection, single-cell

SUMMARY: Single-cell transcriptomics establishes an EBV-associated signature in T-bet⁺ atypical B cells in CIS and a pro-inflammatory phenotype underrepresented in patients with no disease progression.

ABSTRACT

Expansion and pathogenicity of CD19⁺/CD20⁺/CD11c⁺/T-bet⁺ atypical B cells (ABCs) are hallmarks of numerous autoimmune disorders and chronic infections. In many such cases Epstein-Barr virus (EBV) is another associated or etiologic factor, though EBV involvement in these diseases remains poorly understood. Notably, the expansion of pro-inflammatory ABCs and a putative causal role for EBV have been identified independently in multiple sclerosis (MS). A common precipitating event in MS onset is Clinically Isolated Syndrome (CIS), a neuroinflammatory demyelinating condition of which 60-80% of cases progress to relapsing-remitting MS (RRMS). Here we report single-cell gene and surface protein expression (scRNA/CITE-seq) in peripheral B cells collected longitudinally from patients with CIS during the Immune Tolerance Network STAYCIS Trial. We focus on the transcriptomic signatures of ABCs from this cohort, publicly available scRNA-seq datasets from six other autoimmune and chronic infectious diseases, and *in vitro* EBV infection. Conservation of an expanded ABC expression profile across diseases establishes ABC dysregulation as a feature of CIS. Critically, we also observed transcriptomic features that distinguished CIS and *de novo* EBV-infected ABCs from those found in healthy controls and other disease contexts. Outcome stratification of CIS samples revealed a rare yet distinctive pro-inflammatory ABC subset that was significantly underrepresented in long-term non-progressor (LTNP) versus cases with RRMS activity (~5-fold difference). Collectively, this study provides evidence for altered ABC regulation – possibly arising from niche-specific responses to EBV infection – preceding MS onset.

MAIN

Atypical B cells (ABCs) represent a unique immune compartment broadly identified by co-expression of the canonical T cell lineage transcription factor T-bet (encoded by *TBX21*), the plasmacytoid Dendritic Cell marker CD11c (*ITGAX*), and the classical B cell markers CD19 and CD20 (*MS4A1*)^{1,2}. Originally discovered in mice as an age-associated immune subset, these cells accumulate in response to toll-like receptor 7 (TLR7) and interferon gamma (IFN γ) stimulation and exhibit B cell receptor (BCR) anergy in addition to molecular indicators of exhaustion, inhibited adaptive response, and extrafollicular development^{1,3-8}. ABC pools in mice and humans contain high frequencies of autoreactive clones and are expanded in genetic females due at least partially to dosage effects of TLR7, which is encoded on the X chromosome in each species and escapes X-linked inactivation^{1,5,9-11}. Although they are present in healthy individuals and can mediate protective immune responses¹²⁻¹⁴, T-bet⁺ B cells co-expressing combinations of other markers (e.g., FCRL5, CXCR3) are common pathogenic effectors in autoimmune and infectious diseases, many of which exhibit a sex bias toward females^{15,16}. These include systemic lupus erythematosus (SLE)^{9,17-19}, primary Sjögren's Syndrome (PSS)²⁰, rheumatoid arthritis (RA)^{21,22}, and chronic infections including human immunodeficiency virus (HIV) and malaria²³⁻²⁸.

Epstein-Barr virus (EBV) is a gammaherpesvirus that establishes lifelong infection in more than 90% of adults worldwide²⁹. This striking prevalence is largely attributable to the success with which EBV induces and modulates the adaptive immune responses of host B cells to achieve persistent latency in the memory B cell pool via germinal center (GC) dynamics as well as GC-independent routes³⁰⁻³³. In addition to its contribution to the development of multiple cancers³⁴, EBV is frequently associated with each of the autoimmune and chronic infectious diseases that exhibit clonally expanded ABCs³⁵⁻⁴⁶. We recently reported that EBV can infect existing ABCs *de novo*³³ and promote their formation in models of latent infection *in vitro*⁴⁷. However, the manifestation and mechanistic contributions of EBV infection in autoimmune and chronic infectious diseases – including whether they are cell niche-dependent – are incompletely understood.

Multiple sclerosis (MS) is a chronic neuroinflammatory disease with a 3:1 female bias⁴⁸ in which both pathogenic ABCs^{49,50} and epidemiologic evidence of EBV etiology in disease onset (a 32-fold risk from EBV seropositivity after childhood) have been identified^{51,52}. Several mutually inclusive hypotheses have been proposed to explain EBV involvement in MS. These include immune responses targeted toward viral proteins that are cross-reactive with self-antigens (molecular mimicry), EBV-induced licensing of (forbidden) autoreactive B cell clones, and damage to bystander cells caused by pro-inflammatory responses of infected cells⁵³⁻⁵⁵. For example, antibody reactivity and T cell specificity profiles from patients with MS provide evidence for molecular mimicry between the viral latency protein EBNA1 and several CNS-expressed host proteins^{56,57}. CNS inflammation originating from myeloid cell activation by a distinct GM-CSF⁺ B cell subset has also been described⁵⁸. Moreover, multiple EBV antigens and EBV⁺ B cells have been detected in MS brain lesions, though not consistently across studied cases⁵⁹⁻⁶⁴. Intriguingly, a

correlation between neuroinvasive CXCR3⁺ B cells and EBV viral load has been identified recently in patients with MS.⁶⁵ In addition to shedding light on the clinical efficacy of B cell depletion therapy, these and related findings to date underscore substantial consequences of immune dysregulation mediated by EBV in B cells within the CNS. Exactly how the virus gains access to the CNS and the extent to which EBV promotes pathogenesis via the mechanisms described above are open questions. However, these findings and numerous clinical aspects of MS are consistent with the molecular markers and functions of ABCs and suggestive of a role for EBV infection within this niche.

Clinically isolated syndrome (CIS) is an initial clinical episode of neurologic symptoms caused by neuroinflammation and demyelination⁶⁶ that commonly precedes MS. Patients with CIS that also present with brain or spinal cord lesions detected by magnetic resonance imaging (MRI) have a 60-80% chance of subsequent neurologic events and diagnosis with RRMS⁶⁷⁻⁶⁹. Despite its diagnostic timeline and significant clinical relation to MS, the presence, frequency, and phenotypic characteristics of ABCs in CIS cohorts have not been studied, to our knowledge. Similarly, whether signatures of ABC response to EBV infection are present in CIS or across the spectrum of virus-associated diseases described above³³ is not known. Thus, we aimed to transcriptomically profile ABCs from patients with CIS and compare them with ABC compartments from other disease cohorts as well as data from *in vitro* EBV infection.

Peripheral blood mononuclear cells (PBMCs) were collected from a cohort of patients (n = 16; 11 female, 5 male) at two timepoints following initial CIS diagnosis: t₁ (baseline visit; at least 28 days after completion of corticosteroid treatment and within 210 days of CIS presentation) and t₂ (at least 3 months post-t₁; mean = 172 days post-t₁, range = 84-357 days). Beginning at t₁, eleven patients were treated with atorvastatin and five were given a placebo as part of the Immune Tolerance Network (ITN) STAYCIS trial⁷⁰ (NCT00094172). Treatments were given for 12 months or until a participant met the primary endpoint of MS activity, defined as 3 or more new T2 lesions or one clinical exacerbation. Eight of the 16 participants met the primary endpoint within the 12 month treatment window, and two additional participants met the primary endpoint in a follow-up period between 12 and 18 months. Of the remaining six participants, three were subsequently diagnosed with MS on the basis of new T2 lesions. Thus, 13 of 16 participants in the analyzed cohort were diagnosed with MS. The remaining three participants were characterized *post hoc* as long-term non-progressors (LTNP) based on the collective absence of new T2 lesions, Gd-enhancing lesions, and clinical exacerbations in the 18 months following t₁. Of 13 individuals diagnosed with MS, ten received atorvastatin and three received placebo; of three LTNP individuals, two received atorvastatin and one received placebo. PBMCs from each patient and timepoint were prepared as single-cell libraries (scRNA-seq + CITE-seq), which were sequenced, quality controlled, and aligned to generate count matrices. Single-cell data from patients with CIS were then analyzed in conjunction with publicly available scRNA-seq datasets from autoimmune (MS, SLE, PSS, RA)^{49,71-74} and chronic infectious diseases (HIV, malaria)²⁶ in which pathogenic ABCs have been described (**Fig. 1A**). Peripheral B cells were identified

from each dataset based on *CD19* and *MS4A1* (*CD20*) co-expression and isolated for downstream analysis. We also incorporated data from ABCs prior and subsequent to EBV infection *in vitro*³³ to assess transcriptomic similarities among disease and EBV⁺ ABCs (**Fig. 1B**). *CD19*⁺/*MS4A1*⁺ expressing clusters from our CIS and publicly available datasets were merged into an object containing over 60,000 cells from seven diseases, healthy patient controls, and pre-vs-post EBV infection *in vitro* (**Fig. 1C**). The CIS data reported herein contain 20,839 annotated B cells collected from individuals across the two timepoints (n = 32). Of these, we focused on a population of 5,926 cells found via low-resolution unbiased clustering to contain cells expressing the ABC marker *TBX21* (T-bet). While the number of patients from publicly available datasets varied (n_{HIV} = 3; n_{Malaria} = 3; n_{RA} = 4; n_{PSS} = 5; n_{MS_CSF} = 24; n_{MS_PBMC} = 5; n_{SLE_adult} = 7; n_{SLE_child} = 33), a minimum of ~2,000 B cells per disease and healthy controls (n_{adult} = 8; n_{child} = 11) were analyzed. Next, unsupervised methods were used to define 42 clusters for preliminary analysis (**Fig. 1D**). At this resolution, elevated expression of *TBX21* and *CXCR3* (another marker of pathogenic ABCs) was observed in approximately six of eleven SLE, four of five CIS, six of seven MS, two of three HIV, two of four PSS, one of six RA, two of four malaria, and *in vitro* EBV⁺ clusters (**Fig. 1E**).

We next performed hierarchical ordering of the 42 *CD19*⁺/*MS4A1*⁺ clusters based on average expression of genes of interest previously identified in EBV⁺ atMBCs (**Fig. 1F**). These included ABC markers (*TBX21*, *CXCR3*, *FCRL5*, *SOX5*); interferon-induced and antiviral responses genes (e.g., *DDX60*, *IFI6*, *IFI44*, *IFI44L*, *IFIT3*, *ISG15*, *MX1*, *TBKB1*, *TRIM22*); early markers of B cell or other immune subset activation (e.g. *CD69*, *CCR6*, *GRP183*, *TNFRSF19*, *TNFRSF13B*); cytokines and other inflammatory response mediators (e.g., *IL18*, *CD200R1*, *FGR*, *HCK*); and a number of EBV-induced genes associated with neuronal gene ontology (e.g., *ENC1*, *NRCAM*, *PPP1R17*, *RTN4R*)³³. Notably, *in vitro* EBV⁺ ABCs clustered more closely to several disease cell phenotypes than cells prior to infection. Pairwise cluster correlation based on these markers was not strongly disease-dependent, confirming similarities between EBV-infected cells and those from CIS_5, CIS_1, SLE_7, PSS_3, RA_2, MS_2, MS_3, and HIV_1 (**Fig. 1G**). While a limited number of samples were analyzed for Malaria, PSS, RA, and HIV, patient representation within each cluster clearly showed that CIS, MS, and SLE cell similarities with EBV⁺ ABCs were not artificially driven by only one or a few individuals (**Fig. 1H**). Nearly 90% of CIS patient datasets contained cells matching the CIS_1 phenotype (28 of 32 samples), whereas a much smaller cluster, CIS_5, was derived from 11 of 32 patient samples. Roughly 30-50% of 29 analyzed MS patient datasets included cells found within the MS_1 and MS_2 clusters (9 of 29 and 15 of 29 records, respectively). The SLE_7 cluster contained cells from 75% (30 of 40) analyzed datasets from adults and children with SLE. Enumeration of cells co-expressing the ABC markers *CD19*, *MS4A1*, *TBX21*, *ITGAX* (CD11c), and *FCRL5* by sample highlighted the variable frequency of ABCs in disease (**Fig. 1I**). The highest frequencies of ABCs were measured in patient samples from SLE and HIV infection. ABCs were more frequent in patients with CIS than in those with MS, although this may be partially due to the majority of analyzed MS samples being derived from CSF rather than PBMCs.

We aimed to establish deeper phenotypic profiles of ABCs identified in each disease. To do so, we extracted disease clusters with similarity to EBV⁺ ABCs (CIS_5, CIS_1, SLE_7, PSS_3, RA_2, MS_2, MS_3, and HIV_1) and refined them through additional clustering (**Figs. 2A, S1**). This yielded at least one *TBX21*⁺ ABC subset per disease and *in vitro* EBV infection. Single-cell co-expression of *TBX21*, *CD19*, *MS4A1*, *FCRL5*, *ITGAX*, and *SOX5* confirmed hallmark ABCs in each disease, including multiple clusters from patients with CIS (**Fig. 2B**, left top panel). Expression of the transcription factor *FOXP4* was also conserved in ABCs. Notably, ABCs exhibited variable *CXCR3* expression across diseases. This variation may reflect origins from distinct B cell lineages⁷⁵ or resting (*CXCR3*^{lo}) versus stimulation-induced activated or differentiated (*CXCR3*^{hi}) phenotypes^{47,76,77}. We also confirmed expression of top differentially expressed markers of ABCs previously identified in publicly available datasets (*ITGB2*, *MPP6*, *NEAT1*, *PLEK*, *S100A11*, and *TNFRSF1B*) within CIS cohort and *de novo* EBV infection samples (**Fig. 2B**, left middle panel).

We also assayed co-expression of markers from significantly enriched gene ontology (GO) sets, with particular focus on those upregulated upon *de novo* EBV infection of ABCs³³ (**Fig. 2B**). ABCs in each disease displayed higher expression of genes involved in immune cell activation, innate response, and antiviral defense (**Fig. 2B**, left bottom panel). For example, *CD84*, *FCGR2A* (CD32), the EBV EBNA-2 target *FGR*^{78,79}, *GRAP2*, *HCK*, and *TBKBP1* were consistently and specifically expressed by ABCs in every examined context. Others were enriched in the niche across most diseases (*MX1*, *MX2*; *FRMPD3*, *IFITM1*, and *IFITM3*, except in Malaria; *DDX60* and *SLAMF1* (CD150), except in MS; *ISG15*, except in PSS and RA). Genes that were generally expressed in, but not specific to, ABCs included *HSH2D*, *IFI27*, *IFI44*, *IFI44L*, and *IFI6* (except in MS, PSS, and RA); *IFIT1*, and *IFIT3* (except in MS, Malaria, PSS, and RA); and *IFIT2* (except in MS). Expression of certain genes was largely restricted to ABC clusters in a subset of diseases. These include *CCL5*, *EVL*, *GZMM*, *HOXB4*, *RAB3C*, *TREM1*, and *XAF1*. Finally, a limited number of genes with functions in innate immune response displayed comparatively disease-specific expression in ABCs (*AIF1* (IBA1), *GPR174*, and *OASL* in SLE and CIS; *LYST*, *RSAD2* in HIV and SLE; *PTAFR* in HIV and MS; and *RTP4* in HIV and CIS).

Genes with annotated developmental, cell plasticity, or neuronal lineage ontologies that we previously identified in response to EBV infection were expressed by ABCs in one or more autoimmune and chronic infectious diseases (**Fig. 2B**, right top panel). Most of these genes are infrequently expressed in mature B cells and their products support pluripotency (*ALDH1A1*, *POU5F1*, *MLLT3*, *NOTCH4*); cell motility and migration (*CXCR3*, *WASF3*); inflammatory and pro-fibrotic responses (*AREG*, *NOTCH4*, *PDGFD*); and neural cell specification and function (*ENC1*, *GAS7*, *NRCAM*, *NTNG2*, *PLXNC1*, *PPP1R17*, *PRX*). Elevated expression of several of these genes (*ENC1* and *GAS7*) has been identified previously in ABCs, particularly those from patients with SLE^{19,76}. Based on their functional importance to ABC responses, we also examined genes with annotated roles in IFN γ and TLR7 signaling (**Fig. 2B**, right bottom panel). Within

these GO sets, *FCRL3*, *PIK3AP1*, *SCIMP*, and *SLAMF6* were upregulated across ABCs in disease. *TLR7* expression, which is known to be induced upon EBV virion entry into host B cells⁸⁰, was upregulated in ABCs from HIV, SLE, and CIS. Distinctive expression in ABC subsets was observed for *CD300A* (SLE, CIS); *PRKCH* (SLE, CIS, MAL, PSS, and RA); *IL12RB2* (SLE, CIS); and *TXK* (CIS). *IRF8* was enriched within ABCs in HIV and to a lesser degree in MS and SLE. Collectively, this analysis reinforces prior hypotheses and findings that pathogenicity of ABCs in autoimmunity and chronic infection may be mediated by common mechanisms^{26,81}. Moreover, the presence and extent of shared gene expression trends between disease-state and EBV-infected ABCs suggest a potential role for EBV in mediating or inducing ABC pathogenesis.

Curiously, the expression of certain genes (e.g., *ALDH1A1*, *ALPK2*, *DLG4*, *MLLT3*, *PDGFD*, *PPP1R17*, *PRX*, *WASF3*) was virtually exclusive to CIS and *de novo* EBV-infected ABCs. Given this observation, we further investigated genes with shared expression trends in CIS and *de novo* EBV-infected versus healthy donor ABCs in other diseases (**Figs. 2C-D, S2**). Expression of established biomarkers (*CD19⁺/MS4A1⁺/TBX21⁺/ITGAX⁺/FCRL5⁺/SOX5⁺/CR2⁺*) and other conserved transcriptomic features of ABCs (*TNFRSF1B*, *ITGB2*, *FCRL3*, *CD200R1*, *ZEB2*, *FOXP4*, *ENC1*, *FGR*, *HCST*, *MPP6*, *PLEK*, *NEAT1*) provided a reference for differential expression trends specific to CIS (open stars) and EBV⁺ (red asterisk) ABCs (**Fig. 2C**, top panel). Genes that were commonly enriched in CIS EBV-infected ABCs relative to healthy controls (gray and blue dotted circles) included *SUSD3* and *CXCR3* (also up in MS, PSS, HIV, and SLE); *PCDH9* (also up in PSS and HIV); *AREG* (up in malaria and RA); *CREB5* and *NRCAM* (also up in SLE, HIV, and malaria); *MS4A14* and *GAREM2* (also up in HIV); and *AVPI1*, *CD300A*, *FCRL6*, *GPR174*, *HOXC4*, *IL18*, *NUDT17*, *PPP1R17*, *SYTL4*, *THBS3*, *TMEM63C*, and *WASF3* (up in CIS and *de novo* EBV only) (**Fig. 2C**, middle panel). Conversely, genes that were downregulated in CIS and EBV-infected ABCs included *IRF4* and *MANF* (also down in MS, malaria, PSS, and RA) as well as *BACE2*, *CDCA4*, *CXCR5*, *NUS1*, *SLAMF7*, *ST7*, and *XBP1* (also down in ABC subsets from HIV, malaria, MS, and RA) (**Fig. 2C**, bottom panel).

To further focus our comparison of CIS and EBV transcriptomic signatures, we merged ABC clusters from healthy donors (HD, n = 16), pre-EBV infection *in vitro* (resting), post-EBV *in vitro*, two CIS atypical states (CIS_1a, CIS_5) and one CIS *TBX21⁺* state (CIS_1b, representing 15 of 16 CIS patients) (**Fig. 2D**). Beyond the conservation of lineage-defining markers in healthy, EBV-infected, and CIS ABCs, reduced expression of *SLAMF7* (*CD319*, *CRACC*) occurred in response to EBV infection and in CIS (in both atypical and conventional B cells) relative to cells from healthy donors. While *SLAMF7* expression upon conventional B cell activation was demonstrated to promote proliferation and cytokine expression⁸², it negatively regulates pro-inflammatory responses in NK cells⁸³. Reduced *CXCR5* and variable *CXCR4* expression in EBV⁺ and CIS ABCs was also consistent with prior studies^{23,84-86}. Notable genes with higher expression in EBV⁺ and CIS patient ABCs relative to healthy controls included the pro-inflammatory (IFN γ -inducing) cytokine *IL18*⁸⁷; previously described EBV-induced neuronal genes *NRCAM* and *PPP1R17*³³;

and the metastasis-promoting gene *WASF3*^{88,89}. Strikingly, we also observed elevated expression of *CD6* and *TNFRSF1A* in EBV⁺ and CIS atypical subsets relative to healthy control ABCs and conventional B cells (**Fig. 2D**, orange box). This finding was particularly intriguing, given that *CD6* and *TNFRSF1A* are MS susceptibility genes⁹⁰. In particular, CD6 deficiency in mice has been shown to mitigate the severity of experimental autoimmune encephalomyelitis (EAE, a mouse model of MS) and, in T cells, impede immune infiltration of *in vitro* models of brain microvasculature and blunt IFN γ production/Th1 polarization⁹¹. We speculate that similar molecular functions of CD6 expressed by ABCs could contribute to the niche's neuroinvasive and pro-inflammatory potential. Further comparison of *in vitro* EBV⁺ and CIS patient-derived ABCs is provided in a supplementary figure (**Fig. S3**).

Although certain genes expressed upon EBV infection were exclusive to CIS among disease datasets, GO analysis identified similar biological processes across *de novo* EBV infection and disease-state ABCs (**Fig. S4**). Relative to the resting phenotype, each disease ABC state was defined by significantly enriched terms related to viral transcription, gene regulation, and antiviral defense. As expected, terms related to immune cell signaling and activation were also uniformly observed. Notably, significant enrichment of genes associated with response to interferon gamma was observed in ABCs from MS, RA, SLE, and PSS, but not in either chronic infection context. TNF family cytokine production was enriched in ABCs from MS and SLE. Unexpectedly, expression of genes associated with neutrophil activation and degranulation were significantly enriched in ABCs from MS, SLE, PSS, and HIV.

This analysis defined disease-specific transcriptomic signatures of ABCs across autoimmune and chronic infectious diseases relative to EBV-induced expression. It is striking that a presumably idiosyncratic collection of genes upregulated in ABCs following EBV infection are also expressed in disease-associated ABCs but not those from healthy donors and that these transcriptomic similarities appears to be greatest in CIS. The apparent lineage-ectopic nature of this signature may be an artifact of unknown or uncharacterized functional roles for these genes in B cells. Alternatively, it is at least conceivable that their expression, not normally detected in B cells, may be associated with induced reprogramming to promote and maintain aspects of cellular plasticity (*ALDH1A1*, *POU5F1*, *MLLT3*, *NOTCH4*, and *HOXB4*)⁹²⁻⁹⁹ (**Fig. S5**). In this regard, it is arguably worth speculating that ABCs may be primed for even further aberrant phenotypic and functional singularity in response to complex stimuli including viral infection.

Several genes upregulated in EBV⁺ and CIS ABCs relative to healthy donors imply an enhanced capacity for chemotaxis and tissue invasion. For example, *WASF3* (which encodes an actin binding protein that regulates cytoskeletal morphology) promotes cellular metastasis and invasion in breast and prostate cancer contexts^{88,89}. Likewise elevated *CXCR3* expression, which in T cells promotes migration from lymph nodes to inflammatory sites^{100,101}, has been defined as a characteristic of neuroinvasive B cells in several contexts including MS^{65,102,103}. Because *CXCR3*⁺ B cell frequency in MS is positively correlated with EBV viral load⁶⁵ and the virus can induce B cell *CXCR3* mRNA and protein expression^{33,47}, EBV infection *per se* may promote ABC trafficking to the central nervous system. Likewise, the observed induction of *CD6*

expression could conceivably facilitate the niche's neuroinvasive propensity. In this regard, expression of neuronal genes involved in cell-cell interactions and axon guidance (e.g., *NRCAM*, *PPP1R17*, *PRX*) is further suggestive of a neurotropic capacity. Upregulation of *DLG4* within EBV⁺ and CIS ABCs is also noteworthy, as this gene encodes a neuronal signaling regulator that was identified within MS patient leukocytes and exhibited significantly reduced expression during IFN β treatment¹⁰⁴. ABCs even from healthy individuals basally express Src family kinases (*HCK*, *FGR*), which are essential to myeloid cell inflammatory responses¹⁰⁵, but also inflammatory response inhibitors (e.g., *CD200R1*)¹⁰⁶. EBV infection has previously been shown to induce *FGR* expression in an EBNA2-dependent fashion^{78,79}. Upregulated expression of *IL18*, the pro-fibrotic wound healing mediators *AREG*¹⁰⁷ and *PDGFD*^{108,109}, and the inflammation-mediating inhibitory receptor *CD300A*¹¹⁰ within EBV⁺ and CIS phenotypes further indicate potential functions or participation in (neuro)inflammatory responses.

Given the expression profiles of ABCs in CIS patients and their similarities with EBV⁺ ABCs, we further investigated two CIS clusters (CIS_1a and CIS_5) from the original 42 disease-wide clusters that contained appreciable numbers of *CD19*⁺/*MS4A1*⁺/*TBX21*⁺ cells (**Fig. 3A**). CIS_1a and CIS_5 contained 1,371 cells from 30 of 32 samples (15 of 16 patients), which were further partitioned into four subclusters and stratified by timepoint and trial outcome. This population contained 1,298 cells from all patients eventually diagnosed with MS during the 18-month study period (13/13; 100 ± 69 cells per patient) and 73 cells from 2 of 3 long-term non-progressors (LTNP; 24 ± 31 cells per patient). Of the four identified subclusters, cells in CIS_1a ($n = 395$) and CIS_5 ($n = 17$) exhibited the greatest transcriptomic similarity to ABCs infected with EBV *in vitro* (**Fig. S3**). The CIS_1a subcluster accounted for 1.8% of all peripheral B cells in the full CIS cohort ($n = 20,839$). Within CIS_1a, 19 cells originated from 2 of 3 LTNP patients (6.3 ± 5.7 cells per patient) and 376 originated from 13 of 13 eventual MS diagnoses (15 ± 14.7 cells per patient). Although the sample size of CIS_5 was small ($n = 17$ cells; 8 cells from 5 patients at t_1 and 9 cells from 6 patients at t_2), it represented data from seven individuals, all of whom were ultimately diagnosed with RRMS (e.g., 7/13 MS outcomes, 0/3 LTNP outcomes) (**Fig. 3B-C**). In addition to quantifying ABC cluster composition by patient outcome, we identified a limited number of genes that were differentially expressed between MS activity and LTNP groups (**Fig. S6A-B**). Genes with significantly elevated expression in LTNP ABCs included *PEX11G*, *ATL1*, *MTRNR2L8*, *CHI3L2*, *SIRPA*, and *EZH2*. Of these, *MTRNR2L8* has previously been identified as a biomarker with potential predictive diagnostic value in MS, with reduced expression correlating to disease progression¹¹¹. Genes with significantly downregulated expression in LTNP ABCs included *CD200R1*, *TNF*, *CCDC30*, *SYP*, *POU6F1*, and *FOXO4*.

Although the difference in cumulative distributions of all ABCs (CIS_1a + CIS_5) from MS versus LTNP outcomes was not statistically significant (KS test $p = 0.15$, D statistic = 0.51), we sought to further investigate outcome representation in the CIS_5 phenotype. Because CIS_5 constituted a very limited sample size ($n = 17$ cells), we began by mapping all cells by patient ID (**Fig. S7A**) and cells from CIS_5

(**Fig. S7B**) back to the entire CIS B cell population (20,839 cells) to potentially identify a related cell neighborhood. Cells from CIS_5 remained tightly grouped as expected and comprised a subset of a larger cluster (c12, $n = 81$ cells) identified with unbiased methods. Accounting for the number of B cells derived from each sample, the six libraries from LTNP patients (3 patients \times 2 timepoints) accounted for 4,259 (20.4%) of all B cells collected across the CIS cohort. In contrast, c12 contained only 4 B cells (4.9% of c12) from LTNP patient samples. Assuming unbiased sampling from whole blood, we calculated that the c12 phenotype represented 0.094% of peripheral B cells from LTNP patients versus 0.46% from patients ultimately diagnosed with MS, representing a 4.9-fold depletion in LTNPs (**Fig. S7C**). Of the 81 cells in c12, 77 came from 11 of 13 patients diagnosed with MS and 4 were from 2 of 3 LNTN. Poisson simulation determined that the expected number of cells in c12 derived from LTNP patients based on total patient B cell sample sizes and random chance was roughly 17 ($\lambda = 16.6$). Moreover, the cumulative Poisson probability of observing ≤ 4 cells (the empirical number from LTNP samples) was 2.5×10^{-4} . This simulation was repeated to account strictly for *TBX21*⁺ cells within c12 ($n = 63$; 60 from 10/13 MS diagnoses, 3 from 1/3 LTNP) and produced similar results (# LTNP cells observed = 3; expected $[\lambda] = 13$; $\Sigma p[n \leq 3 \text{ cells}] = 1 \times 10^{-3}$) (**Fig. S7D**). Thus, with high probability, the c12 ABC phenotype (a superset of CIS_5) is significantly underrepresented within LTNP versus patients eventually diagnosed within MS.

Cells within CIS_5 exhibited an unusually broad set of surface-expressed genes at both the mRNA and protein levels (**Fig. 3D**). In the scRNA-seq assay, the most distinctive of these included *C5AR1*, *CD14*, *CD33*, *CD274* (PD-L1), *FCAR* (CD89), *HAVCR2* (CD366), *ITGAM* (CD11b), colony stimulating factor family members (*CSF1R*, *CSF2RA*, *CSF3R*), Fc gamma receptor family members (*FCGR1A* (CD64), *FCGR3A*, *FCGR3B*), killer cell lectin-like receptors (*KLRB1*, *KLRD1*, *KLRK1*), and classic T cell markers (*CD4*, *CD8A*, *CD8B*) in addition to ABC hallmarks (*CD19*⁺, *ITGAX*⁺, and *CR2*⁰). This broad CIS_5 mRNA immunophenotype includes features (*CD19*⁺, *PTPRC*(CD45)⁺, *FCER2*(CD23)^{lo}, *CR2*(CD21)^{lo}, *ITGA4*(VLA4)⁺, *ITGAL*(LFA1)⁺, *TLR4*(CD284)⁺, *CD5*^{int}) that are consistent with pro-inflammatory GM-CSF⁺ B cells described in MS⁵⁸, which are closely related but not identical to innate response activator B cells in mice¹¹². Results from the 18 marker CITE-seq assay identified protein-level co-expression of CD19, CD196 (CCR6), CD11b (ITGAM), CD11c (ITGAX), CD64 (FCGR1A), CD45 (PTPRC), and CD14 on CIS_5 cells, confirming the breadth of surface marker expression beyond a single classically defined immune lineage. Hematopoietic lineage marker and transcription factor profiling further supported this surprisingly diverse phenotype (**Figs. 3E, S6C-D**). It was noteworthy that, while EBV infection of ABCs *in vitro* led to elevated *TBX21* expression and modest expression of *NKG7* and *LYZ*, the multi-lineage immunophenotype of CIS_5 atypical cells was much more expansive than the previously described virus-induced phenotype. In this regard, cells within CIS_1a had greater transcriptomic similarity in the context of *de novo* EBV infection of ABCs.

The CIS_5 atypical state was further distinguished by elevated expression of a diverse array of genes encoding chemokine ligands, cytokines, and secreted factors with pro-inflammatory effects (**Fig. 3F-G**).

These included C-C motif chemokine ligands (*CCL3* (MIP-1A), *CCL4* (MIP-1B), *CCL5* (RANTES)); C-X-C motif chemokine ligands (*CXCL2* (MIP-2A), *CXCL3* (MIP-2B), *CXCL8* (IL-8)); granzymes (GZMB, GZMK), *IFNG* (IFN γ); *IL10*, *IL18*, *IL1B*, *NKG7*, and *TNF* (TNF α). Several of these secreted factor genes (*CCL3*, *CCL4*, *CCL5*, *IL10*, *IL18*, *NKG7*) were also significantly upregulated at the mRNA level in the context of *de novo* EBV infection of ABCs (**Fig. 3G**). To further investigate potential contributions of EBV infection to cytokine production, we performed a 27-plex Luminex assay comparing secretion profiles of resting peripheral B cells, early EBV-infected cells, and EBV-immortalized lymphoblastoid cell lines (LCLs) (**Fig. 3H**). Levels of secreted *CCL3*, *CCL4*, *CCL5*, *IFNG*, and *IL-8* were all significantly elevated in *de novo* EBV-infected proliferative B cells relative to resting cells. While not statistically significant, GM-CSF ($p = 0.11$), *IL-10* ($p = 0.051$), and *TNF α* ($p = 0.053$) were also elevated upon EBV infection relative to resting B cells (Welch's two-sided t-tests). Secretion of *CCL5* and *IFNG* remained significantly elevated in LCLs.

We reasoned that the broad immune lineage plasticity and cytokine signatures exhibited by CIS_5 cells may be related to cellular senescence. Extensive expression of pro-inflammatory cytokines is consistent with the senescence-associated secretory phenotype (SASP), through which aged or damaged cells may promote immune cell recruitment in wound healing, regenerative, or developmental contexts¹¹³. Notably, senescent cells promote paracrine induction of cellular plasticity as well as “stemness” of the arrested cell via senescence reversal¹¹⁴⁻¹¹⁷. To this end, we evaluated the expression of hallmark senescence genes within CIS_5 cells relative to other CIS B cell subclusters (**Fig. 3I**). In addition to >85% of cells in G₀/G₁ based on cell cycle marker scoring (**Fig. S1**), CIS_5 cells exhibited upregulation of *CDKN2A* (p16^{INK4A}), *CDCA7*, *CDKN2C* (p18^{INK4C}), and *CCND2*. Further, expression of these four genes and *CDKN2D* (p19^{INK4D}) were highly correlated. Collectively, these data support annotation of CIS_5 as a CXCR3⁺, pro-inflammatory ABC state, potentially associated with cellular senescence. Relative to other CIS B cell clusters, significantly enriched gene ontology terms for the top differentially expressed genes in CIS_5 cells reflect leukocyte chemotaxis, activation of multiple hematopoietic lineages, cytokine secretion, and, specifically, neuroinflammatory response (**Fig. 3J**).

It is important to emphasize that we did not detect EBV reads within B cell scRNA-seq data from the CIS patient cohort, which was also previously reported in a scRNA-seq study of patients with MS⁴⁹. Given that EBV genomes and transcripts have been detected within MS brain lesions⁵⁹⁻⁶⁴ and correlate with neuroinvasive CXCR3⁺ B cells^{50,65}, this absence may be due to one or more factors. These include transcriptome undersampling inherent to current scRNA-seq methods¹¹⁸; inefficient capture chemistry for the most abundant EBV transcripts (EBERs), which are not polyadenylated¹¹⁹; limited viral gene expression owing to restricted latency programs or failure to establish successful infection^{33,120-122}; or a combination thereof. In this case, the transcriptomic profile exhibited by ABCs in CIS and in response to EBV infection *in vitro* may reflect epigenetically encoded responses of the cell niche that EBV, a prevalent stimulus, is capable of evoking. Likewise, although our prior work demonstrated that EBV can infect ABCs based on

viral read detection³³, we cannot definitively rule out the possibility that EBV-induced responses of other infected cells (e.g., inflammatory cytokine production) may indirectly activate ABCs.

In lieu of viral read detection, we used informatic methods to explore potential viral contribution to the observed CIS and EBV⁺ ABC phenotype. We used *cis*-regulatory prediction¹²³ to identify genes associated with EBV nuclear antigen (EBNA) binding sites detected using ChIP-seq and cross-referenced these predictions against genes upregulated in CIS and EBV⁺ ABCs relative to resting cells (**Fig. S8A**). This yielded 82 EBNA-associated genes including *TBX21*, *CXCR3*, *FCRL2*, *FCRL5*, *FCGR2A*, *AREG*, *CCL5*, *CD69*, *CD84*, *CD300A*, *FGR*, *GPR174*, *HCK*, and *TBKBP1*. These 82 genes constituted modest enrichment of GO terms including defense response (GO:0006952, n = 17 genes, FDR = 0.0442), leukocyte activation (GO:0045321, n = 14 genes, FDR = 0.0442), regulation of cell migration (GO:0030334, n = 13 genes, FDR = 0.0468), and leukocyte degranulation (GO:0043299, n = 10 genes, FDR = 0.0468). Analysis of the *TBX21* locus in the lymphoblastoid cell line GM12878 revealed accessible chromatin with activating histone marks (H3K9ac, H3K4me3, H3K27ac), RNA pol II, and binding sites for EBNA-LP and EBNA-2 at the *TBX21* transcription start site (**Fig. S8B**). Collectively, informatic insights suggest that EBV-encoded transcriptional co-transactivators may mediate *TBX21*/T-bet expression and thereby affect ABC functions and responses. We also assayed expression of genes previously found to have high correlation to EBV lytic reactivation¹²⁴ (**Fig. S9**). With few exceptions, host biomarkers of the EBV lytic phase were strongly and broadly expressed in CIS_5, the ABC subset with senescent hallmarks.

In summary, this study provides single-cell transcriptomic and protein-level evidence of ABCs with pathogenic features in patients with CIS. These cells, which constitute almost 2% of peripheral B cells, share many characteristics with ABCs in other disease contexts. Although we did not identify direct evidence of EBV infection within CIS ABCs, these cells were distinguished among diseases by transcriptomic similarities to ABCs in responses to EBV infection *in vitro*. A smaller subset of *CD19⁺/CD20⁺/TBX21⁺/ITGAX⁺* cells from patients with CIS (~0.1% of B cells) that exhibited surprisingly broad immune lineage expression in addition to indicators of cellular senescence was absent or significantly underrepresented in long-term non-progressors. While definitive causality cannot be established in the absence of direct viral detection, the data presented here and prior work implicate EBV in a model wherein infection potentiates ABC migration and neuroinvasion by inducing expression of *CXCR3*^{47,50,65} and possibly other chemotactic receptors. The significant enrichment of neuronal genes including cell adhesion molecules in the EBV-associated ABC phenotype suggest an acquired neurotropic capacity that may facilitate CNS-localized viral antigen presentation and inflammatory responses that precipitate development of MS. Because ABCs differentiate into plasmablasts in response to innate and inflammatory stimuli⁹, it is further conceivable that EBV infection might promote low affinity antibody production by ABCs in the CNS. The likelihood of this pathogenic sequence of events would partially depend on the frequency of ABCs, which increases with age and to greater extent in genetic females¹. In

summary, this model is consistent with epidemiological aspects of the disease including age at time of EBV seroconversion and age- and sex-dependent accumulation of ABCs. Likewise, the clinical efficacy of anti-CD20 therapy¹²⁵ and clinical exacerbations observed upon IFN γ treatment¹²⁶ are each consistent with pathogenicity of CD20⁺ ABCs, which proliferate and differentiate in response to IFN γ . Notably, this model also aligns with empirical findings and prevailing hypotheses^{55,127,128} of EBV etiology in MS: molecular mimicry⁵⁶, pro-inflammatory bystander damage⁵⁸, and the potential for antibody production by autoreactive clones, the latter of which is a frequent fate of extrafollicular activated ABCs^{12,13}.

It must be emphasized that this model and the data reported herein are not dispositive with respect to mechanistic involvement of EBV in CIS and do not preclude alternative causality. Specifically, there is a current lack of insight with respect to: 1) whether the observed age dependence of EBV seroconversion for risk in MS is directly related to the increase in ABC frequency over time; 2) whether EBV promotes or otherwise enhances the accumulation of ABCs *in vivo*; and 3) whether periodic EBV reactivation can result in chronic *de novo* infection of ABCs (or other subsets), consistent with autoimmune disease flares. Moreover, studies are needed to investigate the consequences of EBV infection in different genetic backgrounds with known autoimmune susceptibility variants. Thus, the nature of EBV infection in promoting the formation and activation of ABCs and the relevance of this host-virus relationship in MS and other autoimmune diseases warrants substantial focus in future work. Despite these unknowns, our findings underscore the value and importance of addressing questions of viral pathogenesis related to disease at cellular resolution with high-dimensional assays.

METHODS

Clinical study design (ITN STAYCIS Trial)

Patient PBMC samples used in this work were originally collected as part of the Immune Tolerance Network (ITN) STAYCIS Trial⁷⁰. The clinical objective of the original STAYCIS Trial was to evaluate the safety and efficacy of Atorvastatin in CIS. A total of 81 patients with high risk of conversion to MS were enrolled upon initial CIS diagnosis in a two-arm clinical trial containing experimental (Atorvastatin, n = 49) and control (placebo, n = 32) treatment groups. The average age of the full cohort was 34 years, and the average time since CIS onset to screening was 65 days. All participants had no previous history of neurological disease and were seen within 210 days of the neurologic event. Patients samples were acquired at two timepoints following initial CIS diagnosis: t₁ (baseline visit; at least 28 days after completion of corticosteroid treatment and within 210 days of CIS presentation) and t₂ (at least 3 months post-t₁; mean = 172 days post-t₁, range = 84-357 days). Blood samples were collected and purified to PBMCs from each patient at each endpoint. We note that the criteria for MS diagnosis have been relaxed since the time of the STAYCIS study. Thus, by current standards some of the patients in this trial would have been diagnosed with MS at baseline.

Patient samples for scRNA / CITE-seq

PBMCs from sixteen patients at each of two timepoints (n = 32) were selected for single-cell sequencing. Eleven of the selected patients were female (age = 32 ± 8.2 years) and five were male (age = 38 ± 6.4 years). Of these patients, eight met the defined primary endpoint for MS diagnosis within the 12 month treatment period and two participants met the endpoint criteria within 18 months from t₁. Three additional participants subsequently met these criteria after the 18 month study period. The remaining three participants from the cohort analyzed herein were characterized as long-term non-progressors (LTNP) based on the absence of new T2 lesion, Gd-enhancing lesions, and no clinical exacerbations (3/16 selected cases).

Single-cell library preparation

PBMCs from the 16 patients at each endpoint were incubated with a panel of 18 antibodies with unique conjugated sequencing tags (antibody-derived tags, ADTs) for surface protein expression profiling. The panel contained antibodies against the following surface markers: CD3, CD4, CD8, CD11b (ITGAM), CD11c (ITGAX), CD14, CD16 (FCGR3A), CD19, CD25 (IL2RA), CD27, CD45 (PTPRC), CD56 (NCAM1), CD64 (FCGR1A), CD196 (CCR6), CD294 (PTGDR2), CD366 (HAVCR2), KLRG1, and TCRVa24-JA18. Following ADT labeling, cells were prepared as single-cell RNA-seq (scRNA-seq) libraries using the 10x Genomics Chromium Controller and Single-Cell Gene Expression kit with v3.1 chemistry (5' v1 single index). Cells were normalized to 1x10³ cells/μl, viability assessed, and titered to maximum of ~10,000 cells

per library. Cells are resuspended in master mix that contains reverse transcription reagents and combined with gel beads carrying the sequencing primers, barcodes, unique molecular identifier, and a poly-dT primer for RT. Full length cDNAs are cleaned and assayed to ensure lengths between 200-5000bp. Enzymatic fragmentation of the cDNA was used prior to adapter and sample index ligation; TruSeq read 2 primers are added via End Repair, A-tailing, Adaptor Ligation, and PCR. qPCR will be used to assess P5 and P7 adapter ligation, prior to size assessment of between 400-500bp. We generated 150bp paired end sequencing on an Illumina NovaSeq6000 (read 1 = 26 bp, read 2 = 91 bp, index read = 8 bp) at a target of 50,000 reads/cell for gene expression and 5,000 reads/cell for ADT libraries. Output base calls (.bcl format) from sequencing were then used to assemble single-cell RNA and ADT reads.

Read assembly, QC, and alignment

We first demultiplexed raw FASTQ reads and aligned them to the human transcriptome before using Seurat¹²⁹ to perform QC and analyze data by normalizing on a log scale after filtering for minimum gene and cell frequency cut-offs¹³⁰. We identified and exclude possible multiplets¹³⁰ and reduced noise by removing technical artifacts using regression methods¹³⁰. Principal components (PCs) were calculated using the most variably expressed genes in our dataset¹³⁰ and examined for cell dividing states¹³¹. Significant PCs were determined¹³¹ and carried forward for cell clustering and to enhance visualization¹³⁰.

Data processing and visualization

Upon completing the steps described above, we obtained ~9,000 cells on average ($8,949 \pm 1,559$) from each of the 32 longitudinal CIS patient samples. CIS patient count matrices were prepared as a Seurat^{132,133} object containing 274,277 PBMCs after QC filtering to exclude cells with fewer than 200 unique feature RNAs. Data were log normalized and scaled prior to identification of the top 2000 variable features as per standard Seurat workflows. Normalized, scaled data were dimensionally reduced through principal component analysis (*RunPCA*). Cell neighborhood identification, UMAP projection, and unbiased clustering were performed (*FindNeighbors*, *RunUMAP*, *FindClusters*). Automated cell type annotation delineated 20,839 B cells (7.6% of PBMCs), which were considered for subsequent analysis. Adaptive low-rank approximation (ALRA) was used at this stage to generate an imputed gene expression matrix with preserved biological zeros (true absence of transcript expression) and correction of technical transcript dropout¹³⁴. Imputed expression values are presented unless otherwise stated. Unbiased clustering () at various resolutions was repeated to identify and subset B cells based on *CD19*, *MS4A1*, and *TBX21* expression in downstream analyses. Cell cycle scoring and phase were also assigned based on expression of curated S- and G₂/M-phase gene sets.

Additional disease dataset provenance and curation

In addition to the newly generated CIS cohort data described above, we collected the following publicly available scRNA-seq datasets for autoimmune diseases and chronic infections: GSE133028 and GSE138266 (MS, CSF and peripheral B cells)^{49,71}; GSE157278 (pSS, PBMC)⁷³; GSE196150 (RA, synovial B cells)⁷⁴; GSE135779 (SLE and healthy controls, PBMC)⁷²; GSE149729 (Malaria, HIV, and healthy controls, peripheral B cells)²⁶; and GSE189141 (*in vitro* EBV infection, peripheral B cells)³³. Single cell count matrices from each of these studies were prepared as Seurat objects, from which B cell subsets were identified and extracted based on *CD19* and *MS4A1* (cells exhibiting $\geq 25^{\text{th}}$ percentile assay-wide expression of both genes). B cell scRNA-seq data from each of these datasets were processed and analyzed as described above for the CIS cohort data.

Analysis, statistical methods, and simulation

Differentially expressed genes between cell phenotypes, disease diagnosis, unbiased clusters, and longitudinal outcome were identified using the *FindMarkers()* function in Seurat. For most comparisons, differentially expressed genes were returned if expressed in a minimum of 60% of cells in the group of interest with an average \log_2 fold change > 0.7 . Differential expression data are included as supplementary data files, including calculated p-values with and without multiple hypothesis (Bonferroni) correction. We observed that some genes were identified as differentially expressed in select comparisons across diseases due to discrepancies in gene symbol annotation from publicly available datasets. Such genes were excluded from consideration and figure presentation. We note that this is a conservative approach that ensures the fidelity of presented data but may fail to identify additional differentially expressed genes across ABCs from different diseases. For certain analyses, expression of curated genes of interest previously identified from ABCs before and after EBV infection *in vitro* (using the same approach) were analyzed. For select genes of interest, statistical significance of differential gene expression across groups was calculated via Kolmogorov-Smirnov (KS) test. Poisson simulation was used to randomly sample cells from empirically observed cell frequencies per patient to evaluate outcome representation probabilities within specific phenotypes of interest. Cumulative Poisson probabilities were calculated for empirically observed cell frequency by phenotype in addition to expected frequencies (Poisson λ parameter).

Gene ontology enrichment

Gene ontology (GO) biological process (BP) enrichment analysis was performed from differentially expressed gene sets using the *enrichGO* function from clusterProfiler¹³⁵. Genes with \log_2 fold change values > 0.7 in a given comparison were considered for this analysis, using the whole genome as a background. Resulting terms were filtered to include those with Bonferroni-adjusted $p < 0.05$ and $q < 0.1$. Significantly enriched GO BP terms are presented as bar (barplot), dot (dotplot), and network (emmaplot) plots to represent significance and the number of genes per term set.

Cytokine assays

27-plex Luminex cytokine/chemokine secretion profiling was performed from 48h supernatants of resting peripheral blood CD19⁺ B cells, early EBV⁺ B cells that had undergone at least one population doubling after infection (\geq PD1), and EBV-immortalized LCLs (as in Price *et al.*¹³⁶). Two technical replicates from each of two biological donors were prepared and analyzed for these experiments. Resting B cells were purified from PBMCs by negative isolation (BD iMag kit, BD Biosciences). Early EBV⁺ cells (\geq PD1) were sorted by FACS at 6 days post-EBV infection of PBMCs via CD19 positivity and CFSE staining to track cell proliferation status. Sorted EBV⁺ cells were then re-cultured for two days, after which supernatants were harvested for analysis. LCLs were grown out from PBMC infections with limiting virus dilution and analyzed at five weeks post-transformation. Each of the three sample types were pure B cell populations. The cytokine panel detected the following secreted proteins: PDGFB, IL-1 β , IL-1RA, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, IL-15, IL-17, CCL11 (Eotaxin), FGF2, G-CSF, GM-CSF, IFN γ , CXCL10 (IP-10), CCL2 (MCP-1), CCL3 (MIP-1 α), CCL4 (MIP-1 β), CCL5 (RANTES), TNF α , and VEGF. Statistically significant differences in cytokine secretion between resting B cells and \geq PD1 or LCLs were determined via KS test.

ChIP-seq gene regulatory prediction

The Gene Regulatory Enrichment Analysis Tool (GREAT)¹²³ was used to predict *cis*-linked genes to binding sites for EBNA-2, EBNA-3A, and EBNA-LP identified from ChIP-seq experiments in the GM12878 LCL^{137,138}. Linked gene predictions were cross-referenced against genes that exhibited significant upregulated expression in both CIS and EBV⁺ ABCs relative to resting ABCs. Gene network analysis for EBNA-associated differentially expressed genes was performed using Cytoscape¹³⁹, and EBNA as well as other epigenetic ChIP signals and bulk chromatin accessibility¹⁴⁰ were visualized using IGV¹⁴¹.

DATA AVAILABILITY

All clinical data associated with the STAyCIS Trial is publicly available on the TrialShare website. B cell scRNA + CITE-seq data from the CIS cohort reported herein will be made available to reviewers upon request and made public upon manuscript acceptance.

ACKNOWLEDGMENTS

The authors wish to thank Jay Tourigny for assistance with cytokine secretion profiling experiments reported herein and the Immune Tolerance Network (ITN) for thoughtful feedback in the development of this manuscript. We also thank Gillian Horn and Nico Reinoso in the Luftig laboratory for discussion and feedback.

AUTHOR CONTRIBUTIONS

E.D.S., S.G.G., and M.A.L. conceived and designed the study and wrote the manuscript. E. Haukenfrers, V.J., K.A., & E. Hocke performed single-cell experiments and curated public datasets. E.D.S. and V.J. processed the data and extracted B cell populations. E.D.S. analyzed the data. L.A.C., K.M.H., and S.S.Z. reviewed and edited the manuscript. All authors have approved the final version of the manuscript.

FUNDING

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (NIH NIAID award #UM1AI109565) and an ITN award to S.G.G.. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. E.D.S. wishes to acknowledge funding from the American Cancer Society (ACS; award PF-21-084-01-DMC). M.A.L. acknowledges support from the NIDCR (R01DE025994).

- D)** UMAP of merged B cells annotated by clusters determined through unsupervised methods. Cluster names reflect original disease context.
- E)** Expression of *TBX21* (T-bet) and *CXCR3* in B cells by disease clusters depicted in (D).
- F)** Hierarchical ordering of B cell clusters by average expression of gene signature associated with EBV infection of ABCs from *in vitro* studies. Expression from ABCs before and after *de novo* EBV infection *in vitro* are highlighted with black and red boxes, respectively. Cluster dendrograms depict expression similarity across genes (rows) and similarity across disease clusters (columns).
- G)** Pearson correlation matrix of disease B cell clusters by expression of EBV-induced gene signature. Dashed boxes and triple asterisks (***) denote ABC clusters from *de novo* infection study (EBV⁺ and EBV⁻) and B cell clusters from disease datasets with the strongest correlation to these phenotypes.
- H)** Patient sample composition of disease B cell clusters. Dashed boxes highlight patient sample representation in select clusters with high correlation to EBV⁺ ABCs. Heatmap color encodes the percentage of cells within each cluster derived from patient-level sequence records. CIS samples are listed by de-identified patient records defined in this study; samples from publicly available datasets are listed by GEO submission ID.
- I)** Frequency of ABCs by disease and patient record determined by co-expression of *CD19*, *MS4A1*, *TBX21* (T-bet), *ITGAX* (CD11c), and *FCRL5*.

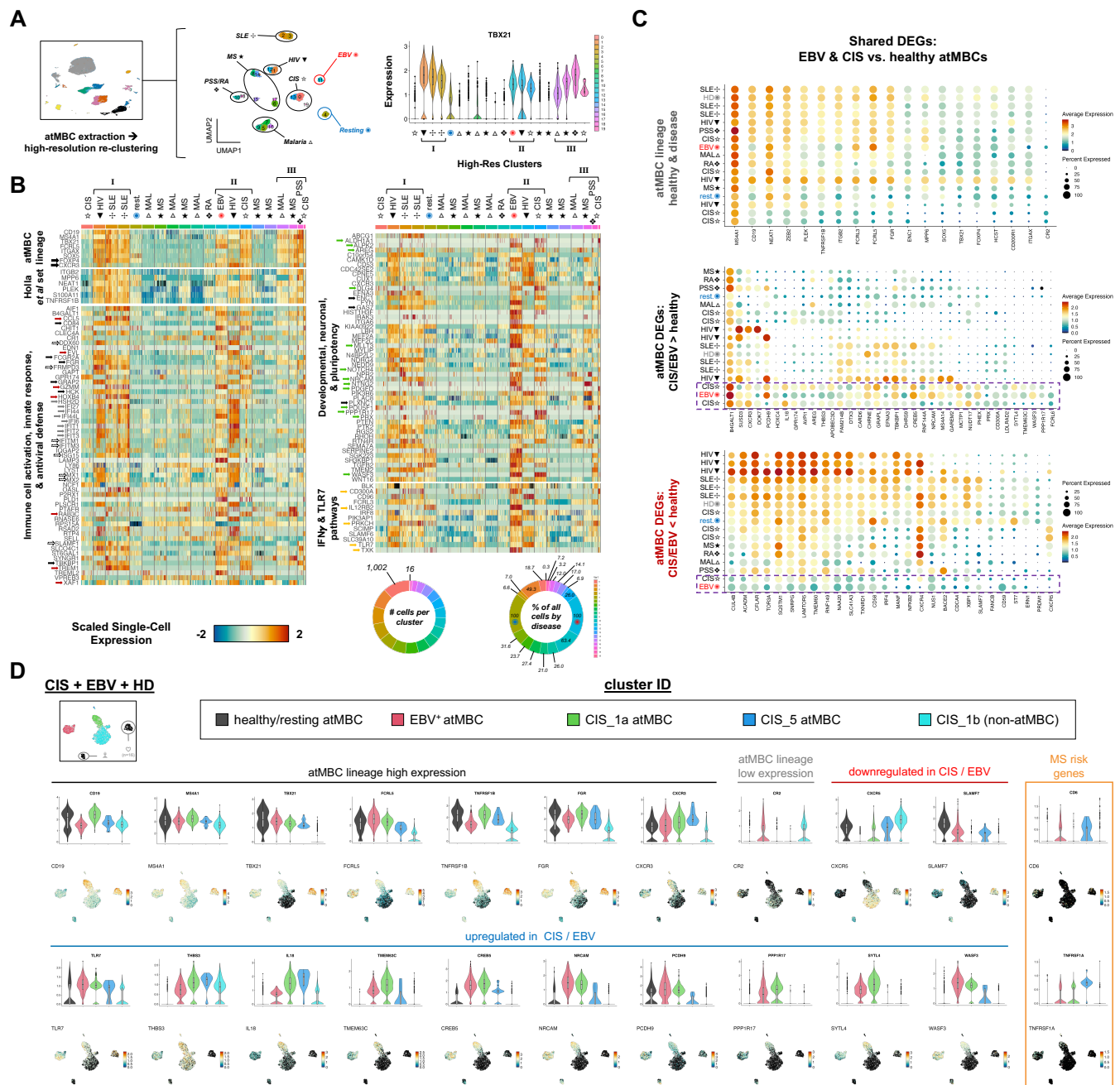


Figure 2. Comparison of ABC transcriptomes by disease identifies shared signature in CIS and *de novo* EBV infection.

A) Extraction and high-resolution re-clustering of B cells with ABC phenotypes across disease contexts (left and middle panels). *TBX21* expression after re-clustering to identify strongest ABC signatures annotated by disease (right panel).

B) Single-cell gene expression in high resolution clusters. Heatmap panels present expression of lineage hallmarks and top markers of ABCs identified in previous studies as well as EBV-induced differentially expressed genes involved in immune cell activation (black arrows), innate (gray arrows) and

inflammatory responses (red arrows), antiviral defense (dashed arrows), neuronal ontology (green arrows) and developmental pluripotency, and interferon gamma and TLR7 pathways (gold arrows).

C) Expression of genes with shared expression trends in ABCs in CIS and *de novo* EBV infection contexts relative to healthy donors across high-resolution disease clusters. Select ABC lineage genes (top panel), genes with elevated expression in CIS and *de novo* EBV contexts (middle panel), and genes with reduced expression in CIS and *de novo* EBV contexts (bottom panel) are presented by expression level and frequency of positive cells.

D) Detailed expression of select genes with common trends in CIS and *de novo* EBV-infected ABC subsets versus healthy donor ABCs and *TBX21*⁺ B cells from CIS samples. Violin and UMAP expression plots for ABC lineage genes (black & gray brackets), upregulated genes (blue bracket), downregulated genes (red bracket), and genes associated with MS risk.

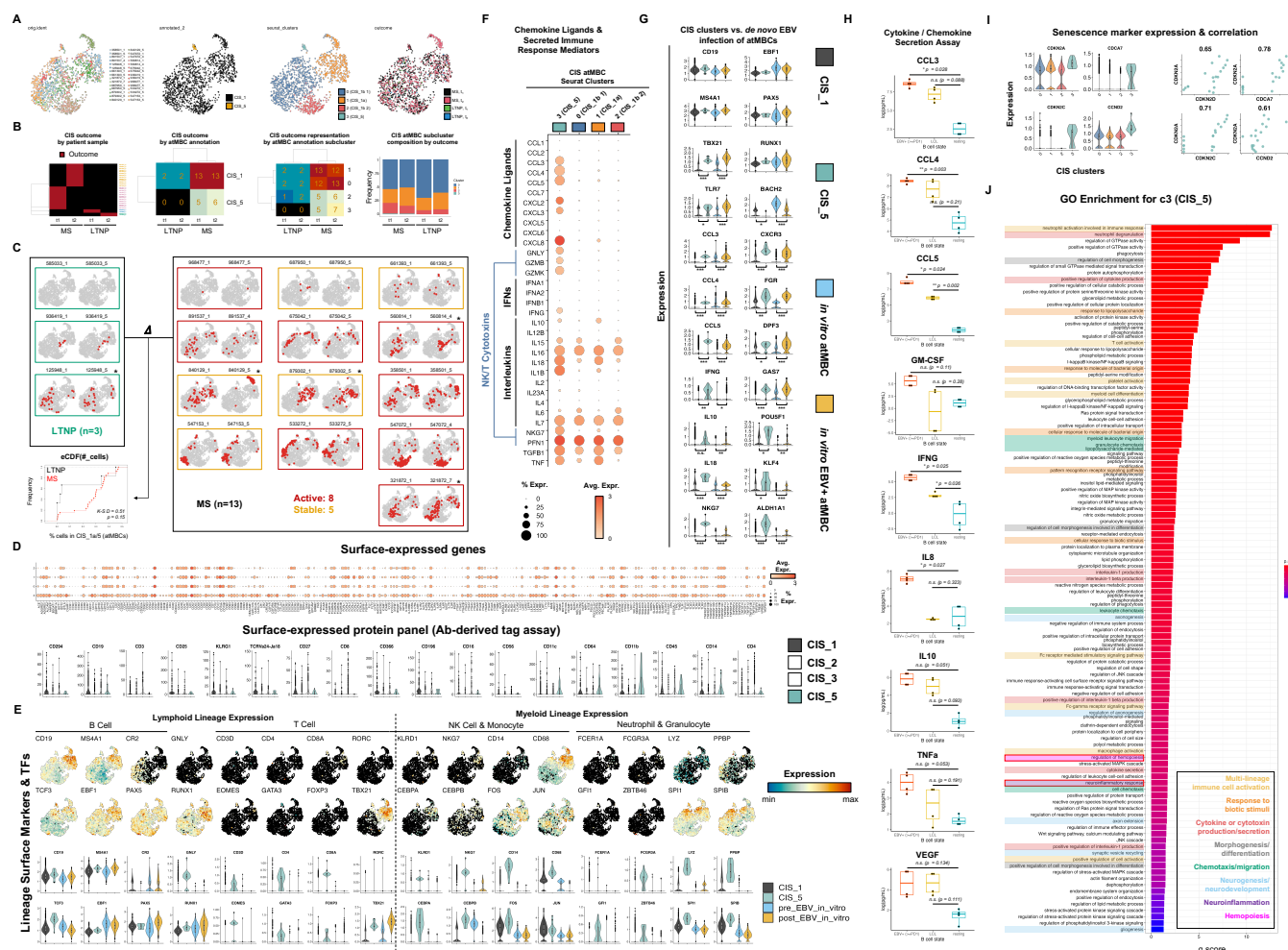


Figure 3. Detection of a rare pro-inflammatory ABC phenotype in CIS stratified by clinical outcome.

A) UMAP visualization of CIS_1 and CIS_5 B cell clusters. Cells are plotted by original time-resolved patient sample (left panel), original (middle-left panel) and high-resolution (middle-right panel) clusters determined by unsupervised methods, and time-resolved clinical outcome (t₁ and t₂ from MS and LTNP stratification).

B) Cluster composition by outcome-stratified patient samples. CIS sample IDs (left panel), number of CIS samples represented in original clusters (middle-left panel), number of CIS samples represented in high-resolution clusters (middle-right panel), and cell frequency in high-resolution clusters (right panel) for each timepoint and outcome.

C) UMAP vizualization of all cell identities in CIS_1 and CIS_5 B cells by patient sample and timepoint. LTNP samples (green), MS active primary endpoint samples (red), and samples from CNS stable patients with post-trial MS activity (gold). Empirical cumulative distributions of cells in CIS_1 and CIS_5 by outcome and statistical comparison (n.s., Kolmogorov-Smirnov test, D-statistic = 0.51, p = 0.15).

- D)** Non-canonical surface gene expression in ABCs from CIS cohort. Single-cell RNA-seq assay of surface-expressed genes in CIS_1 and CIS_5 B cells by high-resolution cluster (top panel) and CITE-seq (antibody-derived tag sequencing) of surface protein expression on CIS_1 and CIS_5 cells relative to two *TBX21*⁺ B cell clusters (CIS_2, CIS_3) from the same cohort (bottom panel).
- E)** Multi-immune lineage biomarker and transcription factor expression in CIS and *de novo* EBV-infected ABCs. In addition to *TBX21* expression, CIS_5 ABCs exhibit diverse immune lineage markers.
- F)** Broad chemokine, cytokine, and cytotoxic scRNA-seq expression profile of CIS_5 ABCs.
- G)** Significant shared scRNA-seq differential expression trends between CIS_5 and *de novo* EBV-infected ABCs indicative of immune activation, inflammatory response, and lineage plasticity.
- H)** Secreted cytokine profiling of B cells in response to *in vitro* EBV infection. Data are presented for purified resting B cells (blue), proliferative B cells in early EBV infection (\geq PD1, red), and latently infected cells (LCL, 5 weeks post-infection). Statistically significant differences in cytokine secretion were calculated using Welch's two-tailed t-test from measurements (n = 4 biological replicates per condition).
- I)** Correlated scRNA-seq expression of select genes associated with cellular senescence in CIS_5 ABCs.
- J)** GO enrichment analysis for top differentially expressed markers of CIS_5 ABCs.

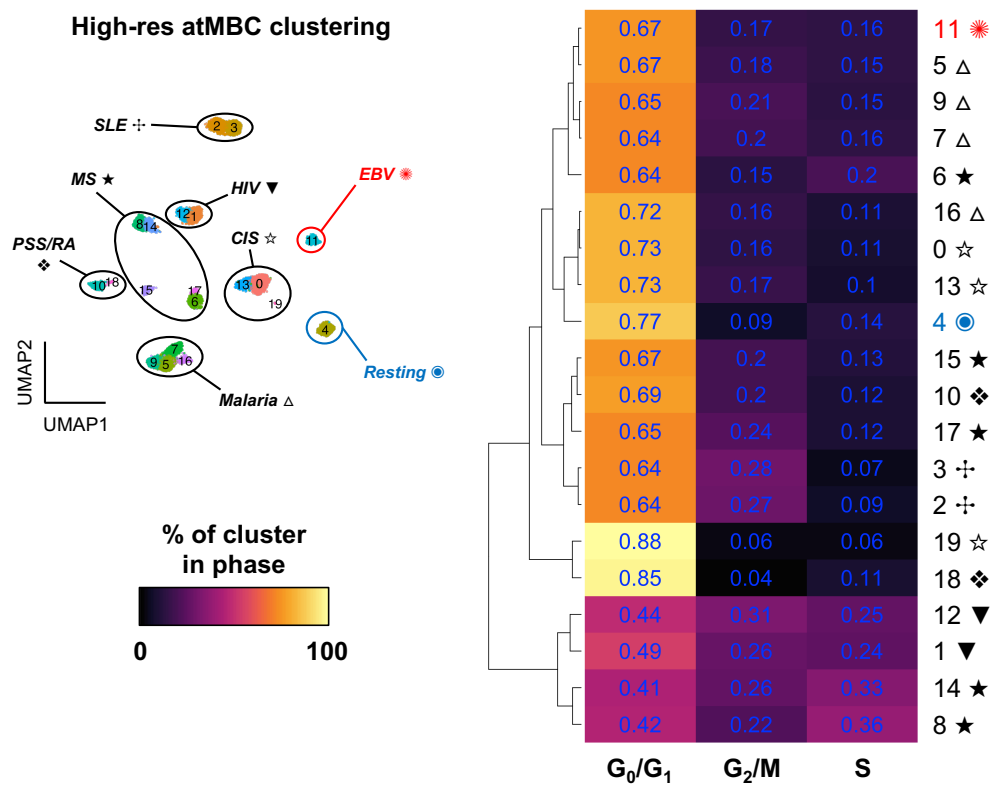


Figure S1. Cell cycle scoring of ABCs by disease.

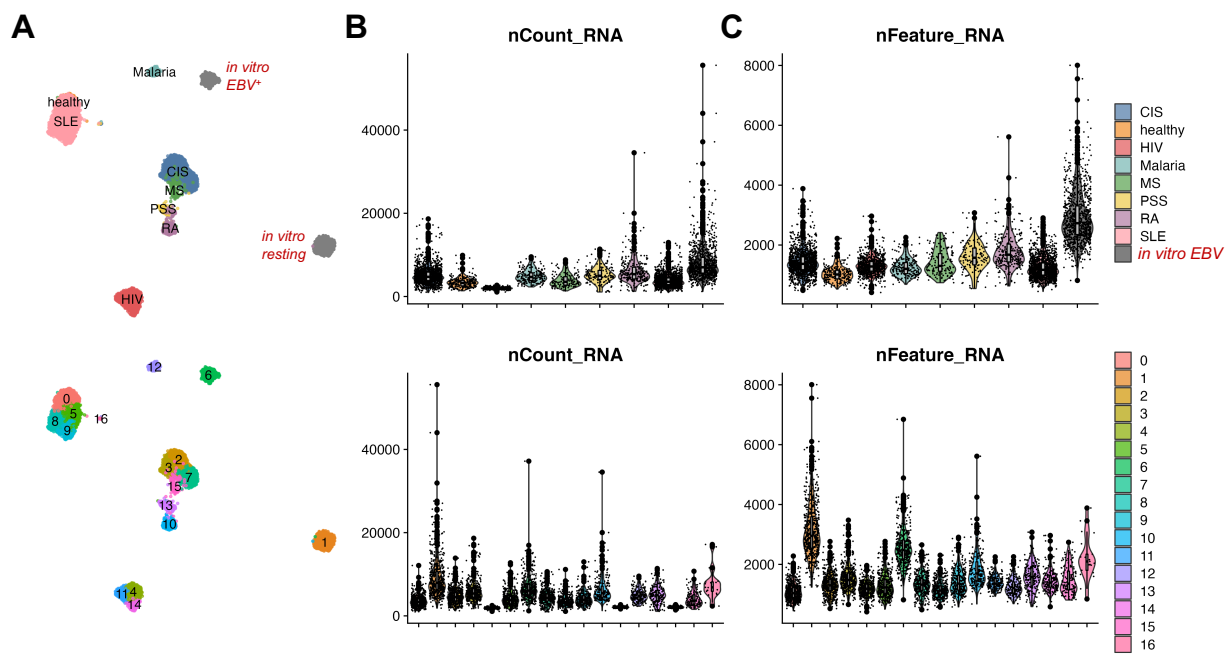


Figure S2. Global feature characterization of B cell clusters by disease dataset.

A) UMAPs of select B cell phenotypes by disease.

B) Distribution of total captured mRNA per cell by disease and cluster.

C) Distribution of unique features (genes) per cell by disease and cluster.

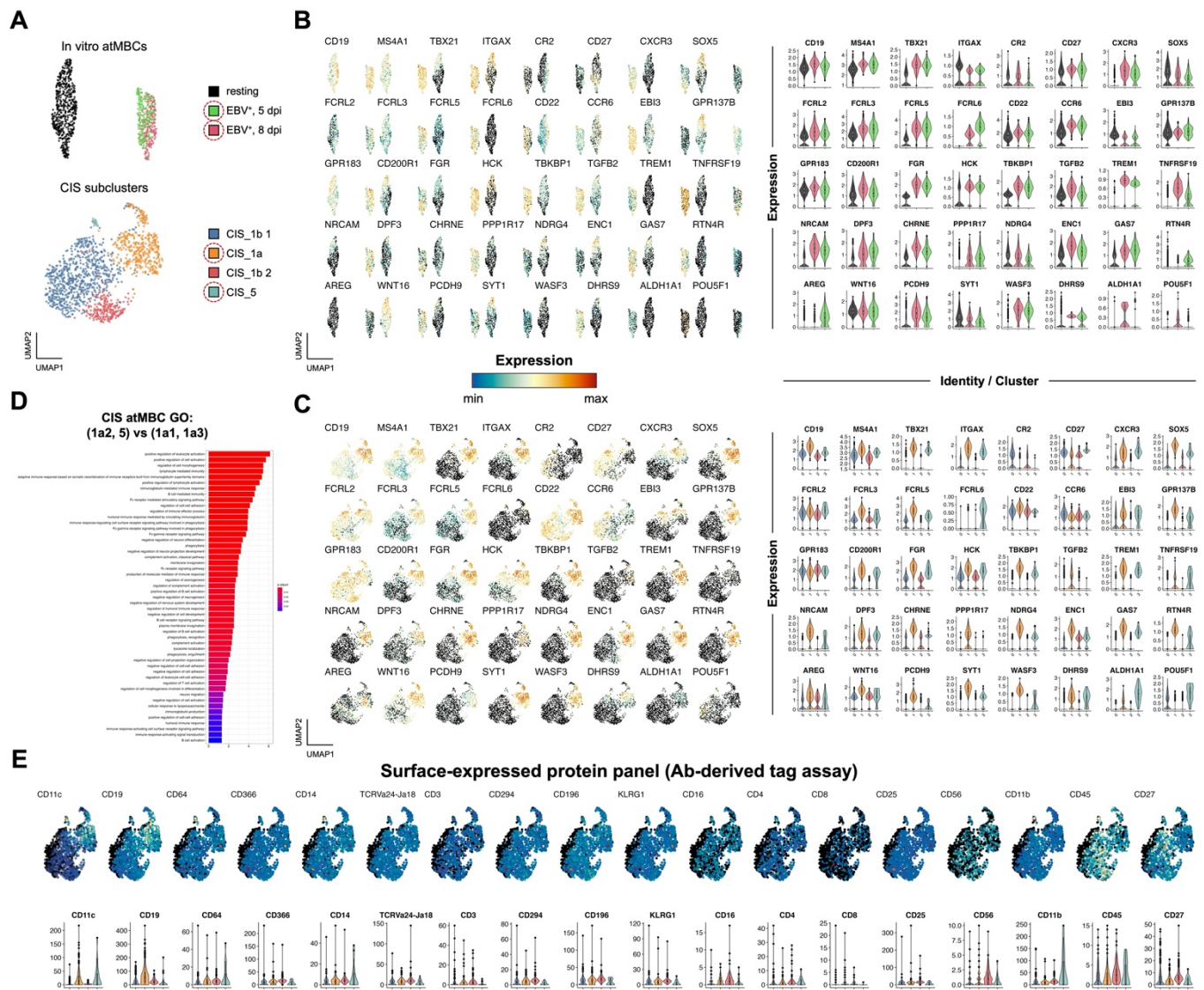


Figure S3. Detail of EBV-induced ABC signature in *de novo* EBV and CIS datasets.

A) UMAPs of ABCs from time-resolved *de novo* EBV infection (top) and CIS cohort (bottom).

B) UMAP and violin plot depiction of ABC gene expression changes in response to EBV infection.

C) UMAP and violin plot depiction of EBV-induced genes shown in B within ABC clusters from CIS dataset.

D) Gene ontology for CIS ABCs clusters (*TBX21*⁺; CIS_5, CIS_1a2) versus non-ABCs.

E) UMAPs and violin plots of surface protein expression in high-resolution CIS clusters.

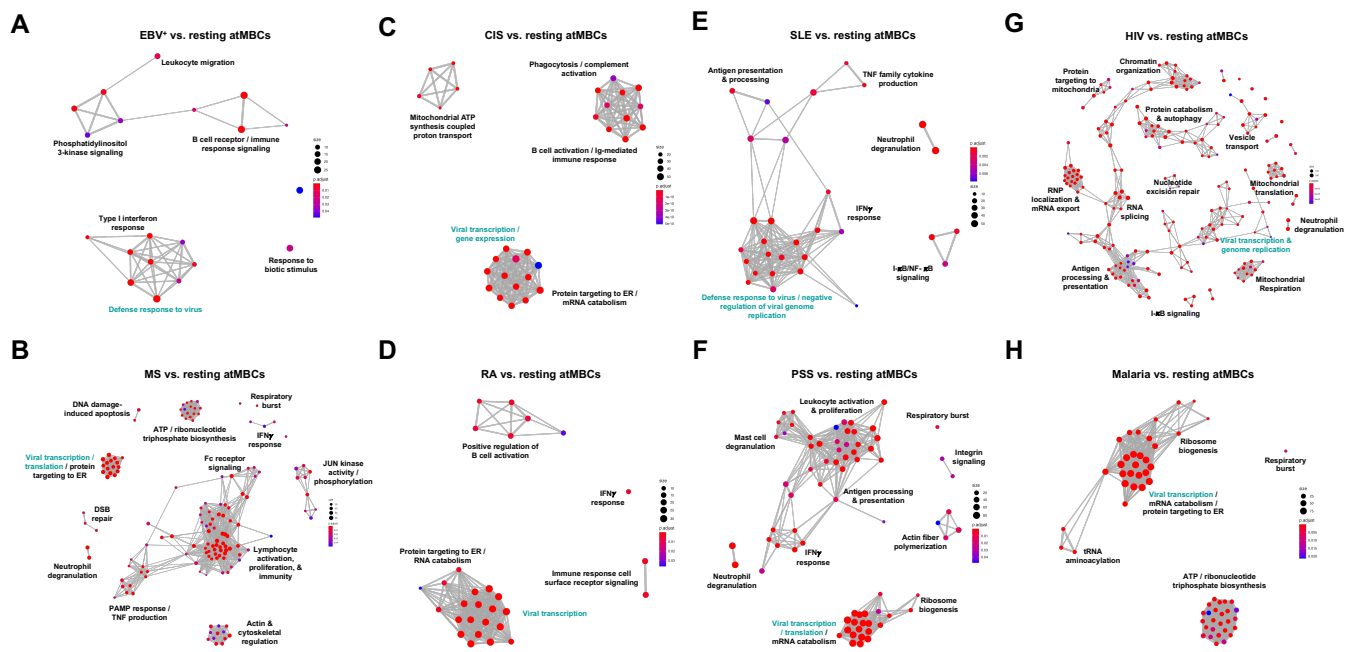


Figure S4. Gene network ontology for ABCs by disease.

A) EBV⁺ versus resting ABCs.

B) MS versus resting ABCs.

C) CIS versus resting ABCs.

D) RA versus resting ABCs.

E) SLE versus resting ABCs.

F) PSS versus resting ABCs.

G) HIV versus resting ABCs.

H) Malaria versus resting ABCs.

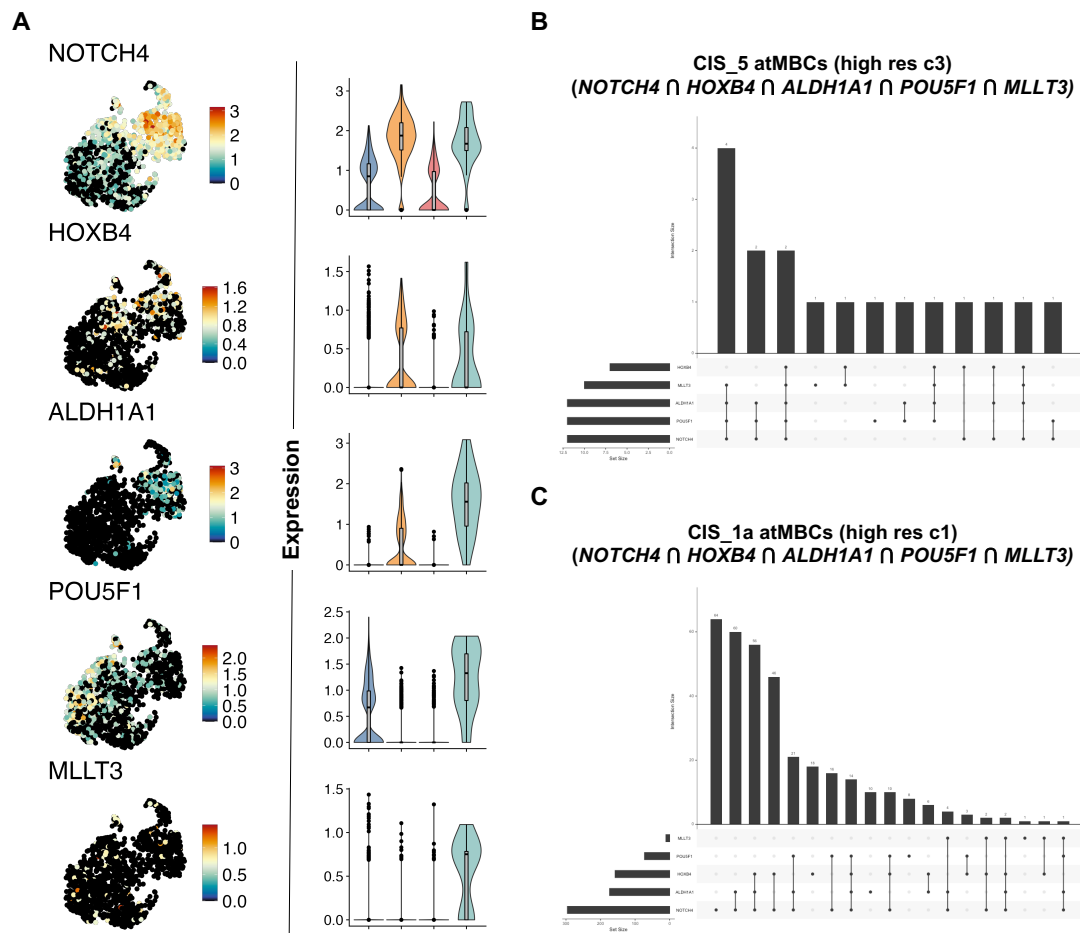


Figure S5. Co-expression of genes involved in maintenance or induction of cellular plasticity within CIS ABCs.

A) UMAP and violin plots of gene expression by high-resolution CIS cluster (CIS_5 depicted in sage green in violin plots).

B) Upset plot depicting co-expression in CIS_5 cells.

C) Upset plot depicting co-expression in CIS_1a cells.

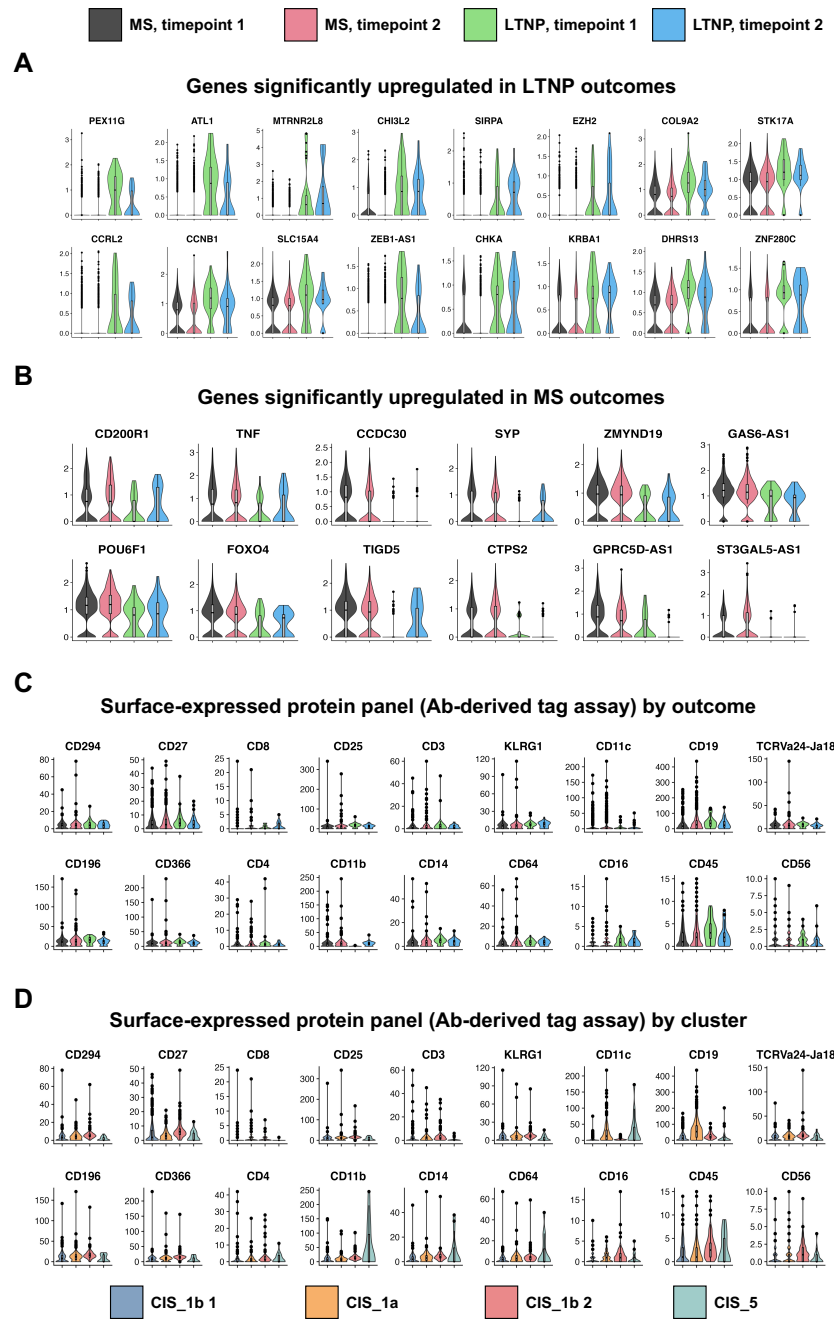


Figure S6. CIS B cell marker genes and surface protein expression by clinical outcome.

A) Top genes enriched in B cells from LTNP samples versus MS samples.

B) Top genes enriched in B cells from MS samples versus LTNP samples.

C) Surface protein expression (CITE-seq assay) stratified by MS / LTNP outcome and sample timepoint.

D) Surface protein expression stratified by high-resolution CIS ABC clusters.

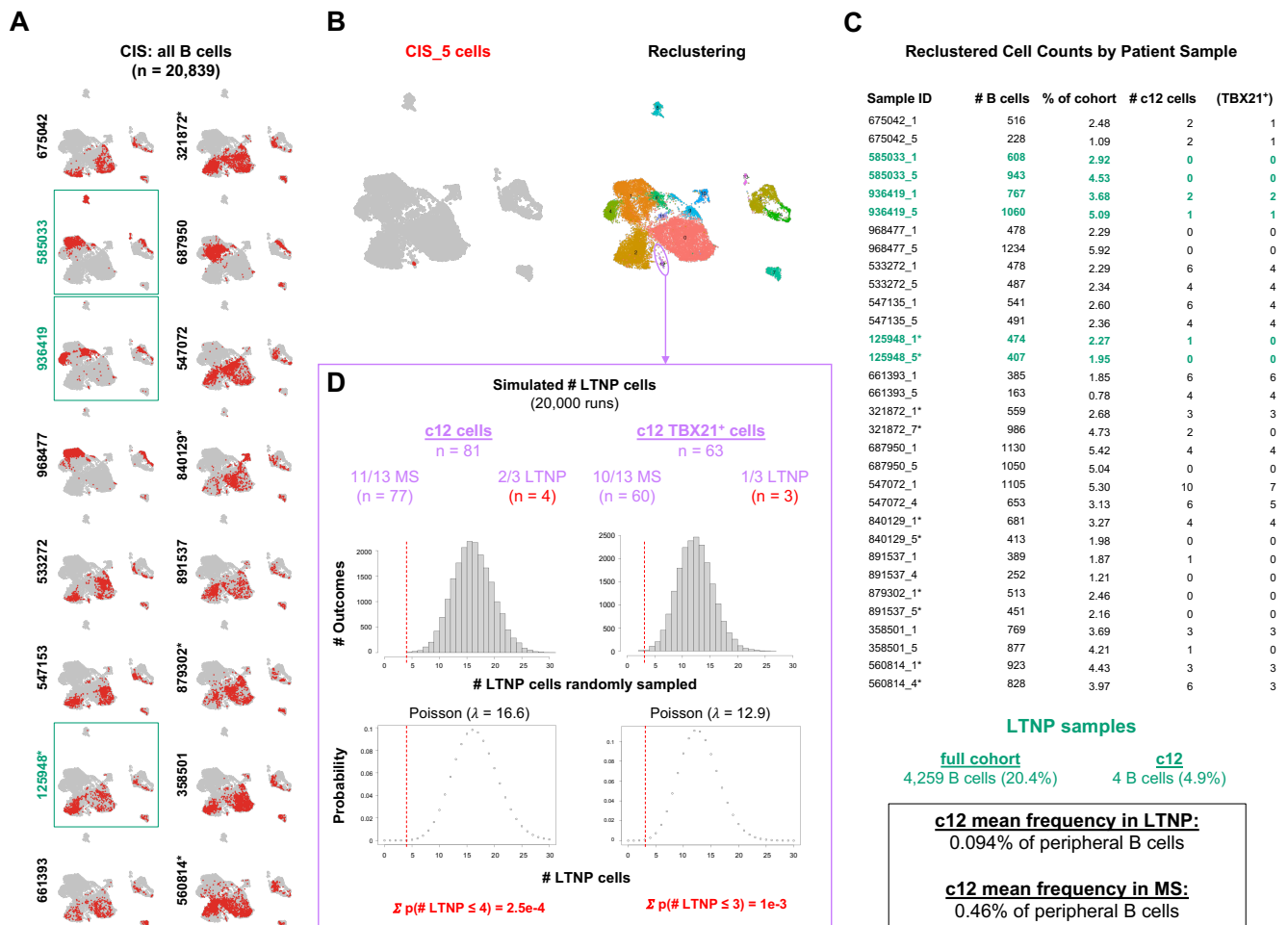


Figure S7. CIS_5 phenotype back-gating to all cohort B cells and associated phenotype frequency by clinical outcome.

- A)** Patient-resolved B cell identities across CIS cohort samples. LTNP samples are depicted in green.
- B)** Location of CIS_5 ABCs (red) within all CIS cohort B cells (left panel) and re-clustering to identify related phenotype (cluster 12, depicted in lavender; right panel).
- C)** B cell representation by patient across full cohort and cluster 12. LTNP samples are depicted in green. Empirically observed frequencies of cells in cluster 12 are presented by clinical outcome (bottom panel).
- D)** Poisson simulation comparing observed versus expected number of cluster 12 cells derived from LTNP samples based on total B cells per patient. For each outcome (MS, LTNP), the number of patients with cells present in cluster 12 as well as the number of cells per outcome are presented. Cumulative probability of observing the empirical number of cluster 12 cells (or fewer) from LTNP samples by random chance is depicted in red.

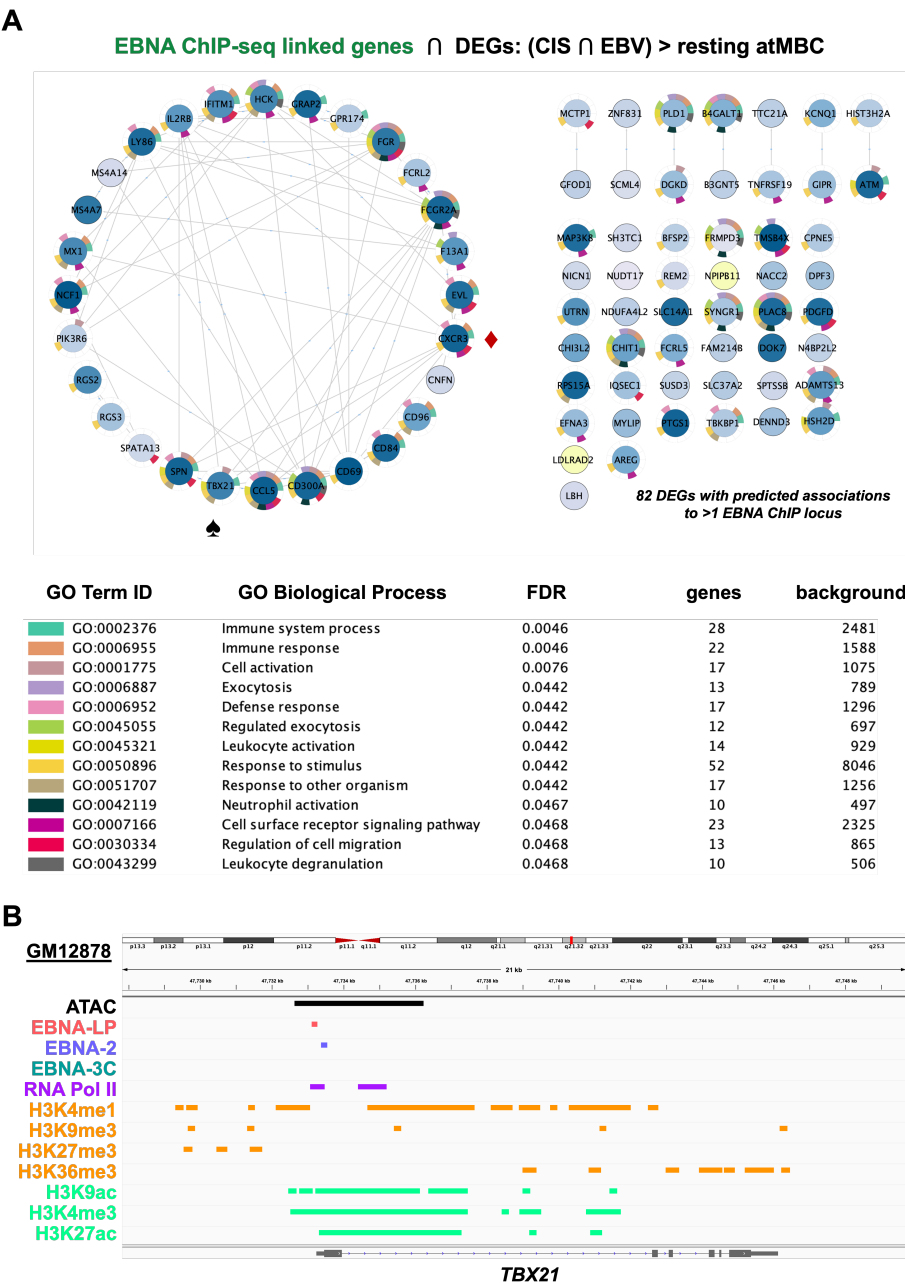


Figure S8. Genes expressed in ABCs from CIS and *de novo* EBV datasets with predicted *cis*-regulatory linkage to EBV nuclear antigen (EBNA) binding sites.

A) Gene network representation and ontology of EBNA binding site-associated genes with elevated expression in CIS and *de novo* EBV-infected versus resting ABCs. These EBNA-linked differentially expressed genes include *TBX21* (T-bet) and *CXCR3* (denoted by spade and diamond icons, respectively).

B) ATAC-seq and ChIP-seq data from the GM12878 cell line (EBV-transformed B cells) identifying open chromatin, EBNA binding sites, and epigenetic marks for active transcription at the *TBX21* transcription start site.

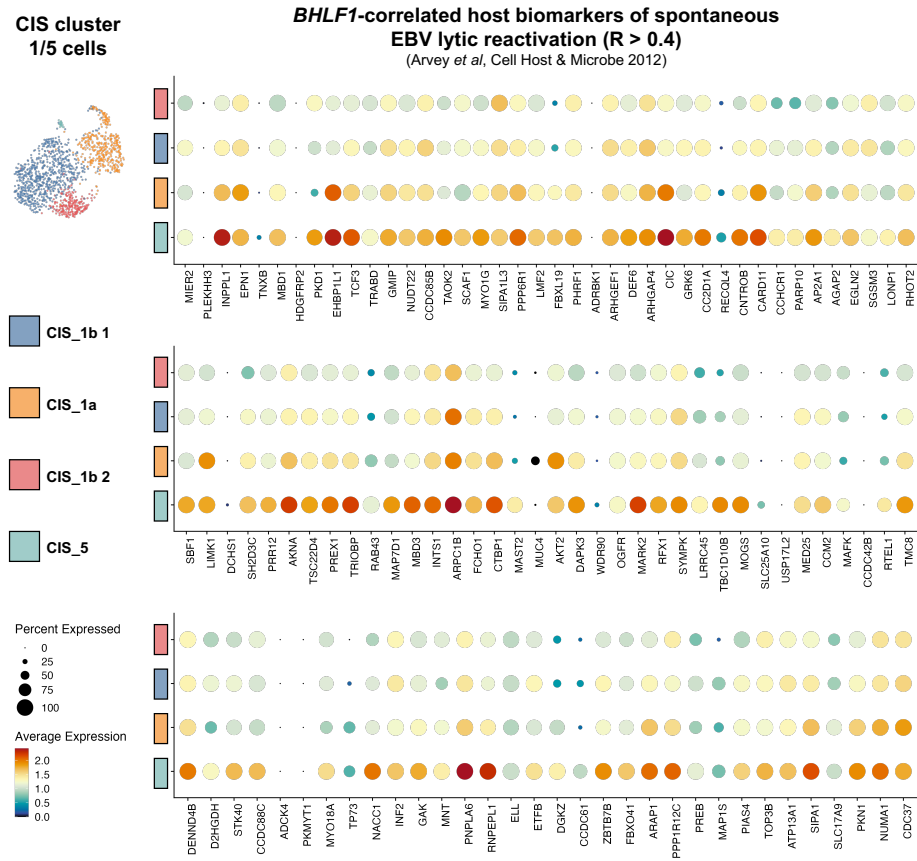


Figure S9. Expression of host cell biomarkers of EBV lytic reactivation from latency by CIS ABC cluster.

REFERENCES

- 1 Rubtsov, A. V. *et al.* Toll-like receptor 7 (TLR7)–driven accumulation of a novel CD11c⁺ B-cell population is important for the development of autoimmunity. *Blood, The Journal of the American Society of Hematology* **118**, 1305-1315 (2011).
- 2 Ehrhardt, G. R. *et al.* Discriminating gene expression profiles of memory B cell subpopulations. *The Journal of experimental medicine* **205**, 1807-1817 (2008).
- 3 Rubtsova, K., Rubtsov, A. V., van Dyk, L. F., Kappler, J. W. & Marrack, P. T-box transcription factor T-bet, a key player in a unique type of B-cell activation essential for effective viral clearance. *Proceedings of the National Academy of Sciences* **110**, E3216-E3224 (2013).
- 4 Rubtsova, K., Rubtsov, A. V., Cancro, M. P. & Marrack, P. Age-associated B cells: a T-bet–dependent effector with roles in protective and pathogenic immunity. *The Journal of Immunology* **195**, 1933-1937 (2015).
- 5 Rubtsova, K., Marrack, P. & Rubtsov, A. V. TLR7, IFN γ , and T-bet: Their roles in the development of ABCs in female-biased autoimmunity. *Cellular immunology* **294**, 80-83 (2015).
- 6 Sohn, H. W., Krueger, P. D., Davis, R. S. & Pierce, S. K. FcRL4 acts as an adaptive to innate molecular switch dampening BCR signaling and enhancing TLR signaling. *Blood, The Journal of the American Society of Hematology* **118**, 6332-6341 (2011).
- 7 Naradikian, M. S. *et al.* Cutting edge: IL-4, IL-21, and IFN- γ interact to govern T-bet and CD11c expression in TLR-activated B cells. *The Journal of Immunology* **197**, 1023-1028 (2016).
- 8 Portugal, S., Obeng-Adjei, N., Moir, S., Crompton, P. D. & Pierce, S. K. Atypical memory B cells in human chronic infectious diseases: An interim report. *Cellular immunology* **321**, 18-25 (2017).
- 9 Nicholas, M. W. *et al.* A novel subset of memory B cells is enriched in autoreactivity and correlates with adverse outcomes in SLE. *Clinical immunology* **126**, 189-201 (2008).
- 10 Isnardi, I. *et al.* Complement receptor 2/CD21– human naive B cells contain mostly autoreactive unresponsive clones. *Blood, The Journal of the American Society of Hematology* **115**, 5026-5036 (2010).
- 11 Souyris, M. *et al.* TLR7 escapes X chromosome inactivation in immune cells. *Science immunology* **3**, eaap8855 (2018).
- 12 Naradikian, M. S., Hao, Y. & Cancro, M. P. Age-associated B cells: key mediators of both protective and autoreactive humoral responses. *Immunological reviews* **269**, 118-129 (2016).
- 13 Myles, A., Sanz, I. & Cancro, M. P. T-bet⁺ B cells: a common denominator in protective and autoreactive antibody responses? *Current opinion in immunology* **57**, 40-45 (2019).
- 14 Johnson, J. L. *et al.* The transcription factor T-bet resolves memory B cell subsets with distinct tissue distributions and antibody specificities in mice and humans. *Immunity* **52**, 842-855. e846 (2020).
- 15 Karnell, J. L. *et al.* Role of CD11c⁺ T-bet⁺ B cells in human health and disease. *Cellular immunology* **321**, 40-45 (2017).
- 16 Knox, J. J., Myles, A. & Cancro, M. P. T-bet⁺ memory B cells: generation, function, and fate. *Immunological reviews* **288**, 149-160 (2019).
- 17 Rubtsova, K. *et al.* B cells expressing the transcription factor T-bet drive lupus-like autoimmunity. *The Journal of clinical investigation* **127**, 1392-1404 (2017).
- 18 Wang, S. *et al.* IL-21 drives expansion and plasma cell differentiation of autoreactive CD11c^{hi}T-bet⁺ B cells in SLE. *Nature communications* **9**, 1-14 (2018).
- 19 Scharer, C. D. *et al.* Epigenetic programming underpins B cell dysfunction in human SLE. *Nature immunology* **20**, 1071-1082 (2019).
- 20 Verstappen, G. M. *et al.* Gene expression profiling of epithelium-associated FcRL4⁺ B cells in primary Sjögren's syndrome reveals a pathogenic signature. *Journal of autoimmunity* **109**, 102439 (2020).

- 21 Tsubaki, T. *et al.* Accumulation of plasma cells expressing CXCR3 in the synovial sublining regions of early rheumatoid arthritis in association with production of Mig/CXCL9 by synovial fibroblasts. *Clinical & Experimental Immunology* **141**, 363-371 (2005).
- 22 Yeo, L. *et al.* Expression of FcRL4 defines a pro-inflammatory, RANKL-producing B cell subset in rheumatoid arthritis. *Annals of the rheumatic diseases* **74**, 928-935 (2015).
- 23 Weiss, G. E. *et al.* Atypical memory B cells are greatly expanded in individuals living in a malaria-endemic area. *The Journal of Immunology* **183**, 2176-2182 (2009).
- 24 Ho, J. *et al.* Two overrepresented B cell populations in HIV-infected individuals undergo apoptosis by different mechanisms. *Proceedings of the National Academy of Sciences* **103**, 19436-19441 (2006).
- 25 Moir, S. *et al.* Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. *The Journal of experimental medicine* **205**, 1797-1805 (2008).
- 26 Holla, P. *et al.* Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria, HIV, and autoimmunity. *Science Advances* **7**, eabg8384 (2021).
- 27 Austin, J. W. *et al.* Overexpression of T-bet in HIV infection is associated with accumulation of B cells outside germinal centers and poor affinity maturation. *Science translational medicine* **11**, eaax0904 (2019).
- 28 Hopp, C. S. *et al.* Atypical B cells up-regulate costimulatory molecules during malaria and secrete antibodies with T follicular helper cell support. *Science Immunology* **7**, eabn1250 (2022).
- 29 Rickinson, A. & Kieff, E. Fields virology. *Epstein-Barr virus* (2001).
- 30 Roughan, J. E. & Thorley-Lawson, D. A. The intersection of Epstein-Barr virus with the germinal center. *Journal of virology* **83**, 3968-3976 (2009).
- 31 Thorley-Lawson, D. A. Epstein-Barr virus: exploiting the immune system. *Nature Reviews Immunology* **1**, 75-82 (2001).
- 32 Babcock, G. J. D., L.L.; Volk, M.; Thorley-Lawson, D.A. EBV Persistence in Memory B Cells In Vivo. *Immunity* **9**, 395-404 (1998).
- 33 SoRelle, E. D. *et al.* Time-resolved transcriptomes reveal diverse B cell fate trajectories in the early response to Epstein-Barr virus infection. *Cell Reports* **40**, 111286 (2022).
- 34 Cohen, J. I., Fauci, A. S., Varmus, H. & Nabel, G. J. Epstein-Barr virus: an important vaccine target for cancer prevention. *Science translational medicine* **3**, 107fs107-107fs107 (2011).
- 35 Ascherio, A. & Munger, K. L. EBV and autoimmunity. *Epstein Barr Virus Volume 1*, 365-385 (2015).
- 36 Gross, A. J., Hochberg, D., Rand, W. M. & Thorley-Lawson, D. A. EBV and systemic lupus erythematosus: a new perspective. *The Journal of Immunology* **174**, 6599-6607 (2005).
- 37 Tosato, G., Steinberg, A. D. & Blaese, R. M. Defective EBV-specific suppressor T-cell function in rheumatoid arthritis. *New England Journal of Medicine* **305**, 1238-1243 (1981).
- 38 Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* **4**, 757-768, doi:10.1038/nrc1452 (2004).
- 39 Fox, R., Pearson, G. & Vaughan, J. Detection of Epstein-Barr virus-associated antigens and DNA in salivary gland biopsies from patients with Sjogren's syndrome. *The Journal of Immunology* **137**, 3162-3168 (1986).
- 40 Saito, I., Servenius, B., Compton, T. & Fox, R. I. Detection of Epstein-Barr virus DNA by polymerase chain reaction in blood and tissue biopsies from patients with Sjogren's syndrome. *The Journal of experimental medicine* **169**, 2191-2198 (1989).
- 41 Tosato, G. *et al.* Abnormally elevated frequency of Epstein-Barr virus-infected B cells in the blood of patients with rheumatoid arthritis. *The Journal of clinical investigation* **73**, 1789-1795 (1984).
- 42 Balandraud, N., Roudier, J. & Roudier, C. Epstein-Barr virus and rheumatoid arthritis. *Autoimmunity reviews* **3**, 362-367 (2004).
- 43 Tsokos, G., Magrath, I. & Balow, J. Epstein-Barr virus induces normal B cell responses but defective suppressor T cell responses in patients with systemic lupus erythematosus. *The Journal of Immunology* **131**, 1797-1801 (1983).

- 44 Kang, I. *et al.* Defective control of latent Epstein-Barr virus infection in systemic lupus erythematosus. *The Journal of Immunology* **172**, 1287-1294 (2004).
- 45 Poole, B. D. *et al.* Aberrant Epstein-Barr viral infection in systemic lupus erythematosus. *Autoimmunity reviews* **8**, 337-342 (2009).
- 46 Larsen, M. *et al.* Exhausted cytotoxic control of Epstein-Barr virus in human lupus. *PLoS pathogens* **7**, e1002328 (2011).
- 47 SoRelle, E. D., Reinoso-Vizcaino, N. M., Horn, G. Q. & Luftig, M. A. Epstein-Barr virus perpetuates B cell germinal center dynamics and generation of autoimmune-associated phenotypes in vitro. *Frontiers in Immunology* **13**, doi:10.3389/fimmu.2022.1001145 (2022).
- 48 Orton, S.-M. *et al.* Sex ratio of multiple sclerosis in Canada: a longitudinal study. *The Lancet Neurology* **5**, 932-936 (2006).
- 49 Ramesh, A. *et al.* A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. *Proceedings of the National Academy of Sciences* **117**, 22932-22943 (2020).
- 50 van Langelaar, J. *et al.* Induction of brain-infiltrating T-bet-expressing B cells in multiple sclerosis. *Annals of neurology* **86**, 264-278 (2019).
- 51 Ascherio, A. *et al.* Epstein-Barr virus antibodies and risk of multiple sclerosis: a prospective study. *Jama* **286**, 3083-3088 (2001).
- 52 Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* (2022).
- 53 Bar-Or, A. *et al.* Epstein-Barr virus in multiple sclerosis: theory and emerging immunotherapies. *Trends in molecular medicine* **26**, 296-310 (2020).
- 54 Pender, M. P. Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases. *Trends in immunology* **24**, 584-588 (2003).
- 55 Soldan, S. S. & Lieberman, P. M. Epstein-Barr virus and multiple sclerosis. *Nature Reviews Microbiology* **21**, 51-64 (2023).
- 56 Lanz, T. V. *et al.* Clonally Expanded B Cells in Multiple Sclerosis Bind EBV EBNA1 and GlialCAM. *Nature*, doi:10.1038/s41586-022-04432-7 (2022).
- 57 Lünemann, J. D. *et al.* Increased frequency and broadened specificity of latent EBV nuclear antigen-1-specific T cells in multiple sclerosis. *Brain* **129**, 1493-1506 (2006).
- 58 Li, R. *et al.* Proinflammatory GM-CSF-producing B cells in multiple sclerosis and B cell depletion therapy. *Science translational medicine* **7**, 310ra166-310ra166 (2015).
- 59 Serafini, B., Rosicarelli, B., Magliozzi, R., Stigliano, E. & Aloisi, F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. *Brain pathology* **14**, 164-174 (2004).
- 60 Serafini, B. *et al.* Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *The Journal of experimental medicine* **204**, 2899-2912 (2007).
- 61 Lassmann, H., Niedobitek, G., Aloisi, F., Middelorp, J. M. & Group, N. E. W. Epstein-Barr virus in the multiple sclerosis brain: a controversial issue—report on a focused workshop held in the Centre for Brain Research of the Medical University of Vienna, Austria. *Brain* **134**, 2772-2786 (2011).
- 62 Moreno, M. A. *et al.* Molecular signature of Epstein-Barr virus infection in MS brain lesions. *Neurology-Neuroimmunology Neuroinflammation* **5** (2018).
- 63 Magliozzi, R. *et al.* B-cell enrichment and Epstein-Barr virus infection in inflammatory cortical lesions in secondary progressive multiple sclerosis. *Journal of Neuropathology & Experimental Neurology* **72**, 29-41 (2013).
- 64 Tzartos, J. S. *et al.* Association of innate immune activation with latent Epstein-Barr virus in active MS lesions. *Neurology* **78**, 15-23 (2012).
- 65 van Langelaar, J. *et al.* The association of Epstein-Barr virus infection with CXCR3+ B-cell development in multiple sclerosis: impact of immunotherapies. *European journal of immunology* **51**, 626-633 (2021).

- 66 Miller, D. H., Chard, D. T. & Ciccarelli, O. Clinically isolated syndromes. *The Lancet Neurology* **11**, 157-169 (2012).
- 67 O'riordan, J. *et al.* The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain: a journal of neurology* **121**, 495-503 (1998).
- 68 Montalban, X. *et al.* MRI criteria for MS in patients with clinically isolated syndromes. *Neurology* **74**, 427-434 (2010).
- 69 Kuhle, J. *et al.* Conversion from clinically isolated syndrome to multiple sclerosis: a large multicentre study. *Multiple Sclerosis Journal* **21**, 1013-1024 (2015).
- 70 Waubant, E. *et al.* Randomized controlled trial of atorvastatin in clinically isolated syndrome: the STAYCIS study. *Neurology* **78**, 1171-1178 (2012).
- 71 Schafflick, D. *et al.* Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nature communications* **11**, 1-14 (2020).
- 72 Nehar-Belaid, D. *et al.* Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nature immunology* **21**, 1094-1106 (2020).
- 73 Hong, X. *et al.* Single-cell RNA sequencing reveals the expansion of cytotoxic CD4⁺ T lymphocytes and a landscape of immune cells in primary Sjögren's syndrome. *Frontiers in immunology*, 3688 (2021).
- 74 Meednu, N. *et al.* Dynamic spectrum of ectopic lymphoid B cell activation and hypermutation in the RA synovium characterized by NR4A nuclear receptor expression. *Cell Reports* **39**, 110766 (2022).
- 75 Sutton, H. J. *et al.* Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. *Cell reports* **34**, 108684 (2021).
- 76 Yang, R. *et al.* Human T-bet governs the generation of a distinct subset of CD11c^{high}CD21^{low} B cells. *Science Immunology* **7**, eabq3277 (2022).
- 77 Keller, B. *et al.* The expansion of human T-bethighCD21^{low} B cells is T cell dependent. *Science Immunology* **6**, eabh0891 (2021).
- 78 Knutson, J. C. The level of c-fgr RNA is increased by EBNA-2, an Epstein-Barr virus gene required for B-cell immortalization. *Journal of virology* **64**, 2530-2536 (1990).
- 79 Cheah, M. S., Ley, T. J., Tronick, S. R. & Robbins, K. C. fgr proto-oncogene mRNA induced in B lymphocytes by Epstein-Barr virus infection. *Nature* **319**, 238-240 (1986).
- 80 Martin, H. J., Lee, J. M., Walls, D. & Hayward, S. D. Manipulation of the toll-like receptor 7 signaling pathway by Epstein-Barr virus. *Journal of virology* **81**, 9748-9758 (2007).
- 81 Ambegaonkar, A. A., Holla, P., Dizon, B. L., Sohn, H. & Pierce, S. K. Atypical B cells in chronic infectious diseases and systemic autoimmunity: puzzles with many missing pieces. *Current Opinion in Immunology* **77**, 102227 (2022).
- 82 Lee, J. K., Mathew, S. O., Vaidya, S. V., Kumaresan, P. R. & Mathew, P. A. CS1 (CRACC, CD319) induces proliferation and autocrine cytokine expression on human B lymphocytes. *The Journal of Immunology* **179**, 4672-4678 (2007).
- 83 Kim, J. R., Horton, N. C., Mathew, S. O. & Mathew, P. A. CS1 (SLAMF7) inhibits production of proinflammatory cytokines by activated monocytes. *Inflammation Research* **62**, 765-772 (2013).
- 84 Nakayama, T. *et al.* Human B cells immortalized with Epstein-Barr virus upregulate CCR6 and CCR10 and downregulate CXCR4 and CXCR5. *Journal of virology* **76**, 3072-3077 (2002).
- 85 Phalke, S. & Marrack, P. Age (autoimmunity) associated B cells (ABCs) and their relatives. *Current Opinion in Immunology* **55**, 75-80 (2018).
- 86 Rubtsov, A. V. *et al.* CD11c-expressing B cells are located at the T cell/B cell border in spleen and are potent APCs. *The Journal of Immunology* **195**, 71-79 (2015).
- 87 Dinarello, C. A., Novick, D., Kim, S. & Kaplanski, G. Interleukin-18 and IL-18 binding protein. *Frontiers in immunology* **4**, 289 (2013).
- 88 Teng, Y., Ross, J. L. & Cowell, J. K. The involvement of JAK-STAT3 in cell motility, invasion, and metastasis. *Jak-stat* **3**, e28086 (2014).

- 89 Teng, Y., Ren, M., Cheney, R., Sharma, S. & Cowell, J. K. Inactivation of the WASF3 gene in prostate cancer cells leads to suppression of tumorigenicity and metastases. *British journal of cancer* **103**, 1066-1075 (2010).
- 90 De Jager, P. L. *et al.* Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nature Genetics* **41**, 776-782, doi:10.1038/ng.401 (2009).
- 91 Li, Y. *et al.* CD6 as a potential target for treating multiple sclerosis. *Proceedings of the National Academy of Sciences* **114**, 2687-2692, doi:doi:10.1073/pnas.1615253114 (2017).
- 92 Ma, I. & Allan, A. L. The role of human aldehyde dehydrogenase in normal and cancer stem cells. *Stem cell reviews and reports* **7**, 292-306 (2011).
- 93 Landen, C. N. *et al.* Targeting Aldehyde Dehydrogenase Cancer Stem Cells in Ovarian CancerTargeting ALDH1A1 Stem Cells in Ovarian Cancer. *Molecular cancer therapeutics* **9**, 3186-3199 (2010).
- 94 Moreb, J. S. Aldehyde dehydrogenase as a marker for stem cells. *Current stem cell research & therapy* **3**, 237-246 (2008).
- 95 Looijenga, L. H. *et al.* POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer research* **63**, 2244-2250 (2003).
- 96 Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).
- 97 Calvanese, V. *et al.* MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature* **576**, 281-286 (2019).
- 98 Amsellem, S. *et al.* Ex vivo expansion of human hematopoietic stem cells by direct delivery of the HOXB4 homeoprotein. *Nature medicine* **9**, 1423-1427 (2003).
- 99 Vercauteren, S. M. & Sutherland, H. J. Constitutively active Notch4 promotes early human hematopoietic progenitor cell maintenance while inhibiting differentiation and causes lymphoid abnormalities in vivo. *Blood* **104**, 2315-2322 (2004).
- 100 Qin, S. *et al.* The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. *The Journal of clinical investigation* **101**, 746-754 (1998).
- 101 Mohan, K., Cordeiro, E., Vaci, M., McMaster, C. & Issekutz, T. B. CXCR3 is required for migration to dermal inflammation by normal and in vivo activated T cells: differential requirements by CD4 and CD8 memory subsets. *European journal of immunology* **35**, 1702-1711 (2005).
- 102 Xu, J. *et al.* Chemokine receptor CXCR3 is required for lethal brain pathology but not pathogen clearance during cryptococcal meningoencephalitis. *Science advances* **6**, eaba2502 (2020).
- 103 Marques, C. P. *et al.* CXCR3-dependent plasma blast migration to the central nervous system during viral encephalomyelitis. *Journal of virology* **85**, 6136-6147 (2011).
- 104 De Felice, B. *et al.* Small non-coding RNA signature in multiple sclerosis patients after treatment with interferon- β . *BMC medical genomics* **7**, 1-9 (2014).
- 105 Kovács, M. *et al.* The Src family kinases Hck, Fgr, and Lyn are critical for the generation of the in vivo inflammatory environment without a direct role in leukocyte recruitment. *Journal of Experimental Medicine* **211**, 1993-2011 (2014).
- 106 Hernangómez, M. *et al.* CD200-CD200R1 interaction contributes to neuroprotective effects of anandamide on experimentally induced inflammation. *Glia* **60**, 1437-1450 (2012).
- 107 Zaiss, D. M., Gause, W. C., Osborne, L. C. & Artis, D. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. *Immunity* **42**, 216-226 (2015).
- 108 Reigstad, L. J., Varhaug, J. E. & Lillehaug, J. R. Structural and functional specificities of PDGF-C and PDGF-D, the novel members of the platelet-derived growth factors family. *The FEBS journal* **272**, 5723-5741 (2005).
- 109 Yang, P. *et al.* Role of PDGF-D and PDGFR- β in neuroinflammation in experimental ICH mice model. *Experimental neurology* **283**, 157-164 (2016).

- 110 Alvarez, Y., Tang, X., Coligan, J. E. & Borrego, F. The CD300a (IRp60) inhibitory receptor is rapidly up-regulated on human neutrophils in response to inflammatory stimuli and modulates CD32a (FcγRIIa) mediated signaling. *Molecular immunology* **45**, 253-258 (2008).
- 111 Iparraguirre, L. *et al.* Whole-transcriptome analysis in peripheral blood mononuclear cells from patients with lipid-specific oligoclonal IgM band characterization reveals two circular RNAs and two linear RNAs as biomarkers of highly active disease. *Biomedicines* **8**, 540 (2020).
- 112 Rauch, P. J. *et al.* Innate Response Activator B Cells Protect Against Microbial Sepsis. *Science* **335**, 597-601, doi:doi:10.1126/science.1215173 (2012).
- 113 Lopes-Paciencia, S. *et al.* The senescence-associated secretory phenotype and its regulation. *Cytokine* **117**, 15-22 (2019).
- 114 Mosteiro, L., Pantoja, C., de Martino, A. & Serrano, M. Senescence promotes in vivo reprogramming through p16 INK 4a and IL-6. *Aging cell* **17**, e12711 (2018).
- 115 Ritschka, B. *et al.* The senescence-associated secretory phenotype induces cellular plasticity and tissue regeneration. *Genes & development* **31**, 172-183 (2017).
- 116 Storer, M. *et al.* Senescence is a developmental mechanism that contributes to embryonic growth and patterning. *Cell* **155**, 1119-1130 (2013).
- 117 Muñoz-Espín, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104-1118 (2013).
- 118 Streets, A. M. & Huang, Y. How deep is enough in single-cell RNA-seq? *Nature biotechnology* **32**, 1005-1006 (2014).
- 119 Lerner, M. R., Andrews, N. C., Miller, G. & Steitz, J. A. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences* **78**, 805-809 (1981).
- 120 McFadden, K. *et al.* Metabolic stress is a barrier to Epstein-Barr virus-mediated B-cell immortalization. *Proceedings of the National Academy of Sciences* **113**, E782-E790 (2016).
- 121 Price, A. M. & Luftig, M. A. To be or not IIb: a multi-step process for Epstein-Barr virus latency establishment and consequences for B cell tumorigenesis. *PLoS Pathog* **11**, e1004656, doi:10.1371/journal.ppat.1004656 (2015).
- 122 Nikitin, P. A., Price, A. M., McFadden, K., Yan, C. M. & Luftig, M. A. Mitogen-induced B-cell proliferation activates Chk2-dependent G1/S cell cycle arrest. *PLoS One* **9**, e87299 (2014).
- 123 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501 (2010).
- 124 Arvey, A. *et al.* An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell host & microbe* **12**, 233-245 (2012).
- 125 Hauser, S. L. *et al.* B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *New England Journal of Medicine* **358**, 676-688 (2008).
- 126 Panitch, H., Haley, A., Hirsch, R. & Johnson, K. Exacerbations of multiple sclerosis in patients treated with gamma interferon. *The Lancet* **329**, 893-895 (1987).
- 127 Fujiwara, S. & Takei, M. Epstein-Barr virus and autoimmune diseases. *Clinical and Experimental Neuroimmunology* **6**, 38-48 (2015).
- 128 Pender, M. P. & Burrows, S. R. Epstein-Barr virus and multiple sclerosis: potential opportunities for immunotherapy. *Clinical & translational immunology* **3**, e27 (2014).
- 129 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).
- 130 Satija, R. *Seurat - Guided Clustering Tutorial*, <<http://satijalab.org/seurat/pbmc-tutorial.html>> (2017).
- 131 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).
- 132 Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902. e1821 (2019).

- 133 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502 (2015).
- 134 Linderman, G. C. *et al.* Zero-preserving imputation of single-cell RNA-seq data. *Nature Communications* **13**, 1-11 (2022).
- 135 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* **16**, 284-287 (2012).
- 136 Price, A. M. *et al.* Analysis of Epstein-Barr virus-regulated host gene expression changes through primary B-cell outgrowth reveals delayed kinetics of latent membrane protein 1-mediated NF- κ B activation. *Journal of virology* **86**, 11096-11106 (2012).
- 137 McClellan, M. J. *et al.* Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming. *PLoS Pathog* **9**, e1003636, doi:10.1371/journal.ppat.1003636 (2013).
- 138 Portal, D. *et al.* Epstein-Barr virus nuclear antigen leader protein localizes to promoters and enhancers with cell transcription factors and EBNA2. *Proceedings of the National Academy of Sciences* **110**, 18537-18542 (2013).
- 139 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).
- 140 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218 (2013).
- 141 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24-26 (2011).