Enhanced EBNA2-dependent activity in EBV-transformed B cells from patients with multiple sclerosis

Marissa Granitto[1,2,3*], Ellie Kim[2,3,4,5*], Carmy Forney[2,3], Cailing Yin[2,3], Arame A. Diouf[2,3], Andrew vonHandorf[2,3], Phillip Dexheimer[2,3], Sreeja Parameswaran[2,3], Xiaoting Chen[2,3], Omer Donmez[2,3], Hope Rowden[2], Casey Swoboda[2,6], Molly Shook[2,3], Katelyn Dunn[2,3], Hania Kebir[7], Maria Vélez-Colón[7], Kenneth Kaufman[2,8,9,10], Danielle Ho[2], Viktoryia Laurynenka[2,8,9], Lee E. Edsall[2,3], Veronica Brennan[2,3], Ben E. Gewurz[11,12,13], Bahram Namjou[2,8], Elizabeth Wilson[10,14], Kristen S. Fisher[15], Aram Zabeti[16], Lucinda P. Lawson[2,3], Jorge I. Alvarez[7], Leah C. Kottyan[2,3,10]**, Matthew T. Weirauch[2,3,6,10,17]**

1. Development, Stem Cells & Regenerative Medicine Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH, 45229, USA.
2. Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
3. Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
4. Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH, 45229, USA.
5. Medical Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH, 45229, USA.
6. Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
7. Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.
8. Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
9. Research Service, US Department of Veterans Affairs Medical Center, Cincinnati, Ohio, USA.

10. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, 45229, USA.
11. Department of Medicine, Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA.
12. Department of Microbiology, Harvard Program in Virology, Harvard Medical School, Boston, MA, 02115, USA.
13. Center for Integrated Solutions to Infectious Diseases, Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA.
14. Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
15. Department of Pediatrics, Division of Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA.
16. Waddell Center for Multiple Sclerosis, College of Medicine, University of Cincinnati, Cincinnati, OH 45220, USA.
17. Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.

*equal contributions

**corresponding authors: Matthew.Weirauch@cchmc.org and Leah.Kottyan@cchmc.org

Running title: Enhanced EBNA2 activity in B cells in MS

Abstract:

Multiple sclerosis (MS) is an immune-mediated demyelinating disease of the central nervous system affecting 2.8 million people worldwide. Both genetic and environmental factors contribute to MS risk, with Epstein-Barr virus (EBV) infection being an important environmental factor. To better clarify the role of EBV in MS, we examined its impact on gene expression, chromatin accessibility, and transcription factor binding in primary B cells and EBV-transformed B cells derived from patients with MS and healthy controls. RNA-seq and ATAC-seq analyses revealed extensive MS-dependent gene expression and chromatin accessibility differences in EBV-transformed, but not in primary B cells. These changes are largely accounted for by the expression levels of EBNA2, an EBV-encoded transcriptional regulator previously implicated in MS. ChIP-seq analysis revealed that EBNA2 binding with its interacting human partners RBPJ, EBF1, and PU.1 is highly enriched at MS genetic risk loci, with extensive EBNA2 allelic binding and increased enrichment at MS genetic risk loci in MS-derived cells. Our findings demonstrate that enhanced EBNA2 activity in MS alters human gene expression, chromatin accessibility, and transcription factor binding in an MS-dependent manner. Collectively, this study provides new insights into the molecular mechanisms through which EBV, particularly EBNA2, interacts with host genetic risk to contribute to MS pathogenesis.

Introduction:

Multiple sclerosis (MS) is an immune-mediated demyelinating disease of the central nervous system, with approximately 2.8 million individuals affected globally (Walton et al. 2020). It is the most common cause of non-traumatic neurological disability in young adults (Filippi et al. 2018). MS is caused by a complex combination of genetic and environmental risk factor interactions (Olsson et al. 2017). Genetic factors account for approximately 30% of disease risk, with the unaccounted risk being attributed to environmental factors, including Epstein-Barr virus (EBV) infection (International Multiple Sclerosis Genetics 2019).

Genome-wide association studies have nominated 233 independent risk loci for MS (International Multiple Sclerosis Genetics 2019). HLA-DRB-15*01 is the only MS risk locus with a moderate effect size (odds ratio ~ 3), with all other variants having small effect sizes (Hollenbach and Oksenberg 2015). Genetic risk variants are strongly enriched within non-coding regions of the genome, including enhancers and promoters (Maurano et al. 2012), suggesting that transcriptional dysregulation is likely key to MS etiology.

The most well-supported environmental factor in the etiology of MS is EBV, a common gammaherpesvirus that infects over 90% of the world's population. However, in patients with MS, the infection rate is consistently higher (over 99%) (DeLorenze et al. 2006). In 2022, a longitudinal study solidified this relationship by demonstrating that EBV infection is required for MS pathogenesis. In this retrospective study, EBV infection in early adulthood resulted in a 32-fold increased risk of MS, and serum levels of neurofilament light chain, a biomarker of neuronal damage, increased only after EBV infection (Bjornevik et al. 2022).

Most cell types driving MS pathophysiology are immune-related (e.g., B and T cells), and the success of anti-CD20 therapies in reducing both the severity of symptoms and relapse rate in MS suggests that B cells likely play important roles (Hauser et al. 2017). EBV infects both CD21+ B cells and epithelial cells, and EBV-encoded nuclear antigens and membrane proteins can evade the host's immune system by controlling the host's cell cycle and preventing apoptosis. EBV can

exploit normal B cell differentiation pathways and cause infected B cells to transition from activated B cell blasts to latently infected resting memory B cells that escape the host immune response (Zhao et al. 2011; Thorley-Lawson 2015). Latency III is the most prevalent viral phase during chronic disease, with the EBV transcriptome facilitating immune evasion (Ressing et al. 2015).

Epstein-Barr virus nuclear antigens expressed during latency, particularly EBNA1 and EBNA2, are hypothesized to play substantial roles in multiple sclerosis (MS). EBNA1 shows molecular mimicry with several human proteins, including GlialCAM, which is found in the central nervous system (Lanz et al. 2022). EBNA2 is a viral nuclear protein that serves as a transcriptional coactivator (Cable et al. 2025). It is thought to contribute to MS pathogenesis through its myriad interactions with the human genome, illustrated by highly enriched binding activity and EBNA2-dependent chromatin alterations at many autoimmune disease risk loci (Cable et al. 2025). EBNA2 binds indirectly to human DNA via human-encoded transcription factors (TFs) such as RBPJ, EBF1, and PU.1 (Waltzer et al. 1994; Johannsen et al. 1995; Murata et al. 2016). Our previous work has demonstrated that 44 MS-associated genetic loci contain a genetic variant located within an EBNA2-occupied region of the genome (Harley et al. 2018). EBNA2 extensively rewires the human genome, altering the expression of over 400 human genes and modifying chromatin accessibility at more than 2,000 genomic regions. Importantly, these regions of the human genome with EBNA2-dependent chromatin accessibility and chromatin looping are highly enriched at MS genetic risk loci (Hong et al. 2021).

In this study, we sought to understand how EBV infection leads to altered gene expression in B cells of patients with MS. To this end, we obtained primary B cells from patients with MS along with matched healthy controls (HC). After measuring gene expression and chromatin accessibility in primary B cells, we uniformly infected cells with the B95-8 strain of EBV. In the resulting transformed B cell lines, we measured gene expression, chromatin accessibility, and genomic binding of EBNA2 and its human-encoded TF interaction partners RBPJ, EBF1, and

PU.1. Our results establish that EBV infection causes extensive, robust MS-dependent changes to human gene expression and chromatin accessibility. EBNA2 expression levels account for virtually all of these differences, and EBNA2 binds along with its human TF partners at differentially expressed genes and differentially accessible chromatin. Building upon past studies, we demonstrate highly enriched EBNA2 occupancy at MS genetic risk loci in MS patient-derived cells, with increased enrichment in patients with MS relative to healthy controls. Altogether, this study establishes molecular mechanisms across the human genome through which EBV rewires B cells in an MS-dependent manner through EBNA2.

Results:

We recruited individuals representing the largest demographic of patients in the United States: MS women of European ancestry between the age of 25-74 (Filippi et al. 2018). Twelve individuals with clinician-diagnosed MS were recruited along with eleven age-matched healthy controls. The demographics and clinical phenotypes of each individual are provided in Supplemental Table S1. Peripheral blood mononuclear cells (PBMCs) were obtained from each subject. CD19+ primary B cells were isolated and EBV-transformed B cell lines were uniformly generated using standard protocols (see Methods, Supplemental Fig. S1).

Extensive MS-dependent gene expression differences in B cells after EBV transformation, which can be accounted for by *EBNA2* expression levels

We performed bulk RNA-seq experiments in primary B cells and EBV-transformed B cells from patients with MS and matched controls (see Methods, Supplemental Tables S2 and S3). We observed very few transcriptional differences in the primary B cells of individuals with MS and matched controls (19 total genes) (Fig. 1A). However, in EBV-transformed B cells, we observed 1,060 differentially expressed genes between MS and HC (Fig. 1B). Based on the established role of EBNA2 in the genetic etiology of MS (Mechelli et al. 2015; Ricigliano et al. 2015; Harley et

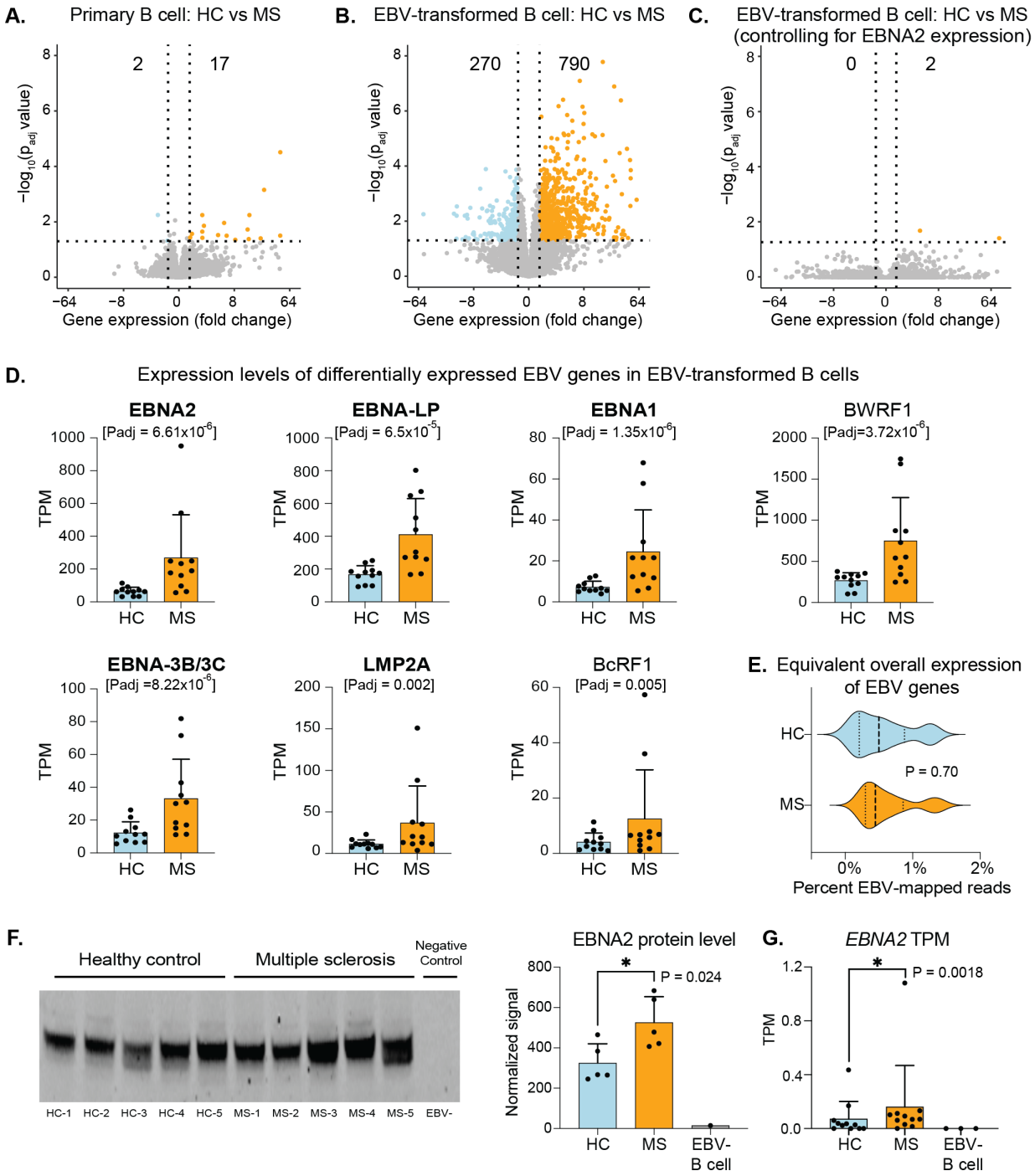

Figure 1. *EBNA2* expression accounts for extensive MS-dependent gene expression differences in B cells after EBV transformation. A-C. Comparison of human gene expression levels (RNA-seq) in healthy controls (HC) and patients with multiple sclerosis (MS). Horizontal dashed lines indicate an adjusted p-value significance threshold of 0.05. Vertical dashed lines indicate a fold-change threshold of 1.5. Results are shown for primary B cells (A), EBV- transformed B cells (B), and EBV- transformed B cells after statistically adjusting for

EBNA2 expression levels (C). D. All significant differentially expressed EBV genes in EBV-transformed B cells from HC and patients with MS. Gene expression (RNA-seq) is shown in transcripts per million (TPM). Each dot represents a cell line derived from one individual. Error bars indicate standard deviation. EBV genes in bold are expressed during latency III. E. The total percent of RNA sequencing reads mapped to the EBV genome in EBV-transformed B cells derived from HC and MS. Vertical dashed lines denote quartiles. F. Western blot for EBNA2 protein in HC and MS cell lines (left) and quantification of EBNA2 protein (right). For full blots of EBNA2 and total protein stains, see Supplemental Fig. S4. G. *EBNA2* gene expression levels (RNA-seq) in primary B cells in transcripts per million (TPM). Each dot represents primary B cells isolated from one individual. A B cell line without EBV was used as a negative control. Error bars indicate standard deviation.

al. 2018; Afrasiabi et al. 2019; Afrasiabi et al. 2020; Keane et al. 2021), we tested the hypothesis that *EBNA2* expression levels, a proxy for EBNA2 activity, might account for these differences. Indeed, *EBNA2* expression levels across cell lines accounted for virtually all MS-dependent differential expression (Fig. 1C). Despite other EBV genes also being dysregulated in MS, no other EBV gene (e.g., *EBNA1*, *EBNA-LP*) could account for as much differential expression as *EBNA2* (Supplemental Fig. S2). Among the altered human genes, EBV-induced molecule 2 (also known as *GPR183*) is a chemotactic receptor that regulates B cell migration (Gatto and Brink 2013). *GPR183* is upregulated in MS EBV-transformed B cells, and *GPR183* expression levels can be explained by *EBNA2* expression levels (Supplemental Fig. S3). *ACKR3,* another gene that is important for B cell chemotaxis, is the most significantly upregulated gene in MS EBV-transformed B cells, and this can also be accounted for by *EBNA2* expression levels (Supplemental Fig. S3). CD40 is a central immune costimulatory receptor that promotes antigen presentation and lymphocyte activation, is implicated in MS neuroinflammation, and is currently being targeted therapeutically to reduce disease activity (Vermersch et al. 2024). While there are no MS-dependent *CD40* expression differences in primary B cells, after EBV transformation *CD40* expression is increased in cells from people with MS relative to controls, with *CD40* expression in the EBV-transformed B cells again being accounted for by *EBNA2* expression levels (Supplemental Fig. S3).

Consistent with these results, *EBNA2* is the second most significant differentially expressed gene of the 84 EBV-encoded genes, with significantly higher levels in people with MS (p adj = 6.61E-06) (Fig. 1D). Among the seven differentially expressed EBV genes, five are expressed during latency III, including *EBNA2*. We confirmed that total EBV gene expression levels are not driving these differences (Fig. 1E). We next performed western blots for EBNA2 in 5 HC and 5 MS cell lines and found significantly increased levels of EBNA2 protein in MS cells (Fig. 1F, Supplemental Fig. S4). Taken together, we conclude that *EBNA2* expression levels can account for the vast majority of MS-dependent gene expression differences in EBV-transformed B cells, independent of overall EBV gene expression levels.

To rule out the possibility that differential prior exposure to EBV might play a role in the observed gene expression differences, we examined our cohort for evidence of prior EBV infection through the detection of EBV-derived RNA or DNA, and through serological evidence of immune responses to EBV. As expected, based on worldwide EBV infection rates (~95%), everyone in our cohort had evidence of prior infection by EBV (Supplemental Table S1, Supplemental Fig. S5).

We next hypothesized that MS-dependent differences in the abundance of individual B cells that were initially transformed by EBV could drive the observed differential expression (Fig. 1B). To test this hypothesis, we performed single cell RNA-sequencing with B cell receptor (BCR) enrichment in cells obtained from 5 patients with multiple sclerosis and 5 healthy controls (Supplemental Fig. S6). To benchmark our single cell dataset against prior studies, we assessed the differential expression of the BCR subunit immunoglobulin heavy chain. Consistent with previous findings in peripheral B cells (von Budingen et al. 2012), *IGHV3-74* is the most upregulated immunoglobulin gene in MS (Supplemental Fig. S6). We then asked if differences in B cell clonality might account for the observed gene expression differences. Because each B cell's BCR is uniquely rearranged, we used unique BCR sequences as a method to measure the clonality of each cell line derived from HC and MS. Based on these analyses, we did not find

evidence of differences in the clonality of cell lines between HC and MS after EBV transformation (Supplemental Fig. S6).

Next, we tested whether genetic risk might account for some of the observed gene expression differences. To this end, we performed whole genome sequencing to enable imputation of the strongest genetic risk locus (HLA-DRB1*15:01) and calculation of polygenic risk scores (PRS). The distribution of HLA-DRB1*15:01 alleles was virtually identical across HC and MS in our cohort (Supplemental Fig. S7), which is consistent with the effect size and the statistical power requirements to identify the original genetic association (Schmidt et al. 2007). We next calculated MS PRS for each member of our cohort to measure the cumulative genetic burden of risk in each person. As expected, PRS was higher on average in the individuals with MS compared to HC (Supplemental Fig. S7). Furthermore, PRS accounted for 16.4% of differential human gene expression, emphasizing the role of genetics and environment in the observed differences of MS-derived EBV-transformed B cells compared to controls (Fig. 1).

We next sought to further examine the small number of differences in primary B cells between patients with MS and controls. These differences could explain the subsequent large number of transcriptional differences after transformation with EBV (Fig. 1B). EBV can only transform B cells that express CD21. We therefore examined a published single cell RNA-seq dataset from an independent cohort of 3 HCs and 10 people with MS (Ramesh et al. 2020) and found that transitional and naïve B cells express the highest amount of *CD21* among B cell subsets, as expected (Supplemental Fig. S8). We found no differences in the proportion of B cells within the CD21+ transitional and naïve B cell compartments between HC and MS (Supplemental Fig. S8). We identified ten genes that are differentially expressed in MS relative to controls within the CD21+ transitional/naïve B cell compartment, including *BACH2, AFF3,* and *BANK1*, which have all previously been implicated in MS (Supplemental Fig. S8; Supplemental Table S4) (International Multiple Sclerosis Genetics et al. 2011; Itoh-Nakadai et al. 2014; Bae and Lee 2017; Tsukumo et al. 2022). *FANCC* is upregulated in MS CD19+ primary B cells (Supplemental Table

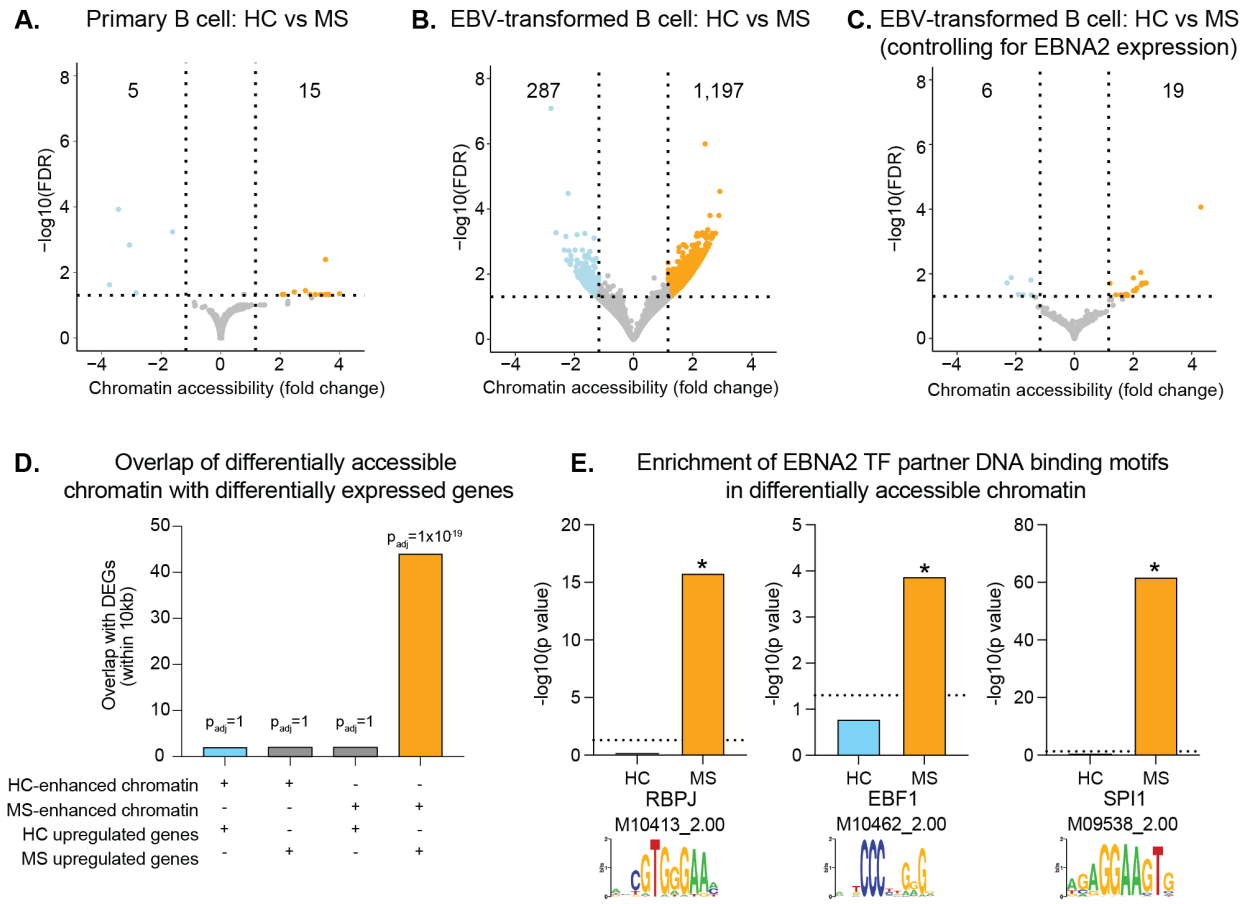

Figure 2. *EBNA2* expression accounts for MS-dependent chromatin accessibility differences in EBV-transformed B cells. A-C. Comparison of chromatin accessibility (ATAC-seq) in healthy controls (HC) and patients with multiple sclerosis (MS). Horizontal dashed lines indicate an adjusted p-value significance threshold of 0.05. Vertical dashed lines indicate a fold-change threshold of 1.5. Results are shown for primary B cells (A), EBV- transformed B cells (B), and EBV- transformed B cells after statistically adjusting for *EBNA2* expression levels (C). D. Overlap of all possible pairs of differentially accessible chromatin (ATAC-seq) and differentially expressed genes with 10 kb padding on either side of the TSS (RNA-seq). Each set of chromatin and differential genes were either up in HC or MS. The statistical enrichment of the overlap (measured by RELI, see Methods) is provided as an adjusted p-value on top of each bar. E. DNA binding motif enrichment of exemplary motifs for EBNA2 TF partners RBPJ, EBF1, and SPI1 (PU.1) in accessible chromatin. Enrichment is calculated as % foreground motifs divided by % background motifs (i.e., the percent of randomly selected genome sequences, matching GC content). Asterisks indicate significant motif enrichment ($P < 0.05$), as calculated by HOMER.

S3). A CRISPR screen demonstrated that *FANCC* expression is essential for EBNA2-dependent transformation of primary B cells to EBV-transformed cell lines (Ma et al. 2017). This gene also demonstrated increased expression in CD21+ transitional/naïve B cells in the independent set of

HC and patients with MS for which we have single cell RNA-seq data, which is trending toward significance (5.2-fold increase, p-value = 0.17). Taken together, these data support a model in which primary B cells from patients with MS have higher expression of *FANCC*, a gene critical for EBV-dependent immortalization of B cells, along with several other genes with prior links to MS and B cell biology. Strikingly, while the primary B cells from patients with MS and controls had equivalent levels of RNA mapped to the EBV genome (Supplemental Fig. S5), primary B cells from patients with MS had significantly higher expression of EBNA2 compared to controls (Mann-Whitney U test, p = 0.0018) (Fig. 1G). The identification of enhanced EBNA2 expression in the primary B cells before transformation for this study suggests a fundamental difference in the B cell compartment of patients with MS relative to controls, a finding consistent with prior studies (Ramesh et al. 2020; Soldan et al. 2024; Sarkkinen et al. 2025).

EBV-transformed B cells from patients with MS have distinct chromatin accessibility patterns compared to controls

Gene expression differences are often driven by differences in chromatin accessibility. To identify chromatin accessibility differences between people with MS and healthy controls, we performed ATAC-seq in primary and EBV-transformed B cells in six HC and six individuals with MS. The resulting data met our stringent quality control criteria, which are based on the ENCODE consortium standards (Luo et al. 2020; Hitz et al. 2023) (Supplemental Table S2). Consistent with the gene expression analysis, we identified only minimal differences in chromatin accessibility in the primary B cells of HC compared to MS (Fig. 2A, Supplemental Table S5). Also consistent with the gene expression analysis, transformation with EBV resulted in large differences between MS and HC, with 1,197 peaks that were stronger in MS and 287 that were stronger in HC (Fig. 2B, Supplemental Table S5). Accounting for *EBNA2* expression levels again removed the vast majority of these disease-dependent differences (Fig. 2C).

We next examined the concordance between chromatin accessibility changes and gene expression changes. To this end, we calculated the statistical significance of the overlap between differentially expressed genes and differentially accessible chromatin using our RELI algorithm (Harley et al. 2018) (see Methods, Supplemental Table S6). Importantly, stronger chromatin accessibility in EBV-transformed B cells from patients with MS corresponded with higher gene expression in these cells (Fig. 2D, right bar). This relationship was not observed in any of the other possible pairwise relationships; for example, we did not observe correspondence between closed chromatin and activated gene expression (Fig. 2D, other bars). As EBNA2 does not directly bind DNA, we also calculated the enrichment of the DNA binding motifs of its TF interaction partners RBPJ, EBF1, and PU.1 (encoded by *SPI1*) in differentially accessible chromatin. The motifs for all three TFs were significantly enriched in chromatin with enhanced accessibility in MS, but not in chromatin that is more accessible in HC (Fig. 2E), supporting a model where increased EBNA2 levels result in increased human gene expression levels through increased chromatin accessibility at regions bound by EBNA2 with its human TF interaction partners.

Binding of EBNA2 and human TF interaction partners in EBV-transformed B cells

Our data reveal extensive MS-dependent chromatin accessibility and gene expression that can be explained by EBNA2 expression levels. To further understand the role of EBNA2 and its TF interaction partners in these differences, we used ChIP-seq to measure the genome-wide binding of EBNA2, RBPJ, EBF1, and SPI1 (PU.1) in five HC and five MS EBV-transformed B cell lines obtained from cohort members with corresponding gene expression and chromatin accessibility measurements. The resulting data met our stringent quality criteria that were derived from the ENCODE consortium (Luo et al. 2020; Hitz et al. 2023) (Supplemental Table S2).

We observed thousands of robust ChIP-seq peaks, with an elevated average peak count in MS compared to controls for all four proteins (Fig. 3A). In cells from patients with MS, EBNA2 binding significantly overlapped genes with increased expression in MS-derived EBV transformed

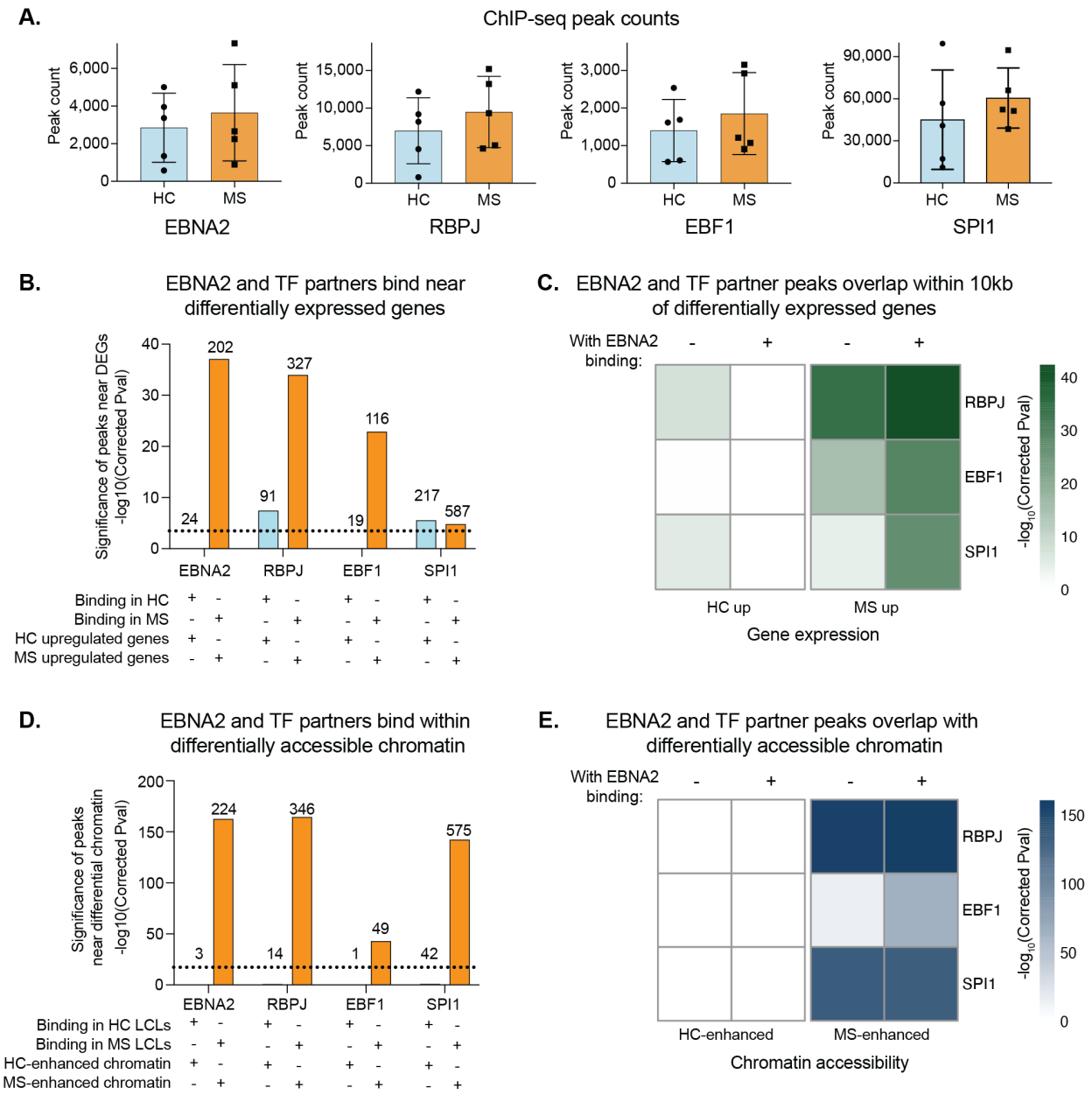

Figure 3. EBNA2 and TF partners bind near genes upregulated in MS EBV-transformed B cells and MS-enriched open chromatin. A. Total number of ChIP-seq peaks for each experiment. Each dot indicates one EBV-transformed cell line from a healthy control (HC) or person with multiple sclerosis (MS). Error bars indicate standard deviation. B. Overlap of all possible pairs of EBNA2, RBPJ, EBF1, or SPI1 (ChIP-seq) and differentially expressed genes with 10 kb padding on either side of the gene (RNA- seq). A combined set of ChIP-seq peaks found in any of the HC or MS-derived EBV-transformed B cells was used for each protein. Differentially expressed genes were higher in either HC or MS-derived EBV-transformed B cells. The statistical enrichment of the overlap (measured by RELI, see Methods) is provided as an adjusted p-value on the y-axis. Dotted line indicates significance threshold. The number of overlaps is provided on top of each bar. C. Co-occurring binding of EBNA2 with each human transcription factor (RBPJ, EBF1, or SPI1) was called within each EBV-

transformed B cell line derived from HC and individuals with MS, using ChIP-seq peaks from the same cell line. For each EBNA2–TF pair, we then compiled a single, pooled set of co-occurrence sites by taking the union of co-occurring peaks observed in any HC- or MS-derived cell line. These co-occurrence peak sets were tested for enrichment near differentially expressed genes (genes upregulated in HC or upregulated in MS). The heatmap shows the significance of each enrichment test, quantified by RELI, for each EBNA2–TF pair and gene set comparison. D. Overlap of all possible pairs of EBNA2, RBPJ, EBF1, or SPI1 (ChIP-seq) and differentially accessible chromatin (ATAC-seq). A combined set of ChIP-seq peaks found in any of the HC or MS-derived EBV-transformed B cells was used for each protein. Chromatin was significantly more accessible in either HC or MS-derived EBV-transformed B cells. The statistical enrichment of the overlap (measured by RELI, see Methods) is provided as an adjusted p-value on the y-axis. The number of overlaps is provided on top of each bar. E. Co-occurrence of EBNA2 and either RBPJ, EBF1, or SPI1 was identified in each EBV-transformed B cell line in both HC and MS. A combined set of co-occurring peaks found in any of the HC or MS-derived cell lines was used. Each set of peaks (ATAC) was significantly more accessible in either HC or MS. The heatmap depicts the significance of the overlap (as measured by RELI) for each pairwise comparison.

B cells (Fig. 3B), binding within 10kb of the TSS for over 25% of genes with MS-dependent expression (Supplemental Table S6). Likewise, the three EBNA2 TF partners also demonstrated significant overlap with genes whose expression increased in MS-derived cell lines (Fig. 3B). When accounting for co-occurrence of EBNA2 and human TF ChIP-seq peaks in cells from the same individual with MS, we again identify the strongest enrichment of overlap of EBNA2-TF pairs at genes with increased expression in MS-derived cells (Fig. 3C).

Next, we compared chromatin accessibility changes with the binding of EBNA2 and identified statistically significant overlap in cells derived from patients with MS and MS enhanced chromatin (Fig. 3D). We also observed strong statistical enrichment for the co-occupancy of EBNA2 with TF partners RBPJ, EBF1, and SPI1 in MS-derived cells within MS-enhanced chromatin, but not for chromatin that is more accessible in HC (Fig. 3E).

In summary, upon assessing EBNA2 binding in cell lines derived from HC, the overlap with HC-dependent genes and HC-enhanced chromatin was minimal, with far less significance relative to MS. These results highlight the fact that many more genes are increased in MS (790 genes) relative to HC (270 genes) (Fig. 1B) and many more ATAC-seq peaks are enhanced in MS (1,197) relative to HC (287) (Fig. 2B). Altogether, these results reveal a model of enhanced

EBNA2 activity in MS compared to HC, leading to alterations in EBNA2 binding, human TF binding, human chromatin accessibility, and human gene expression.

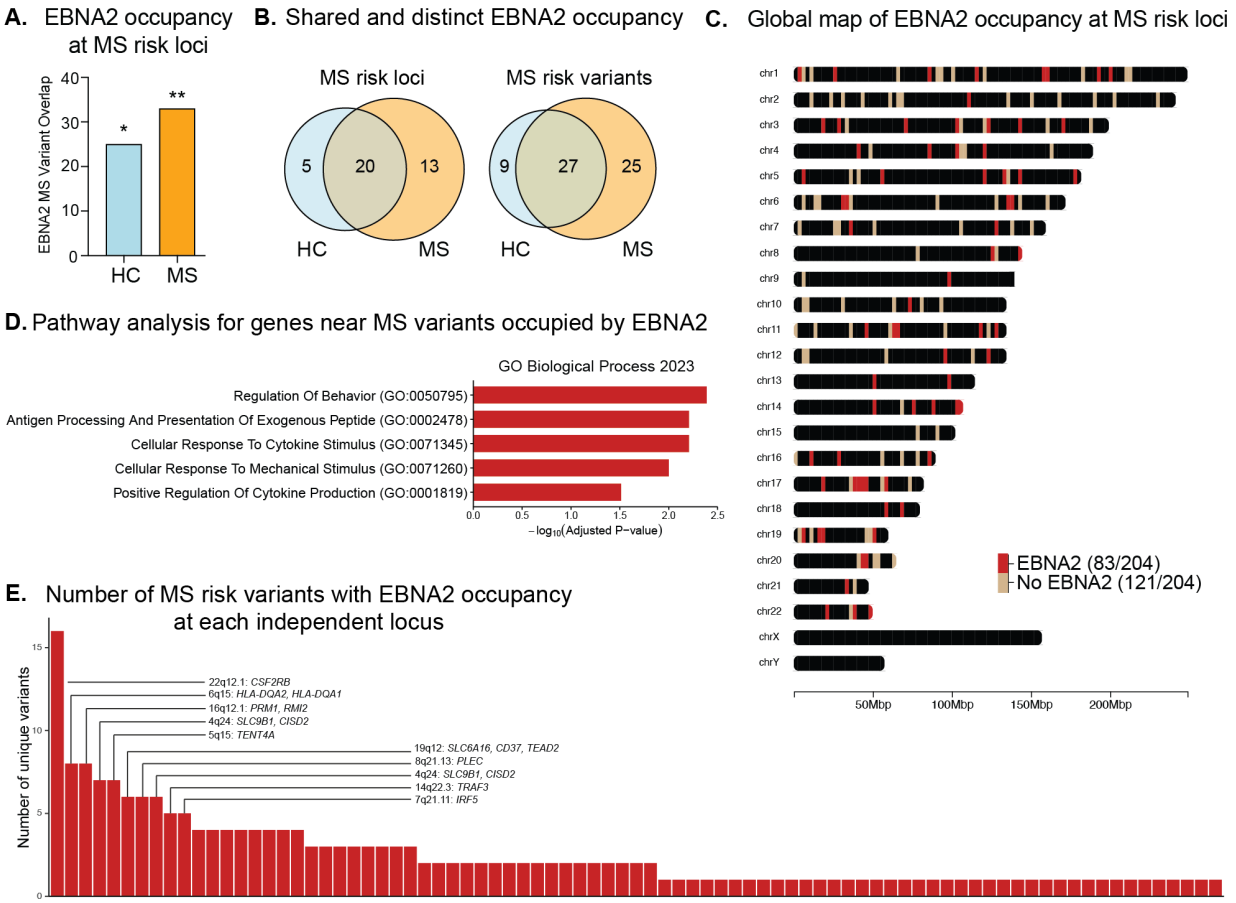

Figure 4. **EBNA2 occupancy at MS genetic risk loci.** A. EBNA2 occupancy at MS risk loci in HC and MS EBV-transformed B cells. Count of EBNA2 ChIP-seq peaks overlapping one or more MS risk variants within each independent locus from HC (left) or MS (right). Overlap significance (as calculated by RELI, see Methods) is indicated with asterisks (*, p = 2.63E-09; **, p = 5.81E-14). B. Comparison of EBNA2 MS risk locus occupancy in HC- or MS-derived cells. (left) Number of MS genetic risk loci where an EBNA2 ChIP-seq peak overlaps with one or more MS risk variant. (right) Total number of these genetic variants (SNPs) (i.e., the counts include multiple variants within a locus). C. Chromomap depicting EBNA2 occupancy at MS risk loci in EBV-transformed B cells. EBNA2 ChIP-seq peaks were collected from a variety of sources, including this study (see Methods). D. Pathway enrichment (Gene Ontology Biological Process) of genes with TSS within 10kb of MS risk loci bound by EBNA2. E. Number of MS risk variants with EBNA2 occupancy at each independent locus.

EBNA2 has increased occupancy at MS genetic risk loci

We next asked whether the enhanced EBNA2 activity in MS-derived cells preferentially targets known MS genetic risk loci. We previously discovered that EBNA2 occupies 44 of the 109

independent MS-associated genetic loci tested (Harley et al. 2018). With the number of established independent genetic risk loci more than doubling in the past few years, and the number of available EBNA2 ChIP-seq datasets likewise increasing (Hong et al. 2021; Viel et al. 2024), we next sought to determine updated EBNA2 occupancy of the expanded set of risk loci in MS-derived cells. Using the RELI algorithm (Harley et al. 2018) and data from the present study, we find that EBNA2 occupies 33 and 25 MS risk loci in EBV-transformed B cells from patients with MS and HC, respectively (Fig. 4A, Supplemental Table S6). Accordingly, the enrichment of EBNA2 at MS risk loci is more significant in MS derived cells (adjusted p-value = 5.81E-14) compared to those from HC (adjusted p-value = 2.63E-09). While most loci and risk variants were shared between HC and MS, many had EBNA2 occupancy only in MS-derived (13 loci; 25 risk variants) or HC-derived (5 loci; 9 risk variants) cells (Fig. 4B).

To further expand the set of EBNA2 human genomic targets, we performed ChIP-seq in 21 additional cell lines in a wide range of B cell types and contexts (Supplemental Table S2). Combining these data with all publicly available EBNA2 genomic binding data, we created an updated global map of EBNA2 occupancy, which reveals EBNA2 occupancy at 83 of the 204 known MS risk loci (Fig. 4C). Pathway enrichment analysis reveals that EBNA2 binds at MS risk loci near genes related to pathways involved in key B cell functions such as antigen presentation, response to cytokine, and cytokine production (Fig. 4D, Supplemental Table S7). While most EBNA2-bound MS risk loci had a single EBNA2 binding event overlapping an MS variant, a subset of MS risk loci had multiple variants bound by EBNA2, such as the *CSF2RB* locus (n=16 EBNA2-bound variants) and *HLA* and others, implicating EBNA2 binding at 204 unique MS variants within 83 independent MS genetic risk loci (Supplemental Table S6) [21].

Genotype-dependent EBNA2 occupancy at MS risk variants

Our analyses indicate that EBNA2 occupancy is significantly enriched at MS risk variants. We next wanted to investigate allele-dependent EBNA2 binding events at these loci. To this end, we

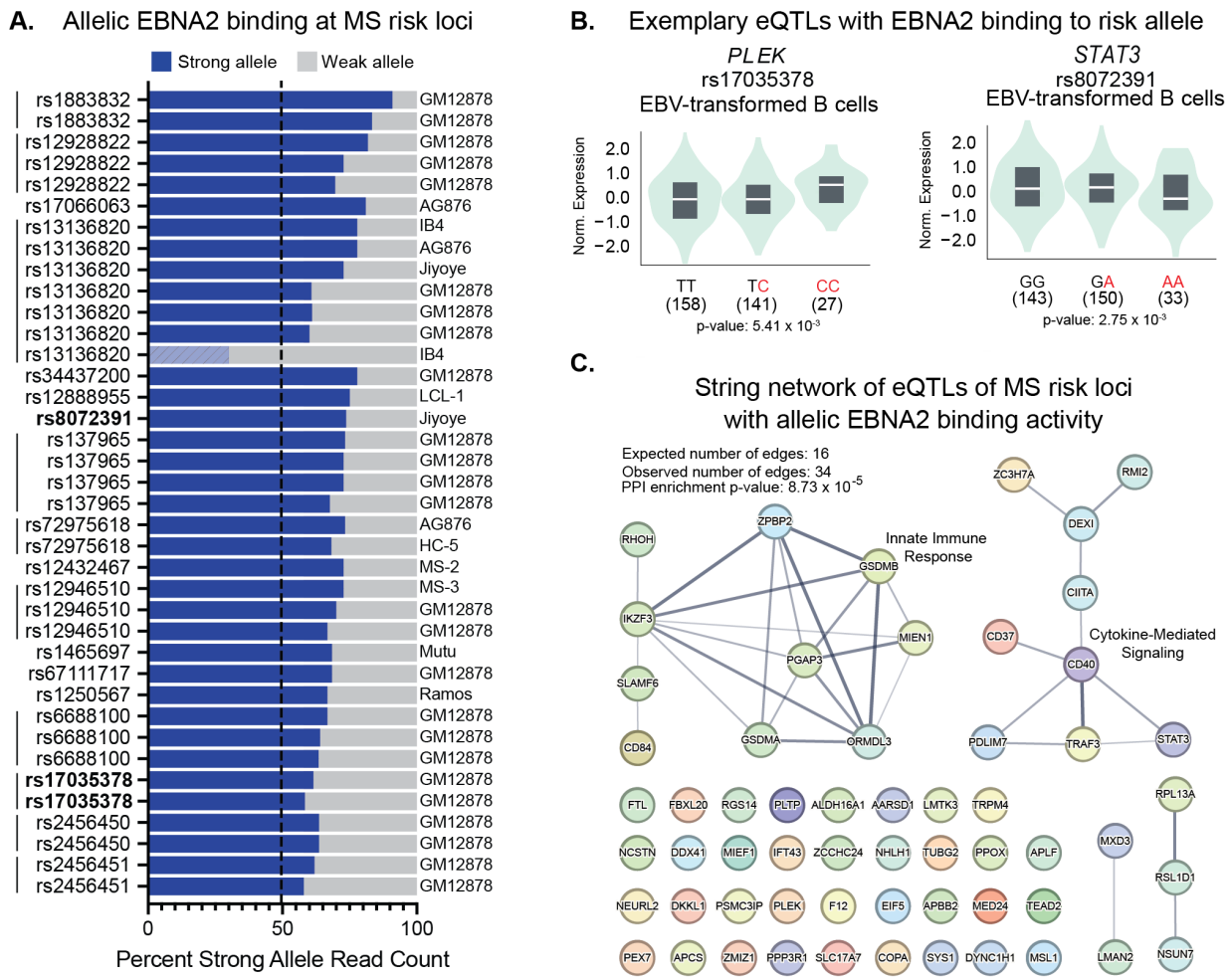

Figure 5. **Allelic EBNA2 occupancy at MS risk variants**. A. MS-associated genetic risk variants with allele-dependent EBNA2 binding. Each variant is heterozygous and located within an EBNA2 ChIP-seq peak to identify allele-dependent occupancy. The x-axis indicates the percent of reads that map to the strong allele vs the weak allele. A strong allele is defined as the allele preferred in the majority of experiments for that variant. A single variant had discordant allelic results in one out of seven experiments (indicated in hatches). Variants shown in eQTL plots in panel B are bolded. Identical variants in different datasets are connected with vertical bars. B. Genotype-dependent expression of *STAT3* and *PLEK* at MS risk variants with allelic EBNA2 binding (identified in GTeX). MS risk alleles are indicated in red. C. eQTLs at MS variants with allelic EBNA2 binding were identified in the eQTL catalogue. A network of these eQTLs was produced using STRING.

used our MARIO pipeline to systematically identify MS risk allele-dependent binding of EBNA2 (Harley et al. 2018). In brief, MARIO identifies heterozygous genetic variants within ChIP-seq peaks with significantly different read counts between the two alleles (see Methods). We combined the EBNA2 ChIP-seq experiments produced in this study with publicly available data

and identified allele-dependent EBNA2 binding events at MS risk variants. In total, we identified 38 instances of allele-dependent EBNA2 binding to 18 unique MS risk variants in EBV-transformed B cells (Fig. 5A, Supplemental Table S8). Among these nominated risk variants, rs17035378 and rs8072391, which both have stronger EBNA2 binding to the MS risk allele, are eQTLs in EBV-transformed B cells for *PLEK* and *STAT3*, respectively (Fig. 5B). When all MS risk variants that are bound by EBNA2 in a genotype-dependent manner are connected through eQTL analysis, including those from this study and previously published variants, a network of genes emerges containing a significant number of interactions (STRING-db analysis, p = 8.73 e-5, Fig. 5C). From the interaction network, we identify a set of genes involved in innate immune response and a second set with roles in cytokine signaling (Fig. 5C). Together, these analyses reveal a large core set of human genes whose expression is likely altered by allele-dependent EBNA2 binding to MS-associated genetic risk variants, revealing important gene/environment disease mechanisms for potential therapeutic targeting.

Discussion:

In this study, we show that the EBV-encoded transcriptional regulator EBNA2 has increased activity in EBV-transformed B cells derived from patients with MS compared to controls, and that this heightened activity amplifies MS-relevant transcriptional and chromatin programs. We observe extensive MS-dependent changes in gene expression and chromatin accessibility that are largely explained by higher EBNA2 expression in MS-derived EBV-transformed B cells (Fig. 1). To explain this phenomenon, we systematically evaluated potential confounders, including B cell clonality, prior EBV exposure, the proportion of CD21+ B cells, overall EBV gene expression, cumulative genetic burden, and HLA-DRB1*15:01 status; none accounted for the MS-dependent gene expression as effectively as *EBNA2* expression levels. Generally, increased EBNA2 expression levels tend to lead to increases in human gene expression levels, affecting genes involved in MS-relevant pathways such as *GPR183* (B cell migration and crosstalk with T cells),

ACKR3 (B cell adhesion and migration), and *CD40* (B cell proliferation and survival). Our MS-derived cells also reveal differential binding of EBNA2, differential chromatin accessibility, and genotype-dependent binding of EBNA2 at MS risk loci.

Overall, we observed dramatic differences between primary B cells before and after EBV-transformation. Prior to EBV infection, we observed only 19 human genes with altered expression between MS and controls, with many involved in B cell activation such as *FANCC*. In an independent single-cell dataset, *FANCC* was again higher in CD21+ transitional/naïve B cells from patients with MS, and a prior CRISPR screen identified *FANCC* as essential for EBV-mediated B cell transformation (Ma et al. 2017). Together, these findings suggest that subtle pre-existing differences in activation programs of CD21+ B cells may prime MS-derived B cells for enhanced EBV transformation and EBNA2-dependent reprogramming.

Subsequent to EBV transformation, we observed substantial differences between patients with MS and matched controls on multiple levels: gene expression levels, chromatin accessibility, and human TF binding. EBNA2 is the most significant differentially expressed EBV gene in MS-derived cells (Fig. 1D), EBNA2 protein levels are higher in MS-derived cells (Fig. 1F), and EBNA2 expression level can account for the extensive gene expression differences seen in HC and MS cell lines (Fig 1C.). Occupancy of EBNA2 and its human TF binding partners is highly enriched for human genes with increased chromatin accessibility and increased expression levels in patients with MS (Fig. 3B, 3D). Many of these EBNA2 binding events also coincide with MS risk loci (Fig. 4), consistent with prior studies from several groups including our own (Ricigliano et al. 2015; Harley et al. 2018; Hong et al. 2021; Viel et al. 2024). In total, EBNA2 occupies 33 and 25 risk loci in EBV-transformed B cells from patients with MS and HC, respectively (Fig. 4A). While the majority of loci were shared between HC and MS, many loci had EBNA2 occupancy only in MS- (13 loci) or HC-derived (5 loci) cells (Fig. 4B). We also identify allelic EBNA2 occupancy at 18 MS risk variants, strengthening our current understanding of MS genotype-dependent biolog. Compared to prior EBNA2 studies that relied on established cell lines, our use of patient-derived

EBV-transformed B cells, coupled with allele-dependent analyses, provides direct evidence that EBNA2 engages MS risk loci in a disease- and genotype-dependent manner.

Our study uses the B95-8 laboratory strain of EBV to infect primary B cells derived from controls and patients with MS, providing a controlled model of latent EBV infection in human B cells with known genetic background and disease status. Consistent with this model, a recent study in systemic lupus erythematosus demonstrated that EBV preferentially infects autoreactive antinuclear B cells and reprograms them into antigen-presenting cells that activate T peripheral helper cells and propagate pathogenic T and B cell responses (Younis et al. 2025). These data provide direct evidence that rare EBV infection events in autoreactive B cells can initiate systemic autoimmunity, aligning with our proposed model for EBV–B cell interactions in MS. They also complement a recent study that used spontaneous lymphoblastoid cell lines to interrogate endogenous EBV in MS-derived B cells, revealing EBV lytic gene expression in active disease compared to both controls and patients with stable disease (Soldan et al. 2024). Those data, together with prior serologic evidence of EBV reactivation during active MS (Wandinger et al. 2000; Farrell et al. 2009), highlight dysregulated lytic infection. In contrast, our work focuses on latent EBV infection under controlled viral and disease conditions, showing that even in the absence of overt lytic reactivation, heightened EBNA2 activity can remodel chromatin and gene expression at MS risk loci. Together, these studies support a model in which impaired control of both latent and lytic EBV infection in B cells contributes to MS pathogenesis. Importantly, both studies reveal reduced control of EBV B cell infection as a potential driver of multiple sclerosis.

The goals of this study are to clarify how interactions between genetic variants and EBNA2 influence MS disease processes, and to define the pathways whose regulation and cell-fate decisions differ in people with MS compared to those without MS. Even common viruses like EBV can contribute to rare non-communicable diseases such as MS. This is likely driven by a combination of host genetics, stochastic events, and environmental factors. In the case of EBV, only a modest minority of people who are infected experience symptoms and develop

mononucleosis (Crawford et al. 2006). Likewise, of the ~95% of people with prior EBV infection, the vast majority will not develop subsequent autoimmune diseases (Ascherio and Munger 2015). While up to 50% of this variation may be due to genetics (International Multiple Sclerosis Genetics 2019), stochasticity also likely plays a role in determining when EBV will lead to MS or other autoimmune diseases. Because only a tiny fraction of B cells are infected with the virus, and the likelihood of those infected B cells possessing a disease-driving BCR is remote (SoRelle et al. 2022), it is possible that the disease-inducing BCR and EBV never occur within the same cell to initiate the cascade of EBNA2-induced changes we found. Likewise, environmental factors like vitamin D level and timing of initial EBV infection will also impact the likelihood of progressing to an MS disease state. Additional studies such as this one are needed to explore these triggers further, for EBV and other viruses.

Our findings reveal a key role for the EBV-encoded EBNA2 protein in transcriptional regulation of MS-derived B cells. Heightened EBNA2 expression in MS-derived cells compared to healthy controls is accompanied by disease-dependent changes in EBNA2 binding, gene expression, and chromatin accessibility. In total, we observe EBNA2 binding at 204 MS genetic risk variants, encompassing 83 distinct MS genetic risk loci, including allele-dependent EBNA2 binding events to 18 different MS-associated variants. Collectively, these results reveal an extensive network of EBNA2-induced MS risk-dependent gene regulatory alterations, providing robust mechanistic targets for therapeutic intervention.

Methods:

Ethics statement

This study was approved by the Cincinnati Children's Hospital Institutional Review Board (IRB #2017-0430) in March of 2017. Both participants with Multiple Sclerosis (MS) and Healthy Controls (HC) signed informed consent forms prior to participation.

Patient recruitment

Patients with a diagnosis of MS (n=12) were recruited from the Waddell Center for Multiple Sclerosis at the University of Cincinnati or through a study advertisement sent out to Cincinnati Children's Hospital employees. To reduce heterogeneity, the study inclusion criteria for subjects with MS was restricted to European American women between the ages of 25 and 74, which is representative of the largest MS demographic. Eleven age- and ancestry- matched controls were recruited from a study advertisement sent out to Cincinnati Children's Hospital employees. Supplemental Table S1 provides the age, sex, self-reported race, ethnicity, and prior EBV infection status of each subject in this study.

Primary B cell isolation

Immediately after sample collection, peripheral blood mononuclear cells (PBMCs) were isolated from blood samples by Ficoll density gradient centrifugation (mean of 132.3 million +/- 84.9 million cells). CD19+ primary B cells were isolated from PBMCs using a Miltenyi negative selection B cell isolation kit (Miltenyi Biotec 130-091-151).

ELISA assays

ELISA assays were performed to evaluate prior history of EBV infection. ELISAs for anti-EBNA1 IgG (cat #EI 2793-9601 G) and EBV Capsid Antigen IgG (cat #EI 2791-9601 G) were performed according to the manufacturer's protocol.

EBV infection of primary B cells

B95-8 EBV was prepared from cell supernatants after culture in 10% FBS-supplemented RPMI-1640 medium for 2 weeks. Viral suspension was filtered twice with 0.45 μm Millipore filters. The concentrated viral stocks were stored at −80°C. Primary B cells were infected with 1 mL viral stock based on infection optimization assays and incubated for 3 h for virus adsorption. After infection,

cells were washed and cultured. Within the first month, we confirmed EBV infection by morphological changes and blasting of the cells. Cells were collected for experiments after at least five passages.

RNA-seq

Total RNA was extracted using the mirVANA Isolation Kit (Ambion) from 11 primary B cells and 11 EBV-transformed B cell lines and DNase treated with the NucleoSpin® RNA Clean-up XS Kit (MACHEREY-NAGEL). RNA libraries were prepared using the SMARTer® Stranded Total RNA-Seq Kit v2 (Takara Bio USA, Inc.), following the manufacturer's protocol, and sequenced 2x150 bp on an Illumina NovaSeq 6000 at Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing Facility (RRID:SCR_022630).

RNA-seq data were analyzed using the nf-core/rnaseq pipeline (version 3.11.1) (doi: 10.1038/s41587-020-0439-x). Adapter sequence contents were removed using cutadapt (Martin 2011) (trimgalore version: 0.6.7) (https://journal.embnet.org/index.php/embnetjournal/article/view/200). RNA-seq reads were aligned to the GRCh38 genome augmented with EBV B95-8 coding sequences (ASM240226v1, GCF_002402265.1) using Spliced Transcripts Alignment to a Reference (STAR, version: 2.7.9a) (https://doi.org/10.1093/bioinformatics/bts635). Ribosomal RNA was removed using the Silva Library (https://www.arb-silva.de/) and SortmeRNA v 4.3.4 (https://bioinfo.lifl.fr/RNA/sortmerna/). All RNA-seq datasets used in this study passed our quality control checks, which were facilitated by FastQC (version: 0.11.9) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Specifically, RseqQC (Wang et al. 2012), Qualimap (Garcia-Alcalde et al. 2012; Okonechnikov et al. 2016), dupRadar (Sayols et al. 2016), and Preseq (https://github.com/smithlabcode/preseq) were used to assess the quality of the datasets. GENCODE (v. 40) (Frankish et al. 2019) annotations for GRCh38 with EBV coding sequence annotations were used to calculate transcript abundance estimates with Salmon (v.1.10.1) (https://doi.org/10.1038/nmeth.4197). The tximport

package (https://doi.org/10.12688/f1000research.7563.2) was used to pre-process quantifications and DESeq2 (Love et al. 2014) was used to perform differential gene expression analysis. Differentially expressed genes were defined as those genes with a fold-change difference greater than 1.5 and an FDR of less than 0.05. DESeq2 output was filtered to only include genes with transcript per million (TPM) values greater than one in at least three replicates of MS or HC cell lines. Variables accounting for gene expression levels (e.g., the expression level of the EBNA2 gene) were statistically controlled for using the DESeq2 design formula " ~log(gene TPM) + disease_status".

Whole genome sequencing and variant calling

Whole genome sequencing was performed on all MS- and HC-derived cells used in this study. DNA was isolated using a PureLink Genomic DNA Kit (Thermo Fisher #K182001). Libraries were sequenced on an Illumina NovaSeq to generate 100-base paired-end reads. Whole genome sequenced files were mapped to the hg38 genome and variants were called using the Sarek v3.4.1 pipeline (Garcia et al. 2020; Hanssen et al. 2024). The Sarek pipeline offers multiple programs for alignment, variant calling, and annotation. These samples were aligned with BWA-mem, variants were called with GATK HaplotypeCaller, and annotated with VEP. In addition, joint variant calling was enabled with the "joint_germline" parameter. To determine prior exposure to EBV, the fraction of reads aligning to the EBV genome was assessed by counting the total number of mapped reads in the duplicate marked CRAM file and comparing to the number of reads mapped to the "chrEBV" contig included as a decoy sequence in the hg38 genome.

Western Blot

Cells were lysed and sonicated, and protein concentration was determined using a BCA protein assay. 10 µg of lysates were loaded in 4-12% NuPAGE Bis-Tris gels, run in MOPS buffer, and transferred to a nitrocellulose membrane using the iBlot machine (Thermo Fisher Scientific). The

membranes were blocked for 1 hour in Intercept Blocking Buffer at room temperature and incubated overnight at 4°C in Intercept Blocking Buffer + 0.1% Tween 20 with anti-EBNA2 (MABE8, Sigma-Aldrich) antibody diluted to 1:1000. The next day, the membranes were washed three times for 5 minutes with PBS + 0.1% Tween 20. The washed membranes were then incubated with rotation for 45 minutes with fluorescent secondary antibodies (diluted to 1:20,000 in Intercept Blocking Buffer + 0.1% Tween + 0.01% SDS). After three more 5-minute washes in PBS + 0.1 Tween 20, the membranes were imaged using an Odyssey DLx Imaging System (Licor Bioscience).

HLA-DRB-15*01 imputation and Polygenic Risk Score (PRS) calculation

The tag variant and allele for HLA-DRB-15*01 is rs3135388 and the MS risk allele is “A” (Goris et al. 2008; Zivkovic et al. 2009; Alcina et al. 2012). The genotype of rs3135388 was obtained from whole genome sequence data of each individual. We applied a previously developed and validated PRS (PGS002726) for multiple sclerosis to individuals recruited to this study (Shams et al. 2023). PGS002726 was obtained from the polygenic risk score catalogue (http://www.pgscatalog.org/). We used the nextflow pipeline “The Polygenic Score Catalog Calculator (pgsc-calc)” to calculate each individual’s PRS from whole genome sequencing data (Lambert et al. 2024).

scRNA-seq

We performed scRNA-seq experiments using cells derived from 5 patients with multiple sclerosis and 5 healthy controls, pooled into one batch. Single-cell 5' immune profiling was performed using the Chromium Next GEM Single Cell 5' Kit v2 (Dual Index, PN-1000263) together with the Single Cell Human BCR Amplification Kit (PN-1000253) from 10x Genomics (Pleasanton, California). Approximately 17,000 cells were loaded onto a Chromium Next GEM Chip K to capture 10,000 cells and processed on a Chromium X Controller to generate barcoded Gel Beads-in-Emulsion

(GEMs). GEM-RT, Dynabeads MyOne SILANE cleanup, and cDNA amplification were performed according to the manufacturer's protocol (CG000331 Rev F). Amplified cDNA was used for both gene expression and BCR V(D)J enrichment and subsequent library construction following the procedures described in the user guide. Final libraries were sequenced on an Illumina NovaSeq 6000 S4 flow cell (PE 100 bp) using the following sequencing parameters: Read 1: 28 cycles; i7 Index: 10 cycles; i5 Index: 10 cycles; Read 2: 90 cycles. Sequencing was performed to a target depth of ~32,500 read pairs per cell for 5' gene expression libraries and ~5,000 read pairs per cell for the BCR libraries.

Initial read alignment and quantification of FASTQ files was performed using Cell Ranger (v 6.1.2) (Zheng et al. 2017). Analysis was performed using Seurat (Butler et al. 2018). Cells with less than 1,000 or greater than 7,000 unique features were removed from downstream analysis. No features that were expressed in 3 or fewer cells were included. Data were normalized using the NormalizeData() command. For clustering and Uniform Manifold Approximation and Projection (UMAP) visualization, 15 components were used; this number was determined based on the strength of PCs visualized by elbow plot and PC heatmaps after PCA was performed. The UMAPs and clusters were generated using the FindNeighbors(), FindClusters(), and RunUMAP() functions (R implementation). The DotPlot was created using the DotPlot() function. Input features for the dot plot were generated through the FindAllMarkers() function provided by Seurat to identify differentially expressed immunoglobulin heavy chain variable region (IGHV) reads.

We also analyzed publicly available single cell RNA-seq data. The dataset GSE133028 was downloaded from the Gene Expression Omnibus, consisting of 3 healthy control and 10 relapsing-remitting multiple sclerosis scRNA samples. Raw fastq files for each patient were aligned to the human genome using Cell Ranger (v. 7.0.1) to make a Seurat object. Cells with greater than 15% mitochondrial genes and cells with less than 500 or greater than 4,500 unique features were removed from analysis. All individual Seurat objects were integrated using Harmony (Korsunsky et al. 2019), and B cells were identified using the gene *MS4A1* (otherwise known as

CD20) and made into a separate Seurat object. 40 components were used to generate UMAP visualization using the FindNeighbors(), FindClusters(), and RunUMAP() functions. B cell subtypes were identified by comparing the expression levels of select genes based on previously published studies (Verstegen et al. 2023; Dai et al. 2024). *CD21* expression across B cell subtypes was analyzed using the VlnPlot() function in Seurat, and the proportion of CD21+ transitional/naïve B cells were extracted to compare between healthy control and MS. Differential gene expression analysis between healthy control and MS was performed using the FindMarkers() command in Seurat (p-value cutoff = 0.05, logFC cutoff = 0.585).

ATAC-seq

Transposase Tn5 with adapter sequences was used to cut accessible DNA (Buenrostro et al. 2013). These accessible DNA sequences with adapter sequences were isolated, and libraries were prepared from ~50,000 primary B cells or EBV-transformed B cells using the OMNI ATAC protocol (Corces et al. 2017). The libraries were sequenced 2x150 bp on an Illumina NovaSeq 6000 at the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing Facility.

ATAC-seq data were processed and aligned to a combined human (hg38) and EBV (B95-8) genome using the ENCODE ATAC-seq pipeline (V2.0.0) (Consortium 2012; Luo et al. 2020; Hitz et al. 2023) (Doi: 10.5281/zenodo.5598324). Aligned reads were then sorted using samtools (v.1.9) and duplicate reads were removed using Picard (v. 2.20.7) (https://broadinstitute.github.io/picard/). Peaks were called within the pipeline using MACS2 (v. 2.2.4) (Zhang et al. 2008). ENCODE's exclusion list regions were removed from the final set of peaks. Differential chromatin accessibility was calculated using DiffBind (v. 3.0.15) (Stark R 2011), with thresholds of a fold-change greater than 1.5 and adjusted p-value less than 0.05. Each individual-derived cell line was used as a replicate. Expression levels of EBNA2 were

statistically controlled for using the DESeq2 design formula "~log(EBNA2 TPM) + disease_status".

ChIP-seq

We performed ChIP-seq in EBV-transformed B cells from 5 MS and 5 HC for EBNA2 (Abcam ab90543; Lot GR3185425-6), RBPJ (Cell Signaling Technology 5313; Lot 3), PU.1 (Cell Signaling Technology 2266; Lot 3), and EBF1 (Santa Cruz sc-137065; Lot J1122) using standard experimental procedures as previously described (Hong et al. 2021). Cells were incubated in a crosslinking solution (1% formaldehyde, 5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0, 10 mM sodium chloride, 0.1 mM ethylenediaminetetraacetic acid (EDTA), and 0.05 mM ethylene glycol tetraacetic acid (EGTA)) in Roswell Park Memorial Institute (RPMI) culture medium with 10% fetal bovine serum (FBS) and placed on a tube rotator at room temperature for 10 min. To stop crosslinking, glycine was added to a final concentration of 0.125 M, and the tubes were rotated at room temperature for 5 min. Cells were washed twice with ice-cold phosphate-buffered saline (PBS), resuspended in lysis buffer 1 (50 mM HEPES pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.25% Triton X-100, and 0.5% NP-40), and incubated for 10 min on ice. Nuclei were harvested after centrifugation at 5,000 rpm for 10 min, resuspended in lysis buffer 2 (10 mM Tris–HCl pH 8.0, 1 mM EDTA, 200 mM NaCl, and 0.5 mM EGTA), and incubated at room temperature for 10 min. Protease and phosphatase inhibitors (Halt™ Protease and Phosphatase Inhibitor Cocktail (100X), Thermo Fisher Scientific, Waltham, MA) were included in both lysis buffers. Nuclei were resuspended in sonication buffer (10 mM Tris [pH 8.0], 1 mM EDTA, and 0.1% sodium dodecyl sulfate (SDS)). An S220 focused ultrasonicator (COVARIS, Woburn, MA) was used to shear chromatin (150–500-bp fragments) with 10% duty cycle, 175 peak power, and 200 bursts per cycle for 7 min. A portion of the sonicated chromatin was run on an agarose gel to verify fragment sizes. Sheared chromatin was pre-cleared with 10 μL of Protein A or G Dynabeads (Thermo Fisher Scientific) at 4 °C for 1 h.

Immunoprecipitation of TF-chromatin complexes was performed with an SX-8X IP-STAR compact automated system (Diagenode). Beads conjugated to antibodies against each TF were incubated with pre-cleared chromatin at 4 °C for 8 h. The beads were then washed sequentially with wash buffer 1 (10 mM Tris–HCl [pH 7.5], 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% NaDOC, and 1% Triton X-100), wash buffer 2 (10 mM Tris–HCl [pH 7.6], 400 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% NaDOC, and 1% Triton X-100), wash buffer 3 (10 mM Tris–HCl [pH 8.0], 250 mM LiCl, 1 mM EDTA, 0.5% NaDOC, and 0.5% NP-40), and wash buffer 4 (10 mM Tris–HCl [pH 8.0], 1 mM EDTA, and 0.2% Triton X-100). Finally, the beads were resuspended in 10 mM Tris–HCl (pH 7.5) and used to prepare libraries via ChIPmentation.

The ChIP-seq libraries were sequenced as single-end 100-base reads on an Illumina NovaSeq 6000 at the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing Facility. ChIP-seq data were processed and aligned to a combined human (hg38) and EBV (B-958) genome using the ENCODE ChIP-seq pipeline (V2.0.0) (Consortium 2012; Luo et al. 2020; Hitz et al. 2023) (Doi: 10.5281/zenodo.5598352). Adapters were trimmed using fastp (v. 0.23.2) (Chen et al. 2018) prior to processing through the pipeline. Aligned reads were then sorted using samtools (v.1.9) and duplicate reads were removed using Picard (v. 2.20.7) (https://broadinstitute.github.io/picard/). Peaks were called using the pipeline's default parameters with MACS2 (v. 2.2.4) (Zhang et al. 2008). ENCODE "exclusion list regions" were removed from the final peak set. QC results are available in Supplemental Table S2.

To identify HC- and MS- specific genomic features for ChIP-seq, all peaks that were found in each condition were combined and overlapping events were merged using bedtools (Quinlan and Hall 2010). Peaks unique to HC or MS were identified using the bedtools command "subtract -A". Shared peaks were determined by removing all condition-specific peaks from the total peak set.

Enrichment analysis for functional genomic datasets

We used the RELI algorithm to estimate the significance of the overlap between the genomic feature sets (e.g., ChIP-seq peaks) generated in this study (Harley et al. 2018). As input, RELI takes the genomic coordinates of peaks from two datasets. RELI then systematically intersects these coordinates with one another, and the number of input regions overlapping the peaks is counted. Next, a p-value describing the significance of this overlap is estimated using a simulation-based procedure in which the peaks from the first dataset are randomly distributed within the union coordinates of open chromatin from human cells. A distribution of expected overlap values is then created from 2,000 iterations of randomly sampling from the negative set, each time choosing a set of negative examples that match the input set in terms of the total number of genomic loci. The distribution of the expected overlap values from the randomized data resembles a normal distribution and can thus be used to generate a Z-score and corresponding p-value estimating the significance of the observed number of input regions that overlap each dataset.

GWAS data set curation

We obtained GWAS data from the NHGRI-EBI GWAS catalog (Buniello et al. 2019) and the supplemental dataset from the International Multiple Sclerosis Genetics Consortium GWAS publication (International Multiple Sclerosis Genetics 2019). A genome-wide significance cutoff of 5×10^{-8} was used to establish the statistical significance of a variant and its association to a given disease or trait. Independent loci were identified using LD-based pruning in PLINK (Purcell et al. 2007) (window size 300,000 kb, SNP shift size 100,000 kb, and $r^2 < 0.2$).

Pathway enrichment analysis

Pathway analysis for MS and HC differentially expressed genes was performed using Enrichr (Chen et al. 2013; Kuleshov et al. 2016).

TF DNA binding motif enrichment analysis

The HOMER software package (v. 4.9) (Heinz et al. 2010) was used to calculate TF DNA binding site motif enrichment for a custom library containing the human CisBP database motifs (build 2.0) (Weirauch et al. 2014; Lambert et al. 2019). HOMER was modified to use a $\log_2$-based likelihood scoring system.

Identification of allele-dependent sequencing reads using MARIO

Allele-dependent behavior was identified in sequencing reads using the MARIO pipeline (Harley et al. 2018). Briefly, the MARIO pipeline identifies allele-dependent behavior by weighing (1) the imbalance between the number of reads that are mapped to each allele, (2) the total number of reads mapped at each variant, and (3) the number and consistency of available replicates. These variables are combined into a single Allelic Reproducibility Score (ARS), which reflects the degree of allelic behavior observed for the given heterozygous variant in the given data set. MARIO ARS values exceeding 0.4 were considered to be allelic, following our previous study (Harley et al. 2018). We implemented the MARIO pipeline for EBNA2 ChIP-seq data from this study along with publicly available EBNA2 binding data (McClellan et al. 2013; Hong et al. 2021). To select for consistent allele-dependent binding activity, we identified and graphed ChIP-seq peaks with at least 9 reads on top of an MS variant and with at least 15% difference between the reads mapping to the strong and weak alleles (Fig. 5A, Supplemental Table S8).

To expand our reach, we assessed additional EBNA2 ChIP-seq experiments performed in our laboratory, including the cell lines GM12878, IB4, AG876, and Jiyoye. Additionally, we included MS risk variants with published allelic EBNA2 binding activity (Supplemental Table S8) (Harley et al. 2018; Hong et al. 2021). All identified MS risk variants that were overlapped by EBNA2 ChIP-seq peaks in an allelic manner were assessed for eQTL function using release 7 of the eQTL Catalogue (Kerimov et al. 2021). Gene targets of these eQTLs (p-value < 0.01) were used as input to STRING (Szklarczyk et al. 2023) (Fig. 5C).

Competing interest statement

The authors declare no competing interests.

Acknowledgements

This research was funded by the National Institutes of Health (NIH) R01 NS099068, P30 AR070549, R01 AR073228, and R01 AI024717 to M.T.W. and L.C.K.; U19 AI070235, R01 DK107502, R01 AI148276, and U01 HG011172 to L.C.K.; R01 HG010730, R01 GM055479, U01 AI130830, U01 AI150748, R01 AI141569, P01 AI150585, and U24 HG013078 to M.T.W. We thank the Single Cell Genomics Core (RRID:SCR_022653) at the Cincinnati Children's Hospital Medical Center, especially Kelly Rangel, for their assistance with scRNA-seq. We thank the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing Facility (RRID:SCR_022630). We thank John Harley for helpful discussions around the role of EBNA2 in autoimmunity.

Author contributions

Marissa Granitto and Ellie Kim contributed equally to this work. Matthew Weirauch and Leah Kottyan co-led this work. Conceptualization: M.G., L.C.K., M.T.W.; Study design: M.G., L.C.K., M.T.W., C.F.; Participant recruitment: C.F., V.B., K.F., A.Z.; Statistical analysis design: M.G., L.C.K, M.T.W.; Computational analysis: M.G., P.J.D., A.V.H. S.P., E.K., C.O.S., X.C., L.E.E., K.K., K.F., B.G.; Directed experiments: M.G., C.F., L.C.K., M.T.W.; Performed Experiments: M.G., C.F., C.Y., M.S.S., O.D., A.A.D., K.D., V.L., D.H.; Writing: M.G., E.K., L.C.K., M.T.W, L.P.L.; Funding: L.C.K., M.T.W. All authors reviewed and approved the final manuscript.

References

Afrasiabi A, Parnell GP, Fewings N, Schibeci SD, Basuki MA, Chandramohan R, Zhou Y, Taylor B, Brown DA, Swaminathan S et al. 2019. Evidence from genome wide association studies implicates reduced control of Epstein-Barr virus infection in multiple sclerosis susceptibility. *Genome Med* **11**: 26.

Afrasiabi A, Parnell GP, Swaminathan S, Stewart GJ, Booth DR. 2020. The interaction of Multiple Sclerosis risk loci with Epstein-Barr virus phenotypes implicates the virus in pathogenesis. *Sci Rep* **10**: 193.

Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernandez O, Ndagire D, Catala-Rabasa A, Ruiz A, Gayan J et al. 2012. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One* **7**: e29819.

Ascherio A, Munger KL. 2015. EBV and Autoimmunity. *Curr Top Microbiol Immunol* **390**: 365-385.

Bae SC, Lee YH. 2017. Association between BANK1 polymorphisms and susceptibility to autoimmune diseases: A meta-analysis. *Cell Mol Biol (Noisy-le-grand)* **63**: 29-35.

Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL et al. 2022. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**: 296-301.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E et al. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**: D1005-D1012.

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**: 411-420.

Cable JM, Grabowski JC, Luftig MA. 2025. EBNA2 and EBNA-LP: The Earliest Viral Latency Proteins. *Curr Top Microbiol Immunol* doi:10.1007/82_2025_335.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**: 128.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884-i890.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**: 959-962.

Crawford DH, Macsween KF, Higgins CD, Thomas R, McAulay K, Williams H, Harrison N, Reid S, Conacher M, Douglas J et al. 2006. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. *Clin Infect Dis* **43**: 276-282.

Dai D, Gu S, Han X, Ding H, Jiang Y, Zhang X, Yao C, Hong S, Zhang J, Shen Y et al. 2024. The transcription factor ZEB2 drives the formation of age-associated B cells. *Science* **383**: 413-421.

DeLorenze GN, Munger KL, Lennette ET, Orentreich N, Vogelman JH, Ascherio A. 2006. Epstein-Barr virus and multiple sclerosis: evidence of association from a prospective study with long-term follow-up. *Arch Neurol* **63**: 839-844.

Farrell RA, Antony D, Wall GR, Clark DA, Fisniku L, Swanton J, Khaleeli Z, Schmierer K, Miller DH, Giovannoni G. 2009. Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. *Neurology* **73**: 32-38.

Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, Rocca MA. 2018. Multiple sclerosis. *Nat Rev Dis Primers* **4**: 43.

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766-D773.

Garcia M, Juhos S, Larsson M, Olason PI, Martin M, Eisfeldt J, DiLorenzo S, Sandgren J, Diaz De Stahl T, Ewels P et al. 2020. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res* **9**: 63.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**: 2678-2679.

Gatto D, Brink R. 2013. B cell localization: regulation by EBI2 and its oxysterol ligand. *Trends Immunol* **34**: 336-341.

Goris A, Walton A, Ban M, Dubois B, Compston A, Sawcer S. 2008. A Taqman assay for high-throughput genotyping of the multiple sclerosis-associated HLA-DRB1*1501 allele. *Tissue Antigens* **72**: 401-403.

Hanssen F, Garcia MU, Folkersen L, Pedersen AS, Lescai F, Jodoin S, Miller E, Seybold M, Wacker O, Smith N et al. 2024. Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. *NAR Genom Bioinform* **6**: lqae031.

Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, Magnusen AF, Lynch A, Chetal K, Yukawa M et al. 2018. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat Genet* **50**: 699-707.

Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, Lublin F, Montalban X, Rammohan KW, Selmaj K et al. 2017. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *N Engl J Med* **376**: 221-234.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Hitz BC, Jin-Wook L, Jolanki O, Kagda MS, Graham K, Sud P, Gabdank I, Strattan JS, Sloan CA, Dreszer T et al. 2023. The ENCODE Uniform Analysis Pipelines. *bioRxiv* doi:10.1101/2023.04.04.535623.

Hollenbach JA, Oksenberg JR. 2015. The immunogenetics of multiple sclerosis: A comprehensive review. *J Autoimmun* **64**: 13-25.

Hong T, Parameswaran S, Donmez OA, Miller D, Forney C, Lape M, Saint Just Ribeiro M, Liang J, Edsall LE, Magnusen AF et al. 2021. Epstein-Barr virus nuclear antigen 2 extensively rewires the human chromatin landscape at autoimmune risk loci. *Genome Res* **31**: 2185-2198.

International Multiple Sclerosis Genetics C. 2019. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**.

International Multiple Sclerosis Genetics C Wellcome Trust Case Control C Sawcer S Hellenthal G Pirinen M Spencer CC Patsopoulos NA Moutsianas L Dilthey A Su Z et al. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**: 214-219.

Itoh-Nakadai A, Hikota R, Muto A, Kometani K, Watanabe-Matsui M, Sato Y, Kobayashi M, Nakamura A, Miura Y, Yano Y et al. 2014. The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program. *Nat Immunol* **15**: 1171-1180.

Johannsen E, Koh E, Mosialos G, Tong X, Kieff E, Grossman SR. 1995. Epstein-Barr virus nuclear protein 2 transactivation of the latent membrane protein 1 promoter is mediated by J kappa and PU.1. *J Virol* **69**: 253-262.

Keane JT, Afrasiabi A, Schibeci SD, Swaminathan S, Parnell GP, Booth DR. 2021. The interaction of Epstein-Barr virus encoded transcription factor EBNA2 with multiple sclerosis risk loci is dependent on the risk genotype. *EBioMedicine* **71**: 103572.

Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, Samovica M, Sakthivel MP, Kuzmin I, Trevanion SJ et al. 2021. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* **53**: 1290-1299.

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**: 1289-1296.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**: W90-97.

Lambert SA, Wingfield B, Gibson JT, Gil L, Ramachandran S, Yvon F, Saverimuttu S, Tinsley E, Lewis E, Ritchie SC et al. 2024. Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet* **56**: 1989-1994.

Lambert SA, Yang AWH, Sasse A, Cowley G, Albu M, Caddick MX, Morris QD, Weirauch MT, Hughes TR. 2019. Similarity regression predicts evolution of transcription factor sequence specificity. *Nat Genet* **51**: 981-989.

Lanz TV, Brewer RC, Ho PP, Moon JS, Jude KM, Fernandez D, Fernandes RA, Gomez AM, Nadj GS, Bartley CM et al. 2022. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* **603**: 321-327.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K et al. 2020. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**: D882-D889.

Ma Y, Walsh MJ, Bernhardt K, Ashbaugh CW, Trudeau SJ, Ashbaugh IY, Jiang S, Jiang C, Zhao B, Root DE et al. 2017. CRISPR/Cas9 Screens Reveal Epstein-Barr Virus-Transformed B Cell Host Dependency Factors. *Cell Host Microbe* **21**: 580-591 e587.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**: 3.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**: 1190-1195.

McClellan MJ, Wood CD, Ojeniyi O, Cooper TJ, Kanhere A, Arvey A, Webb HM, Palermo RD, Harth-Hertle ML, Kempkes B et al. 2013. Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming. *PLoS Pathog* **9**: e1003636.

Mechelli R, Manzari C, Policano C, Annese A, Picardi E, Umeton R, Fornasiero A, D'Erchia AM, Buscarinu MC, Agliardi C et al. 2015. Epstein-Barr virus genetic variants are associated with multiple sclerosis. *Neurology* **84**: 1362-1368.

Murata T, Noda C, Narita Y, Watanabe T, Yoshida M, Ashio K, Sato Y, Goshima F, Kanda T, Yoshiyama H et al. 2016. Induction of Epstein-Barr Virus Oncoprotein LMP1 by Transcription Factors AP-2 and Early B Cell Factor. *J Virol* **90**: 3873-3889.

Okonechnikov K, Conesa A, Garcia-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**: 292-294.

Olsson T, Barcellos LF, Alfredsson L. 2017. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol* **13**: 25-36.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Ramesh A, Schubert RD, Greenfield AL, Dandekar R, Loudermilk R, Sabatino JJ, Jr., Koelzer MT, Tran EB, Koshal K, Kim K et al. 2020. A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. *Proc Natl Acad Sci U S A* **117**: 22932-22943.

Ressing ME, van Gent M, Gram AM, Hooykaas MJ, Piersma SJ, Wiertz EJ. 2015. Immune Evasion by Epstein-Barr Virus. *Curr Top Microbiol Immunol* **391**: 355-381.

Ricigliano VA, Handel AE, Sandve GK, Annibali V, Ristori G, Mechelli R, Cader MZ, Salvetti M. 2015. EBNA2 binds to genomic intervals associated with multiple sclerosis and overlaps with vitamin D receptor occupancy. *PLoS One* **10**: e0119605.

Sarkkinen J, Yohannes DA, Kreivi N, Durnsteiner P, Elsakova A, Huuhtanen J, Nowlan K, Kurdo G, Linden R, Saarela M et al. 2025. Altered immune landscape of cervical lymph nodes reveals Epstein-Barr virus signature in multiple sclerosis. *Sci Immunol* **10**: eadl3604.

Sayols S, Scherzinger D, Klein H. 2016. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17**: 428.

Schmidt H, Williamson D, Ashley-Koch A. 2007. HLA-DR15 haplotype and multiple sclerosis: a HuGE review. *Am J Epidemiol* **165**: 1097-1109.

Shams H, Shao X, Santaniello A, Kirkish G, Harroud A, Ma Q, Isobe N, University of California San Francisco MSET, Schaefer CA, McCauley JL et al. 2023. Polygenic risk score association with multiple sclerosis susceptibility and phenotype in Europeans. *Brain* **146**: 645-656.

Soldan SS, Su C, Monaco MC, Yoon L, Kannan T, Zankharia U, Patel RJ, Dheekollu J, Vladimirova O, Dowling JW et al. 2024. Multiple sclerosis patient-derived spontaneous B cells have distinct EBV and host gene expression profiles in active disease. *Nat Microbiol* **9**: 1540-1554.

SoRelle ED, Dai J, Reinoso-Vizcaino NM, Barry AP, Chan C, Luftig MA. 2022. Time-resolved transcriptomes reveal diverse B cell fate trajectories in the early response to Epstein-Barr virus infection. *Cell Rep* **40**: 111286.

Stark R BG. 2011. DiffBind: differential binding analysis of ChIP-Seq peak data., http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S et al. 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**: D638-D646.

Thorley-Lawson DA. 2015. EBV Persistence--Introducing the Virus. *Curr Top Microbiol Immunol* **390**: 151-209.

Tsukumo SI, Subramani PG, Seija N, Tabata M, Maekawa Y, Mori Y, Ishifune C, Itoh Y, Ota M, Fujio K et al. 2022. AFF3, a susceptibility factor for autoimmune diseases, is a molecular facilitator of immunoglobulin class switch recombination. *Sci Adv* **8**: eabq0008.

Vermersch P, Granziera C, Mao-Draayer Y, Cutter G, Kalbus O, Staikov I, Dufek M, Saubadu S, Bejuit R, Truffinet P et al. 2024. Inhibition of CD40L with Frexalimab in Multiple Sclerosis. *N Engl J Med* **390**: 589-600.

Verstegen NJM, Pollastro S, Unger PA, Marsman C, Elias G, Jorritsma T, Streutker M, Bassler K, Haendler K, Rispens T et al. 2023. Single-cell analysis reveals dynamics of human B cell differentiation and identifies novel B and antibody-secreting cell intermediates. *Elife* **12**.

Viel K, Parameswaran S, Donmez OA, Forney CR, Hass MR, Yin C, Jones SH, Prosser HK, Diouf AA, Gittens OE et al. 2024. Shared and distinct interactions of type 1 and type 2 Epstein-Barr Nuclear Antigen 2 with the human genome. *BMC Genomics* **25**: 273.

von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, Cree BA, Gourraud PA, Schwartzburg A, Huerta G et al. 2012. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest* **122**: 4533-4543.

Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, Robertson N, La Rocca N, Uitdehaag B, van der Mei I et al. 2020. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult Scler* **26**: 1816-1821.

Waltzer L, Logeat F, Brou C, Israel A, Sergeant A, Manet E. 1994. The human J kappa recombination signal sequence binding protein (RBP-J kappa) targets the Epstein-Barr virus EBNA2 protein to its DNA responsive elements. *EMBO J* **13**: 5633-5638.

Wandinger K, Jabs W, Siekhaus A, Bubel S, Trillenberg P, Wagner H, Wessel K, Kirchner H, Hennig H. 2000. Association between clinical disease activity and Epstein-Barr virus reactivation in MS. *Neurology* **55**: 178-184.

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**: 2184-2185.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**: 1431-1443.

Younis S, Moutusy SI, Rasouli S, Jahanbani S, Pandit M, Wu X, Acharya S, Sharpe O, Wijeratne TU, Harris ML et al. 2025. Epstein-Barr virus reprograms autoreactive B cells as antigen-presenting cells in systemic lupus erythematosus. *Sci Transl Med* **17**: eady0210.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhao B, Zou J, Wang H, Johannsen E, Peng CW, Quackenbush J, Mar JC, Morton CC, Freedman ML, Blacklow SC et al. 2011. Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. *Proc Natl Acad Sci U S A* **108**: 14902-14907.

Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**: 14049.

Zivkovic M, Stankovic A, Dincic E, Popovic M, Popovic S, Raicevic R, Alavantic D. 2009. The tag SNP for HLA-DRB1*1501, rs3135388, is significantly associated with multiple sclerosis susceptibility: cost-effective high-throughput detection by real-time PCR. *Clin Chim Acta* **406**: 27-30.

Supplemental Figures and Tables

Supplemental tables.

Supplemental Table S1. Patient demographic information and clinical phenotypes. This contains the demographics and clinical phenotypes of each individual (Healthy Controls and patients with Multiple Sclerosis).

Supplemental Table S2. Functional genomics data quality assessments. This contains the data quality summary report for all ChIP, ATAC, and RNA-seq datasets used.

Supplemental Table S3. RNA-seq raw and analyzed data. This contains the TPM and differential analysis for RNA-seq data.

Supplemental Table S4. Single cell RNA-seq differential analysis of CD21+ Naïve and Transitional B cells. This contains differential expression analysis of the public single cell RNA-seq dataset GSE133028, particularly in the CD21+ Naïve and Transitional B cell compartment.

Supplemental Table S5. ATAC-seq differential analysis. This contains the differential analysis for ATAC-seq data.

Supplemental Table S6. Integrated analysis of genomic datasets using RELI. This contains enrichment analyses performed using RELI including all MS risk variants and unique loci that are overlapped by EBNA2.

Supplemental Table S7. Pathway analysis of genes proximal to EBNA2-bound MS risk loci. This contains the pathway analysis (Enrichr) for the genes proximal to EBNA2-bound MS risk loci.

Supplemental Table S8. Analysis of EBNA2-bound alleles using MARIO. This contains allelic EBNA2 binding events derived from our patient cohort, previously published EBNA2 ChIP-seq datasets, and published EBNA2 binding events from previous studies (Harley et al. 2018; Hong et al. 2021).

Supplemental figures.

Supplemental Figure S1. EBV-transformed B cell line generation.

Supplemental Figure S2. EBNA2 expression level accounts for the largest amount of differentially expressed genes compared to other EBV genes.

Supplemental Figure S3. EBNA2-dependent changes in gene expression.

Supplemental Figure S4. Increased EBNA2 protein in MS-derived EBV-transformed B cells.

Supplemental Figure S5. Evidence of prior exposure of individuals in this study to EBV.

Supplemental Figure S6. Single cell RNA-sequencing with BCR enrichment allows for enumeration of unique clones for each EBV-transformed B cell line.

Supplemental Figure S7. Multiple sclerosis genetic risk variants do not fully account for differential gene expression in EBV-transformed B cells derived from HC and MS.

Supplemental Figure S8. There are few transcriptional differences in CD21+ transitional/naïve B cells between HC and MS.