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

Abstract:

Background. 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.

Results. 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.

Conclusions. 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.

72 **Background:**

73 Multiple sclerosis (MS) is an immune-mediated demyelinating disease of the
74 central nervous system, with approximately 2.8 million individuals affected globally [1]. It
75 is the most common cause of non-traumatic neurological disability in young adults [2].
76 MS is caused by a complex combination of genetic and environmental risk factor
77 interactions [3]. Genetic factors account for approximately 30% of disease risk, with the
78 unaccounted risk being attributed to environmental factors, including Epstein-Barr virus
79 (EBV) infection [4].

80 Genome-wide association studies have nominated 233 independent risk loci for
81 MS [4]. HLA-DRB-15*01 is the only MS risk locus with a moderate effect size (odds ratio
82 ~ 3), with all other variants having small effect sizes [5]. Genetic risk variants are strongly
83 enriched within non-coding regions of the genome, including enhancers and promoters
84 [6], suggesting that transcriptional dysregulation is likely key to MS etiology.

85 The most well-supported environmental factor in the etiology of MS is EBV, a
86 common gammaherpesvirus that infects over 90% of the world's population. However, in
87 patients with MS, the infection rate is consistently higher (over 99%) [7]. In 2022, a
88 longitudinal study solidified this relationship by demonstrating that EBV infection is
89 required for MS pathogenesis. In this retrospective study, EBV infection in early adulthood
90 resulted in a 32-fold increased risk of MS, and serum levels of neurofilament light chain,
91 a biomarker of neuronal damage, increased only after EBV infection [8].

92 Most cell types driving MS pathophysiology are immune-related (e.g., B and T
93 cells), and the success of anti-CD20 therapies in reducing both the severity of symptoms
94 and relapse rate in MS suggests that B cells likely play important roles [9]. 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 [10, 11]. Latency III is the most prevalent viral phase during chronic disease, with the EBV transcriptome facilitating immune evasion [12].

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 [13]. EBNA2 is a viral nuclear protein that serves as a transcriptional coactivator [14]. 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 [14]. EBNA2 binds indirectly to human DNA via human-encoded transcription factors (TFs) such as RBPJ, EBF1, and PU.1 [15-17]. Our previous work has demonstrated that 44 MS-associated genetic loci contain a genetic variant located within an EBNA2-occupied region of the genome [18]. 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 [19].

117 In this study, we sought to understand how EBV infection leads to altered gene
118 expression in B cells of patients with MS. To this end, we obtained primary B cells from
119 patients with MS along with matched healthy controls (HC). After measuring gene
120 expression and chromatin accessibility in primary B cells, we uniformly infected cells with
121 the B95-8 strain of EBV. In the resulting transformed B cell lines, we measured gene
122 expression, chromatin accessibility, and genomic binding of EBNA2 and its human-
123 encoded TF interaction partners RBPJ, EBF1, and PU.1. Our results establish that EBV
124 infection causes extensive, robust MS-dependent changes to human gene expression
125 and chromatin accessibility. EBNA2 expression levels account for virtually all of these
126 differences, and EBNA2 binds along with its human TF partners at differentially expressed
127 genes and differentially accessible chromatin. Building upon past studies, we
128 demonstrate highly enriched EBNA2 occupancy at MS genetic risk loci in MS patient-
129 derived cells, with increased enrichment in patients with MS relative to healthy controls.
130 Altogether, this study establishes molecular mechanisms across the human genome
131 through which EBV rewires B cells in an MS-dependent manner through EBNA2.
132
133 **Results:**
134 We recruited individuals representing the largest demographic of patients in the United
135 States: MS women of European ancestry between the age of 25-74 [2]. Twelve individuals
136 with clinician-diagnosed MS were recruited along with eleven age-matched healthy
137 controls. The demographics and clinical phenotypes of each individual are provided in
138 Additional file 1: Table S1. Peripheral blood mononuclear cells (PBMCs) were obtained
139 from each subject. CD19+ primary B cells were isolated and EBV-transformed B cell lines

140 were uniformly generated using standard protocols (see Methods, Additional file 2: Fig.
141 S1).
142
143 **Extensive MS-dependent gene expression differences in B cells after EBV**
144 **transformation, which can be accounted for by *EBNA2* expression levels**
145 We performed bulk RNA-seq experiments in primary B cells and EBV-transformed B cells
146 from patients with MS and matched controls (see Methods, Additional file 1: Table S2 and
147 Table S3). We observed very few transcriptional differences in the primary B cells of
148 individuals with MS and matched controls (19 total genes) (Fig. 1A). However, in EBV-
149 transformed B cells, we observed 1,060 differentially expressed genes between MS and
150 HC (Fig. 1B). Based on the established role of EBNA2 in the genetic etiology of MS [18,
151 20-24], we tested the hypothesis that *EBNA2* expression levels, a proxy for EBNA2
152 activity, might account for these differences. Indeed, *EBNA2* expression levels across cell
153 lines accounted for virtually all MS-dependent differential expression (Fig. 1C). Despite
154 other EBV genes also being dysregulated in MS, no other EBV gene (e.g., *EBNA1*, *EBNA-*
155 *LP*) could account for as much differential expression as *EBNA2* (Additional file 2: Fig.
156 S2). Among the altered human genes, EBV-induced molecule 2 (also known as *GPR183*)
157 is a chemotactic receptor that regulates B cell migration [25]. *GPR183* is upregulated in
158 MS EBV-transformed B cells, and *GPR183* expression levels can be explained by *EBNA2*
159 expression levels (Additional file 2: Fig. S3). *ACKR3,* another gene that is important for B
160 cell chemotaxis, is the most significantly upregulated gene in MS EBV-transformed B
161 cells, and this can also be accounted for by *EBNA2* expression levels (Additional file 2:
162 Fig. S3). CD40 is a central immune costimulatory receptor that promotes antigen

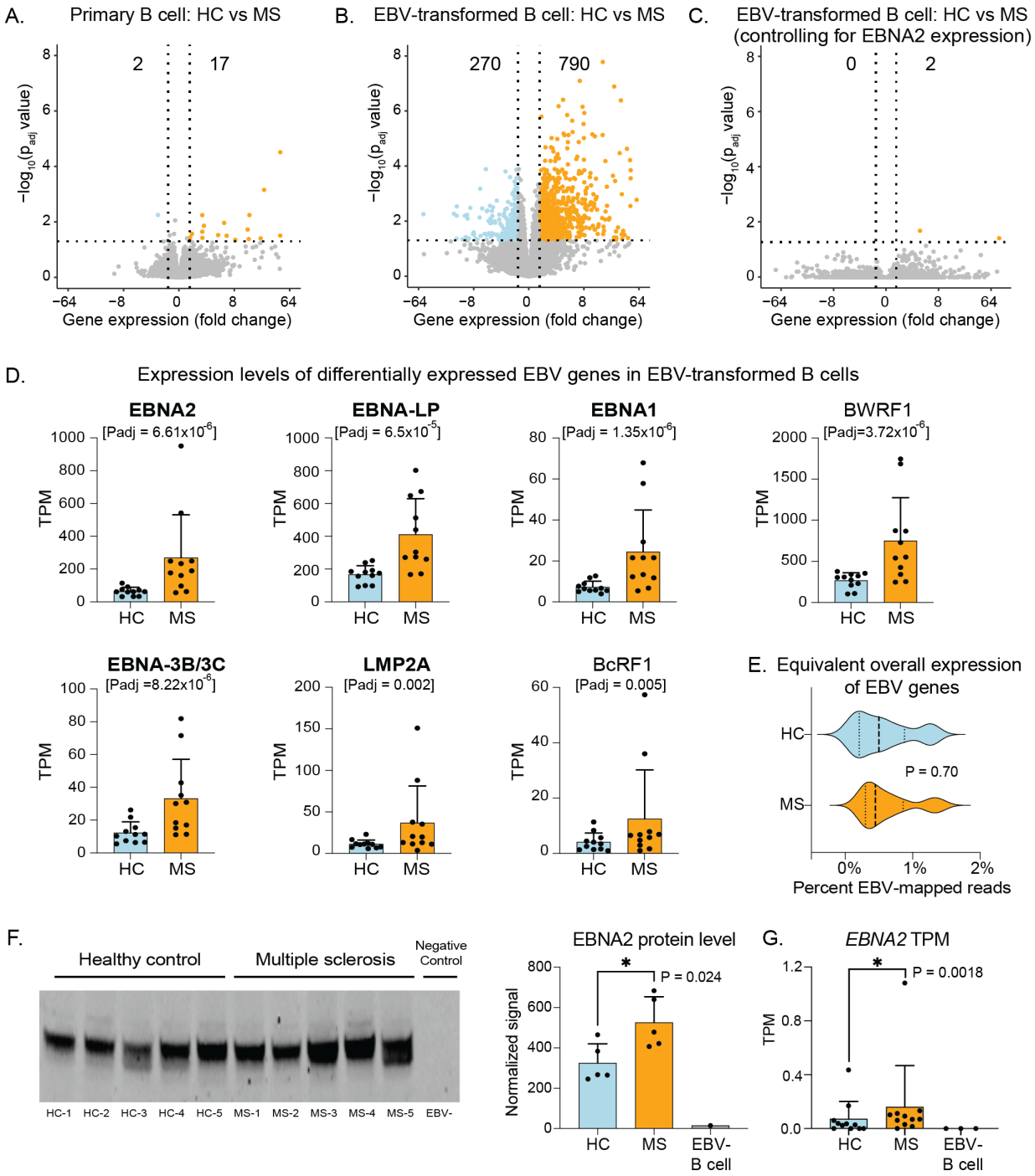


Figure 1. *EBNA2* expression accounts for extensive MS-dependent gene expression differences in B cells after EBV transformation. A-C. Comparison of human gene expression levels (RNA-seq) in healthy controls (HC) and patients with multiple sclerosis (MS). Horizontal dashed lines indicate an adjusted 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 Additional file 2: 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.

presentation and lymphocyte activation, is implicated in MS neuroinflammation, and is currently being targeted therapeutically to reduce disease activity [26]. 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 (Additional file 2: 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, Additional file 2: 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.

182 To rule out the possibility that differential prior exposure to EBV might play a role
183 in the observed gene expression differences, we examined our cohort for evidence of
184 prior EBV infection through the detection of EBV-derived RNA or DNA, and through
185 serological evidence of immune responses to EBV. As expected, based on worldwide
186 EBV infection rates (~95%), everyone in our cohort had evidence of prior infection by EBV
187 (Additional file 1: Table S1, Additional file 2: Fig. S5).

188 We next hypothesized that MS-dependent differences in the abundance of
189 individual B cells that were initially transformed by EBV could drive the observed
190 differential expression (Fig. 1B). To test this hypothesis, we performed single cell RNA-
191 sequencing with B cell receptor (BCR) enrichment in cells obtained from 5 patients with
192 multiple sclerosis and 5 healthy controls (Additional file 2: Fig. S6). To benchmark our
193 single cell dataset against prior studies, we assessed the differential expression of the
194 BCR subunit immunoglobulin heavy chain. Consistent with previous findings in peripheral
195 B cells [27], *IGHV3-74* is the most upregulated immunoglobulin gene in MS (Additional
196 file 2: Fig. S6). We then asked if differences in B cell clonality might account for the
197 observed gene expression differences. Because each B cell's BCR is uniquely
198 rearranged, we used unique BCR sequences as a method to measure the clonality of
199 each cell line derived from HC and MS. Based on these analyses, we did not find evidence
200 of differences in the clonality of cell lines between HC and MS after EBV transformation
201 (Additional file 2: Fig. S6).

202 Next, we tested whether genetic risk might account for some of the observed gene
203 expression differences. To this end, we performed whole genome sequencing to enable
204 imputation of the strongest genetic risk locus (HLA-DRB1*15:01) and calculation of

205 polygenic risk scores (PRS). The distribution of HLA-DRB1*15:01 alleles was virtually
206 identical across HC and MS in our cohort (Additional file 2: Fig. S7), which is consistent
207 with the effect size and the statistical power requirements to identify the original genetic
208 association [28]. We next calculated MS PRS for each member of our cohort to measure
209 the cumulative genetic burden of risk in each person. As expected, PRS was higher on
210 average in the individuals with MS compared to HC (Additional file 2: Fig. S7).
211 Furthermore, PRS accounted for 16.4% of differential human gene expression,
212 emphasizing the role of genetics and environment in the observed differences of MS-
213 derived EBV-transformed B cells compared to controls (Figure 1).

214 We next sought to further examine the small number of differences in primary B
215 cells between patients with MS and controls. These differences could explain the
216 subsequent large number of transcriptional differences after transformation with EBV (Fig.
217 1B). EBV can only transform B cells that express CD21. We therefore examined a
218 published single cell RNA-seq dataset from an independent cohort of 3 HCs and 10
219 people with MS [29] and found that transitional and naïve B cells express the highest
220 amount of *CD21* among B cell subsets, as expected (Additional file 2: Fig. S8). We found
221 no differences in the proportion of B cells within the CD21+ transitional and naïve B cell
222 compartments between HC and MS (Additional file 2: Fig. S8). We identified ten genes
223 that are differentially expressed in MS relative to controls within the CD21+
224 transitional/naïve B cell compartment, including *BACH2, AFF3,* and *BANK1*, which have
225 all previously been implicated in MS (Additional file 2: Fig. S8; Additional file 1: Table S4)
226 [30-33]. *FANCC* is upregulated in MS CD19+ primary B cells (Additional file 1: Table S3).
227 A CRISPR screen demonstrated that *FANCC* expression is essential for EBNA2-

228 dependent transformation of primary B cells to EBV-transformed cell lines [34]. This gene
229 also demonstrated increased expression in CD21+ transitional/naïve B cells in the
230 independent set of HC and patients with MS for which we have single cell RNA-seq data,
231 which is trending toward significance (5.2-fold increase, p-value = 0.17). Taken together,
232 these data support a model in which primary B cells from patients with MS have higher
233 expression of *FANCC*, a gene critical for EBV-dependent immortalization of B cells, along
234 with several other genes with prior links to MS and B cell biology. Strikingly, while the
235 primary B cells from patients with MS and controls had equivalent levels of RNA mapped
236 to the EBV genome (Additional file 2: Fig. S5), primary B cells from patients with MS had
237 significantly higher expression of EBNA2 compared to controls (Mann-Whitney U test, p
238 = 0.0018) (Fig. 1G). The identification of enhanced EBNA2 expression in the primary B
239 cells before transformation for this study suggests a fundamental difference in the B cell
240 compartment of patients with MS relative to controls, a finding consistent with prior studies
241 [29, 35, 36].
242

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

245 Gene expression differences are often driven by differences in chromatin accessibility.
246 To identify chromatin accessibility differences between people with MS and healthy
247 controls, we performed ATAC-seq in primary and EBV-transformed B cells in six HC and
248 six individuals with MS. The resulting data met our stringent quality control criteria, which
249 are based on the ENCODE consortium standards [37, 38] (Additional file 1: Table S2).
250 Consistent with the gene expression analysis, we identified only minimal differences in

251 chromatin accessibility in the primary B cells of HC compared to MS (Fig. 2A, Additional
252 file 1: Table S5). Also consistent with the gene expression analysis, transformation with
253 EBV resulted in large differences between MS and HC, with 1,197 peaks that were
254 stronger in MS and 287 that were stronger in HC (Fig. 2B, Additional file 1: Table S5).
255 Accounting for *EBNA2* expression levels again removed the vast majority of these
256 disease-dependent differences (Fig. 2C).

257 We next examined the concordance between chromatin accessibility changes and
258 gene expression changes. To this end, we calculated the statistical significance of the
259 overlap between differentially expressed genes and differentially accessible chromatin
260 using our RELI algorithm [18] (see Methods, Additional file 1: Table S6). Importantly,
261 stronger chromatin accessibility in EBV-transformed B cells from patients with MS
262 corresponded with higher gene expression in these cells (Fig. 2D, right bar). This
263 relationship was not observed in any of the other possible pairwise relationships; for
264 example, we did not observe correspondence between closed chromatin and activated
265 gene expression (Fig. 2D, other bars). As EBNA2 does not directly bind DNA, we also
266 calculated the enrichment of the DNA binding motifs of its TF interaction partners RBPJ,
267 EBF1, and PU.1 (encoded by *SPI1*) in differentially accessible chromatin. The motifs for
268 all three TFs were significantly enriched in chromatin with enhanced accessibility in MS,
269 but not in chromatin that is more accessible in HC (Fig. 2E), supporting a model where
270 increased EBNA2 levels result in increased human gene expression levels through
271 increased chromatin accessibility at regions bound by EBNA2 with its human TF
272 interaction partners.
273

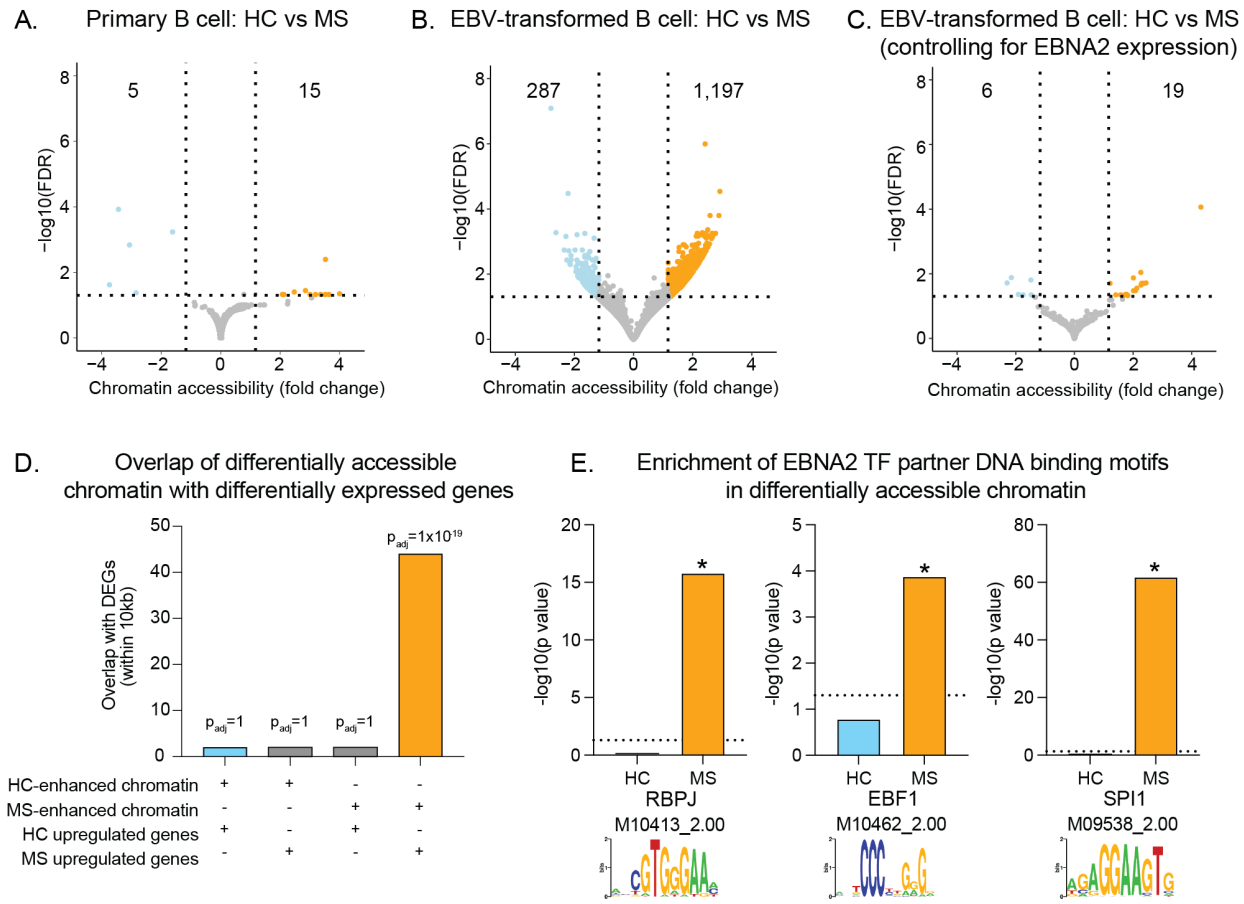


Figure 2. *EBNA2* expression accounts for MS-dependent chromatin accessibility differences in EBV-transformed B cells. A-C. Comparison of chromatin accessibility (ATAC-seq) in healthy controls (HC) and patients with multiple sclerosis (MS). Horizontal dashed lines indicate an adjusted p-value significance 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.

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

278 EBNA2 and its TF interaction partners in these differences, we used ChIP-seq to measure
279 the genome-wide binding of EBNA2, RBPJ, EBF1, and SPI1 (PU.1) in five HC and five
280 MS EBV-transformed B cell lines obtained from cohort members with corresponding gene
281 expression and chromatin accessibility measurements. The resulting data met our
282 stringent quality criteria that were derived from the ENCODE consortium [37, 38]
283 (Additional file 1: Table S2).

284 We observed thousands of robust ChIP-seq peaks, with an elevated average peak
285 count in MS compared to controls for all four proteins (Fig. 3A). In cells from patients with
286 MS, EBNA2 binding significantly overlapped genes with increased expression in MS-
287 derived EBV transformed B cells (Fig. 3B), binding within 10kb of the TSS for over 25%
288 of genes with MS-dependent expression (Additional file 1: Table S6). Likewise, the three
289 EBNA2 TF partners also demonstrated significant overlap with genes whose expression
290 increased in MS-derived cell lines (Fig. 3B). When accounting for co-occurrence of
291 EBNA2 and human TF ChIP-seq peaks in cells from the same individual with MS, we
292 again identify the strongest enrichment of overlap of EBNA2-TF pairs at genes with
293 increased expression in MS-derived cells (Fig. 3C).

294 Next, we compared chromatin accessibility changes with the binding of EBNA2
295 and identified statistically significant overlap in cells derived from patients with MS and
296 MS enhanced chromatin (Fig. 3D). We also observed strong statistical enrichment for the
297 co-occupancy of EBNA2 with TF partners RBPJ, EBF1, and SPI1 in MS-derived cells
298 within MS-enhanced chromatin, but not for chromatin that is more accessible in HC (Fig.
299 3E).

300

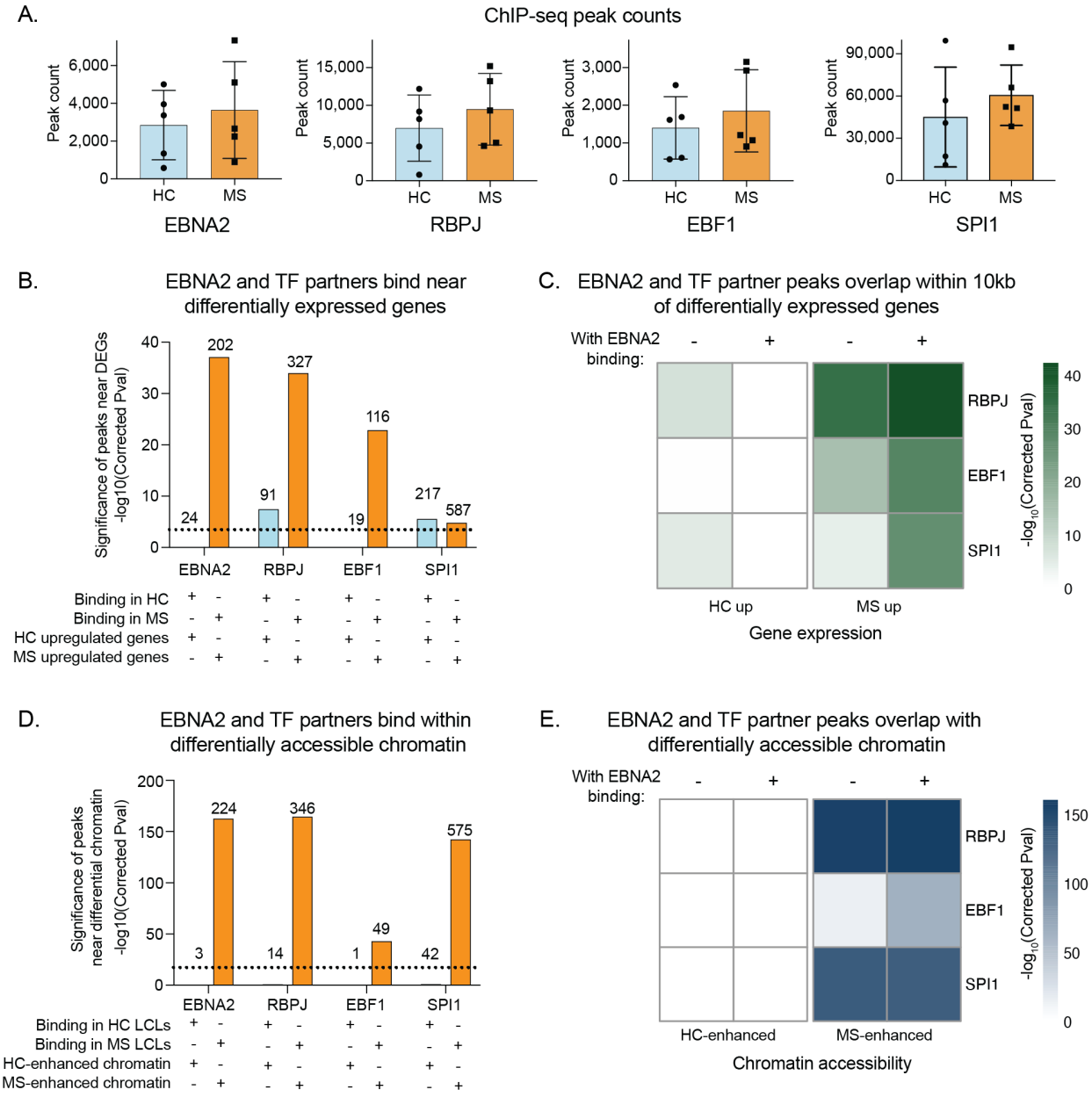


Figure 3. EBNA2 and TF partners bind near genes upregulated in MS EBV-transformed B cells and MS-enriched open chromatin. A. Total number of ChIP-seq peaks for each experiment. Each dot indicates one EBV-transformed cell line from a healthy control (HC) or person with multiple sclerosis (MS). Error bars 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.

302 In summary, upon assessing EBNA2 binding in cell lines derived from HC, the
303 overlap with HC-dependent genes and HC-enhanced chromatin was minimal, with far
304 less significance relative to MS. These results highlight the fact that many more genes
305 are increased in MS (790 genes) relative to HC (270 genes) (Fig. 1B) and many more
306 ATAC-seq peaks are enhanced in MS (1,197) relative to HC (287) (Fig. 2B). Altogether,
307 these results reveal a model of enhanced EBNA2 activity in MS compared to HC, leading
308 to alterations in EBNA2 binding, human TF binding, human chromatin accessibility, and
309 human gene expression.
310

311 **EBNA2 has increased occupancy at MS genetic risk loci**

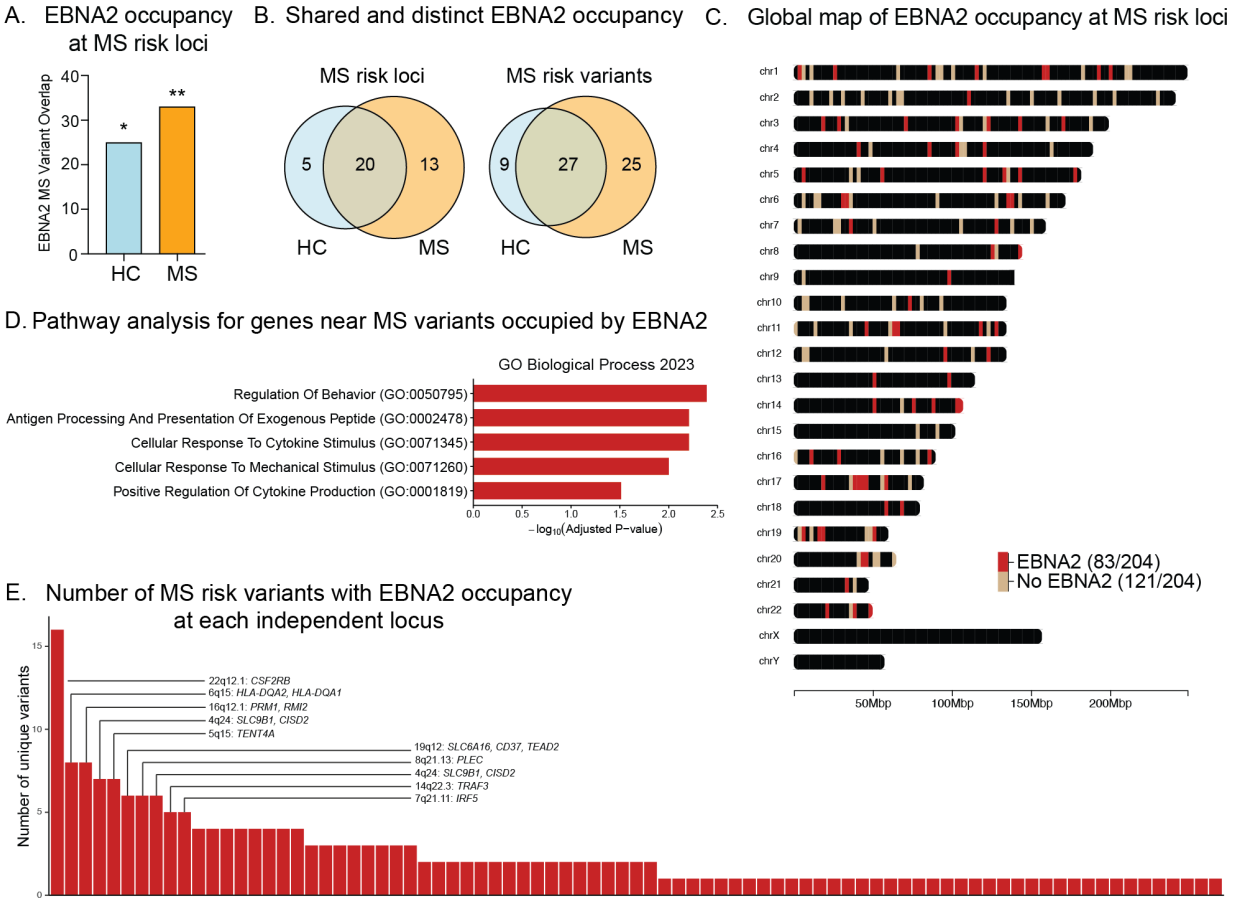
312 We next asked whether the enhanced EBNA2 activity in MS-derived cells preferentially
313 targets known MS genetic risk loci. We previously discovered that EBNA2 occupies 44 of

314 the 109 independent MS-associated genetic loci tested [18]. With the number of
315 established independent genetic risk loci more than doubling in the past few years, and
316 the number of available EBNA2 ChIP-seq datasets likewise increasing [19, 39], we next
317 sought to determine updated EBNA2 occupancy of the expanded set of risk loci in MS-
318 derived cells. Using the RELI algorithm [18] and data from the present study, we find that
319 EBNA2 occupies 33 and 25 MS risk loci in EBV-transformed B cells from patients with
320 MS and HC, respectively (Fig. 4A, Additional file 1: Table S6). Accordingly, the
321 enrichment of EBNA2 at MS risk loci is more significant in MS derived cells (adjusted p-
322 value = 5.81E-14) compared to those from HC (adjusted p-value = 2.63E-09). While most
323 loci and risk variants were shared between HC and MS, many had EBNA2 occupancy
324 only in MS-derived (13 loci; 25 risk variants) or HC-derived (5 loci; 9 risk variants) cells
325 (Fig. 4B).

326 To further expand the set of EBNA2 human genomic targets, we performed ChIP-
327 seq in 21 additional cell lines in a wide range of B cell types and contexts (Additional file
328 1: Table S2). Combining these data with all publicly available EBNA2 genomic binding
329 data, we created an updated global map of EBNA2 occupancy, which reveals EBNA2
330 occupancy at 83 of the 204 known MS risk loci (Fig. 4C). Pathway enrichment analysis
331 reveals that EBNA2 binds at MS risk loci near genes related to pathways involved in key
332 B cell functions such as antigen presentation, response to cytokine, and cytokine
333 production (Fig. 4D, Additional file 1: Table S7). While most EBNA2-bound MS risk loci
334 had a single EBNA2 binding event overlapping an MS variant, a subset of MS risk loci
335 had multiple variants bound by EBNA2, such as the *CSF2RB* locus (n=16 EBNA2-bound

336 variants) and *HLA* and others, implicating EBNA2 binding at 204 unique MS variants

337 within 83 independent MS genetic risk loci (Additional file 1: Table S6) [21].



338

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.

339

340

341 **Genotype-dependent EBNA2 occupancy at MS risk variants**

342 Our analyses indicate that EBNA2 occupancy is significantly enriched at MS risk variants.
343 We next wanted to investigate allele-dependent EBNA2 binding events at these loci. To
344 this end, we used our MARIO pipeline to systematically identify MS risk allele-dependent
345 binding of EBNA2 [18]. In brief, MARIO identifies heterozygous genetic variants within
346 ChIP-seq peaks with significantly different read counts between the two alleles (see
347 Methods). We combined the EBNA2 ChIP-seq experiments produced in this study with
348 publicly available data and identified allele-dependent EBNA2 binding events at MS risk
349 variants. In total, we identified 38 instances of allele-dependent EBNA2 binding to 18
350 unique MS risk variants in EBV-transformed B cells (Fig. 5A, Additional file 1: Table S8).
351 Among these nominated risk variants, rs17035378 and rs8072391, which both have
352 stronger EBNA2 binding to the MS risk allele, are eQTLs in EBV-transformed B cells for
353 *PLEK* and *STAT3*, respectively (Fig. 5B). When all MS risk variants that are bound by
354 EBNA2 in a genotype-dependent manner are connected through eQTL analysis,
355 including those from this study and previously published variants, a network of genes
356 emerges containing a significant number of interactions (STRING-db analysis, p = 8.73
357 e-5, Fig. 5C). From the interaction network, we identify a set of genes involved in innate
358 immune response and a second set with roles in cytokine signaling (Fig. 5C). Together,
359 these analyses reveal a large core set of human genes whose expression is likely altered
360 by allele-dependent EBNA2 binding to MS-associated genetic risk variants, revealing
361 important gene/environment disease mechanisms for potential therapeutic targeting.

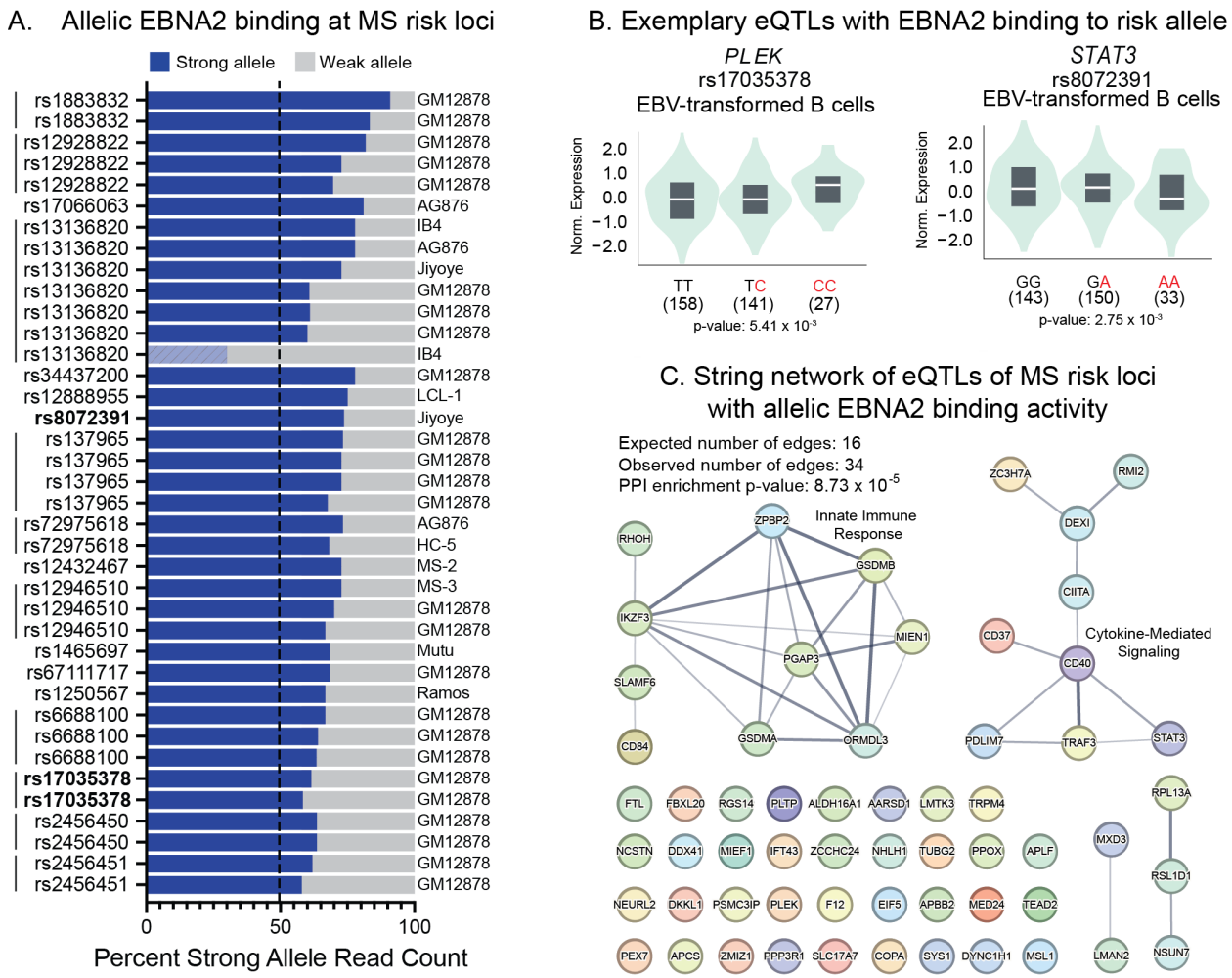


Figure 5. **Allelic EBNA2 occupancy at MS risk variants**. A. MS-associated genetic risk variants with allele-dependent EBNA2 binding. Each variant is heterozygous and located within an EBNA2 ChIP-seq peak to identify allele-dependent occupancy. The x-axis indicates the percent of reads that map to the 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.

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

366 controls, and that this heightened activity amplifies MS-relevant transcriptional and
367 chromatin programs. We observe extensive MS-dependent changes in gene expression
368 and chromatin accessibility that are largely explained by higher EBNA2 expression in MS-
369 derived EBV-transformed B cells (Fig. 1). To explain this phenomenon, we systematically
370 evaluated potential confounders, including B cell clonality, prior EBV exposure, the
371 proportion of CD21+ B cells, overall EBV gene expression, cumulative genetic burden,
372 and HLA-DRB1*15:01 status; none accounted for the MS-dependent gene expression as
373 effectively as *EBNA2* expression levels. Generally, increased EBNA2 expression levels
374 tend to lead to increases in human gene expression levels, affecting genes involved in
375 MS-relevant pathways such as *GPR183* (B cell migration and crosstalk with T cells),
376 *ACKR3* (B cell adhesion and migration), and *CD40* (B cell proliferation and survival). Our
377 MS-derived cells also reveal differential binding of EBNA2, differential chromatin
378 accessibility, and genotype-dependent binding of EBNA2 at MS risk loci.

379 Overall, we observed dramatic differences between primary B cells before and
380 after EBV-transformation. Prior to EBV infection, we observed only 19 human genes with
381 altered expression between MS and controls, with many involved in B cell activation such
382 as *FANCC*. In an independent single-cell dataset, *FANCC* was again higher in CD21+
383 transitional/naïve B cells from patients with MS, and a prior CRISPR screen identified
384 *FANCC* as essential for EBV-mediated B cell transformation [34]. Together, these
385 findings suggest that subtle pre-existing differences in activation programs of CD21+ B
386 cells may prime MS-derived B cells for enhanced EBV transformation and EBNA2-
387 dependent reprogramming.

388 Subsequent to EBV transformation, we observed substantial differences between
389 patients with MS and matched controls on multiple levels: gene expression levels,
390 chromatin accessibility, and human TF binding. EBNA2 is the most significant
391 differentially expressed EBV gene in MS-derived cells (Fig. 1D), EBNA2 protein levels
392 are higher in MS-derived cells (Fig. 1F), and EBNA2 expression level can account for the
393 extensive gene expression differences seen in HC and MS cell lines (Fig 1C.). Occupancy
394 of EBNA2 and its human TF binding partners is highly enriched for human genes with
395 increased chromatin accessibility and increased expression levels in patients with MS
396 (Fig. 3B, 3D). Many of these EBNA2 binding events also coincide with MS risk loci (Fig.
397 4), consistent with prior studies from several groups including our own [18, 19, 21, 39]. In
398 total, EBNA2 occupies 33 and 25 risk loci in EBV-transformed B cells from patients with
399 MS and HC, respectively (Fig. 4A). While the majority of loci were shared between HC
400 and MS, many loci had EBNA2 occupancy only in MS- (13 loci) or HC-derived (5 loci)
401 cells (Fig. 4B). We also identify allelic EBNA2 occupancy at 18 MS risk variants,
402 strengthening our current understanding of MS genotype-dependent biolog. Compared
403 to prior EBNA2 studies that relied on established cell lines, our use of patient-derived
404 EBV-transformed B cells, coupled with allele-dependent analyses, provides direct
405 evidence that EBNA2 engages MS risk loci in a disease- and genotype-dependent
406 manner.

407 Our study uses the B95-8 laboratory strain of EBV to infect primary B cells derived
408 from controls and patients with MS, providing a controlled model of latent EBV infection
409 in human B cells with known genetic background and disease status. Consistent with this
410 model, a recent study in systemic lupus erythematosus demonstrated that EBV

411 preferentially infects autoreactive antinuclear B cells and reprograms them into antigen-
412 presenting cells that activate T peripheral helper cells and propagate pathogenic T and B
413 cell responses [41]. These data provide direct evidence that rare EBV infection events in
414 autoreactive B cells can initiate systemic autoimmunity, aligning with our proposed model
415 for EBV–B cell interactions in MS. They also complement a recent study that used
416 spontaneous lymphoblastoid cell lines to interrogate endogenous EBV in MS-derived B
417 cells, revealing EBV lytic gene expression in active disease compared to both controls
418 and patients with stable disease [35]. Those data, together with prior serologic evidence
419 of EBV reactivation during active MS [42, 43], highlight dysregulated lytic infection. In
420 contrast, our work focuses on latent EBV infection under controlled viral and disease
421 conditions, showing that even in the absence of overt lytic reactivation, heightened
422 EBNA2 activity can remodel chromatin and gene expression at MS risk loci. Together,
423 these studies support a model in which impaired control of both latent and lytic EBV
424 infection in B cells contributes to MS pathogenesis. Importantly, both studies reveal
425 reduced control of EBV B cell infection as a potential driver of multiple sclerosis.

426 The goals of this study are to clarify how interactions between genetic variants and
427 EBNA2 influence MS disease processes, and to define the pathways whose regulation
428 and cell-fate decisions differ in people with MS compared to those without MS. Even
429 common viruses like EBV can contribute to rare non-communicable diseases such as
430 MS. This is likely driven by a combination of host genetics, stochastic events, and
431 environmental factors. In the case of EBV, only a modest minority of people who are
432 infected experience symptoms and develop mononucleosis [44]. Likewise, of the ~95%
433 of people with prior EBV infection, the vast majority will not develop subsequent

434 autoimmune diseases [45]. While up to 50% of this variation may be due to genetics [4],
435 stochasticity also likely plays a role in determining when EBV will lead to MS or other
436 autoimmune diseases. Because only a tiny fraction of B cells are infected with the virus,
437 and the likelihood of those infected B cells possessing a disease-driving BCR is remote
438 [46], it is possible that the disease-inducing BCR and EBV never occur within the same
439 cell to initiate the cascade of EBNA2-induced changes we found. Likewise, environmental
440 factors like vitamin D level and timing of initial EBV infection will also impact the likelihood
441 of progressing to an MS disease state. Additional studies such as this one are needed to
442 explore these triggers further, for EBV and other viruses.
443

444 **Conclusions**

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

454 **Methods:**

455 Ethics statement

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

459

460 Patient recruitment

461 Patients with a diagnosis of MS (n=12) were recruited from the Waddell Center for
462 Multiple Sclerosis at the University of Cincinnati or through a study advertisement sent
463 out to Cincinnati Children's Hospital employees. To reduce heterogeneity, the study
464 inclusion criteria for subjects with MS was restricted to European American women
465 between the ages of 25 and 74, which is representative of the largest MS demographic.
466 Eleven age- and ancestry- matched controls were recruited from a study advertisement
467 sent out to Cincinnati Children's Hospital employees. Additional file 1: Table S1 provides
468 the age, sex, self-reported race, ethnicity, and prior EBV infection status of each subject
469 in this study.

470

471 Primary B cell isolation

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

476

477 ELISA assays

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

481

482 EBV infection of primary B cells

483 B95-8 EBV was prepared from cell supernatants after culture in 10% FBS-supplemented
484 RPMI-1640 medium for 2 weeks. Viral suspension was filtered twice with 0.45 μm
485 Millipore filters. The concentrated viral stocks were stored at −80°C. Primary B cells were
486 infected with 1 mL viral stock based on infection optimization assays and incubated for 3
487 h for virus adsorption. After infection, cells were washed and cultured. Within the first
488 month, we confirmed EBV infection by morphological changes and blasting of the cells.
489 Cells were collected for experiments after at least five passages.

490

491 RNA-seq

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

499

500 RNA-seq data were analyzed using the nf-core/rnaseq pipeline (version 3.11.1) (doi:
501 10.1038/s41587-020-0439-x). Adapter sequence contents were removed using cutadapt
502 [47] (trimgalore version: 0.6.7)
503 (https://journal.embnet.org/index.php/embnetjournal/article/view/200). RNA-seq reads
504 were aligned to the GRCh38 genome augmented with EBV B95-8 coding sequences
505 (ASM240226v1, GCF_002402265.1) using Spliced Transcripts Alignment to a Reference
506 (STAR, version: 2.7.9a) (https://doi.org/10.1093/bioinformatics/bts635). Ribosomal RNA
507 was removed using the Silva Library (https://www.arb-silva.de/) and SortmeRNA v 4.3.4
508 (https://bioinfo.lifl.fr/RNA/sortmerna/). All RNA-seq datasets used in this study passed our
509 quality control checks, which were facilitated by FastQC (version: 0.11.9)
510 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Specifically, RseqQC [48],
511 Qualimap [49, 50], dupRadar [51], and Preseq (https://github.com/smithlabcode/preseq)
512 were used to assess the quality of the datasets. GENCODE (v. 40) [52] annotations for
513 GRCh38 with EBV coding sequence annotations were used to calculate transcript
514 abundance estimates with Salmon (v.1.10.1) (https://doi.org/10.1038/nmeth.4197). The
515 tximport package (https://doi.org/10.12688/f1000research.7563.2) was used to pre-
516 process quantifications and DESeq2 [53] was used to perform differential gene
517 expression analysis. Differentially expressed genes were defined as those genes with a
518 fold-change difference greater than 1.5 and an FDR of less than 0.05. DESeq2 output
519 was filtered to only include genes with transcript per million (TPM) values greater than
520 one in at least three replicates of MS or HC cell lines. Variables accounting for gene
521 expression levels (e.g., the expression level of the EBNA2 gene) were statistically
522 controlled for using the DESeq2 design formula " ~log(gene TPM) + disease_status".

523

524 Whole genome sequencing and variant calling

525 Whole genome sequencing was performed on all MS- and HC-derived cells used in this
526 study. DNA was isolated using a PureLink Genomic DNA Kit (Thermo Fisher #K182001).
527 Libraries were sequenced on an Illumina NovaSeq to generate 100-base paired-end
528 reads. Whole genome sequenced files were mapped to the hg38 genome and variants
529 were called using the Sarek v3.4.1 pipeline [54, 55]. The Sarek pipeline offers multiple
530 programs for alignment, variant calling, and annotation. These samples were aligned with
531 BWA-mem, variants were called with GATK HaplotypeCaller, and annotated with VEP. In
532 addition, joint variant calling was enabled with the "joint_germline" parameter. To
533 determine prior exposure to EBV, the fraction of reads aligning to the EBV genome was
534 assessed by counting the total number of mapped reads in the duplicate marked CRAM
535 file and comparing to the number of reads mapped to the "chrEBV" contig included as a
536 decoy sequence in the hg38 genome.

537

538 Western Blot

539 Cells were lysed and sonicated, and protein concentration was determined using a BCA
540 protein assay. 10 µg of lysates were loaded in 4-12% NuPAGE Bis-Tris gels, run in MOPS
541 buffer, and transferred to a nitrocellulose membrane using the iBlot machine (Thermo
542 Fisher Scientific). The membranes were blocked for 1 hour in Intercept Blocking Buffer at
543 room temperature and incubated overnight at 4°C in Intercept Blocking Buffer + 0.1%
544 Tween 20 with anti-EBNA2 (MABE8, Sigma-Aldrich) antibody diluted to 1:1000. The next
545 day, the membranes were washed three times for 5 minutes with PBS + 0.1% Tween 20.

546 The washed membranes were then incubated with rotation for 45 minutes with fluorescent
547 secondary antibodies (diluted to 1:20,000 in Intercept Blocking Buffer + 0.1% Tween +
548 0.01% SDS). After three more 5-minute washes in PBS + 0.1 Tween 20, the membranes
549 were imaged using an Odyssey DLx Imaging System (Licor Bioscience).
550
551 HLA-DRB-15*01 imputation and Polygenic Risk Score (PRS) calculation
552 The tag variant and allele for HLA-DRB-15*01 is rs3135388 and the MS risk allele is "A"
553 [56-58]. The genotype of rs3135388 was obtained from whole genome sequence data of
554 each individual. We applied a previously developed and validated PRS (PGS002726) for
555 multiple sclerosis to individuals recruited to this study [59]. PGS002726 was obtained
556 from the polygenic risk score catalogue (http://www.pgscatalog.org/). We used the
557 nextflow pipeline "The Polygenic Score Catalog Calculator (pgsc-calc)" to calculate each
558 individual's PRS from whole genome sequencing data [60].
559
560 scRNA-seq
561 We performed scRNA-seq experiments using cells derived from 5 patients with multiple
562 sclerosis and 5 healthy controls, pooled into one batch. Single-cell 5' immune profiling
563 was performed using the Chromium Next GEM Single Cell 5' Kit v2 (Dual Index,
564 PN-1000263) together with the Single Cell Human BCR Amplification Kit (PN-1000253)
565 from 10x Genomics (Pleasanton, California). Approximately 17,000 cells were loaded
566 onto a Chromium Next GEM Chip K to capture 10,000 cells and processed on a
567 Chromium X Controller to generate barcoded Gel Beads-in-Emulsion (GEMs). GEM-RT,
568 Dynabeads MyOne SILANE cleanup, and cDNA amplification were performed according

569 to the manufacturer's protocol (CG000331 Rev F). Amplified cDNA was used for both
570 gene expression and BCR V(D)J enrichment and subsequent library construction
571 following the procedures described in the user guide. Final libraries were sequenced on
572 an Illumina NovaSeq 6000 S4 flow cell (PE 100 bp) using the following sequencing
573 parameters: Read 1: 28 cycles; i7 Index: 10 cycles; i5 Index: 10 cycles; Read 2: 90 cycles.
574 Sequencing was performed to a target depth of ~32,500 read pairs per cell for 5' gene
575 expression libraries and ~5,000 read pairs per cell for the BCR libraries.
576
577 Initial read alignment and quantification of FASTQ files was performed using Cell Ranger
578 (v 6.1.2) [61]. Analysis was performed using Seurat [62]. Cells with less than 1,000 or
579 greater than 7,000 unique features were removed from downstream analysis. No features
580 that were expressed in 3 or fewer cells were included. Data were normalized using the
581 NormalizeData() command. For clustering and Uniform Manifold Approximation and
582 Projection (UMAP) visualization, 15 components were used; this number was determined
583 based on the strength of PCs visualized by elbow plot and PC heatmaps after PCA was
584 performed. The UMAPs and clusters were generated using the FindNeighbors(),
585 FindClusters(), and RunUMAP() functions (R implementation). The DotPlot was created
586 using the DotPlot() function. Input features for the dot plot were generated through the
587 FindAllMarkers() function provided by Seurat to identify differentially expressed
588 immunoglobulin heavy chain variable region (IGHV) reads.
589
590 We also analyzed publicly available single cell RNA-seq data. The dataset GSE133028
591 was downloaded from the Gene Expression Omnibus, consisting of 3 healthy control and

592 10 relapsing-remitting multiple sclerosis scRNA samples. Raw fastq files for each patient
593 were aligned to the human genome using Cell Ranger (v. 7.0.1) to make a Seurat object.
594 Cells with greater than 15% mitochondrial genes and cells with less than 500 or greater
595 than 4,500 unique features were removed from analysis. All individual Seurat objects
596 were integrated using Harmony [63], and B cells were identified using the gene *MS4A1*
597 (otherwise known as CD20) and made into a separate Seurat object. 40 components
598 were used to generate UMAP visualization using the FindNeighbors(), FindClusters(), and
599 RunUMAP() functions. B cell subtypes were identified by comparing the expression levels
600 of select genes based on previously published studies [64, 65]. *CD21* expression across
601 B cell subtypes was analyzed using the VlnPlot() function in Seurat, and the proportion of
602 CD21+ transitional/naïve B cells were extracted to compare between healthy control and
603 MS. Differential gene expression analysis between healthy control and MS was
604 performed using the FindMarkers() command in Seurat (p-value cutoff = 0.05, logFC
605 cutoff = 0.585).
606
607 <u>ATAC-seq</u>
608 Transposase Tn5 with adapter sequences was used to cut accessible DNA [66]. These
609 accessible DNA sequences with adapter sequences were isolated, and libraries were
610 prepared from ~50,000 primary B cells or EBV-transformed B cells using the OMNI ATAC
611 protocol [67]. The libraries were sequenced 2x150 bp on an Illumina NovaSeq 6000 at
612 the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing
613 Facility.
614

615 ATAC-seq data were processed and aligned to a combined human (hg38) and EBV (B95-
616 8) genome using the ENCODE ATAC-seq pipeline (V2.0.0) [37, 38, 68] (Doi:
617 10.5281/zenodo.5598324). Aligned reads were then sorted using samtools (v.1.9) and
618 duplicate reads were removed using Picard (v. 2.20.7)
619 (https://broadinstitute.github.io/picard/). Peaks were called within the pipeline using
620 MACS2 (v. 2.2.4) [69]. ENCODE's exclusion list regions were removed from the final set
621 of peaks. Differential chromatin accessibility was calculated using DiffBind (v. 3.0.15) [70],
622 with thresholds of a fold-change greater than 1.5 and adjusted p-value less than 0.05.
623 Each individual-derived cell line was used as a replicate. Expression levels of EBNA2
624 were statistically controlled for using the DESeq2 design formula "~log(EBNA2 TPM) +
625 disease_status".
626
627 ChIP-seq
628 We performed ChIP-seq in EBV-transformed B cells from 5 MS and 5 HC for EBNA2
629 (Abcam ab90543; Lot GR3185425-6), RBPJ (Cell Signaling Technology 5313; Lot 3),
630 PU.1 (Cell Signaling Technology 2266; Lot 3), and EBF1 (Santa Cruz sc-137065; Lot
631 J1122) using standard experimental procedures as previously described [19]. Cells were
632 incubated in a crosslinking solution (1% formaldehyde, 5 mM 4-(2-hydroxyethyl)-1-
633 piperazineethanesulfonic acid (HEPES) pH 8.0, 10 mM sodium chloride, 0.1 mM
634 ethylenediaminetetraacetic acid (EDTA), and 0.05 mM ethylene glycol tetraacetic acid
635 (EGTA)) in Roswell Park Memorial Institute (RPMI) culture medium with 10% fetal bovine
636 serum (FBS) and placed on a tube rotator at room temperature for 10 min. To stop
637 crosslinking, glycine was added to a final concentration of 0.125 M, and the tubes were

638 rotated at room temperature for 5 min. Cells were washed twice with ice-cold phosphate-
639 buffered saline (PBS), resuspended in lysis buffer 1 (50 mM HEPES pH 8.0, 140 mM
640 NaCl, 1 mM EDTA, 10% glycerol, 0.25% Triton X-100, and 0.5% NP-40), and incubated
641 for 10 min on ice. Nuclei were harvested after centrifugation at 5,000 rpm for 10 min,
642 resuspended in lysis buffer 2 (10 mM Tris–HCl pH 8.0, 1 mM EDTA, 200 mM NaCl, and
643 0.5 mM EGTA), and incubated at room temperature for 10 min. Protease and
644 phosphatase inhibitors (Halt™ Protease and Phosphatase Inhibitor Cocktail (100X),
645 Thermo Fisher Scientific, Waltham, MA) were included in both lysis buffers. Nuclei were
646 resuspended in sonication buffer (10 mM Tris [pH 8.0], 1 mM EDTA, and 0.1% sodium
647 dodecyl sulfate (SDS)). An S220 focused ultrasonicator (COVARIS, Woburn, MA) was
648 used to shear chromatin (150–500-bp fragments) with 10% duty cycle, 175 peak power,
649 and 200 bursts per cycle for 7 min. A portion of the sonicated chromatin was run on an
650 agarose gel to verify fragment sizes. Sheared chromatin was pre-cleared with 10 μL of
651 Protein A or G Dynabeads (Thermo Fisher Scientific) at 4 °C for 1 h.
652
653 Immunoprecipitation of TF-chromatin complexes was performed with an SX-8X IP-STAR
654 compact automated system (Diagenode). Beads conjugated to antibodies against each
655 TF were incubated with pre-cleared chromatin at 4 °C for 8 h. The beads were then
656 washed sequentially with wash buffer 1 (10 mM Tris–HCl [pH 7.5], 150 mM NaCl, 1 mM
657 EDTA, 0.1% SDS, 0.1% NaDOC, and 1% Triton X-100), wash buffer 2 (10 mM Tris–HCl
658 [pH 7.6], 400 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% NaDOC, and 1% Triton X-100),
659 wash buffer 3 (10 mM Tris–HCl [pH 8.0], 250 mM LiCl, 1 mM EDTA, 0.5% NaDOC, and
660 0.5% NP-40), and wash buffer 4 (10 mM Tris–HCl [pH 8.0], 1 mM EDTA, and 0.2% Triton

661 X-100). Finally, the beads were resuspended in 10 mM Tris–HCl (pH 7.5) and used to
662 prepare libraries via ChIPmentation.
663
664 The ChIP-seq libraries were sequenced as single-end 100-base reads on an Illumina
665 NovaSeq 6000 at the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics
666 Sequencing Facility. ChIP-seq data were processed and aligned to a combined human
667 (hg38) and EBV (B-958) genome using the ENCODE ChIP-seq pipeline (V2.0.0) [37, 38,
668 68] (Doi: 10.5281/zenodo.5598352). Adapters were trimmed using fastp (v. 0.23.2) [71]
669 prior to processing through the pipeline. Aligned reads were then sorted using samtools
670 (v.1.9) and duplicate reads were removed using Picard (v. 2.20.7)
671 (https://broadinstitute.github.io/picard/). Peaks were called using the pipeline's default
672 parameters with MACS2 (v. 2.2.4) [69]. ENCODE "exclusion list regions" were removed
673 from the final peak set. QC results are available in Additional file 1: Table S2.
674
675 To identify HC- and MS- specific genomic features for ChIP-seq, all peaks that were found
676 in each condition were combined and overlapping events were merged using bedtools
677 [72]. Peaks unique to HC or MS were identified using the bedtools command "subtract -
678 A". Shared peaks were determined by removing all condition-specific peaks from the total
679 peak set.
680
681 <u>Enrichment analysis for functional genomic datasets</u>
682 We used the RELI algorithm to estimate the significance of the overlap between the
683 genomic feature sets (e.g., ChIP-seq peaks) generated in this study [18]. 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 [73] and the supplemental dataset from the International Multiple Sclerosis Genetics Consortium GWAS publication [4]. 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 [74] (window size 300,000 kb, SNP shift size 100,000 kb, and $r^2 < 0.2$).

Pathway enrichment analysis

706 Pathway analysis for MS and HC differentially expressed genes was performed using
707 Enrichr [75, 76].
708
709 TF DNA binding motif enrichment analysis
710 The HOMER software package (v. 4.9) [77] was used to calculate TF DNA binding site
711 motif enrichment for a custom library containing the human CisBP database motifs (build
712 2.0) [78, 79]. HOMER was modified to use a $\log_2$-based likelihood scoring system.
713
714 Identification of allele-dependent sequencing reads using MARIO
715 Allele-dependent behavior was identified in sequencing reads using the MARIO pipeline
716 [18]. Briefly, the MARIO pipeline identifies allele-dependent behavior by weighing (1) the
717 imbalance between the number of reads that are mapped to each allele, (2) the total
718 number of reads mapped at each variant, and (3) the number and consistency of available
719 replicates. These variables are combined into a single Allelic Reproducibility Score (ARS),
720 which reflects the degree of allelic behavior observed for the given heterozygous variant
721 in the given data set. MARIO ARS values exceeding 0.4 were considered to be allelic,
722 following our previous study [18]. We implemented the MARIO pipeline for EBNA2 ChIP-
723 seq data from this study along with publicly available EBNA2 binding data [19, 40]. To
724 select for consistent allele-dependent binding activity, we identified and graphed ChIP-
725 seq peaks with at least 9 reads on top of an MS variant and with at least 15% difference
726 between the reads mapping to the strong and weak alleles (Fig. 5A, Additional file 1:
727 Table S8).
728

729 To expand our reach, we assessed additional EBNA2 ChIP-seq experiments performed
730 in our laboratory, including the cell lines GM12878, IB4, AG876, and Jiyoye. Additionally,
731 we included MS risk variants with published allelic EBNA2 binding activity (Additional file
732 1: Table S8) [18, 19]. All identified MS risk variants that were overlapped by EBNA2 ChIP-
733 seq peaks in an allelic manner were assessed for eQTL function using release 7 of the
734 eQTL Catalogue [80]. Gene targets of these eQTLs (p-value < 0.01) were used as input
735 to STRING [81] (Fig. 5C).
736
737 **Figure Legends**
738 **Figure 1. *EBNA2* expression accounts for extensive MS-dependent gene**
739 **expression differences in B cells after EBV transformation.** A-C. Comparison of
740 human gene expression levels (RNA-seq) in healthy controls (HC) and patients with
741 multiple sclerosis (MS). Horizontal dashed lines indicate an adjusted p-value significance
742 threshold of 0.05. Vertical dashed lines indicate a fold-change threshold of 1.5. Results
743 are shown for primary B cells (A), EBV- transformed B cells (B), and EBV- transformed B
744 cells after statistically adjusting for *EBNA2* expression levels (C). D. All significant
745 differentially expressed EBV genes in EBV-transformed B cells from HC and patients with
746 MS. Gene expression (RNA-seq) is shown in transcripts per million (TPM). Each dot
747 represents a cell line derived from one individual. Error bars indicate standard deviation.
748 EBV genes in bold are expressed during latency III. E. Percent of RNA sequencing reads
749 that mapped to the EBV genome in EBV-transformed B cells derived from HC and MS.
750 Vertical dashed lines denote quartiles. F. Western blot for EBNA2 protein in HC and MS
751 cell lines (left) and quantification of EBNA2 protein levels (right). For full blots of EBNA2

752 and total protein stains, see Additional file 2: Fig. S4. G. *EBNA2* gene expression levels
753 (RNA-seq) in primary B cells in transcripts per million (TPM). Each dot represents primary
754 B cells isolated from one individual. A B cell line without EBV was used as a negative
755 control. Error bars indicate standard deviation.
756
757 **Figure 2. *EBNA2* expression accounts for MS-dependent chromatin accessibility**
758 **differences in EBV-transformed B cells.** A-C. Comparison of chromatin accessibility
759 (ATAC-seq) in healthy controls (HC) and patients with multiple sclerosis (MS). Horizontal
760 dashed lines indicate an adjusted p-value significance threshold of 0.05. Vertical dashed
761 lines indicate a fold-change threshold of 1.5. Results are shown for primary B cells (A),
762 EBV- transformed B cells (B), and EBV- transformed B cells after statistically adjusting
763 for *EBNA2* expression levels (C). D. Overlap of all possible pairs of differentially
764 accessible chromatin (ATAC-seq) and differentially expressed genes with 10 kb padding
765 on either side of the TSS (RNA-seq). Each set of chromatin and differential genes were
766 either higher in HC or in MS. The statistical enrichment of the overlap (measured by RELI,
767 see Methods) is indicated as an adjusted p-value above each bar. E. DNA binding motif
768 enrichment of exemplary motifs for EBNA2 TF partners RBPJ, EBF1, and SPI1 (PU.1)
769 within accessible chromatin. Enrichment is calculated as % foreground motif occurrence
770 divided by % background motif occurrence (i.e., the percent of random sequences,
771 matching GC content). Horizontal dashed lines indicate significant motif enrichment
772 ($P < 0.05$), as calculated by HOMER.
773

774 **Figure 3. EBNA2 and TF partners bind near genes upregulated in MS EBV-**
775 **transformed B cells and MS-enriched open chromatin.** A. Total number of ChIP-seq
776 peaks for each experiment. Each dot indicates one EBV-transformed cell line from a
777 healthy control (HC) or person with multiple sclerosis (MS). Error bars indicate standard
778 deviation. B. Overlap of all possible pairs of EBNA2, RBPJ, EBF1, or SPI1 (ChIP-seq)
779 and differentially expressed genes with 10 kb padding on either side of the gene (RNA-
780 seq). A combined set of ChIP-seq peaks found in any of the HC or MS-derived EBV-
781 transformed B cells was used for each protein. Differentially expressed genes were higher
782 in either HC or MS-derived EBV-transformed B cells. The statistical enrichment of the
783 overlap (measured by RELI, see Methods) is indicated as an adjusted p-value on the y-
784 axis. Dotted line indicates significance threshold. The number of overlaps is provided
785 above each bar. C. Binding co-occurrence of EBNA2 with each human TF (RBPJ, EBF1,
786 or SPI1) was calculated within each EBV-transformed B cell line derived from HC and
787 individuals with MS, using ChIP-seq peaks obtained from the same cell line (i.e., the same
788 individual). For each EBNA2–TF pair, a single, pooled set of co-occurrence sites was
789 compiled by taking the union of co-occurring peaks observed in any HC- or MS-derived
790 cell line. These co-occurrence peak sets were tested for enrichment near differentially
791 expressed genes (genes upregulated in HC or upregulated in MS). The heatmap
792 indicates the significance of each enrichment test, quantified by RELI, for each EBNA2–
793 TF pair and gene set comparison. D. Overlap of all possible pairs of EBNA2, RBPJ, EBF1,
794 or SPI1 (ChIP-seq) and differentially accessible chromatin (ATAC-seq). A combined set
795 of ChIP-seq peaks found in any of the HC or MS-derived EBV-transformed B cells was
796 used for each protein. “Enhanced chromatin” was significantly more accessible in either

797 HC or MS-derived EBV-transformed B cells. The statistical enrichment of the overlap
798 (measured by RELI, see Methods) is indicated as an adjusted p-value on the y-axis. The
799 number of overlaps is provided above each bar. E. Co-occurrence of EBNA2 and either
800 RBPJ, EBF1, or SPI1 was calculated in each EBV-transformed B cell line in both HC and
801 MS. A combined set of co-occurring peaks found in any of the HC or MS-derived cell lines
802 was used. Each set of peaks (ATAC-seq) was significantly more accessible in either HC
803 or MS. The heatmap depicts the significance of the overlap (as measured by RELI) for
804 each pairwise comparison.
805
806 **Figure 4**. **EBNA2 occupancy at MS genetic risk loci.** A. EBNA2 occupancy at MS risk
807 loci in HC and MS EBV-transformed B cells. Count of EBNA2 ChIP-seq peaks
808 overlapping one or more MS risk variants within each independent locus from HC (left) or
809 MS (right). Overlap significance (as calculated by RELI, see Methods) is indicated with
810 asterisks (*, p = 2.63E-09; **, p = 5.81E-14). B. Comparison of EBNA2 MS risk locus
811 occupancy in HC- or MS-derived cells. (left) Number of MS genetic risk loci where an
812 EBNA2 ChIP-seq peak overlaps with one or more MS risk variant. (right) Total number of
813 these genetic variants (i.e., the counts can include multiple variants within a single locus).
814 C. Chromomap depicting EBNA2 occupancy at MS risk loci in EBV-transformed B cells.
815 EBNA2 ChIP-seq peaks were collected from a variety of sources, including this study
816 (see Methods). D. Pathway enrichment (Gene Ontology Biological Process) of genes with
817 TSS within 10kb of MS risk loci bound by EBNA2. E. Number of MS risk variants with
818 EBNA2 occupancy at each independent locus.
819

820 **Figure 5**. **Allelic EBNA2 occupancy at MS risk variants**. A. MS-associated genetic risk
821 variants with allele-dependent EBNA2 binding. Each variant is heterozygous and located
822 within an EBNA2 ChIP-seq peak (which enables allele-dependent calculations). The x-
823 axis indicates the percent of reads that map to the strong allele vs the weak allele. A
824 strong allele is defined as the allele preferred in the majority of experiments for that
825 variant. A single variant had discordant allelic results in one out of seven experiments
826 (indicated in light hatch marks). Variants shown in eQTL plots in panel B are shown in
827 bold. Identical variants in different datasets are connected with vertical bars. B. Genotype-
828 dependent expression of *STAT3* and *PLEK* at MS risk variants with allelic EBNA2 binding
829 (identified in GTeX). MS risk alleles are indicated in red. C. eQTLs at MS variants with
830 allelic EBNA2 binding were identified in the eQTL catalogue. A network of these eQTLs
831 was produced using STRING.
832
833 **Supplemental Figures and Tables**
834 **Additional file 1. Supplementary tables.**
835 **Table S1**. Patient demographic information and clinical phenotypes. This contains the
836 demographics and clinical phenotypes of each individual (Healthy Controls and patients
837 with Multiple Sclerosis).
838 **Table S2**. Functional genomics data quality assessments. This contains the data quality
839 summary report for all ChIP, ATAC, and RNA-seq datasets used.
840 **Table S3.** RNA-seq raw and analyzed data. This contains the TPM and differential
841 analysis for RNA-seq data.

842 **Table S4**. Single cell RNA-seq differential analysis of CD21+ Naïve and Transitional B
843 cells. This contains differential expression analysis of the public single cell RNA-seq
844 dataset GSE133028, particularly in the CD21+ Naïve and Transitional B cell
845 compartment.
846 **Table S5.** ATAC-seq differential analysis. This contains the differential analysis for ATAC-
847 seq data.
848 **Table S6**. Integrated analysis of genomic datasets using RELI. This contains enrichment
849 analyses performed using RELI including all MS risk variants and unique loci that are
850 overlapped by EBNA2.
851 **Table S7.** Pathway analysis of genes proximal to EBNA2-bound MS risk loci. This
852 contains the pathway analysis (Enrichr) for the genes proximal to EBNA2-bound MS risk
853 loci.
854 **Table S8**. Analysis of EBNA2-bound alleles using MARIO. This contains allelic EBNA2
855 binding events derived from our patient cohort, previously published EBNA2 ChIP-seq
856 datasets, and published EBNA2 binding events from previous studies [18, 19].
857
858 **Additional file 2. Supplementary figures.**
859 **Figure S1.** EBV-transformed B cell line generation.
860 **Figure S2**. EBNA2 expression level accounts for the largest amount of differentially
861 expressed genes compared to other EBV genes.
862 **Figure S3.** EBNA2-dependent changes in gene expression.
863 **Figure S4.** Increased EBNA2 protein in MS-derived EBV-transformed B cells.
864 **Figure S5.** Evidence of prior exposure of individuals in this study to EBV.

865 **Figure S6.** Single cell RNA-sequencing with BCR enrichment allows for enumeration of
866 unique clones for each EBV-transformed B cell line.
867 **Figure S7.** Multiple sclerosis genetic risk variants do not fully account for differential gene
868 expression in EBV-transformed B cells derived from HC and MS.
869 **Figure S8.** There are few transcriptional differences in CD21+ transitional/naïve B cells
870 between HC and MS.

References

1. 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: **Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition.** *Mult Scler* 2020, **26:**1816-1821.
2. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, Rocca MA: **Multiple sclerosis.** *Nat Rev Dis Primers* 2018, **4:**43.
3. Olsson T, Barcellos LF, Alfredsson L: **Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis.** *Nat Rev Neurol* 2017, **13:**25-36.
4. International Multiple Sclerosis Genetics C: **Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility.** *Science* 2019, **365**.
5. Hollenbach JA, Oksenberg JR: **The immunogenetics of multiple sclerosis: A comprehensive review.** *J Autoimmun* 2015, **64:**13-25.
6. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al: **Systematic localization of common disease-associated variation in regulatory DNA.** *Science* 2012, **337:**1190-1195.
7. DeLorenze GN, Munger KL, Lennette ET, Orentreich N, Vogelman JH, Ascherio A: **Epstein-Barr virus and multiple sclerosis: evidence of association from a prospective study with long-term follow-up.** *Arch Neurol* 2006, **63:**839-844.
8. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, Ascherio A: **Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis.** *Science* 2022, **375:**296-301.
9. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, Lublin F, Montalban X, Rammohan KW, Selmaj K, et al: **Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis.** *N Engl J Med* 2017, **376:**221-234.
10. Zhao B, Zou J, Wang H, Johannsen E, Peng CW, Quackenbush J, Mar JC, Morton CC, Freedman ML, Blacklow SC, et al: **Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth.** *Proc Natl Acad Sci U S A* 2011, **108:**14902-14907.
11. Thorley-Lawson DA: **EBV Persistence--Introducing the Virus.** *Curr Top Microbiol Immunol* 2015, **390:**151-209.
12. Ressing ME, van Gent M, Gram AM, Hooykaas MJ, Piersma SJ, Wiertz EJ: **Immune Evasion by Epstein-Barr Virus.** *Curr Top Microbiol Immunol* 2015, **391:**355-381.
13. Lanz TV, Brewer RC, Ho PP, Moon JS, Jude KM, Fernandez D, Fernandes RA, Gomez AM, Nadj GS, Bartley CM, et al: **Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM.** *Nature* 2022, **603:**321-327.
14. Cable JM, Grabowski JC, Luftig MA: **EBNA2 and EBNA-LP: The Earliest Viral Latency Proteins.** *Curr Top Microbiol Immunol* 2025.
15. Waltzer L, Logeat F, Brou C, Israel A, Sergeant A, Manet E: **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* 1994, **13:**5633-5638.
16. Johannsen E, Koh E, Mosialos G, Tong X, Kieff E, Grossman SR: **Epstein-Barr virus nuclear protein 2 transactivation of the latent membrane protein 1 promoter is mediated by J kappa and PU.1.** *J Virol* 1995, **69:**253-262.
17. Murata T, Noda C, Narita Y, Watanabe T, Yoshida M, Ashio K, Sato Y, Goshima F, Kanda T, Yoshiyama H, et al: **Induction of Epstein-Barr Virus Oncoprotein LMP1 by Transcription Factors AP-2 and Early B Cell Factor.** *J Virol* 2016, **90:**3873-3889.
18. Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, Magnusen AF, Lynch A, Chetal K, Yukawa M, et al: **Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity.** *Nat Genet* 2018, **50:**699-707.

19. Hong T, Parameswaran S, Donmez OA, Miller D, Forney C, Lape M, Saint Just Ribeiro M, Liang J, Edsall LE, Magnusen AF, et al: **Epstein-Barr virus nuclear antigen 2 extensively rewires the human chromatin landscape at autoimmune risk loci.** *Genome Res* 2021, **31:**2185-2198.
20. Mechelli R, Manzari C, Policano C, Annese A, Picardi E, Umeton R, Fornasiero A, D'Erchia AM, Buscarinu MC, Agliardi C, et al: **Epstein-Barr virus genetic variants are associated with multiple sclerosis.** *Neurology* 2015, **84:**1362-1368.
21. Ricigliano VA, Handel AE, Sandve GK, Annibali V, Ristori G, Mechelli R, Cader MZ, Salvetti M: **EBNA2 binds to genomic intervals associated with multiple sclerosis and overlaps with vitamin D receptor occupancy.** *PLoS One* 2015, **10:**e0119605.
22. Afrasiabi A, Parnell GP, Fewings N, Schibeci SD, Basuki MA, Chandramohan R, Zhou Y, Taylor B, Brown DA, Swaminathan S, et al: **Evidence from genome wide association studies implicates reduced control of Epstein-Barr virus infection in multiple sclerosis susceptibility.** *Genome Med* 2019, **11:**26.
23. Afrasiabi A, Parnell GP, Swaminathan S, Stewart GJ, Booth DR: **The interaction of Multiple Sclerosis risk loci with Epstein-Barr virus phenotypes implicates the virus in pathogenesis.** *Sci Rep* 2020, **10:**193.
24. Keane JT, Afrasiabi A, Schibeci SD, Swaminathan S, Parnell GP, Booth DR: **The interaction of Epstein-Barr virus encoded transcription factor EBNA2 with multiple sclerosis risk loci is dependent on the risk genotype.** *EBioMedicine* 2021, **71:**103572.
25. Gatto D, Brink R: **B cell localization: regulation by EBI2 and its oxysterol ligand.** *Trends Immunol* 2013, **34:**336-341.
26. Vermersch P, Granziera C, Mao-Draayer Y, Cutter G, Kalbus O, Staikov I, Dufek M, Saubadu S, Bejuit R, Truffinet P, et al: **Inhibition of CD40L with Frexalimab in Multiple Sclerosis.** *N Engl J Med* 2024, **390:**589-600.
27. von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, Cree BA, Gourraud PA, Schwartzburg A, Huerta G, et al: **B cell exchange across the blood-brain barrier in multiple sclerosis.** *J Clin Invest* 2012, **122:**4533-4543.
28. Schmidt H, Williamson D, Ashley-Koch A: **HLA-DR15 haplotype and multiple sclerosis: a HuGE review.** *Am J Epidemiol* 2007, **165:**1097-1109.
29. Ramesh A, Schubert RD, Greenfield AL, Dandekar R, Loudermilk R, Sabatino JJ, Jr., Koelzer MT, Tran EB, Koshal K, Kim K, et al: **A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis.** *Proc Natl Acad Sci U S A* 2020, **117:**22932-22943.
30. Bae SC, Lee YH: **Association between BANK1 polymorphisms and susceptibility to autoimmune diseases: A meta-analysis.** *Cell Mol Biol (Noisy-le-grand)* 2017, **63:**29-35.
31. 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: **Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis.** *Nature* 2011, **476:**214-219.
32. Itoh-Nakadai A, Hikota R, Muto A, Kometani K, Watanabe-Matsui M, Sato Y, Kobayashi M, Nakamura A, Miura Y, Yano Y, et al: **The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program.** *Nat Immunol* 2014, **15:**1171-1180.
33. Tsukumo SI, Subramani PG, Seija N, Tabata M, Maekawa Y, Mori Y, Ishifune C, Itoh Y, Ota M, Fujio K, et al: **AFF3, a susceptibility factor for autoimmune diseases, is a molecular facilitator of immunoglobulin class switch recombination.** *Sci Adv* 2022, **8:**eabq0008.
34. Ma Y, Walsh MJ, Bernhardt K, Ashbaugh CW, Trudeau SJ, Ashbaugh IY, Jiang S, Jiang C, Zhao B, Root DE, et al: **CRISPR/Cas9 Screens Reveal Epstein-Barr Virus-**

Transformed B Cell Host Dependency Factors. *Cell Host Microbe* 2017, **21:**580-591 e587.

35. Soldan SS, Su C, Monaco MC, Yoon L, Kannan T, Zankharia U, Patel RJ, Dheekollu J, Vladimirova O, Dowling JW, et al: **Multiple sclerosis patient-derived spontaneous B cells have distinct EBV and host gene expression profiles in active disease.** *Nat Microbiol* 2024, **9:**1540-1554.
36. Sarkkinen J, Yohannes DA, Kreivi N, Durnsteiner P, Elsakova A, Huuhtanen J, Nowlan K, Kurdo G, Linden R, Saarela M, et al: **Altered immune landscape of cervical lymph nodes reveals Epstein-Barr virus signature in multiple sclerosis.** *Sci Immunol* 2025, **10:**eadl3604.
37. Hitz BC, Jin-Wook L, Jolanki O, Kagda MS, Graham K, Sud P, Gabdank I, Strattan JS, Sloan CA, Dreszer T, et al: **The ENCODE Uniform Analysis Pipelines.** *bioRxiv* 2023.
38. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, et al: **New developments on the Encyclopedia of DNA Elements (ENCODE) data portal.** *Nucleic Acids Res* 2020, **48:**D882-D889.
39. Viel K, Parameswaran S, Donmez OA, Forney CR, Hass MR, Yin C, Jones SH, Prosser HK, Diouf AA, Gittens OE, et al: **Shared and distinct interactions of type 1 and type 2 Epstein-Barr Nuclear Antigen 2 with the human genome.** *BMC Genomics* 2024, **25:**273.
40. McClellan MJ, Wood CD, Ojeniyi O, Cooper TJ, Kanhere A, Arvey A, Webb HM, Palermo RD, Harth-Hertle ML, Kempkes B, et al: **Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming.** *PLoS Pathog* 2013, **9:**e1003636.
41. Younis S, Moutusy SI, Rasouli S, Jahanbani S, Pandit M, Wu X, Acharya S, Sharpe O, Wijeratne TU, Harris ML, et al: **Epstein-Barr virus reprograms autoreactive B cells as antigen-presenting cells in systemic lupus erythematosus.** *Sci Transl Med* 2025, **17:**eady0210.
42. Farrell RA, Antony D, Wall GR, Clark DA, Fisniku L, Swanton J, Khaleeli Z, Schmierer K, Miller DH, Giovannoni G: **Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI.** *Neurology* 2009, **73:**32-38.
43. Wandinger K, Jabs W, Siekhaus A, Bubel S, Trillenberg P, Wagner H, Wessel K, Kirchner H, Hennig H: **Association between clinical disease activity and Epstein-Barr virus reactivation in MS.** *Neurology* 2000, **55:**178-184.
44. Crawford DH, Macsween KF, Higgins CD, Thomas R, McAulay K, Williams H, Harrison N, Reid S, Conacher M, Douglas J, Swerdlow AJ: **A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis.** *Clin Infect Dis* 2006, **43:**276-282.
45. Ascherio A, Munger KL: **EBV and Autoimmunity.** *Curr Top Microbiol Immunol* 2015, **390:**365-385.
46. SoRelle ED, Dai J, Reinoso-Vizcaino NM, Barry AP, Chan C, Luftig MA: **Time-resolved transcriptomes reveal diverse B cell fate trajectories in the early response to Epstein-Barr virus infection.** *Cell Rep* 2022, **40:**111286.
47. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *2011* 2011, **17:**3.
48. Wang L, Wang S, Li W: **RSeQC: quality control of RNA-seq experiments.** *Bioinformatics* 2012, **28:**2184-2185.
49. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A: **Qualimap: evaluating next-generation sequencing alignment data.** *Bioinformatics* 2012, **28:**2678-2679.

1020 50. Okonechnikov K, Conesa A, Garcia-Alcalde F: **Qualimap 2: advanced multi-sample**
1021 **quality control for high-throughput sequencing data.** *Bioinformatics* 2016, **32:**292-
1022 294.
1023 51. Sayols S, Scherzinger D, Klein H: **dupRadar: a Bioconductor package for the**
1024 **assessment of PCR artifacts in RNA-Seq data.** *BMC Bioinformatics* 2016, **17:**428.
1025 52. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu
1026 C, Wright J, Armstrong J, et al: **GENCODE reference annotation for the human and**
1027 **mouse genomes.** *Nucleic Acids Res* 2019, **47:**D766-D773.
1028 53. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for**
1029 **RNA-seq data with DESeq2.** *Genome Biol* 2014, **15:**550.
1030 54. Hanssen F, Garcia MU, Folkersen L, Pedersen AS, Lescai F, Jodoin S, Miller E, Seybold
1031 M, Wacker O, Smith N, et al: **Scalable and efficient DNA sequencing analysis on**
1032 **different compute infrastructures aiding variant discovery.** *NAR Genom Bioinform*
1033 2024, **6:**lqae031.
1034 55. Garcia M, Juhos S, Larsson M, Olason PI, Martin M, Eisfeldt J, DiLorenzo S, Sandgren J,
1035 Diaz De Stahl T, Ewels P, et al: **Sarek: A portable workflow for whole-genome**
1036 **sequencing analysis of germline and somatic variants.** *F1000Res* 2020, **9:**63.
1037 56. Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernandez O, Ndagire D,
1038 Catala-Rabasa A, Ruiz A, Gayan J, et al: **Multiple sclerosis risk variant HLA-**
1039 **DRB1*1501 associates with high expression of DRB1 gene in different human**
1040 **populations.** *PLoS One* 2012, **7:**e29819.
1041 57. Goris A, Walton A, Ban M, Dubois B, Compston A, Sawcer S: **A Taqman assay for high-**
1042 **throughput genotyping of the multiple sclerosis-associated HLA-DRB1*1501 allele.**
1043 *Tissue Antigens* 2008, **72:**401-403.
1044 58. Zivkovic M, Stankovic A, Dincic E, Popovic M, Popovic S, Raicevic R, Alavantic D: **The**
1045 **tag SNP for HLA-DRB1*1501, rs3135388, is significantly associated with multiple**
1046 **sclerosis susceptibility: cost-effective high-throughput detection by real-time PCR.**
1047 *Clin Chim Acta* 2009, **406:**27-30.
1048 59. Shams H, Shao X, Santaniello A, Kirkish G, Harroud A, Ma Q, Isobe N, University of
1049 California San Francisco MSET, Schaefer CA, McCauley JL, et al: **Polygenic risk score**
1050 **association with multiple sclerosis susceptibility and phenotype in Europeans.**
1051 *Brain* 2023, **146:**645-656.
1052 60. Lambert SA, Wingfield B, Gibson JT, Gil L, Ramachandran S, Yvon F, Saverimuttu S,
1053 Tinsley E, Lewis E, Ritchie SC, et al: **Enhancing the Polygenic Score Catalog with**
1054 **tools for score calculation and ancestry normalization.** *Nat Genet* 2024, **56:**1989-
1055 1994.
1056 61. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD,
1057 McDermott GP, Zhu J, et al: **Massively parallel digital transcriptional profiling of**
1058 **single cells.** *Nat Commun* 2017, **8:**14049.
1059 62. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell**
1060 **transcriptomic data across different conditions, technologies, and species.** *Nat*
1061 *Biotechnol* 2018, **36:**411-420.
1062 63. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M,
1063 Loh PR, Raychaudhuri S: **Fast, sensitive and accurate integration of single-cell data**
1064 **with Harmony.** *Nat Methods* 2019, **16:**1289-1296.
1065 64. Dai D, Gu S, Han X, Ding H, Jiang Y, Zhang X, Yao C, Hong S, Zhang J, Shen Y, et al:
1066 **The transcription factor ZEB2 drives the formation of age-associated B cells.**
1067 *Science* 2024, **383:**413-421.
1068 65. Verstegen NJM, Pollastro S, Unger PA, Marsman C, Elias G, Jorritsma T, Streutker M,
1069 Bassler K, Haendler K, Rispens T, et al: **Single-cell analysis reveals dynamics of**

human B cell differentiation and identifies novel B and antibody-secreting cell intermediates. *Elife* 2023, **12**.

66. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** *Nat Methods* 2013, **10:**1213-1218.
67. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al: **An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues.** *Nat Methods* 2017, **14:**959-962.
68. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489:**57-74.
69. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9:**R137.
70. Stark R BG: **DiffBind: differential binding analysis of ChIP-Seq peak data.** http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf; 2011.
71. Chen S, Zhou Y, Chen Y, Gu J: **fastp: an ultra-fast all-in-one FASTQ preprocessor.** *Bioinformatics* 2018, **34:**i884-i890.
72. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.
73. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al: **The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.** *Nucleic Acids Res* 2019, **47:**D1005-D1012.
74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:**559-575.
75. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A: **Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.** *BMC Bioinformatics* 2013, **14:**128.
76. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Res* 2016, **44:**W90-97.
77. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell* 2010, **38:**576-589.
78. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al: **Determination and inference of eukaryotic transcription factor sequence specificity.** *Cell* 2014, **158:**1431-1443.
79. Lambert SA, Yang AWH, Sasse A, Cowley G, Albu M, Caddick MX, Morris QD, Weirauch MT, Hughes TR: **Similarity regression predicts evolution of transcription factor sequence specificity.** *Nat Genet* 2019, **51:**981-989.
80. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, Samovica M, Sakthivel MP, Kuzmin I, Trevanion SJ, et al: **A compendium of uniformly processed human gene expression and splicing quantitative trait loci.** *Nat Genet* 2021, **53:**1290-1299.
81. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al: **The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest.** *Nucleic Acids Res* 2023, **51:**D638-D646.

1121 **Acknowledgements**

1122 This research was funded by the National Institutes of Health (NIH) R01 NS099068, P30
1123 AR070549, R01 AR073228, and R01 AI024717 to M.T.W. and L.C.K.; U19 AI070235,
1124 R01 DK107502, R01 AI148276, and U01 HG011172 to L.C.K.; R01 HG010730, R01
1125 GM055479, U01 AI130830, U01 AI150748, R01 AI141569, P01 AI150585, and U24
1126 HG013078 to M.T.W.

1127

1128 We thank the Single Cell Genomics Core (RRID:SCR_022653) at the Cincinnati
1129 Children's Hospital Medical Center, especially Kelly Rangel, for their assistance with
1130 scRNA-seq. We thank the Cincinnati Children's Hospital Medical Center (CCHMC)
1131 Genomics Sequencing Facility (RRID:SCR_022630). We thank John Harley for helpful
1132 discussions around the role of EBNA2 in autoimmunity.

1133

1134 **Author information**

1135 Author notes

1136 Marissa Granitto and Ellie Kim contributed equally to this work. Matthew Weirauch and
1137 Leah Kottyan co-led this work.

1138

1139 **Author contributions**

1140 Conceptualization: M.G., L.C.K., M.T.W.; Study design: M.G., L.C.K., M.T.W., C.F.;
1141 Participant recruitment: C.F., V.B., K.F., A.Z.; Statistical analysis design: M.G., L.C.K,
1142 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.,
1143 K.K., K.F., B.G.; Directed experiments: M.G., C.F., L.C.K., M.T.W.; Performed
1144 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.

Ethics declarations

Ethics approval and consent to participate

All study participants consented to participate in this study.

Consent for publication

Not applicable.

Competing interests

None.