Systematic investigation of Epstein-Barr virus transcriptional regulator interactions with the human genome

Phillip J. Dexheimer[1,2,#], Matthew R. Hass[1,2,#], Lee E. Edsall[1,2,#], Arame A. Diouf[1,2], Sydney H. Jones[1,2], Omer Donmez[1,2], Cailing Yin[1,2], Katelyn A. Dunn[1,2], Carmy Forney[1,2], Hayley K. Hesse[1,2], Andrew VonHandorf[1,2], Xiaoting Chen[1,2], Sreeja Parameswaran[1,2], Olivia E. Gittens[1,2], Kenyatta C. M. F. Viel[1,2], Kevin Ernst[1,2], Rozita Razavi[3], Benjamin E. Gewurz[4], Bo Zhao[4], Lucinda P. Lawson[1,2], Timothy R. Hughes[3], Leah C. Kottyan[1,2,5,*], Matthew T. Weirauch[1,2,5,6,7*]

1. Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
2. Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.
3. Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; Donnelly Centre, University of Toronto, Toronto, ON, Canada.
4. Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA.
5. Department of Pediatrics, 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. Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA.

#These authors contributed equally.

* Corresponding authors: Matthew.Weirauch@cchmc.org. Leah.Kottyan@cchmc.org

Summary

We systematically investigate interactions between Epstein-Barr virus (EBV) transcriptional regulators (vTRs) and the human genome. Starting with 16 known and candidate vTRs, we identify nine whose introduction into human cells results in substantial alterations to human gene expression. Genome-scale determination of vTR genomic binding events and alterations to chromatin accessibility reveals a detailed map of EBV's functional interactions with the human genome, including >100,000 vTR binding events impacting almost a quarter of all human genes. Notably, BMRF1 emerges as a potent regulator, impacting >7,000 genes and altering >37,000 chromatin regions. Our results provide new evidence that EBV RTA interacts with and stabilizes the binding of human RBPJ. Network analysis reveals that many human genes are targeted by multiple EBV vTRs, highlighting the vast coordinated impact of EBV on human gene expression. Collectively, this study provides a valuable resource for examining EBV-induced alterations to human gene regulation, with data available on multiple platforms.

Keywords: Epstein-Barr virus, EBV, transcriptional regulation, functional genomics, systems biology, regulatory networks, virology, virus-host interactions

Introduction

Epstein-Barr virus (EBV) is a near-ubiquitous gammaherpesvirus that plays an established causative role in mononucleosis (1), Burkitt lymphoma (2), Hodgkin lymphoma (3), and nasopharyngeal carcinoma (4). Compelling emerging evidence also implicates EBV in several autoimmune diseases, including multiple sclerosis (5), systemic lupus erythematosus (6, 7), rheumatoid arthritis (8-10), and inflammatory bowel disease (11). Multiple EBV-based mechanisms are known to contribute to disease processes, including transformation of B cells into malignant cells in cancers (12-14), molecular mimicry between EBV-encoded EBNA1 and GlialCAM in multiple sclerosis (15), and altered regulation of human genes by EBV-encoded EBNA2 in multiple autoimmune diseases (16-19). Despite these advances, it is likely that many additional, currently unknown, mechanisms contribute to EBV-based disease pathogenesis.

EBV survives by manipulating host processes on multiple levels (20-22). EBV encodes 16 known or suspected viral transcriptional regulators (vTRs), the second most of the 419 known human viruses (23). For some of these proteins, such as EBNA2 and ZTA (also known as Z, ZEBRA, or BZLF1), their viral and human genome targets have been characterized by the community over decades of studies (16-19, 24-41). Others, such as RTA (also known as R or BRLF1) and BMRF1, are much less well-studied in the context of their interactions with the human genome. Even for the well-studied proteins, differences in assays, protocols, data processing, and cell types make direct comparison of EBV vTR mechanisms a challenge. Further, only seven of the 16 EBV

vTRs currently have published ChIP-seq data mapped to the human genome in any context (16-18, 37, 42-54). To fully understand the complex layers of gene regulation used by EBV to manipulate human gene expression levels, a uniform systems-level approach is therefore needed.

Here, we employed three functional genomics assays aimed at systematically characterizing the gene regulatory targets and mechanisms of known and suspected EBV vTRs (23). We first examined the transcriptional role of each vTR by transiently expressing each vTR one at a time in human cells. After confirming human gene expression impact (RNA-seq), we assessed the effects of stable vTR expression on human gene expression (RNA-seq), genome binding (ChIP-seq) and chromatin accessibility (ATAC-seq). Altogether, our work encompasses 106 transient RNA-seq experiments, 33 stable RNA-seq experiments, 30 ATAC-seq experiments, and 29 ChIP-seq experiments. We also performed nine ChIP-seq experiments for the human-encoded transcription factor RBPJ due to its known interaction with EBNA2 and its critical role in EBV biology (18). Through a variety of approaches, we confirm that our system produces high quality data and is capable of capturing known EBV biology.

The resulting unprecedented dataset reveals a staggering impact of EBV-encoded vTRs on the human genome, with >15,000 significantly altered human gene expression levels, >64,000 unique vTR-occupied regions of the human genome, and >109,000 regions with significantly altered chromatin accessibility. Based on comprehensive human Transcription Factor (TF) motif enrichment analysis, we discover that the regions

directly bound and opened by each EBV vTR are strongly enriched for AP-1 family motifs, whereas the regions directly bound and closed by eight of the nine vTRs are strongly enriched for CTCF motifs. We predict and confirm that, similar to EBNA2, RTA interacts with the human-encoded RBPJ TF, stabilizing its binding at thousands of human genomic loci. Strikingly, the relatively understudied EBV-encoded BMRF1 protein alters the expression of >7,000 human genes, occupies >12,000 regions of the human genome, and alters chromatin accessibility at >37,000 regions. BMRF1 binding has important implications for human disease; for example, regions occupied by BMRF1 are highly enriched for genetic variants associated with rheumatoid arthritis (RA). Systems-level modeling of vTR interactions revealed a complex landscape where many human genes are targeted by multiple vTRs, including *MYC*, *BCL11B*, and *STAT1*.

Collectively, these data comprise a valuable resource for the systematic study of the impact of EBV vTRs on human gene regulatory mechanisms. The data are freely available on multiple platforms, facilitating future studies aimed at understanding EBV vTR binding events, impacts on chromatin accessibility, alterations of human gene expression, and connections to EBV-implicated human diseases.

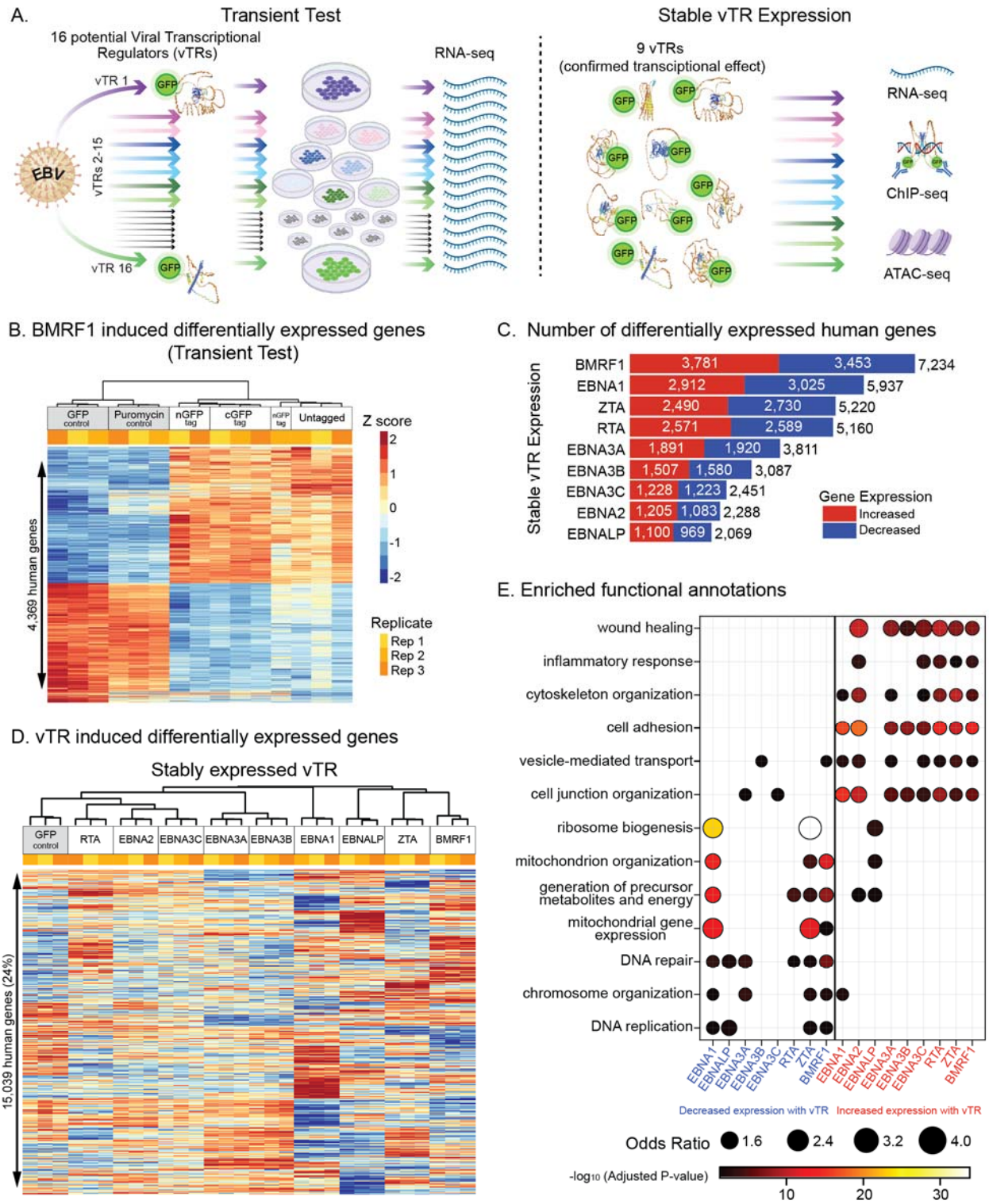
Results

An experimental system to systematically investigate EBV vTR-induced impacts on human gene regulation

We initially sought to identify the role of vTRs encoded by EBV on human transcription. To this end, we focused on eight previously described vTRs (EBNA1, EBNA2, EBNA3A,

EBNA3B, EBNA3C, EBNALP, RTA, and ZTA) and eight less characterized vTRs (BALF2, BCRF1, BDLF3, BDLF4, BFRF2, BGLF3, BMRF1, and BVLF1) that were systematically identified in our catalogue of vTRs from all human viruses (23). We transiently expressed each EBV vTR and measured the impact on human gene expression relative to a control cell line that only expressed GFP (**Figure 1A** and **Supplemental Figure 1**). Of the eight less characterized EBV vTRs, BMRF1 was selected for further study based on its robust and non-redundant impact on gene expression (**Supplemental Figure 1**). Notably, in the transient system, BMRF1 demonstrated the most robust impact on gene expression (**Supplemental Figure 1 panels A and B**), with BMRF1 strongly separated from controls and the other less characterized vTRs in a principal component analysis (PCA) (**Supplemental Figure 1 panel C**).

Figure 1. EBV vTRs induce extensive non-redundant changes to human gene expression. A. Experimental system. **B.** Differential human gene expression in the transient test of one representative vTR (BMRF1). BMRF1-induced differentially expressed human genes are shown for experimental triplicates. Values indicate normalized gene expression levels that have been Z-transformed across rows (conditions). Columns and rows were clustered using Ward's method. **C.** Number of differentially expressed human genes in cells stably expressing each vTR. **D.** Quantification of vTR induced human gene expression changes in cells stably expressing each vTR. All human genes with statistically reproducible differential expression induced by any vTR are shown. Values indicate normalized gene expression levels that have been Z-transformed across rows (conditions). Columns and rows were clustered using Ward's method. **E.** Functional annotations enriched among up and down regulated genes across cell lines stably expressing each vTR. Enriched pathways were identified using ToppGene (see Methods). Up to the top five most significant GO Slim Biological Process categories for each vTR are included. Panel has been curated to reduce redundant pathways (see Supplemental Figure 7 for full results).


A.
Transient Test
16 potential Viral Transcriptional Regulators (vTRs)
vTR 1
vTRs 2-15
vTR 16
EBV
GFP
RNA-seq
Stable vTR Expression
9 vTRs
(confirmed transcriptional effect)
RNA-seq
ChIP-seq
ATAC-seq
B. BMRF1 induced differentially expressed genes (Transient Test)
GFP control
Puromycin control
nGFP tag
cGFP tag
Untagged
Z score
2
1
0
-1
-2
4,369 human genes
Replicate
Rep 1
Rep 2
Rep 3
C. Number of differentially expressed human genes
Stable vTR Expression
BMRF1 3,781 3,453 7,234
EBNA1 2,912 3,025 5,937
ZTA 2,490 2,730 5,220
RTA 2,571 2,589 5,160
EBNA3A 1,891 1,920 3,811
EBNA3B 1,507 1,580 3,087
EBNA3C 1,228 1,223 2,451
EBNA2 1,205 1,083 2,288
EBNALP 1,100 969 2,069
Gene Expression
Increased
Decreased
D. vTR induced differentially expressed genes
Stably expressed vTR
GFP control
RTA
EBNA2
EBNA3C
EBNA3A
EBNA3B
EBNA1
EBNALP
ZTA
BMRF1
15,039 human genes (24%)
E. Enriched functional annotations
wound healing
inflammatory response
cytoskeleton organization
cell adhesion
vesicle-mediated transport
cell junction organization
ribosome biogenesis
mitochondrion organization
generation of precursor metabolites and energy
mitochondrial gene expression
DNA repair
chromosome organization
DNA replication
EBNA1
EBNALP
EBNA3A
EBNA3B
EBNA3C
RTA
ZTA
BMRF1
EBNA1
EBNA2
EBNALP
EBNA3A
EBNA3B
EBNA3C
RTA
ZTA
BMRF1
Decreased expression with vTR
Increased expression with vTR
Odds Ratio
1.6
2.4
3.2
4.0
-log10 (Adjusted P-value)
10
20
30

We next performed a series of experiments aimed at optimizing our reagents for downstream use. In this study, we use GFP as an epitope tag to allow us to perform chromatin immunoprecipitation to measure genome-wide binding (**Figure 1A**). We therefore performed transient transfection tests in which GFP was tagged to each vTR on either the C- or the N-terminus. Using RNA sequencing (RNA-seq), we then measured gene expression with and without each tag for each vTR (**Figure 1B, Supplemental Figure 2**, and **Supplemental Dataset X**). For each protein, the cells expressing the vTR demonstrated a substantially different expression pattern as compared to control cells that only expressed GFP or a puromycin resistance cassette (**Figure 1B and Supplemental Figure 2**). The final location of the GFP tag (**Supplemental Figure 3**) was selected based upon similarity of gene expression effects to the untagged vTR control using both heatmap-based clustering (**Supplemental Figure 2**) and PCA analysis (**Supplemental Figure 4**). AlphaFold (55) predictions affirmed that the selected GFP tag location would likely not disrupt protein folding (**Supplemental Figure 3**).

Once the nine EBV vTRs (eight known vTRs and BMRF1) were selected for further study, we generated isogenic stable cell lines expressing each vTR, with a Flp recombinase enabling integration into a specific genomic location (*FRT*), resulting in consistent vTR expression across the cell lines (**Figure 1A, Supplemental Figure 5, Supplemental Dataset X**, and see Methods). These isogenic stable cell lines were then used to study the impact of EBV vTRs on gene expression (RNA-seq), interactions with the human genome (ChIP-seq), and chromatin accessibility (ATAC-seq) (**Figure 1A**).

EBV vTRs induce extensive non-redundant changes to human gene expression

We first performed biological triplicate RNA-seq of each cell line with stable EBV vTR expression (see Methods). When assessed together, the resulting data demonstrate consistent expression levels and strong quality control metrics (**Figure 1 C and D, Supplemental Dataset X**). We next performed differential analysis of each vTR against the control cell line expressing GFP. Genes with a fold-change of at least 20% and an adjusted p-value less than 0.05 were called differentially expressed. A 20%-fold-change threshold was chosen based on the identification of an inflection point in the number of differential genes across various thresholds that was consistent across all vTRs (**Supplemental Figure 6**). Of the 61,544 human coding and non-coding genes that we examined, 15,039 (24%) were impacted by at least one EBV vTR. Expression of each of the nine selected EBV vTRs resulted in robust changes to gene expression, ranging from 2,069 to 7,234 differentially expressed genes (DEGs) (**Figure 1C**, **Supplemental Dataset X**). Overall, EBV vTRs regulate human gene expression in a balanced fashion: expression of all vTRs results in upregulation and/or downregulation between 45% and 55% of differentially expressed host genes (**Figure 1C**). Unexpectedly, BMRF1 had a greater impact on host gene expression than any of the eight known EBV vTRs (**Figure 1C**). To our knowledge, only a single study of this protein's impact on human gene expression has previously been performed (56).

EBV vTRs broadly regulate different host genes (**Figure 1D, Supplemental Dataset X**), and the biological pathways that are enriched for EBV vTR upregulated genes are largely different from those enriched for downregulated genes (**Figure 1E**, **Supplemental Figure 7**, and **Supplemental Dataset X**). Biological processes enriched for genes that are significantly upregulated by EBV vTR expression were largely shared, and include wound healing, inflammatory response, cytoskeleton organization, and cell adhesion (**Figure 1E** and **Supplemental Figure 7**). By contrast, expression of numerous EBV vTRs led to the downregulation of genes enriched for mitochondrion organization, DNA repair, and chromosome organization (**Figure 1E and Supplemental Figure 7**). We note that most mitochondrial proteins are encoded by the host nuclear genome. Collectively, these results indicate that EBV vTRs have widespread effects on a wide range of human genes and pathways.

<u>EBV vTRs occupy thousands of shared and distinct regions across the human genome.</u> We next sought to systematically catalog the regions of the human genome that are occupied by each of the nine vTRs that substantially alter human gene expression level. To this end, we performed anti-GFP ChIP-seq in the nine cell lines harboring each vTR (see Methods). The resulting data were of high quality, meeting or exceeding the standards established by the ENCODE consortium (57, 58). Fraction of Reads in Peaks (FRiP) scores all exceeded 0.03, every replicate set had at least 1,500 reproducible peaks, and the replicates clustered together tightly and distinctly from GFP-only controls (**Supplemental Dataset X** and **Supplemental Figure 8**). Further, vTR ChIP-seq peaks were enriched for known DNA binding motifs (ZTA, EBNA1) or known interacting human TF motifs (EBNA2 with human RBPJ motifs) (**Supplemental Dataset X**). ChIP-seq

peaks also recapitulate known biology in other experimental systems, with strong EBNA2 ChIP-seq peaks coinciding with previously published EBNA2 ChIP peaks (18) (**Figure 2A**, middle, **Supplemental Figure 9**, **Supplemental Dataset X**). For example, the *MAP3K8* gene is strongly induced by EBNA2 in our system (**Figure 2A**, top), has an EBNA2 ChIP-seq peak in the promoter, and has been shown to be transcriptionally regulated by EBNA2 (17, 26). Additional examples involving known EBNA2-based gene regulatory mechanisms (e.g., *BMF*, *CDKN1A*, *CSRNP1*, and *CXXC5* (17)) are provided in **Supplemental Figure 9.** Taken together, these results indicate that this ChIP experimental system is of high quality and is capable of recapitulating previously established molecular mechanisms.

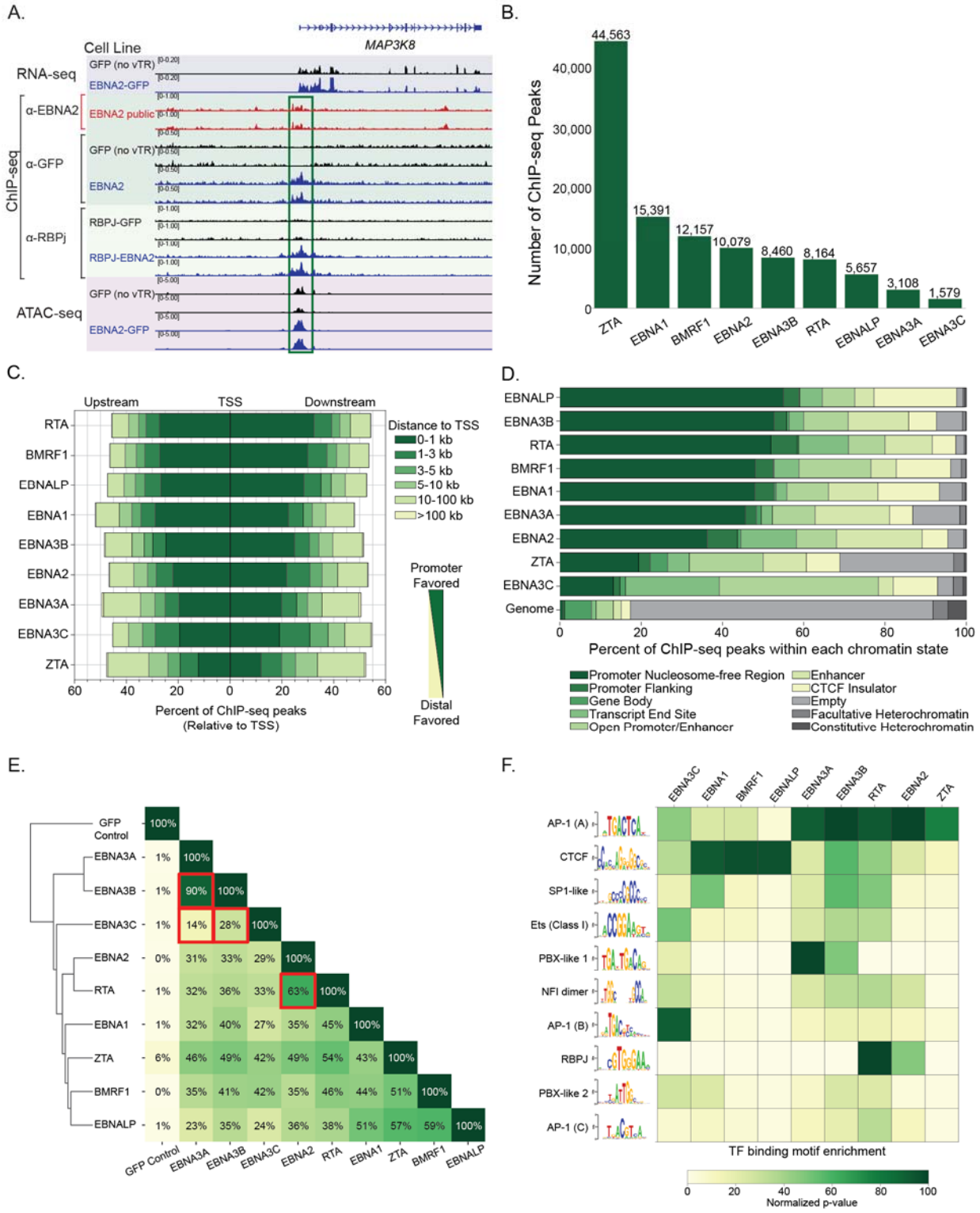

A.
Cell Line
MAP3K8
RNA-seq
GFP (no vTR)
EBNA2-GFP
ChIP-seq
α-EBNA2
EBNA2 public
α-GFP
GFP (no vTR)
EBNA2
α-RBPj
RBPJ-GFP
RBPJ-EBNA2
ATAC-seq
GFP (no vTR)
EBNA2-GFP
B.
Number of ChIP-seq Peaks
44,563
15,391
12,157
10,079
8,460
8,164
5,657
3,108
1,579
ZTA
EBNA1
BMRF1
EBNA2
EBNA3B
RTA
EBNALP
EBNA3A
EBNA3C
C.
Upstream
TSS
Downstream
RTA
BMRF1
EBNALP
EBNA1
EBNA3B
EBNA2
EBNA3A
EBNA3C
ZTA
Distance to TSS
0-1 kb
1-3 kb
3-5 kb
5-10 kb
10-100 kb
>100 kb
Promoter Favored
Distal Favored
Percent of ChIP-seq peaks (Relative to TSS)
D.
EBNALP
EBNA3B
RTA
BMRF1
EBNA1
EBNA3A
EBNA2
ZTA
EBNA3C
Genome
Percent of ChIP-seq peaks within each chromatin state
Promoter Nucleosome-free Region
Promoter Flanking
Gene Body
Transcript End Site
Open Promoter/Enhancer
Enhancer
CTCF Insulator
Empty
Facultative Heterochromatin
Constitutive Heterochromatin
E.
GFP Control
EBNA3A
EBNA3B
EBNA3C
EBNA2
RTA
EBNA1
ZTA
BMRF1
EBNALP
F.
AP-1 (A)
CTCF
SP1-like
Ets (Class I)
PBX-like 1
NFI dimer
AP-1 (B)
RBPJ
PBX-like 2
AP-1 (C)
TF binding motif enrichment
Normalized p-value

Figure 2. EBV vTRs occupy thousands of shared and distinct regions across the human genome. **A.** Example of EBV vTR-induced human gene regulatory changes. Top rows: EBNA2-induced alteration to *MAP3K8* gene expression (RNA-seq). Vertical green outlined box, top-to-bottom: EBNA2 binding in a public dataset (red) and in our system (ChIP-seq), EBNA2-altered RBPJ binding (ChIP-seq), and EBNA2-altered chromatin accessibility (ATAC-seq). **B.** EBV vTR ChIP-seq peak counts in the human genome. **C.** vTR ChIP-seq peak distribution relative to human transcription start sites (TSS). vTRs are sorted top-to-bottom, in order of "promoter favored" to "distal favored." **D.** vTR ChIP-seq peak distribution within human chromatin states. Chromatin states were calculated using a variety of genome-scale epigenetic measurements (see Methods). vTRs are sorted top-to-bottom, in order of "promoter associated" to "non-promoter associated" states. The frequency of each state across the human genome is provided at the bottom for reference. **E.** Pairwise overlap of vTR ChIP-seq peaks. Each entry indicates the percent overlap (number of intersecting peaks divided by the number of peaks in the smaller peak set) between the given pair of vTR ChIP-seq datasets. Rows and columns were hierarchically clustered using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm. vTR pairs discussed in the text are indicated with red boxes. **F.** Human TF binding site motif enrichment within vTR ChIP-seq peaks. The DNA sequences of vTR ChIP-seq peaks were subjected to motif enrichment analysis (see Methods). The top 10 unique motifs across all nine vTRs are shown. Non-redundant motifs were established through manual curation. Normalized p-values indicate the percentage of the maximum p-value obtained from motif enrichment analysis for each vTR. Motifs are sorted top-to-bottom from strongest to weakest cumulative enrichment.

Encouraged by the high quality of our ChIP-seq data, we next took advantage of our systematic approach to compare various biological attributes of the EBV vTRs. In total, EBV vTRs occupy 64,290 distinct regions of the human genome, comprising 62,340,273 total bases (2% of the entire human genome), consistent with their enormous cumulative effect on human gene expression levels (**Figure 1C**). Overall,

ZTA had the largest number of binding events, with over 44,000 reproducible ChIP-seq peaks in the human genome (**Figure 2B**). Surprisingly, the understudied BMRF1 protein had over 12,000 peaks in the human genome (**Figure 2B**), implying a potentially important role in EBV-host interactions. All vTRs bind with high symmetry (upstream vs. downstream) with respect to human transcription start sites (TSS) (**Figure 2C**). Some vTRs prefer binding to promoter regions – for example, >80% of RTA and BMRF1 ChIP-seq peaks fall within 5 kb of a TSS. In contrast, ZTA tends to bind more distal regions, with <60% of its peaks within 5kb of the TSS, consistent with the similarity of its structure and DNA binding motif to human AP-1 TFs, which also tend to occupy enhancers (59).

We next examined the types of functional regions that statistically coincide with vTR binding events in the human genome. To this end, we used previously computed chromatin states obtained by running the ChromHMM method (60) on ChIP-seq data for CTCF and six histone marks, as well as ATAC-seq data (61). Notably, all nine vTRs occupied considerably more regulatory regions of the genome (i.e., “non-empty” and “non-heterochromatin” regions) compared to genomic background (**Figure 2D**), consistent with their roles in gene regulation. Nonetheless, almost 30% of ZTA ChIP-seq peaks coincide with empty or heterochromatin regions, suggesting that it might have similar roles to recently discovered “Dark TFs” (61). Consistent with the TSS analysis, ZTA and EBNA3C both tend to avoid promoters relative to the other vTRs. The binding events of three vTRs (EBNALP, EBNA1, and BMRF1) strongly coincide with CTCF insulators, suggesting possible roles in the alteration of chromatin

architecture through CTCF-based mechanisms. Collectively, these data indicate that the binding events of all nine vTRs strongly localize within regulatory regions of the human genome, with specific vTRs likely participating preferentially in different gene regulatory mechanisms.

We next compared the location of vTR binding events relative to one another. Overall, EBV vTRs occupy largely distinct regions of the genome, with a median pairwise ChIP-seq peak overlap of only 39% (**Figure 2E**). The strongest peak overlap (90%) was observed for EBNA3A and EBNA3B, consistent with their known ability to bind to shared regions of the genome (51). In contrast, EBNA3C had minimal overlap with EBNA3A and EBNA3B (14% and 28%, respectively). EBNA2 and RTA peaks displayed strikingly strong overlap (63%), which is surprising given that they are expressed in mutually exclusive stages of the EBV life cycle (latent vs. lytic, respectively) (62).

Finally, we examined the enrichment of human TF DNA binding motifs within the vTR-occupied regions of the human genome (see Methods). As expected, ZTA enriched for AP-1-like motifs, and EBNA2 enriched for motifs of its well-established binding partner RBPJ (63) (**Figure 2F**, **Supplemental Dataset X**). Surprisingly, RTA peaks displayed even stronger enrichment for RBPJ motifs. To our knowledge, this represents the first evidence that EBV-encoded RTA may engage with RBPJ. The RTA encoded by Kaposi's Sarcoma-associated Herpesvirus (KSHV) is known to interact with RBPJ (64). However, KSHV-RTA shares only limited sequence homology with its EBV counterpart (25% amino acid identity, 38% similarity). Strikingly, all nine vTRs strongly enriched for

either CTCF (EBNA1, BMRF1, EBNALP) or AP-1 (all others) motifs, but not both. These results are consistent with our chromatin state analyses, which implicate these same vTRs at CTCF insulators and enhancers, respectively (**Figure 2D**). Certain other motifs, such as SP1, Ets, PBX, and NFI, are enriched for multiple vTRs, suggesting that they might be shared EBV vTR targets. In summary, these analyses demonstrate that EBV vTRs occupy over 64,000 distinct regions across the human genome, exhibiting variation in genomic localization, preferences for specific chromatin states, and enrichment for human TF motifs.

<u>EBV vTR expression robustly changes human chromatin accessibility</u>

We next examined the effect of each vTR on human chromatin accessibility. To this end, we performed ATAC-seq experiments in cells expressing either GFP alone (as a control) or each vTR independently. Quality control analyses indicated that each dataset was of high quality, with FRiP scores of at least 0.15 (median of 0.32), TSS enrichment scores of at least 4 (median of 11.18), and at least 45,000 ATAC-seq peaks (median of 71,112) (**Supplemental Dataset X**).

We first compared the regions of open chromatin in each of the ten cell lines. All vTRs had at least a moderate impact on human chromatin, with a range of 76% to 93% shared peaks with the GFP control cell line (**Figure 3A**). The lone exception was ZTA, which shared only 63% of its peaks with the control cell line, suggesting that ZTA has potent pioneering-like activity. To further examine these phenomena, we next identified statistically robust chromatin opening and closing events by comparing the ATAC-seq

signal strength of each vTR cell line with the GFP control cell line (see Methods). Next, for each vTR, we partitioned its differential ATAC-seq peaks into peaks that are likely to be directly altered by that vTR and those likely to be impacted by downstream, indirect effects. Specifically, we identified *direct vTR effects on chromatin* as significantly differential vTR-induced ATAC-seq peaks that overlap with ChIP-seq peaks for the same vTR (**Figure 3B, left**). Similarly, we identified *indirect vTR effects on chromatin* as significantly differential vTR-induced ATAC-seq peaks that do not overlap with ChIP-seq peaks for the same vTR (**Figure 3B, right**). Using these definitions, we observe a wide range of effects on human chromatin accessibility across the nine vTRs. ZTA by far has the largest effect on chromatin accessibility, with 16,661 and 5,038 ZTA-induced direct opening and closing effects, respectively (**Figure 3C**). EBNA2 also had a large number of direct chromatin effects, consistent with our previous work (17). Strikingly, BMRF1 also had many direct effects, with 5,787 BMRF1-induced closing events (**Figure 3C**). In total, EBV vTRs directly induce the opening or closing of 23,913 and 9,493 distinct regions of the genome, respectively. As expected due to cascades of downstream regulatory effects, each vTR had more indirect than direct effects on chromatin (**Figure 3D**). Considering both direct and indirect effects, EBV vTRs collectively open or close 109,533 genomic regions, encompassing 97,127,999 unique bases (3.2% of the human genome). These results reveal compelling evidence that EBV vTRs strongly impact the human chromatin accessibility landscape.

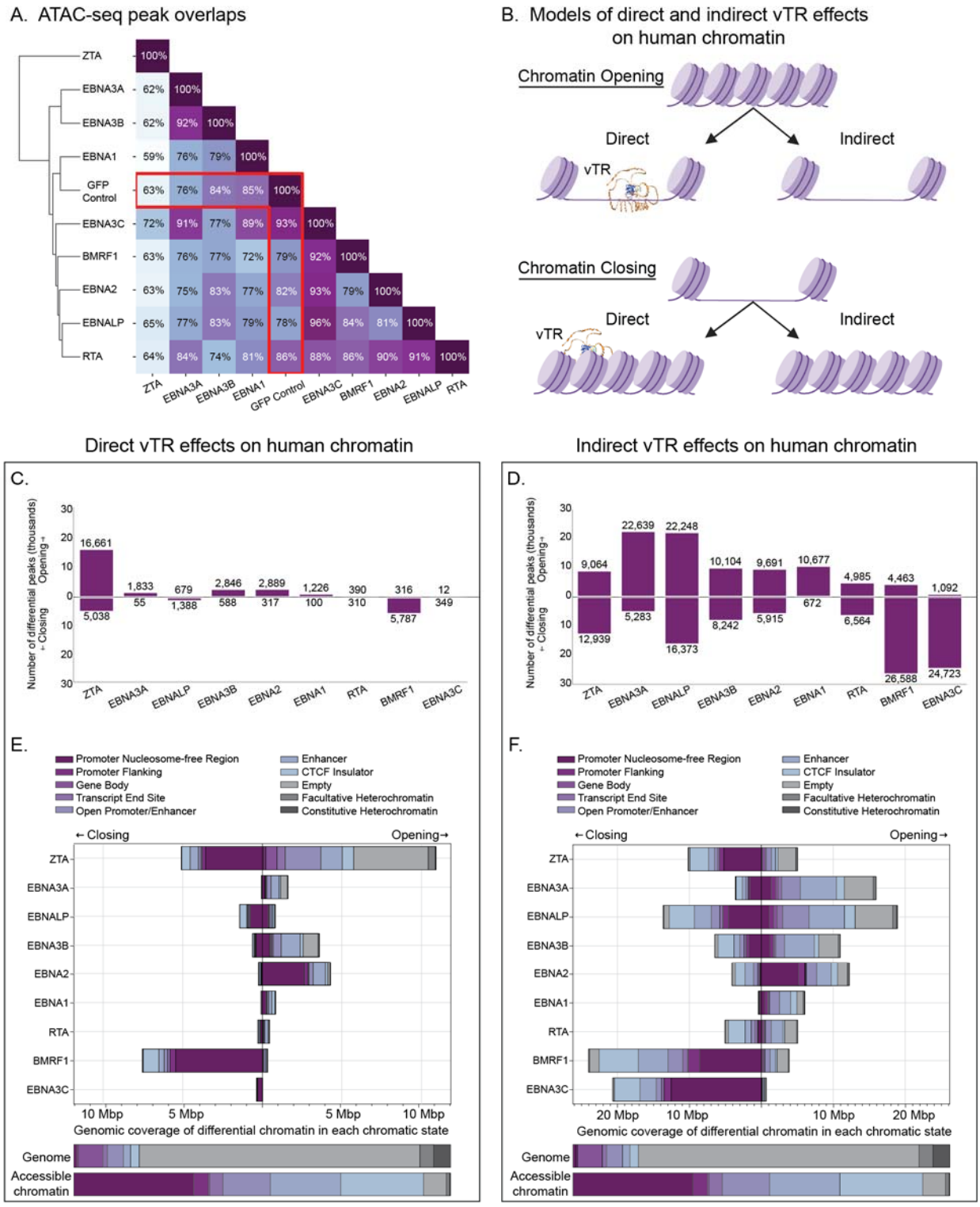

A. ATAC-seq peak overlaps
ZTA 100%
EBNA3A 62% 100%
EBNA3B 62% 92% 100%
EBNA1 59% 76% 79% 100%
GFP Control 63% 76% 84% 85% 100%
EBNA3C 72% 91% 77% 89% 93% 100%
BMRF1 63% 76% 77% 72% 79% 92% 100%
EBNA2 63% 75% 83% 77% 82% 93% 79% 100%
EBNALP 65% 77% 83% 79% 78% 96% 84% 81% 100%
RTA 64% 84% 74% 81% 86% 88% 86% 90% 91% 100%
ZTA EBNA3A EBNA3B EBNA1 GFP Control EBNA3C BMRF1 EBNA2 EBNALP RTA
B. Models of direct and indirect vTR effects on human chromatin
Chromatin Opening
Direct
Indirect
vTR
Chromatin Closing
Direct
Indirect
vTR
Direct vTR effects on human chromatin
Indirect vTR effects on human chromatin
C.
Number of differential peaks (thousands)
Opening→
←Closing
16,661 5,038
1,833 55
679 1,388
2,846 588
2,889 317
1,226 100
390 310
316 5,787
12 349
ZTA EBNA3A EBNALP EBNA3B EBNA2 EBNA1 RTA BMRF1 EBNA3C
D.
9,064 12,939
22,639 5,283
22,248 16,373
10,104 8,242
9,691 5,915
10,677 672
4,985 6,564
4,463 26,588
1,092 24,723
ZTA EBNA3A EBNALP EBNA3B EBNA2 EBNA1 RTA BMRF1 EBNA3C
E.
Promoter Nucleosome-free Region
Promoter Flanking
Gene Body
Transcript End Site
Open Promoter/Enhancer
Enhancer
CTCF Insulator
Empty
Facultative Heterochromatin
Constitutive Heterochromatin
←Closing
Opening→
ZTA EBNA3A EBNALP EBNA3B EBNA2 EBNA1 RTA BMRF1 EBNA3C
10 Mbp 5 Mbp 5 Mbp 10 Mbp
Genomic coverage of differential chromatin in each chromatic state
Genome
Accessible chromatin
F.
Promoter Nucleosome-free Region
Promoter Flanking
Gene Body
Transcript End Site
Open Promoter/Enhancer
Enhancer
CTCF Insulator
Empty
Facultative Heterochromatin
Constitutive Heterochromatin
←Closing
Opening→
ZTA EBNA3A EBNALP EBNA3B EBNA2 EBNA1 RTA BMRF1 EBNA3C
20 Mbp 10 Mbp 10 Mbp 20 Mbp
Genomic coverage of differential chromatin in each chromatic state
Genome
Accessible chromatin

Figure 3. EBV vTR expression robustly changes human chromatin accessibility. A. Pairwise overlap of ATAC-seq peaks in vTR expressing cell lines. Each entry indicates the percent overlap (number of intersecting peaks divided by the number of peaks in the smaller peak set) between the given pair of vTR ATAC-seq datasets. Rows and columns were hierarchically clustered using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm. GFP control, which indicates peak overlap with cells lacking a vTR, is boxed in red. **B.** Models of direct and indirect vTR effects on human chromatin**.** We define genomic regions with direct opening or closing of chromatin as vTR-induced differential ATAC-seq peaks that directly overlap with a ChIP-seq peak for that vTR. All other differential ATAC-seq peaks are defined as indirect. **C and D.** Quantification of vTR-altered ATAC-seq peaks (C, direct; D, indirect). For each vTR, the total number of differential ATAC-seq peaks are indicated (up bars, opening; down bars, closing). **E and F.** vTR-altered ATAC-seq peak distribution within human chromatin states (E, direct; F, indirect)**.** Chromatin states were calculated using a variety of genome-scale epigenetic measurements (see Methods). The frequency of each state across the human genome and within accessible chromatin is provided at the bottom for reference.

We next sought to characterize the potential functional impact of vTR-induced chromatin accessibility changes. To this end, we compared the genomic locations of vTR-induced chromatin opening and closing events to ChromHMM-calculated chromatin states, again distinguishing between direct and indirect effects. The majority of direct EBV vTR-induced chromatin accessibility changes are located within or near human gene promoters (**Figure 3E**). Specifically, EBNA2 tends to open promoters and BMRF1 and ZTA tend to close promoters (**Figure 3E**). The majority of ZTA-induced direct chromatin opening events do not involve promoters but instead consist largely of empty and heterochromatin regions of the genome, consistent with the preferential binding of ZTA to these locations (**Figure 2D**). Examination of indirect vTR effects

further underscores many of these observations, including the predominant effects of EBNA2 opening promoters, ZTA closing promoters, and BMRF1 closing promoters (**Figure 3F**). Although EBNA3C has only limited direct effects on chromatin accessibility (12 and 349 opening and closing events, respectively), it has a substantial number of indirect closing effects (24,723). Collectively, these results reveal that EBV vTRs have substantial and diverse effects on various regions of the human genome, with differential effects on promoters, distal regions, and empty regions of "genomic dark matter" (61).

Finally, to identify the human TFs that are most strongly impacted by these chromatin changes, we performed TF binding motif enrichment analysis in the direct vTR-induced chromatin accessibility regions (**Supplemental Dataset X**). This analysis revealed that AP-1 is the first or second most strongly enriched opened motif for all vTRs, and CTCF is the first or second closed motif for almost every vTR (**Figure 4**). EBNA1 represents an interesting exception by instead strongly opening CTCF motifs and strongly closing Ets motifs. EBNA2, ZTA, and EBNA1 uniquely close AP-1 sites among vTRs. RBPJ motifs are opened by both EBNA2 and RTA, consistent with their strong enrichment within the ChIP-seq peaks of these proteins (**Figure 2F**). Intriguingly, two different classes of homeoboxes are specifically enriched within the opened regions of EBNA3A/3B (PBX-like) and EBNA3C (NKX-like) (**Figure 4**). Collectively, these results indicate that EBV vTRs induce extensive alterations to human chromatin accessibility that primarily impact promoters through a variety of mechanisms (opening vs closing, various impacted classes of human TFs).

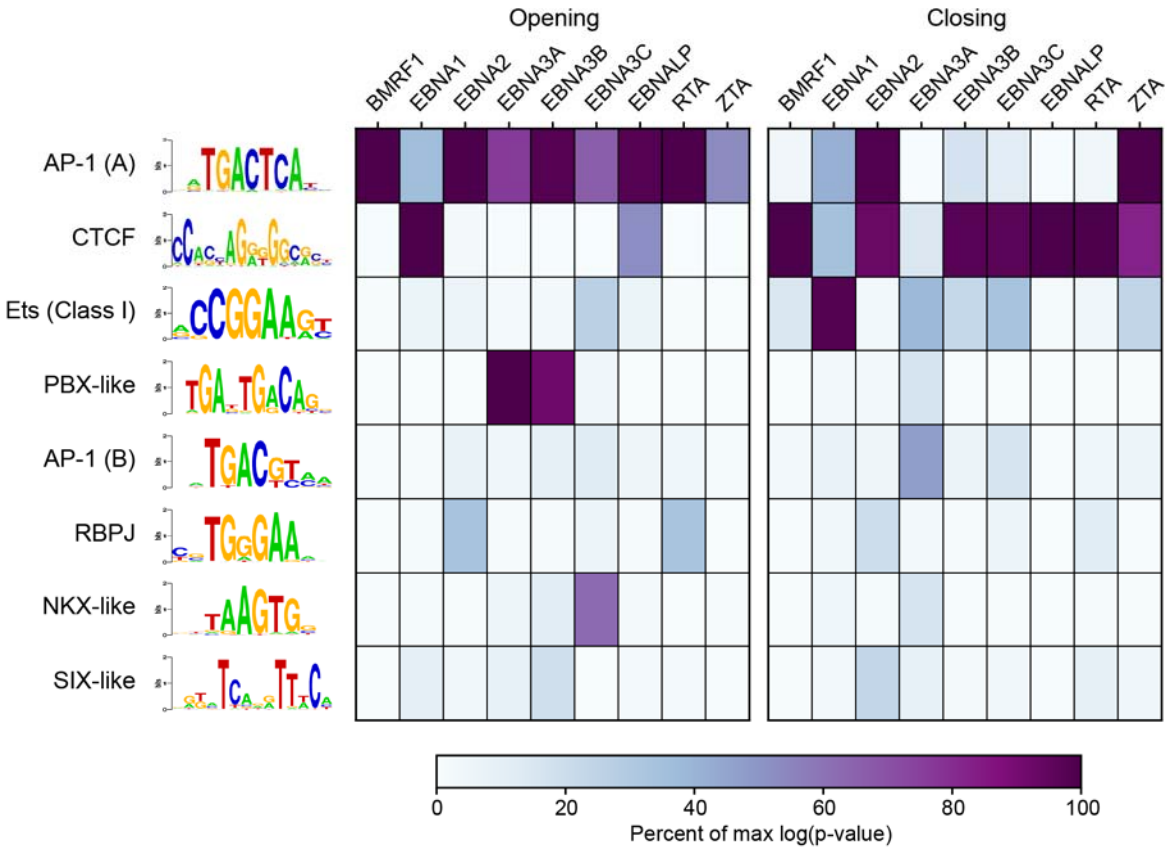


Figure 4. Human TF motif enrichment in regions of the human genome with chromatin accessibility directly altered by EBV vTRs. ATAC-seq experiments were performed in human cells following the expression of one of nine vTRs or GFP as a control. These data were used to identify vTR-induced alterations to chromatin accessibility across the human genome (see Methods). Human TF motif enrichment analysis was performed in these opening or closing chromatin regions for each vTR. Normalized p-values indicate the percentage of the maximum p-value obtained from motif enrichment analysis for each vTR, using a curated non-redundant set of motifs that are strongly enriched for at least one vTR.

An integrated assessment of EBV vTR interactions with the human genome

We next sought to examine the relationships across data types for each vTR. We first asked if vTR induced alterations to human gene expression (RNA-seq) coincided with

vTR binding to the human genome (ChIP-seq). Indeed, for each vTR, we observed strong enrichment of ChIP-seq peaks within a 10kb region of genes induced by that vTR. Binding of each vTR had significant enrichment for both upregulated and downregulated genes ($10^{-190} < P < 10^{-5}$) (**Supplemental Figure 10**, panel A). Likewise, binding of each vTR had significant enrichment for vTR-induced opening and closing chromatin regions ($10^{-215} < P < 0.02$) (**Supplemental Figure 10**, panel B).

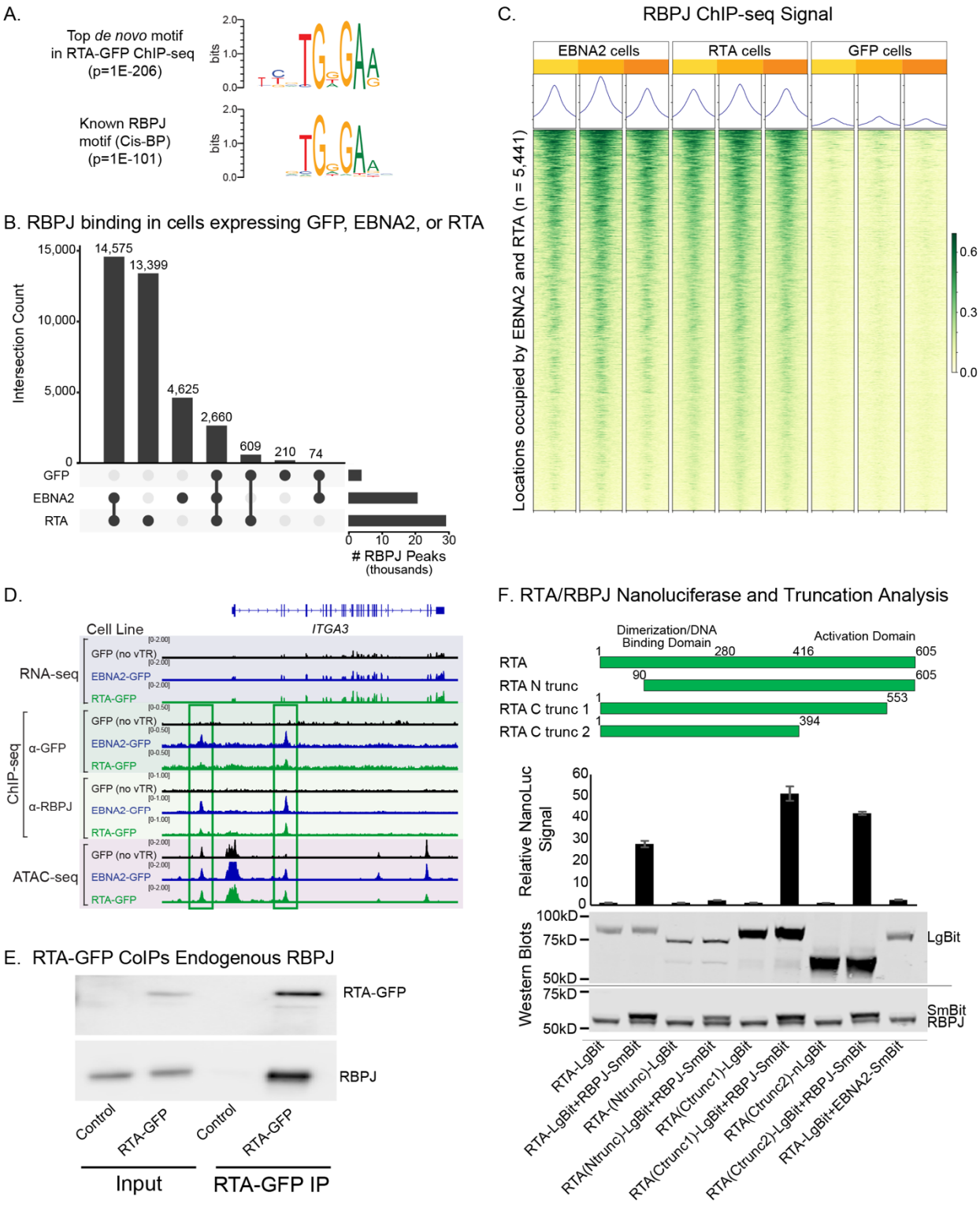
Data integration also reveals multiple compelling examples of EBV vTR-altered gene regulatory mechanisms. As an example, we highlight six regions (**Supplemental Figure 9**) in which EBNA2 binds, increases chromatin accessibility, and induces gene expression of genes that are involved in the interferon response to viruses or are known to be regulated by EBNA2 (65-70). At each locus we also observe RBPJ binding in our system and EBNA2 binding in B cells with whole virus infection (18, 71). Collectively, these results indicate a strong concordance between the data types obtained in this study.

<u>EBNA2 and RTA independently stabilize RBPJ DNA binding to the human genome</u>

Many vTRs exert their regulatory functions through interactions with host factors (23). For example, EBNA2 is a well-established Notch mimic that interacts with and stabilizes RBPJ binding to the host and viral genomes (63). In our integrative analyses, we uncovered several lines of evidence suggesting that a second EBV-encoded protein, RTA, might also interact with RBPJ. First, EBNA2 and RTA ChIP-seq peaks have substantial overlap (63%, ranked second behind the EBNA3A/EBNA3B pair) (**Figure**

2E). Second, RTA ChIP-seq peaks enrich for RBPJ motifs even more strongly than EBNA2 ChIP-seq peaks (**Figure 2F**). Third, human genomic regions that open or close with expression of RTA strongly enrich for RBPJ motifs, similar to what we observe for EBNA2 (**Figure 4**). Fourth, the top enriched *de novo* motif in RTA ChIP-seq peaks is a near-perfect match to the known RBPJ motif (**Figure 5A**, **Supplemental Dataset X**). To further examine the possibility that RTA and RBPJ might interact, we performed ChIP-seq for RBPJ in human cells expressing EBNA2, RTA, or GFP only (as a control). The resulting data met our high quality control standards, with FRiP scores exceeding 0.09, peak counts above 20,000, very strong enrichment for RBPJ motifs, and replicates clustering together tightly and distinctly from GFP-only controls (**Supplemental Dataset X**, **Supplemental Figure 11**).

Figure 5. EBNA2 and RTA independently stabilize RBPJ DNA binding to the human genome. A. RTA-GFP ChIP enriches the RBPJ motif. Top: #1 *de novo* enriched motif obtained from RTA ChIP-seq peaks. Bottom: the known RBPJ DNA binding motif. **B.** RBPJ ChIP-seq peak overlap counts in cells expressing GFP (control), EBNA2, or RTA. **C.** Normalized RBPJ ChIP-seq signal in EBNA2, RTA, or GFP (control) cells at sites ChIPed by both EBNA2-GFP and RTA-GFP. Values indicate counts per million. Window size is –1Kb to 1Kb for each column. **D**. Genome browser screenshot of the *ITGA3* gene, which is upregulated in both EBNA2 and RTA expressing cells. Track sets show (top-to-bottom): RNA-seq, anti-GFP ChIP-seq, anti-RBPJ ChIP-seq, and ATAC-seq signal. For each data type, data are presented in cells expressing (top-to-bottom): GFP, EBNA2, or RTA. Green boxes indicate genomic regions discussed in the text. **E**. Co-immunoprecipitation of RTA-GFP with endogenous RBPJ. Lane order (left to right): input from control cells, input from RTA-GFP cells, GFP immunoprecipitation from control cells, GFP immunoprecipitation from RTA-GFP cells that were probed for RTA-GFP (top blot) or RBPJ (bottom blot). **F**. Split NanoLuc luciferase assays examining the interaction between RTA and RBPJ. Top: Schematic of RTA protein and RTA protein truncations. Middle: Luciferase signal obtained from cells expressing the indicated proteins. Bottom: Western blots on the same cells. Note the similar expression levels of the RTA-LgBit and the RTA truncations, equivalent expression levels of RBPJ-SmBit across cells, and similar endogenous RBPJ expression levels.


A.
Top de novo motif in RTA-GFP ChIP-seq (p=1E-206)
Known RBPJ motif (Cis-BP) (p=1E-101)
bits
B. RBPJ binding in cells expressing GFP, EBNA2, or RTA
Intersection Count
14,575
13,399
4,625
2,660
609
210
74
GFP
EBNA2
RTA
RBPJ Peaks (thousands)
C.
RBPJ ChIP-seq Signal
EBNA2 cells
RTA cells
GFP cells
Locations occupied by EBNA2 and RTA (n = 5,441)
D.
Cell Line
ITGA3
RNA-seq
ChIP-seq
α-GFP
α-RBPJ
ATAC-seq
GFP (no vTR)
EBNA2-GFP
RTA-GFP
E. RTA-GFP CoIPs Endogenous RBPJ
RTA-GFP
RBPJ
Control
Input
RTA-GFP IP
F. RTA/RBPJ Nanoluciferase and Truncation Analysis
Dimerization/DNA Binding Domain
Activation Domain
RTA
RTA N trunc
RTA C trunc 1
RTA C trunc 2
Relative NanoLuc Signal
Western Blots
LgBit
SmBit RBPJ
RTA-LgBit
RTA-LgBit+RBPJ-SmBit
RTA-(Ntrunc)-LgBit
RTA(Ntrunc)-LgBit+RBPJ-SmBit
RTA(Ctrunc1)-LgBit
RTA(Ctrunc1)-LgBit+RBPJ-SmBit
RTA(Ctrunc2)-nLgBit
RTA(Ctrunc2)-LgBit+RBPJ-SmBit
RTA-LgBit+EBNA2-SmBit

We next compared the binding profiles of RBPJ in the three conditions (cells expressing GFP, EBNA2, or RTA). Notably, RBPJ binding is greatly increased in the EBNA2- and RTA-expressing cells (22,978 and 32,130 peaks, respectively) compared to the GFP-only condition (3,744 peaks) (**Figure 5B**). This is likely due to the transient nature of RBPJ binding when it lacks a cofactor such as Notch (72) or EBNA2 (73). Further, the largest class of RBPJ ChIP-seq peaks represents places where RBPJ binds the human genome both in the presence of EBNA2 and in the presence of RTA, but not in the presence of GFP (**Figure 5B**). Indeed, the presence of either EBNA2 or RTA leads to consistently stronger RBPJ ChIP-seq signal within regions at which both EBNA2 and RTA independently localize (**Figure 5C**). An example genomic region is shown in **Figure 5D** for the *ITGA3* gene, which is a known RBPJ target (74). *ITGA3* gene expression levels are significantly increased relative to GFP controls with expression of EBNA2 or RTA (**Figure 5D**, top; **Supplemental Dataset X**). Two regions proximal to the *ITGA3* TSS contain binding events for both EBNA2 and RTA that coincide with RBPJ binding events that only occur in the presence of EBNA2 or RTA (and not in the presence of GFP only) (**Figure 5D**, middle). The intronic region also displays both EBNA2- and RTA-dependent chromatin accessibility, while the promoter region does not (**Figure 5D**, bottom). Additional examples of EBNA2- and RTA-dependent RBPJ binding are provided in **Supplemental Figure 12**. Collectively, these results reveal that both EBNA2 and RTA likely stabilize RBPJ binding at thousands of regions of the human genome.

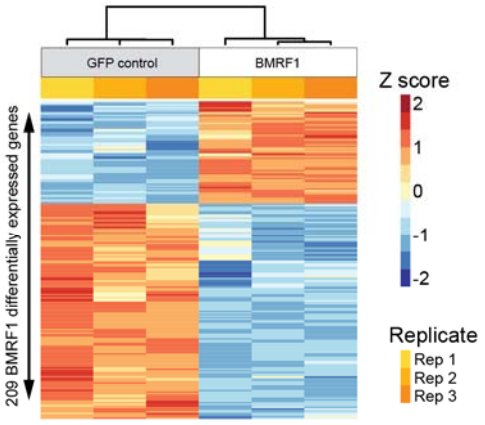
To investigate whether RTA and RBPJ proteins physically interact, we conducted a series of biochemical assays. To this end, we first confirmed that RTA directly interacts with endogenous RBPJ through co-immunoprecipitation assays (**Figure 5E**). Notably, the amount of RBPJ does not change in cells stably expressing RTA compared to control cells (**Figure 5E**, bottom left), indicating that the massive increase in RBPJ binding events in RTA-expressing cells (**Figures 5B and 5C**) is not simply due to increased amounts of available RBPJ protein. To further characterize this interaction, we employed split NanoLuc luciferase assays, wherein the two NanoLuc luciferase halves (SmBit and LgBit) were fused to (1) RBPJ and (2) one of several RTA forms, including full length RTA and various truncated forms (**Figure 5F**). Co-expression of RTA-LgBit and RBPJ-SmBit resulted in a significant increase in luminescence compared to RTA-LgBit alone, confirming an interaction between the two proteins. (**Figure 5F**). Analysis of the truncated forms of RTA revealed that a small N-terminal truncation (aa 1-89) in the region containing the RTA DNA binding domain significantly reduced luciferase signal. In contrast, deletions of C-terminal regions (amino acids 554–605 and 395–605) did not impair the interaction (i.e., the signal was retained). Collectively, these results confirm that RTA colocalizes with RBPJ on the human genome and reveal that the RTA DNA binding domain is necessary for this colocalization.

<u>EBV BMRF1 regulates mitochondrion organization genes with enriched binding at Rheumatoid Arthritis genetic risk loci</u>

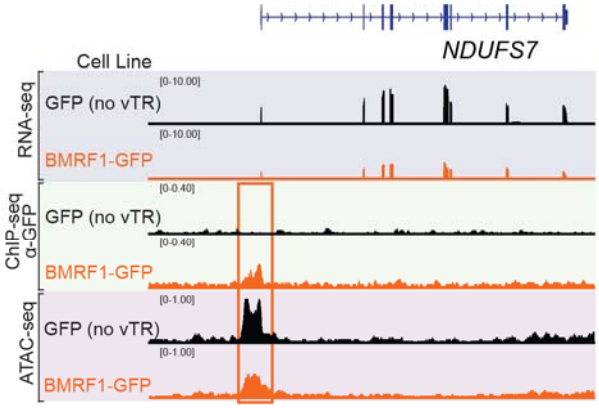
Through a comprehensive analysis of BMRF1's impact on human gene regulation, this study offers a valuable first glimpse into the function of a largely uncharacterized EBV vTR. Of all EBV vTRs examined, BMRF1 had the largest impact on gene expression (**Figure 1C**), with 1,396 genes that are uniquely regulated by this protein. Notably, the majority of these genes were downregulated in the presence of BMRF1. Integrated multi-omic data support a model in which BMRF1 generally binds to promoters, closes chromatin, and reduces gene expression levels (**Figure 2**, **Figure 3**, **Supplemental Figure 10**).

As little is known about this protein, we focused our efforts on deeper characterization of the genes and pathways impacted by BMRF1. Our data implicate an important role for BMRF1 in the reduction of mitochondrial organization gene expression levels (**Figure 1E**). In total, 209 genes involved in this process are regulated by BMRF1 (**Figure 6A**)**.** As an example, *NDUFS7* expression levels are reduced with BMRF1 promoter binding, with a concomitant decrease in chromatin accessibility (**Figure 6B**). NDUFS7 is a component of the "mitochondrial complex I", with mutations in the protein implicated in Leigh Syndrome, a rare neurodegenerative disorder (75, 76). There are many additional examples of mitochondrial organization genes with similar BMRF1-based mechanisms (**Supplemental Figure 13**).

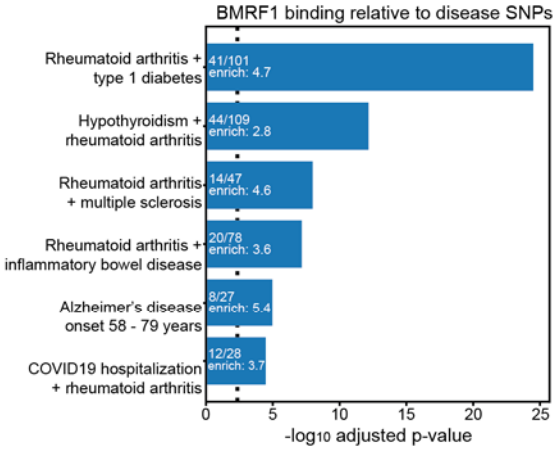
A. Mitochondrion organization genes



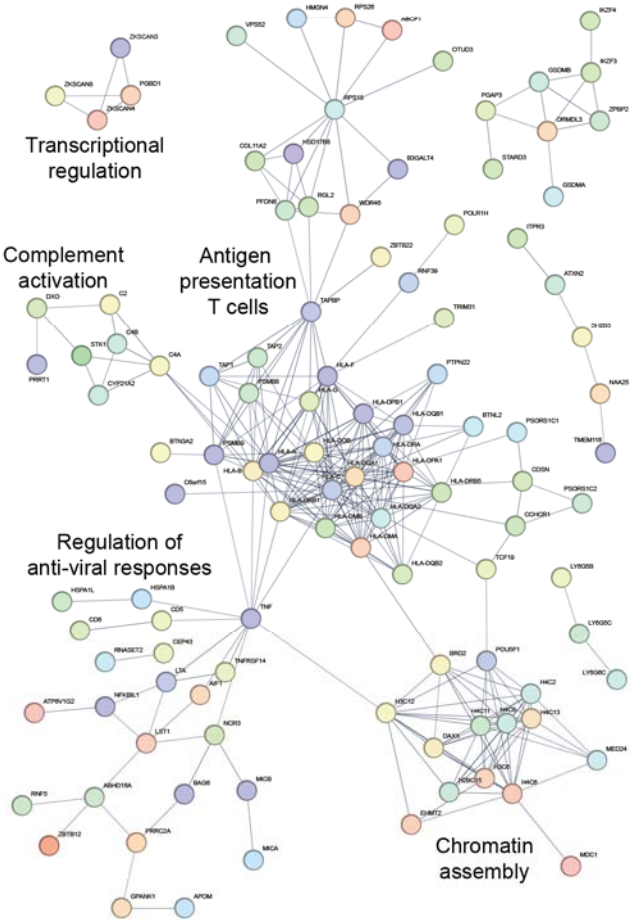
B. Browser shot of *NDUFS7*



C. Enriched BMRF1 binding at disease risk variants



D. Rheumatoid arthritis variant interaction network



E. BMRF1 ChIP-seq allelic imballance

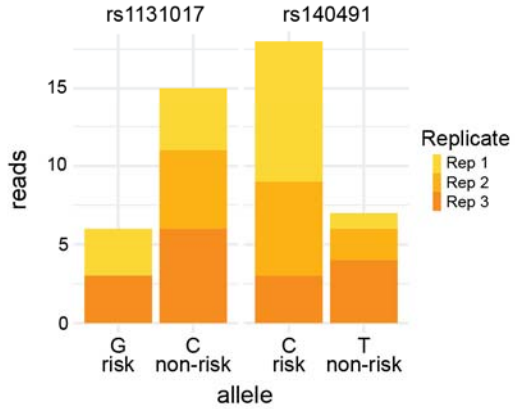


Figure 6. BMRF1 regulates mitochondrion organization genes, with enriched binding at rheumatoid arthritis genetic risk loci. A. Quantification of BMRF1 induced differential genes in the mitochondrion organization process in cells stably expressing BMRF1 or a GFP control. Values indicate normalized gene expression levels that have been Z-transformed across rows (replicates). Columns and rows were clustered using Ward's method. **B.** Genome browser screenshot of the *NDUFS7* gene, which is downregulated and shows reduced promoter chromatin accessibility with BMRF1 expression. Track sets show (top-to-bottom): RNA-seq, anti-GFP ChIP-seq, and ATAC-seq signal. For each data type, data are presented in cells expressing (top-to-bottom) GFP or BMRF1-GFP. Orange box indicates the promoter region discussed in the text. **C.** Enrichment of BMRF1-GFP ChIP-seq peak sets at disease risk loci. All assessed datasets with significant enrichment (adjusted p-values less than p = 0.01) are shown. The –log10 adjusted p-value is graphed on the x-axis and the enrichment is provided within each bar. A dotted line indicates the threshold for statistical enrichment. **D.** STRING representation of genes linked to BMRF1-GFP ChIP-seq peaks at rheumatoid arthritis (RA) risk loci, as identified using the eQTLs (98). Connected gene clusters were labeled based on pathway enrichment. **E.** Allelic BMRF1-GFP ChIP-seq at two RA risk variants as assessed using the MARIO tool (see Methods). The number of reads corresponding to the reference and non-reference allele are quantified for each replicate.

Our work and others have established an important role for the binding of EBV vTRs at genetic risk loci that mediate chronic human diseases (16, 18, 19, 23, 77, 78). Building on this foundation, we conducted an unbiased, systematic analysis using the GWAS Catalog (see Methods, (79)) and found that BMRF1 binding events are highly enriched at genetic risk variants for a range of diseases, including rheumatoid arthritis, type 1 diabetes, hypothyroidism, multiple sclerosis, inflammatory bowel disease, Alzheimer's disease, and SARS-CoV-2 hospitalization (**Figure 6C**, **Supplemental Dataset X**). Many genetic risk loci are shared between diseases, and some GWAS focus on these

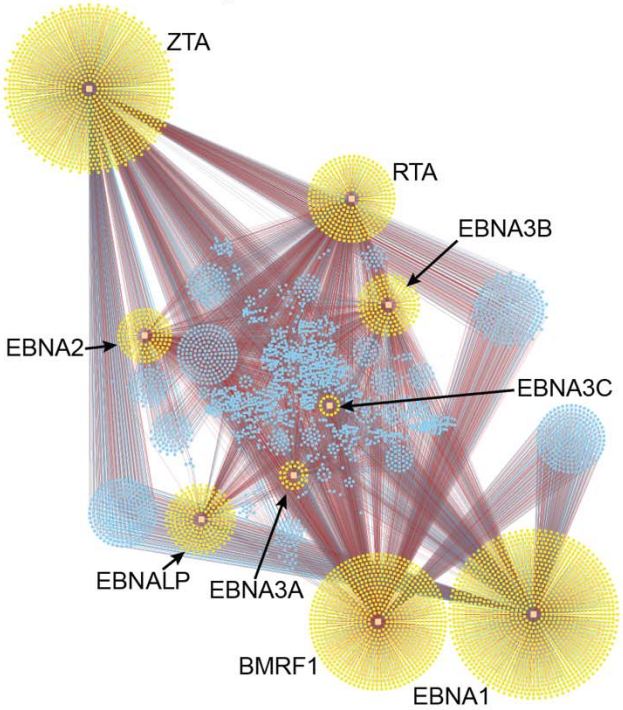
pleotropic risk loci. Impressively, BMRF1 was implicated at rheumatoid arthritis loci in five out of the six GWAS with significant BMRF1 enrichment (**Figure 6C**). When these BMRF1-bound variants are connected to genes through expression quantitative trait loci analysis (see Methods), a network emerges of interacting genes that impact processes implicated in the pathoetiology of rheumatoid arthritis, including complement activation (80), antigen presentation to T cells (81), chromatin assembly (82), and anti-viral responses (83) (**Figure 6D**, **Supplemental Dataset X**). The role of BMRF1 in mitochondrial biology might have direct implications on RA, as mitochondria are an important source of intracellular reactive oxygen species, and facilitate the activation of the inflammasome (84), which produces cytokines linked to disease symptoms in rheumatoid arthritis (85, 86).

Allele-dependent biology plays an important role in diseases with genetic components. When a genetic variant is heterozygous, both alleles are present in cells, allowing allele-dependent measurements. We used the MARIO computational tool (16) to identify allelic BMRF1 binding events to RA variants in our ChIP-seq data (see Methods). These analyses revealed two compelling allele-dependent BMRF1 binding examples at risk variants within a single rheumatoid arthritis risk locus (**Figure 6E**). Notably, one of these variants (rs140491) has previously been shown to result in allele-dependent reporter activity and allele-dependent CTCF binding in EBV infected B cells (87), suggesting a possible disease mechanism involving BMRF1 and CTCF at this locus. Collectively, these results implicate BMRF1 in rheumatoid arthritis-based genetic mechanisms.

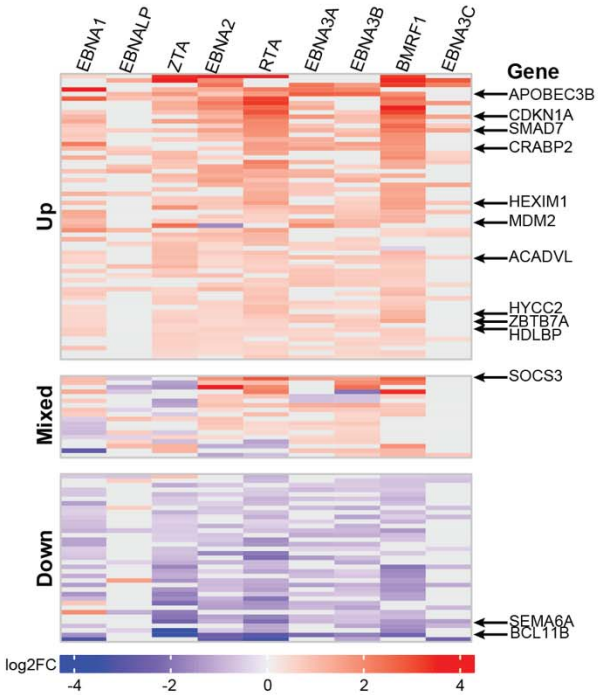
EBV vTRs consistently impact a coordinated network of host genes

Finally, we sought to obtain a systems level understanding of the complex EBV vTR human genome interaction landscape. To this end, we created a network linking each vTR to its likely human gene targets using gene expression and vTR binding data (see Methods) (**Figure 7A**). This network reveals that a substantial number of human genes (2,438) are regulated by multiple EBV vTRs. To further investigate this phenomenon, we examined the expression patterns of all human genes regulated by five or more vTRs (**Figure 7B**). Strikingly, the vast majority of these genes can be classified as either consistently upregulated (53%) or downregulated (31%) across vTRs, but not a mixture of the two (only 16%). The most heavily targeted genes (seven or more vTRs) include *SOCS3* (mixed regulation), *BCL11B (*down regulation), and *SMAD7* (up regulation). Dysregulation of SOCS3 function has been implicated in numerous chronic disease types including allergy, autoimmunity, and cancer (88). BCL11B controls cellular proliferation and promotes apoptosis in the context of viral infection (89). SMAD7 promotes the differentiation of numerous immune cells and has anti-inflammatory effects (90, 91); additionally, SMAD7 has roles in epithelial cell differentiation, which is directly involved in the EBV lytic cycle (92). Consistent with its upregulation by most EBV vTRs, SMAD7 expression is increased in viral infection (93). The interaction network of genes regulated and bound by five or more EBV vTRs also illustrates the connectivity of consistently regulated genes (**Figure 7C**). Intriguingly, antiviral effectors STAT1 and STAT3 are consistently upregulated. Collectively, our data reveal a vast human gene regulatory network coordinately regulated by a complex system of EBV vTR interactions with the human genome.

A. EBV vTR human gene interaction network



B. Genes regulated and bound by 5 or more EBV vTRs



C. Interaction network of genes regulated and bound by 5 or more EBV vTRs

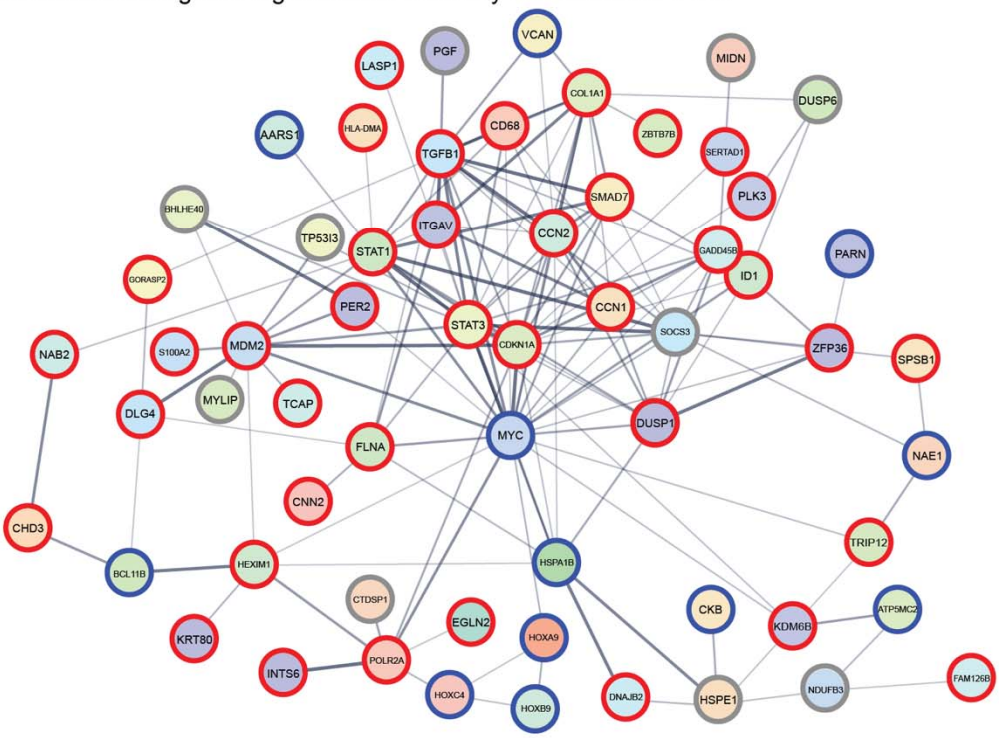


Figure 7. EBV vTRs consistently impact a coordinated network of host genes. A. Network representation of genes regulated by vTRs. Nodes indicate vTRs (large squares) or human genes (small squares) with vTR dependent gene expression and a proximal ChIP-seq peak (within 5 kb of TSS) for that vTR. Edges connect vTRs to regulated genes. Edges are colored based on the direction of differential expression (red = up, blue = down). Yellow genes are regulated by a single vTR. Blue genes are regulated by multiple vTRs. **B.** Heatmap of genes (n=118) regulated by five or more vTRs. Colors indicate $\log_2$ gene expression fold change. Gray cells indicate that the gene was not differentially expressed and bound by the vTR. Genes were initially k-means clustered (k=3), then hierarchically clustered using Ward's method within each k-means cluster. Columns were clustered using Ward's method. Clusters are named according to the direction of differential expression. Genes regulated by at least seven vTRs are indicated. **C.** Network showing interactions between the genes represented in panel B obtained from the STRING database. Colored rings indicate membership in the three clusters displayed in panel B (red=up, blue=down, gray=mixed).

A new resource for EBV-induced alterations to human gene regulation

The data produced in this study represent a novel and unique resource for examining the human gene regulatory mechanisms impacted by EBV vTRs. In total, our study contains 139 RNA-seq, 30 ATAC-seq, and 38 ChIP-seq datasets. To facilitate access to this massive resource, we have made the data available through three freely-available sources: 1) The Gene Expression Omnibus (GEO) database; 2) A UCSC Genome Browser (94) session (in progress); 3) A project-specific web page for easy browsing of the available datasets and easy download of the functional genomics data produced in this study, in a variety of formats (in progress). Constructs for the tagged and untagged EBV vTRs presented in this study have been deposited in Addgene. We anticipate this unprecedented resource will be of high value to the virology, bioinformatics, functional genomics, and disease genetics communities.

Discussion

In this study, we conducted the first systematic genome-scale analysis of EBV vTRs. Our experimental system allowed us to compare results across vTRs with isogenic expression, providing well-controlled conditions that allow us to assess causality. As anticipated, we observed binding at many previously characterized regions of the genome by various vTRs. For example, EBNA2 enriched for the motif of RBPJ, its known DNA binding partner, and bound near its known target *MAP3K8* (17, 26). Similarly, ZTA enriched its known motif and exhibited ChIP-seq peaks near established targets such as *FOSB, PDYN* and *SCIMP (49)*. Additionally, we observed robust chromatin changes induced by EBNA2 and ZTA expression, contextualizing our findings within the broader landscape of EBV functional genomics (16-18, 42-52).

Our data reveal extensive binding of EBV vTRs across the human genome, with substantial direct and indirect effects on human chromatin accessibility and gene expression. Each vTR regulates a different set of genes. Notably, upregulated genes are enriched in consistent pathways, as are downregulated genes, with minimal overlap between the two. This suggests that the virus has evolved to consistently up- or down-regulate particular biological processes. We also found that RBPJ genomic binding is stabilized by both EBNA2 and RTA, with many shared binding locations despite the distinct expression windows of EBNA2 and RTA in latency and lytic phases of EBV infection, respectively. Our findings thus suggest a "hand-off" mechanism between these proteins facilitated by RBPJ, deepening our understanding of the

strategies used by EBV to regulate human genes. A related virus, Kaposi Sarcoma Herpes Virus (KSHV), also expresses an RTA protein that is known to interact with RBPJ (95). Another vTR, BMRF1, emerged from this study as a potent regulator of human genes, occupying promoters, closing chromatin, and lowering gene expression for numerous genes. BMRF1 disease implications are profound, with enriched and allelic binding at genetic risk variants of rheumatoid arthritis.

Upon assessing the genomic sequences occupied by vTRs, we generated models of human transcription factors with which EBV vTRs might bind or compete. For example, places where vTRs occupied and opened chromatin were strongly enriched for AP-1 motifs, while loci bound and closed were enriched for CTCF. We further found that all nine vTRs' genomic binding events are enriched for either CTCF motifs (EBNA1, BMRF1, EBNA-LP) or AP-1 motifs (all others), but not both. When we assessed the context of EBV vTR genomic occupation, each of the vTRs, and especially ZTA was found to occupy the so called "dark matter" regions of the genome. These enigmatic regions have recently been shown to be occupied by the equally enigmatic human C2H2 zinc finger transcription factors (61).

A previous group identified important viral reprograming of the transcription in the lytic cycle that was driven by a viral pre-initiation complex of many of the less characterized vTRs that were also assessed in this study (96). The complex binds a TATT[T/A]AA, TATA box-like sequence and activates transcriptional initiation across tens of thousands of host genes. We only assess one EBV vTR at a time, which is a limitation

of our study. While we did not pursue all of the less characterized EBV vTRs we initially tested, we did measure some transcriptional changes from each of them. It is likely that many of the previously described impacts of these proteins are only realized when they come together in a complex.

Our findings have significant implications for understanding how EBV and other viruses impact the human gene regulatory landscape. By linking changes to specific vTRs, we provide a detailed map of EBV regulatory protein interactions with the human genome. Our network analysis elucidates that individual EBV vTRs regulate many human genes, painting a portrait of a complex regulatory landscape. The consistency in gene regulation and directionality across vTRs for a small core set of human genes is striking, highlighting the vast coordinated impact of EBV on the human genome.

In conclusion, this study represents a major advancement in the field of viral genomics, providing holistic insights into the regulatory mechanisms of EBV vTRs. Similar approaches applied to other less well-characterized viruses will also likely yield robust insights and discoveries. Our study offers a new resource for examining EBV-induced alterations to human gene regulation, with experimental reagents and data available and cross referenced on multiple platforms. These data will enable future research in virology, bioinformatics, functional genomics, and disease genetics, paving the way for new discoveries and therapeutic strategies.

Acknowledgements

This research was funded by National Institutes of Health (NIH) R01 HG010730, U01 AI130830, P01 AI15058501, and U24 HG013078 to M.T.W.; R01 AR073228, R01 NS099068, and R01 AI024717 to M.T.W. and L.C.K.; R01 AI148276, U01 HG011172, and U19 AI070235, and P30 AR070549 to L.C.K.; R01 AI164709 and DE033907 to B.E.G. We would like to acknowledge the assistance of Cincinnati Children's Hospital Medical Center Genomics Sequencing Facility (RRID SCR_022630) for high-throughput sequencing services, Ally Yang for experimental advice, and the members of the Kottyan and Weirauch labs for constructive feedback and discussion.

Author contributions

Conceptualization: LCK, MTW; Formal analysis: PJD, LEE, SP, KCMFV, XC; Funding acquisition: LCK, MTW; Investigation: MRH, AAD, SHJ, OD, CY, KAD, CF, HKH, OEG; Methodology: MRH, LCK, MTW; Resources: RR, BEG, BZ, TRH; Software: AV, KE; Supervision: MRH, LCK, MTW; Visualization: PJD, MRH, LEE, SP, LPL; Writing (original draft): PJD, MRH, LEE, LCK, MTW; Writing (review & editing): all authors.

Declaration of interests

The authors declare no competing interests.

Supplemental Figures

Supplemental Figure 1. Transcriptional assessment of less characterized EBV vTRs using transient expression of untagged proteins. **A.** All human genes with statistically reproducible differential expression induced by any vTR are shown. Values indicate

normalized gene expression levels that have been Z-transformed across rows (conditions). Columns and rows are clustered using Ward's method. **B.** Number of differentially expressed human genes in cells transiently expressing each untagged vTR. **C.** Principal component analysis of genes with statistically reproducible gene expression by any vTR.

Supplemental Figure 2. Differential gene expression from transient tests for EBV vTRs with and without GFP tags. For EBNALP, we only have the N-terminus construct because we were unable to successfully generate a construct with the C-terminus tag. For each vTR, all human genes with statistically reproducible differential expression induced by the untagged vTR are shown. Values indicate normalized gene expression levels that have been Z-transformed across rows (conditions). Columns and rows are clustered using Ward's method.

Supplemental Figure 3. AlphaFold3 (97) renderings of vTRs with and without a GFP tag. The coding sequences of the untagged and GFP fusions (N- and C-term) were submitted to the AlphaFold server to produce the predicted structures on the left of each panel. The confidence of the structure at each position is indicated by a dark green color in the chart on the right for each panel.

Supplemental Figure 4. Principal component analysis of human gene expression from transient tests for EBV vTRs with and without GFP tags. For each vTR, all human

genes with statistically reproducible differential expression induced by the untagged vTR are shown.

Supplemental Figure 5. Western blot confirmation of expected protein sizes. Confirmation that GFP fusion expression in stable cell lines produces proteins of the expected sizes for the various fusions, along with confirmation of equivalent protein loading through reprobing the blot with an antibody against beta-actin.

Supplemental Figure 6. Data driven determination of gene expression fold change thresholds. For each vTR, the number of differentially expressed genes was assessed at the indicated fold change. A clear inflection point emerges at 20% that is consistent across all examined vTRs.

Supplemental Figure 7. Complete results of functional annotations enriched among up and down regulated genes across cell lines stably expressing each vTR. Enriched pathways were identified using ToppGene (see Methods). Up to the top five most significant GO Slim Biological Process categories are included for each vTR.

Supplemental Figure 8. Overlap analysis of individual vTR ChIP-seq replicates. Each entry indicates the percent overlap (number of intersecting peaks divided by the number of peaks in the smaller peak set) between the given pair of vTR ChIP-seq replicates. Peak sets for each replicate contained peaks with q-value < 0.05. Rows and columns

were hierarchically clustered using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm.

Supplemental Figure 9. Genome browser screenshots showing examples of EBNA2 binding in our vTR system and in EBV infected B cells at select genes with EBNA2-dependent expression.

Supplemental Figure 10. Integration of vTR binding data (ChIP-seq) with vTR induced gene (RNA-seq) and chromatin (ATAC-seq) changes. **A.** Enrichment of overlap between ChIP-seq and up / down regulated genes (within 10 kilobases of the transcription start site). Enrichment for each vTR was assessed using RELI (see Methods). **B.** Enrichment of overlap between ChIP-seq peaks and opening / closing ATAC-seq peaks using RELI.

Supplemental Figure 11. Overlap analysis of individual αRBPJ ChIP-seq replicates. Each entry indicates the percent overlap (number of intersecting peaks divided by the number of peaks in the smaller peak set) between the given pair of αRBPJ ChIP-seq replicates. Peak sets for each replicate contain peaks with q-value < 0.05. Rows and columns were hierarchically clustered using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm

Supplemental Figure 12. Genome browser screenshots of select genes whose expression is upregulated in both EBNA2 and RTA expressing cells. Each example

exhibits overlapping binding events for EBNA2 and RTA, along with enhanced RBPJ binding in both EBNA2 and RTA expressing cells.

Supplemental Figure 13. Genome browser screenshots of select genes whose expression is altered by BMRF1, along with closing chromatin and BMRF1 binding at the promoter.

Supplemental Dataset X. Explanation of supplemental datasets.

References

1. Dunmire SK, Hogquist KA, Balfour HH. Infectious Mononucleosis. Curr Top Microbiol Immunol. 2015;390(Pt 1):211-40. Epub 2015/10/02. doi: 10.1007/978-3-319-22822-8_9. PubMed PMID: 26424648; PubMed Central PMCID: PMC4670567.
2. Rochford R, Moormann AM. Burkitt's Lymphoma. Curr Top Microbiol Immunol. 2015;390(Pt 1):267-85. Epub 2015/10/02. doi: 10.1007/978-3-319-22822-8_11. PubMed PMID: 26424650.
3. Vockerodt M, Cader FZ, Shannon-Lowe C, Murray P. Epstein-Barr virus and the origin of Hodgkin lymphoma. Chin J Cancer. 2014;33(12):591-7. Epub 2014/11/25. doi: 10.5732/cjc.014.10193. PubMed PMID: 25418190; PubMed Central PMCID: PMC4308654.
4. Tham T, Machado R, Russo DP, Herman SW, Teegala S, Costantino P. Viral markers in nasopharyngeal carcinoma: A systematic review and meta-analysis on the detection of p16(INK4a), human papillomavirus (HPV), and Ebstein-Barr virus (EBV). Am J Otolaryngol. 2021;42(1):102762. Epub 20201020. doi: 10.1016/j.amjoto.2020.102762. PubMed PMID: 33202328.
5. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science. 2022;375(6578):296-301. Epub 20220113. doi: 10.1126/science.abj8222. PubMed PMID: 35025605.
6. James JA, Kaufman KM, Farris AD, Taylor-Albert E, Lehman TJ, Harley JB. An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. J Clin Invest. 1997;100(12):3019-26. doi: 10.1172/JCI119856. PubMed PMID: 9399948; PubMed Central PMCID: PMC508514.
7. McClain MT, Heinlen LD, Dennis GJ, Roebuck J, Harley JB, James JA. Early events in lupus humoral autoimmunity suggest initiation through molecular mimicry. Nat Med. 2005;11(1):85-9. Epub 20041226. doi: 10.1038/nm1167. PubMed PMID: 15619631.

8. Banko A, Cirkovic A, Jeremic I, Basaric M, Grk M, Miskovic R, et al. Uncovering the Role of Epstein-Barr Virus Infection Markers for Remission in Rheumatoid Arthritis. Biomedicines. 2023;11(9). Epub 20230824. doi: 10.3390/biomedicines11092375. PubMed PMID: 37760816; PubMed Central PMCID: PMC10525384.
9. Erre GL, Mameli G, Cossu D, Muzzeddu B, Piras C, Paccagnini D, et al. Increased Epstein-Barr Virus DNA Load and Antibodies Against EBNA1 and EA in Sardinian Patients with Rheumatoid Arthritis. Viral Immunol. 2015;28(7):385-90. Epub 20150617. doi: 10.1089/vim.2015.0035. PubMed PMID: 26083265.
10. Robinson WH, Younis S, Love ZZ, Steinman L, Lanz TV. Epstein-Barr virus as a potentiator of autoimmune diseases. Nat Rev Rheumatol. 2024;20(11):729-40. Epub 20241010. doi: 10.1038/s41584-024-01167-9. PubMed PMID: 39390260.
11. Nandy A, Petralia F, Porter C, Elledge S, Viruses, Ibd Study g, et al. Epstein-Barr Virus Exposure Precedes Crohn;s Disease Development. Gastroenterology. 2025. Epub 20250224. doi: 10.1053/j.gastro.2025.01.247. PubMed PMID: 40010579.
12. Cao Y, Xie L, Shi F, Tang M, Li Y, Hu J, et al. Targeting the signaling in Epstein-Barr virus-associated diseases: mechanism, regulation, and clinical study. Signal Transduct Target Ther. 2021;6(1):15. Epub 20210112. doi: 10.1038/s41392-020-00376-4. PubMed PMID: 33436584; PubMed Central PMCID: PMC7801793.
13. Lu J, Tang M, Li H, Xu Z, Weng X, Li J, et al. EBV-LMP1 suppresses the DNA damage response through DNA-PK/AMPK signaling to promote radioresistance in nasopharyngeal carcinoma. Cancer Lett. 2016;380(1):191-200. Epub 20160530. doi: 10.1016/j.canlet.2016.05.032. PubMed PMID: 27255972.
14. Xiao L, Hu ZY, Dong X, Tan Z, Li W, Tang M, et al. Targeting Epstein-Barr virus oncoprotein LMP1-mediated glycolysis sensitizes nasopharyngeal carcinoma to radiation therapy. Oncogene. 2014;33(37):4568-78. Epub 20140324. doi: 10.1038/onc.2014.32. PubMed PMID: 24662831; PubMed Central PMCID: PMC4162460.
15. Lanz TV, Brewer RC, Ho PP, Moon JS, Jude KM, Fernandez D, et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. Nature. 2022;603(7900):321-7. Epub 20220124. doi: 10.1038/s41586-022-04432-7. PubMed PMID: 35073561; PubMed Central PMCID: PMC9382663.
16. Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet. 2018;50(5):699-707. Epub 20180416. doi: 10.1038/s41588-018-0102-3. PubMed PMID: 29662164; PubMed Central PMCID: PMC6022759.
17. Hong T, Parameswaran S, Donmez OA, Miller D, Forney C, Lape M, et al. Epstein-Barr virus nuclear antigen 2 extensively rewires the human chromatin landscape at autoimmune risk loci. Genome Res. 2021;31(12):2185-98. Epub 20211119. doi: 10.1101/gr.264705.120. PubMed PMID: 34799401; PubMed Central PMCID: PMC8647835.
18. Viel K, Parameswaran S, Donmez OA, Forney CR, Hass MR, Yin C, 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(1):273. Epub 20240312. doi: 10.1186/s12864-024-10183-8. PubMed PMID: 38475709; PubMed Central PMCID: PMC10935964.

19. 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. Epub 20210903. doi: 10.1016/j.ebiom.2021.103572. PubMed PMID: 34488019; PubMed Central PMCID: PMC8426200.
20. Damania B, Kenney SC, Raab-Traub N. Epstein-Barr virus: Biology and clinical disease. Cell. 2022;185(20):3652-70. Epub 20220915. doi: 10.1016/j.cell.2022.08.026. PubMed PMID: 36113467; PubMed Central PMCID: PMC9529843.
21. Ding W, Wang C, Narita Y, Wang H, Leong MML, Huang A, et al. The Epstein-Barr Virus Enhancer Interaction Landscapes in Virus-Associated Cancer Cell Lines. J Virol. 2022;96(18):e0073922. Epub 20220912. doi: 10.1128/jvi.00739-22. PubMed PMID: 36094314; PubMed Central PMCID: PMC9517713.
22. Guo R, Liang JH, Zhang Y, Lutchenkov M, Li Z, Wang Y, et al. Methionine metabolism controls the B cell EBV epigenome and viral latency. Cell Metab. 2022;34(9):1280-97 e9. doi: 10.1016/j.cmet.2022.08.008. PubMed PMID: 36070681; PubMed Central PMCID: PMC9482757.
23. Liu X, Hong T, Parameswaran S, Ernst K, Marazzi I, Weirauch MT, et al. Human Virus Transcriptional Regulators. Cell. 2020;182(1):24-37. doi: 10.1016/j.cell.2020.06.023. PubMed PMID: 32649876; PubMed Central PMCID: PMC7346790.
24. Beer S, Wange LE, Zhang X, Kuklik-Roos C, Enard W, Hammerschmidt W, et al. EBNA2-EBF1 complexes promote MYC expression and metabolic processes driving S-phase progression of Epstein-Barr virus-infected B cells. Proc Natl Acad Sci U S A. 2022;119(30):e2200512119. Epub 20220720. doi: 10.1073/pnas.2200512119. PubMed PMID: 35857872; PubMed Central PMCID: PMC9335265.
25. Feng J, Zhang P, Yao P, Zhang H. EBNA2 mediates lipid metabolism and tumorigenesis through activation of ATF4 pathway. Am J Cancer Res. 2023;13(4):1363-76. Epub 20230415. PubMed PMID: 37168348; PubMed Central PMCID: PMC10164800.
26. Glaser LV, Rieger S, Thumann S, Beer S, Kuklik-Roos C, Martin DE, et al. EBF1 binds to EBNA2 and promotes the assembly of EBNA2 chromatin complexes in B cells. PLoS Pathog. 2017;13(10):e1006664. Epub 20171002. doi: 10.1371/journal.ppat.1006664. PubMed PMID: 28968461; PubMed Central PMCID: PMC5638620.
27. Leopizzi M, Mundo L, Messina E, Campolo F, Lazzi S, Angeloni A, et al. Epstein-Barr virus-encoded EBNA2 downregulates ICOSL by inducing miR-24 in B-cell lymphoma. Blood. 2024;143(5):429-43. doi: 10.1182/blood.2023021346. PubMed PMID: 37847858; PubMed Central PMCID: PMC10862363.
28. Lu F, Wiedmer A, Martin KA, Wickramasinghe P, Kossenkov AV, Lieberman PM. Coordinate Regulation of TET2 and EBNA2 Controls the DNA Methylation State of Latent Epstein-Barr Virus. J Virol. 2017;91(20). Epub 20170927. doi: 10.1128/JVI.00804-17. PubMed PMID: 28794029; PubMed Central PMCID: PMC5625499.
29. Muhe J, Wang F. Species-specific functions of Epstein-Barr virus nuclear antigen 2 (EBNA2) reveal dual roles for initiation and maintenance of B cell immortalization. PLoS Pathog. 2017;13(12):e1006772. Epub 20171220. doi:

10.1371/journal.ppat.1006772. PubMed PMID: 29261800; PubMed Central PMCID: PMC5754137.
30. Ponnusamy R, Khatri R, Correia PB, Wood CD, Mancini EJ, Farrell PJ, et al. Increased association between Epstein-Barr virus EBNA2 from type 2 strains and the transcriptional repressor BS69 restricts EBNA2 activity. PLoS Pathog. 2019;15(7):e1007458. Epub 20190708. doi: 10.1371/journal.ppat.1007458. PubMed PMID: 31283782; PubMed Central PMCID: PMC6638984.
31. Su C, Lu F, Soldan SS, Lamontagne RJ, Tang HY, Napoletani G, et al. EBNA2 driven enhancer switching at the CIITA-DEXI locus suppresses HLA class II gene expression during EBV infection of B-lymphocytes. PLoS Pathog. 2021;17(8):e1009834. Epub 20210805. doi: 10.1371/journal.ppat.1009834. PubMed PMID: 34352044; PubMed Central PMCID: PMC8370649.
32. Yanagi Y, Okuno Y, Narita Y, Masud H, Watanabe T, Sato Y, et al. RNAseq analysis identifies involvement of EBNA2 in PD-L1 induction during Epstein-Barr virus infection of primary B cells. Virology. 2021;557:44-54. Epub 20210221. doi: 10.1016/j.virol.2021.02.004. PubMed PMID: 33639481.
33. Zhang J, Sommermann T, Li X, Gieselmann L, de la Rosa K, Stecklum M, et al. LMP1 and EBNA2 constitute a minimal set of EBV genes for transformation of human B cells. Front Immunol. 2023;14:1331730. Epub 20231219. doi: 10.3389/fimmu.2023.1331730. PubMed PMID: 38169736; PubMed Central PMCID: PMC10758421.
34. Zheng X, Wang J, Zhang X, Fu Y, Peng Q, Lu J, et al. RNA m(6) A methylation regulates virus-host interaction and EBNA2 expression during Epstein-Barr virus infection. Immun Inflamm Dis. 2021;9(2):351-62. Epub 20210112. doi: 10.1002/iid3.396. PubMed PMID: 33434416; PubMed Central PMCID: PMC8127537.
35. Bernaudat F, Gustems M, Gunther J, Oliva MF, Buschle A, Gobel C, et al. Structural basis of DNA methylation-dependent site selectivity of the Epstein-Barr virus lytic switch protein ZEBRA/Zta/BZLF1. Nucleic Acids Res. 2022;50(1):490-511. doi: 10.1093/nar/gkab1183. PubMed PMID: 34893887; PubMed Central PMCID: PMC8754650.
36. Duraisamy B, Pramanik D. Influence of DNA Sequences on the Thermodynamic and Structural Stability of the ZTA Transcription Factor horizontal line DNA Complex: An All-Atom Molecular Dynamics Study. J Phys Chem B. 2025;129(18):4282-97. Epub 20250423. doi: 10.1021/acs.jpcb.4c07713. PubMed PMID: 40266646.
37. Godfrey A, Osborn K, Sinclair AJ. Interaction sites of the Epstein-Barr virus Zta transcription factor with the host genome in epithelial cells. Access Microbiol. 2021;3(11):000282. Epub 20211126. doi: 10.1099/acmi.0.000282. PubMed PMID: 35018326; PubMed Central PMCID: PMC8742585.
38. Ray S, Tillo D, Assad N, Ufot A, Porollo A, Durell SR, et al. Altering the Double-Stranded DNA Specificity of the bZIP Domain of Zta with Site-Directed Mutagenesis at N182. ACS Omega. 2022;7(1):129-39. Epub 20211228. doi: 10.1021/acsomega.1c04148. PubMed PMID: 35036684; PubMed Central PMCID: PMC8756438.
39. Ray S, Tillo D, Ufot A, Assad N, Durell S, Vinson C. bZIP Dimers CREB1, ATF2, Zta, ATF3|cJun, and cFos|cJun Prefer to Bind to Some Double-Stranded DNA Sequences Containing 5-Formylcytosine and 5-Carboxylcytosine. Biochemistry.

2020;59(38):3529-40. Epub 20200920. doi: 10.1021/acs.biochem.0c00475. PubMed PMID: 32902247; PubMed Central PMCID: PMC8842497.
40. Zheng H, Zhao H, Zhang W, Zhang Y, Zhang Y, Ye B, et al. Epstein-Barr virus immediate-early protein Zta mediates the proliferation and migration of HER2-overexpressing cancer cells. Arch Virol. 2023;168(5):150. Epub 20230503. doi: 10.1007/s00705-023-05774-x. PubMed PMID: 37133552.
41. Zhou Y, Heesom K, Osborn K, AlMohammed R, Sweet SM, Sinclair AJ. Identifying the Cellular Interactome of Epstein-Barr Virus Lytic Regulator Zta Reveals Cellular Targets Contributing to Viral Replication. J Virol. 2020;94(3). Epub 20200117. doi: 10.1128/JVI.00927-19. PubMed PMID: 31694936; PubMed Central PMCID: PMC7000981.
42. Gunnell A, Webb HM, Wood CD, McClellan MJ, Wichaidit B, Kempkes B, et al. RUNX super-enhancer control through the Notch pathway by Epstein-Barr virus transcription factors regulates B cell growth. Nucleic Acids Res. 2016;44(10):4636-50. Epub 20160215. doi: 10.1093/nar/gkw085. PubMed PMID: 26883634; PubMed Central PMCID: PMC4889917.
43. Jiang S, Willox B, Zhou H, Holthaus AM, Wang A, Shi TT, et al. Epstein-Barr virus nuclear antigen 3C binds to BATF/IRF4 or SPI1/IRF4 composite sites and recruits Sin3A to repress CDKN2A. Proc Natl Acad Sci U S A. 2014;111(1):421-6. Epub 20131216. doi: 10.1073/pnas.1321704111. PubMed PMID: 24344258; PubMed Central PMCID: PMC3890834.
44. Lu F, Wikramasinghe P, Norseen J, Tsai K, Wang P, Showe L, et al. Genome-wide analysis of host-chromosome binding sites for Epstein-Barr Virus Nuclear Antigen 1 (EBNA1). Virol J. 2010;7:262. Epub 20101007. doi: 10.1186/1743-422X-7-262. PubMed PMID: 20929547; PubMed Central PMCID: PMC2964674.
45. McClellan MJ, Wood CD, Ojeniyi O, Cooper TJ, Kanhere A, Arvey A, et al. Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming. PLoS Pathog. 2013;9(9):e1003636. Epub 20130912. doi: 10.1371/journal.ppat.1003636. PubMed PMID: 24068937; PubMed Central PMCID: PMC3771879.
46. Moquin SA, Thomas S, Whalen S, Warburton A, Fernandez SG, McBride AA, et al. The Epstein-Barr Virus Episome Maneuvers between Nuclear Chromatin Compartments during Reactivation. J Virol. 2018;92(3). Epub 20180117. doi: 10.1128/JVI.01413-17. PubMed PMID: 29142137; PubMed Central PMCID: PMC5774889.
47. Paschos K, Bazot Q, Ho G, Parker GA, Lees J, Barton G, et al. Core binding factor (CBF) is required for Epstein-Barr virus EBNA3 proteins to regulate target gene expression. Nucleic Acids Res. 2017;45(5):2368-83. doi: 10.1093/nar/gkw1167. PubMed PMID: 27903901; PubMed Central PMCID: PMC5389572.
48. Portal D, Zhou H, Zhao B, Kharchenko PV, Lowry E, Wong L, et al. Epstein-Barr virus nuclear antigen leader protein localizes to promoters and enhancers with cell transcription factors and EBNA2. Proc Natl Acad Sci U S A. 2013;110(46):18537-42. Epub 20131028. doi: 10.1073/pnas.1317608110. PubMed PMID: 24167291; PubMed Central PMCID: PMC3832032.
49. Ramasubramanyan S, Osborn K, Al-Mohammad R, Naranjo Perez-Fernandez IB, Zuo J, Balan N, et al. Epstein-Barr virus transcription factor Zta acts through distal

regulatory elements to directly control cellular gene expression. Nucleic Acids Res. 2015;43(7):3563-77. Epub 20150316. doi: 10.1093/nar/gkv212. PubMed PMID: 25779048; PubMed Central PMCID: PMC4402532.
50. Tempera I, De Leo A, Kossenkov AV, Cesaroni M, Song H, Dawany N, et al. Identification of MEF2B, EBF1, and IL6R as Direct Gene Targets of Epstein-Barr Virus (EBV) Nuclear Antigen 1 Critical for EBV-Infected B-Lymphocyte Survival. J Virol. 2016;90(1):345-55. Epub 20151014. doi: 10.1128/JVI.02318-15. PubMed PMID: 26468528; PubMed Central PMCID: PMC4702570.
51. Wang A, Welch R, Zhao B, Ta T, Keles S, Johannsen E. Epstein-Barr Virus Nuclear Antigen 3 (EBNA3) Proteins Regulate EBNA2 Binding to Distinct RBPJ Genomic Sites. J Virol. 2015;90(6):2906-19. Epub 20151230. doi: 10.1128/JVI.02737-15. PubMed PMID: 26719268; PubMed Central PMCID: PMC4810642.
52. Zhao B, Zou J, Wang H, Johannsen E, Peng CW, Quackenbush J, 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(36):14902-7. Epub 20110711. doi: 10.1073/pnas.1108892108. PubMed PMID: 21746931; PubMed Central PMCID: PMC3169132.
53. Arvey A, Tempera I, Tsai K, Chen HS, Tikhmyanova N, Klichinsky M, et al. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. Cell Host Microbe. 2012;12(2):233-45. doi: 10.1016/j.chom.2012.06.008. PubMed PMID: 22901543; PubMed Central PMCID: PMC3424516.
54. Schmidt SC, Jiang S, Zhou H, Willox B, Holthaus AM, Kharchenko PV, et al. Epstein-Barr virus nuclear antigen 3A partially coincides with EBNA3C genome-wide and is tethered to DNA through BATF complexes. Proc Natl Acad Sci U S A. 2015;112(2):554-9. Epub 20141224. doi: 10.1073/pnas.1422580112. PubMed PMID: 25540416; PubMed Central PMCID: PMC4299249.
55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. Epub 20210715. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PubMed Central PMCID: PMC8371605.
56. Holley-Guthrie EA, Seaman WT, Bhende P, Merchant JL, Kenney SC. The Epstein-Barr virus protein BMRF1 activates gastrin transcription. J Virol. 2005;79(2):745-55. doi: 10.1128/JVI.79.2.745-755.2005. PubMed PMID: 15613302; PubMed Central PMCID: PMC538557.
57. Hitz BC, Jin-Wook L, Jolanki O, Kagda MS, Graham K, Sud P, et al. The ENCODE Uniform Analysis Pipelines. bioRxiv. 2023. Epub 20230406. doi: 10.1101/2023.04.04.535623. PubMed PMID: 37066421; PubMed Central PMCID: PMC10104020.
58. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48(D1):D882-D9. doi: 10.1093/nar/gkz1062. PubMed PMID: 31713622; PubMed Central PMCID: PMC7061942.
59. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature.

2014;507(7493):455-61. doi: 10.1038/nature12787. PubMed PMID: 24670763; PubMed Central PMCID: PMC5215096.
60. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215-6. Epub 20120228. doi: 10.1038/nmeth.1906. PubMed PMID: 22373907; PubMed Central PMCID: PMC3577932.
61. Razavi R, Fathi A, Yellan I, Brechalov A, Laverty KU, Jolma A, et al. Extensive binding of uncharacterized human transcription factors to genomic dark matter. bioRxiv. 2024. Epub 20241112. doi: 10.1101/2024.11.11.622123. PubMed PMID: 39605320; PubMed Central PMCID: PMC11601254.
62. Kenney SC, Mertz JE. Regulation of the latent-lytic switch in Epstein-Barr virus. Semin Cancer Biol. 2014;26:60-8. Epub 20140120. doi: 10.1016/j.semcancer.2014.01.002. PubMed PMID: 24457012; PubMed Central PMCID: PMC4048781.
63. Zimber-Strobl U, Strobl LJ. EBNA2 and Notch signalling in Epstein-Barr virus mediated immortalization of B lymphocytes. Semin Cancer Biol. 2001;11(6):423-34. doi: 10.1006/scbi.2001.0409. PubMed PMID: 11669604.
64. Papugani A, Coleman T, Jones C, Zhang L. The interaction between KSHV RTA and cellular RBP-Jkappa and their subsequent DNA binding are not sufficient for activation of RBP-Jkappa. Virus Res. 2008;131(1):1-7. Epub 20070911. doi: 10.1016/j.virusres.2007.07.019. PubMed PMID: 17850910; PubMed Central PMCID: PMC2225583.
65. Shen RR, Hahn WC. Emerging roles for the non-canonical IKKs in cancer. Oncogene. 2011;30(6):631-41. Epub 20101101. doi: 10.1038/onc.2010.493. PubMed PMID: 21042276; PubMed Central PMCID: PMC3235643.
66. Schmid S, Sachs D, tenOever BR. Mitogen-activated protein kinase-mediated licensing of interferon regulatory factor 3/7 reinforces the cell response to virus. J Biol Chem. 2014;289(1):299-311. Epub 20131125. doi: 10.1074/jbc.M113.519934. PubMed PMID: 24275658; PubMed Central PMCID: PMC3879553.
67. Ma S, Wan X, Deng Z, Shi L, Hao C, Zhou Z, et al. Epigenetic regulator CXXC5 recruits DNA demethylase Tet2 to regulate TLR7/9-elicited IFN response in pDCs. J Exp Med. 2017;214(5):1471-91. Epub 20170417. doi: 10.1084/jem.20161149. PubMed PMID: 28416650; PubMed Central PMCID: PMC5413332.
68. Lin CS, Kuo HH, Chen JY, Yang CS, Wang WB. Epstein-barr virus nuclear antigen 2 retards cell growth, induces p21(WAF1) expression, and modulates p53 activity post-translationally. J Mol Biol. 2000;303(1):7-23. doi: 10.1006/jmbi.2000.4121. PubMed PMID: 11021966.
69. Guan J, Fan Y, Wang S, Zhou F. Functions of MAP3Ks in antiviral immunity. Immunol Res. 2023;71(6):814-32. Epub 20230607. doi: 10.1007/s12026-023-09401-4. PubMed PMID: 37286768; PubMed Central PMCID: PMC10247270.
70. Contreras AU, Mebratu Y, Delgado M, Montano G, Hu CA, Ryter SW, et al. Deacetylation of p53 induces autophagy by suppressing Bmf expression. J Cell Biol. 2013;201(3):427-37. doi: 10.1083/jcb.201205064. PubMed PMID: 23629966; PubMed Central PMCID: PMC3639396.
71. Lu F, Chen HS, Kossenkov AV, DeWispeleare K, Won KJ, Lieberman PM. EBNA2 Drives Formation of New Chromosome Binding Sites and Target Genes for B-

Cell Master Regulatory Transcription Factors RBP-jkappa and EBF1. PLoS Pathog. 2016;12(1):e1005339. Epub 20160111. doi: 10.1371/journal.ppat.1005339. PubMed PMID: 26752713; PubMed Central PMCID: PMC4709166.
72. Wang H, Zang C, Taing L, Arnett KL, Wong YJ, Pear WS, et al. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. Proc Natl Acad Sci U S A. 2014;111(2):705-10. Epub 20131227. doi: 10.1073/pnas.1315023111. PubMed PMID: 24374627; PubMed Central PMCID: PMC3896193.
73. Portal D, Zhao B, Calderwood MA, Sommermann T, Johannsen E, Kieff E. EBV nuclear antigen EBNALP dismisses transcription repressors NCoR and RBPJ from enhancers and EBNA2 increases NCoR-deficient RBPJ DNA binding. Proc Natl Acad Sci U S A. 2011;108(19):7808-13. Epub 20110425. doi: 10.1073/pnas.1104991108. PubMed PMID: 21518914; PubMed Central PMCID: PMC3093489.
74. Lilly B, Kennard S. Differential gene expression in a coculture model of angiogenesis reveals modulation of select pathways and a role for Notch signaling. Physiol Genomics. 2009;36(2):69-78. Epub 20081104. doi: 10.1152/physiolgenomics.90318.2008. PubMed PMID: 18984672; PubMed Central PMCID: PMC2636923.
75. Triepels RH, van den Heuvel LP, Loeffen JL, Buskens CA, Smeets RJ, Rubio Gozalbo ME, et al. Leigh syndrome associated with a mutation in the NDUFS7 (PSST) nuclear encoded subunit of complex I. Ann Neurol. 1999;45(6):787-90. doi: 10.1002/1531-8249(199906)45:6<787::aid-ana13>3.0.co;2-6. PubMed PMID: 10360771.
76. Lebon S, Rodriguez D, Bridoux D, Zerrad A, Rotig A, Munnich A, et al. A novel mutation in the human complex I NDUFS7 subunit associated with Leigh syndrome. Mol Genet Metab. 2007;90(4):379-82. Epub 20070201. doi: 10.1016/j.ymgme.2006.12.007. PubMed PMID: 17275378.
77. Hong T, Parameswaran S, Donmez OA, Miller D, Forney C, Lape M, et al. Epstein-Barr virus nuclear antigen 2 extensively rewires the human chromatin landscape at autoimmune risk loci. Genome Res. 2021. Epub 2021/11/21. doi: 10.1101/gr.264705.120. PubMed PMID: 34799401; PubMed Central PMCID: PMC8647835.
78. Lu X, Chen X, Forney C, Donmez O, Miller D, Parameswaran S, et al. Global discovery of lupus genetic risk variant allelic enhancer activity. Nat Commun. 2021;12(1):1611. Epub 2021/03/14. doi: 10.1038/s41467-021-21854-5. PubMed PMID: 33712590; PubMed Central PMCID: PMC7955039.
79. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D12. Epub 2018/11/18. doi: 10.1093/nar/gky1120. PubMed PMID: 30445434; PubMed Central PMCID: PMC6323933.
80. Okroj M, Heinegard D, Holmdahl R, Blom AM. Rheumatoid arthritis and the complement system. Ann Med. 2007;39(7):517-30. doi: 10.1080/07853890701477546. PubMed PMID: 17852027.
81. Cope AP, Schulze-Koops H, Aringer M. The central role of T cells in rheumatoid arthritis. Clin Exp Rheumatol. 2007;25(5 Suppl 46):S4-11. PubMed PMID: 17977483.

82. Weinand K, Sakaue S, Nathan A, Jonsson AH, Zhang F, Watts GFM, et al. The chromatin landscape of pathogenic transcriptional cell states in rheumatoid arthritis. Nat Commun. 2024;15(1):4650. Epub 20240531. doi: 10.1038/s41467-024-48620-7. PubMed PMID: 38821936; PubMed Central PMCID: PMC11143375.
83. Edilova MI, Akram A, Abdul-Sater AA. Innate immunity drives pathogenesis of rheumatoid arthritis. Biomed J. 2021;44(2):172-82. Epub 20200708. doi: 10.1016/j.bj.2020.06.010. PubMed PMID: 32798211; PubMed Central PMCID: PMC8178572.
84. Yin H, Liu N, Sigdel KR, Duan L. Role of NLRP3 Inflammasome in Rheumatoid Arthritis. Front Immunol. 2022;13:931690. Epub 20220627. doi: 10.3389/fimmu.2022.931690. PubMed PMID: 35833125; PubMed Central PMCID: PMC9271572.
85. Clayton SA, MacDonald L, Kurowska-Stolarska M, Clark AR. Mitochondria as Key Players in the Pathogenesis and Treatment of Rheumatoid Arthritis. Front Immunol. 2021;12:673916. Epub 20210429. doi: 10.3389/fimmu.2021.673916. PubMed PMID: 33995417; PubMed Central PMCID: PMC8118696.
86. Li JC, Huang WS, Yang DH, He QF, Sun W. Assessing causality between mitochondrial-associated proteins with musculoskeletal diseases: A Mendelian randomization study. Medicine (Baltimore). 2025;104(10):e41731. doi: 10.1097/MD.0000000000041731. PubMed PMID: 40068079; PubMed Central PMCID: PMC11903026.
87. Gopalakrishnan J, Tessneer KL, Fu Y, Pasula S, Pelikan RC, Kelly JA, et al. Variants on the UBE2L3/YDJC Autoimmune Disease Risk Haplotype Increase UBE2L3 Expression by Modulating CCCTC-Binding Factor and YY1 Binding. Arthritis Rheumatol. 2022;74(1):163-73. Epub 20211213. doi: 10.1002/art.41925. PubMed PMID: 34279042; PubMed Central PMCID: PMC8712360.
88. Yin Y, Liu W, Dai Y. SOCS3 and its role in associated diseases. Hum Immunol. 2015;76(10):775-80. Epub 20150930. doi: 10.1016/j.humimm.2015.09.037. PubMed PMID: 26429311.
89. Qiu L, Yang T, Guo Q, Hua T, Bi Y, Chu P, et al. C(2)H(2)-type zinc-finger protein BCL11B suppresses avian Leukosis virus subgroup J replication by regulating apoptosis. Int J Biol Macromol. 2024;275(Pt 1):133644. Epub 20240702. doi: 10.1016/j.ijbiomac.2024.133644. PubMed PMID: 38964687.
90. Zhu L, Chen S, Chen Y. Unraveling the biological functions of Smad7 with mouse models. Cell Biosci. 2011;1:44. Epub 20111228. doi: 10.1186/2045-3701-1-44. PubMed PMID: 22204639; PubMed Central PMCID: PMC3275527.
91. Hu Y, He J, He L, Xu B, Wang Q. Expression and function of Smad7 in autoimmune and inflammatory diseases. J Mol Med (Berl). 2021;99(9):1209-20. Epub 20210531. doi: 10.1007/s00109-021-02083-1. PubMed PMID: 34059951; PubMed Central PMCID: PMC8367892.
92. Iempridee T, Das S, Xu I, Mertz JE. Transforming growth factor beta-induced reactivation of Epstein-Barr virus involves multiple Smad-binding elements cooperatively activating expression of the latent-lytic switch BZLF1 gene. J Virol. 2011;85(15):7836-48. Epub 20110518. doi: 10.1128/JVI.01197-10. PubMed PMID: 21593157; PubMed Central PMCID: PMC3147924.

93. Liu N, Jiao T, Huang Y, Liu W, Li Z, Ye X. Hepatitis B virus regulates apoptosis and tumorigenesis through the microRNA-15a-Smad7-transforming growth factor beta pathway. J Virol. 2015;89(5):2739-49. Epub 20141224. doi: 10.1128/JVI.02784-14. PubMed PMID: 25540364; PubMed Central PMCID: PMC4325757.
94. Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2025 update. Nucleic Acids Res. 2025;53(D1):D1243-D9. doi: 10.1093/nar/gkae974. PubMed PMID: 39460617; PubMed Central PMCID: PMC11701590.
95. Liang Y, Chang J, Lynch SJ, Lukac DM, Ganem D. The lytic switch protein of KSHV activates gene expression via functional interaction with RBP-Jkappa (CSL), the target of the Notch signaling pathway. Genes Dev. 2002;16(15):1977-89. doi: 10.1101/gad.996502. PubMed PMID: 12154127; PubMed Central PMCID: PMC186409.
96. Ungerleider NA, Roberts C, O'Grady TM, Nguyen TT, Baddoo M, Wang J, et al. Viral reprogramming of host transcription initiation. Nucleic Acids Res. 2024;52(9):5016-32. doi: 10.1093/nar/gkae175. PubMed PMID: 38471819; PubMed Central PMCID: PMC11109974.
97. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500. Epub 20240508. doi: 10.1038/s41586-024-07487-w. PubMed PMID: 38718835; PubMed Central PMCID: PMC11168924.
98. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet. 2021;53(9):1290-9. Epub 20210906. doi: 10.1038/s41588-021-00924-w. PubMed PMID: 34493866; PubMed Central PMCID: PMC8423625.
99. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015;12(10):963-5. Epub 20150817. doi: 10.1038/nmeth.3542. PubMed PMID: 26280331; PubMed Central PMCID: PMC4589892.
100. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14(10):959-62. Epub 20170828. doi: 10.1038/nmeth.4396. PubMed PMID: 28846090; PubMed Central PMCID: PMC5623106.
101. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21 9 1- 9 9. Epub 20150105. doi: 10.1002/0471142727.mb2129s109. PubMed PMID: 25559105; PubMed Central PMCID: PMC4374986.
102. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7. Epub 20121015. doi: 10.1093/bioinformatics/bts611. PubMed PMID: 23071270.
103. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. Epub 20121025. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PubMed Central PMCID: PMC3530905.
104. Mudge JM, Carbonell-Sala S, Diekhans M, Martinez JG, Hunt T, Jungreis I, et al. GENCODE 2025: reference gene annotation for human and mouse. Nucleic Acids Res.

2025;53(D1):D966-D75. doi: 10.1093/nar/gkae1078. PubMed PMID: 39565199; PubMed Central PMCID: PMC11701607.
105. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 20090608. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMC2723002.
106. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184-5. Epub 2012/06/30. doi: 10.1093/bioinformatics/bts356. PubMed PMID: 22743226.
107. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292-4. Epub 2015/10/03. doi: 10.1093/bioinformatics/btv566. PubMed PMID: 26428292; PubMed Central PMCID: PMC4708105.
108. Sayols S, Scherzinger D, Klein H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. BMC Bioinformatics. 2016;17(1):428. Epub 2016/10/23. doi: 10.1186/s12859-016-1276-2. PubMed PMID: 27769170; PubMed Central PMCID: PMC5073875.
109. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. Nat Methods. 2013;10(4):325-7. Epub 20130224. doi: 10.1038/nmeth.2375. PubMed PMID: 23435259; PubMed Central PMCID: PMC3612374.
110. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-9. Epub 20170306. doi: 10.1038/nmeth.4197. PubMed PMID: 28263959; PubMed Central PMCID: PMC5600148.
111. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
112. Chen J, Xu H, Aronow BJ, Jegga AG. Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics. 2007;8:392. Epub 20071016. doi: 10.1186/1471-2105-8-392. PubMed PMID: 17939863; PubMed Central PMCID: PMC2194797.
113. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1). doi: 10.1093/genetics/iyad031. PubMed PMID: 36866529; PubMed Central PMCID: PMC10158837.
114. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9. doi: 10.1038/75556. PubMed PMID: 10802651; PubMed Central PMCID: PMC3037419.
115. R. K. pheatmap: Pretty Heatmaps. R package version 1.0.12. 2025.
116. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5. Epub 20160413. doi: 10.1093/nar/gkw257. PubMed PMID: 27079975; PubMed Central PMCID: PMC4987876.

117. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92. Epub 20120419. doi: 10.1093/bib/bbs017. PubMed PMID: 22517427; PubMed Central PMCID: PMC3603213.
118. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. Epub 20100128. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278; PubMed Central PMCID: PMC2832824.
119. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMC2898526.
120. Lambert SA, Yang AWH, Sasse A, Cowley G, Albu M, Caddick MX, et al. Similarity regression predicts evolution of transcription factor sequence specificity. Nat Genet. 2019;51(6):981-9. Epub 20190527. doi: 10.1038/s41588-019-0411-1. PubMed PMID: 31133749.
121. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481(7381):389-93. Epub 20120104. doi: 10.1038/nature10730. PubMed PMID: 22217937; PubMed Central PMCID: PMC3272464.
122. Cerezo M, Sollis E, Ji Y, Lewis E, Abid A, Bircan KO, et al. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res. 2025;53(D1):D998-D1005. doi: 10.1093/nar/gkae1070. PubMed PMID: 39530240; PubMed Central PMCID: PMC11701593.
123. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, 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(D1):D638-D46. doi: 10.1093/nar/gkac1000. PubMed PMID: 36370105; PubMed Central PMCID: PMC9825434.

STAR Methods

Key resources table

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GFP antibody	Abcam	Cat#ab290
GFP antibody (Western)	Cell Signaling Technology	Cat#2956S
RBPJ antibody (Western)	Cell Signaling Technology	Cat#5313S
β-Actin Antibody (Western)	Cell Signaling Technology	Cat#4967L
IRDye® 800CW Goat anti-Rabbit IgG Secondary Antibody (0.5 mg)	LI-COR Biosciences	Cat#926-32211
Bacterial and virus strains		
Biological samples		
Chemicals, peptides, and recombinant proteins		
DMEM	Thermo Fisher Scientific	Cat#11965-118
Fetal Bovine Serum (FBS), Premium, heat-inactivated	Thermo Fisher Scientific	Cat#A5670801
Anti-Anti (100X)	Thermo Fisher Scientific	Cat#15240-062
Normocin	Invivogen	Cat#ANT-NR-2
0.25% Trypsin-EDTA (1X)	Thermo Fisher Scientific	Cat#25200-056
Puromycin	Invivogen	Cat#ANT-PR-1
Hygromycin	Thermo Fisher Scientific	Cat#10687010
Zeocin	Thermo Fisher Scientific	Cat#R25001
Doxycycline hyclate, ≥93.5% (HPLC), Powder	Sigma-Aldrich	Cat#D9891
Opti-Mem	Thermo Fisher Scientific	Cat#31-985-088; Cat#11-058-021
Lipofectamine 2000	Thermo Fisher Scientific	Cat#11668019
mirVana™ RNA Isolation Kit	Thermo Fisher Scientific	Cat#AM1560
RIPA Lysis Buffer, 10X	Millipore Sigma	Cat#20-188
Pierce™ BCA® Protein Assay Kits and Reagents	Thermo Fisher Scientific	Cat#23227
4-12% NuPAGE Bis-Tris Gel	Thermo Fisher Scientific	Cat#NP0323BOX
NuPAGE MOPS SDS Running Buffer (20X)	Thermo Fisher Scientific	Cat#NP0001
Intercept (PBS) Blocking Buffer	LI-COR Biosciences	Cat#927-70003

Tween® 20	Sigma-Aldrich	Cat#P9416-50ML
Sodium dodecyl sulfate solution,BioUltra, for molecular biology, 20% in H2O	Sigma-Aldrich	Cat#05030-500ML-F
Formaldehyde	Electron Microscopy Sciences	Cat#15686
NaCl	Quality Biological	Cat#351-036-101
EDTA	Promega Corporation	Cat#V4231
EGTA	Thermo Fisher Scientific	Cat#NC0300614
Hepes (pH 8.0)	Boston BioProducts	Cat#BBH-80
Glycine	Thermo Fisher Scientific	Cat#BP381-1
PBS	Corning	Cat#21-040-CV
Glycerol	Teknova	Cat#G1796
10% Triton X-100	Sigma-Aldrich	Cat#648463
10% NP-40	Abcam	Cat#AB142227
Protease inhibitors	Thermo Fisher Scientific	Cat#78440
Tris (pH 8.0)	Thermo Fisher Scientific	Cat#J22638AE
TE	Sigma-Aldrich	Cat#T9285
SDS	Sigma-Aldrich	Cat#71736
NaDOC	Bioworld	Cat#40430018-1
Dynabeads Protein A	Invitrogen	Cat#10008D
Dynabeads Protein G	New England Biolabs	Cat#S1430S
TrisHCl	Alfa Aesar	Cat#J22638-AE
Digitonin	Promega Corporation	Cat#G9441
Tagment DNA Enzyme and Buffer Large Kit	Illumina	Cat#20034198
NEBNext High-Fidelity 2X PCR Master Mix	New England Biolabs (NEB)	Cat#M0541L
KCl	Fisher Scientific	Cat#BP366-500g
NaF	Sigma-Aldrich	Cat#S7920-100g
GFP antibody magnetic bead conjugate	Cell Signaling Technology	Cat#67090S
Enhanced chemiluminesent HRP substrate	Thermo Fisher Scientific	Cat#PI34095
Bromophenol Blue	Sigma-Aldrich	Cat#B0126-25g
Critical commercial assays		
Nano-Glo Live Cell Assay System	Promega Corporation	Cat#N2012
Deposited data		
Raw and analyzed data	This paper	GEO XXX
Human reference genome GRCh38	Genome Reference Consortium	https://www.gencodegenes.org/human/release_40.html

GENCODE human release 40 annotations	GENCODE	https://www.gencodegenes.org/human/release_40.html
Public EBNA2 ChIP-seq data	PMID 38475709	GEO GSM7854739; GEO GSM7854740
Experimental models: Cell lines		
Flp-In 293 Cell Line	Thermo Fisher Scientific	Cat#R75007
Flp-In 293 sublines expressing GFP-EBNA1, GFP-EBNA2, GFP-EBNA3A, GFP-EBNA3B, GFP-EBNA3C, RTA-GFP, BMRF-GFP, or GFP-ZTA	This paper	N/A
Flp-In T-REx-293	Thermo Fisher Scientific	Cat#R78007
Flp-In T-REx-293 subline expressing GFP-EBNALP	This paper	N/A
Experimental models: Organisms/strains		
Oligonucleotides		
Sequencing Primers	Integrated DNA Technologies	N/A
Recombinant DNA		
pOG44 Flp-Recombinase expression vector	Thermo Fisher Scientific	Cat#K650001
pcDNA5/FRT/TO vector	Thermo Fisher Scientific	Cat#K650001
pcDNA5_FRT_TO_GFP	This paper	N/A
pcDNA5_FRT_TO_EBNA2	This paper	N/A
pcDNA5_FRT_TO_GFP_EBNA2	This paper	N/A
pcDNA5_FRT_TO_EBNA1	This paper	N/A
pcDNA5_FRT_TO_GFP_EBNA1	This paper	N/A
pcDNA5_FRT_TO_EBNA3A	Twist Bioscience	N/A
pcDNA5_FRT_TO_GFP_EBNA3A	Twist Bioscience	N/A
pcDNA5_FRT_TO_EBNA3B	Twist Bioscience	N/A
pcDNA5_FRT_TO_GFP_EBNA3B	Twist Bioscience	N/A
pcDNA5_FRT_TO_EBNA3C	This paper	N/A
pcDNA5_FRT_TO_GFP_EBNA3C	This paper	N/A
pcDNA5_FRT_TO_EBNALP	This paper	N/A
pcDNA5_FRT_TO_GFP_EBNALP	This paper	N/A
pcDNA5_FRT_TO_ZTA	This paper	N/A
pcDNA5_FRT_TO_GFP_ZTA	This paper	N/A
pcDNA5_FRT_TO_RTA	Twist Bioscience	N/A
pcDNA5_FRT_TO_RTA_GFP	Twist Bioscience	N/A
pcDNA5_FRT_TO_BMRF1	Twist Bioscience	N/A

pcDNA5_FRT_TO_BMRF1_GFP	Twist Bioscience	N/A
pcDNA5_FRT_TO_LgBit_RTA	This paper	N/A
pcDNA5_FRT_TO_LgBit_RTA_NtermTruncation	This paper	N/A
pcDNA5_FRT_TO_LgBit_RTA_CtermTruncationSmall	This paper	N/A
pcDNA5_FRT_TO_LgBit_RTA_CtermTruncationBig	This paper	N/A
pcDNA_FRT_TO_SmBit_RBPJ	This paper	N/A
Software and algorithms		
nf-core/rnaseq pipeline (version 3.11.1)	PMID 32055031 and PMID 28398311	https://doi.org/10.5281/zenodo.7789554
FastQC (version 0.11.9)	citation: Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Cutadapt (version 3.4)	citation: https://doi.org/10.14806/ej.17.1.200	https://cutadapt.readthedocs.io/en/stable/
Trim Galore (version 0.6.7)	citation: https://doi.org/10.5281/zenodo.5127899	https://github.com/FelixKrueger/TrimGalore
SortMeRNA (version 4.3.4)	PMID 23071270	https://github.com/sortmerna/sortmerna
STAR (version 2.7.9a)	PMID 23104886	https://github.com/alexdobin/STAR
SAMtools (version 1.16.1)	PMID 19505943	https://www.htslib.org/
RSeQC (version 3.0.1)	PMID 22743226	https://rseqc.sourceforge.net/
Qualimap 2 (version 2.2.2-dev)	PMID 26428292	http://qualimap.conesalab.org/
dupRadar (version 1.28.0)	PMID 27769170	https://bioconductor.org/packages/release/bioc/html/dupRadar.html
preseq (version 3.1.1)	PMID 23435259	https://github.com/smithlabcode/preseq
Salmon (version 1.10.1)	PMID 28263959	https://github.com/COMBINE-lab/salmon
DESeq2 (version 1.34.0)	PMID 25516281	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ToppFun from the ToppGene Suite	PMID 17939863	https://toppgene.cchmc.org/enrichment.jsp
pheatmap (version 1.0.12)	cite Kolde R (2025). pheatmap: Pretty Heatmaps. R package version 1.0.12, https://github.com/raivokolde/pheatmap	https://github.com/raivokolde/pheatmap
ComplexHeatmap (version 2.20.0)	PMID 38868715	https://github.com/jokergoo/ComplexHeatmap
bamCoverage tool from the deepTools 3.5.5 software package	PMID 27079975	https://deeptools.readthedocs.io/en/stable/content/tools/bamCoverage.html
Other		
Resource website for this publication	This paper	In progress

UCSC Genome Browser session	This paper	https://genome.ucsc.edu/s/Ledsall/EBV_vTR
S220 Focused-ultrasonicator	Covaris	N/A
SX-8X IP-STAR compact automated system robot	Diagenode (Precision System Science Co. Ltd)	N/A
Tape Station 4150	Agilent Technologies	Cat#G2992A
iBLOT	Invitrogen	Cat#IB21001
Odyssey Dlx Imaging System	LI-COR Biosciences	Cat#2645119
GloMax Discover	Promega Corporation	Cat#GM3000

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Flp-In-293 and Flp-In T-REx-293 cells were maintained in DMEM supplemented with 10% FBS, 0.1% Anti-anti and 0.2% normocin. Similarly, the vTR stable cell lines containing the GFP fusions were grown in DMEM supplemented with 10% FBS, 0.1% Anti-anti, 0.2% normocin, and 200 µg/mL of hygromycin B. All cells were maintained in incubators at 37°C and 5% CO2. The cells were routinely tested and confirmed to be negative for mycoplasma contamination.

METHOD DETAILS

<u>vTR plasmid production</u>

The pcDNA5/FRT/TO backbone and DNA sequences for EBV vTRs from strain B95-8 were supplied to Twist Bioscience to generate untagged, N-terminally GFP tagged and C-terminally GFP tagged plasmids. The coding sequence was codon optimized if necessary for production by Twist Bioscience. Some of the vTR constructs could not be generated by Twist Bioscience, so they were made through restriction digest or TOPO cloning into the same plasmid backbone instead (labeled in the resource table as “This paper”). We used AlphaFold3 (97) to determine whether the N-terminal or C-terminal

fusion of GFP would have noticeable effects on the structure of each vTR protein. Plasmids utilized in this manuscript have been deposited in Addgene for distribution to other labs.

Transient transfection

Flp-In-293 cells were plated in 6 well plates at 500,000 cells/well in DMEM with 10% FBS and 0.1% Anti-anti. The following day cells were transfected (3 wells per vTR plasmid) with Lipofectamine 2000, according to the manufacturer's protocol, using 0.8 mg of vTR plasmid and 0.2 µg of puromycin plasmid. The day after the transfection, each 6 well of cells was transferred to a well of a 12 well plate in media containing 2 µg/mL puromycin to kill off non-transfected cells. After 24 hours of selection, the media was changed to regular growth media and cells were allowed to recover for 24 hours before RNA collection using a mirVana RNA isolation kit, following the manufacturer's protocol.

Stable cell generation

Flp-In-293 cells were plated on 6 well plates in DMEM supplemented with 10% FBS, 0.1% Anti-anti, and 0.2% normocin. Two days later, the cells were transfected using Lipofectamine 2000, according to the manufacturer's protocol, with 0.25 µg of vTR plasmid and 2.25 mg of the pOG44 Flp-Recombinase expression vector. Within three days, the media was changed to media containing hygromycin at 50 mg/mL and the hygromycin was gradually increased to 200 mg/mL to produce a pool of cells stably expressing the vTR. The purity of the pool was confirmed by microscopy for GFP and

Western blotting using an anti-GFP antibody to confirm expression of vTR fusions of the expected size. Doxycycline inducible Flp-In T-REx-293 cells were used to generate cells stably expressing EBNALP and were cultured in media also containing 100µg/mL zeocin. These cells were cultured in 1µg/mL doxycycline for overnight prior to collection.

Genomic assay sample collection

To ensure that the different vTR expressing cell lines were collected under similar culture conditions, stable cell lines were grown to confluency in four T175 flasks before transferring 90 million cells to six 15 cm dishes (15 million cells/dish). After incubating overnight, cells were trypsinized and two dishes/replicates were combined and standardized to 23 million cells per replicate for RNA-seq, ChIP-seq, and ATAC-seq experiments.

RNA-seq

Three million cells were washed with PBS and total RNA was isolated using the mirVana RNA extraction kit according to manufacturer's instructions. RNA quality was measured using the 4150 TapeStation System and samples were sent to either the Cincinnati Children's Hospital Medical Center (CCHMC) Genomics Sequencing Facility or Novogene for sequencing on an Illumina NovaSeq (paired-end, 150 bp read length).

ChIP-seq

20 million cells were resuspended in 20 mL of media for a final concentration of 1 million cells/mL. To crosslink the cells, 2 mL of 11% formaldehyde, 0.1M NaCl, 1 mM EDTA,

0.5 mM EGTA, and 50mM HEPES (pH 8.0) were added to the 20 mL of cells in media and crosslinked while rotating at room temperature for 10 minutes. The crosslinking reaction was stopped by adding 1.1 mL of 2.5M glycine and rotating for 5 minutes at room temperature. The cells were washed 2 times with cold PBS and then pelleted by centrifugation at 5000 RPM for 5 minutes and then stored in an Ultra-low freezer until further processing for ChIP. Crosslinked nuclei were prepared by washing the cell pellet with 1 mL of ChIP L1 buffer (10mM HEPES pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.25% Triton X-100, and 0.5% NP-40) plus protease inhibitors and rotating at 4 °C for 10 minutes and pelleted by centrifuging at 10,000 g for 5 minutes. Then the cell pellets were washed with 1 mL of ChIP L2 buffer (10 mM Tris pH 8.0, 1 mM EDTA, 200 mM NaCl, and 0.5 mM EGTA) plus protease inhibitors by rotating at room temperature for 10 minutes and the nuclei were pelleted by spinning at 10,000 g for 5 minutes. The nuclei were then washed with 0.5 mL of sonication buffer (TE + 0.1% SDS) with protease inhibitors and again pelleted by centrifugation. The pelleted nuclei were resuspended in 1 mL of sonication buffer containing protease inhibitors and sonicated using a Covaris S220 focused-ultrasonicator at 10% duty cycle, 175 peak power, 200 burst/cycle for 7 minutes. The sonicated chromatin was centrifuged at maximum speed for 10 minutes and the supernatant of soluble chromatin was collected. The salt and detergents were adjusted to final concentrations of 150 mM NaCl, 0.1% NaDOC, 1% Triton X-100, 5% glycerol and protease inhibitors. The chromatin was pre-cleared of Ig by adding 20 µL of Dynabeads Protein A or G and rotating at 4 °C for 40 minutes and centrifuged for 10,000 g for 10 minutes. We performed the ChIP using 200 mL of the pre-cleared chromatin on a SX-8X IP-STAR compact automated system robot from

Diagenode along with the antibody against the protein of interest and Dynabeads. For the vTRs, we used the Abcam anti-GFP antibody. For RBPJ we used the Cell Signaling Technologies RBPSUH (D10A4) XP antibody. The ChIP material was eluted using 10 mM TrisHCl and libraries were prepared by ChIPmentation (99) and sent to the CCHMC Genomics Sequencing Facility for sequencing on an Illumina NovaSeq (single-end, 100 bp read length).

ATAC-seq

We used 50,000 cells and followed the OMNI ATAC protocol (100) to generate the libraries of open chromatin that were subsequently sequenced. Briefly, transposase (Tn5) loaded with sequencing adapter sequences was used to cut the genome and insert sequencing adapters to tag the accessible DNA, as detailed in Buenrostro et al. (101). The resulting accessible DNA sequences were isolated, and libraries were prepared using the OMNI ATAC protocol (100). The libraries were sent to the CCHMC Genomics Sequencing Facility for sequencing on an Illumina NovaSeq (paired-end, 150 bp read length).

Western blots

Cells were lysed in RIPA-DOC-PPI and sonicated, and protein concentration was determined using a BCA protein assay. 10 µg of lysates were loaded in 4-12% NuPAGE Bis-Tris gels, run in MOPS buffer, and transferred to nitrocellulose using an iBlot. The membranes were blocked for 1 hour in Intercept Blocking Buffer and incubated overnight at 4°C in Intercept Blocking Buffer + 0.1% Tween 20 with anti-GFP or actin

antibodies diluted to 1:1000. The next day, the membranes were washed 2-3 times for 10 minutes with PBS + 0.1% Tween 20. The washed membranes were then incubated with rotation for 45 minutes with fluorescent secondary antibodies (diluted to 1:10,000 in Intercept Blocking Buffer + 0.1% Tween + 0.01% SDS). After three more 10 minute washes in PBS + 0.1 Tween 20, the membranes were imaged using an Odyssey DLx Imaging System.

Co-immunoprecipitation assays

10 million Flp-In-293 parental cells (control) or RTA-GFP expressing cells were lysed in 1 mL of co-immunoprecipitation buffer (25 mM HEPES, pH 7; 200 mM KCl; 20 mM NaF; 0.2mM EGTA; 1% NP40). The lysates were sonicated in a water bath sonicator for 4 minutes of 10 seconds on/off cycles and then centrifuged for 5 minutes at 12,000 RPM to pellet debris. We combined 100 µL of the soluble lysate with 100 µL of 2X Laemmli sample buffer to use as inputs and then used 500 µL of the soluble lysates for the co-immunoprecipitation through the addition of 25 µL of mouse anti-GFP antibody conjugated to magnetic beads and incubated rotating at 4 degrees for 2 hours. The beads were isolated with a magnet and washed 5 times with 500 µL of co-immunoprecipitation buffer. The immunoprecipitated material was eluted off the beads with 50 µL of 2X Laemmli sample buffer. We then loaded 10 µL of inputs and IP elutions on 4-12% NuPAGE Bis-Tris gradient gels, transferred to nitrocellulose and probed with rabbit anti-GFP or rabbit anti-RBPJ antibodies. The blots were washed 3x for 10 minutes, incubated with HRP conjugated anti-rabbit secondary antibodies, washed again and then imaged using enhanced chemiluminescent HRP substrate.

Split NanoLuc luciferase protein complementation assays

For the split NanoLuc luciferase experiments, 200,000 Flp-In-293 cells were plated per well of a 12 well plate. The next day, the cells were transfected with 0.2 mg/well of the indicated split NanoLuc luciferase constructs (LgBit fused to the various RTA constructs and SmBit fused to RBPJ). The next day the wells were trypsinized and 50,000 cells were plated/well of a 96 well in quadruplicate wells. After allowing the cells to attach for 24 hours, the media was changed to 50 mL of Opti-Mem + 12.5 mL of Promega Nano-Glo Live Cell reagents. The plate was then read over a 10 second integration on a GloMax (Promega) plate reader. The signals from all the wells were normalized to the average of the cells transfected with only the RTA-LgBit fusion to control for background signal.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq data analysis

RNA sequencing data were processed using the nf-core/rnaseq pipeline (version 3.11.1) https://doi.org/10.5281/zenodo.7789554). Initial quality control of the raw sequencing data was performed using FastQC (version 0.11.9) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/]). Low-quality bases and adapter sequences were then trimmed and filtered from the reads using Cutadapt (version 3.4) (https://doi.org/10.14806/ej.17.1.200) and Trim Galore (version 0.6.7) (https://doi.org/10.5281/zenodo.5127899), respectively. Ribosomal RNA sequences were subsequently removed from the aligned data using SortMeRNA (version 4.3.4)

(102) against the default libraries augmented with human rRNA sequences to eliminate potential contamination from non-target RNA species. STAR (version 2.7.9a) (103) was used to align the trimmed reads to the GRCh38 genome (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/GRCh38.primary_assembly.genome.fa.gz) with GENCODE release 40 annotations (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/gencode.v40.annotation.gtf.gz) (104). The resulting alignments were sorted and indexed using SAMtools (version 1.16.1) (105). To further evaluate the quality of the RNA sequencing data, several tools were employed, including RSeQC (version 3.0.1) (106), Qualimap 2 (version 2.2.2-dev) (107), dupRadar (version 1.28.0) (108), and preseq (version 3.1.1) (109). These tools provide comprehensive assessments of various quality metrics, such as read distribution, GC content, duplication rates, and library complexity, ensuring reliable data for downstream analysis. For transcript quantification, Salmon (version 1.10.1) (110) was used to estimate the expression levels of transcripts. Differential gene expression analysis was performed using DESeq2 (version 1.34.0) (111). Genes were considered differentially expressed if they had a 20% fold change compared to the GFP control and an adjusted p-value threshold of less than 0.05. The EBNALP stable cell line was compared to the Flp-In T-REx-293 GFP control cell line. The other stable cell lines were compared to the Flp-In-293 GFP control cell line. For the transient transfections, the untagged vTR was compared to the transient GFP control. The input count files were the quant.sf files generated by Salmon. The gene information came from the salmon_tx2gene.tsv file and contained 61,544 entries. Pathway enrichment analyses for the differentially expressed genes were performed using ToppFun from the

ToppGene Suite (112) and the GO Consortium's Generic GO subset (goslim_generic) terms annotated as a biological process ((113, 114) and https://doi.org/10.5281/zenodo.14083199). The parameters used for ToppFun were "Probability density function" for the pValue Method, "GO: Biological Process for the Feature", "None" for the Correction, "1.0" for the p-Value cutoff, and gene limits of 5 and 2,000. The ToppFun calculated background set of 20,262 genes was used. Each set of DEGs for a specific vTR and direction (e.g. BMRF1 up-regulated genes) was processed separately, using the Ensembl IDs as the input set. The results were filtered against the GO slim terms and an adjusted p-value was calculated using the Benjamini-Hochberg Procedure. The odds ratio was calculated using the formula from Enrichr (113, 114), specifically (1.0 * a * d) / max(1.0 * b * c, 1), where a is "the number of overlapping genes"; b is "the number of genes in the annotated set" - "the number of overlapping genes"; c is "the number of genes in the input set" - "the number of overlapping genes"; and d is "the number of genes in the background set" - "the number of genes in the annotated set" - "the number of genes in the input set" + "the number of overlapping genes". An adjusted p-value threshold of 0.05 was used to generate the pathway enrichment figures. Heatmaps were made with pheatmap (version 1.0.12) (115), except for Figure 7B, which was made with ComplexHeatmap (version 2.20.0) (24).

Data Visualization

Visualization tracks for each dataset were created for the UCSC Genome Browser (94). Signal tracks (in bigWig format) were created using the bamCoverage tool from the deepTools 3.5.5 software package (116) with the parameters --normalizeUsing CPM

and --binSize 10. For the RNA-seq signal tracks, the command was run twice in order to maintain strandedness. For the forward strand track, the parameter --filterRNAstrand forward was included. For the reverse strand track, the parameters --filterRNAstrand reverse and --scaleFactor -1 were included. Genome browser screenshots were made with the Integrative Genomics Viewer (117).

Chromatin State Analysis

Regions corresponding to particular chromatin states in HEK-293 cells were derived previously (61). Briefly, a ChromHMM model was trained with the histone marks H3K4me1, H3K4me3, H3K36me3, H3K27ac, H3K9me3, and H3K27me3, CTCF ChIP-seq, ATAC-seq, and GENCODE promoter regions. ChIP- and ATAC-seq peaks from our analyses were assigned to a ChromHMM state using “bedtools intersect” (version 2.31.1) (118), and total coverage was derived by using “bedtools groupby” to group intersected regions by their label and sum the sizes of intersecting peaks. Chromatin state distribution over accessible chromatin was calculated in the same way, using the ATAC-seq peaks from the monoGFP (control) experiment. The total size of each chromatin state across the genome was assessed similarly by first removing the ENCODE blacklist regions from the ChromHMM states using “bedtools subtract” and then aggregating with “bedtools groupby”.

Transcription factor motif enrichment analyses

HOMER v4.9.1 (119) was used for all motif enrichment analyses. Known motif enrichment using CisBP v3.00 motifs (120) was computed for each peakset with

findMotifsGenome.pl and the “-size given” parameter. The enrichment heatmaps in Figures 2F and 4 were constructed by first identifying the top (lowest p-value) 10 motifs in each peakset that come from distinct motif families. The Pearson correlation of the normalized motif p-values for each motif was calculated, and the motif with the highest correlation was iteratively removed until the highest correlation coefficient was below 0.85. *De novo* motif enrichment of RTA was performed with the HOMER “-size given” parameter and the CisBP 3.0 motifs in the “-mcheck” parameter.

ChIP-seq peak calling

ChIP-seq alignment and peak calling was performed using a version of the ENCODE ChIP-seq pipeline (57, 58, 120) slightly modified to run within a Nextflow environment. All alignments were performed to the hg38 genome. Each vTR used the GFP-expressing cell line as a background control for peak calling. For all downstream analyses, “IDR conservative” peaks were used.

ATAC-seq peak calling and differential analysis

ATAC-seq alignment and peak calling was performed using a version of the ENCODE ATAC-seq pipeline (57, 58, 120) slightly modified to run within a Nextflow environment. All alignments were performed to the hg38 genome. For all downstream analyses, “IDR conservative” peaks were used.

Differentially accessible regions were identified using DiffBind v3.12.0 (121) with an FDR threshold of 0.05. Assignments to “direct” and “indirect” chromatin effects were

made using bedtools v2.31.1 (118) and command lines similar to “bedtools intersect -a $open_bed -b $chip_peaks -wa -u > ${vtr}_direct_opening.bed” or “bedtools intersect -a $open_bed -b $chip_peaks -v > ${vtr}_indirect_opening.bed”.

BMRF1 disease associations

The Regulatory Element Locus intersection (RELI) algorithm (16) was used to compare BMRF1 ChIP-seq peaks with tag SNPs for phenotypes in the GWAS Catalog (122) (downloaded on February 18, 2025). The final set of 439 tested phenotypes consisted of those that 1) describe disease processes, 2) have at least 10 tag SNPs which are not in linkage disequilibrium ($r^2 < 0.2$), and 3) were assessed in populations of primarily European ancestry (to match the ancestral background of our cell line).

The BMRF1 interaction network (Figure 6D) was constructed using STRING (123), with expert curation to assign group function. Variants that were overlapped by BMRF1 ChIP-seq peaks and were tag SNPs for one of the five rheumatoid arthritis-related phenotypes in Figure 6C were assessed for eQTL function using release 7 of the eQTL Catalogue (98). Gene targets of these eQTLs (p-value < 1E-5) were used as input to STRING.

ADDITIONAL RESOURCES

Project website: in progress

UCSC Genome Browser session: https://genome.ucsc.edu/s/Ledsall/EBV_vTR