Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Syst. Author manuscript; available in PMC 2024 June 21.

Published in final edited form as:

Cell Syst. 2023 June 21; 14(6): 482–500.e8. doi:10.1016/j.cels.2023.05.008.

High-Throughput Discovery and Characterization of Viral Transcriptional Effectors in Human Cells

Connor H. Ludwig[1,2], **Abby R. Thurm**[3], **David W. Morgens**[4], **Kevin J. Yang**[5], **Josh Tycko**[6,7], **Michael C. Bassik**[6], **Britt A. Glaunsinger**[4,5,8], **Lacramioara Bintu**[1,9,*]

[1]Bioengineering Department, Stanford University, Stanford, CA 94305, USA

[2]Octant Inc., Emeryville, CA 94608 (current employer)

[3]Biophysics Graduate Program, Stanford University, Stanford, CA 94305, USA

[4]Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA 94720, USA

[5]Department of Molecular and Cell Biology, UC Berkeley, Berkeley, CA 94720, USA

[6]Department of Genetics, Stanford University, Stanford, CA 94305, USA

[7]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA

[8]Howard Hughes Medical Institute, UC Berkeley, Berkeley, CA 94720, USA

[9]Lead Contact

Summary

Viruses encode transcriptional regulatory proteins critical for controlling viral and host gene expression. Given their multifunctional nature and high sequence divergence, it is unclear which viral proteins can affect transcription and which specific sequences contribute to this function. Using a high-throughput assay, we measured the transcriptional regulatory potential of over 60,000 protein tiles across ~1,500 proteins from 11 coronaviruses and all nine human herpesviruses. We discovered hundreds of transcriptional effector domains, including a conserved

*Correspondence: lbintu@stanford.edu.

Author Contributions

CHL and LB designed the study, with significant intellectual contributions from ART. CHL designed the libraries. CHL and ART cloned the libraries. CHL performed the vTR/CoV tiling, HHV tiling, HHV perturbation, and HHV chemical inhibition screens. ART performed the Spike perturbation screen. CHL analyzed all screen data. CHL and ART performed RNA-seq and analyzed the data. CHL and ART generated plasmids and cell lines and performed individual recruitment assay experiments. DWM and KJY performed CRISPR/Cas9 experiments with KSHV and analyzed the data. CHL analyzed all other data, with contributions from ART and LB. CHL and LB wrote the manuscript, with significant contributions from ART and contributions from DWM and BG. JT and MCB provided technical advice and materials at the beginning of the project. LB supervised the project.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests

CHL, ART, JT, MCB, and LB have filed a provisional patent related to this work through Stanford University. CHL is an employee and shareholder of Octant Inc. JT, MCB, and LB are co-founders of Stylus Medicine. MCB and LB are members of the scientific advisory board of Stylus Medicine. BAG is a member of the scientific advisory board of Imunon. The remaining authors declare no competing interests.

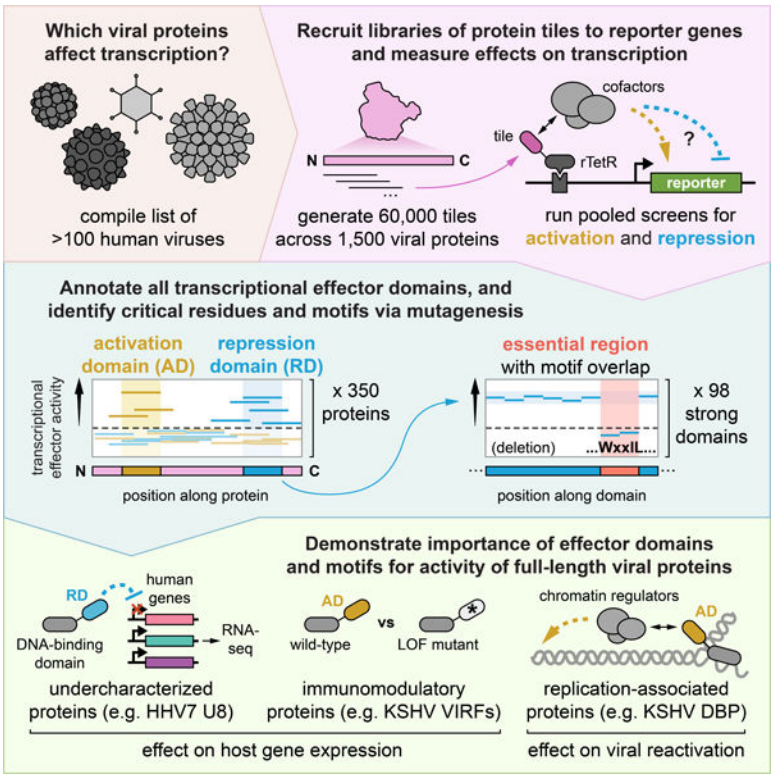
Inclusion and Diversity

We support inclusive, diverse, and equitable conduct of research.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

repression domain in all coronavirus Spike homologs, dual activation-repression domains in VIRFs, and an activation domain in six herpesvirus homologs of the single-stranded DNA-binding protein that we show is important for viral replication and late gene expression in KSHV. For the effector domains we identified, we investigated their mechanisms via high-throughput sequence and chemical perturbations, pinpointing sequence motifs essential for function. This work massively expands viral protein annotations, serving as a springboard for studying their biological and health implications and providing new candidates for compact gene regulation tools.

Graphical Abstract



eTOC Blurb

Ludwig et al. measure transcriptional activation and repression for tens of thousands of viral protein sequences and mutants using a high-throughput reporter assay in live human cells. This study massively expands viral protein annotations, benefitting basic virology and synthetic biology efforts.

Introduction

There are more than 200 viruses that infect humans, many of which are known etiological agents of disease[1] and have been responsible for major global health crises, including the most recent COVID-19 pandemic. Key to this pathogenicity are interactions between viral factors and host cellular machinery[2]. Viruses encode transcriptional regulatory proteins,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which are critical for the precise temporal control of viral gene expression and the extensive rewiring of host gene expression programs necessary for creating a cellular environment conducive to productive infection[3]. Viral transcriptional regulators (vTRs) are thus attractive targets for therapeutic intervention[4].

Given the multifunctional nature of many viral proteins, which have evolved so due to virion and genome size constraints[5], and their relatively high sequence divergence[6], it is not clear which viral proteins can affect host transcription. A recently published meta-analysis compiled a census of viral proteins with evidence for nucleic acid binding and/or transcriptional regulation and examined their properties, secondary functions, and genomic targets for the small subset of proteins for which data was available[7]. While this represents the best compilation of vTRs to date, many of the entries within the vTR census lack direct experimental evidence of transcriptional regulation, most of their effector domains have not yet been defined, and the census as a whole likely only represents a fraction of all vTRs due to historical technical limitations that have precluded systematic experimental investigation of transcriptional effector function.

In this study, we use a recently developed high-throughput approach[8] to test tens of thousands of protein sequences for their effect on gene expression when recruited at reporter genes. This method allows us to identify and characterize viral transcriptional regulators and their effector domains. We start with entries from the vTR census to demonstrate feasibility. We then extend this approach to discover previously undescribed effector domains within the proteins of 11 coronaviruses, including SARS-CoV-2, and all nine human herpesviruses. For the hundreds of effector proteins that we identify, we investigate the sequence determinants of transcriptional regulation, their mechanisms of action using high-throughput measurements, and for a small subset of them their consequences on host gene expression.

Results

High-throughput identification of activation and repression domains across a curated library of putative viral transcriptional regulators

We have recently developed a high-throughput method (HT-recruit) that allows us to measure the activity of thousands of transcriptional activators and repressors at reporter genes (Fig. 1A, Fig. S1A)[8]. We do this by cloning a library of putative regulators as fusions to the doxycycline (dox)-inducible rTetR DNA binding domain and delivering them to K562 cells by lentivirus at low multiplicity of infection such that each cell contains a single library member. By adding dox, we can recruit candidate effector domains to a minimal promoter (minCMV) to identify activators or to a constitutive promoter (EF1α) to identify repressors (Fig. 1A). The reporter genes encode both a fluorescent protein for visualization as well as a surface marker for rapid and robust magnetic separation based on reporter transcriptional state (ON or OFF). Following magnetic separation, we extract genomic DNA from cells in the ON and OFF populations, prepare libraries for next generation sequencing, and compute quantitative enrichment scores for each library member based on their frequencies in the two populations. This method allows us to measure the activity of tens of thousands of candidate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

effector domains, each 80 amino acids (aa) long (the current limit of DNA synthesis for pooled libraries of this size).

In order to test this method with viral proteins, we designed a library that contains strong positive controls for both activation and repression as well as proteins that have been proposed as vTRs but lack strong experimental evidence[7]. This library consists of 80aa-long protein tiles (sampled every 10aa) across 377 putative vTRs encoded by 116 non-BSL4 human viruses[7] as well as 80aa-long random sequences to serve as negative controls (Fig. 1B, Table S1, Methods). Activation and repression measurements of this library were reproducible between biological replicates (Fig. 1C&D). For the activation screen (Fig. S1B), we computed enrichment in the ON versus the OFF population for all library members and used the scores of negative controls to define a detection threshold (Fig. 1C, Methods). We identified 586 activator tiles, including those from the well-known activators E1A (from human adenovirus), RTA (from human gammaherpesviruses), and VP16 (from alphaherpesviruses) (Fig. 1C, Table S1). To assess the accuracy of our assay, we performed individual recruitment experiments for a set of hits and non-hits and found good correlation between the fraction of cells ON by flow cytometry and the HT-recruit enrichment score (Spearman r = 0.86), with 95% (20 of 21) of the individually recruited hit tiles measurably activating transcription (Fig. S1C–E, Table S7). Similarly, we screened the same library for tiles that could repress the constitutive reporter gene (Fig. S1F), defined a detection threshold of OFF versus ON enrichment scores based on the random negative control scores, and identified 476 repressor tiles, including those from the well-known repressors E1A, ICP0 (from herpesvirus), and LT (from human polyomavirus) (Fig. 1D, Table S1). Screen enrichment scores correlated well with individual recruitment experiments (Spearman r = 0.92), with all 21 of the individually recruited hit tiles measurably repressing transcription (Fig. S1G–I, Table S7), giving us confidence that our high-throughput method can reliably measure transcriptional activity.

Proteins are typically composed of structural and functional subunits called domains that are modular and can evolve independently[9]. Identifying protein domains can provide useful annotations, structural clarity, and mechanistic insight for protein and drug design purposes. One distinct advantage of screening protein tiling libraries is the ability to pinpoint the domains that are responsible for the measured function. For our assay, we defined a transcriptional effector domain as any set of two or more consecutive hit tiles or as any single hit tile positioned at the N- or C-terminus (Fig. 1E), and the strongest tile from each domain was used in subsequent analyses. Applying these criteria yielded 87 activation domains (Fig. 1F) and 106 repression domains (Fig. 1G) across a total of 117 proteins (Table S3).

For VP16, one of the better known proteins associated with transcriptional activation and responsible for immediate early gene activation during alphaherpesvirus infection[10], we recovered known activation domains and, in addition, discovered previously unannotated transcriptional effector domains in some homologs (Fig. 1H, Fig. S1J–L). Specifically, we detected the well-described and highly conserved tandem C-terminal activation domains present in human herpes simplex virus 1 (HSV1) and HSV2 and absent in varicella zoster virus (VZV), which instead possesses a potent N-terminal activation domain (Fig. S1K)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that shares sequence homology with part of the HSV1 and HSV2 C-terminal activation domains[11]. We also detected C-terminal activation domains in the VP16 homologs of cercopithecine herpesvirus (CeVH) 1 and CeHV2, whose natural hosts are macaque monkeys, as well as weak N-terminal activation domains that, to our knowledge, were not previously described (Fig. S1L). We did not detect any activation domains in the homolog from Suid herpesvirus 1 (SuHV1), which primarily infects pigs and other non-primate animals. We also identified weak repression domains - some of which overlap with activation domains - within the HSV2, CeHV1, CeHV2, and SuHV1 VP16 homologs (Fig. 1H, SuHV1 data not shown, Table S3), suggesting that they may act as transcriptional repressors in certain contexts or at least engage with co-repressors.

Some of the strongest activation and repression domains we measured originate from homologs of human adenovirus (HAdV) E1A (Fig. 1C&D, Fig. 1F&G), a highly multifunctional protein involved in cell cycle deregulation, immune evasion, and oncogenesis and known to bind over 50 cellular factors[12]. We identified effector domains in all six E1A homologs included in the vTR census (Fig. 1I, Fig. S1M–O, representative examples) and found that most of these domains aligned with conserved regions (CRs) previously described as having transcriptional function (Fig. 1I). Specifically, we identified potent transcriptional activation domains aligning with the p300-binding CR1 and the TBP/TAF-binding CR3[13]. We also identified repression domains aligning with CR4 in all homologs except HAdV9 E1A, which had a single very weak repressive tile in that region (Fig. S1N).

CR4 from HAdV5 E1A, which is the best studied homolog of those in the vTR library, has been shown to contain three regions that are important for interferon response suppression and that bind the CtBP corepressor, the adaptor protein DCAF7, and FOXK transcription factors (Fig. 1J)[14,15]. However, FOXK binding appears to be specific to HAdV5 E1A (Fig. 1J), suggesting that it is dispensable for the repressive activity we measured across homologs. Indeed, deletion of the FOXK-binding sequence had no effect on silencing (Fig. 1K). In contrast, mutating the DCAF7-binding region (R262E)[15] or deleting the CtBP-binding region (PLDLS) partially reduced silencing to similar degrees, and perturbing both regions abolished silencing altogether. Consistent with these results, deletion of the CtBP-binding sequence in the weaker repressive CR4 domain from HAdV9 E1A completely abolished silencing, while installing a E159R mutation within the DCAF7-binding region to resemble the HAdV5 E1A sequence (Fig. 1J) increased silencing (Fig. 1K). These data support the observation that the combined activities of DCAF7 and CtBP are important for transcriptional repression function across E1A homologs and that the exact E1A sequence may modulate affinity for these cofactors.

Within the vTR library, we found a significant enrichment of effector domains within proteins from DNA viruses compared to RNA viruses, especially dsDNA viruses (Fig. 1F–G, Fig. S1P&Q). This supports the observation that there is generally concordance between viral genome type and the target of encoded viral transcriptional regulators[7]. In further support of this, an unbiased tiling screen of all 368 proteins encoded by 11 human and bat coronaviruses (RNA viruses) identified primarily weak activation tiles and few moderate-strength repression tiles (Fig. S2A–C). About one third of repression domains mapped to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

one region across all Spike protein homologs (Fig. S2D&E), and a screen perturbing the sequence of this region identified critical residues for silencing within a likely monomeric leucine zipper that faces inward in the normal trimeric state (Fig. S2F–K). It remains to be determined whether this portion of the Spike protein affects chromatin regulators in the cytoplasm or transcription in the nucleus in the context of the full-length protein during infection.

The largest enrichment of effector domains in the vTR library was within proteins from the dsDNA herpesvirus family (Fig. 1L), which account for 30% of the vTR library (114 of 377) but represent 46% of proteins containing effector domains (54 of 117) (OR: 2.86, 95% CI: 1.80–4.54, Fisher's $p < 0.0001$). Overall, the correlation between HT-recruit screen scores and individual flow cytometry experiments, as well as the recovery of tiles from well-described transcriptional effectors, demonstrates that our high-throughput method can quantitatively measure transcriptional activation and repression domains within viral proteins.

Unbiased identification of activators and repressors from herpesviruses

Given their dominance in the vTR screens, we next focused on herpesviruses (HHVs), which are important in human health and disease, are ubiquitous[16], have a chromatinized dsDNA genome that persists for life[17], and encode more proteins than most viruses[18]. As such, we took a discovery-based approach to identify herpesvirus-encoded transcriptional effectors beyond those included in the vTR census, tiling nearly all known proteins (891) encoded by nine human herpesviruses and the porcine SuHV1 (hereafter HHV tiling library) (Fig. 2A, Fig. S3A, Table S2). We found good reproducibility between replicate screens with this library (Fig. S3B–G), as well as a strong correlation between individual flow cytometry experiments measuring the fraction of cells ON or OFF and the screen enrichment scores (Fig. 2B&C, Table S7).

We identified 72 activation domains and 196 repression domains across 178 proteins (Table S3). Several proteins contain both types of domains (Fig. 2D), and sometimes activation and repression domains overlap: a subset of activator tiles spanning across all activation scores also act as weak repressors (Fig. S3H–J, Table S2). Among the herpesvirus species tested, human cytomegalovirus (HCMV) encoded the most proteins with transcriptional regulatory activity, although a higher percentage of the proteins from the gammaherpesviruses Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV) contain transcriptional effector domains (Fig. 2D).

There are 67 herpesvirus proteins that are common to the vTR and HHV tiling libraries (identical UniProt identifiers), which allows us to assess the consistency of our measurements across screens. At the tile level, we observed a strong correlation between vTR and HHV tile measurements for each of the activation (Fig. S3K) and repression (Fig. S3L) screens, with the HHV activation screen exhibiting greater sensitivity than the vTR activation screen. Additionally, 31 of the 34 (91%) herpesvirus proteins with at least one effector domain in the vTR screen also had the same effector domain in the HHV screen (Fig. 2E). This overlap includes well-known activators, such as VP16, RTA,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and alphaherpesvirus ICP4 homologs, as well as repressors, such as KSHV KbZIP and alphaherpesvirus ICP0 homologs (Fig. 2F&G).

We identified an additional 147 herpesvirus proteins unique to the HHV tiling library that possessed measurable transcriptional regulatory potential (Fig. 2E–G), nearly 5-fold more than the herpesvirus proteins for which we measured activity in the vTR screen. These newly identified effectors spanned a similar range of scores (Fig. S3M&N) and were validated with individual flow cytometry experiments (Fig. S3F&G). To better understand what was already known about these proteins and what new functional information our screen could provide, we examined the UniProt biological process (BP) GO term annotations for our hits. While two-thirds of these proteins had some annotation, only 9.5% (14 proteins) were reported to be involved in the regulation of gene expression (Fig. 2H) (e.g. HSV1 UL46, HHV6A IE2, HCMV UL117), with only a few of these (VP16 and ICP4 homologs) having defined effector domains in UniProt. For instance, HHV7 U84 is annotated as having a role in transcriptional regulation based on sequence homology to HCMV UL117[19], but this activity has never been measured. Our assay identifies a strong repression domain in U84 (Fig. 2I), which also has a predicted DNA-binding domain[20], suggesting a role as a viral transcription factor. Indeed, expression of full-length U84 for 48 hours produced significant changes in host gene expression profiles compared to negative control cells expressing mCitrine as measured by RNA-seq (Fig. 2I, Fig. S3O&P, Table S5). Thus, for many of the proteins that do have BP GO terms related to regulating gene expression, our study provides the first experimental evidence supporting this annotation and defines the domain responsible for this activity. Additional experimental evidence is needed for these proteins to determine which of these effects on transcription are the result of direct binding to DNA and activation/repression of the target gene or indirect actions, such as competition with human transcription factor binding or competition for a limited pool of coactivators and corepressors.

The remaining effector proteins with at least one BP GO term annotation fell into several categories associated with other biological processes, including DNA replication (e.g. DNA polymerase, helicase, and DNA-binding protein homologs), viral entry (e.g. envelope glycoprotein homologs), immune suppression (e.g. HCMV UL18, EBV BLRF2, KSHV ORF52), and virion assembly (e.g. capsid assembly and tegument proteins). This finding of an additional function is consistent with the observation that viral proteins tend to be multifunctional[7,21].

One-third of the transcriptional effector proteins identified in our screen (49 proteins) were not associated with any BP GO term in UniProt (Fig. 2H), meaning that our dataset provides the first functional annotation for these un- and under-characterized proteins. For example, the previously uncharacterized RL5A protein from HCMV harbors a moderately strong repression domain but lacks a predicted DNA-binding domain and produces modest changes in host gene expression when expressed in its full-length form for 48 hours (Fig. 2J, Table S5). Most of the differentially expressed genes are upregulated, suggesting that the repressive domain of RL5A might bind repressive cofactors and sequester them away from their target genes, leading to mild de-repression. Since it lacks a DNA binding domain, RL5A may require additional DNA-associated factors or function in a complex with other

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

viral proteins to exert potentially stronger transcriptional regulatory activity. In contrast, the previously uncharacterized U8 protein from HHV7 harbors a strong repression domain and a predicted DNA-binding domain, and expression of the full-length protein for 48 hours produced significant changes in host gene expression (Fig. 2K, Table S5), supporting a role for this protein as a viral transcription factor. Taken together, these findings demonstrate that our high-throughput, unbiased tiling approach can discover viral transcriptional regulators and annotate their effector domains.

Sequence analyses and systematic perturbation of herpesviruses transcriptional effectors

As evident in the E1A example, small differences in protein sequence can produce substantial differences in transcriptional effector activity. Understanding which amino acids within transcriptional activation and repression domains are critical to and modulate function enables us to begin to understand their mechanisms of action, predict the functional consequences of viral mutations, and identify potential drug targets (Fig. 3A). Many eukaryotic transcriptional activation domains consist of interspersed acidic and hydrophobic residues[22–24], while repressors fall into more categories not defined by common sequence composition[25]. In line with this, nearly all activator tiles from the HHV tiling screen have a net negative charge, with stronger activator tiles typically having greater negative charge (Fig. 3B). In contrast, herpesvirus repressor tiles appear to be equally likely to have net positive or negative charge (Fig. 3B). Both activators and repressors have an intermediate non-polar content (30–60%), and tiles with extremely low or high net charge or non-polar content generally do not exhibit effector activity (Fig. 3B).

To better understand the sequence bases for the diverse range of transcriptional regulatory activities of herpesvirus proteins, we examined residue frequencies across effector domains. Since dual effector tiles share more sequence properties with pure activators than pure repressors (Fig. S4A&B), consistent with their behavior as stronger activators than repressors (Fig. S3H), we grouped activator and dual effectors for all subsequent sequence analyses. Overall, activation domains are generally enriched in acidic residues and depleted in basic residues, consistent with their overall negative net charge (Fig. 3B), repression domains are enriched in acidic residues, and both domain types are depleted in certain non-polar residues (Fig. 3C). However, these enrichments do not necessarily mean that these amino acids are important for transcriptional regulatory function of these domains.

To directly measure which residues and regions are important for transcriptional activation and repression, we systematically perturbed the amino acid sequences of the maximum-strength tiles within effector domains. In this set of high-throughput perturbation measurements, we focused on tiles that we estimated could activate or repress at least 40% of cells (Fig. 2B&C, Table S2) so that we could measure appreciable differences in activity and test more perturbations for a smaller set of tiles. Specifically, we mutated the residues enriched in our effector domains, as well as others that have been implicated in transcriptional regulation in human cells[25]: acidic (D, E), basic (K, R), aromatic (W, F, Y), and others (S, T, Q, P). In addition, we performed deletion scanning with 5aa deletions every 5aa to identify critical regions and residues in a more unbiased manner.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Our activation and repression screens with this HHV perturbation library were reproducible (Fig. S4C&D, Fig. S4G&H), individual validation experiments showed a strong correlation between percent ON or OFF and screen scores (Fig. S4E&F, Fig. S4I&J, Table S7), and these data identified functionally important sequences within each domain: essential regions whose deletion breaks function, as well as regions whose deletion reduces or enhances function (Fig 3D&E, Table S4). For example, the HHV6A U95 dual effector domain is a strong activator with mild repressive activity that has an essential region containing neighboring phenylalanine and tryptophan residues critical for both activation (Fig. 3D) and repression (Data S1). This essential region overlaps a short alpha helix that is predicted by JPred4[26] (Fig. 3D, top). In another example, our assay identified a lysine-rich essential region within HHV7 U84 harboring several critical residues whose individual substitutions were sufficient to abolish repression altogether (Fig. 3E). These residues mapped onto one face of a basic alpha helix that likely engages a corepressor complex (AlphaFold, data not shown). In general, essential regions within both activation and repression domains were more likely to overlap JPred4-predicted alpha helices (Fig. S4K&L), which could stabilize binding interfaces and particular side chain conformations required for activity.

A high-level analysis of the functional consequences of single-residue substitutions and deletions revealed a critical role for tryptophan in transcriptional effector activity (Fig. 3F–I) that has not been described before. Substitution of tryptophan to alanine reduced or abolished activation and repression in 73% and 67% of cases, respectively (Fig. 3F&G), consistent with the fact that tryptophan was enriched in the essential regions of both activators and repressors (Fig. 3H&I). Substitution of the other aromatic residues also broke or reduced function, though less frequently (Fig. 3F&G). Substitution of acidic residues reduced or abolished activation and repression in approximately 30–40% of cases, while substitution of basic residues generally only negatively affected repression and not activation (Fig. 3G), consistent with our findings above (Fig. 3C). In general, sequence bias was stronger within the essential regions of activation domains than repression domains (Fig. 3H&I), most likely reflecting the greater complexity and more diverse modes of transcriptional repression also observed in human transcriptional repressors[25].

To connect the above sequence features with how this set of effector domains might modulate transcription through the recruitment of co-repressors (CoRs) and co-activators (CoAs), we first searched for well-defined cofactor interaction motifs compiled in the ELM database and those identified in recent publications[25,27,28] (Table S6), ultimately focusing on those enriched in reducing/breaking regions in an initial search (Methods). In essential regions of repression domains, we found several instances of SUMOylation sites, which have been connected to transcriptional repression in human cells[29]. For both activation and repression essential regions, we identified SUMO-interaction motifs (SIMs), which may bind to SUMOylated CoRs and CoAs[30,31] (Fig. 3J).

In activation domain essential regions, we found several instances of the multifunctional nuclear receptor (NR) box motif (i.e. LxxLL), which is known to engage CoAs such as p300/CBP and TFIID[32] (Fig. 3J). Previous research reported instances of modified NR motifs in human proteins that can still bind their targets despite having other non-polar residues in place of leucines in the LxxLL consensus[33–35]. We also found these types

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of motifs, which we termed flexiNR box motifs, in the essential regions of our effector domains (Fig. S4M, Table S6). With the addition of the flexiNR box motif to our list, the majority of our effector domains contain a motif for binding to a candidate cofactor: 50% of the activators and 67% of the repressors.

For 26% of activation domains and 10% of repression domains, there was no essential region whose deletion broke function (Fig. 3J). For nearly all of these domains, we identified two or more of the motifs listed above (Table S6); thus, it is possible that upon deletion of a single motif, the other motifs may compensate to avoid total loss of activity. Conversely, for 24% of activation domains and 23% of repression domains, we identified at least one essential region but could not identify a known motif (Fig. 3J). While there were too few of these sequences for de novo motif finding, we found several critical acidic and aromatic residues within the activation essential regions and critical tryptophan residues within the repression essential regions, consistent with our above analysis (Fig. 3H&I).

As an orthogonal approach to the identification of potential cofactors, we performed screens in the presence of chemical inhibitors of chromatin-modifying enzymes classically associated with gene activation and silencing: celastrol, an inhibitor of the KIX domain of the CoA p300/CBP[36] (KIX directly binds NR box-containing proteins like the human transcription factor MYB); SGC-CBP30, an inhibitor of the bromodomains of p300/CBP[37] (bromodomains bind acetyllysine residues); tazemetostat, an inhibitor of the histone methylation activity of the polycomb repressive complex 2 (PRC2)-associated enzyme EZH2[38] (no known motif); and TMP269, an inhibitor of class IIa histone deacetylases (HDACs)[39] that generally act as CoRs (no known motif). All chemical inhibition screens were reproducible, and when we compared the results of each to a DMSO-control screen, we uncovered a set of tiles exhibiting differential activation or repression with treatment (Fig. 3K, Table S4). In particular, 81% of tiles that could activate the reporter within 24 hours (148 of 183 normally strong activators) exhibited reduced activation under p300/CBP KIX inhibition, with 60% of the sensitive tiles (89/148) containing flexiNR or NR box motifs that may directly bind the KIX domain (Fig. 3L). Among these were tiles from the EBV EBNA2 (explored in detail later) and KSHV RTA activation domains, both of which recruit p300/CBP via the KIX domain for their activities[40,41]. The remaining 40% of sensitive tiles that lacked these motifs may bind intermediary proteins containing flexiNR or NR motifs that then recruit p300/CBP. In contrast, only 18 tiles exhibited reduced activation upon p300/CBP bromodomain inhibition (top 10 in Fig. 3K); the majority of these tiles are normally weak activators. These hits seem to depend on histone acetylation for their activity and include tiles from DBP, a family which we examine in more detail in a later section. For the EZH2 and HDAC IIa inhibition screens, we identified tiles from 20 and 83 proteins, respectively, that exhibited reduced repression upon inhibitor treatment, although these changes were modest (top 10 in Fig. 3K, Table S4). Among the effectors sensitive to EZH2 inhibition, we find sequences from U84, a protein in which we discovered a strong repression domain (Fig. 2I, S3G), as well as sequences from better studied proteins, including EBNA3, IE1, and IE2 (Fig. 3K). Among the tiles sensitive to HDAC IIa inhibition, we find many sequences (64%, 82/129 tiles) that contain the NR or flexiNR motifs (Fig. 3L), suggesting these motifs recruit CoRs associated with the deacetylation pathway. These chemical screens, in conjunction with the sequence perturbations, can serve as a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

springboard for in-depth investigation of the molecular mechanisms associated with each effector domain.

Sequence and functional comparison of EBNA family effector domains

The HHV tiling screen identified transcriptional effector domains of varying strengths (weak to very strong) within natural variants of the EBNA family proteins from different EBV strains. EBV strains are broadly classified into two subtypes, where EBV type 1, which includes the prototypical B95-8 strain, is associated with greater prevalence and malignancy than EBV type 2, which includes the prototypical AG876 strain (Fig. 4A)[42,43]. This typing classification is primarily driven by sequence differences between type 1 and type 2 homologs of the EBNA family proteins (Fig. 4B)[44]. For homologs of both types, we identified activation and repression domains in EBNA2 and EBNA3 as well as repression domains in EBNA1, EBNA4, and EBNA6 (Fig. 4C, Fig. S5A). These findings are consistent with previous studies that identified transcriptional corepressors and coactivators as interaction partners of the EBNA proteins[45]. While most analogous effector domains across the EBV subtypes are comparable in strength, EBNA2 type 1 and type 2 homologs exhibit pronounced differences in activation and repression domain strengths (Fig. 4C–E).

The EBNA2 transcriptional activation domain we identified at the C-terminus is well-characterized, contains a mixture of acidic and non-polar residues typical of activators, and is critical for cellular transformation (Fig. 4F, Fig. 4H) - a function that is preserved even when this domain is substituted with the VP16 activation domain[46]. Our perturbation screen found two critical tryptophan residues in each of the EBNA2 activation domains of both prototypical EBV subtypes (W444 and W458 for B95-8; W411 and W425 for AG876). The latter critical tryptophan residue is part of a flexiNR box motif in both homologs that has been shown to bind both TFIIH and p300/CBP[40], has also previously been shown to abolish transcriptional activation in a reporter assay, and is essential for B cell transformation in a system with recombinant virus[46]. Mutation of serine to aspartic acid at residue 409 in the weaker AG876 EBNA2 homolog restored activation levels to that of the stronger type 1 homolog (Fig. 4H starred, Fig. S5B) whose natural sequence includes an aspartic acid at residue 442.

In contrast, less is known about the transcriptional effector potential of the EBNA2 N-terminal region where we discovered a repression domain. This region is annotated as containing a self-interaction domain that was recently reported to be important in phase separation and chromatin reorganization[47], as well as two conserved regions present in EBV and non-human primate lymphocryptovirus EBNA2 homologs[48] (Fig. 4F). Our perturbation screen identified two essential regions for transcriptional repression that almost perfectly map to conserved regions 3 and 4 (Fig. 4G). Deletion of these conserved regions has been shown to result in no or poor transformation, respectively, in a system with recombinant virus[49]. The second essential region contains a methionine-rich, predicted alpha helix that resembles portions of the KSHV VIRF3 and KbZIP proteins that are important for transcriptional repression, suggesting that methionine-rich sequences may be able to coordinate interactions with corepressors. In particular, this region of EBNA2 (aa121-213) is known to bind DDX20, an RNA helicase reported to have transcriptional repressive

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function[50], although it remains to be seen whether this factor contributes to the repression we measure in our assay.

Given that EBNA2 is known to associate with chromatin via its interactions with CBF1 (ubiquitously expressed) and EBF1 (not expressed in K562 cells) to affect target genes[51], we reasoned that differences in effector domain strengths measured at a synthetic reporter gene with our tiling screen might translate to measurable differences in host gene expression upon expression of full-length EBNA2 proteins. To test this hypothesis, we expressed full-length type 1 or type 2 EBNA2 for 48 hours and harvested cells for gene expression profiling by RNA-seq and differential expression analysis by DESeq2[52] (Methods, Table S5). For both types, we observed significant changes in genes enriched in GO terms related to aspects of the immune response (Table S5). Type 1 EBNA2 produced greater changes in expression compared to type 2 (Fig. 4I, Fig. S5C&D), consistent with our tiling screen measurements of stronger type 1 effector domains (Fig. 4C). Wild-type type 2 EBNA2 produced gene expression changes more similar to a chimera containing the type 2 effector domains and type 1 CBF1-binding domain than to a chimera containing the type 1 effector domains and type 2 CBF1-binding domain (Fig. S5C&D), suggesting that these effector domains can influence genomic targets. Taken together, these data support our assay's ability to measure differences in transcriptional effector activity between natural sequence variants both in the context of recruiting protein fragments and expressing the full-length protein.

Investigating the importance of cofactor interaction motifs on VIRF protein functions

We identified some of the strongest herpesvirus activator, repressor, and dual effector domains in three of the KSHV viral interferon regulatory factors (VIRFs) (Fig. S6A–C), which are homologous to and interact with the human IRF proteins to modulate immune signaling[53,54]. Despite the homology between the viral and human IRF N-terminal DNA-binding domains, the effector domains of VIRF2, VIRF3, and VIRF4 differ substantially in sequence from their human counterparts (Fig. 5A).

The VIRFs also differ from each other in the type and number of domains they have (Fig. 5A, Fig. S6A–C) and the sequences necessary for their activity (Fig. 5A–D). For instance, VIRF2-4 each have a dual effector domain that activates minCMV and represses pEF (Fig. 5A red, Fig. S6A–C). The dual effector domains of VIRF2 and VIRF3 are structurally similar (predicted alpha helices), and each contain two regions that affect their function: two NR box motifs for VIRF2 (Fig. 5B) and one NR box motif and a methionine-rich sequence (MDMLM) in VIRF3 (Fig. 5C). Deletion of the NR box motif within the VIRF3 dual effector domain completely abolishes activation and repression, while deletion of either of these two motifs in the VIRF2 dual effector domain abolishes repression but only somewhat reduces activation (Fig. 5E–F, Fig. 5H, Fig. S6D–G). These results suggest that this motif is either bound by a single cofactor with dual transcriptional effector activities or competitively by multiple CoAs and CoRs. One candidate cofactor for the first scenario is p300/CBP, which, in addition to its well-described CoA function, can mediate transcriptional repression when SUMOylated by recruiting HDAC6[55]. Deletion of the methioninerich region from the dual domain of VIRF3 produces a similar effect to deletion of the NR box motif: it decreases activator potential and breaks repressive function.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

VIRF4 has four effector domains (Fig. 5A, Fig. S6C), none of which have been described before nor shown to interact with specific cofactors: a weak repression domain; an unstructured dual effector domain containing four critical tryptophans and several important aromatic residues (Fig. S6H); a moderate-strength repression domain containing key aspartic acid and threonine residues (Fig. S6I); and a strong activation domain consisting of an alpha helix with an essential tryptophan (W671) adjacent to a flexiNR box motif (LxxIL). (Fig. 5D&G, Fig. S6J).

In order to understand whether the essential regions and key residues identified in our HHV perturbation screen are functionally relevant in the context of the full-length proteins, we compared the consequences of expressing full-length wild-type or mutant VIRFs on host gene expression (Fig. 5I, Table S5). First, to understand how wild-type VIRF proteins modulate type I interferon (IFN), we expressed these proteins with or without IFN-beta treatment and performed RNA-seq. We first confirmed that IFN-beta treatment in the absence of viral protein expression led to the activation of classical interferon-stimulated genes (ISGs) (Fig. S6K). The total number of differentially expressed genes was smaller with IFN-beta treatment (n = 193, log2 fold change cutoffs of -1.5 and 1.5) than any of the conditions in which VIRF2, VIRF3, or VIRF4 were overexpressed on their own (n = 1,291, n = 970, and n = 215, respectively). Overexpression of VIRF2 or VIRF3 with IFN-beta treatment prevents the activation of a small number of genes (n = 10 and n = 9, respectively) that were weakly activated by IFN-beta treatment alone (Fig. S6L), but has no effect on ISGs that are normally strongly activated. VIRF2, VIRF3, and VIRF4 overexpression alone actually activates a subset of ISGs, suggesting that these proteins can partially phenocopy IFN-beta treatment (Fig. S6M–O). In particular, VIRF3 overexpression increases the expression of human IRF7.

Next, we performed RNA-seq expression measurements when overexpressing the full length VIRF2 and VIRF4 proteins with mutations in the transcriptional effector domains we identified through the tiling and mutational screens. Overall, of the genes that are up- or down-regulated upon WT VIRF2 expression, including the ISGs, fewer of them change significantly upon expression of an NR box mutant (deletion of residues L321-S325), and the changes are smaller (Fig. 5J), consistent with this mutation decreasing activation and abolishing repression (Fig. 5E–F). Similarly, of the genes upregulated by WT VIRF4, including the ISGs, fewer are upregulated and to a lesser extent by the W671A mutant (Fig. 5K) as expected from this mutation abolishing one of the VIRF4 activation domains (Fig. 5G&H). These results are not due to differences in protein levels between the WT and mutant VIRFs (Fig. S6P). This finding suggests that indeed the same amino acids that we identified to be important for reporter activation are also important for controlling endogenous genes in the context of the full-length protein.

The herpesvirus DBP C-terminus regulates late gene expression and replication

Tiling across all proteins from herpesviruses identified previously unannotated, moderate-strength C-terminal transcriptional activation domains within six of the ten homologs of the herpesvirus single-stranded DNA-binding protein (DBP) (Table S3, Fig. 6A–C, Fig. S7A–C). The DBPs are classically associated with herpesvirus genome replication, which

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

is required for expression of late genes that encode proteins important in virion assembly[56]. Although the vTR library contained several DBPs, their inclusion in the census was due to their ability to bind single-stranded DNA rather than direct evidence for modulation of transcription.

We identified this conserved activation domain in all four alphaherpesvirus homologs (HSV1, HSV2, VZV, and SuHV1) (Fig. 6A, S7A–C), one betaherpesvirus homolog (HHV7) (Fig. 6B), and one gammaherpesvirus homolog (KSHV) (Fig. 6C). We also detected mild repression potential in the same domains from HSV1, HSV2, and HHV7 (Fig. 6A&B, Fig. S7A). While these domains resemble typical activation domains in that they consist of hydrophobic residues interspersed with acidic ones, the six homologs with activity do differ in sequence across alpha, beta, and gammaherpesvirus subfamilies and in their essential regions (Fig. 6D, Fig. S7D–G). For example, HHV7 DBP residues 1112–1116 overlap a SUMOylation site (LKCE) that we generally find in repressive domains, and its deletion strongly increases activation (Fig. S7F–G). In contrast, KSHV DBP contains four flexiNR box motifs, three of which overlap essential regions (Fig. 6E). Moreover, tiles from the different DBP homologs show different sensitivity to the p300/CBP bromodomain inhibitor, with the KSHV DBP C-terminal activation domain being the most sensitive activation domain in that screen (Fig. 3M, Fig. 6F). Taken together, we hypothesize that these C-terminal activation domains are biologically relevant and functionally conserved despite sequence and even mechanistic divergence.

Deletion within the C-terminal region of the HSV1 DBP homolog has been shown to inhibit both replication and late gene expression[57], but this has not been shown for beta or gammaherpesvirus DBP homologs. In order to test our hypothesis that the C-terminal activation domain of the gammaherpesvirus KSHV DBP (residues 1053–1132) is important for late gene transcription, we used CRISPR/Cas9 to perturb this region of the KSHV genome on a bacterial artificial chromosome (BAC) in iSLK cells[58]. In this system, latently infected cells harbor an mIFP2 reporter under the control of a late gene promoter, and the expression of a dox-inducible RTA protein can reactivate these cells from latent to lytic infection (active production of KSHV virions) (Fig. 6G). Complete knockout of DBP in this cell culture model has been shown to prevent late gene expression as measured by a lack of mIFP2 48-hr after reactivation[58]. In this study, sgRNAs targeting positions corresponding to residues 1017, 1032, 1052, and 1076 (within one of the critical regions determined in our perturbation screen (Fig. 6E)) each reduced mIFP2 levels to the same degree as sgRNAs targeting the beginning of the DBP gene, indicating that deletion of this C-terminal region is functionally equivalent to complete knockout of DBP (Fig. 6H). EdU staining showed that viral replication was also impaired (Fig. 6H). Taken together, these and prior data suggest that the KSHV DBP C-terminus is critical for viral genome replication and late gene expression. However, it remains unclear whether DBP's function in transcriptional activation is independent of its effect on DNA replication since we have not identified a mutant that only affects late gene transcription. Since the action of this domain depends on p300/CBP (Fig. 6F), and p300/CBP recruitment generally leads to both chromatin opening and gene activation, it is possible that p300/CBP recruitment by DBP helps viral replication by opening the local chromatin.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Discussion

Viral proteins can control transcription of viral genomes and reprogram it in host cells. However, outside of a small set of viral transcriptional effector proteins that has been deeply characterized over the past several decades, most viral proteins lack functional and domain annotations supported by experimental evidence. While experimental throughput has historically been a limiting factor in gleaning this knowledge, advancements in DNA synthesis and sequencing have enabled quantitative measurements of protein functions at scale in human cells. Here, we employ high-throughput quantitative approaches to investigate transcriptional regulation across over 60,000 protein fragments across more than 1,500 proteins that span the entire proteomes of 11 coronaviruses and nine human herpesviruses. Specifically, we identify the proteins that harbor activating or repressive transcriptional domains, determine where in the proteins these domains are, and interrogate the sequence features responsible for these functions. Moreover, for a subset of these proteins, we investigated the mechanistic details and consequences of these activities on host cell mRNA expression and the viral life cycle.

We first investigated a set of putative and known viral transcriptional regulators, or vTRs, to assess the efficacy of HT-recruit in recovering transcriptional effector domains. For example, we were able to identify the well-described activation domains within the HSV1 VP16 C-terminus and VZV VP16 N-terminus, as well as several effector domains that had been described for HAdV5 E1A proteins. In addition, we localized transcriptional repression activity to the N-terminus of four VP16 homologs. To our knowledge, this study is the first to directly compare the strengths of multiple VP16 and E1A effector domains across homologs. Overall, our assay identified transcriptional regulatory domains in over one hundred proteins included in the vTR census (117/377). While all vTR members have some evidence supporting their inclusion in the census (such as DNA or RNA binding), it is possible that members of the vTR census in which we do not identify any effector domain with our method either: 1) have effector domains where the necessary sequence is larger than 80aa, 2) require other viral or human cofactors that are not present in our cells, or 3) bind DNA but do not contain transcriptional effector domains, and enact their function in cells by competing with human transcription factors for binding across the genome. The approaches developed here can be further extended to address these questions. In addition, the vTR library contains a mixture of proteins that may act on DNA and/or RNA substrates, yet we only measured the ability to affect transcription from a dsDNA template. Indeed, the authors of the vTR census show that the viral genome type that encodes a protein is generally concordant with the protein's particular substrate[7]. This agrees with the enrichment of transcriptional effector domains identified in dsDNA virus proteins and the relative dearth identified in RNA virus proteins in our reporter assays. A similar high-throughput approach involving recruitment of viral proteins to RNA reporters could be used to measure their ability to affect RNA degradation or translation into protein.

In our unbiased screens for coronavirus-encoded transcriptional regulators, we identified relatively few hits, which, as discussed above, is unsurprising for a family of RNA viruses. However, we found that all 11 Spike homologs tiled in this library harbored a repression domain mapping to heptad repeat 1 (HR1). These findings are unexpected given the classical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

role of viral transmembrane glycoproteins as critical factors in tropism, fusion, and entry[59]. In the native context, this region (Spike-095) appears to stabilize trimers of the Spike S2 fragments through hydrophobic interactions at the trimerization interface. However, deletion and deep mutational scanning of Spike-095 when fused to rTetR suggest that the normally inaccessible face of monomeric Spike-095 is primarily responsible for the activity we measured in our assay. This region is a leucine zipper that may heterodimerize with other leucine zipper-containing human repressors. While Spike could be cleaved by human proteases to liberate a fragment containing the repression domain we identified[60,61], it is unclear how such a fragment could escape from the endolysosomal pathway. Recently, the Spike protein has been detected in the nucleus, albeit at low frequency[62]. However, there has not been prior evidence for Spike being involved in transcriptional regulation. Therefore, it remains to be determined whether our findings about small Spike protein fragments are physiologically relevant upon coronavirus infection. This raises an important point: protein domains that are transcriptional activators or repressors at a reporter gene should be interpreted in combination with information regarding domain exposure within the full-length protein structure, protein localization, and protein-protein interaction data.

Our unbiased screens for herpesvirus-encoded transcriptional effectors identified activation or repression domains in 178 of the 891 proteins (20%) included in the tiling library, a high hit rate. At the most basic level, viruses need to enter a host cell, replicate their genome, and produce structural components to package these genome copies. Only a small set of genes is required for these processes (e.g. five total for rabies virus), yet herpesviruses encode 80–200 genes each, with many of these genes harboring transcriptional effector domains as measured in our assay. This finding suggests that together these proteins may enable more complex regulation of viral and host gene expression. This may be reflected in the unique features of herpesviruses, such as their near universal prevalence, lifelong infection with repeated transitions between active and inactive states, their immune evasiveness, and their implication in chronic, autoimmune, and neurodegenerative diseases[63–65]. Some of the strongest effectors we identified were late gene proteins and latency factors from gammaherpesviruses, which can infect many cells but establish latency in B cells. It is possible that our K562 model cells, which are also derived from bone marrow, express similar transcriptional and chromatin cofactors. Follow-up studies that screen herpesvirus proteins in different cell types will provide mechanistic insight into cell-type specific consequences of infection.

It is also important to understand how the activities of isolated effector domains connect to the function of the full-length protein (Fig. 7), especially in models relevant to viral infection. For example, domains that activate or repress transcription when recruited to a reporter gene (Fig. 7A&F) could act as direct activators or repressors of viral or human genes if there is a DNA-binding domain in the viral protein (Fig 7B&G). Alternatively, since many coactivators and corepressors are chromatin regulators, these domains could be involved in remodeling chromatin during other chromatin-templated processes, such as viral replication or latency (Fig. 7C&H). Finally, in some cases, reporter-identified activation or repression domains could indirectly lead to the opposite effect on human genes by sequestering cofactors away (Fig. 7D&I), or, within a protein that has a strong domain of the opposite type, by contributing to its localization to active or repressed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

genes (Fig. 7E&J). To understand the role of the domains we discovered in the context of full-length proteins one should first check the full-length protein structure to determine if the domain is exposed, and then check for predicted/known DNA-binding domains, and protein localization. We did find a positive correlation between effector strength and the likelihood that the full-length protein contains a predicted nuclear localization signal (Fig. S3Q&R), suggesting that many of these proteins are likely to act in the nucleus. Moreover, when we expressed individual full-length proteins that contain predicted or known DNA-binding domains, we measured significant changes in host gene expression (e.g. HHV7 U8 and U84 in Fig. 3I, VIRFs in Fig. 6J&K), and these changes were affected by mutations that impaired effector domain function (VIRFs in Fig. 6J&K). In the case of the EBV EBNA2 proteins, which indirectly associate with DNA via the human CBF1 repressive protein, we found that the effector domain sequences themselves appear to influence genomic targets (Fig. S5D). It is conceivable that the repression domain we detected in our assay for EBNA2 reflects a protein-protein interaction important for its localization to repressed genes (similar to the CBF1 binding domain), and once there, the strong C-terminal activation domain allows EBNA2 to de-repress these genes. Finally, for one member of the family of activation domains we discovered in the herpesvirus DBPs, we tested its role within the context of the full-length protein and a KSHV viral model, and showed that deletions of this domain impair viral replication and late-gene expression and that its action depends on the histone acetyltransferase p300/CBP. We hope this large quantitative dataset (Data S1) will massively expand herpesvirus protein annotations, which are largely lacking, and that similar integrative approaches with full-length viral proteins and mutants that lack transcriptional activity will help virologists interpret and build upon these findings.

Sequence analyses and perturbations revealed that herpesvirus activators and repressors share some properties with human ones, which is unsurprising given that they must work with host machinery. One key finding unique to herpesvirus effectors is the importance of tryptophan residues to both activation and repression: 73% and 67%, respectively, of all single substitutions reduced or completely abolished activity. These critical tryptophans tend to be surrounded by acidic residues and some hydrophobic residues for many effectors, suggesting that there may be a way to predict critical regions of activity and determine whether these rules extend to other dsDNA viruses. Related to this trend, we identified several variations of the multifunctional NR box motif (LxxLL) within essential regions of herpesvirus activation and repression domains, as well as essential regions with no known motif. Future studies using wild-type and mutant proteins lacking these essential regions can elucidate the interaction partners associated with the sequence motifs we discovered and responsible for these effector activities. More broadly, understanding the rules that underlie activation and repression would enable protein engineers to design novel, synthetic transcriptional effectors. In the meantime, this study provides a rich repertoire of short activator, repressor, and dual effector domains spanning a range of strengths and acting through a variety of cofactors that should expand and improve the synthetic gene regulation toolkit.

The inclusion of homologs from different viral species and strains allow us to appreciate 1) how function can be conserved despite natural sequence variation (e.g. DBP activation domain) as well as 2) how homolog-specific functions can arise despite high sequence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

similarity (e.g. EBNA2 repression domain). The ability to design and simultaneously test the functional consequences of thousands of deletions and substitutions allows us to map essential regions of activity for hundreds of effector domains, something that has not been possible in systems with live virus. Additional screens with chemical inhibition of chromatin-modifying enzymes and investigation of host gene expression changes in the presence of individually expressed viral proteins help further elucidate the mechanisms and consequences of these transcriptional regulatory activities. Thus, these high-throughput, quantitative synthetic biology approaches provide a powerful way to understand the physical basis for viral protein function and complement traditional virological methods, with the added benefit of enabling investigation of proteins from viruses that otherwise cannot be easily grown in cell culture. This knowledge will facilitate *in silico* drug screening and the development of antivirals and vaccines. In summary, our catalog of viral protein sequences that act as transcriptional effectors in human cells, together with their functional mutants (Data S1), can serve as a resource for interpreting viral protein function and sourcing components for synthetic biology tools.

STAR Methods

Lead Contact and Materials Availability

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lacramioara Bintu (lbintu@stanford.edu).

Materials Availability—The lentiviral vector for inducible transgene expression generated in this study has been deposited at Addgene as pCL040 lenti TRE-3xFLAG-LibCloneSite pEF-rTetR(SE-G72P)-VP48-T2A-mCherry-BSD-WPRE, catalog number #198054. Information for the previously published reporter plasmids and rTetR-fusion recruitment vectors is listed in the key resources table.

Data and Code Availability

- Raw and processed RNA sequencing files (i.e. FASTQ, counts, and DESeq2 outputs) have been deposited at NCBI GEO and at NCBI SRA. Raw HT-Recruit next-generation sequencing files (i.e. FASTQ) have been deposited at NCBI SRA. All sequencing data are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper also analyzes existing, publicly available data. Accession numbers for these datasets are listed in the key resources table.
- All other data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at https://github.com/bintulab/Viral_Ludwig_2022 and is publicly available as of the date of publication. The DOI for the corresponding release is listed in the key resources table.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental Model and Subject Details

Cell lines and cell culture—K562 cells (ATCC #CCL-243), which were used for all recruitment and gene expression profiling experiments, were cultured in RPMI 1640 (Gibco #11875–119) supplemented with 10% FBS (Omega Scientific #FB-15) and 1% Penicillin-Streptomycin-Glutamine (Gibco #10378–016). K562 reporter cell lines were the same as those used in the original HT-recruit study[8] and were generated by TALEN-mediated homology-directed repair to integrate donor constructs (JT039 with EF1a reporter: Addgene #161927; DY032 with minCMV reporter: Addgene #161928) into the *AAVS1* locus using hAAVS1 1L TALEN (Addgene #35431) and hAAVS1 1R TALEN (Addgene #35432). These cell lines were not authenticated. HEK293T Lenti-X cells (Takara #632180), which were used to package lentivirus as described in a later section, were cultured in DMEM with GlutaMAX (Gibco #10566–024) supplemented with 10% FBS (Omega Scientific) and 1% Penicillin-Streptomycin (Gibco #15140–122). iSLK cells latently infected with a version of the BAC16 KSHV genome[66] and modified to express Cas9 and contain a reporter of late gene activity (K8.1 promoter driving an mIFP2 fluorescent cassette)[58] were used for measuring KSHV late gene expression and replication defects as described in a later section. These cells were cultured in DMEM (Gibco #11965–092) supplemented with 10% FBS (Peak Serum), 1% Penicillin-Streptomycin (Gibco), 1X GlutaMAX (Gibco #35050–061), 1 μg/mL puromycin, 50 μg/mL G418, 10 μg/mL blasticidin, and 125 μg/mL hygromycin. All cell lines were cultured in a controlled humidified incubator at 37°C with 5% CO_2 and tested negative for mycoplasma.

Method Details

Lentiviral transduction—For small-scale lentivirus production, HEK293T Lenti-X cells were seeded into 6-well plates at 1 x 10^6 cells per well in 2mL. The next day, cells were transfected with 750ng of an equimolar mixture of the three third-generation envelope and packaging plasmids (pMD2.G: Addgene #12259; pRSV-Rev: Addgene #12253; pMDLg/pRRE: Addgene #12251, all gifts from Didier Trono) and 750ng of lentiviral transfer plasmid encoding the transgene of interest after a 15 minute incubation of these plasmids with 10μL of polyethylenimine (PEI, Polysciences #23966). Lentivirus-containing culture supernatant was harvested 72 hours after transfection, passed through a 0.45μm PES filter (CELLTREAT #229749), and added undiluted to K562 cells for a final cell concentration of 3–4 x 10^5 cells/mL for pJT126-based effector recruitment vectors (Addgene #161926) or 1–2 x 10^5 cells/mL for pCL040-based inducible protein expression vectors (Addgene #198054) to account for differences in infection efficiency. The cells were spinfected as follows: the cell-virus suspension was centrifuged in 15mL conicals at 1,000 x g for 2 hours at 33°C, after which the supernatant was removed from the cells (decanted), decontaminated, and discarded; cells were subsequently cultured for two days in fresh media to allow for integration and expression of mCherry and blasticidin resistance. Cells were treated with 10μg/mL Blasticidin S HCl (Gibco #A1113903) from days 2–10 post-infection to select for successfully transduced cells. Starting at day 3 post-infection, selection efficiency was

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

monitored regularly by measuring mCherry positivity on a Bio-Rad ZE5 Cell Analyzer (Bio-Rad #12004278).

For screen-scale lentivirus production, HEK293T Lenti-X cells were seeded into 15-cm dishes at 13 x 10^6 cells per dish in 30mL (approximately one dish per 5,000 library elements). The next day, cells were transfected with 11μg of an equimolar mixture of the three third-generation envelope and packaging plasmids and 11μg of the library of lentiviral transfer plasmids with 150μL of PEI. A full media change was performed 24 hours after transfection, and lentivirus-containing culture supernatant was harvested at 72 hours after transfection, applied to a 0.45μm PES filter unit (Thermo Scientific #1680045), and a fraction was used for titration to determine the appropriate dilution for approximately 25% mCherry-positive cells (equivalent to an MOI of 0.3 where approximately 90% of infected cells only receive one library member).

vTR tiling library design—Protein sequences and metadata for the 419 human virus transcriptional regulators included in Table S2 of the vTR census study7 were downloaded from UniProt; however, only the 377 proteins from non-BSL4 viruses were considered for tiling due to safety concerns. Protein tiles of 80aa in length were generated in 10aa increments along each protein, and duplicates were removed, using generate_tiles_v2.py. Protein tile sequences were reverse-translated and codon-optimized using the Python package DNAchisel in domains_to_codon_opt_oligos_v2.py. Our codon optimization approach matched codon usage to natural human frequencies, excluded BsmBI sites, excluded C homopolymers greater than seven in length, enforced a local GC content of 20–70% within a 50bp window, and enforced an initial maximum global GC content of 65% that was incrementally relaxed by 1% if optimization failed. To the resulting 13,133 tiles, we added 1,500 80aa-long random negative controls whose codon usage matched natural human frequencies (generateRandomers.py), and 386 additional sequences that would serve as fiducial controls across screens (Table S1). To the 5’ and 3’ ends of all sequences, we appended BsmBI restriction sites for scarless Golden Gate cloning and library-specific primer handles for amplification by PCR, yielding a final length of 300nt for every oligonucleotide in the library. The vTR, CoV, and HHV libraries (design discussed below) were ordered as a single oligonucleotide pool from Twist Biosciences.

CoV tiling library design—Protein sequences and metadata for the entire proteomes of 11 human and closely related bat coronaviruses were downloaded from UniProt, with most, but not all, of these sequences reviewed. For all ORF1a and ORF1ab polyprotein entries, we used the PTM/Processing > Chain information in UniProt to extract each of the individual non-structural protein sequences for tiling (polyprotein2chains.py). For BtCoV-RaTG13 Orf1ab, which lacks Chain information in UniProt, we used the annotations of the near identical SARS-CoV-2 Orf1ab, accounting for the insertion of an isoleucine at residue 1023 for the SARS-CoV-2 homolog. Protein tiling, codon optimization, and appending restriction site and primer handle sequences were performed as above. The final CoV library comprised 7,564 unique coronavirus protein tiles, 850 80aa-long random negative controls, and 391 additional fiducial control sequences (Table S1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SARS-CoV-2 Spike perturbation library design—A multiple sequence alignment for all 11 full-length Spike homologs was performed with Clustal Omega to define the non-SARS-CoV-2 WT Spike sequences aligning to the repressive SARS-CoV-2 Spike-095 tile identified in the primary CoV tiling screen (UniProt P0DTC2, residues 941–1020). All other non-control library members were perturbations of the SARS-CoV-2 Spike-095 tile sequence and were generated in the following way: 1) 5aa deletion scanning with a step size of 1aa across the entire 80aa tile; 2) double and triple alanine scanning across the entire 80aa tile; and 3) a deep mutational scan of the 'core' region (residues 941–980) representing the intersection of all repressive tiles within the domain. Altogether, the library comprised 9 WT protein sequences, 68 deletions, 69 consecutive double alanine substitutions, 64 consecutive triple alanine substitution, 760 single substitution elements in the core region, 100 non-consecutive double alanine substitutions within the core region along the external trimer face, and one non-consecutive 15-residue alanine substitution of this same face. To assess the consequence of codon variation on protein function, three alternatively encoded oligonucleotides were designed for each of the unique WT sequence and perturbations described above, except for the SARS-CoV-2 WT Spike-095 sequence, which we alternatively encoded 10 different ways. The standard deviation in screen score measurements for the differentially coded tiles was small (mean ~0.1 screen score unit). The final library comprised 3,217 Spike-related elements, 381 80aa-long random negative controls from the vTR and CoV screens, and 100 additional fiducial control sequences. Codon optimization and appending restriction site and primer handle sequences were performed as above, except in the case of the deletion scanning elements, for which we added a filler sequence between the restriction site and primer handle sequence in order to maintain a uniform final oligonucleotide length of 300nt. This library (Table S1) was ordered as an oligonucleotide pool from Twist Bioscience.

HHV tiling library design—Protein sequences and metadata for nearly the entire proteomes of 9 human herpesviruses were downloaded from UniRef90, which collapses UniProt entries on 90% sequence identity and represents each resulting protein cluster with a single, reviewed sequence, using the following search term: uniprot:(herpesvirus host:human NOT molluscum reviewed:yes) AND identity:0.9. A similar search on UniRef90 was performed for Suid herpesvirus 1, which primarily infects pigs but is a commonly used model for studying alphaherpesvirus biology. Two human herpesvirus protein sequences contained at least one X (HHV6B Q1: UniProt Q9QJ11; HHV6B Q2: UniProt P0DOE1), which required manual correction based on other entries in the cluster and the literature. Protein tiling, codon optimization, and appending restriction site and primer handle sequences were performed as above. The final library comprised 11,856 unique alphaherpesvirus protein tiles, 13,679 unique betaherpesvirus protein tiles, 7,434 unique gammaherpesvirus protein tiles, 3,650 80aa-long random negative controls, and 413 additional fiducial control sequences (Table S2).

HHV perturbation library design—Transcriptional effector domains identified in the primary HHV tiling screen were represented by their strongest tile. Screen scores were converted into estimated percent activation or repression based on the fit to the individual validation data described by the following logistic function:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PercentON=A/(1+ê(-k*ScreenScoreLog2(ON:OFF)))+B for activation, PercentOFF=A/(1+ê(-k*ScreenScoreLog2(OFF:ON)))+B for repression, where A, B and k are fitting parameters.

Only tiles whose percent activation or repression was estimated to be at least 40% (based on their screen scores and the equations above) were considered for perturbation in order to be able to measure appreciable differences in activity and to be able to test a larger set of perturbations for each tile. This criterion yielded 43 activator and 55 repressor tiles. While 21 of these activator and repressor tiles had some degree of dual effector activity, only eight met the 40% threshold for both activation and repression and were considered strong dual effector tiles. The protein-level deletions and substitutions described in the main text were generated with Python scripts, these sequences were reverse-translated and codon-optimized using DNAchisel, and restriction site and primer handle sequences were appended as above. Altogether, the library comprised 97 WT sequences, 1,567 deletions, and 6,129 substitutions: 268 F:A, 112 W:A, 171 Y:A, 557 D:A, 557 D:N, 621 E:A, 621 E:Q, 384 R:A, 218 K:A, 218 K:Q, 218 K:R, 645 S:D, 506 T:D, 327 Q:A, and 706 P:A. Given the low variability observed in the SARS-CoV-2 Spike perturbation screen between alternatively coded library members, each HHV perturbation library member was only encoded one way. To these 7,794 elements, we added 320 80aa-long random negative controls from the HHV tiling screens and 100 additional fiducial control sequences (Table S4). This library was ordered as an oligonucleotide pool from Twist Bioscience.

HHV hits library design for chemical inhibition screens—All library members that were above the activation or repression detection thresholds in the primary HHV tiling screens were included in the HHV hits library without modification (i.e. identical DNA sequence) except for the primer handle sequences. To the 194 activation hits, 74 dual effector hits, and 553 repression hits, we added 50 80aa-long random negative controls from the vTR and CoV screens and 71 additional fiducial control sequences. This library (Table S4) was ordered as an oligonucleotide pool from Twist Bioscience.

Tiling and perturbation library cloning—Twist oligonucleotide pools were resuspended to a concentration of 10ng/μL in 10mM Tris-HCl pH 8.0 with 1mM EDTA. Individual libraries (e.g. vTR, CoV, HHV, etc.) were selectively PCR amplified using library-specific primers annealing to the primer handle sequences flanking each oligonucleotide. Between two to six 50μL PCR reactions were performed for each library to produce enough product for downstream cloning steps, and all reactions were prepared in a pre-PCR hood to mitigate DNA contamination. Each 50μL reaction consisted of 10μL of 5X Herculase II Reaction Buffer (Agilent #600675), 34.5μL of nuclease-free water, 0.5μL of 10ng/μL template (5ng total), 1μL of each 10μM primer, 1μL of DMSO, 1μL of 10nM dNTPs, and 1μL of Herculase II Fusion DNA Polymerase, added in that order. The thermocycling protocol was as follows: an initial denaturation at 98°C for 3 minutes; between 17 and 21 cycles of 98°C for 20s, 61°C for 20s, and 72°C for 30s; and a final extension at 72°C for 3 minutes. Initial small-scale test PCRs were performed to determine library-specific cycle numbers that yielded clean, visible amplicons suitable for gel extraction and not at saturation. Amplified oligonucleotide libraries were run on a 2%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TBE gel, the 300bp bands were excised, and DNA was extracted from the agarose using the QIAquick Gel Extraction Kit (Qiagen #28704).

The pJT126 lentiviral recruitment vector (Addgene #161926) was pre-digested with 10,000 U/mL Esp3I (NEB #R0734L) at a ratio of 1μL of enzyme per 5μg of plasmid at 37°C for 15 minutes, followed by heat inactivation at 65°C for 20 minutes. Pre-digested pJT126 was run on a 0.5% TAE gel long enough to cleanly excise the digested product, which was subsequently extracted from the agarose. Oligonucleotide libraries were cloned into this vector using the GoldenGate cloning method, with between 10 to 16 20μL reactions per library. Each 20μL GoldenGate reaction consisted of 2μL of 10X T4 DNA Ligase Reaction Buffer (NEB #B0202S), nuclease-free water, 75ng of pre-digested pJT126, 5ng of amplified oligonucleotide library, and 2μL of NEBridge Golden Gate Assembly Kit (BsmBI-v2) (NEB #E1602L), added in that order. The NEBridge kit contains both BsmBI-v2 (an isoschizomer of Esp3I) and T4 DNA ligase. GoldenGate reaction conditions were 65 cycles of 42°C for 5 minutes then 16°C for 5 minutes, followed by a final digest at 42°C for 5 minutes and heat inactivation at 70°C for 20 minutes. Reactions were pooled, purified and concentrated with the MinElute PCR Purification Kit (Qiagen #28004), and eluted in 6μL of nuclease-free water.

Endura Electrocompetent Cells (Lucigen #60242–2) were thawed on ice for 10 minutes, then 25μL of cells were mixed with 2μL of the purified/concentrated GoldenGate product and transferred to a Gene Pulser/MicroPulser Electroporation Cuvettes, 0.1cm gap (Bio-Rad #1652089). Cells were electroporated on a Gene Pulser Xcell Total System (Bio-Rad #1652660) with the following conditions: 1.8kV, 10μF, 600Ω, and 0.1cm distance. Immediately after, 2mL of 37°C SOC Recovery Medium (NEB #B9020) were added to the cuvette, the contents of which were mixed by gentle pipetting and subsequently transferred to a 14mL round-bottom tube for a 1-hour recovery in a 37°C bacterial shaker. After recovery, cells were plated across four 10” x 10” luria broth agar plates with 100μg/mL carbenicillin, with a small amount of the recovery reserved for 1:100 dilution plating in triplicate to estimate library coverage. Plates were incubated in a warm room (approximately 33°C) for 14 to 18 hours, after which colonies were harvested by addition of luria broth and scraping. Cells were pelleted at 4,000 x g for 20 minutes, and plasmid pools were extracted using the Qiagen Plasmid Maxi Kit (Qiagen #12162). To assess library quality and representation bias, library members were amplified from the plasmid pool by PCR with primers containing Illumina adaptors for readout by next generation sequencing.

High-throughput recruitment assay with vTR, CoV, Spike perturbation, and HHV libraries—K562 reporter cells were infected with lentiviral libraries by centrifugation at 1,000 x g for 2 hours. Infection was performed with two replicates per library and the following number of starting cells per replicate per reporter line: 45 x 10^6 cells for the pooled vTR and CoV libraries; 12.5 x 10^6 cells for the Spike perturbation library; 45 x 10^6 cells for the HHV tiling library; 15 x 10^6 cells for the HHV perturbation library; and 2.5 x 10^6 cells for the HHV hits library screened in the presence of chemical inhibitors. Estimates of infection coverage (the average number of cells infected with a given library member) are as follows: 420X and 330X for the minCMV and EF1a reporter lines, respectively, for the pooled vTR and CoV libraries; 900X for the EF1a reporter line infected with the Spike

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

perturbation library; 320X and 290X for the minCMV and EF1a reporter lines, respectively, infected with the HHV tiling library; 250X and 200X for the minCMV and EF1a reporter lines, respectively, infected with the HHV perturbation library; and 250X and 350X for the minCMV and EF1a reporter lines, respectively, infected with the HHV hits library. Cells were treated with 10μg/mL blasticidin (Gibco #A1113903) starting two days after infection for approximately five to seven days total when at least 80% of cells were mCherry positive as monitored by daily flow cytometry (cells were analyzed no earlier than three days after infection in compliance with safe lentivirus practice).

For the vTR/CoV and HHV libraries, cells were maintained in 1L spinner flasks with constant, gentle paddle rotation. For the Spike perturbation, HHV perturbation, and HHV hits libraries, cells were maintained in vented T175, T225, and T25 flasks, respectively. All cultures were maintained in log growth conditions with daily half-volume media changes to dilute cells back to approximately 5 x 10^5 cells/mL, making sure to never drop the maintenance coverage (the number of cells harboring a given library member) below the initial infection coverage. By the end of antibiotic selection, cells had been stably expressing the rTetR-tile fusion proteins for between seven to nine days. Recruitment of the rTetR-tile fusion proteins was induced by treating the cells with 1000ng/mL doxycycline hyclate (Tocris #4090) for two days for activation screens or for five days for repression screens. Half the amount of doxycycline was replenished each day under the assumption of a 24-hour half-life. For screens with the HHV hits library, cells were treated with 10μM SGC-CBP30 (Selleck Chemicals #S7256), 10μM tazemetostat (Selleck Chemicals #S7128), 10μM TMP269 (Selleck Chemicals #S7324), or DMSO (vehicle) for 24 hours prior to doxycycline addition and throughout the recruitment timecourse, with these chemical inhibitors replenished daily during media changes. For the activation screen with the HHV hits library and p300/CBP KIX domain inhibition, cells were first treated with 1μM celastrol (Tocris #3203) for 2 hours prior to doxycycline addition, and cells were harvested only 24 hours later because prolonged celastrol treatment is toxic. A time-matched DMSO control screen was also performed.

Magnetic separation for high-throughput recruitment assays—At assay endpoint, a volume of cells equivalent to 10,000X coverage was pelleted and washed twice with DPBS (Gibco #14190–250) to remove immunoglobulins from the FBS. Cell pellets were resuspended in magnetic separation wash buffer (2% BSA in DPBS) at a concentration of 23 x 10^6 cells/mL, and a small volume was reserved as an ‘input’ sample for analysis by flow cytometry (described below). In parallel, a volume of paramagnetic Dynabeads M-280 Protein G (Thermo Fisher, #10003D) of 3–9μL per 1 x 10^6 cells (scaled based on the rarity of the bound population) were diluted in five volumes of wash buffer, incubated on a magnetic stand for 2 minutes, cleared of the supernatant, and resuspended in the cell suspension. The cell-bead suspension was incubated at room temperature for 75 minutes on a nutator to allow adequate time for cells expressing the IgG surface marker to bind the protein G-functionalized Dynabeads, and the suspension was subsequently incubated on a magnetic stand for 5 minutes to separate bead-bound and unbound fractions. The unbound fraction was transferred to a new tube, which was incubated again on a magnetic stand for 5 minutes to clear the suspension of any remaining beads, and the unbound fraction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

was transferred to a final tube. All beads in the original and second tube were pooled by resuspending in a volume of wash buffer equivalent to the initial volume. This tube was incubated at room temperature for 15 additional minutes on a nutator, subsequently incubated on a magnetic stand for 5 minutes, and the unbound 'wash' fraction was transferred to a new tube. The remaining bead-bound cell fraction was resuspended in a volume of wash buffer equivalent to the initial volume, and all three fractions (unbound, wash, and bead-bound) as well as the input sample were run on a Bio-Rad ZE5 Cell Analyzer to assess the effectiveness of magnetic separation and to estimate the total number of cells recovered in each fraction. As expected with the dual surface marker-mCitrine reporter, the unbound fraction had low mCitrine and the bead-bound fraction had high mCitrine. In every screen, the wash fraction (typically less than 5% of the total sample) mCitrine distribution resembled that of the input sample, and thus this fraction was discarded. The unbound and bead-bound fractions were pelleted by centrifugation at 600 x g for 5 minutes, decanted, and frozen at −20°C.

Library preparation and sequencing—For all high-throughput recruitment assays, genomic DNA was extracted for pelleted cell fractions with one of the following: DNeasy Blood & Tissue Kit (Qiagen #69504) for fractions with fewer than 5 x 10^6 cells; QIAamp DNA Blood Midi Kit (Qiagen #51183) for fractions with 5–20 x 10^6 cells; and Blood & Cell Culture DNA Maxi Kit (Qiagen #13362) for fractions with 20–100 x 10^6 cells. During genomic DNA extraction, bead-bound fractions were incubated on a magnetic stand to remove beads prior to loading lysate onto the silica columns. Genomic DNA was eluted in Buffer EB (Qiagen #19086) rather than the provided Buffer AE (Qiagen #19077) to avoid inhibition of PCR.

Library members were amplified by PCR with primers containing Illumina adaptor extensions at a final concentration of 500nM. All PCRs were prepared in a pre-PCR hood to reduce the likelihood of contamination by amplicons and plasmids. Small-volume test PCRs across a range of cycle numbers were performed and visualized by gel electrophoresis to identify the optimal cycle number that yielded sufficient material for extraction without reaching saturation and without producing non-specific bands. Final DNA template concentrations of 100–200ng/μL were used when possible and standardized for a given fraction (unbound or bound) across screen replicates, and at least one-third of all extracted genomic DNA was used as input for PCR to preserve library representation, resulting in a variable number of PCRs for each fraction. PCRs for screens with the vTR and CoV libraries were performed using NEBNext High-Fidelity 2X PCR Master Mix (NEB #M0541L) with 33 cycles. PCRs for screens with all other libraries were performed using the NEBNext Ultra II Q5 Master Mix (NEB #M0544L) with 21 to 24 cycles. Thermocycling and subsequent steps were performed outside of the pre-PCR hood. The thermocycling protocol was as follows: an initial denaturation at 98°C for 3 minutes; the aforementioned number of cycles of 98°C for 10s, 63°C for 30s, and 72°C for 30s; and a final extension at 72°C for 3 minutes. All reaction products for a given fraction were pooled and mixed, and 150μL were subsequently run on a 1% TAE gel, extracted, purified using the QIAquick Gel Extraction Kit, and eluted with 30μL of Buffer EB. The concentrations of each sample were quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher #Q32854) on a Qubit

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4 Fluorometer (Thermo Fisher #Q33238), pooled with 15% PhiX Control v3 (Illumina #FC-110-3001), and sequenced one of the following ways: on an Illumina NextSeq 550 with 1 x 75 or 1 x 150 cycles, on an Illumina HiSeq 2000 with 2 x 150 cycles, or on an Illumina MiSeq with 2 x 150 cycles.

Sequencing analysis—Sequencing data was processed and analyzed using the HT-recruit Analyze code first described in the original study8 and available on GitHub (https://github.com/bintulab/HT-recruit-Analyze). Briefly, reads were demultiplexed with bcl2fastq (Illumina), aligned with 'makeCounts.py' to a reference (made using 'makeIndices.py'), and used to compute enrichment scores between unbound (OFF) and bound (ON) fractions for each library member using 'makeRhos.py'. Library members with fewer than five reads in both fractions for a given replicate were filtered out, while those with fewer than five reads in one fraction had their reads adjusted to five reads for that fraction to avoid inflation of enrichment scores. For all screens, library members with a sum of fewer than 50 reads between both fractions for both replicates were filtered out, as these would produce noisy enrichment scores. For all screens, the detection threshold above which we estimated we could measure transcriptional effector activity was set at two standard deviations above the mean enrichment score of the negative random control population.

Individual recruitment assay validations by flow cytometry—Library members selected for individual validation experiments were ordered as gene fragments from Twist or IDT and cloned into the pJT126 lentiviral recruitment vector using Golden Gate cloning. Lentivirus was prepared and used to transduce reporter cells in replicate as described above. Following selection, cells were split into two wells, one of which was untreated and the other treated with 1μg/mL doxycycline for 2 days for the activation assay or 5 days for the repression assay. Half-media changes were performed daily, replenishing doxycycline for the appropriate wells, and 10,000 cells from each well were passed through a 40μm filter to be analyzed on a Bio-Rad ZE5 Flow Analyzer daily to monitor changes to reporter transcriptional state. Data was analyzed using Cytoflow67, first gating events for viability and mCherry expression. Gates for mCitrine expression were set based on an rTetR-only negative control to compute the fraction of mCitrine ON and OFF cells on each day of doxycycline treatment. Additional analyses and visualizations were performed with custom Python scripts.

Estimation of 3xFLAG-tagged protein levels by anti-FLAG staining and flow cytometry—All effector recruiter fusions (pJT126 vector) and full-length viral proteins for inducible expression (pCL040 vector) were designed as fusions to a 3xFLAG epitope tag to enable estimation of protein levels by anti-FLAG staining. Briefly, Fix Buffer I (BD Biosciences #557870) was pre-warmed to 37°C for 15 minutes, and Perm Buffer III (BD Biosciences #558050) was pre-chilled on ice. Approximately 1 x 10^6 cells were harvested by centrifugation at 300 x g for 5 minutes and washed once with DPBS. Cells were resuspended in 50μL of Fix Buffer I, incubated at 37°C for 15 minutes for fixation, pelleted by centrifugation, and washed once with 500μL cold DBPS with 10% FBS. Cells were resuspended in 50μL of Perm Buffer III, incubated on ice for 30 minutes for permeabilization, pelleted by centrifugation, and washed once with 500μL cold DBPS with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

10% FBS. Cells were resuspended in an antibody solution containing 5μL DYKDDDDK Epitope Tag Alexa Fluor 647-conjugated Antibody (R&D Systems #IC8529R) and 45μL DBPS + 10% FBS and incubated in the dark at room temperature for one hour. Cells were pelleted by centrifugation, washed with 500μL cold DBPS with 10% FBS, resuspended in 250μL DBPS with 10% FBS, and filtered through a 40μm filter prior to analysis on a Bio-Rad ZE5 Flow Analyzer. Data was analyzed using Cytoflow, first gating samples for viability, and, in the case of effector recruiter fusions, for mCherry expression. Gates for FLAG positivity were set based on wild-type or uninduced cells lacking the 3xFLAG epitope. Additional analyses and visualizations were performed with custom Python scripts.

Multiple sequence alignment—Protein sequences to be aligned were compiled into a single FASTA file either manually or by querying multiple identifiers in UniProt and downloading a FASTA file of the compiled results. These files were run through Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the default settings and the 'ClustalW with character counts' output format. Alignment files were downloaded and either visualized in JalView or with custom python scripts using Biopython.

Bulk RNA-seq—Gene fragments encoding full-length wild-type or mutant viral proteins were ordered from Twist or IDT and cloned into the pCL040 lentiviral inducible expression vector. Lentivirus was prepared and used to transduce wild-type cells in replicate as described above. Following selection, cells were cultured in 12-well plates and treated with 1μg/mL doxycycline to induce expression of viral transgenes. For the experiment investigating VIRF activities under type I IFN signaling, cells were treated with 10ng/mL interferon beta (PeproTech, #300–02BC) for 24 hours starting 24 hours after initial doxycycline treatment. On day 2 post-induction for all experiments, approximately 1 x 10^6 cells were harvested by centrifugation at 300 x g for 5 minutes. RNA was extracted using the RNeasy Mini Kit (Qiagen #74104), with a volume of 600μL Buffer RLT for cell lysis and with the QIAshredder columns (Qiagen #79654) for lysate homogenization. For all samples, the RNA integrity number was 10 as assessed by the Stanford Protein and Nucleic Acid (PAN) Biotechnology Facility using the RNA Nano Kit (Agilent #5067–1511) on an Agilent Bioanalyzer. A total of 500ng of purified RNA was used as input for the NEBNext Ultra II RNA Library Prep Kit (NEB #E7770S), which first involved enrichment of polyadenylated mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490). All steps were performed in accordance with the NEB protocol, with nine PCR cycles used for library amplification. Library size distributions were determined using the High Sensitivity DNA Kit (Agilent #5067–4626), and sample concentrations were quantified using the Qubit dsDNA HS Assay Kit on a Qubit 4 Fluorometer. Samples were pooled at equimolar ratios and sequenced on either a NextSeq 550 with 2 x 37 cycles or a MiSeq with 2 x 150 cycles.

Sequencing reads were demultiplexed with bcl2fastq. A FASTA of the GRCh38 human reference genome build was modified to include the viral transgenes of interest as separate chromosomes. The resulting FASTA was used to construct both a custom reference transcriptome using hisat2-build and a custom GTF genome annotation file using the script 'make_transgene_gtf.py'. Paired reads were aligned to the custom reference using hisat2, and output SAM files were converted to BAM files using samtools. A differential

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

expression analysis was performed in R with the Bioconductor DESeq2 package52 using a set of custom R scripts that were largely based on the workflow and commands described in the following tutorial: http://bioconductor.org/help/course-materials/2016/CSAMA/lab-3-rnaseq/rnaseq_gene_CSAMA2016.pdf. Additional analyses and data visualization were performed in python with custom scripts.

Motif finding—Regular expressions describing short linear motifs (SLiMs) associated with gene regulation were pulled from the Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org/) and used by custom Python scripts for pattern matching within protein sequences of interest. For the HHV perturbation screen data, an initial search with 40 motifs was conducted, specifically aimed at comparing motif frequencies in regions whose deletion either 1) had no effect or enhanced effector activity, or 2) reduced or completely broke effector activity. Sixteen motifs were found at a higher rate within the latter category and were used for a second search focused on annotating the overlap between these motifs and the effector domain essential regions (those whose deletion completely breaks activity) to identify potential cofactors (Table S6).

The initial motif search included a proposed motif that we termed the flexiNR box based on its similarity to the traditional NR box motif (LxxLL). This motif was included on the basis of reported flexibility of the NR box in other human proteins32-35 and our own observations when examining the data, and it initially tolerated V, L, I, W, F, or Y at every position in the original NR box motif containing an L. The regular expression in the initial search was:

([^P][VIWFY][^P][^P][VLIWFY][VLIWFY][^P])|([^P][VLIWFY][^P][^P][VIWFY][VLIWFY][^P])|([^P][VLIWFY][^P][^P][VLIWFY][VIWFY][^P])

Logos of the motif instances in the no effect/enhancing regions versus reducing/breaking regions in activation and repression domains were generated using the ‘logomaker’ Python package. From these, we determined that position 1 of the motif rarely contained Y, position 4 rarely contained W, and the position 4 rarely contained W or Y, resulting in the final pattern:

([^P][VIWF][^P][^P][VLIFY][VLIF][^P])|([^P][VLIWF][^P][^P][VIFY][VLIF][^P])|([^P][VLIWF][^P][^P] [VLIFY][VIF][^P])

Sequence similarity analysis of EBV B95-8 and AG876 proteins—The Biopython Entrez package was used to fetch the genome and protein sequences for the EBV type 1 B95-8 strain (NCBI Reference Sequence accession number NC_007605.1) and the type 2 AG876 strain (GenBank accession number DQ279927.1). A pairwise protein sequence alignment between the B95-8 and AG876 homologs was performed for each protein, and the number of mismatches was divided by the length of the shorter homolog to calculate a mismatch per residue rate, which is plotted in Fig. 4B.

CRISPR/Cas9 targeting of the KSHV DBP gene (ORF6)—CRISPR/Cas9 sgRNAs were introduced to cells and their effects on late gene expression and replication were measured as previously described in58. Briefly, sgRNAs targeting along the ORF6 gene were cloned into a mU6-driven guide expression plasmid (Addgene #89359) and delivered

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

via lentiviral transduction at high MOI to the Cas9+ iSLK-BAC16-K8.1pr-mIFP2 cells. To analyze late gene expression, cells were were treated with 5 μg/mL doxycycline, which induces expression of KSHV RTA (ORF50) for lytic reactivation, and 1 mM sodium butyrate, an HDAC inhibitor that facilitates reactivation. Forty-eight hours after reactivation, cells were fixed in 4% PFA and quantified using flow cytometry (BD LSRFortessa). To analyze viral DNA replication, cells were similarly reactivated, and 48 hours later were treated with 30 μM 5-ethynyl-2'-deoxyuridine (EdU, Invitrogen #A10044) for two hours. Cells were then trypsinized and fixed using 4% PFA. EdU was labeled with Alexa Fluor 647 using the Click-iT EdU Alexa Fluor 647 Flow Cytometry Assay Kit (Invitrogen #C10424) and quantified by flow cytometry (BD Accuri C6 Plus). All experiments were performed in four replicates from different days and independent reactions.

To monitor ORF6 mRNA levels, cells were reactivated, and 24 hours post-reactivation, RNA was harvested using QuickExtract RNA Extraction Solution (Biosearch Technologies #QER090150). TURBO DNase (Invitrogen #AM2238) was used to remove DNA. AMV Reverse Transcriptase (Promega #M5101) with 9 bp random primers was used for reverse transcription. qPCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad #1725120) amplifying the coding region of ORF6 along with primers targeting the host 18S RNA (ORF6_fwd: GCTTGGACAAAGGAGCAATC; ORF6_rev: GCTCTGGCTATCCTGACCTG; 18S_fwd:CCTGCGGCTTAATTTGACTC; 18S_rev: ATGCCAGAGTCTCGTTCGTT). To monitor editing at the ORF6 locus, genomic DNA was extracted from non-reactivated cells using QuickExtract DNA Extraction Solution (Biosearch #QE09050). The ORF6 locus was amplified using GoTaq DNA Polymerase (Promega #M3001) and Sanger sequenced. Editing outcomes were then determined with the Synthego Inference of CRISPR Edits (ICE) Analysis tool (https://www.synthego.com/help/synthego-ice-analysis).

Quantification and Statistical Analysis—Statistical analyses were performed in Python with the SciPy package[68], are two-sided (where applicable), and are indicated in the text and/or figure legends. The "n" for each analysis is indicated in the main text or in figure legends of relevant analyses. Significance was set at $p < 0.05$, or $q < 0.05$ when correcting for multiple comparisons. No methods were used to determine whether the data met assumptions of the statistical approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Michaela M. Hinks and Benjamin R. Doughty for sequencing assistance, Adi Mukund for data visualization advice, and members of the Bintu lab for helpful conversations and assistance. We thank Twist Biosciences for oligonucleotide library synthesis. This work was supported by NIH-NIGMS MIRA R35GM12894701 (LB), NIH-NHGRI R01HG011866 (LB, MCB), NIH T32GM145402 (ART), NIH-NIAID R01AI122528 (BAG), Sarafan Chem-H Chemistry-Biology Interface Training Grant (ART), Howard Hughes Medical Institute and Life Sciences Research Foundation Postdoctoral Fellowship (DWM), and NIH-4K00DK126120-03 (JT). BAG is an investigator of the Howard Hughes Medical Institute.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. Fields BN (1983). How Do Viruses Cause Different Diseases? JAMA 250, 1754–1756. 10.1001/jama.1983.03340130072038. [PubMed: 6310171]
2. Brito AF, and Pinney JW (2017). Protein–Protein Interactions in Virus–Host Systems. Front. Microbiol 8.
3. Latchman DS (1993). Transcriptional regulation of viral gene expression. Rev. Med. Virol 3, 115–122. 10.1002/rmv.1980030208.
4. Nečasová I, Stojaspal M, Motyčáková E, Brom T, Janovič T, and Hofr C (2022). Transcriptional regulators of human oncoviruses: structural and functional implications for anticancer therapy. NAR Cancer 4, zcac005. 10.1093/narcan/zcac005. [PubMed: 35252867]
5. DiMaio D (2012). Viruses, Masters at Downsizing. Cell Host Microbe 11, 560–561. 10.1016/j.chom.2012.05.004. [PubMed: 22704616]
6. Sanjuán R, and Domingo-Calap P (2021). Genetic Diversity and Evolution of Viral Populations. Encycl. Virol, 53–61. 10.1016/B978-0-12-809633-8.20958-8.
7. Liu X, Hong T, Parameswaran S, Ernst K, Marazzi I, Weirauch MT, and Fuxman Bass JI (2020). Human Virus Transcriptional Regulators. Cell 182, 24–37. 10.1016/j.cell.2020.06.023. [PubMed: 32649876]
8. Tycko J, DelRosso N, Hess GT, Aradhana Banerjee, A., Mukund A, Van MV, Ego BK, Yao D, Spees K, et al. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. Cell 183, 2020–2035.e16. 10.1016/j.cell.2020.11.024. [PubMed: 33326746]
9. Bagowski CP, Bruins W, and te Velthuis AJW (2010). The Nature of Protein Domain Evolution: Shaping the Interaction Network. Curr. Genomics 11, 368–376. 10.2174/138920210791616725. [PubMed: 21286315]
10. Fan D, Wang M, Cheng A, Jia R, Yang Q, Wu Y, Zhu D, Zhao X, Chen S, Liu M, et al. (2020). The Role of VP16 in the Life Cycle of Alphaherpesviruses. Front. Microbiol 11.
11. Moriuchi H, Moriuchi M, Pichyangkura R, Triezenberg SJ, Straus SE, and Cohen JI (1995). Hydrophobic cluster analysis predicts an amino-terminal domain of varicella-zoster virus open reading frame 10 required for transcriptional activation. Proc. Natl. Acad. Sci. 92, 9333–9337. 10.1073/pnas.92.20.9333. [PubMed: 7568128]
12. Glavina J, Román EA, Espada R, de Prat-Gay G, Chemes LB, and Sánchez IE (2018). Interplay between sequence, structure and linear motifs in the adenovirus E1A hub protein. Virology 525, 117–131. 10.1016/j.virol.2018.08.012. [PubMed: 30265888]
13. Avvakumov N, Kajon AE, Hoeben RC, and Mymryk JS (2004). Comprehensive sequence analysis of the E1A proteins of human and simian adenoviruses. Virology 329, 477–492. 10.1016/j.virol.2004.08.007. [PubMed: 15518825]
14. Zemke NR, and Berk AJ (2017). The Adenovirus E1A C Terminus Suppresses a Delayed Antiviral Response and Modulates RAS Signaling. Cell Host Microbe 22, 789–800.e5. 10.1016/j.chom.2017.11.008. [PubMed: 29241042]
15. Cohen MJ, Yousef AF, Massimi P, Fonseca GJ, Todorovic B, Pelka P, Turnell AS, Banks L, and Mymryk JS (2013). Dissection of the C-Terminal Region of E1A Redefines the Roles of CtBP and Other Cellular Targets in Oncogenic Transformation. J. Virol. 87, 10348–10355. 10.1128/JVI.00786-13. [PubMed: 23864635]
16. Whitley RJ (1996). Herpesviruses. In Medical Microbiology, Baron S, ed. (University of Texas Medical Branch at Galveston).
17. Pellet PE, and Roizman B (2007). Fields Virology Knipe DM and Howley, eds.
18. Renner DW, and Szpara ML (2018). Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. J. Virol 92, e00908–17. 10.1128/JVI.00908-17. [PubMed: 29046445]
19. Nicholas J (1996). Determination and analysis of the complete nucleotide sequence of human herpesvirus. J. Virol 70, 5975–5989. 10.1128/jvi.70.9.5975-5989.1996. [PubMed: 8709220]
20. Szilágyi A, and Skolnick J (2006). Efficient prediction of nucleic acid binding function from low-resolution protein structures. J. Mol. Biol 358, 922–933. 10.1016/j.jmb.2006.02.053. [PubMed: 16551468]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

21. Faust TB, Binning JM, Gross JD, and Frankel AD (2017). Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. Annu. Rev. Virol 4, 241–260. 10.1146/annurev-virology-101416-041654.
22. Arnold CD, Nemčko F, Woodfin AR, Wienerroither S, Vlasova A, Schleiffer A, Pagani M, Rath M, and Stark A (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. EMBO J 37, e98896. 10.15252/embj.201798896. [PubMed: 30006452]
23. Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, and Cohen BA (2018). A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. Cell Syst 6, 444–455.e6. 10.1016/j.cels.2018.01.015. [PubMed: 29525204]
24. Sanborn AL, Yeh BT, Feigerle JT, Hao CV, Townshend RJ, Lieberman Aiden E, Dror RO, and Kornberg RD (2021). Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. eLife 10, e68068. 10.7554/eLife.68068. [PubMed: 33904398]
25. DelRosso N, Tycko J, Suzuki P, Andrews C, Aradhana Mukund, A., Liongson I, Ludwig C, Spees K, Fordyce P, et al. (2022). Large-scale mapping and systematic mutagenesis of human transcriptional effector domains 2022.08.26.505496. 10.1101/2022.08.26.505496.
26. Drozdetskiy A, Cole C, Procter J, and Barton GJ (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43, W389–W394. 10.1093/nar/gkv332. [PubMed: 25883141]
27. Alerasool N, Leng H, Lin Z-Y, Gingras A-C, and Taipale M (2022). Identification and functional characterization of transcriptional activators in human cells. Mol. Cell 82, 677–695.e7. 10.1016/j.molcel.2021.12.008. [PubMed: 35016035]
28. Klaus L, Almeida B.P. de, Vlasova A, Nemčko F, Schleiffer A, Bergauer K, Rath M, and Stark A (2022). Identification and characterization of repressive domains in Drosophila transcription factors 2022.08.26.505062. 10.1101/2022.08.26.505062.
29. Verger A, Perdomo J, and Crossley M (2003). Modification with SUMO. EMBO Rep 4, 137–142. 10.1038/sj.embor.embor738. [PubMed: 12612601]
30. Cho G, Lim Y, and Golden JA (2009). SUMO Interaction Motifs in Sizn1 Are Required for Promyelocytic Leukemia Protein Nuclear Body Localization and for Transcriptional Activation. J. Biol. Chem 284, 19592–19600. 10.1074/jbc.M109.010181. [PubMed: 19416967]
31. Wang C-H, Hung P-W, Chiang C-W, Lombès M, Chen C-H, Lee K-H, Lo Y-C, Wu M-H, Chang W-C, and Lin D-Y (2019). Identification of two independent SUMO-interacting motifs in Fas-associated factor 1 (FAF1): Implications for mineralocorticoid receptor (MR)-mediated transcriptional regulation. Biochim. Biophys. Acta BBA - Mol. Cell Res 1866, 1282–1297. 10.1016/j.bbamcr.2019.03.014.
32. Plevin MJ, Mills MM, and Ikura M (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem. Sci 30, 66–69. 10.1016/j.tibs.2004.12.001. [PubMed: 15691650]
33. Huang N, vom Baur E, Garnier J-M, Lerouge T, Vonesch J-L, Lutz Y, Chambon P, and Losson R (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J 17, 3398–3412. 10.1093/emboj/17.12.3398. [PubMed: 9628876]
34. Li D, Desai-Yajnik V, Lo E, Schapira M, Abagyan R, and Samuels HH (1999). NRIF3 Is a Novel Coactivator Mediating Functional Specificity of Nuclear Hormone Receptors. Mol. Cell. Biol 19, 7191–7202. [PubMed: 10490654]
35. Hu X, and Lazar MA (1999). The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. Nature 402, 93–96. 10.1038/47069. [PubMed: 10573424]
36. Uttarkar S, Frampton J, and Klempnauer K-H (2017). Targeting the transcription factor Myb by small-molecule inhibitors. Exp. Hematol 47, 31–35. 10.1016/j.exphem.2016.12.003. [PubMed: 28017646]
37. Hammitzsch A, Tallant C, Fedorov O, O'Mahony A, Brennan PE, Hay DA, Martinez FO, Al-Mossawi MH, de Wit J, Vecellio M, et al. (2015). CBP30, a selective CBP/p300 bromodomain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inhibitor, suppresses human Th17 responses. Proc. Natl. Acad. Sci 112, 10768–10773. 10.1073/pnas.1501956112. [PubMed: 26261308]

38. Bratkowski M, Yang X, and Liu X (2018). An Evolutionarily Conserved Structural Platform for PRC2 Inhibition by a Class of Ezh2 Inhibitors. Sci. Rep 8, 9092. 10.1038/s41598-018-27175-w. [PubMed: 29904056]
39. Choi SY, Kee HJ, Jin L, Ryu Y, Sun S, Kim GR, and Jeong MH (2018). Inhibition of class IIa histone deacetylase activity by gallic acid, sulforaphane, TMP269, and panobinostat. Biomed. Pharmacother 101, 145–154. 10.1016/j.biopha.2018.02.071. [PubMed: 29482060]
40. Chabot PR, Raiola L, Lussier-Price M, Morse T, Arseneault G, Archambault J, and Omichinski JG (2014). Structural and Functional Characterization of a Complex between the Acidic Transactivation Domain of EBNA2 and the Tfb1/p62 Subunit of TFIIH. PLOS Pathog 10, e1004042. 10.1371/journal.ppat.1004042. [PubMed: 24675874]
41. Gwack Y, Byun H, Hwang S, Lim C, and Choe J (2001). CREB-Binding Protein and Histone Deacetylase Regulate the Transcriptional Activity of Kaposi's Sarcoma-Associated Herpesvirus Open Reading Frame 50. J. Virol 75, 1909–1917. 10.1128/JVI.75.4.1909-1917.2001. [PubMed: 11160690]
42. Rickinson AB, Young LS, and Rowe M (1987). Influence of the Epstein-Barr virus nuclear antigen EBNA 2 on the growth phenotype of virus-transformed B cells. J. Virol 61, 1310–1317. 10.1128/JVI.61.5.1310-1317.1987. [PubMed: 3033261]
43. Cancian L, Bosshard R, Lucchesi W, Karstegl CE, and Farrell PJ (2011). C-Terminal Region of EBNA-2 Determines the Superior Transforming Ability of Type 1 Epstein-Barr Virus by Enhanced Gene Regulation of LMP-1 and CXCR7. PLoS Pathog 7, e1002164. 10.1371/journal.ppat.1002164. [PubMed: 21857817]
44. Farrell PJ (2015). Epstein–Barr Virus Strain Variation. In Epstein Barr Virus Volume 1: One Herpes Virus: Many Diseases Current Topics in Microbiology and Immunology, Münz C, ed. (Springer International Publishing), pp. 45–69. 10.1007/978-3-319-22822-8_4.
45. Bhattacharjee S, Ghosh Roy S, Bose P, and Saha A (2016). Role of EBNA-3 Family Proteins in EBV Associated B-cell Lymphomagenesis. Front. Microbiol 7.
46. Cohen JI (1992). A region of herpes simplex virus VP16 can substitute for a transforming domain of Epstein-Barr virus nuclear protein 2. Proc. Natl. Acad. Sci. U. S. A 89, 8030–8034. 10.1073/pnas.89.17.8030. [PubMed: 1325641]
47. Yang Y, Ye X, Dai R, Li Z, Zhang Y, Xue W, Zhu Y, Feng D, Qin L, Wang X, et al. (2021). Phase separation of Epstein-Barr virus EBNA2 protein reorganizes chromatin topology for epigenetic regulation. Commun. Biol 4, 1–14. 10.1038/s42003-021-02501-7. [PubMed: 33398033]
48. Peng R, Gordadze AV, Fuentes Pananá EM, Wang F, Zong J, Hayward GS, Tan J, and Ling PD (2000). Sequence and Functional Analysis of EBNA-LP and EBNA2 Proteins from Nonhuman Primate Lymphocryptoviruses. J. Virol 74, 379–389. 10.1128/JVI.74.1.379-389.2000. [PubMed: 10590127]
49. Cohen JI, Wang F, and Kieff E (1991). Epstein-Barr virus nuclear protein 2 mutations define essential domains for transformation and transactivation 10.1128/jvi.65.5.2545-2554.1991.
50. Gillian AL, and Svaren J (2004). The Ddx20/DP103 dead box protein represses transcriptional activation by Egr2/Krox-20. J. Biol. Chem 279, 9056–9063. 10.1074/jbc.M309308200. [PubMed: 14699164]
51. Beer S, Wange LE, Zhang X, Kuklik-Roos C, Enard W, Hammerschmidt W, Scialdone A, and Kempkes B (2022). EBNA2-EBF1 complexes promote MYC expression and metabolic processes driving S-phase progression of Epstein-Barr virus–infected B cells. Proc. Natl. Acad. Sci 119, e2200512119. 10.1073/pnas.2200512119. [PubMed: 35857872]
52. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. 10.1186/s13059-014-0550-8. [PubMed: 25516281]
53. Offermann MK (2007). Kaposi sarcoma herpesvirus-encoded interferon regulator factors. Curr. Top. Microbiol. Immunol 312, 185–209. 10.1007/978-3-540-34344-8_7. [PubMed: 17089798]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

54. Hwang S-W, Kim D, Jung JU, and Lee H-R (2017). KSHV-encoded viral interferon regulatory factor 4 (vIRF4) interacts with IRF7 and inhibits interferon alpha production. Biochem. Biophys. Res. Commun 486, 700–705. 10.1016/j.bbrc.2017.03.101. [PubMed: 28342865]
55. Girdwood D, Bumpass D, Vaughan OA, Thain A, Anderson LA, Snowden AW, Garcia-Wilson E, Perkins ND, and Hay RT (2003). p300 Transcriptional Repression Is Mediated by SUMO Modification. Mol. Cell 11, 1043–1054. 10.1016/S1097-2765(03)00141-2. [PubMed: 12718889]
56. Gruffat H, Marchione R, and Manet E (2016). Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. Front. Microbiol 7, 869. 10.3389/fmicb.2016.00869. [PubMed: 27375590]
57. Gao M, and Knipe DM (1991). Potential role for herpes simplex virus ICP8 DNA replication protein in stimulation of late gene expression. J. Virol 65, 2666–2675. 10.1128/jvi.65.5.2666-2675.1991. [PubMed: 1850040]
58. Morgens DW, Nandakumar D, Didychuk AL, Yang KJ, and Glaunsinger BA (2022). A Two-tiered functional screen identifies herpesviral transcriptional modifiers and their essential domains. PLOS Pathog 18, e1010236. 10.1371/journal.ppat.1010236. [PubMed: 35041709]
59. Belouzard S, Millet JK, Licitra BN, and Whittaker GR (2012). Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. Viruses 4, 1011–1033. 10.3390/v4061011. [PubMed: 22816037]
60. Bollavaram K, Leeman TH, Lee MW, Kulkarni A, Upshaw SG, Yang J, Song H, and Platt MO (2021). Multiple sites on SARS-CoV-2 spike protein are susceptible to proteolysis by cathepsins B, K, L, S, and V. Protein Sci. Publ. Protein Soc 30, 1131–1143. 10.1002/pro.4073.
61. Zhao M-M, Yang W-L, Yang F-Y, Zhang L, Huang W-J, Hou W, Fan C-F, Jin R-H, Feng Y-M, Wang Y-C, et al. (2021). Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. Signal Transduct. Target. Ther 6, 1–12. 10.1038/s41392-021-00558-8. [PubMed: 33384407]
62. Sattar S, Kabat J, Jerome K, Feldmann F, Bailey K, and Mehedi M (2022). Nuclear translocation of spike mRNA and protein is a novel pathogenic feature of SARS-CoV-2 (Microbiology) 10.1101/2022.09.27.509633.
63. Readhead B, Haure-Mirande J-V, Funk CC, Richards MA, Shannon P, Haroutunian V, Sano M, Liang WS, Beckmann ND, Price ND, et al. (2018). Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. Neuron 99, 64–82.e7. 10.1016/j.neuron.2018.05.023. [PubMed: 29937276]
64. Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, Magnusen AF, Lynch A, Chetal K, Yukawa M, et al. (2018). Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat. Genet 50, 699–707. 10.1038/s41588-018-0102-3. [PubMed: 29662164]
65. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science 375, 296–301. 10.1126/science.abj8222. [PubMed: 35025605]
66. Brulois KF, Chang H, Lee AS-Y, Ensser A, Wong L-Y, Toth Z, Lee SH, Lee H-R, Myoung J, Ganem D, et al. (2012). Construction and Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial Artificial Chromosome Clone. J. Virol 86, 9708–9720. 10.1128/JVI.01019-12. [PubMed: 22740391]
67. Teague B (2022). Cytoflow: A Python Toolbox for Flow Cytometry 2022.07.22.501078. 10.1101/2022.07.22.501078.
68. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. 10.1038/s41592-019-0686-2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- We annotate activator and repressor domains and mutants across ~1,500 viral proteins
- 1/5 of herpesvirus proteins have effector domains, including VIRFs and replication DBPs
- Systematic mutagenesis uncovers motifs essential for function similar to human effectors
- Herpesvirus effector activity often depends on key tryptophans and flexiNR box motifs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

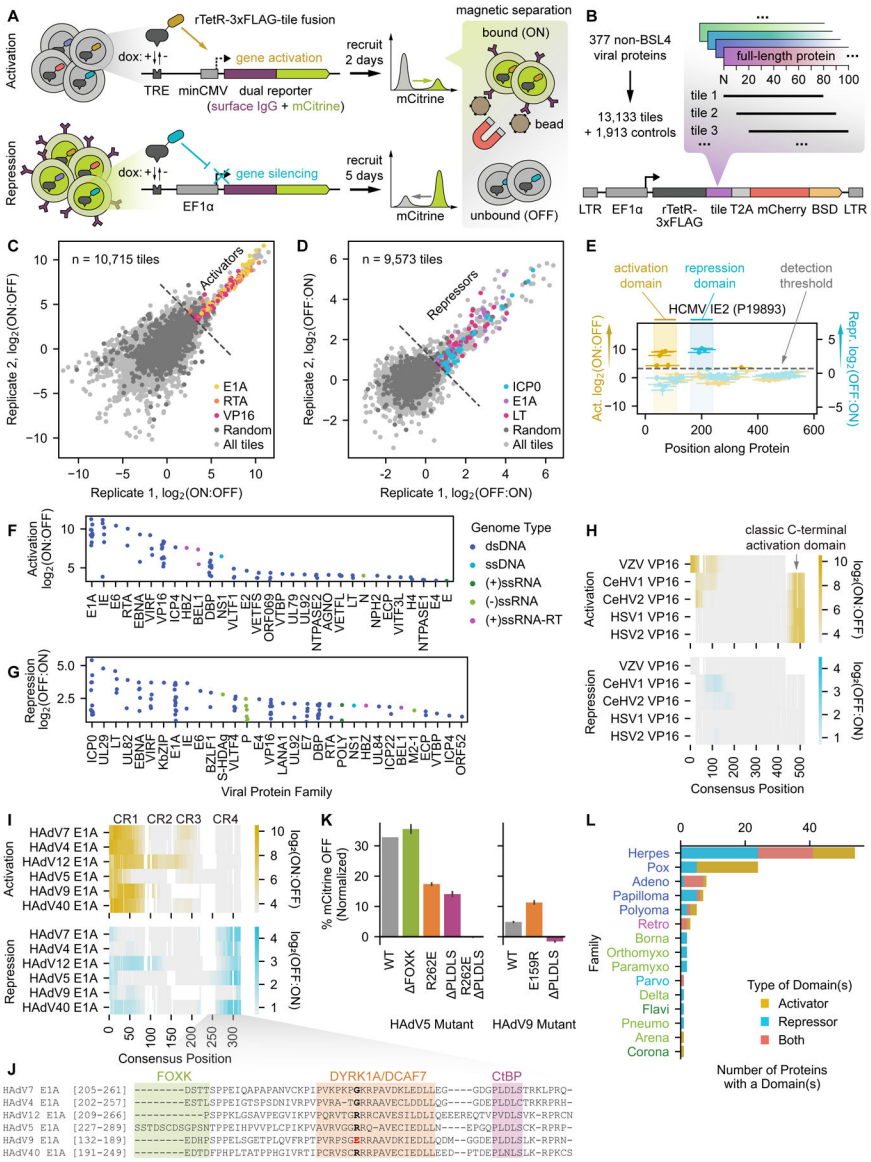


Fig. 1 |. HT-recruit recovers hundreds of protein domains with transcriptional effector activity among a set of known or putative viral transcriptional regulators.

(**A**) Schematic of the high-throughput recruitment (HT-recruit) approach. Library members are synthesized, cloned as fusions to the doxycycline (dox)-inducible rTetR DNA-binding domain, and delivered to cells harboring a dual reporter gene encoding both mCitrine and a surface marker that enables magnetic sorting of cells by reporter transcriptional state. Library member frequencies in the bead-bound (ON) and unbound (OFF) populations are determined by next generation sequencing to compute enrichment scores. Pooled screens are performed in cells whose reporter is under the control of a weak minimal cytomegalovirus (minCMV) promoter to measure transcriptional activation (top) or a strong EF1a promoter to measure repression (bottom). TRE = tetracycline response element (nine copies of the TetO sequence that can be bound by rTetR). (**B**) Composition of the viral transcriptional regulator (vTR) library, which includes 80aa-long tiles sampled every 10aa for 377 proteins with known or putative transcriptional regulatory potential.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

BSD = blasticidin resistance gene. (**C**) Reproducibility of activation enrichment scores, log2(ON:OFF), across two replicates, with hit tiles from the well-described activators E1A, RTA, and VP16 indicated. For all analyses, the detection thresholds (dashed lines in C-G) are set as two standard deviations above the mean of the random negative controls. (**D**) Reproducibility of repression enrichment scores, log2(OFF:ON), across two replicates, with hit tiles from the repressors E1A, ICP0, and LT indicated. (**E**) Calling activation and repression domains from tiling measurements, using the HCMV IE2 protein as an example. The dashed line represents the detection threshold, and higher scores above it correspond to stronger activation (yellow) or repression (blue). Vertical spans indicate the maximum strength tile within each domain. (**F-G**) Summary of identified activation (F) and repression (G) domains, represented by their strongest tile, stratified by viral protein family, and colored by the genome type of the virus that encodes them. ds = double-stranded, ss = single-stranded, (+) = positive-sense, (-) = negative-sense, and RT = reverse-transcribed. (**H**) Multiple sequence alignment (MSA) of five herpesvirus VP16 homologs, with activation log2(ON:OFF) and repression log2(OFF:ON) enrichment scores represented as yellow and blue color mappings, respectively. (**I**) MSA of six human adenovirus (HAdV) E1A homologs with their conserved regions (CRs) indicated. Color mappings reflect enrichment scores as in (H). (**J**) Zoomed alignment of CR4 showing known cofactor binding regions for HAdV5 E1A. A critical residue within the DYRK1A/DCAF7-binding region is bolded, with the HAdV9 mutant highlighted in red. (**K**) Quantification of the fraction of cells OFF by flow cytometry (normalized to no dox) after 5 days of recruitment of wild-type (WT) and mutant CR4 sequences from HAdV5 and HAdV9 E1A. (**L**) Summary of vTR proteins with at least one effector domain, stratified by viral family and colored by effector type. Viral family names are colored by genome type (see legend of F&G).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

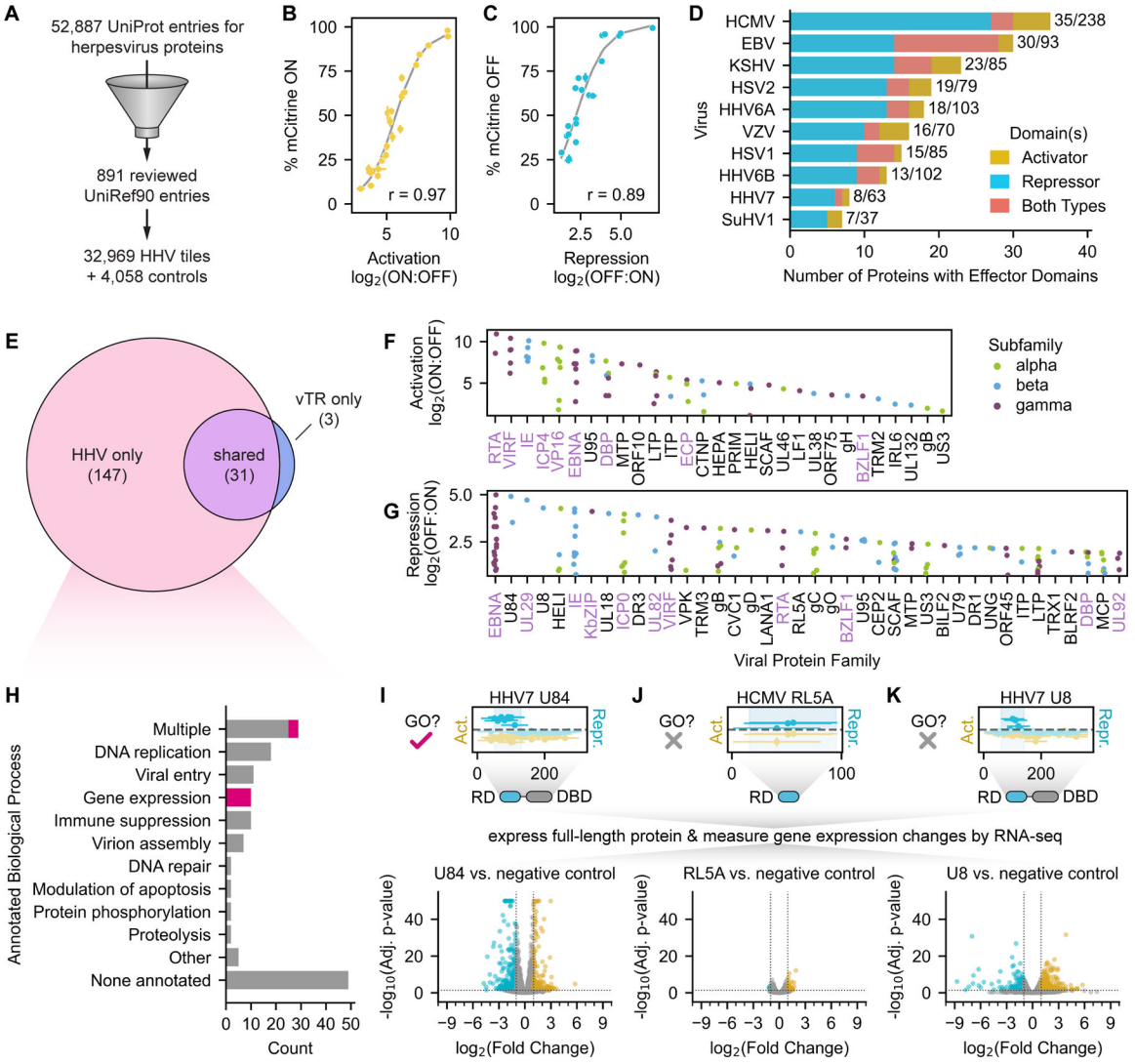


Fig. 2 |.

Unbiased identification of activator and repressor domains from herpesviruses. (**A**) Herpesvirus tiling library design: we compiled all 891 herpesvirus protein sequences listed in UniRef90, which collapses proteins on 90% sequence identity to limit redundancy and represent related sequences with a single high-confidence reviewed sequence, and generated a library of 32,969 tiles supplemented with 4,058 controls. (**B**) Relationship between activation log2(ON:OFF) enrichment score from the screen and the percent of mCitrine-positive (ON) cells as measured by flow cytometry after two days of recruitment for a set of individually validated tiles. Logistic fit in gray with Spearman r = 0.97. (**C**) Relationship between repression log2(OFF:ON) enrichment score from the screen and the percent of mCitrine-negative (OFF) cells as measured by flow cytometry after five days of recruitment for a set of individually validated tiles. Logistic fit in gray with Spearman r = 0.89. (**D**) Summary of herpesvirus proteins with at least one effector domain, stratified by viral species and colored by the effector domain type(s). Proteins with two or more domains of different functions or a single dual effector domain are categorized as ‘Both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Types'. Bar label numerators indicate the total number of proteins with domains, and bar label denominators indicate the total number of proteins tiled for each virus (effective proteome size). (**E**) Venn diagram of all herpesvirus proteins shared between the vTR and HHV libraries in which we identified effector domains. Number of proteins per group indicated in parentheses. (**F-G**) Summary of identified activation (D) and repression (E) domains, represented by their strongest tile and stratified by viral protein family. Only the top 40 families are shown for repressors due to space constraints. Protein families labeled in lavender have at least one protein homolog for which we also measured transcriptional effector activity in the vTR screen (**H**) Summary of the biological processes associated with the 147 effector proteins uniquely identified in the HHV tiling screen. Biological process gene ontology terms associated with each protein were pulled from UniProt and assigned to high-level categories for this analysis. Proteins with gene expression-related gene ontology terms are colored in magenta. (**I-K**) Top: tiling plots from the HHV tiling screen, with a schematic showing the repression domain (RD) and the presence or absence of a predicted DNA-binding domain (DBD) for HHV7 U84 (I), HCMV RL5A (J), and HHV7 U8 (K). U84 is annotated with a gene expression-related GO term, while RL5A and U8 are not. Bottom: volcano plots from RNA-seq with expression of the full-length proteins show significantly upregulated (yellow) and downregulated (blue) genes 48 hours after induction of viral protein expression compared to a negative control expressing mCitrine instead of the viral protein.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

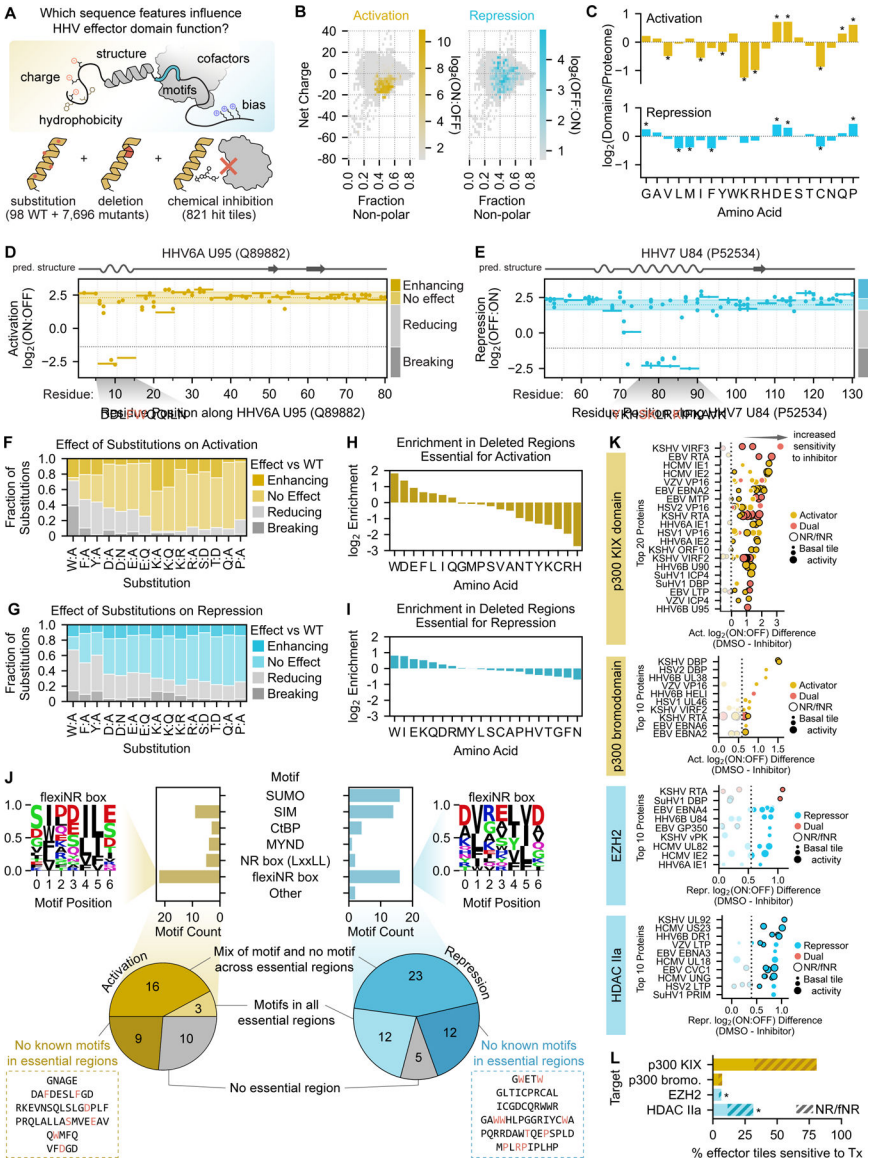


Fig. 3 |. Sequence analysis and systematic perturbation of herpesvirus transcriptional effector sequences.

(**A**) Overview of the sequence features examined and the perturbations performed to connect herpesvirus (HHV) effector domain sequences to their functions. (**B**) Two-dimensional histograms of net charge versus the fraction of non-polar residues for all tiles in the HHV tiling screen. Bins are colored with their maximum activation (yellow, left) or repression (blue, right) screen score as indicated by the color bars. (**C**) Barplots of the log2-transformed ratios of amino acid frequencies in activation (top) or repression (domains) relative to their proteome frequencies. Positive values represent an enrichment in effector domains while negative values represent a depletion. Significant differences in amino acid frequencies were determined by the Welch's T test (Bonferroni-corrected) and are indicated as stars. (**D, E**) Perturbation tiling plots mapping the effects of single-residue substitutions (dots) and 5aa deletions (horizontal spans) on the maximum-strength tiles from the activation domain of HHV6A U95 (D) and repression domain of HHV7 U84 (E). JPred4-predicted secondary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

structures are shown above the plots, with alpha helices as squiggles, beta sheets as arrows, and other (including unstructured) as a straight line. The shaded horizontal span represents the wild-type screen score mean plus/minus two times the estimated error (mean of all wild-type tiles shown as the yellow horizontal dotted line within) for U95 (D) and U84 (E). Perturbations with scores within these regions are considered to have 'no effect', while those above and below are considered 'enhancing' and 'reducing', respectively. The gray horizontal dotted lines represent the detection thresholds, and thus perturbations whose scores are below this threshold are considered 'breaking'. Deleted regions below the detection threshold are deemed essential, and their sequences are displayed below the plot, with red residues indicating single-residue substitutions that abolish activity. (**F-G**) Effect of single-residue substitutions on activation (F) or repression (G) as measured in the perturbation screen. (**H-I**) Barplots of the log2-transformed ratios of amino acid frequencies in regions whose activity is essential to activation (H) or repression (I) relative to their proteome frequencies. (**J**) Top: counts of motifs that are enriched in essential regions. Logo of the newly proposed flexiNR box motif from all essential regions in activators (top left) or repressors (top right). Other motifs follow ELM definitions (Methods). Also shown are examples of essential sequences with no known overlapping motif (inside dashed boxes), with the residues most sensitive/critical to activity as determined by single-residue substitution in red. (**K**) Summary of the top 20 herpesvirus proteins with tiles sensitive to p300/CBP KIX domain inhibition with celastrol (top) or the top 10 herpesvirus proteins with tiles sensitive to p300/CBP bromodomain inhibition with SGC-CBP30 (upper middle), EZH2 inhibition with tazemetostat (lower middle), and class IIa HDAC inhibition with TMP269 (bottom). Each dot is a tile from the viral protein indicated on the y-axis and is colored based on its effector activity. Dot size indicates the strength of the tile's transcriptional effect in the DMSO control screen, and a black outline indicates the presence of at least one NR or flexiNR box motif (NR/fNR) in the tile. The x-axis shows the difference between screen scores in the DMSO control screen versus in the screen with inhibitor, with increasing positive values indicating increased sensitivity to the inhibitor (i.e. greater impairment of activation or repression with treatment). The dashed lines in the p300/CBP inhibition screens represent the sensitivity threshold set at the mean plus two standard deviations of the repressor scores (expected to have no activity), and vice versa for the EZH2 and class IIa HDAC inhibition screens. (**L**) Summary of the percent of activator tiles (yellow) or repressor tiles (blue) whose activities are significantly reduced upon treatment with the inhibitors in (K). The hashed portion indicates the fraction of sensitive tiles that contain an NR box or flexiNR box motif (NR/fNR). Stars indicate significant depletion of these motifs in tiles sensitive to EZH2 inhibition (OR: 0.31, 95% CI: 0.13–0.73) or significant enrichment in tiles sensitive to class IIa HDAC inhibition (OR: 1.58, 95% CI: 1.03–2.42).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

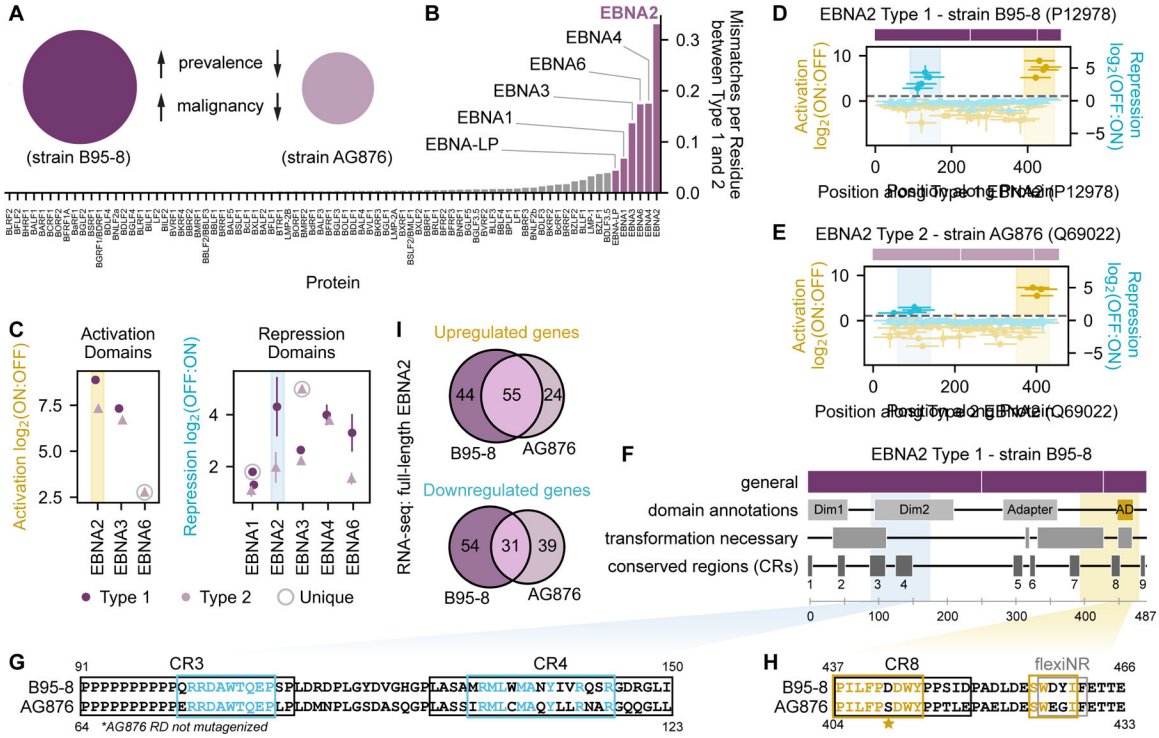


Fig. 4 |. Sequence and functional comparison of EBNA family effector domains.

(**A**) EBV type 1 and type 2 are associated with different clinicopathological features. (**B**) Sequence dissimilarity between type 1 (B95-8 strain) and type 2 (AG876 strain) homologs of each EBV protein, quantified as the number of mismatches per residue. EBNA family proteins (highlighted) exhibit the greatest sequence dissimilarity between EBV type 1 and type 2. (**C**) Summary of activation and repression domain strengths across type 1 (dark circles) and type 2 (light triangles) homologs of EBNA proteins. The activation and repression domains of EBNA2 are highlighted with yellow and blue spans, respectively. 'Unique' refers to effector domains detected in only the type 1 or type 2 homolog. (**D-E**) Tiling plots of type 1 (D) and type 2 (E) EBNA2, with simple schematics showing rough domain boundaries (see (F) for details). (**F**) Schematic of type 1 EBNA2 with relevant annotations. Colored spans indicate the maximum strength tile coordinates for repression (blue) and activation (yellow). CBF1-BD = CBF1 binding domain (for indirect association with DNA); Dim = dimerization domain; transform. ess. = essential for cellular transformation; AD = activation domain. (**G-H**). Pairwise alignments of EBNA2 sequences from the B95-8 (Type 1) and AG876 (Type 2) strains. Consensus essential regions as determined from deletion scanning are indicated for the repression domain in (G) with blue boxes or for the activation domain in (H) with yellow boxes, where colored letters indicate residues within these regions that are identical between strains. The repression domain (RD) of the AG876 EBNA2 homolog was not mutagenized. Star represents the S to D mutation that restores AG876 EBNA2 activity to the B95-8 level. CR = conserved region. (**I**) Venn diagrams showing the number of unique and shared significantly upregulated (left) and downregulated (right) genes as measured by RNA-seq in cells expressing either full-length B95-8 or AG876 EBNA2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

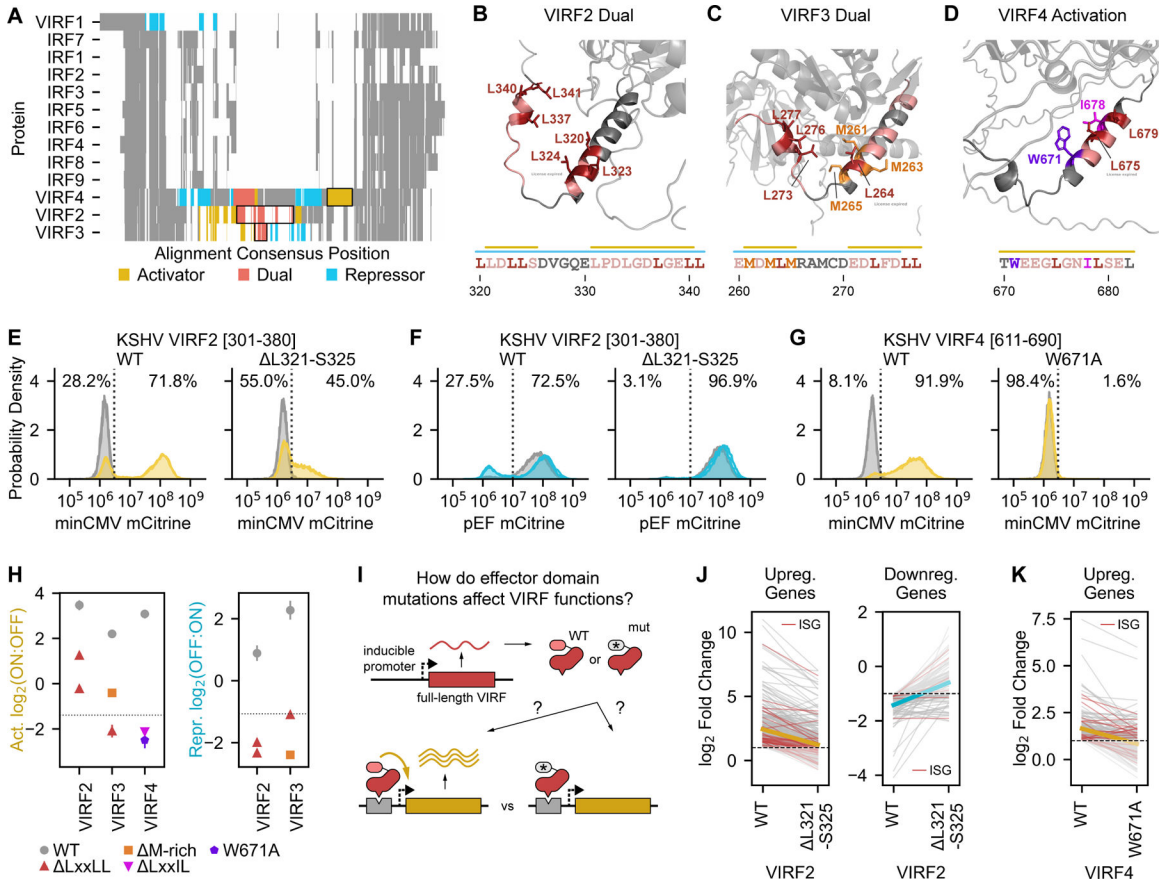


Fig. 5 |. Investigating the importance of cofactor interaction motifs on VIRF protein functions.

(**A**) Multiple sequence alignment of nine human interferon-regulatory factors (IRFs) and their viral homologs (VIRFs) from KSHV, showing homology across all homologs in the N-terminal DNA-binding domain and in the C-terminal region and high sequence divergence in the middle, where the VIRF2, VIRF3, and VIRF4 effector domains are located. Effector domains highlighted in subsequent panels are outlined with black boxes. (**B-D**) AlphaFold2-predicted structures highlighting portions of the dual effector domain of VIRF2 (B), the dual effector domain of VIRF3 (C), and the strong activation domain of VIRF4 (D). Residues in essential regions are colored, with key residues in the NR box (LxxLL), flexiNR box, and methionine-rich motifs in specific colors: leucines in red, methionines in orange, and isoleucine in magenta. The critical tryptophan in VIRF4 (D) is in purple. Bars above the sequences indicate regions that reduce or break activation (yellow bars) or repression (blue bars) when deleted. (**E-F**) Flow cytometry distributions of cells after recruitment for two days to a minCMV-mCitrine reporter (E) or for five days to a pEF-mCitrine reporter (F) of a wild-type or mutant (L321-S325 deletion) tile from the dual effector domain of VIRF2. The percentages of the dox-treated cells (yellow or blue) that are mCitrine-negative (OFF) and mCitrine-positive (ON) are indicated to the left and right, respectively, of the dashed line. Distributions for untreated cells are in gray. (**G**) Flow cytometry distributions of cells after recruitment for two days to a minCMV-mCitrine reporter of a wild-type or mutant (W671A) tile from the strong activation domain of VIRF4. (**H**) Summary of effector domain activity for WT VIRFs (gray) and mutants with deletions overlapping the LxxLL motif, the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

M-rich region, a flexiNR motif (LxxIL), or the W671A substitution. (**I**) Full-length VIRF proteins harboring either WT or mutant effector domains are expressed in K562 cells from a dox-inducible promoter (Methods) to measure differential effects on host gene expression. (**J**-**K**) Comparison of K562 gene expression upon overexpression of WT or mutant VIRFs, as measured by RNA-seq and presented as log-2 fold changes of significantly up- or downregulated genes relative to an mCitrine-expressing negative control. The thick yellow and blue lines indicate the mean change of all genes. Genes normally up- or downregulated by interferon beta treatment (i.e. interferon-stimulated genes or ISGs) are indicated in red.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

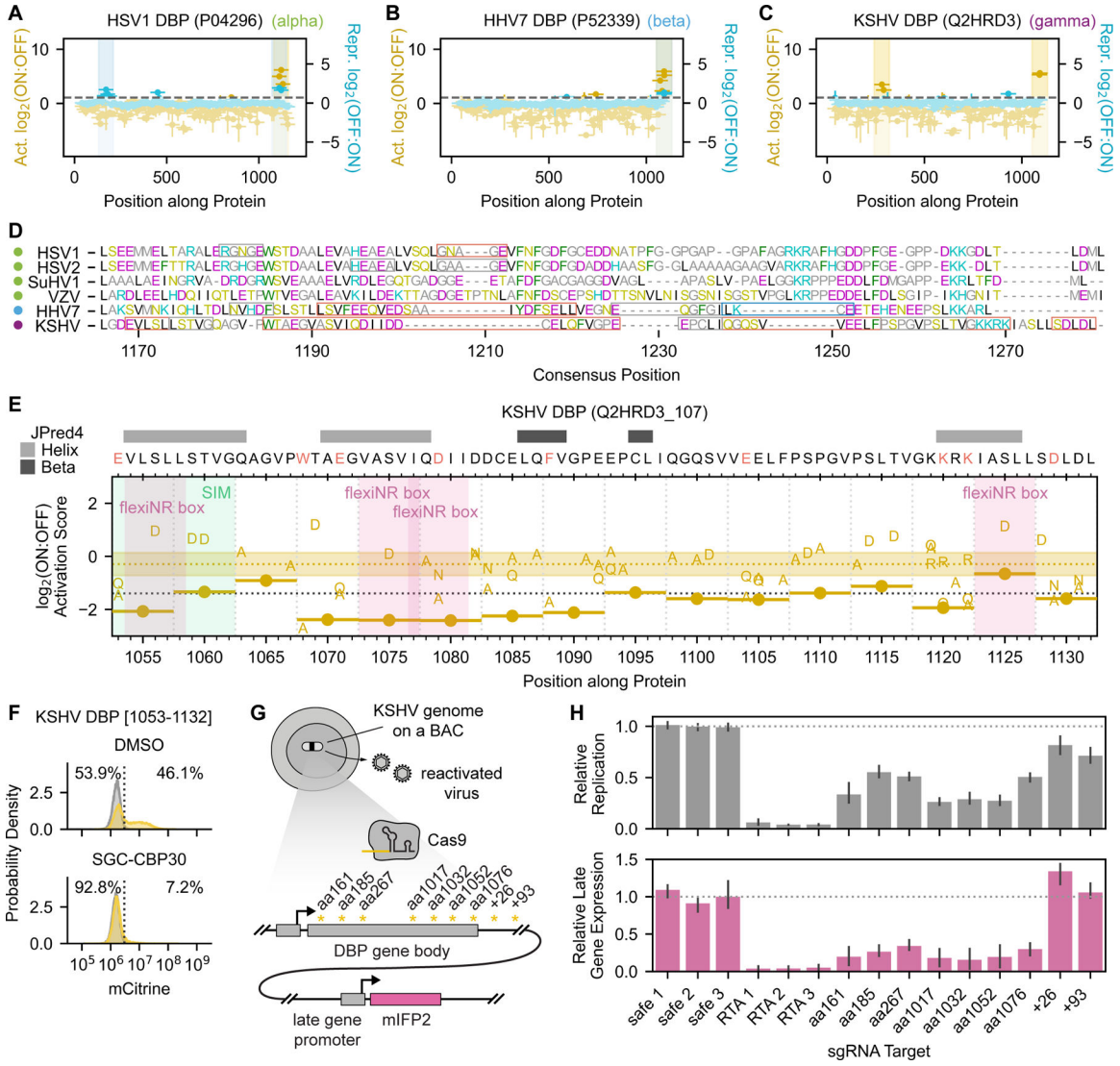


Fig. 6 |. The herpesvirus DBP C-terminus regulates late gene expression and replication.

(**A-C**) Tiling plots showing the C-terminal activation domain of the alphaherpesvirus HSV1 homolog (A), betaherpesvirus HHV7 homolog (B), and gamma KSHV homolog (C). (**D**) Multiple sequence alignment of the DBP C-terminal region for all homologs with activation potential, with amino acids colored by biochemical similarity. Colored dots to the left of the virus names indicate the herpesvirus subfamily: alpha (green), beta (blue), and gamma (purple). Red boxes indicate essential regions whose deletion breaks activity, gray boxes indicate sensitive regions whose deletion reduces activity by approximately two-fold, and the right-most blue box in HHV7 DBP indicates the SUMOylation site (LKCE) whose deletion increases activation. (**E**) Tiling plot showing the effects of 5aa deletions and single-residue substitutions on KSHV DBP activation domain activity. A SUMO-interaction motif (SIM) is highlighted with a green vertical span, and four FlexiNR box motifs are highlighted with pink vertical spans. (**F**) Flow cytometry distributions of cells with the minCMV-citrine reporter after two days of recruitment of the KSHV DBP C-terminal activation domain (aa1053-1132) with DMSO or 10μM SGC-CBP30 (p300/CBP inhibitor). (**G**) Overview

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of assay to measure the consequence of Cas9-induced KSHV DBP truncations on late gene expression during KSHV reactivation. Latently infected iSLK cells harbor the KSHV genome on a bacterial artificial chromosome (BAC). Targeting Cas9 to various regions of the DBP gene body produces truncated gene products whose effects on late gene expression can be measured using a KSHV genome-integrated mIFP2 reporter gene under the control of a late gene promoter. sgRNA targets are indicated by asterisks, with the approximate position indicated (residue position in gene body as 'aa' and base pairs past the stop codon as '+'). (**H**) Quantification of EdU incorporation during viral genome replication (top) and mIFP2-positivity (bottom) at 48 hours after reactivation from latency. Safe sgRNAs (safe 1–3) targeting a non-functional locus of the KSHV genome serve as negative controls that have minimal effects on DBP expression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

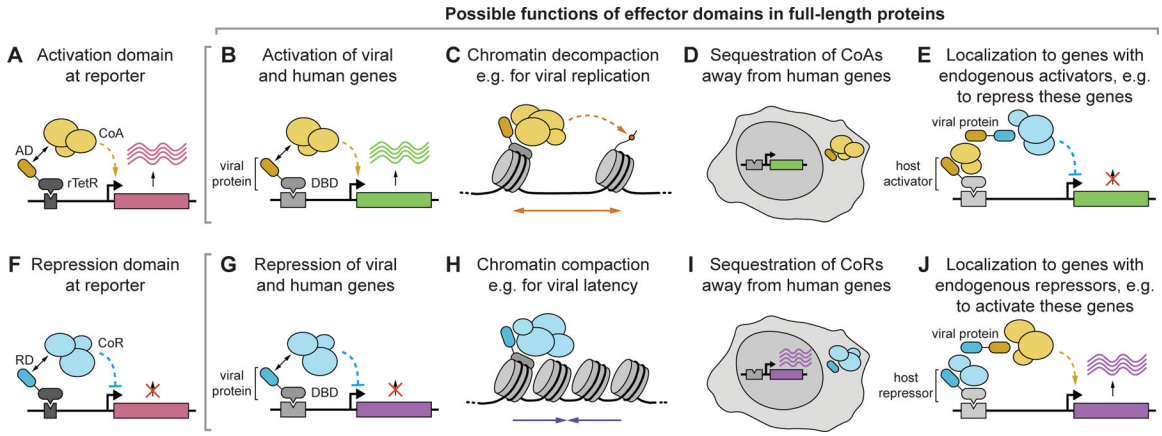


Fig. 7 |. Possible roles of transcriptional effector domains discovered with a reporter assay within full-length proteins.

(**A**) Domain from viral protein (AD, yellow) activates a reporter by recruiting coactivators (CoA). (**B**) If the full-length viral protein has a DNA binding domain, the AD can directly activate viral or human genes. (**C**) The viral protein could use the AD to recruit CoAs to chromatin, leading to permissive chromatin. For example, chromatin decompaction could enhance viral replication rather than controlling transcription directly. (**D**) The AD could sequester CoAs from endogenous genes (either at other places in the nucleus, or in the cytoplasm) leading to an indirect decrease of human gene expression. (**E**) The AD could bind CoAs to localize the viral protein to actively transcribed regions, and use a dominant repressive domain (blue) to repress these genes. (**F**) Domain from viral protein (RD, blue) represses a reporter by recruiting corepressors (CoR). (**G**) If the full-length viral protein has a DNA binding domain, the RD can directly repress viral or human genes. (**H**) The viral protein could use the RD to recruit CoRs to chromatin, leading to repressive chromatin, as might be needed for example for viral latency. (**I**) The RD could sequester CoRs from endogenous genes leading to an indirect increase of human gene expression. (**J**) The RD could bind CoRs to localize the viral protein to repressed regions, and use a dominant activation domain (yellow) to activate these genes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
DYKDDDDK Epitope Tag Alexa Fluor 647-conjugated Antibody	R&D Systems	IC8529R
Bacterial and virus strains		
Endura Electrocompetent Cells	Lucigen	60242–2
Chemicals, peptides, and recombinant proteins		
Doxycycline hyclate	Tocris	4090
Celastrol	Tocris	3203
SGC-CBP30	Selleck Chemicals	S7256
Tazemetostat	Selleck Chemicals	S7128
TMP269	Selleck Chemicals	S7324
Human Interferon beta	PeproTech	300–02BC
Critical commercial assays		
Dynabeads M-280 Protein G	Thermo Fisher	10003D
NEBNext Ultra II RNA Library Prep Kit	New England Biolabs	E7770S
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490
Deposited data		
Raw HT-Recruit sequencing files	This manuscript	BioProject PRJNA930640
Raw and analyzed RNA-seq data	This manuscript	BioProject PRJNA930640; GEO GSE224325
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
EBV type 1 B95-8 strain reference sequence	NCBI	NC_007605.1
EBV type 2 AG876 strain reference sequence	GenBank	DQ279927.1
Experimental models: Cell lines		
Human: HEK293T Lenti-X	Takara	632180
Human: K562	ATCC	CCL-243
Human: K562 with pJT039 pEF reporter	Tycko et al.[8]	N/A
Human: K562 with pDY32 minCMV reporter	Tycko et al.[8]	N/A
Human: Cas9+ iSLK-BAC16-K8.1pr-mIFP2	Morgens et al.[58]	N/A
Oligonucleotides		
vTR and CoV tiling libraries	Twist; see Table S1	N/A
Spike DMS library	Twist; see Table S1	N/A
HHV tiling library	Twist; see Table S2	N/A
HHV perturbation library	Twist; see Table S3	N/A
HHV hits library	Twist; see Table S3	N/A
HT-Recruit NGS library prep primers	Tycko et al.[8]	N/A
Recombinant DNA		

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pJT039 AAVS1-PuroR-9xTetO-pEF-IGKleader-hIgG1_FC-Myc-PDGFRb-T2A-Citrine-PolyA	Tycko et al.[8]	Addgene #161927
pDY32 AAVS1-PuroR-9xTetO-minCMV-IGKleader-hIgG1_FC-Myc-PDGFRb-T2A-Citrine-PolyA	Tycko et al.[8]	Addgene #161928
pJT126 lenti pEF-rTetR(SE-G72P)-3XFLAG-LibCloneSite-T2A-mCherry-BSD-WPRE	Tycko et al.[8]	Addgene #161926
pCL040 lenti TRE-3xFLAG-LibCloneSite pEF-rTetR(SE-G72P)-VP48-T2A-mCherry-BSD-WPRE	This manuscript	Addgene #198054
pMD2.G	Didier Trono	Addgene #12259
pRSV-Rev	Didier Trono	Addgene #12253
pMDLg/pRRE	Didier Trono	Addgene #12251
Software and algorithms		
HT-Recruit analysis pipeline	Tycko et al.[8]	https://github.com/bintulab/HT-recruit-Analyze
Cytoflow	Teague[67]	https://cytoflow.github.io/
DESeq2	Love et al.[52]	DOI: 10.18129/B9.bioc.DESeq2
Custom analyses	This manuscript	https://github.com/bintulab/Viral_Ludwig_2022 (version of record deposited at DOI: 10.5281/zenodo.7874429)
Other		
Miscellaneous data and resources	This manuscript	https://github.com/bintulab/Viral_Ludwig_2022 (version of record deposited at DOI: 10.5281/zenodo.7874429)
Antibodies		
Rabbit monoclonal anti-Snail	Cell Signaling Technology	Cat#3879S; RRID: AB_2255011
Mouse monoclonal anti-Tubulin (clone DM1A)	Sigma-Aldrich	Cat#T9026; RRID: AB_477593
Rabbit polyclonal anti-BMAL1	This manuscript	N/A
Bacterial and virus strains		
pAAV-hSyn-DIO-hM3D(Gq)-mCherry	Krashes et al.[1]	Addgene AAV5; 44361-AAV5
AAV5-EF1a-DIO-hChR2(H134R)-EYFP	Hope Center Viral Vectors Core	N/A
Cowpox virus Brighton Red	BEI Resources	NR-88
Zika-SMGC-1, GENBANK: KX266255	Isolated from patient (Wang et al.[2])	N/A
Staphylococcus aureus	ATCC	ATCC 29213
Streptococcus pyogenes: M1 serotype strain: strain SF370; M1 GAS	ATCC	ATCC 700294
Biological samples		
Healthy adult BA9 brain tissue	University of Maryland Brain & Tissue Bank; http://medschool.umaryland.edu/btbank/	Cat#UMB1455
Human hippocampal brain blocks	New York Brain Bank	http://nybb.hs.columbia.edu/
Patient-derived xenografts (PDX)	Children's Oncology Group Cell Culture and Xenograft Repository	http://cogcell.org/
Chemicals, peptides, and recombinant proteins		
MK-2206 AKT inhibitor	Selleck Chemicals	S1078; CAS: 1032350–13-2

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SB-505124	Sigma-Aldrich	S4696; CAS: 694433–59-5 (free base)
Picrotoxin	Sigma-Aldrich	P1675; CAS: 124–87-8
Human TGF-β	R&D	240-B; GenPept: P01137
Activated S6K1	Millipore	Cat#14-486
GST-BMAL1	Novus	Cat#H00000406-P01
Critical commercial assays		
EasyTag EXPRESS 35S Protein Labeling Kit	PerkinElmer	NEG772014MC
CaspaseGlo 3/7	Promega	G8090
TruSeq ChIP Sample Prep Kit	Illumina	IP-202-1012
Deposited data		
Raw and analyzed data	This manuscript	GEO: GSE63473
B-RAF RBD (apo) structure	This manuscript	PDB: 5J17
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Nanog STILT inference	This manuscript; Mendeley Data	http://dx.doi.org/10.17632/wx6s4mj7s8.2
Affinity-based mass spectrometry performed with 57 genes	This manuscript; Mendeley Data	Table S8; http://dx.doi.org/10.17632/5hvpvspw82.1
Experimental models: Cell lines		
Hamster: CHO cells	ATCC	CRL-11268
D. melanogaster: Cell line S2: S2-DRSC	Laboratory of Norbert Perrimon	FlyBase: FBtc0000181
Human: Passage 40 H9 ES cells	MSKCC stem cell core facility	N/A
Human: HUES 8 hESC line (NIH approval number NIHhESC-09-0021)	HSCI iPS Core	hES Cell Line: HUES-8
Experimental models: Organisms/strains		
C. elegans: Strain BC4011: srl-1(s2500) II; dpy-18(e364) III; unc-46(e177)rol-3(s1040) V.	Caenorhabditis Genetics Center	WB Strain: BC4011; WormBase: WBVar00241916
D. melanogaster: RNAi of Sxl: y[1] sc[*] v[1]; P{TRiP.HMS00609}attP2	Bloomington Drosophila Stock Center	BDSC:34393; FlyBase: FBtp0064874
S. cerevisiae: Strain background: W303	ATCC	ATTC: 208353
Mouse: R6/2: B6CBA-Tg(HDexon1)62Gpb/3J	The Jackson Laboratory	JAX: 006494
Mouse: OXTRfl/fl: B6.129(SJL)-$Oxtr^{tm1.1Wsy}$/J	The Jackson Laboratory	RRID: IMSR_JAX:008471
Zebrafish: Tg(Shha:GFP)t10: t10Tg	Neumann and Nuesslein-Volhard[3]	ZFIN: ZDB-GENO-060207-1
Arabidopsis: 35S::PIF4-YFP, BZR1-CFP	Wang et al.[4]	N/A
Arabidopsis: JYB1021.2: pS24(AT5G58010)::cS24:GFP(-G):NOS #1	NASC	NASC ID: N70450
Oligonucleotides		
siRNA targeting sequence: PIP5K I alpha #1: ACACAGUACUCAGUUGAUA	This manuscript	N/A
Primers for XX, see Table SX	This manuscript	N/A
Primer: GFP/YFP/CFP Forward: GCACGACTTCTTCAAGTCCGCCATGCC	This manuscript	N/A

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Morpholino: MO-pax2a GGTCTGCTTTGCAGTGAATATCCAT	Gene Tools	ZFIN: ZDB-MRPHLNO-061106-5
ACTB (hs01060665_g1)	Life Technologies	Cat#4331182
RNA sequence: hnRNPA1_ligand: UAGGGACUUAGGGUUCUCUCUAGGGACUUAGGGUUCUCUCUAGGGA	This manuscript	N/A
Recombinant DNA		
pLVX-Tight-Puro (TetOn)	Clonetech	Cat#632162
Plasmid: GFP-Nito	This manuscript	N/A
cDNA GH111110	Drosophila Genomics Resource Center	DGRC:5666; FlyBase:FBcl0130415
AAV2/1-hsyn-GCaMP6-WPRE	Chen et al.[5]	N/A
Mouse raptor: pLKO mouse shRNA 1 raptor	Thoreen et al.[6]	Addgene Plasmid #21339
Software and algorithms		
ImageJ	Schneider et al.[7]	https://imagej.nih.gov/ij/
Bowtie2	Langmead and Salzberg[8]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al.[9]	http://samtools.sourceforge.net/
Weighted Maximal Information Component Analysis v0.9	Rau et al.[10]	https://github.com/ChristophRau/wMICA
ICS algorithm	This manuscript; Mendeley Data	http://dx.doi.org/10.17632/5hvpvspw82.1
Other		
Sequence data, analyses, and resources related to the ultra-deep sequencing of the AML31 tumor, relapse, and matched normal	This manuscript	http://aml31.genome.wustl.edu
Resource website for the AML31 publication	This manuscript	https://github.com/chrisamiller/aml31SuppSite
Chemicals, peptides, and recombinant proteins		
QD605 streptavidin conjugated quantum dot	Thermo Fisher Scientific	Cat#Q10101MP
Platinum black	Sigma-Aldrich	Cat#205915
Sodium formate BioUltra, ≥99.0% (NT)	Sigma-Aldrich	Cat#71359
Chloramphenicol	Sigma-Aldrich	Cat#C0378
Carbon dioxide (^{13}C, 99%) (<2% ^{18}O)	Cambridge Isotope Laboratories	CLM-185-5
Poly(vinylidene fluoride-co-hexafluoropropylene)	Sigma-Aldrich	427179
PTFE Hydrophilic Membrane Filters, 0.22 μm, 90 mm	Scientificfilters.com/Tisch Scientific	SF13842
Critical commercial assays		
Folic Acid (FA) ELISA kit	Alpha Diagnostic International	Cat# 0365–0B9
TMT10plex Isobaric Label Reagent Set	Thermo Fisher	A37725
Surface Plasmon Resonance CM5 kit	GE Healthcare	Cat#29104988
NanoBRET Target Engagement K-5 kit	Promega	Cat#N2500
Deposited data		
B-RAF RBD (apo) structure	This manuscript	PDB: 5J17

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Structure of compound 5	This manuscript; Cambridge Crystallographic Data Center	CCDC: 2016466
Code for constraints-based modeling and analysis of autotrophic *E. coli*	This manuscript	https://gitlab.com/elad.noor/sloppy/tree/master/rubisco
Software and algorithms		
Gaussian09	Frish et al.[1]	https://gaussian.com
Python version 2.7	Python Software Foundation	https://www.python.org
ChemDraw Professional 18.0	PerkinElmer	https://www.perkinelmer.com/category/chemdraw
Weighted Maximal Information Component Analysis v0.9	Rau et al.[2]	https://github.com/ChristophRau/wMICA
Other		
DASGIP MX4/4 Gas Mixing Module for 4 Vessels with a Mass Flow Controller	Eppendorf	Cat#76DGMX44
Agilent 1200 series HPLC	Agilent Technologies	https://www.agilent.com/en/products/liquid-chromatography
PHI Quantera II XPS	ULVAC-PHI, Inc.	https://www.ulvac-phi.com/en/products/xps/phi-quantera-ii/