

Supplement Information

Reorganization of the host epigenome by a viral oncogene

Roberto Ferrari^{1,8}, Trent Su^{1,2}, Bing Li¹, Giancarlo Bonora^{1,3}, Amit Oberai¹, Yvonne Chan¹,
Rajkumar Sasidharan⁷ Arnold J. Berk^{5,6}, Matteo Pellegrini^{7,8}
and Siavash K. Kurdistani^{1,4,6,8}

Departments of ¹Biological Chemistry, ²Division of Oral Biology and Medicine, School of Dentistry, ³UCLA Bioinformatics Interdepartmental Degree Program and ⁴Pathology and Laboratory of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA.

⁵Department of Microbiology, Immunology and Molecular Genetics and ⁶Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

⁷Department of Molecular, Cellular, and Developmental Biology, University of California, Los Angeles, CA 90095, USA.

⁸Eli and Edythe Broad Centre of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA 90095, USA

Supplemental Methods

Generation of H3K18ac antibody

We used a cysteine-conjugated peptide (GGKAPRK[Ac]QLASK-C) for injection of eight young male rabbits. Initial injection of 250 µg peptide into each rabbit using Freund's complete adjuvant was followed with a boost of 250 µg with Freund's incomplete. Injection series were as follows: i) Day 0 (prebleed) we used for the initial inoculation of 0.25 ml conjugate emulsified with 0.5 ml Freunds Complete for three injections along the back of eight rabbits. ii) Day 21, we performed a boost of 0.25 ml conjugated emulsified with 0.5 ml Freunds incomplete three injections along the back of eight rabbits. iii) Day 28 (exanguination). The anti-sera from one of the rabbits (814) were fully characterized by ELISA, western blotting and ChIP against histone H3 lysine mutations (see Fig S1).

Antibodies used for ChIP

H3K18ac (814) 1:250 dilution, H3K9ac (07-352, Upstate) at 1:250 dilution, RB1 (4H1) (9309L, Cell Signaling) at 1:400, Santa Cruz (C20) at 1:400, Santa Cruz RBL1 (C18) at 1:400, H3K4me1 Abcam (ab8895) at dilution 1:250, EP300 Santa Cruz (sc-C15) and CREBBP Santa Cruz (sc-A22) at dilution 1:100.

Yeast ChIP validation

ChIP validation was performed as described in (Suka et al. 2001). Primers used for PCR are listed below.

iYDR224C-F 5'-ACTATGGTTAGACGCTCAATGTCG-3'

iYDR224C-R 5'-AGATGCCCTTCTTACCAATC-3'

TEL-F 5'-GCGTAACAAAGCCATAATGCCTCC-3'

TEL-R 5'-CTCGTTAGGATCACGTTCGAATCC-3'

ChIP-qPCR

Primers used for qPCR are listed below.

CCNE2-F 5'-CCTTCGCTGCCTCTATGAAT-3'

CCNE2-R 5'-ATCTTGTTCCCGGAGCTGT-3'

POLD3-F 5'-GAGCCTTGACCCCTCTGTCTG-3'

POLD3-R 5'-ACTTCCGTGTGCTGGAG-3'

COL6A3-F 5'-CACTTCTGAGCAGCCAATG-3'

COL6A3-R 5'-AAGGGTGATCCCACAGAATGC-3'

CCNE2_EP300/CREBBP-F 5'-ATGACCCCCAGTCGTCTAGTT-3'

CCNE2_EP300/CREBBP-R 5'-GGGATGTGGGAGGAAAGAAAG-3'

POLD3_EP300/CREBBP-F 5'-CGTCTCCCTGACTAGTTGACC-3'

POLD3_EP300/CREBBP-R 5'-CCTCTAACCCCTCTCTCAATCA-3'

COL6A3_EP300/CREBBP-F 5'-CATTTCACAGGACTGCCTGTGT-3'

COL6A3_EP300/CREBBP-R 5'-GGCTGTAGAATTATGCCAGATG-3'

GAPDH-F 5'-GCCAATCTCAGTCCCTTCCC-3'

GAPDH-R 5'-CGCCCGTAAAACCGCTAGTAG-3'

CDC7-F 5'-GTGAGTTCCGACGGTTGT-3'

CDC7-R 5'-AAACCGAAGTCACGATCCAG-3'

FOS gene primers are from MAGnify Chromatin immunoprecipitation System (Life Technology 49-2024).

Statistical Analysis

External data source

Raw data for H3K4me1, H3K9ac and H3K18ac in asynchronous IMR90 cells was downloaded from the NCBI epigenome roadmap: (<http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/>).

Raw data for EP300 and CREBBP binding was downloaded from NCBI GEO (Ramos et al. 2010) (GSE21026). Raw data for EP300 in human H9 stem cells was downloaded from GEO (GSE24447) (Rada-Iglesias et al. 2011). Raw data for E2F1 binding in MCF7 cells was downloaded from GEO (GSE28286). All the datasets were processed using the algorithm described above.

Genomic annotation

Coordinates for all the genomic annotation regions used in this work were downloaded from UCSC Human genome (hg19) tables.

Enrichment calculation for Intergenic Regions (IRs)

Intergenic regions (IRs) were defined as genomic locations with no gene annotation on either the (+) or (-) strand based on Refseq gene annotation. To calculate enrichment and coverage of specific regions of interest, we coded a script that compared: i) a .bed format file which provides the chromosomal coordinates for the feature of interest, for example the coding region of a gene or double intergenic regions, and ii) a .wig file which provides the coordinates for regions that are enriched in the ChIP-seq experiment in 100-bp segments. The output calculates: i) the size in bp of each feature, ii) the coverage for the feature, and iii) the enrichment of reads over the feature. These values were calculated for each feature across the genome. The probe coverage for the feature is the number of base pairs over the feature that have reads mapped to them, while the enrichment of reads over the feature quantifies the number of reads matched to each base pair within the feature.

CEAS analysis

Cis-regulatory Elements Annotation System (CEAS) was used to create average profiling of specific list of genes (Fig 5A-D) and overlaps of significant peaks with genomic annotation regions (Fig 5F-G). For p-value calculations, please refer to publication⁸. For tiling profiles within CEAS we used wig files generated as described above. Only counts for significant windows were retained in these wig files.

Sitepro analysis

Sitepro as part of the Cistrome Analysis pipeline (<http://cistrome.dfci.harvard.edu/ap/>) was used to profile levels of histone modifications for defined genomic intervals (Fig. 5A-F). Wig files provided for these analyses were same as described above.

Seqpos analysis

Seqpos algorithm from CEAS package was used to determine transcription factor binding motifs enrichments for genomic coordinates 600 bp upstream and downstream of the acetylation peak center using p-value threshold of 0.0001.

Cluster Z-scoring

The z scores of expression for each cluster were computed by calculating the average RPKM values of all genes within a cluster, subtracting the average of all genes in our dataset and dividing the resulting number by the standard deviation of the whole dataset. Each score has been then multiplied by the square root of the number of genes within the cluster.

$$Z = (\text{mean(cluster)} - \text{mean(all)}) / (\text{sd} / \sqrt{n}) = [(\text{mean(cluster)} - \text{mean(all)}) / \text{sd}] * \sqrt{n}$$

Reference

Chicas A, Wang X, Zhang C, McCurrach M, Zhao Z, Mert O, Dickins RA, Narita M, Zhang M, Lowe SW. 2010. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**(4): 376-387.

Ferrari R, Pellegrini M, Horwitz GA, Xie W, Berk AJ, Kurdistani SK. 2008. Epigenetic reprogramming by adenovirus e1a. *Science* **321**(5892): 1086-1088.

Horwitz GA, Zhang K, McBriar MA, Grunstein M, Kurdistani SK, Berk AJ. 2008. Adenovirus small e1a alters global patterns of histone modification. *Science* **321**(5892): 1084-1085.

Kleindienst R, Moeller L, Sinzinger S. 2010. Highly efficient refractive Gaussian-to-tophat beam shaper for compact terahertz imager. *Appl Opt* **49**(10): 1757-1763.

Pilch DR, Redon C, Sedelnikova OA, Bonner WM. 2004. Two-dimensional gel analysis of histones and other H2AX-related methods. *Methods Enzymol* **375**: 76-88.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**(7333): 279-283.

Ramos YF, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, van Galen M, van Dam H, van Ommen GJ, den Dunnen JT, Zantema A et al. 2010. Genome-wide assessment of differential roles for EP300 and CREBBP in transcription regulation. *Nucleic Acids Res* **38**(16): 5396-5408.

Suka N, Suka Y, Carmen AA, Wu J, Grunstein M. 2001. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol Cell* **8**(2): 473-479.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

Xu G, Deng N, Zhao Z, Judeh T, Flemington E, Zhu D. 2011. SAMMate: a GUI tool for processing short read alignments in SAM/BAM format. *Source Code Biol Med* **6**(1): 2.

Table S1. Comparison of histone acetylation peak number and coverage area in mock- vs. *d*/1500-infected cells.

H3K18ac	Mock (M)	e1a (E)	Ratio (E/M)
Number of peaks	16660	5500	0.33
Peaks coverage (Kb)	3899	914	0.23
H3K9ac	Mock (M)	e1a (E)	Ratio (E/M)
Number of peaks	24464	31354	1.28
Peaks coverage (Kb)	7794	8717	1.12

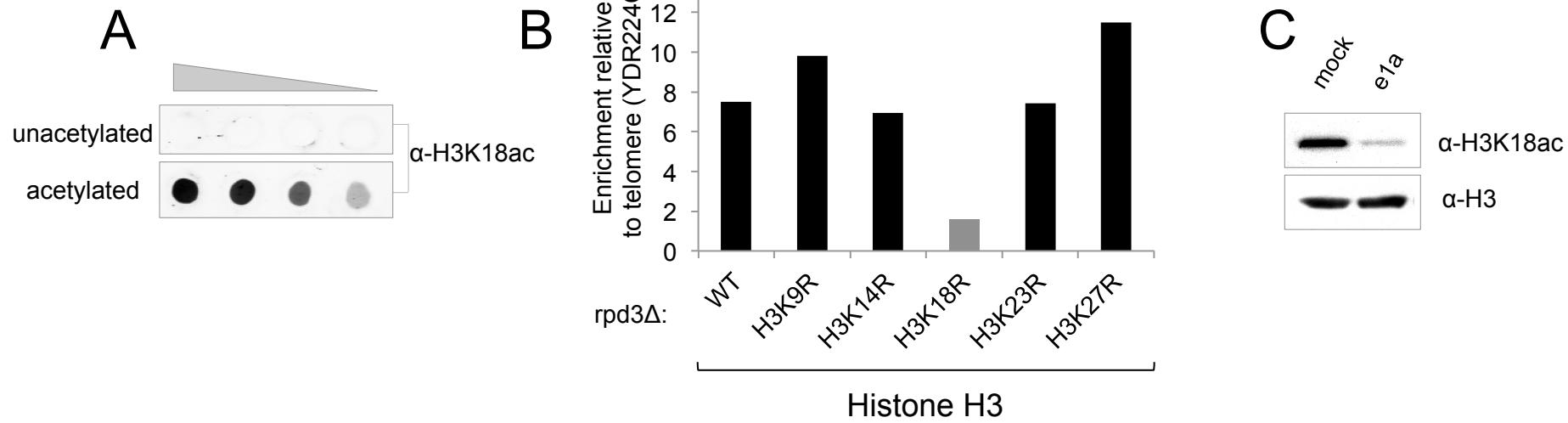


Fig. S1. Characterization of H3K18ac (custom 814) antibody. A) Peptide blot for unacetylated and acetylated H3K18 peptide. Decreasing amounts of peptide were blotted (grey triangle). B) Yeast strains with the indicated histone mutations were used to determine the specificity of H3K18ac antibody in ChIP at the YDR224C promoter region. The strains are as follows: NSY115 (WT), NSY119 (K9R), NSY120 (K14R), NSY130 (K18R), NSY131 (K23R), NSY132 (K27R). All strains lack the *RPD3* HDAC to increase acetylation globally. ChIP with 814 antibody in *rpd3* Δ -H3K18R shows strongly reduced enrichment compared to WT. In contrast, other mutations in different H3 lysines show enrichment comparable to WT. C) Western blot for H3K18ac (814) in mock and e1a-infected cells showing the specificity of the antibody in recognizing the decrease in H3K18ac upon infection.

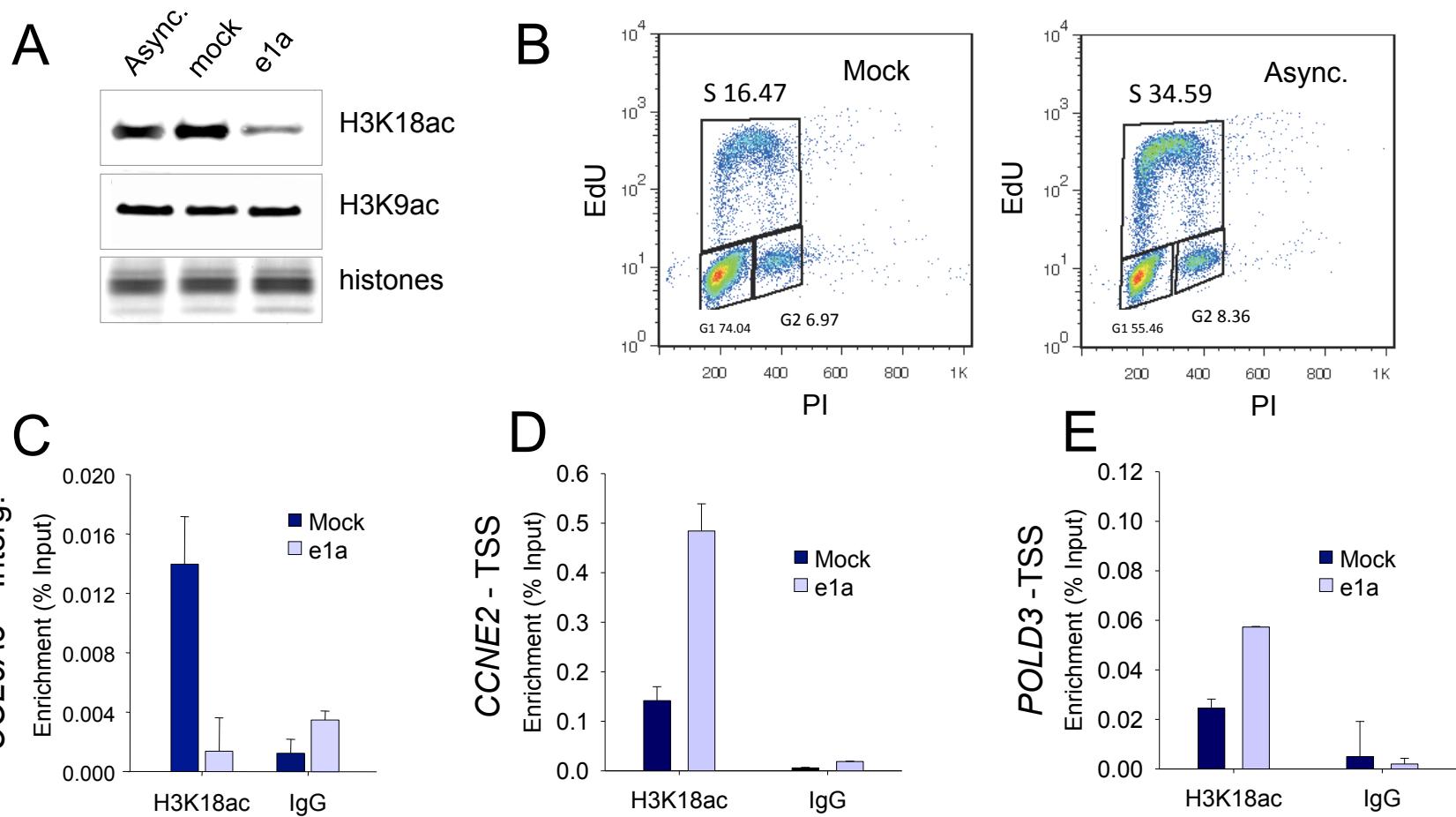


Fig. S2. Global level of H3K18ac and ChIP-quantitative PCR (ChIP-qPCR) validate ChIP-seq enrichments. A) Western blot for H3K18ac (814) and H3K9ac in asynchronous IMR90, mock- and *d/1500*-infected cells showing that a global decrease of H3K18ac is only observed upon *e1a* expression. Global levels of H3K9ac are comparable in all conditions. Total histones are used as loading control. **B)** FACS analysis (5-ethynyl-2'-deoxyuridine (EdU) versus propidium iodide (PI) staining) of contact-inhibited (mock) and asynchronous IMR90 cells showing a greater percentage of S-phase (S) cells in the asynchronous population compared to the mock-infected cells. **C-E)** H3K18ac ChIP was analyzed by qPCR for selected regions. The selected regions represent higher level of H3K18ac in *d/1500*-infected cells (*CCNE2*, *POLD3*) and higher level of H3K18ac in mock-infected cells (*COL6A3*). **C)** Collagen 6A3 (*COL6A3*) intergenic region. **D)** Cyclin E2 (*CCNE2*) promoter. **E)** DNA Polymerase D3 (*POLD3*) promoter.

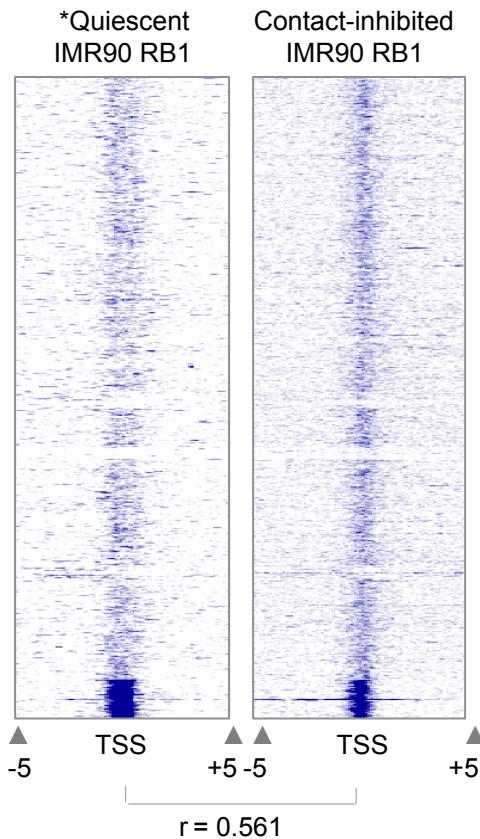


Fig. S3. RB1 binding in contact-inhibited (mock-infected) and quiescent* IMR90 fibroblast cells. Heat maps showing the binding of RB1 in the two conditions over a 10-kb region spanning the TSS of all annotated transcripts. The majority of the significant peaks (shades of dark blue) grouped in the same cluster.

*Chicas A. et al., Cancer Cell. 2010

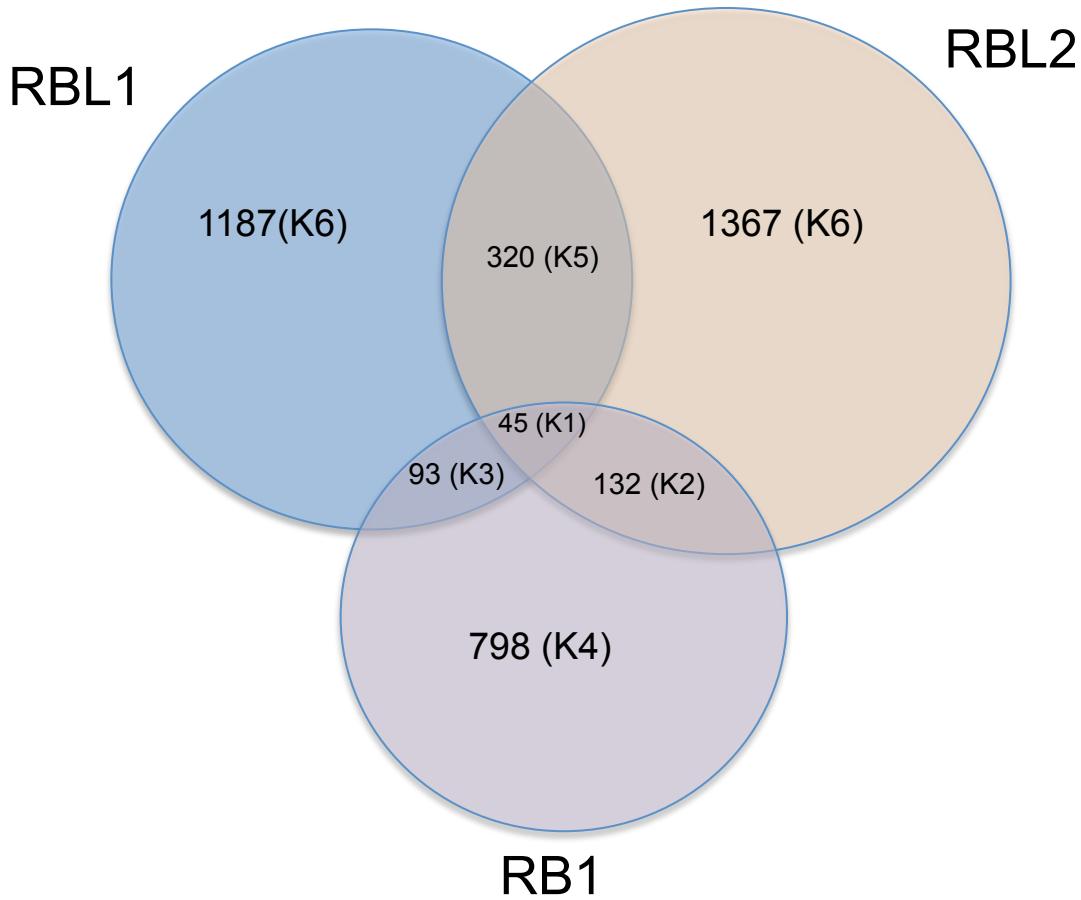


Fig. S4. RB family members share overlapping but distinct set of target promoter regions. Promoters having at least one significant peak of binding for any of the RB-family members RB1, RBL1 and RBL2 were selected and a Venn diagram representing the overlap among the bound promoters is shown. The numbers correspond to number of genes in each of the clusters (K) in figure 3D.

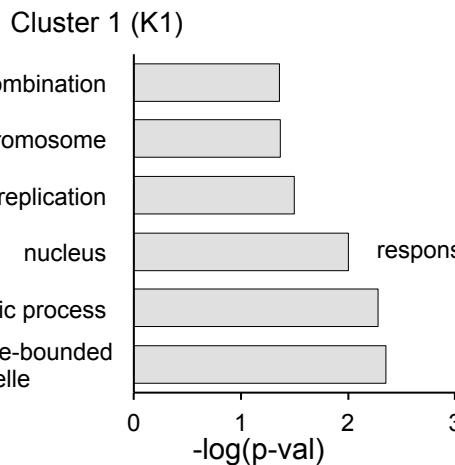
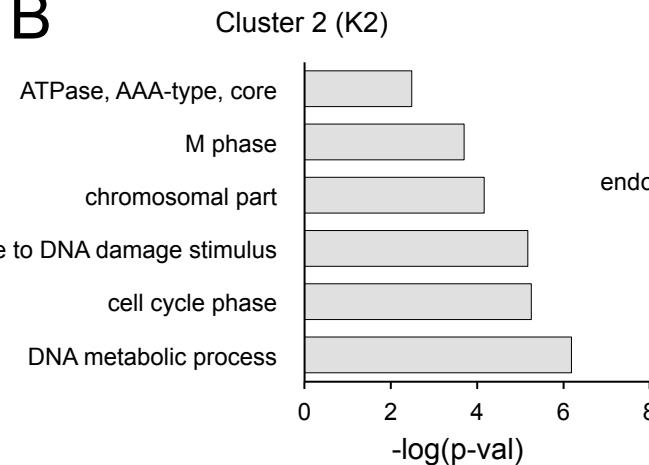
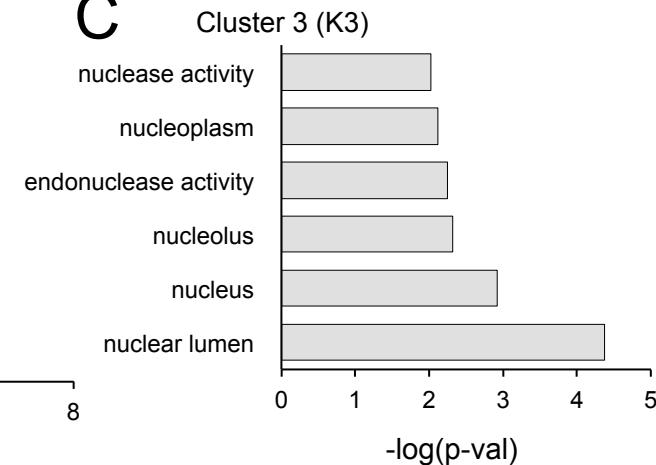
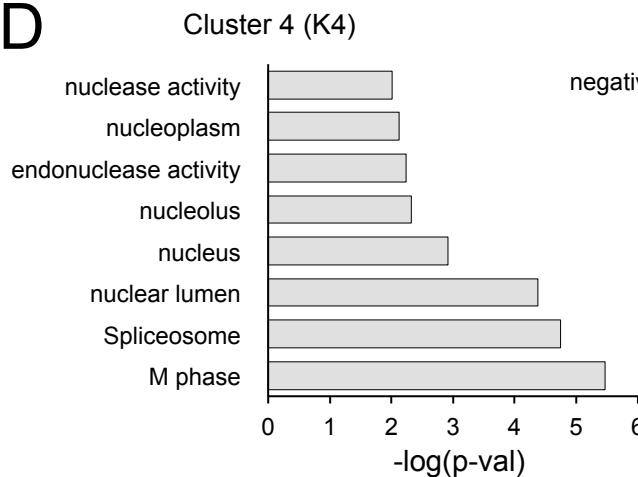
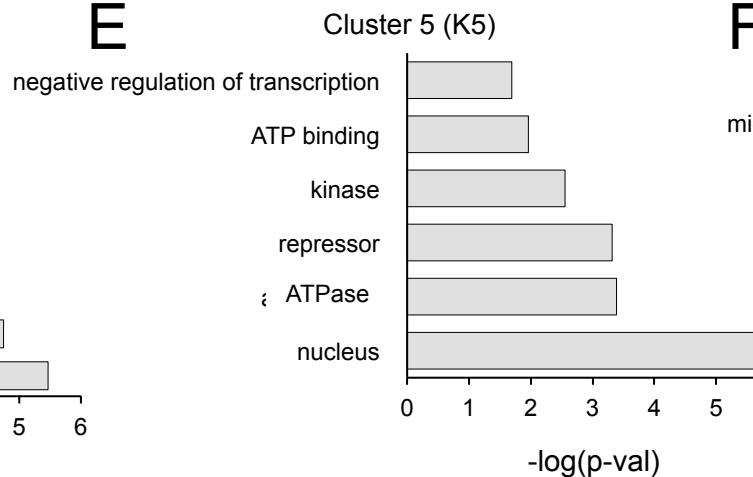
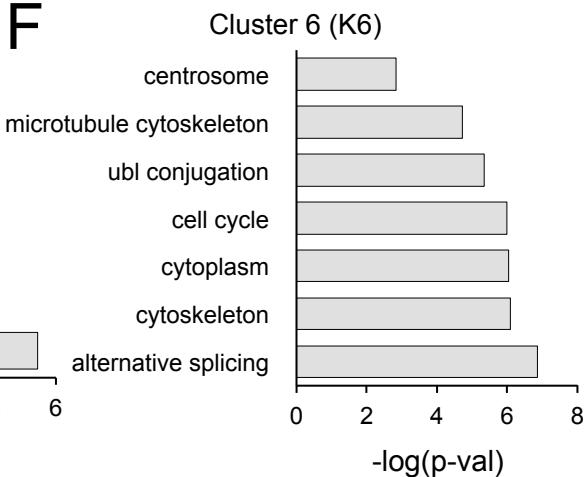
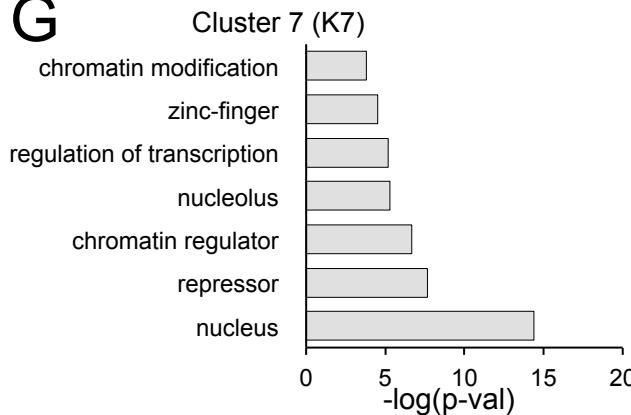
A**B****C****D****E****F****G**

Fig. S5. Gene ontology (GO) enrichment analysis of the RB-family target genes. Genes from clusters in figure 3D were used for GO analysis. Bars represent $-\log_{10}$ of the p-value for the selected GO terms. A-G refer to clusters 1-7 in figure 3D, respectively.

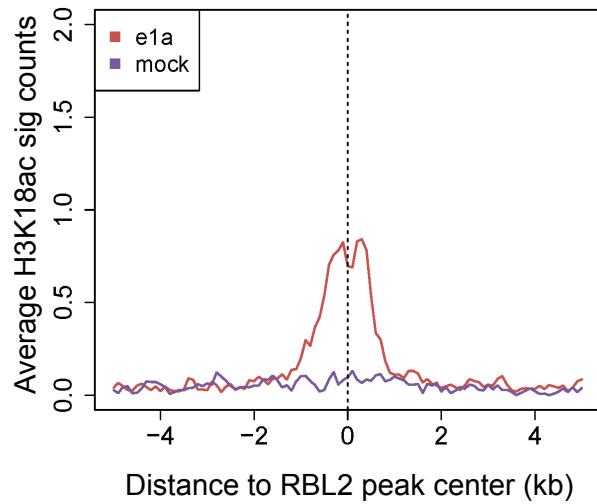


Fig. S6. The regions that are bound by RBL2 in contact-inhibited fibroblasts are targets of H3K18ac in e1a-expressing cells. Average levels (sig. counts) of H3K18ac in mock-infected and e1a-expressing cells are shown relative to RBL2-centered peaks that are 5 kb away from RB1 peaks.

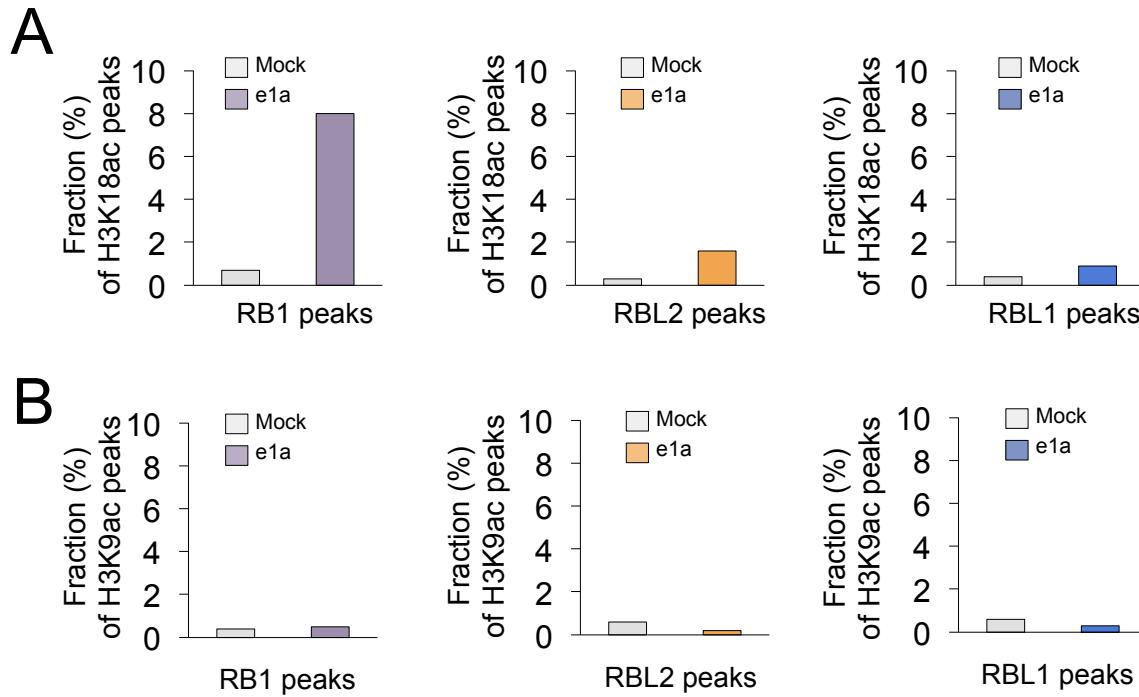


Fig. S7. H3K18ac in *d*/1500-infected cells is preferentially enriched precisely at regions bound by RB1 in uninfected cells. A) The fraction of H3K18ac peaks that precisely overlap with RB1, RBL2 and RBL1 peaks in mock- and *d*/1500-infected cells is shown. **B)** The same analysis as in A) for H3K9ac.

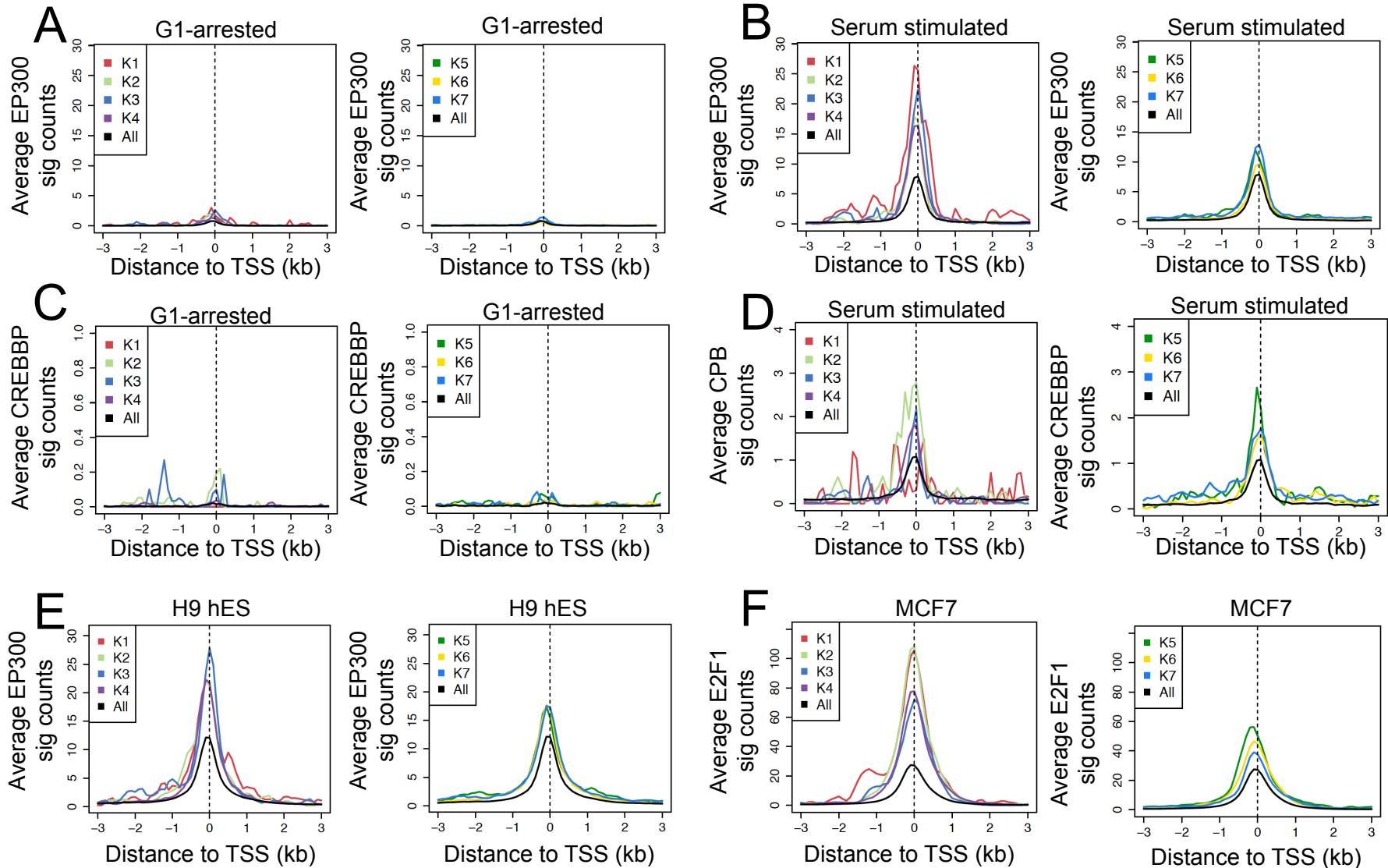


Fig. S8. Promoters bound by RB1 show increased levels of EP300 after serum stimulation in T98G glioblastoma and H9 hES cells.

A-B) Average levels (significant (sig.) counts) of EP300 binding in the seven RB-clusters (see Fig. 3D) for G1-arrested and serum stimulated T98G cells. **C-D)** Average levels of CREBBP binding in the seven RB-clusters for G1-arrested and serum stimulated T98G cells. **E)** Average levels of EP300 binding in the seven RB-clusters for H9 human embryonic stem cells (hES). **F)** Average levels of E2F1 binding in the seven RB-clusters for MCF7 cells.

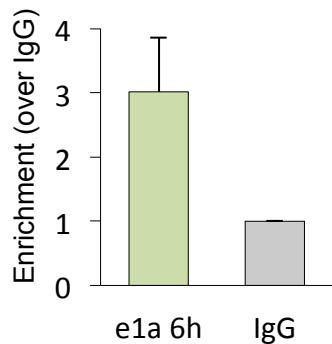
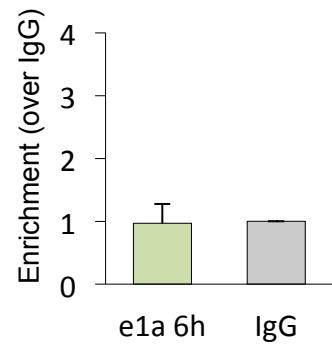
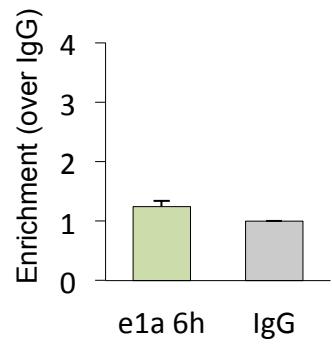
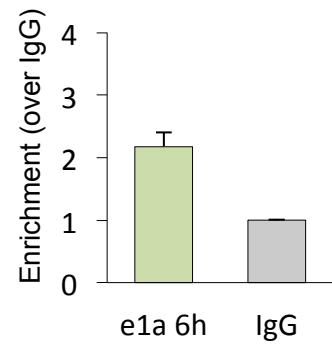
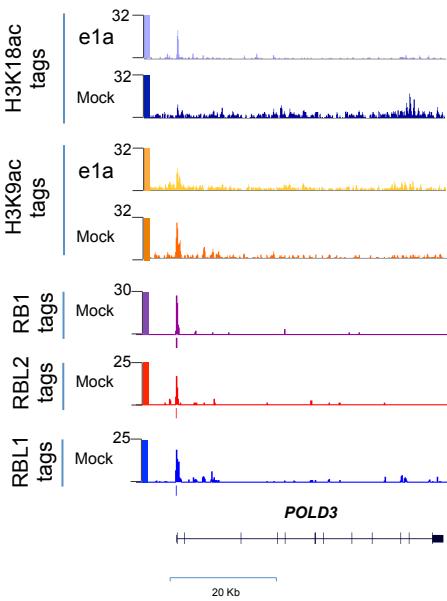
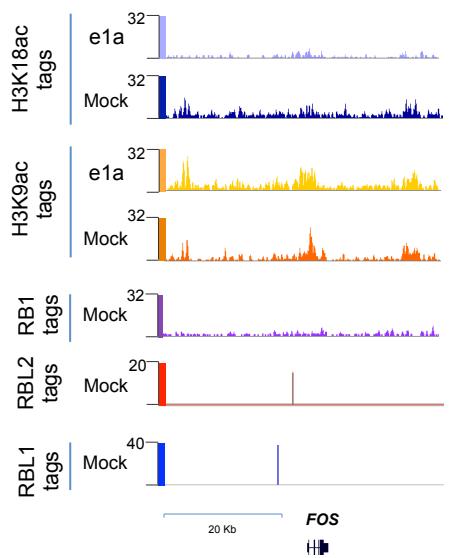
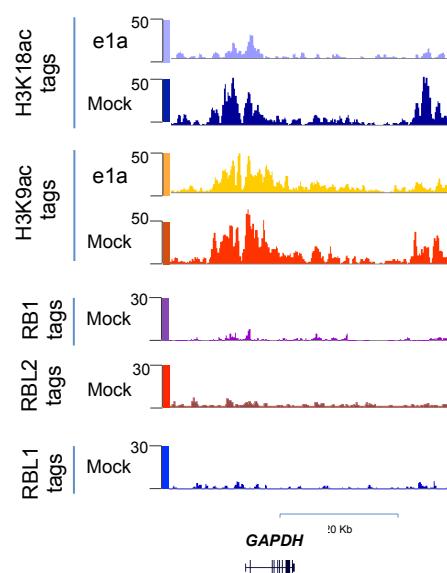
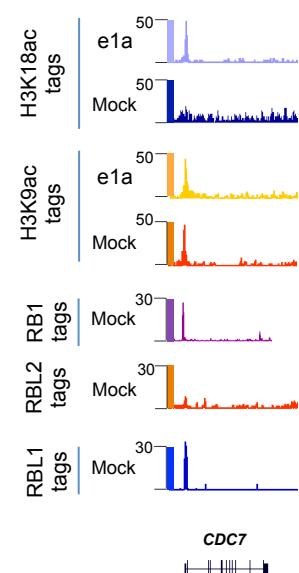
A**B****C****D****E****F****G****H**

Fig. S9. Regions bound by RB1 in contact-inhibited fibroblasts are also targets of e1a binding. **A-D)** Small e1a binding at 6 h post infection is shown for *POLD3*, *FOS*, *GAPDG* and *CDC7* loci compared to IgG control. The 6 h time point was chosen based on transient and early binding of e1a to cell cycle regulated genes (Ferrari et al. 2008). **E-H)** Patterns of H3K18ac, H3K9ac in mock- and *d*1500-infected cells and RB1, RBL2 and RBL1 at *POLD3*, *FOS*, *GAPDG* and *CDC7* loci are also shown. For each histone modification and TF binding, the y-axis indicates the number of input-normalized ChIP-seq reads across the locus (x-axis).

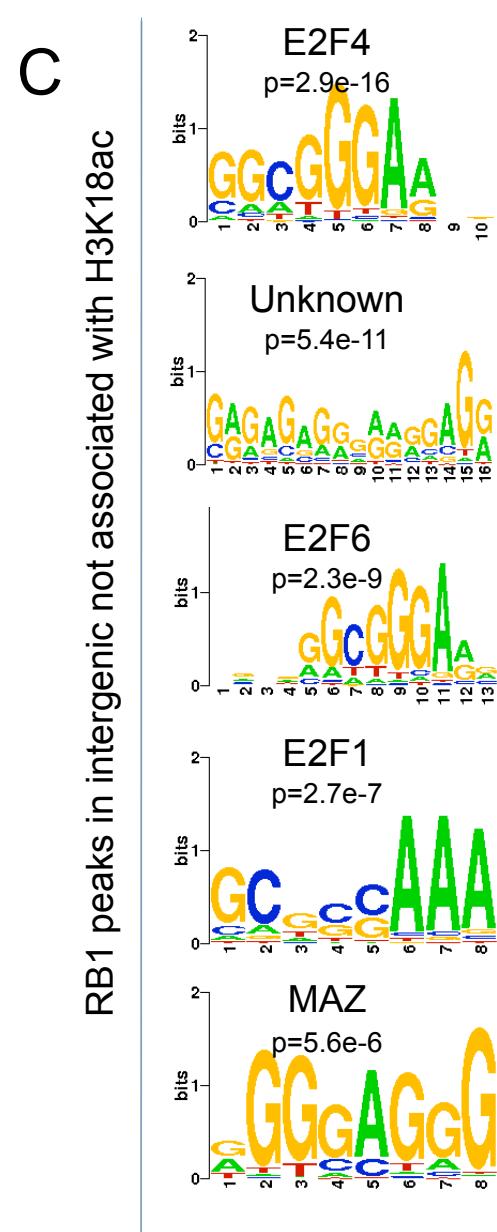
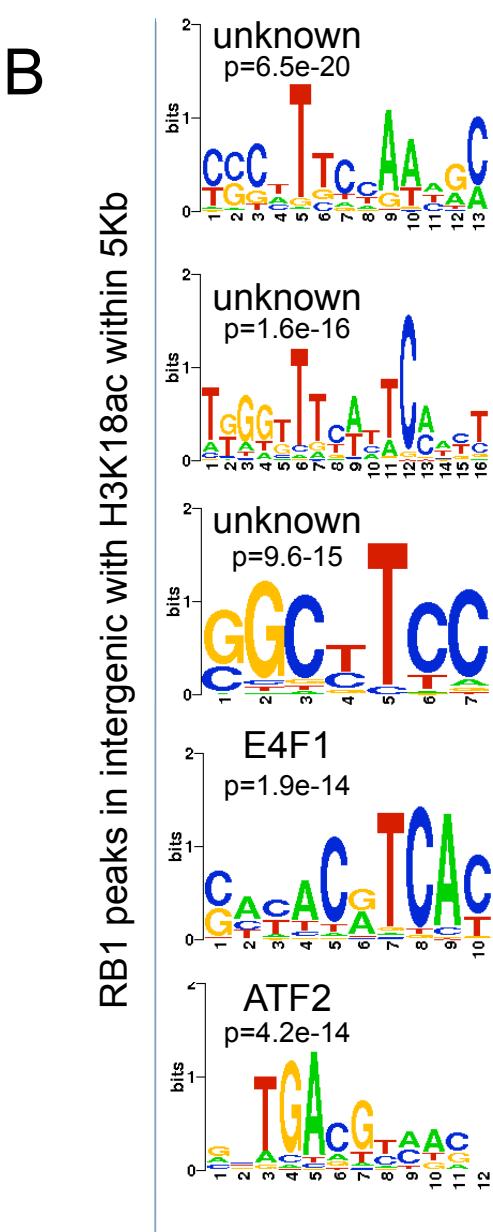
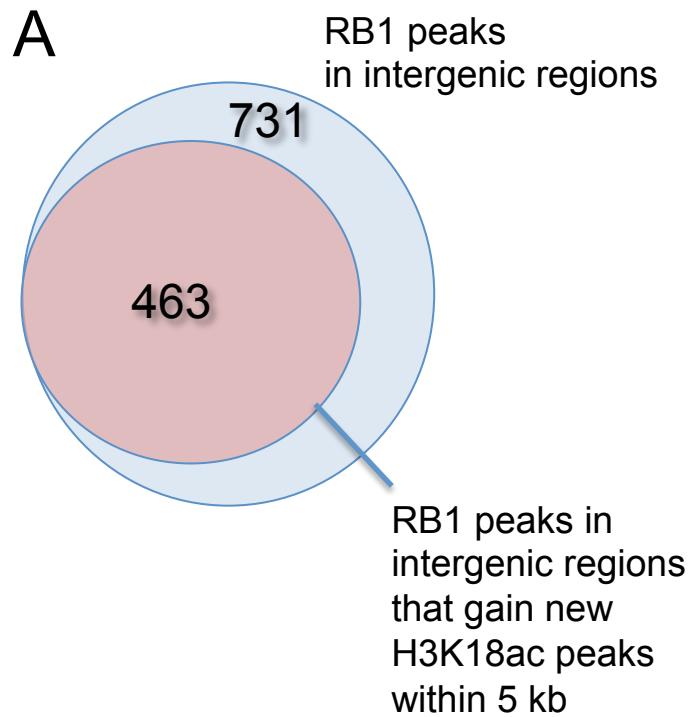


Fig. S10. H3K18ac peaks are associated with differential enrichment for transcription factor binding signatures in RB1-bound intergenic regions (IRs). A) ~63% of RB1 peaks in IRs are associated with H3K18ac. **B)** The 600-bp region around the peaks of IR RB1 associated with H3K18ac in e1a-expressing cells were analyzed for TF binding motifs using Seqpos (Galaxy). **C)** The same for IR RB1 peaks not associated with H3K18ac in e1a-expressing cells.

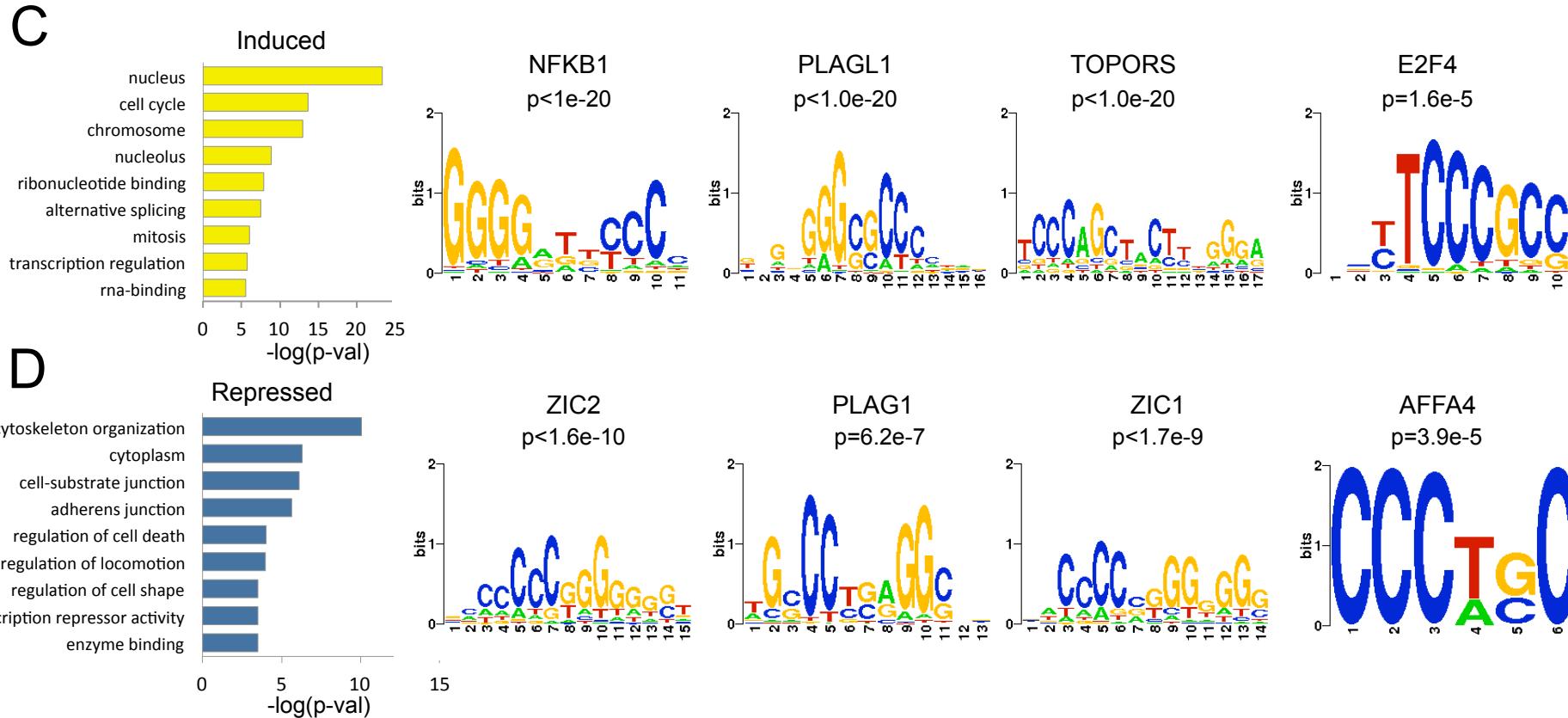
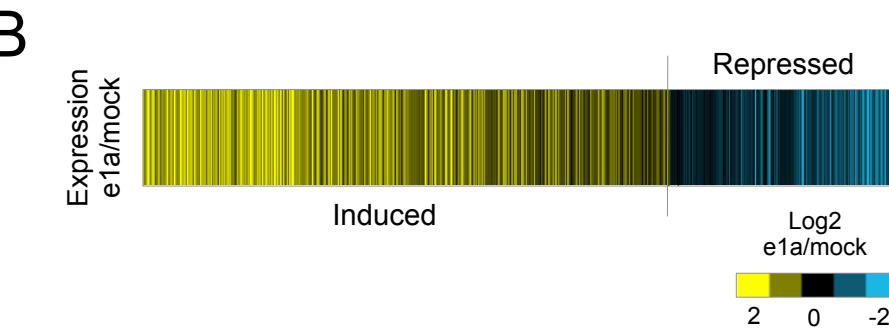
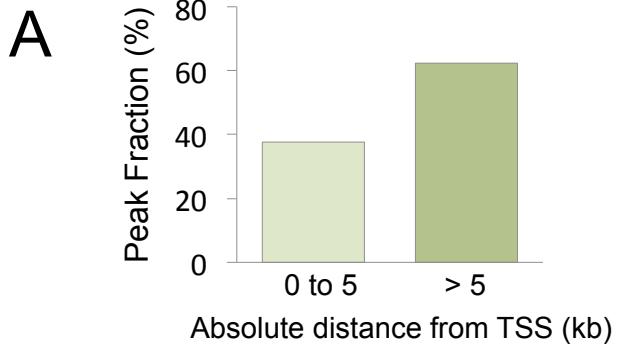


Fig. S11. H3K18ac peaks away from RB1 bound regions are relevant to the effect of e1a on the host cell. 3894 peaks of H3K18ac in e1a-expressing cells were found to be at least 5 kb away from any RB1 peak. **A)** GREAT analysis of peak locations. 60% of the 3894 peaks are located more than 5 kb from the TSS. **B)** Heatmap of log2 ratio between e1a-expressing versus mock-infected cells for the 2533 genes associated with peaks in A). **C-D)** DAVID gene ontology and Seqpos analysis for TFs motifs in over-expressed and repressed genes as in (B).

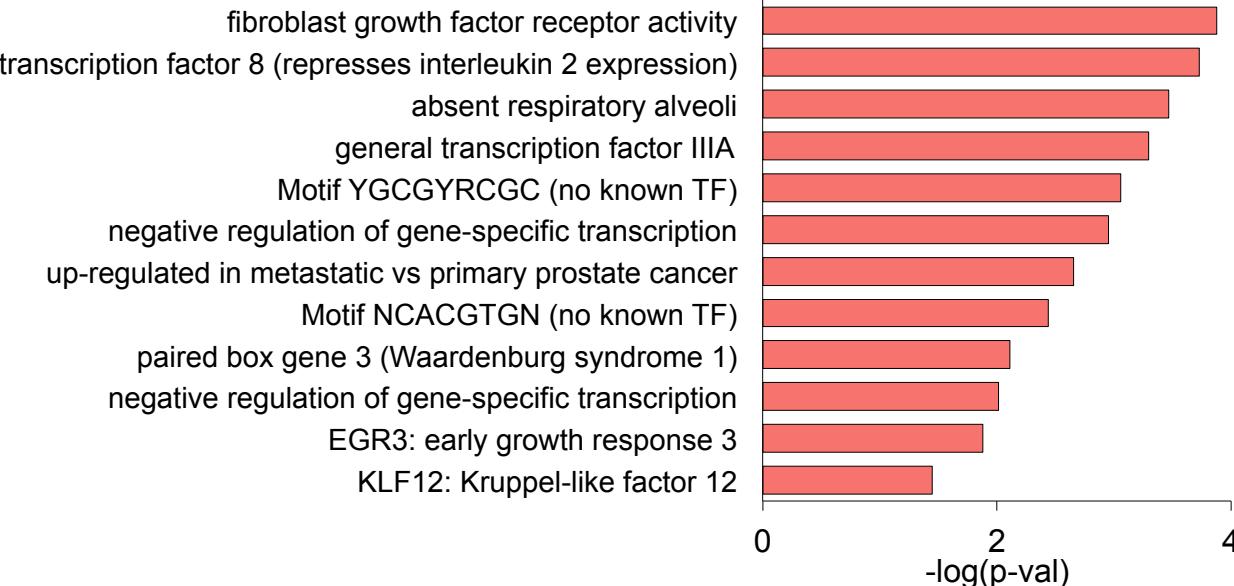
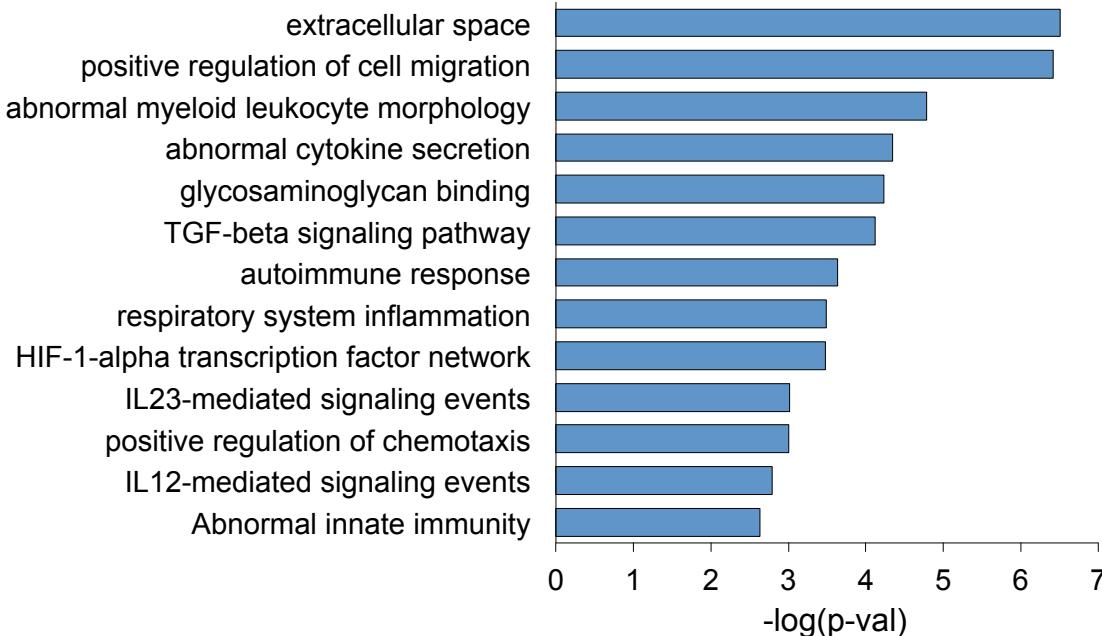
A**e1a-IRs****B****mock-IRs**

Fig. S12. Gene ontology (GO) enrichment analysis of genes associated with peaks of H3K18ac in intergenic regions. GREAT was used to find peak-gene associations for the peaks of H3K18ac in intergenic regions of mock-infected (mock-IRs) and e1a-expressing (e1a-IRs) cells. Bars represent $-\log_{10}$ of the p-value for the selected GO terms. **A)** Cluster e1a-IRs, figure 5A. **B)** Cluster mock-IRs, figure 5A.

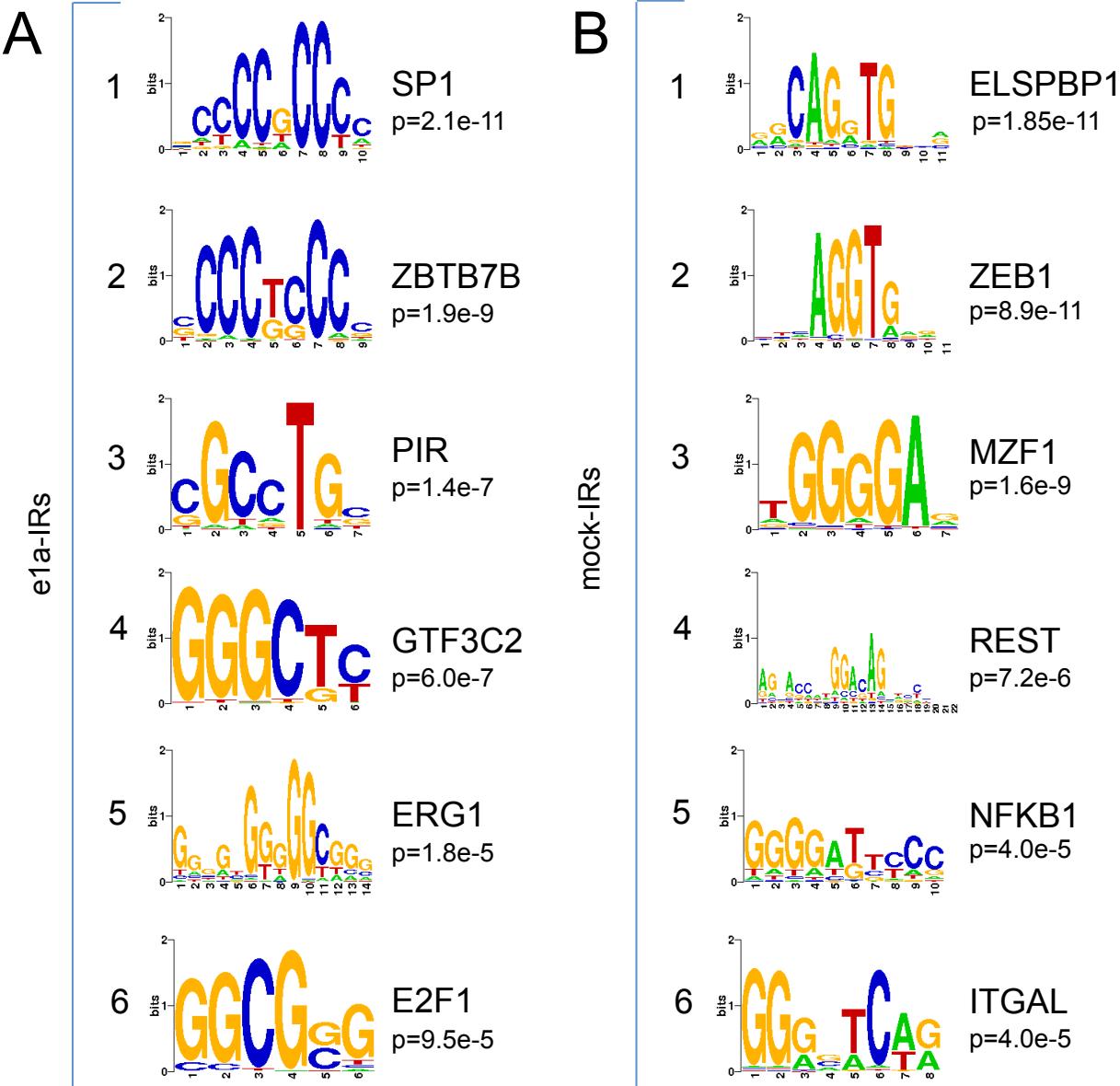


Fig. S13. Sequence logo representation of TFs binding motifs associated with intergenic peaks of H3K18ac. The 600-bp region around the peaks of H3K18ac in the intergenic regions of mock- (mock-IRs) and e1a-expressing (e1a-IRs) cells (Fig. 5A) were analyzed for TFs binding motifs using Seqpos (Galaxy). **A**) Selected significant TFs binding motifs for e1a-IRs. **B**) Selected significant TFs binding motifs for mock-IRs.

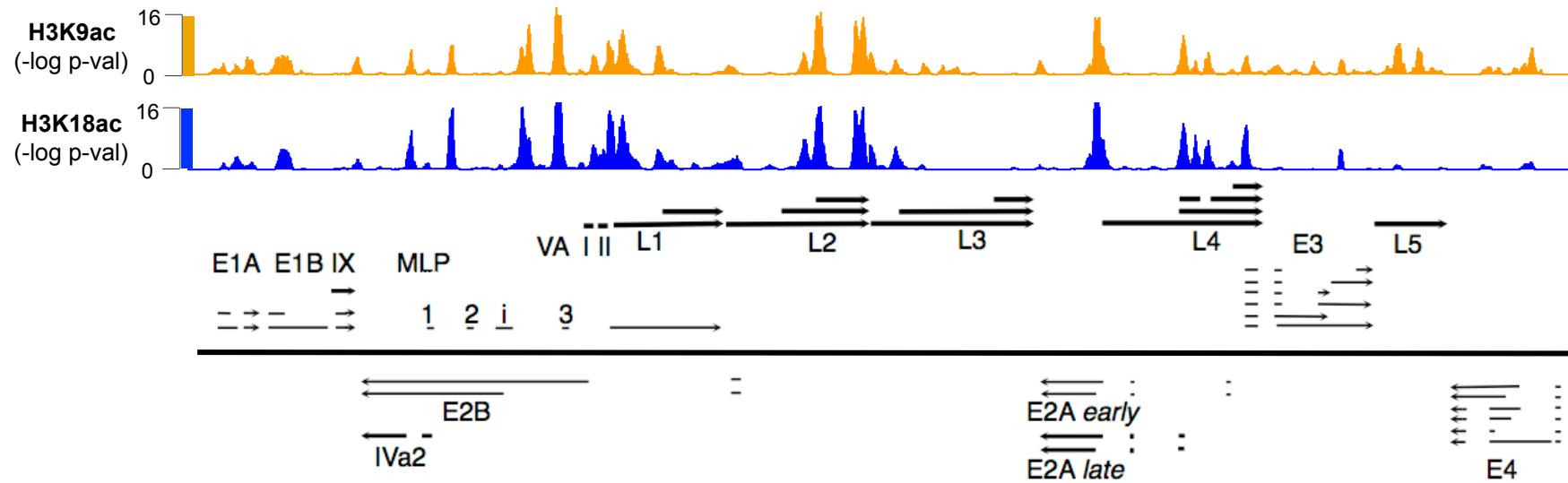


Fig. S14. Histones associated with the adenovirus genome are acetylated. Shown is the $-\log p$ -value of input-normalized tags across the entire adenovirus genome for H3K9ac (yellow track) and H3K18ac (dark blue track). Black arrows represent adenoviral transcripts and direction of the arrow corresponds to direction of transcription.