Title
Dynamics of Host Cell Transcriptome in DNA Tumor Virus Infection

Author
Su, Yu-Chyuan Trent

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Dynamics of Host Cell Transcriptome in DNA Tumor Virus Infection

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Oral Biology

by

Yu-Chyuan T Su

2014

ABSTRACT OF THE DISSERTATION

Dynamics of Host Cell Transcriptome in DNA Tumor Virus Infection

by

Yu-Chyuan T Su

Doctor of Philosophy in Oral Biology

University of California, Los Angeles, 2014

Professor Siavash Kurdistani, Co-Chair

Professor Wenyuan Shi Co-Chair

Adenovirus e1a expression in contact-inhibited cells forces G1 to S phase transition in preparation to create suitable cellular environment for viral replication. To determine genomic regions targeted by e1a to bind and manipulate chromatin landscape of host cells, ChIP-seq of e1a was performed in e1a-expressing primary fibroblast cells at 6 and 24 hours post infection. Critical regions in the genome are occupied by e1a proteins since early in infection including active enhancers, gene body of growth inhibition genes, and cell cycle gene promoters. Binding of e1a at active enhancers results in drastic reduction of H3K18ac and H3K27ac histone modifications. Through co-localization with massive levels of P300/CBP, activators of TGF-beta pathway are strongly repressed with high density of e1a binding throughout gene

promoter and gene body to prevent G1 arrest. Finally, promoters of cell cycle genes bound by Rb in contact-inhibited cells are targeted by e1a at both 6 and 24 hours post infection to relieve Rb repression of E2F-regulated cell cycle genes. Overall, highly-expressed cell-type-specific genes are repressed by e1a binding and genes promoting S phase entry are stimulated by e1a. Surprisingly, e1a binding sites at both 6 and 24 hours PI are also early replication origin found in growing normal fibroblast cells. Furthermore, e1a represses host cell replication as 95% of newly synthesized DNA in e1a expressing cells originated from viral genome. By binding to host cell chromatin, e1a creates S phase environment in host cell to stimulate viral replication.

The dissertation of Yu-Chyuan T Su is approved.

Cun-Yu Wang

Shen Hu

Wenyuan Shi, Committee Co-Chair

Siavash K. Kurdistani, Committee Co-Chair

University of California, Los Angeles

2014

TABLE OF CONTENTS

FIGURES

ACKNOWLEDGEMENTS

Reprints of journal article in Chapter 2 was published in *Cell host and microbes* (2014 Nov;16:663-676). Permission to reproduce this material from the publisher was granted through the Copyright Clearance Center.

Reprints of journal article in Chapter 3 was published in *Epigenetics* (2014 Feb;9(2):257-67). Permission to reproduce this material was granted by publisher for non-commercial use.

Reprints of journal article in Chapter 4 was published in *Molecular Cell* (2013 Jan 24;49(2):310-21). Permission to reproduce this material from the publisher was granted through the Copyright Clearance Center.

Reprints of journal article in Chapter 5 was published in *Genome Research* (2012 Jul;22(7):1212-21). Permission to reproduce this material was granted by publisher for non-commercial use.

Reprints of journal article in Chapter 6 was published in *PLOS One* (2011;6(11):e26894) Permission to reproduce this material was granted by publisher for non-commercial use.

My graduate training started with the Division of Oral Biology and Medicine in School of Dentistry at University of California, Los Angeles (UCLA) where I met friends and colleagues who gave me a lot of help and support: Dr. David Bae, Dr. Supanigar

Ruangsri Sermswatsri, and Dr. Jeffrey Kim. I am also very grateful toward administrative staff in the Oral Biology PhD program, Megan Scott and Pauli Nuttle, for all the help that I received. I also had a great mentor during my graduate rotation, Prof. Shen Hu.

To pursuit my interest in studying the process of oncogenesis using viral oncoprotein e1a as a model system, I received most of my graduate training in the Kurdistani lab (Department of Biological Chemistry at David Geffen School of Medicine at UCLA). I would like also to acknowledge all the help and support that I have received from past and current members of the Kurdistani Lab: Dr. Maria Vogelauer, Dr. Roberto Ferrari, Dr. Bing Li, Dr. Jinyu Li, Dr. Amit Oberai, Dr. Matthew McBrian, Dr. Ben Macadangdang, Dr. Tanya Spektor, Dr. Fang Sheng, Dr. Yong Xue, Dr. Hossein Khaedemian, Kunwu Li, Chen Cheng, Oscar Campos, Narsis Attar, and Maggie. I especially would like to thank the e1a project members in our lab, Dr. Roberto Ferrari and Dr. Bing Li, for their insightful inputs and advices on innovative approaches to study e1a biology. The e1a project is a highly productive collaborative effort with Prof. Arnold Berk's lab. I am also very thankful for all the help and support from members of the Berk lab including Dr. Dawei Gou, Dr. Gauri Jawdekar, Dr. Jordan Parker, Sarah Ann Johnson, Miguel Nava, Robert Zemke, and most importantly Professor Arnold Berk whose knowledge in e1a biology is unparalleled. I also want to give my most sincere gratitude to all members of my committee: Prof. Cun-Yu Wang, Prof. Wenyuan Shi, and Prof. Shen Hu for their kind help and advices during my graduate training. I especially would like to thank my mentor Professor Siavash Kurdistani for his innovative thinking and intellectual guidance throughout my graduate training.

VITA

EDUCATION and EXPERIENCE

2008-present	Graduate Student Researcher Department of Biological Chemistry David Geffen School of Medicine University of California, Los Angeles Principal investigator: Prof. Siavash Kurdistani
2007-2008	Graduate Student Researcher Division of Oral Biology and Medicine School of Dentistry University of California, Los Angeles Principal investigator: Prof. Shen Hu
2005-2007	Programmer Analyst School of Information and Computer Science University of California, Irvine Principal investigator: Prof. Eric Mjolsness
2003-2005	MS, Computer Science and Information Systems School of Information and Computer Science University of California, Irvine Advisor: Prof. Eric Mjolsness
2000-2003	Research Assistant College of Natural Resources, Genomics Facility University of California, Berkeley
2000-2003	Research Assistant Department of Nutritional Science and Toxicology College of Natural Resources University of California, Berkeley Principal investigator: Chris Vulpe lab
1995-2000	BA, Biochemistry and Molecular Biology College of Letters and Science University of California, Berkeley

PUBLICATION

R. Ferrari*****, D. Gou*****, G. Jawdekar*****, S. Johnson*****, M. Nava*****, **T. Su***, A. F. Yousef*****, N. R. Zemke*****, M. Pellegrini, S. K. Kurdistani, A. J. Berk, 'Adenovirus Small E1A Employs the Lysine Acetylases p300/CBP and Tumor Suppressor Rb to Repress Select Host Genes and Promote Productive Virus Infection', *Cell Host & Microbe*, 16 (2014), 663-676.

B. Li, **T. Su**, R. Ferrari, J. Y. Li, and S. K. Kurdistani, 'A Unique Epigenetic Signature Is Associated with Active DNA Replication Loci in Human Embryonic Stem Cells', *Epigenetics*, 9 (2014), 257-67.

Y. Xue, A. A. Vashisht, Y. Tan, **T. Su**, and J. A. Wohlschlegel, 'Prb1 Is Required for Clipping of the Histone H3 N Terminal Tail in Saccharomyces Cerevisiae', *PLoS One*, 9 (2014), e90496.

Y. D. Chai, L. Zhang, Y. Yang, **T. Su**, P. Charugundla, J. Ai, D. Messadi, D. T. Wong, and S. Hu, 'Discovery of Potential Serum Protein Biomarkers for Lymph-Node Metastasis in Oral Cancer', *Head Neck* (2014).

M. A. McBrian, I. S. Behbahan, R. Ferrari, **T. Su**, T. W. Huang, K. Li, C. S. Hong, H. R. Christofk, M. Vogelauer, D. B. Seligson, and S. K. Kurdistani, 'Histone Acetylation Regulates Intracellular Ph', *Mol Cell*, 49 (2013), 310-21.

R. Ferrari, **T. Su**, B. Li, G. Bonora, A. Oberai, Y. Chan, R. Sasidharan, A. J. Berk, M. Pellegrini, and S. K. Kurdistani, 'Reorganization of the Host Epigenome by a Viral Oncogene', *Genome Res*, 22 (2012), 1212-21.

C. Marban*****, **T. Su*,** R. Ferrari, B. Li, D. Vatakis, M. Pellegrini, J. A. Zack, O. Rohr, and S. K. Kurdistani, 'Genome-Wide Binding Map of the Hiv-1 Tat Protein to the Human Genome', *PLoS One*, 6 (2011), e26894.

B. Compani, **T. Su**, I. Chang, J. Cheng, K. H. Shah, T. Whisenant, Y. Dou, A. Bergmann, R. Cheong, B. Wold, L. Bardwell, A. Levchenko, P. Baldi, and E. Mjolsness, 'A Scalable and Integrative System for Pathway Bioinformatics and Systems Biology', *Adv Exp Med Biol*, 680 (2010), 523-34.

H. Chen, Z. K. Attieh, H. Gao, G. Huang, **T. Su**, W. Ke, and C. D. Vulpe, 'Age-Related Changes in Iron Homeostasis in Mouse Ferroxidase Mutants', *Biometals*, 22 (2009), 827-34.

H. Chen, G. Huang**, T. Su**, H. Gao, Z. K. Attieh, A. T. McKie, G. J. Anderson, and C. D. Vulpe, 'Decreased Hephaestin Activity in the Intestine of Copper-Deficient Mice Causes Systemic Iron Deficiency', *J Nutr*, 136 (2006), 1236-41.

H. Chen, Z. K. Attieh, **T. Su**, B. A. Syed, H. Gao, R. M. Alaeddine, T. C. Fox, J. Usta, C. E. Naylor, R. W. Evans, A. T. McKie, G. J. Anderson, and C. D. Vulpe, 'Hephaestin Is a Ferroxidase That Maintains Partial Activity in Sex-Linked Anemia Mice', *Blood*, 103 (2004), 3933-9.

J. M. De Freitas, J. H. Kim, H. Poynton, **T. Su**, H. Wintz, T. Fox, P. Holman, A. Loguinov, S. Keles, M. van der Laan, and C. Vulpe, 'Exploratory and Confirmatory Gene Expression Profiling of Mac1delta', *J Biol Chem*, 279 (2004), 4450-8.

H. Chen, **T. Su**, Z. K. Attieh, T. C. Fox, A. T. McKie, G. J. Anderson, and C. D. Vulpe, 'Systemic Regulation of Hephaestin and Ireg1 Revealed in Studies of Genetic and Nutritional Iron Deficiency', *Blood*, 102 (2003), 1893-9.

Y. M. Kuo, **T. Su**, H. Chen, Z. Attieh, B. A. Syed, A. T. McKie, G. J. Anderson, J. Gitschier, and C. D. Vulpe, 'Mislocalisation of Hephaestin, a Multicopper Ferroxidase Involved in Basolateral Intestinal Iron Transport, in the Sex Linked Anaemia Mouse', *Gut*, 53 (2004), 201-6.

CHAPTER 1

Introduction:

Host cell reprogramming toward S phase entry

by

Adenovirus e1a

Tumorigenic Adenovirus

Human adenovirus type 12 was found capable of inducing tumor formation in Hamsters (Huebner, Rowe et al. 1962). Additionally, the adenoviral protein e1a was shown to collaborate with G12V HRAS oncogene and transform primary rat embryonic fibroblast cells (Land, Parada et al. 1983; Ruley 1983; Ruley 1990). The co-expression of Adenovirus type 5 early region genes E1A and E1B were also able to transform Baby rat kidney (BRK) cells (van den Elsen, de Pater et al. 1982; Branton, Bayley et al. 1985) By observing dynamic changes in molecular processes leading to cellular transformation in normal human cells, adenovirus e1a stands as a very useful tool in enhancing our understanding of tumorigenesis with its capacity to cooperate with either viral or host cell proteins to accomplish cellular transformation and immortalization.

Adenovirus e1a

Adenovirus early region 1A expresses two alternatively spliced mRNAs 12S and 13S which encodes two highly related proteins of 289 and 243 amino acids. Large and small e1a differs by small e1a lacking the CR3 (conserved region 3) region (Figure 1-1) (Berk 2005). Expression of adenovirus small e1a alone is sufficient in forcing stationary contact-inhibited fibroblastic cells into S phase of the cell (Howe, Mymryk et al. 1990; Ghosh and Harter 2003). This oncogenic event is driven by small e1a through its interaction with multiple key cellular proteins in the infected host cell including Rb family of pocket proteins (Nevins 2001; Burkhart and Sage 2008; Choi, Kim et al. 2014), P300/CBP(Gayther, Batley et al. 2000; Berk 2005), and P400 (Dick 2007; Macher-Goeppinger, Bermejo et al. 2013).

Rb Family of Pocket Proteins

Rb is the first tumor suppressor being isolated through study of cancerous retina tissue, Retinoblastoma (Murphree and Benedict 1984). The primary function of Rb is to regulate cell cycle progression from G1 to S phase by binding to the transactivation domain of E2F transcription factors through its pocket domain and inhibit induction of E2F target genes (Goodrich, Wang et al. 1991). E2F transcription factors promote expression of genes required for S phase entry. Therefore, Rb is regarded as the primary protein guarding the G1 to S cell cycle checkpoint. In addition to Rb, RBL1 (p107) and RBL2 (p130) both contain homology in the pocket domain with Rb protein. The pocket domain mediates interaction with members of the E2F family of transcription factors. In normal cell cycle progression from G1 to S phase, Rb family proteins are phosphorylated sequentially by CDK4/6 and CDK2 (Rubin, Gall et al. 2005). Phosphorylated Rb family proteins are then disassociated from E2F binding and relieve repression on E2F activation (Figure 1-2). Activation of E2F transcription factors lead to transcription of cell cycle genes required for S phase entry.

LXCXE Motif

Forcing contact-inhibited primary cells into S phase entry can be accomplished by expressing small e1a in contact-inhibited primary cells. The LXCXE peptide motif located in the CR2 region of small e1a protein enables high affinity binding between e1a and Rb. The LXCXE motif is shared among multiple viral oncoproteins including HPV E7 and SV40 large T antigen (Dahiya, Gavin et al. 2000; Liu and Marmorstein 2007). The LXCXE motif is important in oncogenesis induced by DNA tumor viruses as viral oncoproteins use LXCXE motif to bind Rb family of proteins and disable regulation of cell cycle progression. Small e1a with LXCXE motif is one of the DNA tumor viral

protein equipped of such function. Through interaction with Rb family of proteins, small e1a can bind and remove Rb family of proteins from E2F transcription factors and induce expression of cell cycle genes. The ability to bind Rb is essential for viral replication since small e1a with mutations at CR2 regions are incapable of binding to Rb family of proteins, therefore are unable to displace Rb from Rb-bound E2F promoters to force contact-inhibited normal cells into S phase (Figure 1-3).

The LXCXE motif can also be found in a number of cellular proteins including histone deacetylases HDAC1, HDAC2 (Dahiya, Gavin et al. 2000). Histone proteins H2A, H2B, H3, H4 forms nucleosomes which are protein octamers that allows genomic DNA to wrap around and organize the DNA- bound nucleosomes into chromatin structure. Chromatin can be further rearranged into higher order structures through interaction with linker histones, histone chaperones, and histone modifying enzymes. By removing acetyl group from lysine residues of N-terminus histone tails, HDAC proteins remodel chromatin structure into closed and more condensed conformation which prevents access and binding of transcriptional activator (Kurdistani and Grunstein 2003). Rb interacts with HDAC1 and HDAC2 through binding to the LXCXE motif. Histone deactylase activity of HDAC1 and HDAC2 was shown to be important in Rb mediated repression of transcription (Dahiya, Gavin et al. 2000). Recent studying using microscopy and DNA array in CHO cells (Ferrari, Gou et al. 2014) showed that complex formed between P300, e1a, and RB (P300-e1a-Rb) is able to condense chromatin and the inhibitor of histone deacetylase activity prevented chromatin condensation which revealed that repression mediated by Rb-e1a-P300 required active histone deacetylase activity.

P300/CBP

P300 and CBP are structurally-similar transcription co-activators. P300 was discovered through its binding to adenovirus e1a (Eckner, Ewen et al. 1994). P300/CBP participates in regulation of transcription through its function as a histone acetyltransferase (HAT). By adding acetyl groups to N-terminus tails of histone H3, P300/CBP remodels and relaxes the chromatin structure which permits more binding of activator proteins leading to transcription activation. In MEF cells, Jin Q et al. showed that P300/CBP is the main histone acetyl transferase enzyme responsible for the acetylation of lysine at position 18 (H3K18) and 27 (H3K27) on histone (Jin, Yu et al. 2011). Abundance of H3K18ac and H3K27ac near the transcription start site is associated with open chromatin and active transcription. In addition to its co-activator function at transcriptional start site, P300/CBP also plays essential roles in process of cell differentiation through its requirement for the maintenance of active enhancers.

Role of P300/CBP at enhancers

Enhancers are short stretches of genomic DNA regions which permit binding of activators to induce gene transcription through long range interaction with gene promoters. The interaction can be either intra-chromosomal or inter-chromosomal. Enhancer regions can be defined epigenetically by regions in the genome bound by nucleosomes with presence of histone modification H3K4Me1 and simultaneously lack H3K4Me3 (Heintzman, Hon et al. 2009; Creyghton, Cheng et al. 2010). Enhancer regions can be in either active or poised state. Enhancers in poised state are characterized with binding of nucleosomes with H3K27Me3 and H3K4Me1 histone modifications. Additionally, poised enhancers are incapable of influencing expression of

target genes. Active enhancer regions are open and relaxed regions of chromosome characterized by its sensitivity to DNAse I enzyme digestion which only digests highly accessible regions of the chromatin (Gevry, Hardy et al. 2009). In these highly accessible regions, activator proteins are allowed to bind and interact with mediator complex to promote activation of transcription on. By acetylating H3K18ac and H3K27ac on nucleosomes at enhancer regions, EP300 maintains the accessibility of the enhancer region. Therefore, genes neighboring enhancers bound by nucleosomes with highly acetylated H3K18ac or H3K27ac are associated with higher levels of gene expression (Creyghton, Cheng et al. 2010; Zentner, Tesar et al. 2011).

Interaction between P300/CBP and e1a

In e1a expressing IMR90 cells, e1a causes global deacetylation of H3K18ac (Horwitz, Zhang et al. 2008) and H3K27ac (Ferrari, Gou et al. 2014) by binding to TAZ2 domain of P300/CBP. Regions with high level of H3K18ac and H3K27ac are locate within important functional regions in the genome such as enhancers and gene promoters, e1a induced histone deacetylation in these regions resulted in reduction of gene transcription (Ferrari, Su et al. 2012). Furthermore, complex formed by e1a and P300/CBP interaction was found to mediate strong gene repression and interaction between e1a and P300/CBP at TAZ2 domain is essential in maintaining such repressive complex (Ferrari, Gou et al. 2014). Mutants of e1a unable to bind to TAZ2 domains were less effective in exerting strong gene repression (Ferrari, Gou et al. 2014).

Remodeling of epigenetic landscape at gene promoters by e1a

High-throughput DNA microarray technology in combination with chromatin immune-precipitation (ChIP-on-chip) permitted monitoring of e1a binding and epigenetic

changes at wide range of genomic regions. Focusing on genomic regions 5 kb upstream of transcription start site (TSS) to 2 kb downstream of TSS, the study in Ferrari et al. 2007 unveiled the dynamic binding pattern of small e1a on promoter regions of infected IMR90 cells. Small e1a induced relocation of Rb and EP300/CBP to antiviral response genes and resulted in reduction of antiviral gene expression. Most importantly, epigenetic changes through gaining of H3K18ac at promoter of cell cycle genes after e1a removes Rb from E2F regulated genes was revealed. By gaining activating histone modification, H3K18ac, transcription of cell cycle genes became highly activated (Ferrari, Pellegrini et al. 2008). Recent advances in DNA sequencing technology such as ChIP-sequencing (ChIP-seq) and RNA-sequencing (RNAseq) enabled genome-wide studies to have unprecedented resolution and coverage.

ChIP-sequencing and RNA-seq

Chromatin immunoprecipitaiton inconjunction with cDNA microarray (ChIP on chip) allowed detailed genome-wide studies in species with small genomes. However, ChIP-on-chip is not a cost-effective solution for high resolution mapping of protein-DNA interactions for species with large genomes such as human and mouse. Enabled by recent advances in high-throughput sequencing, genome-wide mapping of protein-DNA interaction at unparalleled resolution is now possible using ChIP-seq technology. Briefly, cells are treated with crosslinking agent to fix protein-DNA complexes. After cell lysis and shearing of genomic DNA into smaller pieces through sonication, antibody targeting protein of interest is used to immune-precipitate (IP) protein-DNA complex of interest. Also an aliquot of the sonicated sample was used as input sample to control for background level of DNA enrichment. The crosslinking between protein and DNA is

then reversed and DNA species in region of interest is isolated. Adapter required for sequencing process is attached to purified DNA through ligation reactions to make high-throughput DNA sequencing library. Sequencing libraries are then amplified with PCR reactions and purified for subsequent sequencing reaction. Each sequencing reaction currently generates around 200 million reads. The generated data was then aligned to human genome using aligner software. Reads that map to multiple locations in the genome are usually discarded to ensure unique identification of genomic location. To determine genomic regions bound by protein of interest, enrichment of protein binding region is determined through comparing reads obtained from input and IP samples using statistical software.

For RNA-sequencing, total RNA is isolated from cells and ribosomal RNAs are isolated and discarded from total RNA. The total RNAs free of ribosomal RNAs are then fragmented through chemical cleavage. The fragmented RNA are then reverse transcribed into cDNA and tagged with sequencing adapters to make DNA sequencing libraries. Sequencing libraries are then amplified with PCR reactions and purified for subsequent sequencing reaction. Data generated from sequencing machines are then aligned to genome with software that permits gap alignment to accommodate spliced RNAs.

Genome-wide study of e1a biology using ChIP-seq and RNA-seq methods

ChIP-on-chip method revealed the gain of H3K18ac histone modification on cell cycles active by H3K18ac and decrease in H3K18ac at gene promoters being repressed by e1a binding (Ferrari, Pellegrini et al. 2008). Since promoter region accounts for small percentage of the genome, global loss of H3K18ac in infected cells

remained partially explained using ChIP-on-chip method. Study in dl1500 infected IMR90 cells by Ferrari et al. 2012 (Ferrari, Su et al. 2012) performed ChIP-seq on H3K18ac and Rb family of proteins (Rb, p107, p130) and mRNA-seq was used to measure global changes in gene expression. The study found that majority of genomic regions enriched with H3K18ac histone modifications are located within enhancer. Additionally, e1a expression greatly reduced H3K18ac levels in enhancer regions to decrease expression of its neighboring genes composed of mostly fibroblast-specific genes. Reduction of cell-type specific gene expression implied de-differentiation and loss of cell-specific function (Frisch and Mymryk 2002). ChIP-seq of Rb family members precisely defined binding positions of Rb, p107, and p130 on gene promoters and promoters of cell cycle genes bound by Rb was found to gain H3K18ac at 24 hours after infection.

Through ChIP-seq and RNAseq method, recent study on e1a mediated gene repression through formation of P300-e1a-Rb complex in infected cells revealed a cluster of TGF-β activator genes being bound by the repressive complex throughout the promoter and gene body (Ferrari, Gou et al. 2014). Mutants of e1a incapacitated with ability to bind to either P300 or Rb were generated. Global expression profiles of cells infected with different e1a mutants were measured using RNA-seq and different gene clusters requiring different interactions between e1a and host proteins were generated as P300 binding mutant of e1a was unable to effectively repress TGF-β activator genes. ChIP-seq of H3K27ac revealed differential regulation of H3K18ac and H3K27ac in infected cells. Even though P300/CBP is responsible for acetylation of both H3K18ac and H3K27ac, while both modifications decreased dramatically in infected cells,

differential regulation of H3K18ac and H3K27ac levels were observed in infected cells. H3K27ac was abundant at most promoters in mock infected cells including cell cycle genes. At 24 hours PI, H3K27ac decreased greatly at most gene promoters except for cell cycle genes. In contrast, H3K18ac levels are low in most promoters especially at cell cycle genes, but promoter of cell cycle genes gained H3K18ac at 24 hours PI. ChIP-seq and RNA-seq methods enabled detailed observations on dynamic changes incurred by e1a in cellular chromatin environment of host cells.

Rb and DNA replication

ChIP-seq study of Rb binding in growing and senescent IMR90 cells identified unique role of Rb in binding and repressing replication genes in senescent condition induced by elevated levels of ras (Chicas, Wang et al. 2010). In the study, DNA replication was detected cells with RNAi knockdown of Rb but not p107 or p130 protein. DNA replication requires assembling of pre-Replication Complex (pre-RC). Pre-replication complex is composed of Orc (1-6), CDC6, CDT1, and MCM (2-7) proteins and the complex is bound at replication origins (Mechali 2010). Pre-RC components were expressed and found on chromatin only in Rb-knockdown senescent cells. This study suggests Rb binding in senescent cells regulates expression of genes required in DNA replication. Study by Mendoza-Maldonado et al. (Mendoza-Maldonado, Paolinelli et al. 2010) showed a more direct relationship between Rb and pre-RC, as direct binding between hypo-phosphorylated Rb and Orc1 was observed and E2F1 competes binding to Rb with Orc1 at replication origins.

Myc and e1a

Another protein recently found to bind at replication origin is Myc (Dominguez-Sola, Ying et al. 2007). Elevated level of Myc protein was found to cause increased activity at replication origins. However, Myc levels are usually low in normal cells but are elevated in cancer cells. Interestingly, e1a was found to stabilize Myc through binding with P400 (Tworkowski, Chakraborty et al. 2008). P400 was originally isolated through binding with e1a as P400 binding mutant of e1a was unable to mediate cell transformation (Fuchs, Gerber et al. 2001) However, transformation defect of e1a P400 mutant was able to be overcame by over-expressing Myc suggesting Myc regulates the downstream effectors of e1a-P400 interactions. While e1a was shown to interact with Myc through P400, the relationship between e1a and Myc interaction on chromatin remains unclear.

P400 and e1a

Interaction between P400 and N-terminus of e1a is required for e1a induced cell transformation. A part of SWI/SNF (SWItch/Sucrose NonFermentable) chromatin remodeling family of proteins, P400 functions by exchange histone variant H2A.Z into nucleosomes to activate transcription (Gevry, Hardy et al. 2009) or facilitate DNA repair (Xu, Ayrapetov et al. 2012). Additionally, function of P400 is essential in keeping stem identity as P400 was shown to repressing developmental genes in stem cells (Fazzio, Huff et al. 2008). With multitude of functional roles being discovered for P400, the molecular changes exerted by interaction between P400 and e1a on host cell chromatin to induce oncogenic transformation is still unclear.

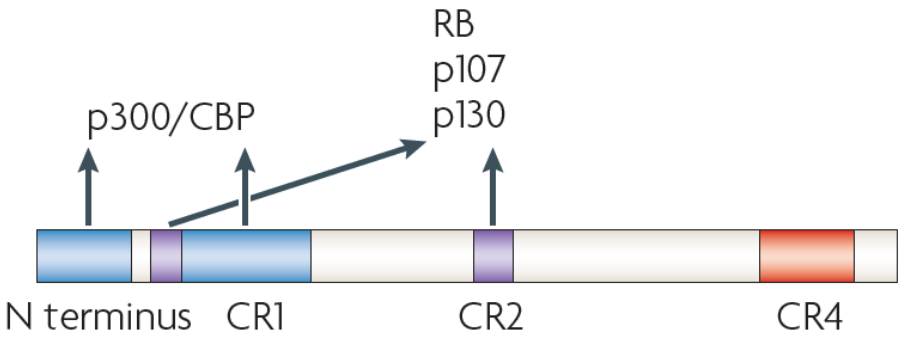


Figure 1-1 Schematic structure of small e1a and its interactions with cellular proteins required for S phase entry. Compare to large E1A, splicing isoform, small E1a lacks Conserved Region 3 (CR3). Small e1a binds to P300/CBP through its N terminus and CR1 region. Binding to RB family of proteins (RB, p107, p130) requires CR2 region and parts of CR1 region. Figure from (Ferrari, Berk et al. 2009)

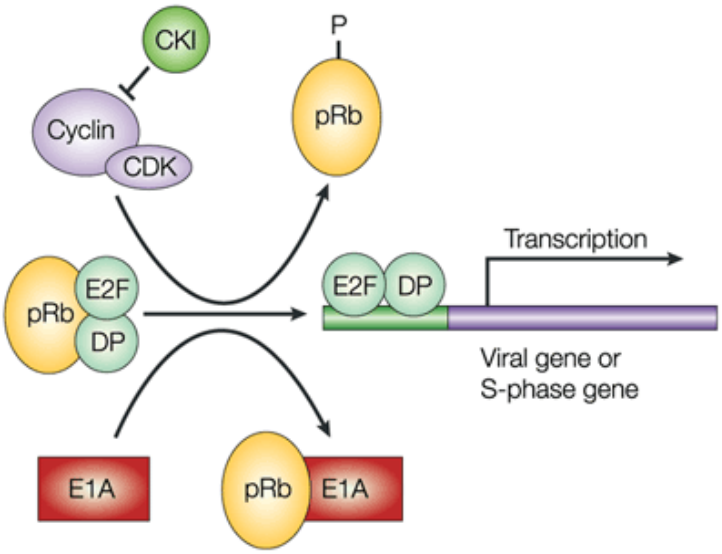
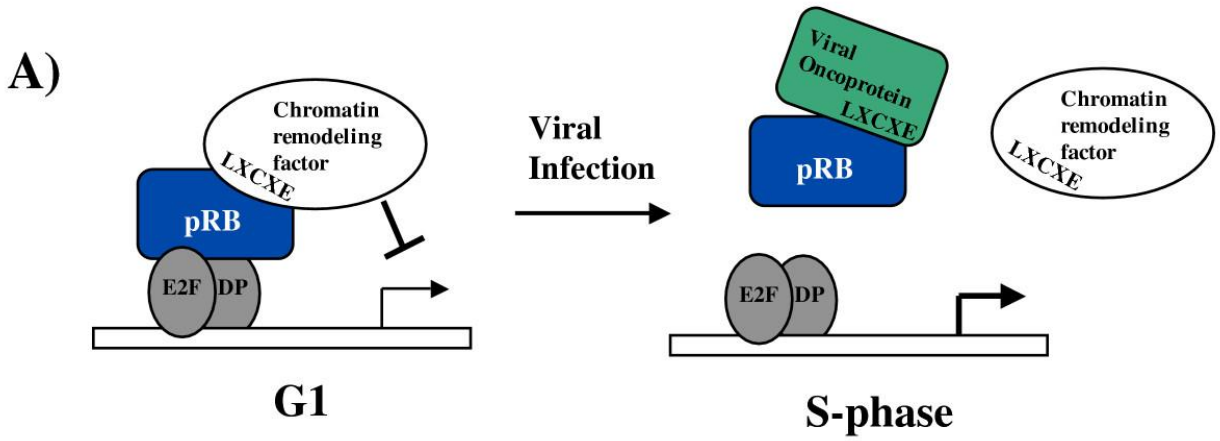


Figure 1-2 Rb binding in normal versus viral induced S phase.

Rb in normal S phase gets phosphorylated by CDKs and relieve from binding to E2F. Rb in e1a expressing cells are bound and sequestered by e1a to induce activation of E2F genes. Figure from (Frisch and Mymryk 2002)



B)

		700	715	752	764
H. sapiens	pRB	LDQIMMCSMYGICKVK		IIVFYNSVFMQRL	
H. sapiens	p107	LDQLLLCAFYIMAKVT		LIKFYNTIYVGRV	
H. sapiens	p130	LDQLLMCAIYVMAKVT		LIQFYNNIYIKQI	
Z. mays	RRB1	IDQLILCCLYGVAKVC		IITFYNEVFVPAA	
D. melanogaster	RBF	LDQNIMCAIYIYIRVK		IIHFYNHTYVPLM	
C. elegans	LIN-35	LDQILLCCVFVIMKIN		IIKYYNIEFRDRI	
C. reinhardtii	Mat3p	IDQIMLSTLYGYCKVH		IIGFYNAVFVPAM	
		* *		* **	

Figure 1-3 LXCXE motif in Rb pocket family of proteins

In repressive state, Rb interacts with histone deacetylases such as HDAC1 and HDAC2 and binds at cell cycle gene promoters to repress transcription. In viral infected cells, Rb is seized by viral oncoproteins such as e1a to allow host cell entry into S phase. Figure from (Dick 2007)

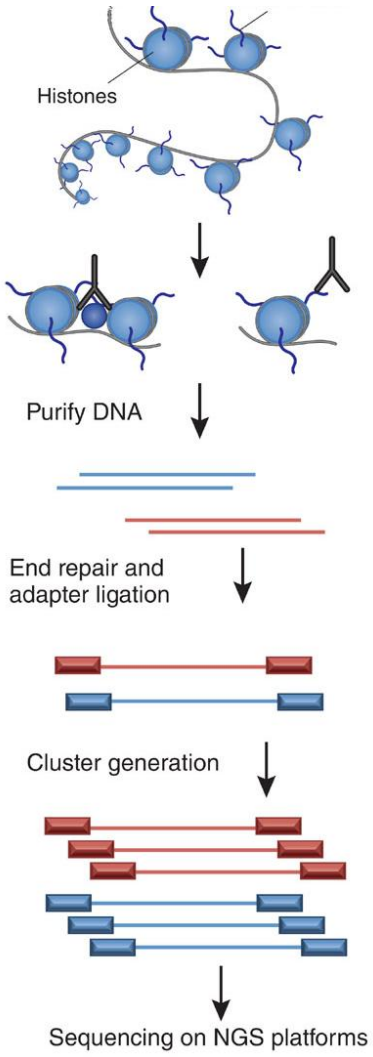


Figure 1-4 Process of ChIP-Sequencing

Proteins-DNA complexes are crosslinked and sonicated to smaller pieces. Protein-DNA complex of interest is isolated by immunoprecipitation using target-specific antibody. DNA from protein-bound region is purified and tagged with sequencing adapters to make DNA sequencing library for high-throughput sequencing. Figure from (Kidder, Hu et al. 2011)

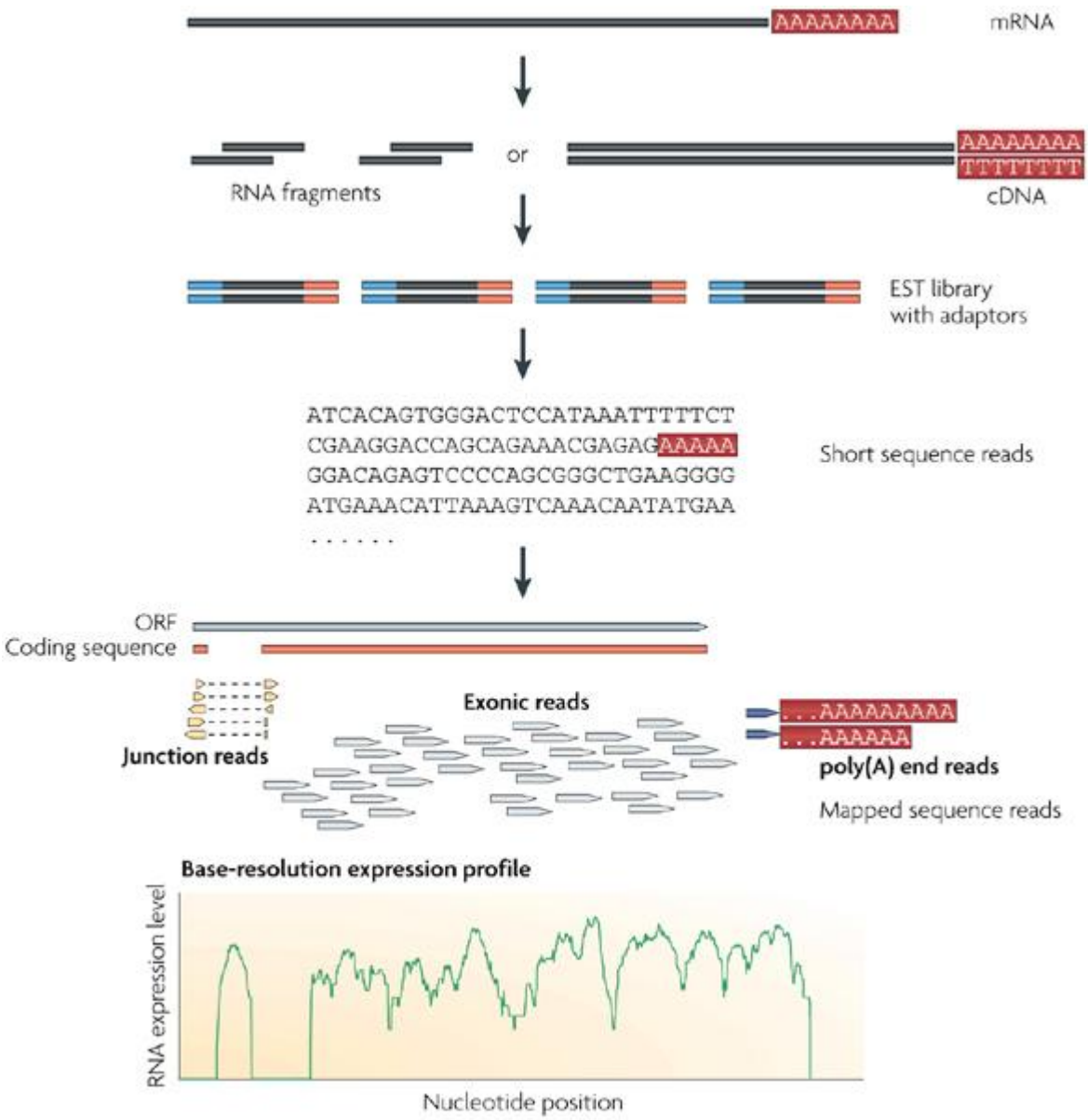


Figure 1-5 Process of RNA-Sequencing

RNA species free of ribosomal RNA (mRNA or total RNA) are fragmented and converted into cDNA. Sequencing adapters are ligated cDNA fragments to make sequencing libraries. After amplification and purification, sequencing libraries are sent for high-throughput sequencing. Figure from (Wang, Gerstein et al. 2009)

Chapter 2

Adenovirus small e1a employs the lysine acetylases p300/CBP and tumor suppressor Rb to repress select host genes and promote productive virus infection

This chapter was originally published in *Cell Host and Microbe*, 2014 Nov;16:663-676 with supplemental material. Mutants of e1a that are unable to bind to Rb or P300 was generated and mRNAseq was performed in contact-inhibited IMR90 cells infected with various e1a mutants. Through comparing differential gene expression between IMR90 cells infected with different e1a mutants, gene clusters requiring e1a-p300 or e1a-Rb interaction for modulation of gene expression was revealed. Furthermore, by mapping genome-wide distribution of Rb, P300, H3K27ac, H3K4Me1 using ChIP-seq method, host cell gene repression requiring simultaneous binding of e1a to both Rb and p300 was revealed. The p300-e1a-Rb complex was shown acetylate Rb and condense DNA-arrays in microscopy through HDAC activity. This study revealed that simultaneous interaction between Rb, e1a, and P300 is required for strong repression of host cell gene transcription.

Cell Host & Microbe
Article

Adenovirus Small E1A Employs the Lysine Acetylases p300/CBP and Tumor Suppressor Rb to Repress Select Host Genes and Promote Productive Virus Infection

Roberto Ferrari,[1,7] Dawei Gou,[2,3,7] Gauri Jawdekar,[2,7] Sarah A. Johnson,[2,7] Miguel Nava,[3,7] Trent Su,[4,7] Ahmed F. Yousef,[2,7,8] Nathan R. Zemke,[2,7] Matteo Pellegrini,[1,2,5] Siavash K. Kurdistani,[1,2,4,6] and Arnold J. Berk[2,3,*]
[1]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
[2]Molecular Biology Institute
[3]Department of Microbiology, Immunology and Molecular Genetics
[4]Department of Biological Chemistry
[5]Department of Molecular, Cellular, and Developmental Biology
[6]Department of Pathology and Laboratory of Medicine
UCLA David Geffen School of Medicine, Los Angeles, CA 90095-1570, USA
[7]Co-first authors
[8]Present address: Department of Chemical and Environmental Engineering, Masdar Institute of Science and Technology, Abu Dhabi, UAE
*Correspondence: berk@mbi.ucla.edu
http://dx.doi.org/10.1016/j.chom.2014.10.004

SUMMARY

Oncogenic transformation by adenovirus small e1a depends on simultaneous interactions with the host lysine acetylases p300/CBP and the tumor suppressor RB. How these interactions influence cellular gene expression remains unclear. We find that e1a displaces RBs from E2F transcription factors and promotes p300 acetylation of RB1 K873/K874 to lock it into a repressing conformation that interacts with repressive chromatin-modifying enzymes. These repressing p300-e1a-RB1 complexes specifically interact with host genes that have unusually high p300 association within the gene body. The TGFβ-, TNF-, and interleukin-signaling pathway components are enriched among such p300-targeted genes. The p300-e1a-RB1 complex condenses chromatin in a manner dependent on HDAC activity, p300 lysine acetylase activity, the p300 bromodomain, and RB K873/K874 and e1a K239 acetylation to repress host genes that would otherwise inhibit productive virus infection. Thus, adenovirus employs e1a to repress host genes that interfere with viral replication.

INTRODUCTION

Adenovirus (Ad) E1A is a classic DNA virus oncogene (Weinberg, 2013). When expressed alone, small E1A (hereafter called "e1a") (Figure 1A) causes G_1-arrested cells to enter S phase (Ghosh and Harter, 2003). In cooperation with Ad E1B (Branton et al., 1985) or G12V *HRAS* (Ruley, 1983), e1a stably transforms rodent cells. Two interactions with host cell proteins are essential for e1a-induced cell transformation in cooperation with G12V *HRAS*: an interaction with RB family proteins (RB1, RBL1 [p107], and RBL2 [p130], hereafter referred to as "RBs") and an interaction with the closely related nuclear lysine acetylases (KATs) p300 and CBP (Pelka et al., 2008). (Hereafter we refer to both p300 and CBP simply as "P300.").

E2F transcription factors regulate genes required to enter S phase (Dick and Rubin, 2013). In G_1, G_0 end-differentiated (Chong et al., 2009), and senescent cells (Chicas et al., 2010), RB proteins bind to E2F activation domains (ADs) (Lee et al., 2002; Xiao et al., 2003) repressing E2F-regulated genes by both masking the AD and by inducing repressive chromatin structure through interactions with chromatin-modifying enzymes (Dick and Rubin, 2013). In cycling cells, RBs are phosphorylated by cyclin D-CDK4/6 and cyclin E/A-CDK2, causing them to change conformation and dissociate from E2Fs and repressing chromatin-modifying complexes (Dick and Rubin, 2013), derepressing hundreds of genes required to enter S phase. e1a derepresses these same genes by directly displacing unphosphorylated RBs from E2Fs (Bagchi et al., 1990; Fattaey et al., 1993; Ikeda and Nevins, 1993), which explains why the e1a-RB interaction promotes entry into S phase. In contrast, the explanation for why e1a must bind P300 to transform cells is less clear.

Interestingly, an e1a mutant defective for binding RBs and a second e1a mutant defective for binding P300 do not complement for transformation, suggesting that a single e1a molecule must interact with both P300 and RBs to transform cells (Wang et al., 1995). Although there is negative cooperativity in the binding of both RB1 and a CBP TAZ2 domain to the same e1a molecule (Ferreon et al., 2013), such trimeric complexes form in vivo (Wang et al., 1995) and in vitro (Ferreon et al., 2013). In such complexes, e1a promotes acetylation of RB1 K873/K874 by P300, inhibiting binding of cyclin-CDKs to RB1 and hence RB1 phosphorylation during the cell cycle (Chan et al., 2001). P300 also acetylates e1a K239, inhibiting binding of importin-α3 to a NLS at the e1a C terminus (Madison et al., 2002).

How do the e1a-RB, e1a-P300 interactions and RB-e1a-P300 complexes influence cellular gene expression? To address these questions, we constructed Ad vectors for wild-type (WT) e1a and

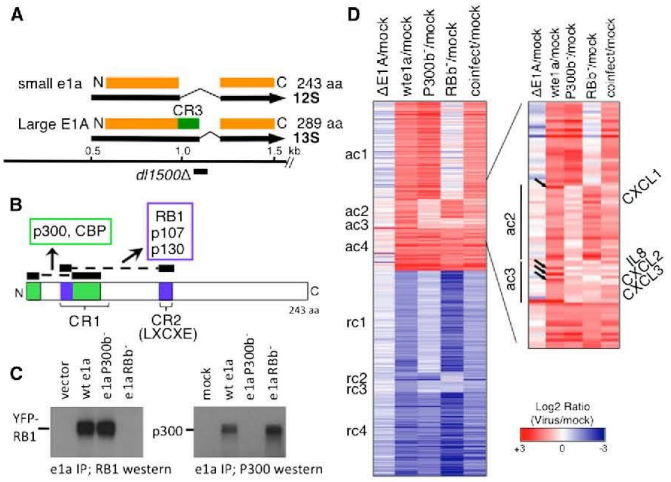


Figure 1. e1a-Regulated Host Cell Gene Expression

(A) Major E1A early mRNAs and proteins and 9 bp dl1500 deletion.

(B) Regions of e1a that bind to RBs (blue) and p300/CBP (green).

(C) e1a mutant interactions with RB1 and P300. Extracts of HeLa cells transfected with expression vectors for YFP-RB1 (used in Figure 7) and the indicated e1a mutants (left) or infected with Ad vectors for the indicated e1a mutants (right) were immunoprecipited with anti-e1a mAb M73 and immunobloted with anti-RB1 (left) or anti-p300 (right) antibody.

(D) Heat map of RNA increased (red) or decreased (blue) compared to mock-infected cells. See also Figure S1 and Tables S1, S2, S3, and S4.

mutants completely defective for interactions with the RBs or P300. We performed RNA sequencing (RNA-seq) when infected cells were entering S phase. Mechanisms underlying the observed changes in expression were explored by chromatin immunoprecipitation sequencing (ChIP-seq) of RNA polymerase II (pol2), p300, RB1, and posttranslationally modified histones. We found that RB1 is enriched at E2F sites of genes activated by the e1a-RB interaction, much more than p130 or p107. This may help to explain why *RB1*, and not *RBL1* or *RBL2*, is a tumor suppressor. Further, the results suggest why the e1a-P300 interaction is required for transformation. Importantly, e1a did not completely inhibit histone acetylation by P300. Instead, e1a regulated P300 HAT activity differently at different promoters and enhancers. We discovered an unexpected mechanism of e1a repression by targeting hypophosphorylated RB1 and p300 to the gene bodies of repressed genes. Fluorescently tagged proteins allowed direct visualization of e1a-driven chromatin condensation by a p300-e1a-RB1 complex, dependent on p300 KAT activity, the p300 bromodomain, and acetylation of RB1 and e1a. Our data suggest that e1a exploits the RBs displaced from E2Fs, locked into a repressing conformation by P300 acetylation, to repress host cell genes in pathways that would otherwise inhibit viral replication. Further, our data may explain why primate Ads express small e1a as well as large E1A.

RESULTS

Mutants Defective for Binding Either RBs or P300

Small e1a binds RBs through two interactions: one with the N-terminal ~10 aa of conserved region 1 (CR1) and one through the LXCXE region of CR2 (Lee et al., 1998; Liu and Marmorstein, 2007) (Figure 1B, blue). To completely eliminate e1a binding to RBs, we deleted CR2 (e1a aa 112–128) and mutated L43, L46, and Y47 to A, mutations that individually reduce the affinity of e1a CR1 for the RB1-pocket domain by 10-fold or more (Liu and Marmorstein, 2007). To eliminate the e1a-P300 TAZ2 (CH3) domain interaction, we constructed a multisite mutant based on the high-resolution structure of the e1a-CBP TAZ2 complex (Ferreon et al., 2009): R2G, E59A, V62A, F66A, and E68A. These mutations eliminate several e1a-TAZ2 electrostatic and hydrophobic interactions and mutate the N-terminal region that binds to the other side of TAZ2 relative to e1a residues 53–83 in CR1 (Figure 1B). We call the mutants e1aRB-binding minus (e1aRBb) and e1a P300-binding minus (e1aP300b) because they fail to coimmunoprecipitate RBs or p300/CBP, respectively, from extracts of transfected HeLa cells or infected IMR90 cells but bind the alternative interacting protein comparably to WT e1a (Figure 1C; Figures 1A–S1G available online). These mutants were incorporated into Ad5-vectors in the dl1500 background with a deletion of the unique E1A 13S mRNA splice site (Montell et al., 1984). Since large E1A is primarily responsible for activating other viral promoters (Montell et al., 1984; Winberg and Shenk, 1984), these vectors express only very low levels of the other viral early regions compared to WT Ad5. The vector expressing WT e1a drove contact-inhibited primary IMR90 fibroblasts into S phase ~20 hr postinfection (p.i.) (Figures S1H and S1I).

e1a-Regulated Cellular mRNA Expression

Contact-inhibited IMR90 cells arrested in G_1 were mock infected; infected with dl312, an Ad5 mutant with a deletion of E1A (Jones and Shenk, 1979), with the Ad vectors expressing WT or mutant e1a; or coinfected with the e1aRBb and e1aP300b vectors. e1aRBb accumulated to lower level than e1a WT or e1aP300b when the vectors were infected at the same multiplicity of infection (MOI). Consequently, the e1aRBb vector was used at 4-fold higher MOI to achieve nearly equal levels of WT and mutant e1a's (Figure S1J). RNA-seq was performed with two biological replicates at 24 hr p.i. (Table S1). Genes with a difference in expression between mock- and WT e1a vector-infected cells of 2-fold or more and $p < 0.01$ for WT e1a between the two experiments are shown in Figure 1D and Table S2. Genes were clustered according to whether the change in expression required the e1a-RB, e1a-P300, both, or neither interaction. Expression of a representative gene and boxplots of expression levels for each cluster are shown in Figures S1K and S1L.

The e1a-RB interaction controlled expression of the largest cluster of activated genes (Figure 1D, ac1). The ac1 gene ontology is overwhelmingly enriched for S phase genes (Figure 2A; Table S3), amply confirming the generalization that most genes required for S phase are regulated by E2F activators repressed in G_1 and G_0 by RB-proteins. Detailed studies of the time course of changes in host gene expression following infection with WT Ad2 and closely related Ad5 have been performed in G1-arrested IMR90 cells and primary human foreskin fibroblasts using microarrays (Miller et al., 2007; Zhao et al., 2003), as well as RNA-seq at 12 and 24 hr p.i. (Zhao et al., 2012). The most highly induced genes had gene ontologies of DNA replication and cell cycle. These authors suggested this was due to E1A displacement of RBs from E2Fs. Our data show that this is indeed the case. Figure S2A compares genes regulated by WT Ad2 at 24 hr p.i to our data with the WT e1a vector. While there was considerable overlap in genes induced/repressed by the WT e1a vector compared to Ad2, larger numbers of genes were induced/repressed by Ad2. This is probably because of expression of all viral genes in Ad2-infected cells by 24 hr p.i. Also, we used very stringent criteria for classifying genes as induced/repressed ($p < 0.01$ in duplicate experiments) to maximize the opportunity of detecting similar trends in histone modifications and pol2, p300, and RB1 association among genes in the individual clusters.

ChIP-seq for pol2 (Figure 2B) showed that at ac1 promoters in mock-infected, G_1-arrested cells there were on average small peaks of pol2 at ~+50 to +100 and ~−100 to −200, presumably due to pol2 that initiated transcription in the sense and antisense directions and then paused (Core et al., 2008; Seila et al., 2008). In dl1500-infected cells expressing e1a, there was a large increase in the pol2 peak for sense strand transcription, but not for antisense transcription. This e1a-induced increase in sense-oriented pol2 near the TSS strongly suggests that ac1 mRNAs increase because of increased transcription.

We anticipated that ac1 genes would have a peak of E2Fs near the TSS. Consequently we analyzed available ChIP-seq data for E2Fs 1 and 4 from HeLa cells (Bernstein et al., 2012), because we expected that genes required for S phase would be activated by E2Fs similarly in different human cell types. For example, similar genes are activated in RB1-deficient fibroblasts and pituitary and thyroid tumor cells (Black et al., 2003). Indeed, peaks of E2F1 and E2F4 were observed well above the average for all genes at ac1 promoters (Figure 2C) and not at genes in the other e1a-activated (Figure S2B) or repressed clusters (data not shown). Analysis of our ChIP-seq data for all three RBs in arrested IMR90 cells (Ferrari et al., 2012) showed peaks at ac1 promoters coincident with the E2F peaks (Figures 2C, 2D, and S3). Comparing the significance of peak heights relative to all ChIP data across the genome showed that RB1 association with ac1 promoters was significantly greater than for RBL2 (p130) or RBL1 (p107). Although each of the RB-family members was immunoprecipitated with a different specific antibody, the average ChIP-seq p value across all promoters was similar for each RB (Figure S2C). The much larger signal for RB1 at ac1 genes suggests that RB1 is the predominant RB family member at ac1 promoters. Examples include *CCNE2* (cyclin E), the critical regulator of S phase entry, *MCM2*, and *MCM3* (Figure 2E). e1a expression following infection with dl1500 decreased the average signal for RB1 at ac1 promoters more than 2-fold compared to cells infected with the E1A mutant (Figures 2E and 2F), demonstrating that e1a displaces RB1 from E2Fs in vivo, as it does in vitro.

The average peak of p300 at ac1 TSSs doubled in response to e1a (Figures 2G and S3). H3K18 and H3K27 are acetylated primarily by P300 (Horwitz et al., 2008; Jin et al., 2011). Surprisingly, although ac1 genes were repressed by RBs in the G_1-arrested cells, H3K27 was acetylated to a significant extent at ac1 promoters (Figure 2H, 3A, and S3). The average H3K27ac downstream peak at ac1 promoters was slightly reduced by e1a, while the upstream peak fell considerably (Figure 2H). This differs from the profile in asynchronous IMR90 cells with ~50% of cells in S phase, where the upstream H3K27ac peak was higher (Hawkins et al., 2010) (Figure S2S). In contrast to ac1 genes, e1a decreased H3K27ac at most other promoters (Figures S2K, S2L, and S4C), including promoters of the other e1a-activated clusters (Figure S4B), intergenic regions, and introns, resulting in extensive global H3K27 deacetylation (Figures 3B–3E), even though there was little change in the sharp peaks of p300 association in intergenic regions (e.g., Figure S5).

H3K18 is the other histone tail lysine acetylated primarily by P300. In contrast to H3K27, H3K18ac was low at ac1 promoters in G_1-arrested mock-infected cells and increased greatly in response to e1a, primarily in the downstream direction (Figures 2I, 3A, and S3). Again, this was in contrast to asynchronous IMR90 cells where the upstream H3K18ac peak was higher (Figure S2T). Consequently, acetylation of the two histone tail substrates for P300, H3K18 and H3K27, was regulated differently by e1a at the activated promoters, whereas both H3K18ac and H3K27ac decreased dramatically at repressed promoters, intergenic regions, and introns (Figures 3B–3E, S2M, S2N, S4, and S5).

Like H3K27, H3K9 was acetylated at ac1 promoters in the G_1-arrested mock-infected cells (Figure 2J). H3K9ac did not change significantly when the ac1 promoters were derepressed by e1a displacement of RB1. In contrast to H3K27ac and H3K18ac, e1a did not appreciably alter H3K9 at promoters of the other activated gene clusters (Figure S4A) or most intergenic regions (Figures 3B, 3C, and S4D). Similarly, H3K4me1 changes were modest in response to e1a (Figures 2K, S3, and S5).

RNA from ac4 genes increased 2-fold or more in response to WT e1a and both of the e1a mutants (Figures 1D and S1L). These genes may be regulated by e1a interactions with other host proteins besides RBs or P300 (Pelka et al., 2008). In this regard, it is interesting that while E2F binding motifs were highly enriched in ac1 promoters, as expected, other TF binding motifs were enriched in other clusters (Table S4). While E2Fs do not appear to be the major activators for clusters ac2–ac4, they may contribute to activation of some genes in these clusters, accounting for the small reduction in RNA in cells expressing e1aRBb compared to WT e1a (Figure 1D, ac4), the small peaks of E2F association at the TSS in the average E2F profiles (Figure S2B), and the detection of E2F sites with less significant p values at ac3 and ac4 promoters (Table S4).

e1a Regulation of mRNA Stability

In contrast to ac1 and ac4 genes, e1a did not greatly increase pol2 or modify chromatin at ac2 and ac3 genes (Figures 4A–4E), even though their RNAs increased by >2. This suggests

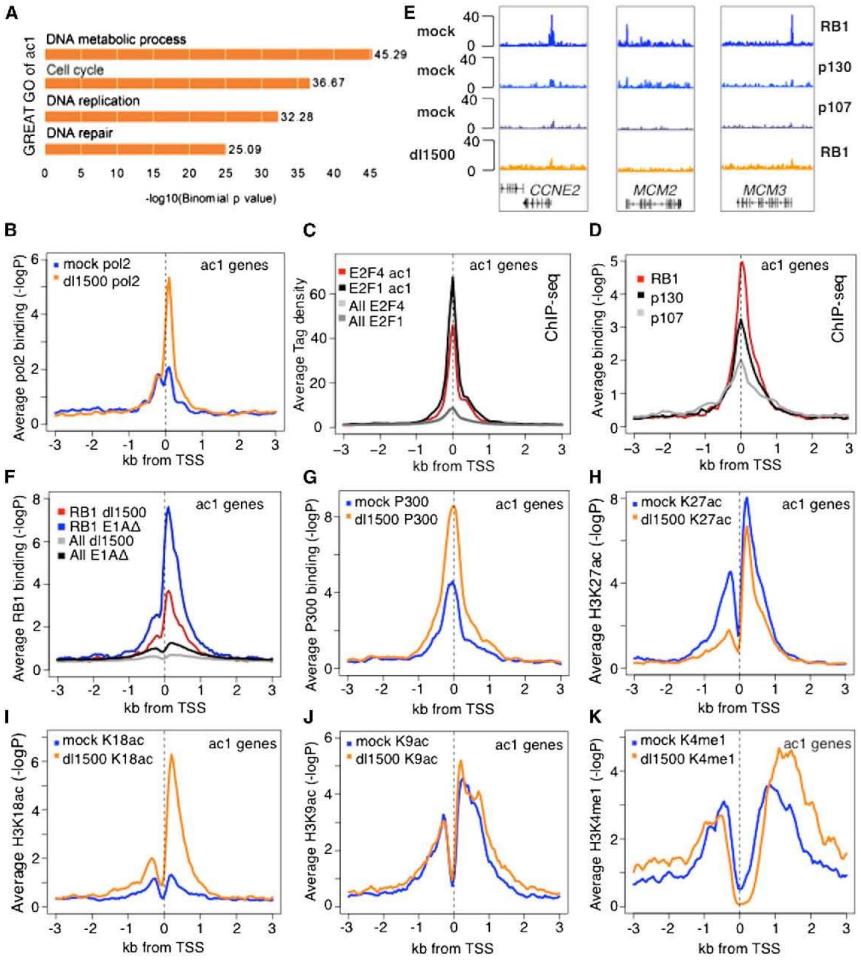


Figure 2. Gene Ontology and ChIP-Seq Data for ac1 Genes

(A) Gene ontology determined by GREAT.

(B–D and F–K) Plots of average $-\log_{10}$ poissonP or tag density relative to TSS for the indicated proteins for ac1 genes in mock and dl1500-infected cells.

(E) Genome browser maps of ChIP-seq data (seq tags) for RBs for three ac1 genes.

that the increased RNA results from posttranscriptional regulation. Indeed, the stability of mRNA from *CXCL1* in ac2 and *CXCL2* in ac3 encoding small cytokines increased dramatically in cells expressing e1a, as indicated by Actinomycin D chase experiments (Figures 4F and 4G). *CXCL1* was assigned to ac2 because e1aRBb induced it slightly more than 2-fold, but its activation by WT e1a and the e1a mutants was very similar to that of *CXCL2*, *CXCL3*, and *IL8*, another CXCL cytokine in cluster ac3 (Figure S2E). Remarkably, induction of these cytokines was not observed after infection with the e1aRBb or e1aP300b vectors or by coinfection of the two vectors (Figures 1D, right, and S2E). As discussed above, this implies that induction of these cytokines requires RB and P300 binding to the same e1a molecule.

e1a-Mediated Repression

e1a significantly repressed slightly more host genes than it activated (Figure 1D). The largest cluster of repressed genes (rc1) required the e1a-P300 interaction for repression. rc1 genes are enriched for secreted glycoproteins comprising the extracellular matrix and were expressed at higher level than any of the other e1a-regulated clusters (Table S3; Figure S1L). Since production

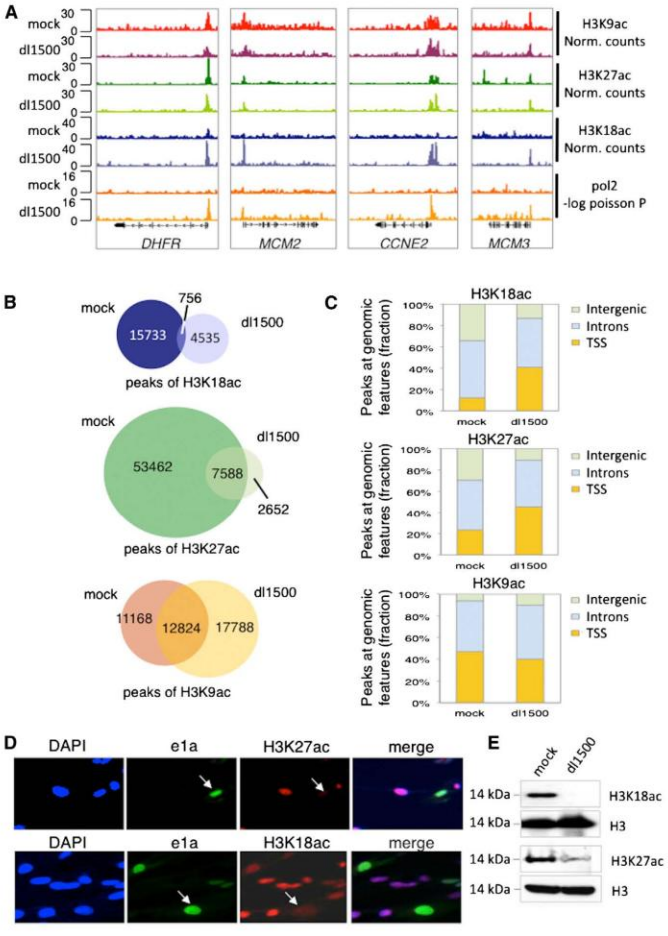


Figure 3. ChIP-Seq for ac1 Genes and Global H3K27 Hypoacetylation in Response to e1a

(A) Gene browser plots of H3K18ac, H3K27ac, H3K9ac, and pol2.

(B) Venn diagrams of total significant peaks of H3K27ac in mock- and dl1500-infected cells and, for comparison, for H3K18ac and H3K9ac (Ferrari et al., 2012).

(C) Fraction of significant peaks shown in (B)1 at TSSs (3 kb from TSSs), intergenic regions (>3 kb upstream of TSSs and >3 kb downstream of polyA sites), and introns in mock- and dl1500-infected cells.

(D) Immunostaining of H3K27ac and H3K18ac in cells infected at MOI = 0.5.

(E) Western blots of total cell H3K27ac, H3K18ac, and total H3.

of extracellular matrix is a major function of fibroblasts, repression of these genes contributes to the dedifferentiation of cells induced by e1a (Pelka et al., 2008). This repression of abundantly expressed cellular genes may make more of the host cell RNA and protein synthesis capacity available for expression of viral genes. e1a decreased pol2 at the TSS and throughout the transcribed regions of rc1, rc3, and rc4 genes and at the TSS of rc2 genes (Figure 5A). Consequently, e1a repression is largely the result of decreased transcription. p300 association with the repressed promoters was either unchanged or increased (Figure 5B), although H3K18ac and H3K27ac, as well as H3K9ac, were reduced by e1a expression (Figure S4A–S4C).

The pol2 ChIP-seq data revealed that the number of genes classified as repressed by WT e1a on the basis of the RNA-seq data is probably an underestimate of the number of transcriptionally repressed genes. For example, for the most abundantly expressed gene in IMR90 cells, *COL1A1*, pol2 association with the TSS and transcribed region in dl1500-infected cells fell to less than half the level in mock-infected cells, whereas *COL1A1* RNA fell less than 2-fold (Figure S6A). This is probably because *COL1A1* mRNA is relatively stable, and consequently, the fall in *COL1A1* RNA occurs more slowly following inhibition of transcription than for less stable mRNAs. To better estimate the number of repressed and activated genes, we summed the pol2 ChIP-seq signal from the annotated TSS to the TTS from mock- and dl1500-infected cells. Of 24,507 annotated human genes, 3,944 (16%) had a drop in total pol2 ChIP-seq signal across the gene to ≤ 0.5 the mock-infected level. A total of 1,874 (7.6%) had an increase in total pol2.

e1a Control of Genes Induced in Response to Ad Infection

RNA-seq of cells infected with the E1A deletion mutant dl312 revealed cellular genes induced or repressed by Ad infection in the absence of e1a (Figure 5C; Table S5). e1a prevented induction/repression of the majority, but not all, of these genes. The e1a-regulated genes were clustered according to whether e1aP300b$^-$ or e1aRBb$^-$ also activated or repressed expression $\geq$2-fold. Of the genes whose induction was inhibited by e1a dependent on the e1a-P300 interaction (ka1), *CDKN1A* encoding the p21^{CIP} inhibitor of cyclin-CDKs was the most abundantly expressed. ka1 genes also include *SESN2*, an mTOR inhibitor (Figures S6B and S6C).

Most of the cellular genes repressed by infection with the E1A mutant are in cluster kr1 (Figure 5C). e1a prevents repression of these genes via the e1a-RB interaction. The gene ontology of kr1 genes is overwhelmingly related to the cell cycle ($p = 1.4 \times 10^{-34}$). These genes probably are repressed by infection with the E1A mutant, because a DNA damage response is activated by the termini of the viral DNA in the absence of E1B and E4 functions (Weitzman and Ornelles, 2005).

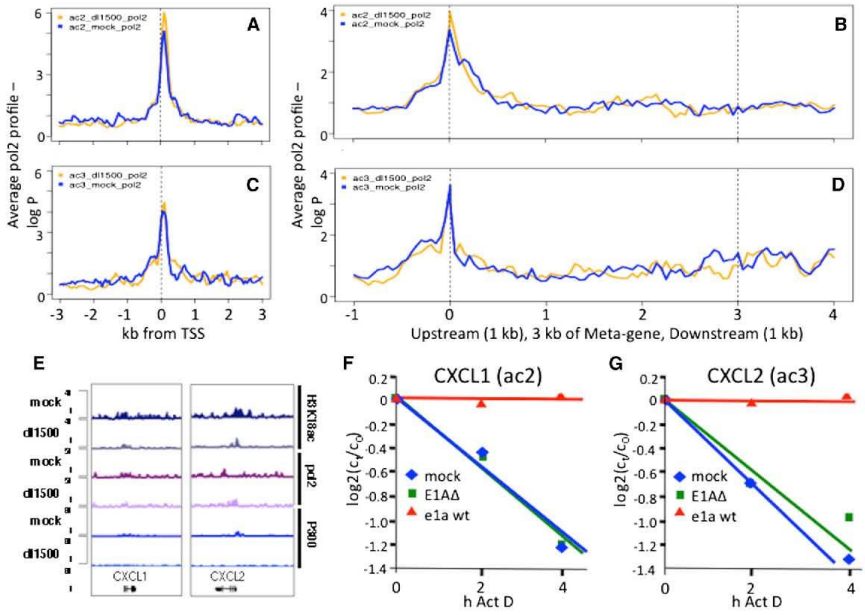


Figure 4. e1a-Induction of *CXCL1* and *2* by mRNA Stabilization

(A–D) Average pol2 ChIP-seq signal from mock- (blue) and dl1500-infected (gold) ac2 ([A] and [B]) and ac3 ([C] and [D]) genes.

(E) Gene browser plots of *CXCL1/2*.

(F and G) Plots of RNA concentration ($\log2(c_t/c_0)$) versus time of exposure to actinomycin D in cells infected with the E1AΔ mutant, the WT e1a vector, and mock-infected cells.

High p300 and RB1 throughout the Transcribed Region of Repressed Genes

Repression of the largest cluster of e1a-repressed genes (rc1) required the e1a-P300 interaction (Figure 1D) and was associated with hypoacetylation of H3K18 and H3K27 at their promoters and associated intronic and intergenic regions (Figures 3B, 3C, S4B, S4C, and S5). These results suggest that e1a inhibits P300 HAT activity at repressed promoters and intergenic regions. But since not all genes are repressed, what distinguishes repressed genes?

In analyzing p300 association with the e1a-repressed genes, we noted that the average p300-association within their transcribed regions was higher than for a group of genes expressed at similar level but whose expression was not altered by e1a (Figure 6A). This surprising association of p300 throughout the transcribed region of repressed genes, in contrast to the sharp peaks of p300 in most other regions of the genome (e.g., Figure S5), was particularly striking for 76 genes in TRAIL, TNF, TGFβ, and interleukin-signaling pathways that might otherwise inhibit fibroblast cell cycling and viral replication (Figures 6B and 6E; Table S6). e1a induced greater p300 association with these genes at the promoter and a few kb upstream, and even more so throughout the transcribed region (Figure 6B). RB1 also was observed at the promoter and spread throughout the bodies of these genes, and also increased in response to e1a (Figure 6C). At the same time, e1a substantially diminished pol2 association with these genes (Figure 6D). Examples of repressed genes in this group include *THBS1* and *CTGF*, important activators of TGFβ signaling (Figure 6F). The e1aP300b and e1aRBb mutants failed to cause the increase in RB1 association (Figure S6B). These results indicate that e1a represses genes that have high P300 association within the gene body before infection and that for genes such as *THBS1*, *CTGF*, *CYR61*, *KLF6*, and *KLF10* (Table S6) in the TGFβ signaling pathway, repression is associated with increased p300 and RB1 throughout the transcribed region.

This high level of p300 and RB1 association with the transcribed region was most obvious for the 76 genes in Table S6. However, repression of all the genes in the large rc1 cluster was significantly greater in cells infected with the e1a WT vector than in cells coinfected with the e1aP300b and e1aRBb vectors (Figure 6G), even though the level of e1a proteins was similar in both groups of cells (Figure S1J). This result suggests that, while repression of rc1 genes required primarily the e1a-P300 TAZ2 domain interaction, the e1a-RB interaction contributed to the full extent of repression by WT e1a. Genes in cluster rc3 required e1a interaction with both P300 and RBs for >2-fold repression (Figures 1D and S1L), including proapoptotic *IFIT1* and *IFIT2*, repressed much less by e1aP300b plus e1aRBb than by WT e1a (Figures S6D and S6E).

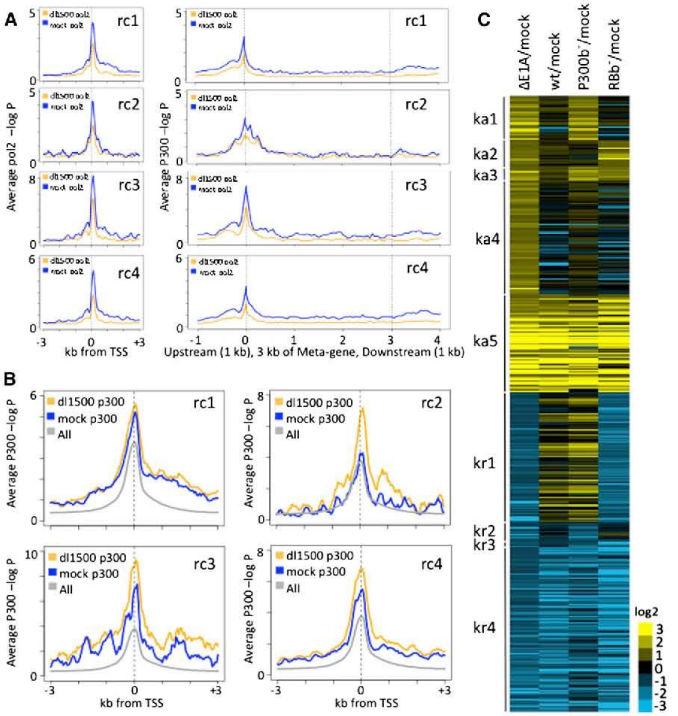


Figure 5. e1a Repression and Inhibition of Activation and Repression Induced by an E1AΔ Mutant

(A) Plots of average pol2 ChIP-seq signal relative to TSS and Meta-gene plots from TSS to TTS for e1a-repressed clusters.

(B) Average p300 ChIP-seq $-\log_{10}$ poisson P for e1a-repressed clusters.

(C) Heat map of genes activated (yellow) or repressed (blue) by E1AΔ-infection, clustered by whether the e1a-P300 (ka1), e1a-RB (ka2), both (ka3), or neither (ka4) interactions were required to inhibit induction 2-fold; whether induction was not inhibited by e1a (ka5); whether the e1a interaction with RB (kr1), P300 (kr2), or both (kr3) were required to inhibit repression by E1AΔ; or whether repression was not blocked by e1a (kr4). See also Table S5.

Chromatin Condensation by a P300-e1a-RB Complex, Dependent on P300 Acetylation of RB1 and e1a

Earlier, using ChIP-chip, we observed that cross-linking of total histone H3 increased in response to e1a at promoter regions of repressed genes with increased p300 and RB1/p130 association (Ferrari et al., 2008). We therefore asked whether e1a causes chromatin condensation, accounting for the increased H3 cross-linking. To assay chromatin condensation, we used a microscopic method that allowed direct observation of chromatin condensation by HP1 (Verschure et al., 2005). The method utilizes CHO cells (RRE.1) engineered to contain ~10^4 lac operators amplified over a region of ~1 Mb (*lacO* array) in one chromosome. When LacI-mCherry with an N-terminal SV40 NLS (NLM) was expressed in these cells, the *lacO* array was visualized by confocal fluorescence microscopy spread through ~5%–10% of the nuclear volume (Figure 7A). When e1a fused to NLS-LacI-mCherry (e1a-NLM) was expressed, the *lacO* array condensed into a much smaller volume in the confocal slice with the largest area of red fluorescence (Figure 7A). Array areas measured in ~100 cells had a mean value in cells expressing e1a-NLM ~1/2 that of the area in cells expressing NLM, the difference being highly significant ($p < 0.0001$) (Figure 7D). However, no difference was observed when e1aRBb or e1aP300b mutants were fused to NLM, when e1aRBb -NLM and e1aP300b -NLM were coexpressed, or when e1a was not fused to LacI (Figures 7D and S7A–S7C). Thus, WT e1a binding to the *lacO* array caused ~4-fold condensation in volume, dependent on simultaneous e1a interactions with both P300 and RBs. Further, no condensation was observed when e1a with eight alanines substituted into the e1a p400 binding region from 25–36 (Fuchs et al., 2001) (e1aP400b) was fused to NLM (Figure S7D).

We found that HDAC activity also was required for chromatin condensation, as treatment of transfected cells with the HDAC inhibitor trichostatin A for 4 hr before fixation reversed e1a-NLM condensation (Figure 7E). Interestingly, a large E1A-NLM fusion failed to condense the *lacO* array (Figure S7D). This may be because E1A CR3 (Figure 1A) associates with several HATs contributing to CR3-dependent activation (Ablack et al., 2012; Pelka et al., 2009a, 2009b). This may reverse histone hypoacetylation required for e1a-induced condensation.

E1A induces P300 acetylation of RB1 at K873 and K874 near the RB1 C terminus by targeting p300 KAT activity to these lysines in a p300-E1A-RB1 complex (Chan et al., 2001). Acetylation at these sites in RB1 inhibits RB1 phosphorylation by cyclin-CDKs, inhibiting progression into S phase. Consistent with this, we observed coelution of p300, e1a, and RB1 by gel filtration of nuclear extract from e1a-expressing 293 cells, but not from HeLa cells lacking e1a (Figure 7B). Also, in IMR90 cells stimulated to enter S phase by infection with dl1500, RB1 phosphorylation at several cyclin-CDK target sites was inhibited, despite induction of cyclin E (Figure S1K) and high CDK2 kinase activity in extracts of the same cells (Figure 7C). To determine if p300 KAT activity required to acetylate these RB1 sites is required for chromatin condensation, we overexpressed a p300 AT2 mutant having greatly attenuated KAT activity (Kraus et al., 1999). This prevented condensation by e1a-NLM but did not alter the size of the larger array observed with NLM alone (Figure 7F). Interestingly, p300 with a bromodomain deletion also interfered with e1a-induced chromatin condensation (Figure 7F). In addition, expression of RB1 with arginine substitutions at

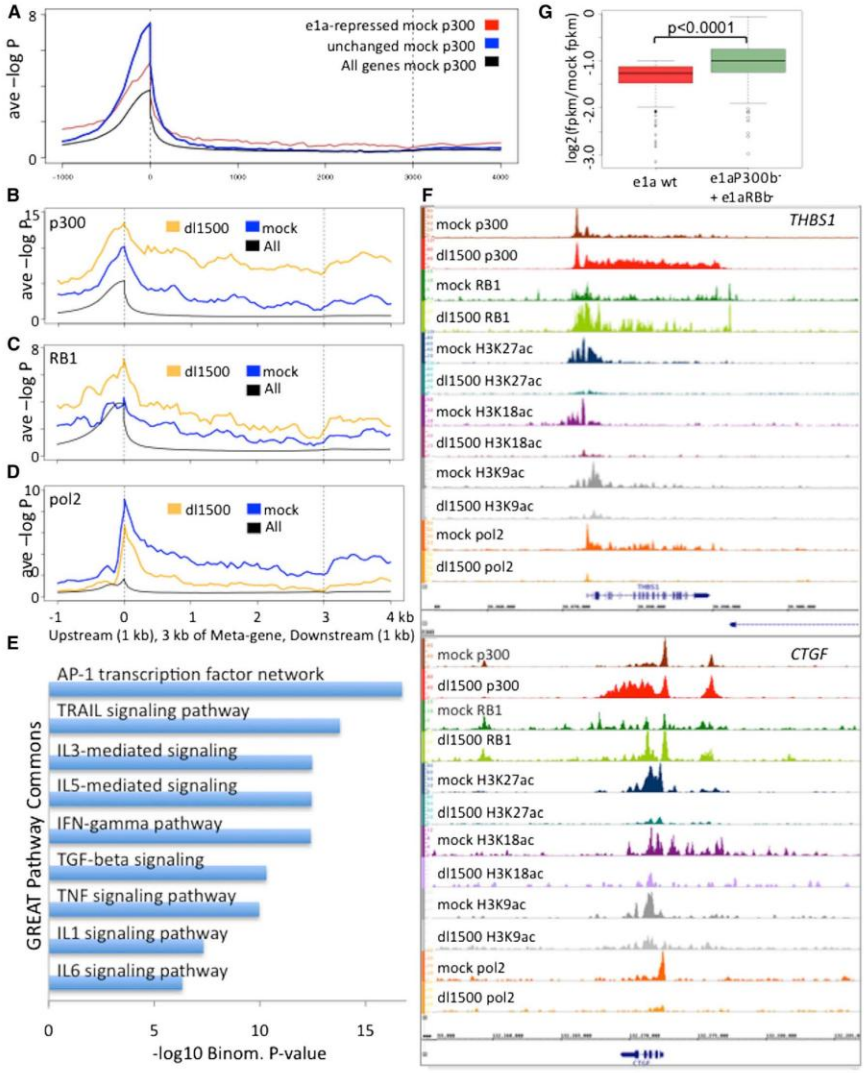


Figure 6. Repressed Genes with p300 and RB1 throughout the Gene Body

(A) Average p300 ChIP-seq signal for rc1 genes from mock-infected cells in Meta-gene plots (red), the control group (blue), and the average of all annotated genes (black).

(B–D) Average ChIP-seq data for 76 genes (Table S6) that had high p300 ChIP-seq signal in the gene body in mock-infected cells. Meta-Gene plots of ChIP-seq data for p300 (B), RB1 (C), and pol2 (D).

(E) GREAT gene ontology probabilities categorized as pathways for this cluster of 76 genes.

(F) Examples of Gene Browser plots of ChIP-seq data from mock- and dl1500-infected cells for *THBS1* and *CTGF*.

(G) Boxplots of $\log_2$ fold repression of rc1 genes from cells infected with the WT e1a vector (red) or coinfected with the e1aRBb$^-$ and e1aP300b$^-$ vectors (green). See also Table S6.

In these and subsequent box plots, the horizontal dark line is the median of the distribution; the box includes 50% of the data; the whiskers include 75% of the data. The p value for the significance of the difference between two distributions indicated by brackets was calculated using one-way ANOVA and a Tukey's HSD post hoc comparison.

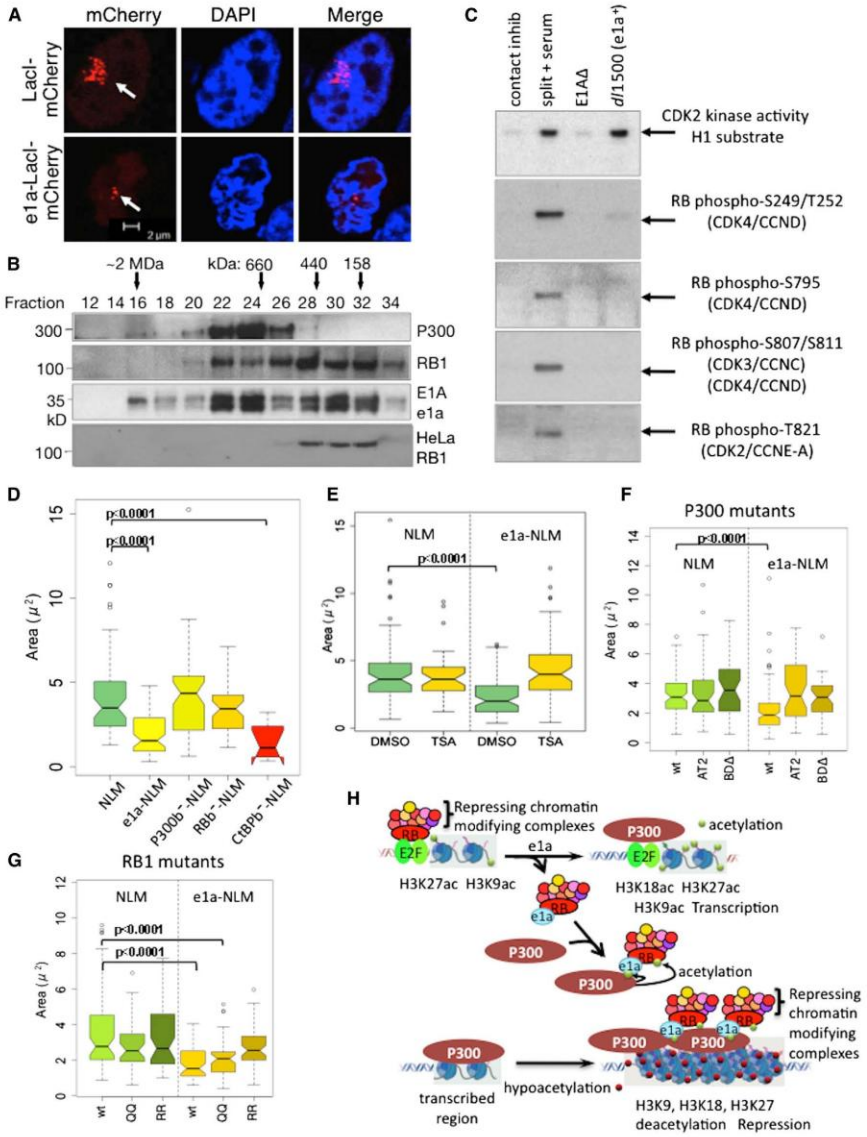


Figure 7. e1a-Induced Chromatin Condensation

(A) Confocal micrographs of mCherry fluorescence in RRE.1 cells transfected with vectors for NLS-LacI-mCherry (NLM) and e1a-NLS-LacI-mCherry (e1a-NLM).

(B) Western blots of Superose 6 column fractions of 293 or HeLa nuclear extract.

(C) Top panel: autoradiogram of gel for CDK2 kinase activity. Lower panels: immunoblots for the indicated phospho-RB1 sites. Lanes from left to right were from mock-infected cells, cells that were split 1 to 3 into fresh media with 20% FBS 24 hr earlier, and cells infected with the E1AΔ mutant or dl1500.

(D–G) Boxplots of mCherry-fluorescent areas in the confocal slice with the largest area.

(H) Model for e1a regulation of host cell gene activation and repression through interactions with RBs and P300. See Discussion and Figure S7.

K873/874 also inhibited condensation, but not expression of RB1 K873Q/K874Q, chemical mimics of acetylated lysines (Figure 7G).

P300 acetylates e1a at K239 (Zhang et al., 2000; Madison et al., 2002). e1a K239R also prevented condensation, as well as greatly reducing colocalization of YFP-P300 with e1a-NLM at the array, while mutant K239Q did not (Figures S7E–S7G). e1a K239 acetylation was reported to inhibit binding of corepressors CtBP1/2 to the e1a C terminus (Zhang et al., 2000), although contradictory results have been reported (Madison et al., 2002). In either case, CtBP binding is not required for condensation, since mutation of the e1a PXDLS consensus CtBP binding site (233–237) to ALAAA did not inhibit array condensation (Figure 7D). We conclude that chromatin condensation by a complex of P300/p400-e1a-RB1 requires acetylation of RB1 K873/K874 and e1a K239 by p300 and the p300 bromodomain.

DISCUSSION

This analysis of host cell gene expression in response to Ad small e1a and e1a mutants reveal how the e1a interactions with RBs and the TAZ2 (CH3) domain of p300/CBP (P300) regulate host cell gene expression. For the most part, the two interactions regulate genes independently. Genes in the largest group of activated genes, ac1 (Figure 1D), are activated by the e1a-RB interactions independently of e1a-P300 interactions. The e1a-P300 TAZ2 domain interaction is responsible for most e1a host cell gene repression (Figure 1D, rc1). In addition, e1a activation and repression of a smaller number of host genes requires e1a interactions with both RBs and P300. Many of these are not properly regulated by a combination of e1aRBb and e1aP300b , suggesting that e1a regulation of this class of genes requires formation of a trimeric P300-e1a-RB complex (Wang et al., 1995). Also, our mutants may possibly affect one of the many other e1a-host protein interactions reported (Pelka et al., 2008).

e1a interactions with RB1 and CBP are examples of binding-induced folding of an intrinsically unstructured polypeptide (Ferreon et al., 2009, 2013). Nonetheless, there is great complexity in the spatially and chemically complementary interactions at the e1a-RB and e1a-TAZ2 interfaces. Because of the specificity of these interactions, it is likely that genes activated by WT e1a but not by e1aRBb (ac1 genes, Table S2) comprise the cellular genes regulated principally by E2F-RB complexes in G_1-arrested primary IMR90 cells. RB1 is the primary RB-family member binding to E2Fs at promoters of these genes. This helps to explain why RB1 is a tumor suppressor, and not p130 or p107. Further, Chicas et al. (2010) found that induction of senescence is more dependent on RB1 than p130 or p107, likely contributing to its unique function as a tumor suppressor.

e1a Regulates Histone Acetylation by P300 Differently at Different Loci

e1a was reported to inhibit the histone acetylase (HAT) activity of P300 in vitro (Chakravarti et al., 1999). But this is due to competitive inhibition by e1a, which is also a substrate, as opposed to allosteric regulation (Madison et al., 2002). Our results indicate that in the cell, e1a regulation of P300 is subtler than simply inhibiting their HAT activities indiscriminately. H3K18 and H3K27 are acetylated principally by P300 in vivo (Horwitz et al., 2008; Jin et al., 2011). At ac1 promoters, e1a induced hyperacetylation of H3K18, while acetylation of H3K27 was unexpectedly high at ac1 promoters in the G_1-arrested cells and remained almost unchanged (Figures 2H and 2J). At ac2, ac3, and ac4 promoters, e1a increased H3K18ac but decreased H3K27ac (Figures S4B and S4C). Finally, e1a caused extensive H3K18 and H3K27 hypoacetylation at most other sites in the genome, including intergenic regions and introns that contain enhancers and at e1a-repressed promoters (Figures 3B, 3D, S5B, and S5C). In e1a-expressing cells, the total number of significant peaks of H3K27ac across the genome fell to only ~17% the level in uninfected cells. H3K18ac also fell in intergenic and intronic regions (Ferrari et al., 2012). How e1a might inhibit P300 HAT activity at some locations, but activate it at others, is presently unclear. We note that although H3K27ac is a mark of active enhancers (Creyghton et al., 2010), RNA- and pol2 ChIP-seq data indicate that for the majority of genes transcription was reduced <2-fold by this extensive reduction in H3K27ac in intergenic and intronic regions where enhancers reside.

Promoter H3K18ac Is Linked to Transcriptional Activation

It is remarkable that transcription of e1a-regulated genes correlated best with H3K18ac at promoters and not with H3K9ac, H3K27ac, or H3K4me1 (Figure S4). Surprisingly, promoter proximal nucleosomes of E2F-RB-regulated genes (cluster ac1) were highly acetylated on H3K9 and H3K27 in the contact-inhibited cells where these promoters are repressed by RB1. These promoters may be in a chromatin structure that is poised for activation in primary fibroblasts, since these genes must be activated rapidly during one of the most important fibroblast functions—wound healing—which requires prompt fibroblast replication. What mechanism would couple final H3K18ac to transcription? TRIM33, a coactivator of SMAD TFs, has a PHD-bromodomain cassette that binds H3K18ac through its bromodomain and a neighboring H3K9me3 through its PHD domain, displacing HP1 to derepress TGFβ target genes (Xi et al., 2011). H3K9 is trimethylated at RB-repressed *Cdc6* and *CcnA* promoters before e1a activation in serum-depleted NIH 3T3 cells (Ghosh and Harter, 2003; Sha et al., 2010), so TRIM33 may bind H3K18ac in these cells. Further studies will be required to identify what interacts with H3K18ac at e1a-activated ac1 promoters in IMR90 cells.

Activation of Small CXCL Cytokine Genes by mRNA Stabilization

e1a increased *CXCL1* and *2* mRNAs by stabilizing them (Figure 4). Induction of these genes as well as *IL8* and *CXCL3* requires the e1a interactions with both RBs and P300 (Figure 1D). All of these homologous cytokines bind and activate the same GPCR, CXCR2 (Rosenkilde and Schwartz, 2004). Stabilization of these mRNAs is a response activated by e1a rather than a cellular response to viral infection because they were not induced by infection with an E1A deletion or the vectors for e1aP300b or e1aRBb . It seems surprising that adenovirus would evolve a mechanism to activate expression of these proinflammatory cytokines. However, these cytokines stimulate expression of viral receptors and consequently increase infection of neighboring cells (Lütschg et al., 2011).

e1a probably stabilizes the CXCL cytokine mRNAs indirectly by inducing genes for proteins that stabilize mRNAs with AU-rich 3′ UTRs and repressing genes for proteins that cause degradation of such mRNAs. In this regard, it is interesting that e1a induced an isoform of ELAVL2 (NM_001171197) nearly 10-fold and repressed ZFP36L1 more than 2-fold, since ELAV-family proteins stabilize mRNAs with AU-rich elements and ZFP36-family proteins destabilize them (Mukherjee et al., 2014; Rattenbacher and Bohjanen, 2012). The requirement for e1a binding to both RBs and P300 may be for complete repression of genes that destabilize these mRNAs.

e1a Inhibition of Cellular Responses to Viral Infection may Explain Why the e1a-P300 Interaction and a P300-e1a-RB Complex Are Required for Transformation

Previously, it was not clear why e1a must interact with P300 to transform primary cells. Results from cells infected with the E1AΔ mutant provide potential answers. Genes whose induction by E1AΔ is inhibited by WT e1a, dependent on the e1a-P300 interaction, include *CDKN1A* and stress-induced *SESN2*, an mTOR inhibitor (Budanov and Karin, 2008) (Figures S6B and S6C). Genes whose induction by E1AΔ is inhibited dependent on simultaneous interactions between e1a and both P300 and RBs include *IFIT1* and *IFIT2*, interferon-induced proapoptotic proteins (Reich, 2013) (Figures S6D and S6E). e1a's ability to inhibit induction of these genes following the stress associated with transfection likely contributes to the requirement for these simultaneous interactions for transformation.

A Complex of p300-e1a-RB1 Condenses Chromatin

Perhaps the most unanticipated results in these studies came from searching for a mechanism that accounts for e1a-repression of some, but not all, genes. The largest class of repressed genes requires the e1a-P300 interaction for repression (rc1, Figures 1D and S1L). This might suggest that e1a targets genes for repression that have P300 associated with their control regions. However, this includes virtually all expressed genes (Visel et al., 2009). We noted that e1a-repressed genes have a higher average p300 association within their transcribed regions, the "gene body," compared to genes expressed at a comparable level that were unchanged in their expression by e1a (Figure 6A). A high level of p300 spread throughout the gene body in uninfected cells was observed in gene browser views of 76 genes (Table S6) enriched for genes of the TRAIL; TGFβ; TNF; and IL1, IL3, and IL5 signaling pathways (Figures 6B and 6E). p300 association with the bodies of these genes increased to high levels in response to e1a. Repression of the TGFβ and TNF pathways by Ad2 and Ad5 infection was noted earlier (Zhao et al., 2012 and references therein). Remarkably, many of the same genes showed similar association of p300 with the gene body when they were induced in serum-starved T98G glioblastoma cells (Ramos et al., 2010) (Figure S6G). However, this was associated with strong activation in T98G cells, whereas these same genes were repressed by e1a in IMR90 cells. This is probably because e1a also induced RB1 association, H3 hypoacetylation, and reduced pol2 within these gene bodies (Figures 6B, 6D, 6F, and S4A–S4C). Based on these data, we suggest that one mechanism to target e1a-repression is for e1a to associate with P300 in genes that have a high level of P300 throughout the gene body, coupled with e1a cobinding to hypophosphorylated, repressing RB proteins associated with repressing chromatin modifying enzymes. The association of P300 and RB1 with these gene bodies in uninfected IMR90 and T98G cells, and the increase in p300 and RB1 in response to serum in uninfected T98G cells (Figures 6B, 6C, 6F, and S6G), suggest that similar mechanisms regulate these genes in uninfected cells. Ad may exploit this cellular mechanism to target hypo-phosphorylated RBs to these genes, which inhibit cell cycling and promote apoptosis and cytokine secretion, repressing them.

While this high level of p300 and RB1 in gene bodies was observed clearly at only 76 genes, including three lncRNAs of unknown function (Table S6), we note that repression of the large rc1 cluster was significantly greater for WT e1a than for e1aP300b plus e1aRBb expressed at similar level (Figure 6G). This suggests that, although the e1a-P300 interaction is primarily responsible for e1a repression of rc1 genes, assembly of P300-e1a-RB complexes contributes to their maximal repression by WT e1a. The stronger repression of rc1 and rc4 genes by e1aRBb compared to WT e1a (Figure 1D) may have resulted from the higher level of this mutant in these experiments (Figure S1J) or because the absence of negative cooperativity in the cobinding of RB and P300 (Ferreon et al., 2013) results in binding of the e1aRBb mutant to a larger fraction of total cellular P300.

To explore the mechanism of e1a repression further, we employed a microscopic, cell biological method for directly visualizing chromatin condensation (Figure 7A). When WT e1a was bound to a large, extended array of *lacO* sites in RRE.1 CHO cells (Verschure et al., 2005) by expression of an e1a-NLS-LacI-mCherry fusion (e1a-NLM), the volume occupied by the *lacO* array condensed to ~1/4 the volume visualized with NLM alone (Figures 7A and 7D). Moreover, this ability to condense chromatin required interaction of both P300 and RBs with the same e1a molecule, as well as the e1a interaction with p400, HDAC activity, the KAT activity of P300, the P300 bromo domain, and P300 acetylated lysines in RB1 and e1a. RB1 remained hypophosphorylated even when CDK2-cyclin E activity was induced, presumably because e1a-induced P300 acetylation of RB1 K873/K874 inhibits RB1 phosphorylation by cyclin-CDKs (Chan et al., 2001). This is expected to maintain RB1 (and probably the other RBs) in a repressing conformation that interacts with repressing chromatin-modifying enzymes (Dick and Rubin, 2013). We propose that this causes chromatin associated with P300/p400-e1a-RB to become hypoacetylated and condensed, repressing transcription. In this way Ad appears to exploit RB repression complexes released from E2Fs by e1a by targeting them to genes that would otherwise inhibit viral replication (Figure 7H). P400 is a SWI/SNF chromatin remodeler. SWI/SNF complexes also function in other examples of repression (Martens and Winston, 2003). The requirement of the e1a p400 binding region (aa 25–36) (Fuchs et al., 2001) for direct visualization of e1a-mediated chromatin condensation (Figure S7D) is consistent with the model that remodelers can "close" chromatin as well as "open" it.

Importantly, large E1A did not induce chromatin condensation (Figure S7D), revealing an activity of small e1a that is not shared with large E1A, despite the presence of all of the small e1a sequence in the larger protein. This may provide an explanation

for why all primate Ads express both proteins via alternative RNA splicing and why a mutant virus that cannot express small e1a because of a point mutation in the 12S mRNA 5′ splice site replicates as well as wild-type virus in cycling primary cells, but not in G_1-arrested cells (Spindler et al., 1985).

The finding that the p300 bromo domain is required for e1a-mediated chromatin condensation is interesting in light of the required p300 acetylation of e1a and RB1. It is unlikely that the p300 bromo domain is required for binding to acetylated histones, since the repressed genes become extensively hypoacetylated (Figure 6F). P300 bromo domains might bind e1a K239ac and/or RB1 K873ac/K874ac to form of a lattice of multiple P300-e1a-RB complexes through such interactions in addition to those diagrammed in Figure 1B (Figure 7H). Such a network of interactions might help to explain how p300 and RB1 virtually "coat" the genes in Table S6. It might also explain why YFP-P300 showed reduced association with e1a-K239R at the *lacO* array in RRE cells (Figures S7F and S7G). Further studies will be required to test these ideas.

EXPERIMENTAL PROCEDURES

Cell Culture

IMR-90 primary human fetal lung fibroblasts (ATCC Number: CCL-186) were obtained from the ATTC and Sigma-Aldrich. They were grown at 37°C in Dulbecco's modified Eagle's medium plus 10% FBS, penicillin, and streptomycin in a 5% CO_2 incubator until they reached confluence. Cells were then incubated 2 days more and were either mock infected or infected with the indicated Ad5-based vectors.

RNA-Seq

Low-passage IMR-90 cells were mock-infected or infected with Ad5 E1A-E1B-substituted, E3-deleted vectors expressing WT Ad2 small E1A proteins from the dl1520 deletion removing the 13S E1A mRNA 5′ splice site (Montell et al., 1984) as described in the text, 2 days after reaching confluence. RNA was isolated 24 hr p.i. using QIAGEN RNeasy Plus Mini Kit. Eluted RNA was treated with Ambion DNase Treatment and Removal reagent and then Ambion TRIzol reagent, precipitated with isopropanol, and dissolved in sterile water. RNA concentration was measured with a Qubit fluorometer. One microgram of RNA was copied into DNA and PCR amplified with bar-coded primers for separate samples to prepare sequencing libraries using the Illumina TruSeq RNA Sample Preparation procedure. Libraries were sequenced using the Illumina Hiseq-2000 to obtain 50-base-long reads. Sequences were aligned to the hg19 human genome sequence using TopHat v2. Fpkm (fragments per kb per million mapped reads) for each annotated hg19 RefSeq transcript was determined using Cuffdiff v2 from Cufflinks RNA-Seq analysis tools at http://cufflinks.cbcb.umd.edu.

ChIP-Seq

Preparation of cross-linked chromatin free of RNA, sonication, and immunoprecipitation was as described in (Ferrari et al., 2012). ChIP of RB1 was done using formaldehyde and DSG cross-linking as described (Chicas et al., 2010). Sequencing libraries were constructed from 1 ng of immunoprecipitated and input DNA using the NuGen Ovation Ultralow DR Multiplex System 1-8 kit. Analysis of sequence data was as described in Ferrari et al. (2012), except that the genome was tiled into 50 bp windows.

Confocal Microscopy of Transfected RRE.1 Cells

RRE.1.1 cells (Verschure et al., 2005) were plated on fibronectin-coated glass bottom 35 mm dishes (MatTek Corporation) and transfected with 2 μg expression vector for e1a or e1a mutant-NLS-LacI-mCherry, or cotransfected with 2 μg WT e1a-NLS-LacI-mCherry and 2 μg YFP-p300 or YFP-P300 mutant, or YFP-RB1, or YFP mutant RB1 in the pCDNA expression vector using the CMV IE enhancer/promoter with Lipofectamine 2000 (Invitrogen). After 24 hr cells were fixed in 1.6% formaldehyde, washed in 1× PBS, mounted onto slides, and imaged for colocalization of mCherry and YFP using a Leica TCS SP2 AOBS single-photon confocal microscope using a 63×1.4-numerical-aperture oil immersion objective. Micrographs were analyzed with ImageJ software to subtract all background that was not at least 25% of maximum fluorescence and subjected to particle analysis in ImageJ to identify foci of fluorescence and measure their areas in μm^2. The bromodomain deletion in YFP-P300 BDΔ included p300 amino acids 1,071–1,241. p values for differences between the distributions of data shown in boxplots was calculated using Kaleidograph to perform one way ANOVA with a Tukey's HSD post hoc comparison. The "p400b " mutant fused to SV40 NLS-lac I-mCherry (NLM) used in Figure S7D has e1a mutations E25A, E26A, V27A, L28A, D30A, L32A, P35A, and S36A. The "CtBPb " mutant fused to NLM used in Figure 7D has e1a mutations P233A, D235A, L236A, and S237A.

ACCESSION NUMBERS

All RNA- and ChIP-seq data has been uploaded to NCBI GEO accession number GSE59693.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2014.10.004.

AUTHOR CONTRIBUTIONS

All authors contributed to experimental design and data interpretation. G.J. constructed the Ψ5 expression vectors. D.G. performed RNA-seq and ChIP-seq for RB1, R.F. for p300, T.S. for H3K27ac and H3K4me1, and M.N. for pol2. R.F., M.P., and N.Z. performed bioinformatic analysis; A.Y. and M.N. performed gel filtration westerns of Figure 7C; and S.J. performed the microscopy of Figures 7 and S9.

ACKNOWLEDGMENTS

Supported by CA25235 to A.J.B., NIH Director's Innovator Award (1-DP2-OD006515) to S.K.K., GM095656 to M.P., NRSAs T32HG02536 to M.N., T32GM007185 to S.J., and T32AI060567 to N.Z. We thank Professor Ulf Pettersson (Uppsala University) for providing bam files from Zhao et al. 2012.

Received: May 1, 2014
Revised: July 25, 2014
Accepted: September 7, 2014
Published: November 12, 2014

REFERENCES

Ablack, J.N., Cohen, M., Thillainadesan, G., Fonseca, G.J., Pelka, P., Torchia, J., and Mymryk, J.S. (2012). Cellular GCN5 is a novel regulator of human adenovirus E1A-conserved region 3 transactivation. J. Virol. *86*, 8198–8209.

Bagchi, S., Raychaudhuri, P., and Nevins, J.R. (1990). Adenovirus E1A proteins can dissociate heteromeric complexes involving the E2F transcription factor: a novel mechanism for E1A trans-activation. Cell *62*, 659–669.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Black, E.P., Huang, E., Dressman, H., Rempel, R., Laakso, N., Asa, S.L., Ishida, S., West, M., and Nevins, J.R. (2003). Distinct gene expression phenotypes of cells lacking Rb and Rb family members. Cancer Res. *63*, 3716–3723.

Branton, P.E., Bayley, S.T., and Graham, F.L. (1985). Transformation by human adenoviruses. Biochim. Biophys. Acta *780*, 67–94.

Budanov, A.V., and Karin, M. (2008). p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. Cell *134*, 451–460.

Chakravarti, D., Ogryzko, V., Kao, H.Y., Nash, A., Chen, H., Nakatani, Y., and Evans, R.M. (1999). A viral mechanism for inhibition of p300 and PCAF acetyltransferase activity. Cell *96*, 393–403.

Chan, H.M., Krstic-Demonacos, M., Smith, L., Demonacos, C., and La Thangue, N.B. (2001). Acetylation control of the retinoblastoma tumour-suppressor protein. Nat. Cell Biol. *3*, 667–674.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R.A., Narita, M., Zhang, M., and Lowe, S.W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. Cancer Cell *17*, 376–387.

Chong, J.L., Wenzel, P.L., Sáenz-Robles, M.T., Nair, V., Ferrey, A., Hagan, J.P., Gomez, Y.M., Sharma, N., Chen, H.Z., Ouseph, M., et al. (2009). E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. Nature *462*, 930–934.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Dick, F.A., and Rubin, S.M. (2013). Molecular mechanisms underlying RB protein function. Nat. Rev. Mol. Cell Biol. *14*, 297–306.

Fattaey, A.R., Harlow, E., and Helin, K. (1993). Independent regions of adenovirus E1A are required for binding to and dissociation of E2F-protein complexes. Mol. Cell. Biol. *13*, 7267–7277.

Ferrari, R., Pellegrini, M., Horwitz, G.A., Xie, W., Berk, A.J., and Kurdistani, S.K. (2008). Epigenetic reprogramming by adenovirus e1a. Science *321*, 1086–1088.

Ferrari, R., Su, T., Li, B., Bonora, G., Oberai, A., Chan, Y., Sasidharan, R., Berk, A.J., Pellegrini, M., and Kurdistani, S.K. (2012). Reorganization of the host epigenome by a viral oncogene. Genome Res. *22*, 1212–1221.

Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2009). Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. Proc. Natl. Acad. Sci. USA *106*, 13260–13265.

Ferreon, A.C., Ferreon, J.C., Wright, P.E., and Deniz, A.A. (2013). Modulation of allostery by protein intrinsic disorder. Nature *498*, 390–394.

Fuchs, M., Gerber, J., Drapkin, R., Sif, S., Ikura, T., Ogryzko, V., Lane, W.S., Nakatani, Y., and Livingston, D.M. (2001). The p400 complex is an essential E1A transformation target. Cell *106*, 297–307.

Ghosh, M.K., and Harter, M.L. (2003). A viral mechanism for remodeling chromatin structure in G0 cells. Mol. Cell *12*, 255–260.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479–491.

Horwitz, G.A., Zhang, K., McBrian, M.A., Grunstein, M., Kurdistani, S.K., and Berk, A.J. (2008). Adenovirus small e1a alters global patterns of histone modification. Science *321*, 1084–1085.

Ikeda, M.A., and Nevins, J.R. (1993). Identification of distinct roles for separate E1A domains in disruption of E2F complexes. Mol. Cell. Biol. *13*, 7029–7035.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. *30*, 249–262.

Jones, N., and Shenk, T. (1979). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Natl. Acad. Sci. USA *76*, 3665–3669.

Kraus, W.L., Manning, E.T., and Kadonaga, J.T. (1999). Biochemical analysis of distinct activation functions in p300 that enhance transcription initiation with chromatin templates. Mol. Cell. Biol. *19*, 8123–8135.

Lee, J.O., Russo, A.A., and Pavletich, N.P. (1998). Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. Nature *391*, 859–865.

Lee, C., Chang, J.H., Lee, H.S., and Cho, Y. (2002). Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. Genes Dev. *16*, 3199–3212.

Liu, X., and Marmorstein, R. (2007). Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. Genes Dev. *21*, 2711–2716.

Lütschg, V., Boucke, K., Hemmi, S., and Greber, U.F. (2011). Chemotactic antiviral cytokines promote infectious apical entry of human adenovirus into polarized epithelial cells. Nat. Commun. *2*, 391.

Madison, D.L., Yaciuk, P., Kwok, R.P., and Lundblad, J.R. (2002). Acetylation of the adenovirus-transforming protein E1A determines nuclear localization by disrupting association with importin-alpha. J. Biol. Chem. *277*, 38755–38763.

Martens, J.A., and Winston, F. (2003). Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Curr. Opin. Genet. Dev. *13*, 136–142.

Miller, D.L., Myers, C.L., Rickards, B., Coller, H.A., and Flint, S.J. (2007). Adenovirus type 5 exerts genome-wide control over cellular programs governing proliferation, quiescence, and survival. Genome Biol. *8*, R58.

Montell, C., Courtois, G., Eng, C., and Berk, A. (1984). Complete transformation by adenovirus 2 requires both E1A proteins. Cell *36*, 951–961.

Mukherjee, N., Jacobs, N.C., Hafner, M., Kennington, E.A., Nusbaum, J.D., Tuschl, T., Blackshear, P.J., and Ohler, U. (2014). Global target mRNA specification and regulation by the RNA-binding protein ZFP36. Genome Biol. *15*, R12.

Pelka, P., Ablack, J.N., Fonseca, G.J., Yousef, A.F., and Mymryk, J.S. (2008). Intrinsic structural disorder in adenovirus E1A: a viral molecular hub linking multiple diverse processes. J. Virol. *82*, 7252–7263.

Pelka, P., Ablack, J.N., Shuen, M., Yousef, A.F., Rasti, M., Grand, R.J., Turnell, A.S., and Mymryk, J.S. (2009a). Identification of a second independent binding site for the pCAF acetyltransferase in adenovirus E1A. Virology *391*, 90–98.

Pelka, P., Ablack, J.N., Torchia, J., Turnell, A.S., Grand, R.J., and Mymryk, J.S. (2009b). Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP. Nucleic Acids Res. *37*, 1095–1106.

Ramos, Y.F., Hestand, M.S., Verlaan, M., Krabbendam, E., Ariyurek, Y., van Galen, M., van Dam, H., van Ommen, G.J., den Dunnen, J.T., Zantema, A., and 't Hoen, P.A. (2010). Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Res. *38*, 5396–5408.

Rattenbacher, B., and Bohjanen, P.R. (2012). Evaluating posttranscriptional regulation of cytokine genes. Methods Mol. Biol. *820*, 71–89.

Reich, N.C. (2013). A death-promoting role for ISG54/IFIT2. J. Interferon Cytokine Res. *33*, 199–205.

Rosenkilde, M.M., and Schwartz, T.W. (2004). The chemokine system — a major regulator of angiogenesis in health and disease. APMIS *112*, 481–495.

Ruley, H.E. (1983). Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. Nature *304*, 602–606.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Sha, J., Ghosh, M.K., Zhang, K., and Harter, M.L. (2010). E1A interacts with two opposing transcriptional pathways to induce quiescent cells into S phase. J. Virol. *84*, 4050–4059.

Spindler, K.R., Eng, C.Y., and Berk, A.J. (1985). An adenovirus early region 1A protein is required for maximal viral DNA replication in growth-arrested human cells. J. Virol. *53*, 742–750.

Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E., Belmont, A.S., and van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. Mol. Cell. Biol. *25*, 4552–4564.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Wang, H.G., Moran, E., and Yaciuk, P. (1995). E1A promotes association between p300 and pRB in multimeric complexes required for normal biological activity. J. Virol. *69*, 7917–7924.

Weinberg, R. (2013). The Biology of Cancer. (New York, NY: Garland Science).

Weitzman, M.D., and Ornelles, D.A. (2005). Inactivating intracellular antiviral responses during adenovirus infection. Oncogene *24*, 7686–7696.

Winberg, G., and Shenk, T. (1984). Dissection of overlapping functions within the adenovirus type 5 E1A gene. EMBO J. *3*, 1907–1912.

Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., et al. (2011). A poised chromatin platform for TGF-β access to master regulators. Cell *147*, 1511–1524.

Xiao, B., Spencer, J., Clements, A., Ali-Khan, N., Mittnacht, S., Broceño, C., Burghammer, M., Perrakis, A., Marmorstein, R., and Gamblin, S.J. (2003). Crystal structure of the retinoblastoma tumor suppressor protein bound to E2F and the molecular basis of its regulation. Proc. Natl. Acad. Sci. USA *100*, 2363–2368.

Zhang, Q., Yao, H., Vo, N., and Goodman, R.H. (2000). Acetylation of adenovirus E1A regulates binding of the transcriptional corepressor CtBP. Proc. Natl. Acad. Sci. USA *97*, 14323–14328.

Zhao, H., Granberg, F., Elfineh, L., Pettersson, U., and Svensson, C. (2003). Strategic attack on host cell gene expression during adenovirus infection. J. Virol. *77*, 11006–11015.

Zhao, H., Dahlö, M., Isaksson, A., Syvänen, A.C., and Pettersson, U. (2012). The transcriptome of the adenovirus infected cell. Virology *424*, 115–128.

Cell Host & Microbe, Volume *16*

Supplemental Information

Adenovirus Small E1A Employs the Lysine Acetylase p300/CBP and Tumor Suppressor Rb to Repress Select Host Genes and Promote Productive Virus Infection

Roberto Ferrari, Dawei Gou, Gauri Jawdekar, Sarah A. Johnson, Miguel Nava, Trent Su, Ahmed F. Yousef, Nathan R. Zemke, Matteo Pellegrini, Siavash K. Kurdistani, and Arnold J. Berk

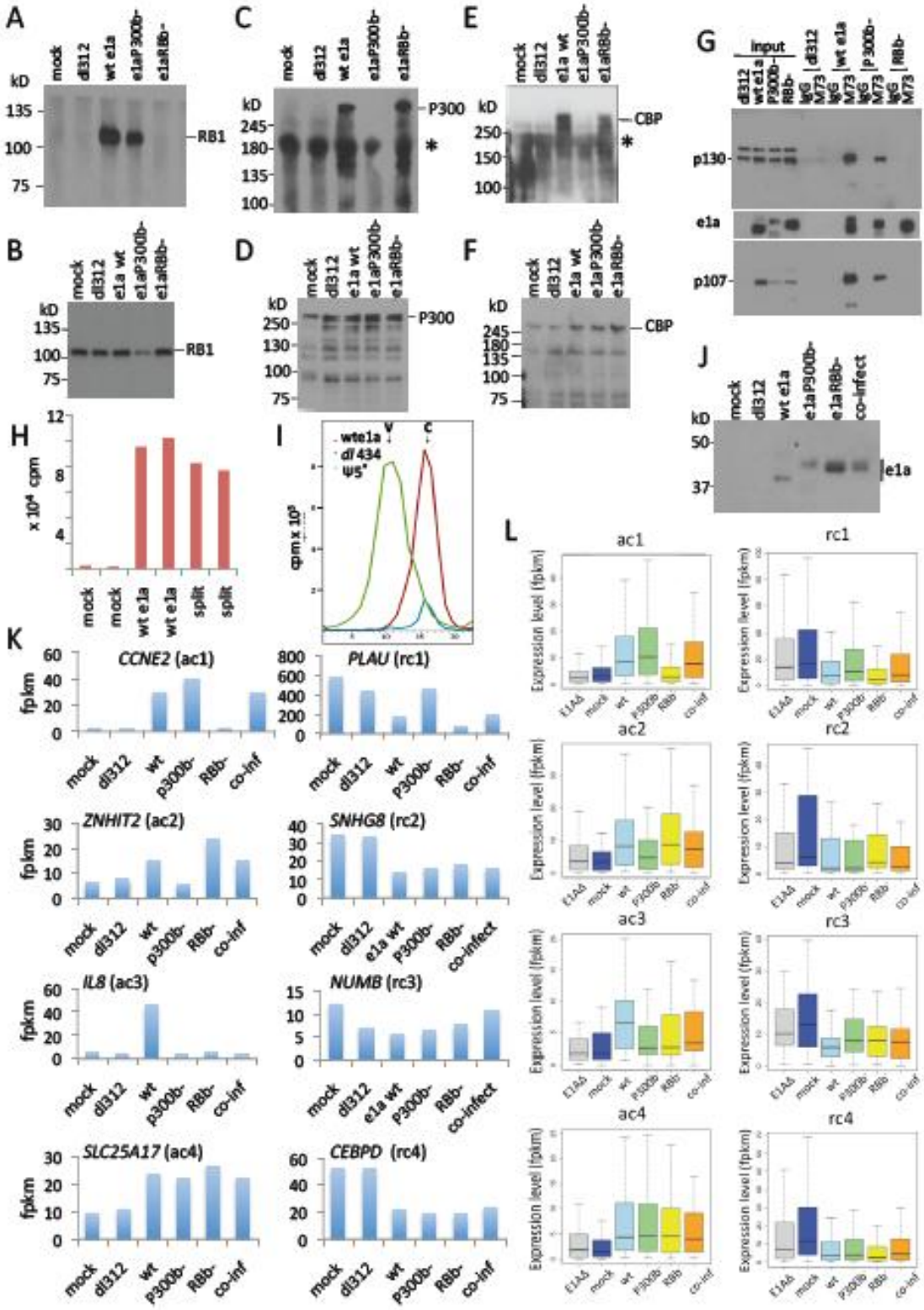

A
mock
dl312
wt e1a
e1aP300b-
e1aRBb-
kD
135
100
75
RB1
B
mock
dl312
e1a wt
e1aP300b-
e1aRBb-
kD
135
100
75
RB1
C
mock
dl312
wt e1a
e1aP300b-
e1aRBb-
kD
245
180
135
100
P300
*
D
mock
dl312
e1a wt
e1aP300b-
e1aRBb-
kD
250
130
100
75
P300
E
mock
dl312
e1a wt
e1aP300b-
e1aRBb-
kD
250
150
100
CBP
*
F
mock
dl312
e1a wt
e1aP300b-
e1aRBb-
kD
245
180
135
100
75
CBP
G
input
dl312
wt e1a
P300b-
RBb-
IgG
M73
p130
e1a
p107
H
x 10⁴ cpm
12
8
4
mock
mock
wt e1a
wt e1a
split
split
I
wte1a
dl 434
ψ5*
V
C
cpm x 10³
J
mock
dl312
wt e1a
e1aP300b-
e1aRBb-
co-infect
kD
50
37
e1a
K
CCNE2 (ac1)
PLAU (rc1)
ZNHIT2 (ac2)
SNHG8 (rc2)
IL8 (ac3)
NUMB (rc3)
SLC25A17 (ac4)
CEBPD (rc4)
fpkm
mock
dl312
wt
p300b-
RBb-
co-inf
co-infect
e1a wt
L
ac1
rc1
ac2
rc2
ac3
rc3
ac4
rc4
Expression level (fpkm)
E1A5
mock
wt
P300b
RBb
co-inf

Fig. S1

Fig. S1 related to Fig 1D: Mutant e1a binding to RBs and p300/CBP, induction of cellular DNA synthesis, and regulation of cellular mRNAs in contact-inhibited IMR90 cells. (A) Contact-inhibited IMR90 cells were infected with Ad5 expression vectors ($\psi 5^*$, derived from the ψ5-vector (Hardy et al., 1997) by deletion of the CMV-IE promoter/enhancer so that the indicated e1a mutants were expressed from the Ad5 E1A promoter and enhancer). 24 h p.i. extracts were prepared and immunoprecipitated with anti-e1a mAb M73 (Harlow et al., 1985), and the IP was western blotted with anti-RB1 mAb 4H1 (Cell Signaling). **(B)** Inputs for IP-western in (A). Infection with dl312 and the vectors for wt e1a and e1aRBb$^-$ had little effect on the level of RB1, while e1aP300b$^-$ caused an ~2-4-fold decrease. **(C)** same as (A) but western with anti-p300 rabbit polyclonal antibody. * here and below indicates a non-specific band. **(D)** Inputs for IP-western in (C). Infection with dl312 and vectors for wt and mutant e1a's caused a modest increase in total cellular p300 compared to mock-infected, arrested cells. **(E)** Same as (A) but western with anti-CBP. **(F)** Inputs for IP-western in E. Expression of wt and mutant e1a's caused an ~2-fold increase in total cellular CBP. **(G)** Inputs and IP-westerns for p130, p107 and e1a wt and mutants. Extracts were prepared from HeLa cells 24 h p.i. with the indicated vectors. The same IPs and inputs were resolved on three gels. One was blotted with anti-p130 (top), one with anti-e1a mAb M73 (middle), and one with anti-p107 (bottom). Control IPs were with non-immune mouse IgG. Wt e1a and e1aRBb$^-$ increased p130 ~2 fold, but not e1aP300b$^-$. Wt e1a induced p107 in the G_1-arrested cells, while e1aP300b$^-$ and e1aRBb$^-$ did so to a lesser extent. A lower level of p130 and p107 were IPed from cells infected with the e1aP300b$^-$ vector compared to the wt e1a vector, presumably because of the lower level of e1aP300b$^-$ compared to wt e1a in this experiment. **(H)** Contact-inhibited IMR90 cells were infected with the wt e1a Ad vector, mock-infected, or split 1:3 into fresh medium with 10% FBS in duplicate. At 22 h p.i. cells were labeled with ^{3}H-thymidine for 2 h, and total TCA-precipitable cpm were measured. **(I)** Contact-inhibited IMR90 cells were infected with the wt e1a vector, or the E1A-E1B deletion mutant dl434 (Klessig et al., 1982), or complementing 293 cells were infected with the Ad vector $\psi 5^*$, and cells were labeled with ^{3}H-thymidine from 22-24 h p.i. DNA was isolated and subjected to CsCl gradient equilibrium buoyant density centrifugation in parallel gradients. Fractions were collected from the bottom, subjected to TCA-precipitation, and precipitated cpm counted. The arrows marked V and C show the position of Ad vector DNA labeled in 293 cells (green), and of IMR90 cellular DNA from cells infected with the wt e1a vector (red), or dl434 (blue). These results show the higher level of cellular DNA synthesis in cells expressing wt e1a compared to the control dl434, and that the cpm shown in (H) lanes wt e1a had the density of cellular DNA. **(J)** Western blot of extracts from contact-inhibited IMR90 24 h p.i. with the indicated vectors, or co-infected with the e1aP300b$^-$ and e1aRBb$^-$ vectors. These extracts were from duplicate plates infected and harvested at the same time as plates used for isolation of RNA for one of the two replicate RNA-seq experiments. Wt e1a has a faster mobility in an SDS gel than the mutants, perhaps because wt e1a is acetylated by P300 on K239 (Madison et al., 2002; Zhang et al., 2000) resulting in a higher

net negative charge. **(K)** fpkm of RNA from the indicated genes following infection with (from left to right) mock, the E1AΔ mutant dl312 (Jones and Shenk, 1979), e1a wt vector, e1aP300b$^-$ vector, e1aRBb$^-$ vector, and co-infection with the e1aP300b$^-$ and e1aRBb$^-$ vectors. One representative of each Fig 1D cluster is shown. **(L)** Boxplots of the distribution of fpkm for all the genes in each Fig 1D cluster following infection as in (K). The line in the box is the median; the box contains 50% of the data; the whiskers include 75% of the data.

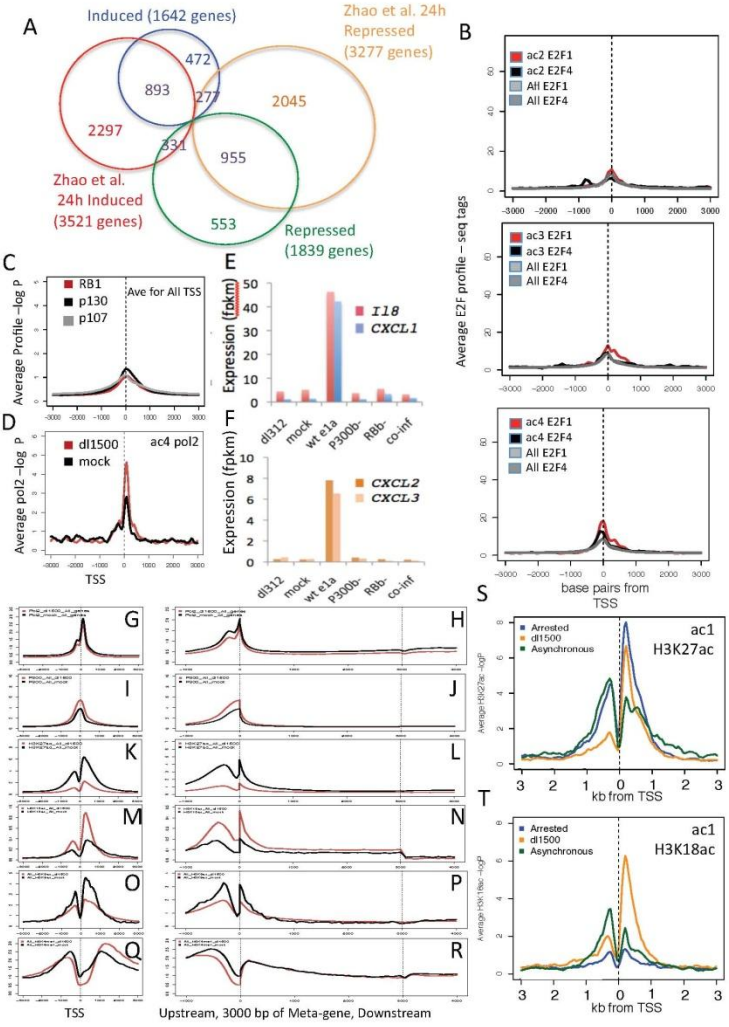

A
Induced (1642 genes)
Zhao et al. 24h Repressed (3277 genes)
Zhao et al. 24h Induced (3521 genes)
Repressed (1839 genes)
472
893
277
2045
2297
331
955
553
B
ac2 E2F1
ac2 E2F4
All E2F1
All E2F4
ac3 E2F1
ac3 E2F4
ac4 E2F1
ac4 E2F4
Average E2F profile – seq tags
base pairs from TSS
C
RB1
p130
p107
Ave for All TSS
Average Profile –log P
D
dl1500
mock
ac4 pol2
Average pol2 –log P
TSS
E
Expression (fpkm)
Il8
CXCL1
F
CXCL2
CXCL3
dl312
mock
wt e1a
p300b-
RBb-
co-inf
S
Arrested
dl1500
Asynchronous
ac1
H3K27ac
kb from TSS
T
H3K18ac
G
H
I
J
K
L
M
N
O
P
Q
R
TSS
Upstream, 3000 bp of Meta-gene, Downstream

Fig S2

Fig S2 related to Fig 1D: ChIP-seq and expression data. (A) Comparison of data reported here with the results of Zhao et al. (2012) for expression of IMR90 genes at 24 h p.i. with wild-type Ad2. Bam files from Zhao et al. (2012) were analyzed by Cufflinks to determine fpkm for the most abundant transcript for each gene. Fpkms from 24 h p.i. with wt Ad2 were divided by fpkm from mock-infected cells. Genes were excluded that did not have an fpkm >0.1 for the most abundant transcript from mock-infected cells (for activated genes) or from e1a wt vector-infected cells (for repressed genes). Genes with 24 h Ad2 fpkm/mock fpkm >2 (induced) and <0.5 (repressed) were compared with our data for RNA isolated 24 h p.i. with the wt e1a vector in duplicate experiments. Venn diagrams show the overlap of activated and repressed genes in the two data sets. There are more activated and repressed genes in the data from Zhao et al., probably because large E1A as well as all other adenovirus genes, early and late, were expressed in cells infected with wt Ad2, whereas only small e1a and very low levels of the other early regions were expressed in cells infected with the wt e1a vector. Also, because we required a factor of two different from wild-type in replicate experiments and $p < 0.01$ for fpkm from wt e1a vector-infected cells comparing the two biological replicates, we may have eliminated more false positives. **(B)** Average seq tags for E2F1 and E2F4 at TSSs of ac2 (top), ac3 (middle), and ac4 (bottom) genes. The plots were prepared using data from the P. Farnham laboratory at University of Southern California reported in (Bernstein et al., 2012): HeLa E2F1 (wgEncodeEH000699) and E2F4 (wgEncode EH000689) at http://www.ncbi.nlm.nih.gov/geo/ **(C)** Average -log poissonP relative to the TSS for all annotated TSSs for RB1, RBL1 (p107), and RBL2 (p130) from Ferrari et al. (2012): GSE32340. **(D)** Average -log poissonP for pol2 at ac4 genes relative to their alligned TSSs in mock- and dl1500-infected cells. **(E)** Expression levels in fpkm for *IL8, CXCL1, CXCL2, and CXCL3* at 24 h p.i. with the indicated Ad vector. **(G)** Average pol2 ±3 kb from the TSS for all genes from mock (black) and wt e1a-vector infected cells (red). **(H)** Metagene plots of pol2 from mock (black) and wt e1a-vector infected cells (red). **(I, J)** same for p300. **(K, L)** same for H3K27ac. **(M, N)** same for H3K18ac. **(O, P)** same for H3K9ac. **(Q, R)** same for H3K4me1. **(S)** Average H3K27ac –log poissonP values for ac1 genes from mock- (Arrested) and dl1500-infected cells compared to the same calculated from ChIP-seq data for asynchronous IMR90 cells reported by the Bing Ren laboratory, University of California, San Diego as reported in (Hawkins et al., 2010) (GSM469965_UCSD.IMR90.H3K27ac.LL230.wig). **(T)** Same as (S) for H3K18ac, (GSM469965_UCSD.IMR90.H3K18ac.LL230.wig).

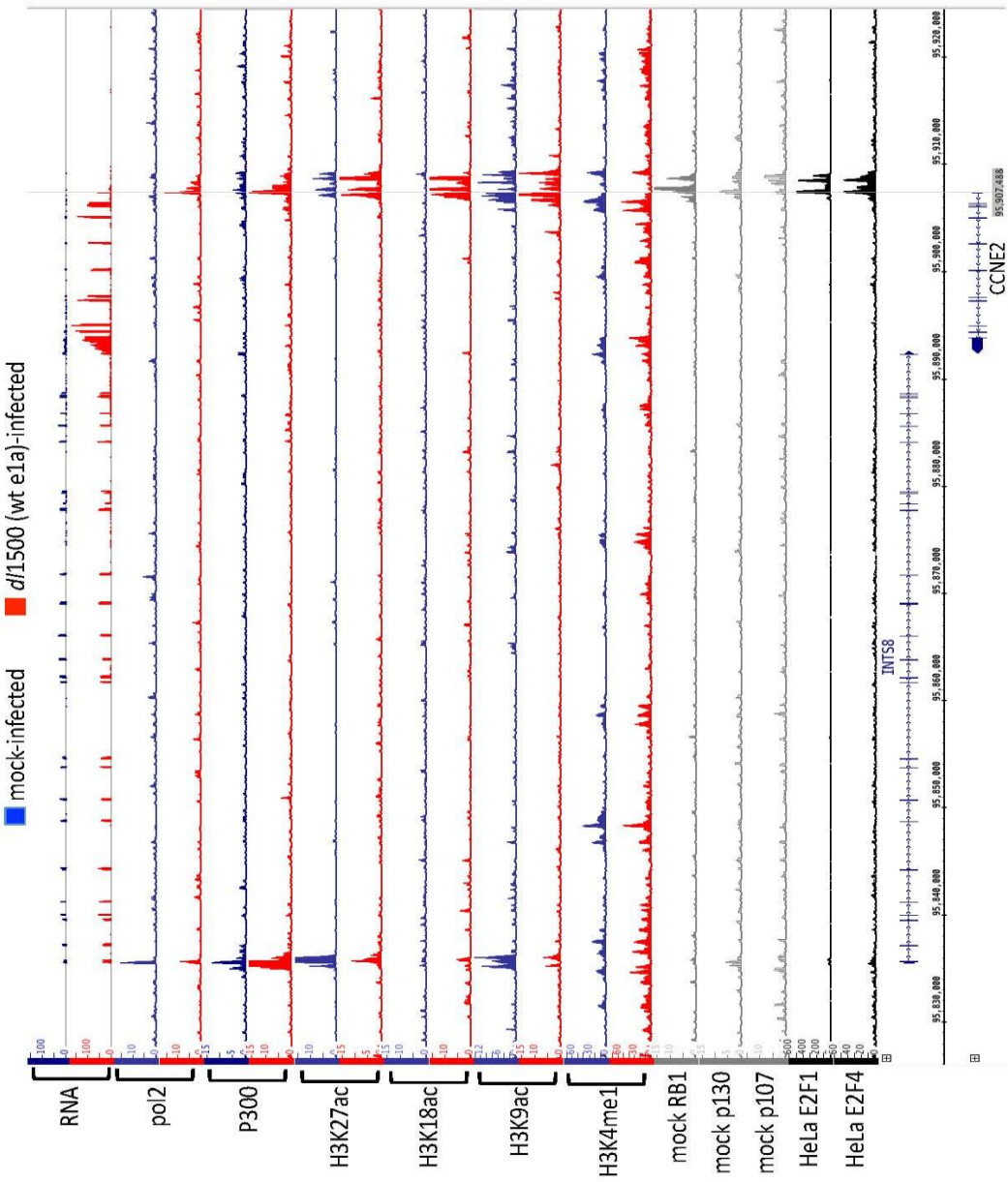


Fig S3 related to Figs. 1-3: Gene browser plots of RNA- and ChIP-seq data for *CCNE2* (cyclin E) and neighboring gene *INTS8*.

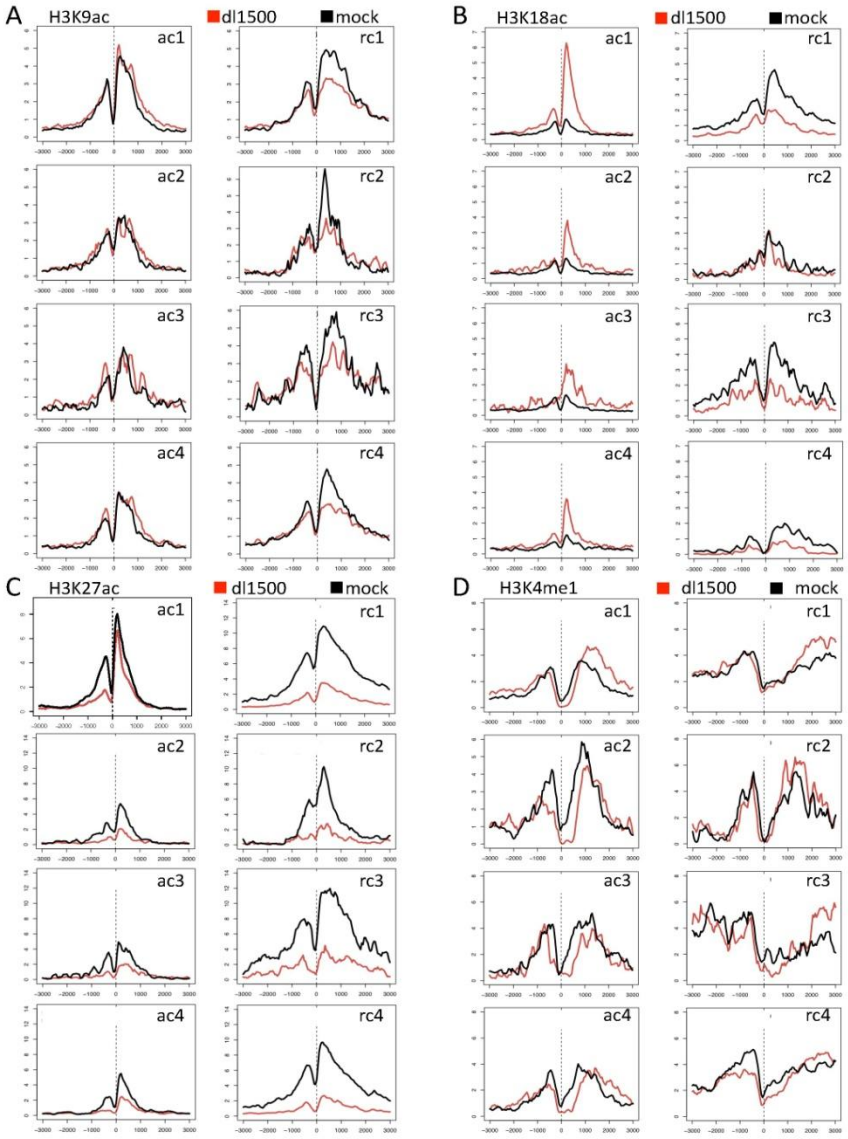


Fig. S4 related to Fig 1D: Histone modifications at e1a-regulated genes. (A-D) Average ChIP-seq signal ±3 kb from the TSS (0) for the modification indicated at the upper left and the e1a-regulated cluster (Table S1) indicated at the upper right of each plot.

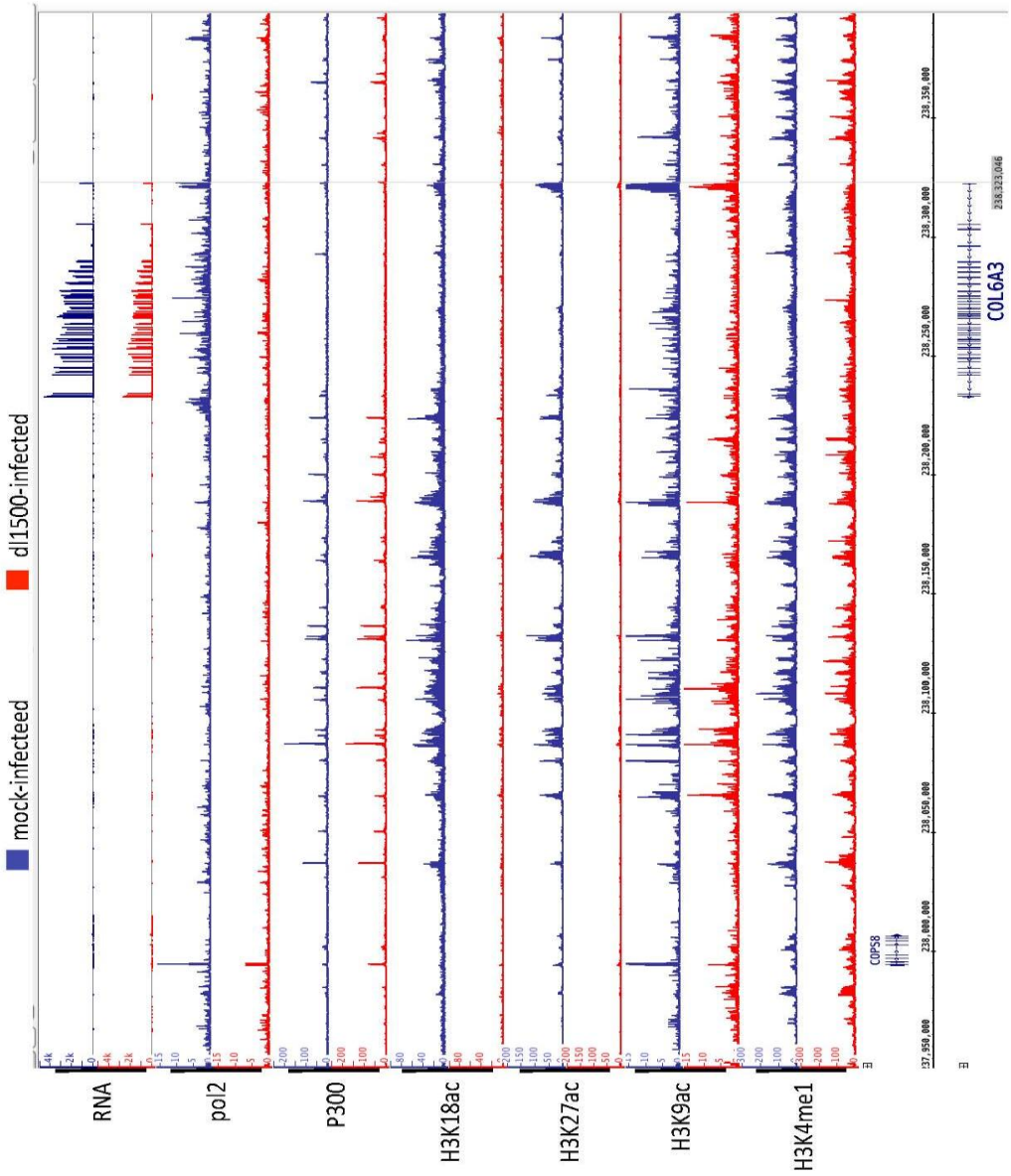


Fig. S5 related to Figs. 1-3: RNA- and ChIP-seq data plotted on the genome browser map of the *COPS8* and *COL6A3* region.

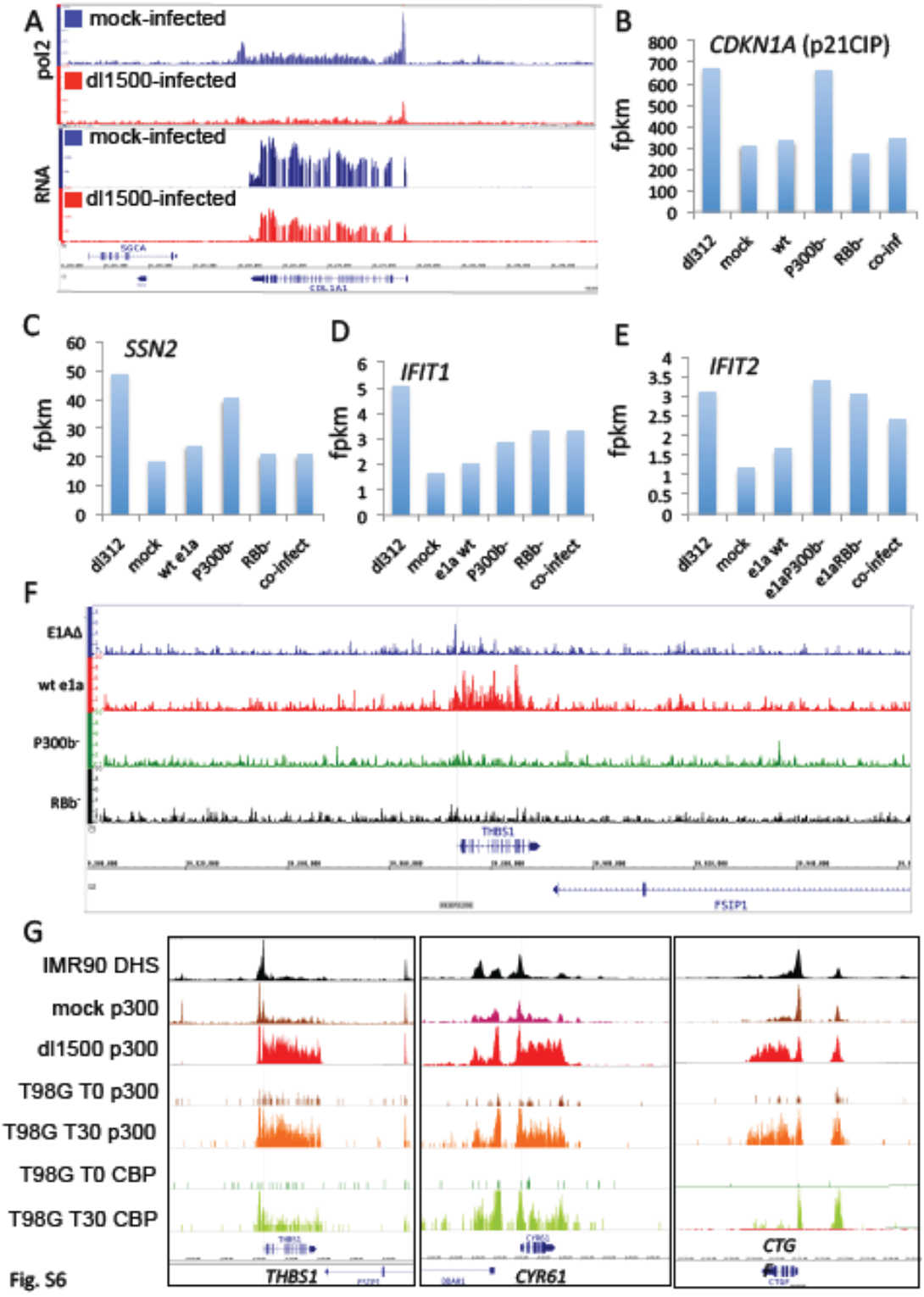

A
pol2
mock-infected
dl1500-infected
RNA
mock-infected
dl1500-infected
B
CDKN1A (p21CIP)
fpkm
dl312
mock
wt
P300b-
RBb-
co-inf
C
SSN2
fpkm
dl312
mock
wt e1a
P300b-
RBb-
co-infect
D
IFIT1
fpkm
dl312
mock
e1a wt
P300b-
RBb-
co-infect
E
IFIT2
fpkm
dl312
mock
e1a wt
e1aP300b-
e1aRBb-
co-infect
F
E1AΔ
wt e1a
P300b-
RBb-
THBS1
FSIP1
G
IMR90 DHS
mock p300
dl1500 p300
T98G T0 p300
T98G T30 p300
T98G T0 CBP
T98G T30 CBP
THBS1
CYR61
CTG

Fig. S6

Fig S6 related to Figs 1D and 6: RNA-seq, ChIP-seq and expression data. (A) pol2 ChIP-seq and RNA-seq data plotted on the genome browser map of the *COL1A1* gene. **(B-E)** Expression of genes induced by infection with an E1AΔ mutant (dl312 (Jones and Shenk, 1979)) whose induction is prevented by wt e1a and their expression in cells infected with vectors for the indicated e1a mutants. **(F)** RB1 ChIP-seq –log poissonP plotted on a gene browser view of *THBS1* following infection with Ad vectors for the indicated e1a mutants. **(G)** Gene browser graphs of DNase I hypersensitive sites in cycling IMR90 cells (DHS) (GSM1008586 from Duke University), mock-infected contact-inhibited IMR90 p300, dl1500-infected IMR90 p300, p300 in serum starved T98G glioblastoma cells (T98G T0 p300), p300 in T98G cells treated for 30 min with 20% FBS (TG98 T30 p300), CBP in serum starved T98G cells (T98G T0 CBP), and CBP in T98G cells treated with 20% FBS for 30 min (T98G T30 CBP) at the *THBS1*, *CYR61*, and *CTGF* genes. ChIP-seq data for p300 and CBP in T98G cells from (Ramos et al., 2010) NCBI GEO GSE21026.

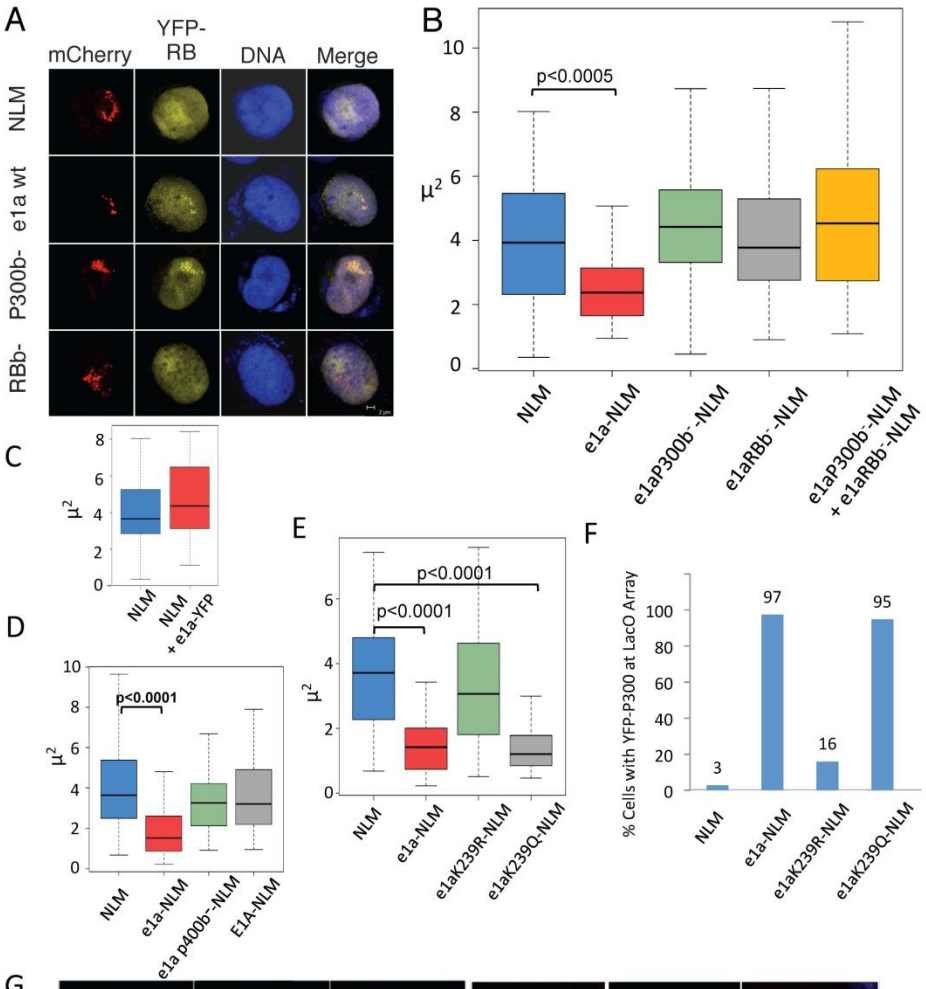

A
mCherry
YFP-RB
DNA
Merge
NLM
e1a wt
P300b-
RBb-
B
p<0.0005
μ²
NLM
e1a-NLM
e1aP300b⁻-NLM
e1aRBb⁻-NLM
e1aP300b⁻-NLM + e1aRBb⁻-NLM
C
NLM
NLM + e1a-YFP
D
p<0.0001
NLM
e1a-NLM
e1a p400b⁻-NLM
E1A-NLM
E
p<0.0001
p<0.0001
NLM
e1a-NLM
e1aK239R-NLM
e1aK239Q-NLM
F
% Cells with YFP-P300 at LacO Array
3
97
16
95
NLM
e1a-NLM
e1aK239R-NLM
e1aK239Q-NLM
G
NLM
YFP-p300
Merge + DAPI
K239R-NLM
YFP-p300
Merge + DAPI
e1a-NLM
YFP-p300
Merge + DAPI
K239Q-NLM
YFP-p300
Merge + DAPI

Fig S7

Fig S7 related to Fig 7: *lacO* array areas in RRE.1 cells transfected with expression vectors for the indicated fusion proteins. (A) Examples of confocal micrographs of e1a wt and mutants fused to NLS-LacI-mChery (NLM) and YFP-RB expressed in transfected RRE.1 cells. **(B)** Boxplots as in Fig 7. Brackets indicate a statistically significant difference (p value indicated) between the distribution of areas from NLM and the indicated e1a wt (e1a) or mutant fused to NLM. **(C)** Boxplots of array areas in RRE.1 cells transfected with the NLM vector or the NLM vector plus a vector for wt e1a fused at its C-terminus to YFP. An e1a-YFP fusion was used in order to visualize cells expressing both fusion proteins. **(D,E)** As in (B) for fusions to the indicated e1a mutants. For comparison, data from Fig. 7A for NLM and e1a-NLM was plotted together with data for e1a p400b$^-$-NLM and large E1A-NLM. **(F)** Co-localization of YFP-P300 with NLM or the indicated e1a mutants fused to NLM at the *lacO* array in RRE.1 cells. **(G)** Confocal micrographs of RRE.1 cells transfected with expression vectors for the indicated proteins. K239R and K239Q refer to the respective e1a mutants.

Table S3 related to Fig 1D: DAVID Gene Ontology of e1a-regulated clusters

		P-value
ac1	DNA metabolic process	8.2 E-51
	DNA replication	2.6 E-44
	Cell cycle	1.1 E-27
ac2	no term <E-3	
ac3	small chemokine, CXC	1.4 E-4
ac4	Glucose/ribitol dehydrogenase	2.9 E-4
	Short-chain dehydrogenase/reductase SDR	4.3 E-4
rc1	glycosylation site:N-linked (GlcNAc...)	1.7 E-11
	regulation of cell proliferation	2.1 E-10
	regulation of cell migration	2.6 E-8
	wound healing	8.2 E-7
	cell-matrix adhesion	5.9 E-6
rc2	no term <E-3	
rc3	no term <E-3	
rc4	regulation of cell proliferation	2.6 E-6
	anchoring junction	4.1 E-6
	cell migration	4.8 E-5
	cytoskeleton	6.5 E-5
	regulation of transcription	6.0 E-4

Table S6 related to Fig 6: Genes with high P300 throughout the gene body in mock-infected IMR90 cells

ACTB
ACTG1
ADAMTS1
ADAMTSL4
ADM
AMOTL2
ATF3
AX747619
BBC3
BHLHE40
BTBD19
CCDC85B
CD44
CEBPB
CITED2
CTGF
CYR61
DLX2
DUSP1
EGR1
ERRFI1
F3
DLGAP1
FOS
FOSB
FOSL1
FTL
GADD45A
GADD45B
HCFC1R1
HES1
HIST2H2AB
HIST2H2AC
HIST2H2BE
ID3
IER2
IL6
IRF2BP2
JUN
JUNB
KDM6B
KLF10
KLF6
KLHL21
LINC01011

LMNA
MAFF
MAFK
MALAT1
MAP2K3
MCL1
MT2A
MYH9
NEAT1
NFKBIA
NR4A1
NR4A3
PFN1
PHLDA1
PLAU
PLK3
S100A6
SLC2A1
SMAD3
SQSTM1
STC2
TGFB2
TGIF1
THBS1
TIMP3
TIPARP
TNFRSF12A
TPM1
TRIB1
VIM
ZFP36

Detailed Experimental Procedures

Cell culture: IMR-90 primary human fetal lung fibroblasts (ATCC® Number: CCL-186 ™) were obtained from the ATTC and Sigma-Aldrich. They were grown at 37°C in DMEM plus 10% FBS, penicillin and streptomycin in a 5% CO_2 incubator until they reached confluence. Cells were then incubated two days and were either mock-infected or infected with the indicated Ad5-based vectors.

Ad vectors: Vectors were similar to the Ψ5 vector that uses in vivo Cre-mediated site specific recombination to substitute the E1A and most of the E1B region, leaving the IX gene intact (Hardy et al., 1997). However, the vectors used (called Ψ5*) have the CMV IE promoter/enhancer deleted. The sequence is wt Ad5 from 1 to 555, including the wt Ad5 E1A enhancer and TSS, followed by CGAAGCT, followed by dl1500 sequence (Montell et al., 1984) from 540 to 1574 in the Ad2 sequence containing the Ad2 E1A AUG at 559, the 12S mRNA 5' splice site at 973, a 9 bp deletion encompassing the 13S 5' splice site at 1111, the 3' splice site for E1A mRNAs at 1225, and the E1A UAA termination codon at 1540. Ad2 1574 is followed by the SV40 late poly A region (SV40 nt 2643-2557), followed by a *loxP* site, followed by dl309 sequence (Jones and Shenk, 1979) from the Bgl II site at 3327 to the right end of the dl309 genome. Complete sequences of the vectors are available upon request to AJB. The "p400b-" mutant fused to SV40 NLS-lac I-mCherry (NLM) used in Fig S9 D has e1a mutations E25A, E26A, V27A, L28A, D30A, L32A, P35A, and S36A. The "CtBPb-" mutant fused to NLM used in Fig. 7D has e1a mutations P233A, D235A, L236A, and S237A.

Infections: For RNA-seq experiments the moi was 40 for wt e1a and e1aP300b$^-$ vectors, 160 for the e1aRBb$^-$ vector, and an moi of 10 each for e1aP300b$^-$ and e1aRBb$^-$ vectors for co-infections. These multiplicities of infection yielded approximately equal amounts of wt and mutant e1a proteins as judged by western blotting with mAb M73 (Fig S1 J). For ChIP-seq experiments, cells were infected at an moi of 200 with dl1500 (Montell et al., 1984) or the Ψ5* vectors, or mock-infected in parallel and incubated for 24 h at 37°C when formaldehyde was added to the media to final concentration 1% and plates were further incubated at 37° for 10 min before preparation of cross-linked chromatin.

RNA-seq: RNA was isolated from infected and mock-infected cells at 24 h p.i. using Qiagen RNeasy Plus Mini Kit. The eluted RNA was then treated with Ambion DNase Treatment and Removal reagent and then treated with Ambion TRIzol reagent and finally precipitated with isopropanol and dissolved in sterile water. RNA concentration was measured using a Qubit fluorometer. 1 μg RNA was then copied into DNA and PCR-amplified with bar-coded primers for separate samples to prepare sequencing libraries using the Illumina TruSeq RNA Sample Preparation procedure. Libraries were sequenced using the Illumina HIseq-2000 to obtain 50-base-long reads. Sequences were aligned to the hg19 human genome sequence using TopHat v2. Fpkm (fragments per kilobase per million mapped reads) for each annotated hg19 RefSeq transcript was determined using Cuffdiff v2 from Cufflinks RNA-Seq analysis tools at http://cufflinks.cbcb.umd.edu .

ChIP-seq: Preparation of cross-linked chromatin free of RNA, sonication, and immunoprecipitation was as described in (Ferrari et al., 2012). ChIP of RB1 was done using formaldehyde and DSG cross-linking as described (Chicas et al., 2010). Sequencing libraries were constructed from 1 ng of immunoprecipitated and input DNA using the NuGen Ovation Ultralow DR Multiplex System 1-8 kit. Analysis of sequence data was as described in Ferrari et al. (2012), except that the genome was tiled into 50 bp windows. Average ChIP-seq signals of the 50 bp windows 3kb upstream and downstream of annotated TSSs and Meta-Gene plots of average ChIP-seq signals across gene bodies were calculated using the CEAS: cis-regulatory annotation system (Shin et al., 2009). *P*-values for the significance of ChIP-seq counts compared to input DNA were calculated as described (Pellegrini and Ferrari, 2012).

Antibodies: H3K18ac (814) prepared and validated as described (Ferrari et al., 2012); H3K27ac, a gift from Michael Grunstein; H3K9ac (Upstate 07-352); H3K4me1 (Abcam ab8895); pol2 N20 (Santa Cruz sc-899x); RB1 (4H1 mAb) (Cell Signaling 9309L); p107 C18 (Santa Cruz sc-318); p130 C20 (Santa Cruz sc-317); p300 N15 (Santa Cruz sc-584); CBP (Santa Cruz sc-369); anti-E1A mAb M73 (Harlow et al., 1985).

External data sources: ChIP-seq wig files and original data for H3K18ac and H3K9ac in mock- and dl1500-infected two day confluent IMR90 cells 24 h p.i. and for RB1, RBL1 (p107), and RBL2 (p130) in mock-infected two day confluent IMR90 cells were from Ferrari et al. (2012) and can be downloaded from NCBI GEO accession number GSE32340 at http://www.ncbi.nlm.nih.gov/geo/. ChIP-

seq E2F1 and E2F4 wig files from cycling HeLa cells were downloaded from NCBI GEO accession numbers GSM935365 and GSM935366, respectively, from the laboratory of Peggy Farnham, University of Southern California (Bernstein et al., 2012). ChIP-seq data for H3K18ac and H3K27ac from asynchronous IMR90 cells were from the Bing Ren laboratory, University of California, San Diego as reported in Hawkins et al. (2010) downloaded from NCBI GEO accession numbers (GSM469965) and (GSM469965), respectively. Bam files from Zhao et al. (2012) were provided by H. Zhou and Professor Ulf Pettersson, University of Uppsala. Wig files of DNase I hypersensitive sites in IMR90 were downloaded from NCBI GEO accession number GSM1008586 from the Duke Genome Center. Wig files for p300 and CBP from serum starved T98G glioblastoma cells and T98 cells treated with fetal calf serum and phorbol ester for 30 min were downloaded from Supplemental Material of (Ramos et al., 2010), and can also be downloaded from NCBI GEO accession number GSE21026.

Confocal microscopy of transfected RRE.1 cells (Verschure et al., 2005) was performed as described earlier for A03 cells (Balamotis et al., 2009). The bromodomain deletion in YFP-P300 BDΔ included P300 amino acids 1071—1241. P-values for differences between the distributions of data shown in boxplots was calculated using Kaleidograph® to perform one way ANOVA and a Tukey's HSD post-hoc comparison.

Gene Ontologies (Fig. 2A) and GREAT Pathways Commons (Fig. 6F) were determined using GREAT at http://bejerano.stanford.edu/great/public/html/ (McLean et al., 2010). Gene Ontologies shown in Table S2 were determined

using DAVID Bioinformatics Resources 6.7 at

http://david.abcc.ncifcrf.gov/home.jsp (Huang da et al., 2009a, b).

Transcription factor motifs were searched using a window of ± 1 kb from the TSS of the e1a-regulated cluster genes in Fig 2D and Table S5 using Cistrome (Liu et al., 2011) and MEME-ChIP (Machanick and Bailey, 2011). Log10 p-values were taken from Cistrome for TF sites found only by Cistrome or from Cistrome when TF sites were found by both Cistrome and MEME-ChIP, or calculated from E-values given as N e^{-n} by MEME-ChIP.

References

Balamotis, M.A., Pennella, M.A., Stevens, J.L., Wasylyk, B., Belmont, A.S., and Berk, A.J. (2009). Complexity in transcription control at the activation domain-mediator interface. Sci Signal *2*, ra20.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

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

Ferrari, R., Su, T., Li, B., Bonora, G., Oberai, A., Chan, Y., Sasidharan, R., Berk, A.J., Pellegrini, M., and Kurdistani, S.K. (2012). Reorganization of the host epigenome by a viral oncogene. Genome Res *22*, 1212-1221.

Hardy, S., Kitamura, M., Harris-Stansil, T., Dai, Y., and Phipps, M.L. (1997). Construction of adenovirus vectors through Cre-lox recombination. J Virol *71*, 1842-1849.

Harlow, E., Franza, B.R., Jr., and Schley, C. (1985). Monoclonal antibodies specific for adenovirus early region 1A proteins: extensive heterogeneity in early region 1A products. J Virol *55*, 533-546.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., *et al.* (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Jones, N., and Shenk, T. (1979). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc Natl Acad Sci U S A *76*, 3665-3669.

Klessig, D.F., Quinlan, M.P., and Grodzicker, T. (1982). Proteins containing only half of the coding information of early region 1b of adenovirus are functional in human cells transformed with the herpes simplex virus type 1 thymidine kinase gene and adenovirus type 2 DNA. J Virol *41*, 423-434.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., *et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics *29*, 1830-1831.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

Madison, D.L., Yaciuk, P., Kwok, R.P., and Lundblad, J.R. (2002). Acetylation of the adenovirus-transforming protein E1A determines nuclear localization by disrupting association with importin-alpha. J Biol Chem *277*, 38755-38763.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Montell, C., Courtois, G., Eng, C., and Berk, A. (1984). Complete transformation by adenovirus 2 requires both E1A proteins. Cell *36*, 951-961.

Pellegrini, M., and Ferrari, R. (2012). Epigenetic analysis: ChIP-chip and ChIP-seq. Methods Mol Biol *802*, 377-387.

Ramos, Y.F., Hestand, M.S., Verlaan, M., Krabbendam, E., Ariyurek, Y., van Galen, M., van Dam, H., van Ommen, G.J., den Dunnen, J.T., Zantema, A., *et al.*

(2010). Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Res *38*, 5396-5408.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.

Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E., Belmont, A.S., and van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. Mol Cell Biol *25*, 4552-4564.

Zhang, Q., Yao, H., Vo, N., and Goodman, R.H. (2000). Acetylation of adenovirus E1A regulates binding of the transcriptional corepressor CtBP. Proc Natl Acad Sci U S A *97*, 14323-14328.

Zhao, H., Dahlo, M., Isaksson, A., Syvanen, A.C., and Pettersson, U. (2012). The transcriptome of the adenovirus infected cell. Virology *424*, 115-128.

Chapter 3

A unique epigenetic signature
Is associated with active DNA replication loci
in human embryonic stem cells

This chapter was originally published in *Epigenetics*, 2014 Feb;9(2):257-67. with supplemental material. This study performed BrdUseq by labeling actively replicating DNA with BrdU followed by high throughput sequencing in both H1 human embryonic stem (ES) cells and IMR90 human primary lung fibroblast cells. By comparing epigenetic histone modification patterns with BrdUseq results, BrdU peaks common between all cell lines were found to associate with H3K18ac, H3K56ac, and H4K20me1 histone marks in human ES cells. However, the same histone modifications did not correlate with BrdU peaks in normal fibroblasts or cancer cell lines. Furthermore, by knocking down EP300 and CBP in human ES cells, the number of BrdU incorporating peaks decreased drastically suggesting that epigenetic signature found in ES cells might be required for normal replication in ES cells.

RESEARCH PAPER

Epigenetics 9:2, 257–267; February 2014; © 2014 Landes Bioscience

A unique epigenetic signature is associated with active DNA replication loci in human embryonic stem cells

Bing Li[1], Trent Su[1,2], Roberto Ferrari[1,3,†], Jing-Yu Li[1,3], and Siavash K Kurdistani[1,3,4,5]

[1]Department of Biological Chemistry; University of California; Los Angeles, CA USA; [2]Division of Oral Biology and Medicine; School of Dentistry; University of California; Los Angeles, CA USA; [3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research; David Geffen School of Medicine; University of California; Los Angeles, CA USA; [4]Department of Pathology and Laboratory Medicine; University of California; Los Angeles, CA USA; [5]Molecular Biology Institute; University of California; Los Angeles, CA USA

[†]Current Address: Department of Molecular, Cellular and Developmental Biology; University of California; Los Angeles, CA USA

Keywords: epigenetics, histone modifications, chromatin, DNA replication, embryonic stem cells, histone acetyltransferase, P300, CBP

Abbreviations: BrdU, 5-bromo-2′-deoxyuridine; CEAS, cis-regulatory element annotation system; FDR, false discovery rate; GEO, Gene Expression Omnibus; GREAT, Genomic Regions Enrichment of Annotations Tool; HAT, histone acetyltransferase; hESCs, human embryonic stem cells; iPS cells, induced pluripotent stem cells; KD, knockdown; LADs, lamina-associated domains; MCM, mini chromosome maintenance; NS, nascent strand; rmsk, repeat masker; ORC, origin recognition complex; ORCA, origin recognition complex associated; TSS, transcription start sites

The cellular epigenetic landscape changes as pluripotent stem cells differentiate to somatic cells or when differentiated cells transform to a cancerous state. These epigenetic changes are commonly correlated with differences in gene expression. Whether active DNA replication is also associated with distinct chromatin environments in these developmentally and phenotypically diverse cell types has not been known. Here, we used BrdU-seq to map active DNA replication loci in human embryonic stem cells (hESCs), normal primary fibroblasts and a cancer cell line, and correlated these maps to the epigenome. In all cell lines, the majority of BrdU peaks were enriched in euchromatin and at DNA repetitive elements, especially at microsatellite repeats, and coincided with previously determined replication origins. The most prominent BrdU peaks were shared between all cells but a sizable fraction of the peaks were specific to each cell type and associated with cell type-specific genes. Surprisingly, the BrdU peaks that were common to all cell lines were associated with H3K18ac, H3K56ac, and H4K20me1 histone marks only in hESCs but not in normal fibroblasts or cancer cells. Depletion of the histone acetyltransferases for H3K18 and H3K56 dramatically decreased the number and intensity of BrdU peaks in hESCs. Our data reveal a unique epigenetic signature that distinguishes active replication loci in hESCs from normal somatic or malignant cells.

Introduction

Patterns of histone modifications and DNA methylation create epigenetic signatures on chromatin that are associated with and may regulate many molecular activities including gene expression and DNA replication. DNA replication initiates at specific regions of the genome known as replication origins.[1-3] In *S. cerevisiae*, replication origins have sequence specificity but additional genetic or chromatin features are required to define functional origins.[4] In multicellular organisms, however, no clear sequence specificity is found for replication origins but certain genetic elements, such as CpG islands and G-quadruplex DNA motifs, are preferentially associated with origins.[5] It has been proposed that replication origins may also be associated with specific epigenetic states.[6] Early activated origins tend to localize in actively transcribed, hyperacetylated chromatin regions while late activated origins are found mainly in heterochromatin with hypo-acetylated histones. In mammalian cells, histone acetyltransferase (HAT) HBO1 mediates H4 acetylation of histones surrounding origins in a cell cycle-dependent manner and promotes loading of MCM complex onto chromatin.[7] Histone H3 lysine 56 acetylation (H3K56ac) also plays roles in DNA replication in both yeast and multicellular organisms. In yeast, H3K56ac is catalyzed by Rtt109, which has genetic interactions with multiple replication machinery proteins including ORC2, CDC45, PCNA, and DNA polymerase α.[8-10] In Drosophila and

*Correspondence to: Siavash K Kurdistani; Email: skurdistani@mednet.ucla.edu
Submitted: 08/26/2013; Revised: 10/08/2013; Accepted: 10/18/2013; Published Online: 10/29/2013
http://dx.doi.org/10.4161/epi.26870

Table 1. Summary of BrdU peaks and blocks

Cells	Total peaks	FDR	Peaks overlapping with NS DNA	Peaks mapped to repeat sequences	Total blocks	Coverage (Mbp)	Peaks in blocks
H1	5086	2.04%	3417 (67.18%)	3850 (75.7%)	296	321 (10.38%)	3062 (60.2%)
IMR90	5743	2.54%	4774 (83.13%)	4063 (70.75%)	282	358 (11.56%)	4294 (74.77%)
Saos-2	6272	2.87%	N/A	5250 (83.71%)	265	373 (12.03%)	4533 (72.27%)

NS, nascent strand

human cells, H3K56ac may promote the packaging of newly synthesized DNA into chromatin in DNA repair and replication.[11] In yeast, the Rpd3 histone deacetylase delays replication origin firing by deacetylating histones around replication origins. Targeting of H3K18ac and H2Bac to a late activated origin promotes loading of replication factor CDC45 and leads to its earlier activation, while deletion of *RPD3* promotes H3K18 and H4 acetylation and activation or increased firing efficiency at origins.[12] In humans, H4K20me2/3 recruit ORC to chromatin through direct binding with ORC1 and ORCA.[13,14] Certain histone modifications also correlate with replication timing and ORC binding in *Drosophila*.[15]

In human cells, the genome-wide relationship between histone modifications and DNA replication has been less understood. It has been unclear whether the DNA replication profile and the associated histone modifications are comparable in cells that are phenotypically different. In this study, we generated DNA replication profile in asynchronously growing H1 hESCs, normal human fetal IMR90 lung fibroblasts and Saos-2 osteosarcoma cells by 5-bromo-2′-deoxyuridine (BrdU) DNA immunoprecipitation followed by massively-parallel sequencing (BrdU-seq).[16] We found that the BrdU peaks overlap significantly with replication origins that have high activity as defined by nascent strand (NS) DNA sequencing,[5] indicating that BrdU peaks represent active DNA replication loci including a subset of active origins. The distributions of BrdU peaks in all 3 cell types were by and large similar, with majority of the peaks enriched in DNA repetitive elements within introns and intergenic regions in early replicating regions. But a considerable number of replication loci were unique to each cell type and associated with cell type-specific genes. In hESCs the histone marks H3K18ac, H3K56ac, and H4K20me1 were preferentially present at regions that had highest levels of BrdU incorporation, were highly conserved in vertebrates and were also sites of replication in differentiated cells. While BrdU was still incorporated in the same regions in fibroblasts and cancer cells, the histone modifications were absent. Finally, knockdown (KD) of EP300 and CREBBP HATs in hESCs led to a global decrease in H3K18 and H3K56 acetylation and in number and intensity of BrdU peaks, indicating a role for these HATs and their acetylated substrates in DNA replication in hESCs. The association of specific histone marks with active DNA replication loci in hESCs reveals an epigenetic signature that uniquely distinguishes pluripotent DNA replication from that of differentiated or cancerous cells.

Results

Global BrdU incorporation pattern delineates active replication loci in human cells and overlaps with a subset of known replication origins

To determine whether the chromatin environment associated with DNA replication changes during cellular differentiation, we first mapped the locations of BrdU incorporation across the human genome by BrdU-seq in control and EP300 and CREBBP KD H1 hESCs as well as normal human lung fibroblast IMR90 cells and osteosarcoma Saos-2 cancer cells which have a high degree of aneuploidy. To enable a direct comparison of BrdU incorporation and epigenetic modifications that were generated from asynchronously growing cells,[17-19] we performed BrdU-seq also in asynchronous cell populations. We verified the specificity of the anti-BrdU antibody by flow cytometry and BrdU dot blot and its utility in immunoprecipitation by showing that significant amounts of DNA were immunoprecipitated from cells in S phase but not in G1 phase of the cell cycle (**Figs. S1A and B**). Input and immunoprecipitated DNA from each experiment was sequenced, and the obtained reads were aligned uniquely to the human genome (hg19) allowing for up to two mismatches (**Table S1**). We applied a stringent computational criterion to define BrdU enriched regions with high confidence. We divided the human genome into 100 bp windows and calculated a *P* value for Poisson distribution of enriched immunoprecipitated DNA relative to input for each window. Significant peaks were defined as those windows with a *P* value < 10^{-4} and with 2 neighboring windows at the same significance. Based on these criteria, we identified 5086, 5743, and 6272 high confident BrdU peaks in hESCs, IMR90 and Saos-2 cells respectively (**Table 1**) with false discovery rates (FDR) < 3%. A less restrictive *P* value < 10^{-3} identified approximately twice as many peaks for each cell line with similar distributions but the FDR was >5% (**Table S2**). We therefore performed subsequent analyses with the peaks defined at *P* value < 10^{-4}.

While NS DNA analysis is the method of choice for identification of replication origins,[5] the BrdU peaks identified in our study showed high concordance with published NS DNA distribution in H9 hESCs and IMR90 cells. Genome-wide, 67.2% and 83.1% of BrdU peaks in H1 hESCs and IMR90 cells respectively overlapped with NS DNA peaks (**Table 1**; no NS data are available for Saos-2 cells). **Figure 1A** shows the pattern of BrdU incorporation (red track) at a representative region of the genome for hESCs, IMR90 fibroblasts and Saos-2 cells. The peaks of BrdU coincided with elevated sequence tag counts in the input genomic DNA (black track) and with peaks of NS

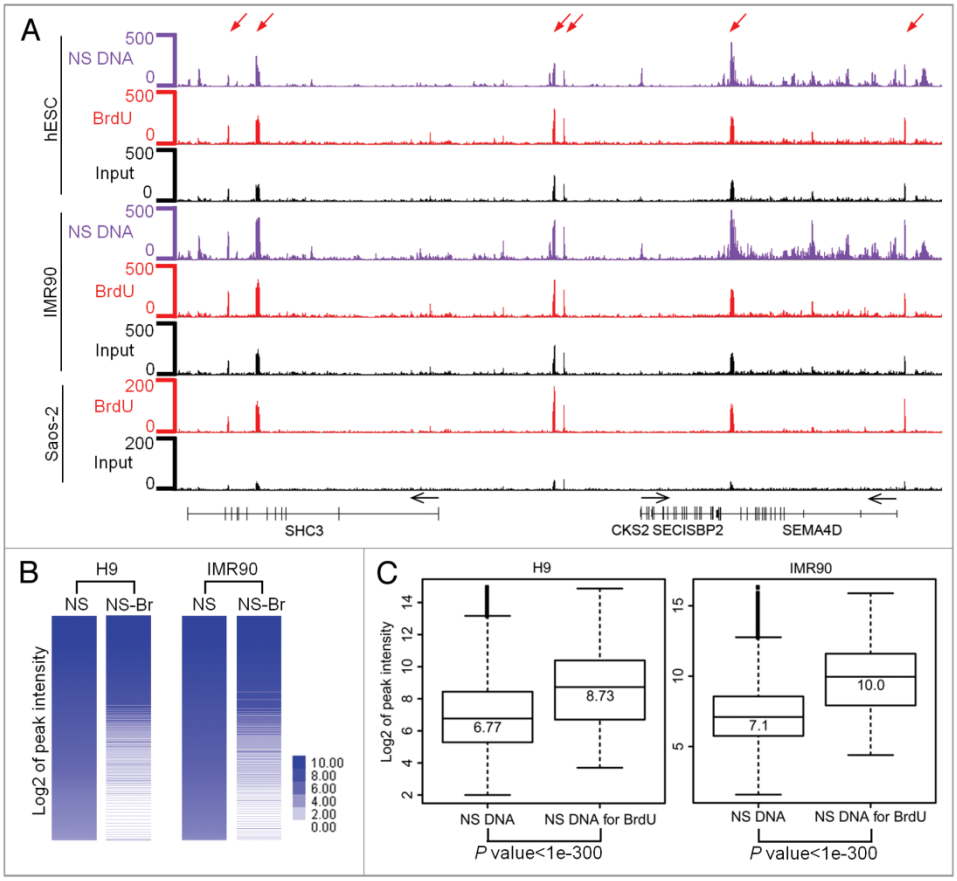


Figure 1. BrdU peaks correspond to actively replicating origins in human cells. (**A**) Genome browser views of NS DNA, BrdU peaks and input DNA tracks along a representative region of chromosome 9. No NS DNA data were available for Saos-2 cells. Red arrows indicate peaks that occurred in all cell lines and also mapped to microsatellite repeats. Black arrows indicate direction of transcription of the indicated genes. (**B**) Heat maps of nascent strand DNA abundance at replication origins in H9 hESCs and IMR90 cells and the corresponding NS DNA value for the BrdU peaks identified (NS-Br). (**C**) Box plots show the distributions of nascent strand DNA values at replication origins in H9 hESCs and IMR90 cells and the corresponding values for BrdU peaks in this study. The thick horizontal line in each box represents the median value which is indicated. *P* values for 2-sample Kolmogorov–Smirnov test are shown.

DNA (purple track). Furthermore, BrdU-seq peaks identified the highly active origins. This was evident when we centered and sorted NS DNA peaks based on their abundance and calculated the NS DNA value for the corresponding BrdU-seq peaks. The identified BrdU peaks coincided with the upper portion of NS DNA distribution (**Fig. 1B**) and had higher NS DNA values in both H1 hESCs and IMR90 cells (**Fig. 1C**). This is because for BrdU peaks to meet the significance criterion, the number of sequence reads from immunoprecipitated DNA has to be significantly enriched relative to input DNA which, at sites of DNA replication, contains higher sequence reads compared with the genome average (**Fig. S1C**, left panel). This constraint does not apply to NS DNA analysis. Thus, the locations where NS DNA is detected but BrdU is not preferentially incorporated are not considered as BrdU peaks (**Fig. S1C**, right panel). But when no statistical cutoff is applied, the total number of sequence reads in BrdU immunoprecipitated sample parallels those measured by NS DNA (**Fig. S1D**). Taken together, these data indicate that the BrdU peaks significantly overlap with highly active replications origins. However, since BrdU is incorporated in regions of the genome that are being replicated at the time of the BrdU pulse, our BrdU-seq data identifies regions of active DNA replication regardless of whether these regions function as replication origins.

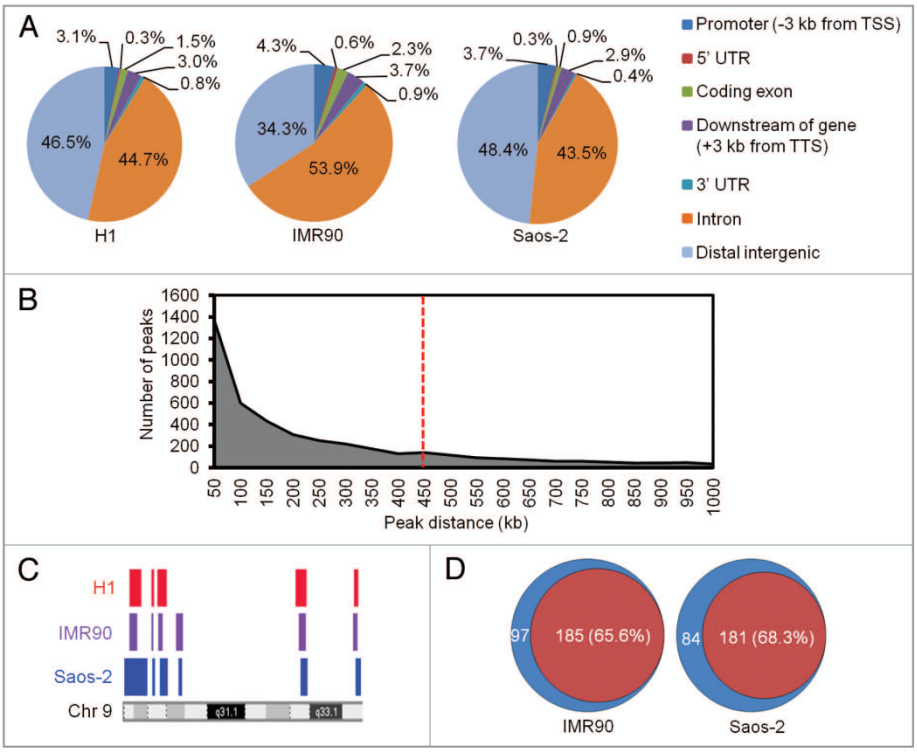


Figure 2. The majority of BrdU peaks occur in proximity of each other. (**A**) Genomic distribution of BrdU peaks in each cell line is shown as pie charts. Distal intergenic regions are defined as being at least 3 kb away from the start and end of genes. (**B**) The chart shows the distribution of distances between consecutive BrdU peaks. Red dashed line indicates the 450 kb cutoff which includes >70% of the BrdU peaks. (**C**) Locations of BrdU blocks as colored bars along a region of chromosome 9 are shown for each of the three cell types. (**D**) The Venn diagrams show the extent of overlap of BrdU blocks between IMR90 or Saos-2 cells and H1 hESCs. Red, common blocks; blue, unique blocks.

Genome-wide distribution of BrdU peaks

Analysis of BrdU peak distribution by the cis-regulatory element annotation system (CEAS) software[20] revealed similar genomic distributions in the 3 cell types with most enrichment in intron and distal intergenic regions. BrdU peak distribution in hESCs was more similar to Saos-2 cancer cells than IMR90 fibroblasts (**Fig. 2A**). Greater than 70% of BrdU peaks in all cell lines occurred at DNA repetitive elements as defined by repeat masker (rmsk) (**Table 1**).[21] Of the numerous types of repetitive elements, microsatellite repeats especially $(TG)_n$ repeats contained the majority of BrdU peaks (**Table S3**). The sequenced reads were uniquely aligned to these loci due to the polymorphisms present in microsatellites.[22] DNA motif analysis also revealed that BrdU peaks were enriched at TG- and G-rich sequences in all cell types (**Fig. S2A**).

The BrdU peaks were not randomly distributed throughout the genome but clustered mostly in vicinity of each other. Plotting the distribution of distances between consecutive BrdU peaks indicated that more than 70% of peaks have distances smaller than 450 kb (**Fig. 2B**). This is consistent with the fact that the origins of replication occur in DNA replicons ranging from 50–450 kb in length.[23] We therefore scanned the genome using a moving 450-kb window to define BrdU blocks of enrichment. We identified >250 BrdU blocks in each cell line, covering over 320 Mb of the genome and accounting for the majority of BrdU peaks (**Table 1**). The blocks showed a high degree of overlap with 65.6% and 68.3% of blocks in IMR90 and Saos-2 cells respectively overlapping with those in hESCs. Thus a large subset of BrdU blocks were conserved between hESCs and differentiated normal and cancer cells (**Fig. 2C and D**).

The BrdU blocks were located in early replicating euchromatin (**Fig. S2B–D**) and depleted from lamina-associated domains (LADs), heterochromatic regions that replicate late in S phase (**Fig. S2E and F**).[24] The blocks also showed differential methylation status in ESCs vs. IMR90 fibroblasts. The DNA methylation level of BrdU blocks in hESCs was not significantly different from randomly selected regions while the BrdU blocks contained significantly more DNA methylation than might be expected

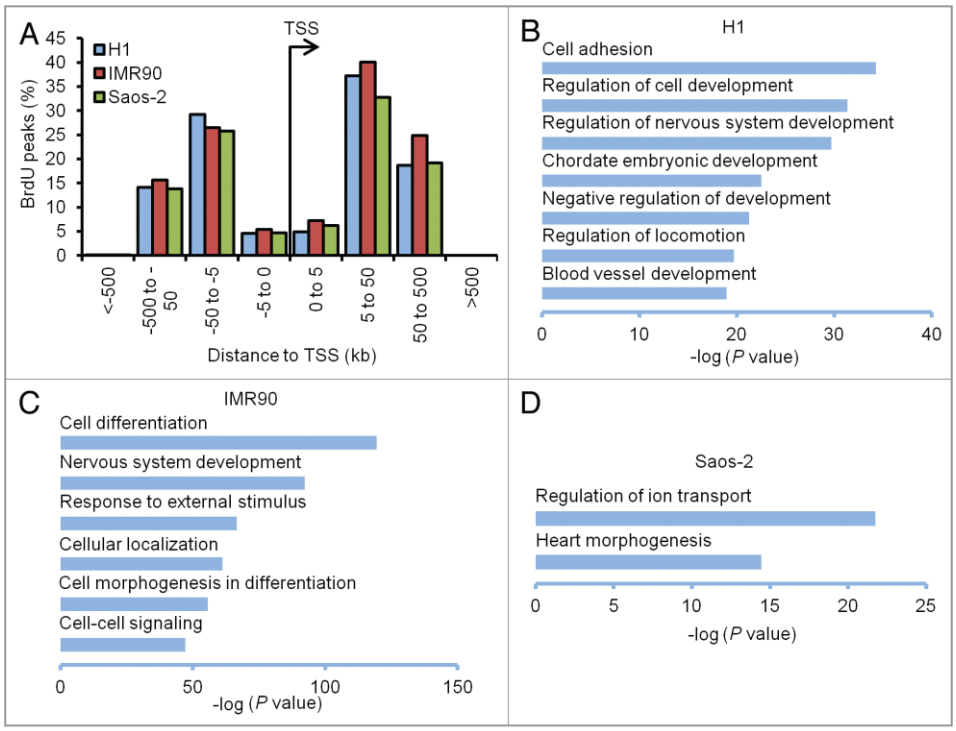


Figure 3. BrdU peaks are associated with cell type-specific genes. (**A**) The bar chart shows the distribution of the distances of BrdU peaks from TSS. (**B–D**) GO analysis of BrdU peaks in each cell line as indicated. Bars represent -$\log_{10}$ of the binomial raw *P* values.

based on background levels in IMR90 cells (binomial *P* value < 1e-300; **Fig. S2G**).[17] We found a similar relationship between DNA methylation and the published early replicating regions in BG02 hESCs and BJ fibroblasts (data not shown).[25] Thus, the early replicating regions in differentiated fibroblasts are preferentially enriched for DNA methylation which is not the case in hESCs.

BrdU peaks are associated with cell type-specific genes

The majority of BrdU peaks was found ± 5 kb or further away from transcription start sites (TSS) in all three cell lines (**Fig. 3A**). We therefore used the Genomic Regions Enrichment of Annotations Tool (GREAT)[26] to determine the genes associated with distant peaks of BrdU. In hESCs, BrdU peaks occurred near genes that are involved in early development and maintenance of pluripotency such as OCT4, NODAL, and BMPR1A (**Fig. 3B**; **Fig. S3**). In contrast, BrdU peaks associated with genes that function in differentiation and morphogenesis in IMR90 fibroblasts such as HES1 and ACTB. In Saos-2 cells, BrdU peaks were linked to ion channels and transporters (**Fig. 3C** and **D**), which are involved in cancer progression.[27,28] Therefore, BrdU peaks tend to associate with genes that are functionally relevant in each cell type.

Co-occupancy of histone H3K18ac, H3K56ac and H4K20me1 with BrdU peaks in hESCs

Since histone acetylation plays a role in DNA replication,[7,12,15] we compared the distribution of histone acetylation marks to BrdU peaks. As shown in **Figure 4A and B** for a representative region on chromosome 18 and genome-wide, respectively, there were significant overlaps between H3K18ac and H3K56ac with BrdU peaks only in hESCs. Distribution of H3K56ac in H9 hESCs also overlapped with BrdU peaks in H1 hESCs, indicating this co-occupancy may be a general feature of hESCs (**Fig. 4A**). We also found strong occupancy of H4K20me1 at BrdU peaks in hESCs, consistent with a role for H4K20 methyltransferase, PR-SET7, in regulation of certain replication origins.[29] Globally, 23.4%, 28%, and 50.5% of BrdU peaks in hESCs had H3K18ac, H3K56ac, and H4K20me1 nearby while only 9%, 6.55%, and 13.2% of BrdU peaks in IMR90 cells were associated with these marks, respectively. Other histone acetylation sites such as H3K9ac and H3K27ac showed somewhat opposite patterns with no enrichment in hESCs and slight enrichment in fibroblasts' BrdU incorporating regions, indicating the high specificity of H3K18ac and H3K56ac co-occupancy with BrdU peaks in hESCs (**Fig. 4B**; **Fig. S4A**).

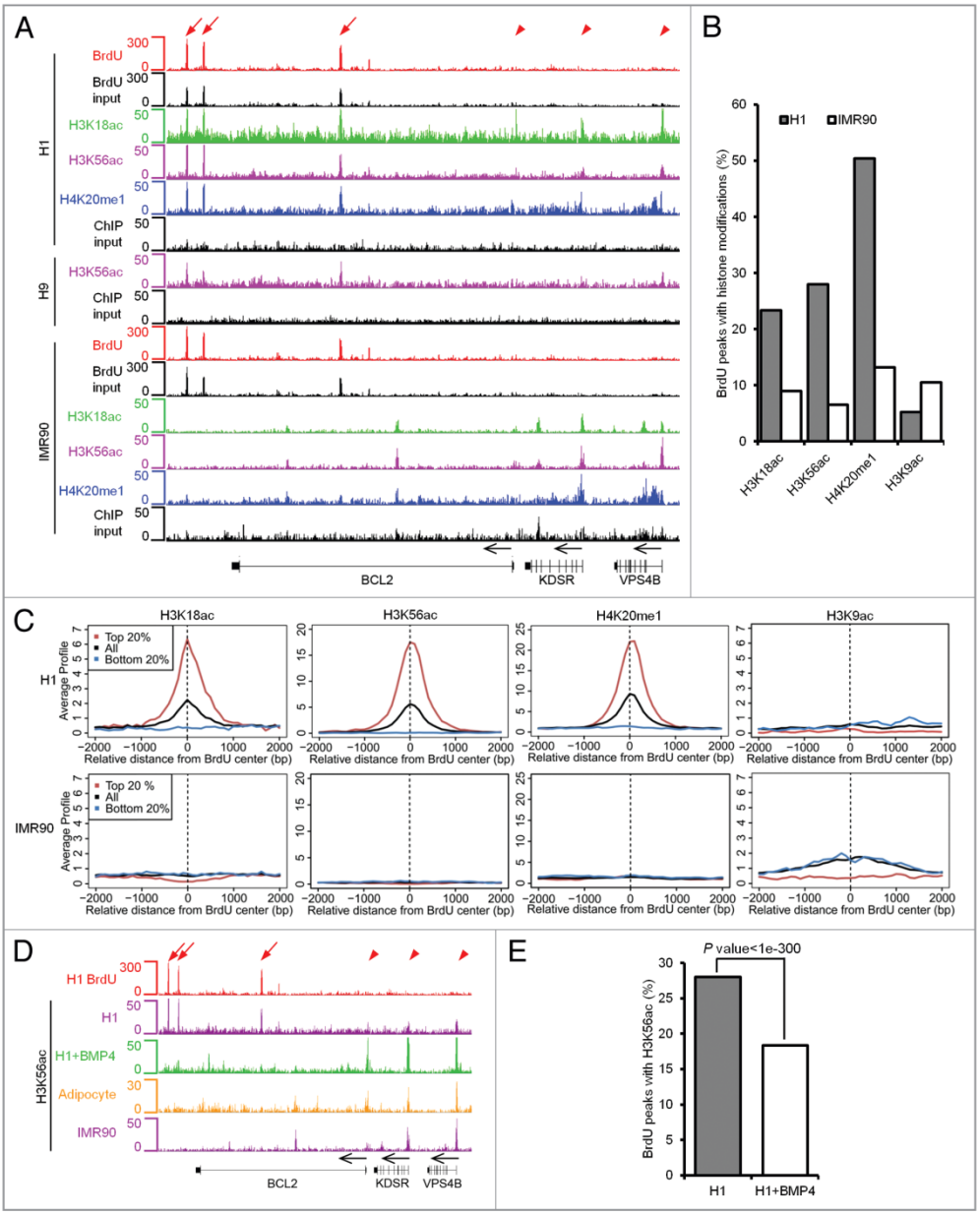


Figure 4. Co-occupancy of BrdU peaks with histone modifications is ESCs specific. (**A**) Shown are genome browser views of the indicated features along a representative region of chromosome 18. Red arrows indicate BrdU peaks that overlap with histone modifications. Arrowheads indicate the histone modification peaks at the TSS. Black arrows indicate direction of transcription. (**B**) The percentage of BrdU peaks that overlap with histone modifications. (**C**) The average levels of the indicated histone modifications (significant reads counts) relative to the center of BrdU incorporating regions are shown for the top and bottom 20% according to the BrdU signal intensity in comparison to all BrdU incorporating regions. (**D–E**) Levels of H3K56ac at BrdU incorporating regions in H1, BMP4-induced differentiated H1 cells and adipocytes are shown for (**D**) a representative locus on chromosome 18 and (**E**) genome-wide.

To further characterize the relationship between BrdU peaks and H3K18ac, H3K56ac, and H4K20me1, we grouped genomic regions that have BrdU peaks based on peak intensity and centered all regions. We then plotted the average levels of histone modifications in hESCs and IMR90 cells within ± 2 kb of each BrdU region center. As shown in **Figure 4C** and **Figure S4A**, the regions with highest levels of BrdU incorporation (top 20%) also had higher levels of H3K18ac, H3K56ac, and H4K20me1, but not H3K9ac, H3K27ac, or H4K5ac in hESCs. In IMR90 cells, there was no preferential enrichment of H3K18ac, H3K56ac, and H4K20me1 within BrdU labeled regions. When we analyzed published H3K56ac data sets in adipocytes and BMP4-induced mesendoderm cells differentiated from H1 hESCs, we also observed loss of H3K56ac at replication origins upon differentiation. This was not due to complete absence of H3K56ac as new peaks of H3K56ac appeared at similar places in mesendoderm as in IMR90 cells (**Fig. 4D**; **Fig. S4B**). Globally, the number of BrdU peaks that overlapped with H3K56ac was dramatically reduced upon differentiation (binomial *P* value < 1e-300, **Fig. 4E**). Similarly, we did not detect enrichment of H4K20me1 levels in HeLa S3 and K562 leukemia cells at the genomic regions that in H1 hESCs the BrdU peaks overlapped with H4K20me1 (**Fig. S4B**). Taken together, our analysis shows that H3K18ac, H3K56ac, and H4K20me1 are enriched at the regions of active DNA replication mainly in hESCs and correlate positively with the intensity of BrdU peaks.

The H1-IMR90-Saos-2 shared peaks of BrdU are associated with high level of histone marks in hESCs and conserved DNA sequences across vertebrates

To rule out that the association of histone modifications with DNA replication occur at H1-specific BrdU peaks, we divided the BrdU peaks into those that are shared between H1, IMR90, and Saos-2 cells and those that are unique to H1 (1167 vs 2813 BrdU peaks, respectively). The average levels of BrdU intensity was much higher in shared regions than in H1 specific regions. The shared regions also had higher levels of H3K18ac, H3K56ac, and H4K20me1 while H1 specific regions showed low levels of histone modifications (**Fig. 5A**). Gene ontology analysis indicated that the shared peaks were associated with differentiation genes while H1 unique peaks were linked to embryonic development genes (**Fig. 5B**). Furthermore, the shared BrdU regions had higher degree of DNA sequence conservation among vertebrates compared with H1-specific BrdU peaks (**Fig. 5C**). Thus the shared BrdU regions account for most of the enrichment observed for the associated histone modifications, indicating that patterns of histone modifications are indeed different at the same replicating regions between hESCs and differentiated cells.

EP300/CREBBP depletion decreases global BrdU incorporation in hESCs

EP300 and CREBBP are the main HATs for H3K18 and H3K56.[11,30] We therefore tested whether EP300 and CREBBP were required for DNA replication in hESCs. Since EP300 and CREBBP are largely redundant in acetylation of H3K18,[30] we co-transfected siRNAs against both genes into hESCs and performed BrdU-seq (**Table S1**). As shown in **Figure 6A**, EP300 and CREBBP protein levels decreased ~85–90% upon knockdown. Levels of H3K18ac and H3K56ac also decreased dramatically after knockdown, indicating that EP300/CREBBP were responsible for the bulk of H3K18ac and H3K56ac in hESCs.

The number of BrdU peaks decreased from 5086 in control siRNA to 1755 in EP300/CREBBP KD hESCs. Depletion of EP300/CREBBP also significantly decreased BrdU blocks from 296 to 33. The remaining BrdU peaks after KD were similarly distributed across the genome with 42.2% and 49.5% of peaks in introns and distal intergenic regions, respectively (**Fig. 6B**). The majority (64%) of the BrdU peaks in EP300/CREBBP KD cells still occurred within DNA repeats (**Table S3**) and were associated with negative regulators of cell proliferation such as cyclin-dependent kinase inhibitors, genes involved in cell morphogenesis and cell fate commitment (**Fig. 6C**). Among the 1755 BrdU peaks in KD cells, we detected 1489 new peaks compared with control. However, a closer examination of these 1489 regions revealed smaller BrdU peaks in the control sample, which had not reached our statistical threshold for significance to be included in the analysis (**Fig. 6D**). Globally, the level of BrdU incorporation over the newly detected regions was much lower in control KD cells than in EP300/CREBBP KD cells (**Fig. 6E**), suggesting that these regions were being replicated by a minority of hESCs prior to EP300/CREBBP KD. These data reveal that depletion of EP300/CREBBP significantly decreases BrdU peaks and block formation and changes genome-wide replication pattern, revealing an essential role for these HATs in regulation of DNA replication in hESCs.

Discussion

We have compared active DNA replication patterns in pluripotent stem cells, the fully differentiated IMR90 primary normal fibroblasts and the Saos-2 cancer cell line. Our genome-wide BrdU maps overlapped partially with the early origins of DNA replication as determined previously by nascent strand detection and early timing of replication.[5,25] The genomic regions with higher levels of NS DNA also showed preferential BrdU incorporation demonstrating that BrdU peaks overlap with highly active replication origins. Overall the distribution of BrdU peaks were similar between all cell types indicating that DNA replication is a highly conserved process even in cells with vastly different phenotypes or with significant chromosomal abnormalities such as Saos-2 cells.[31] In all cell types, the bulk of BrdU peaks were at $(TG)_n$ dinucleotide microsatellite repeats and poly G sequences. A similar trend for significantly overrepresented sequences at replication origins in Drosophila and mouse cells has also been observed.[32]

Despite the similarity of BrdU incorporation patterns in the 3 cell types, the epigenetic marks at or around the BrdU peaks were quite distinct. We observed significant enrichment of H3K18ac, H3K56ac, and H4K20me1 at BrdU peaks specifically in hESCs but not at the same replicating regions in fibroblasts or cancer cells, suggesting that active DNA replication is associated with distinct epigenetic marks in hESCs. While certain histone

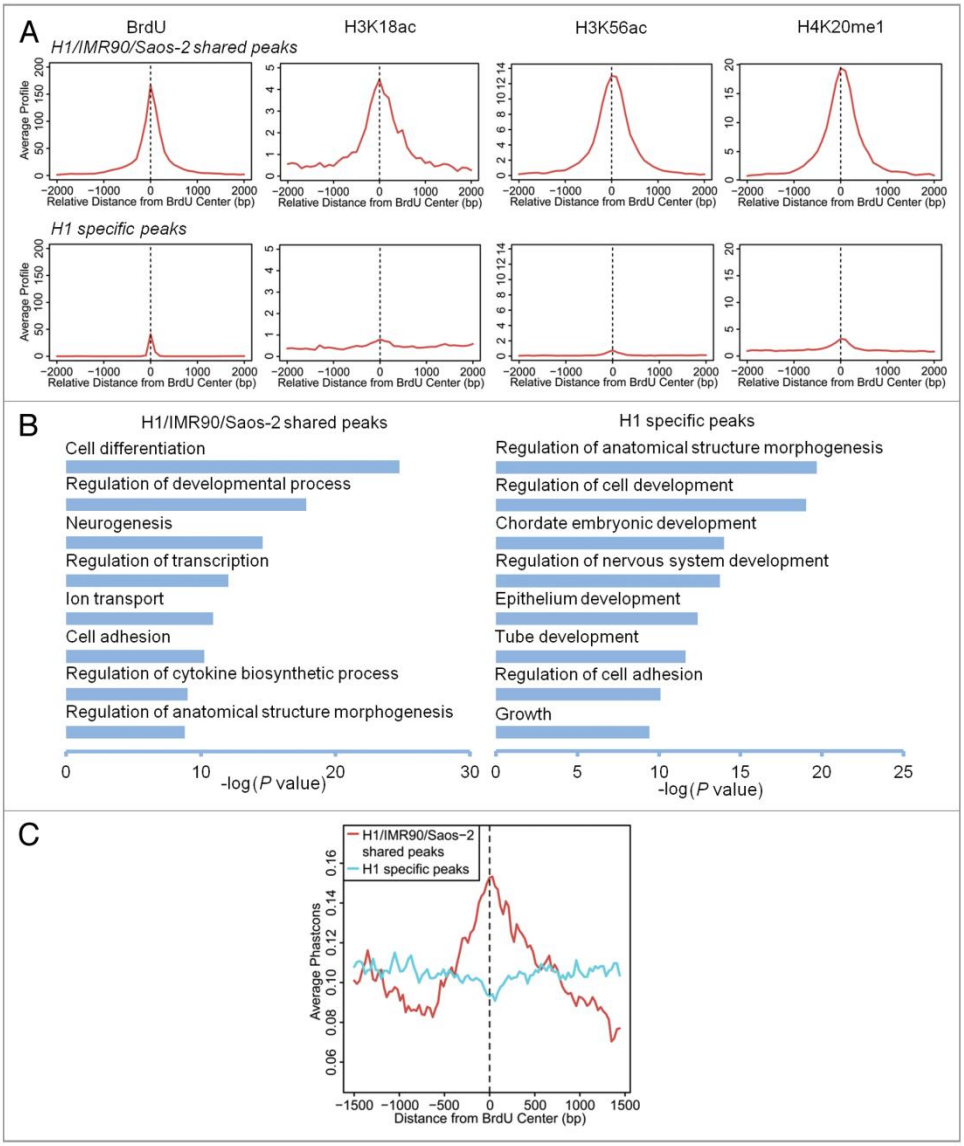


Figure 5. The BrdU peaks shared by the three cell types are preferentially associated with histone marks in hESCs and conserved DNA sequences. (**A**) Average levels of BrdU incorporation and histone modifications in hESCs at the center of BrdU peaks that are shared between H1, IMR90 and Saos-2 cells or that are unique to H1 cells. (**B**) GO analysis of shared (left panel) and H1-specific (right panel) BrdU peaks. (**C**) Conservation plots of shared and H1-specific BrdU peaks.

marks have been previously associated with DNA replication origins,[12,15,29] the cell type specificity of such association is unexpected. Several possibilities may underlie these observations. First, the same histone marks may still associate with DNA replication in fibroblasts but only transiently in a narrower window of cell cycle which would be missed when examining asynchronously growing cells. Second, these histone marks may have been erased at some point in the differentiation lineage from hESCs to fibroblasts or

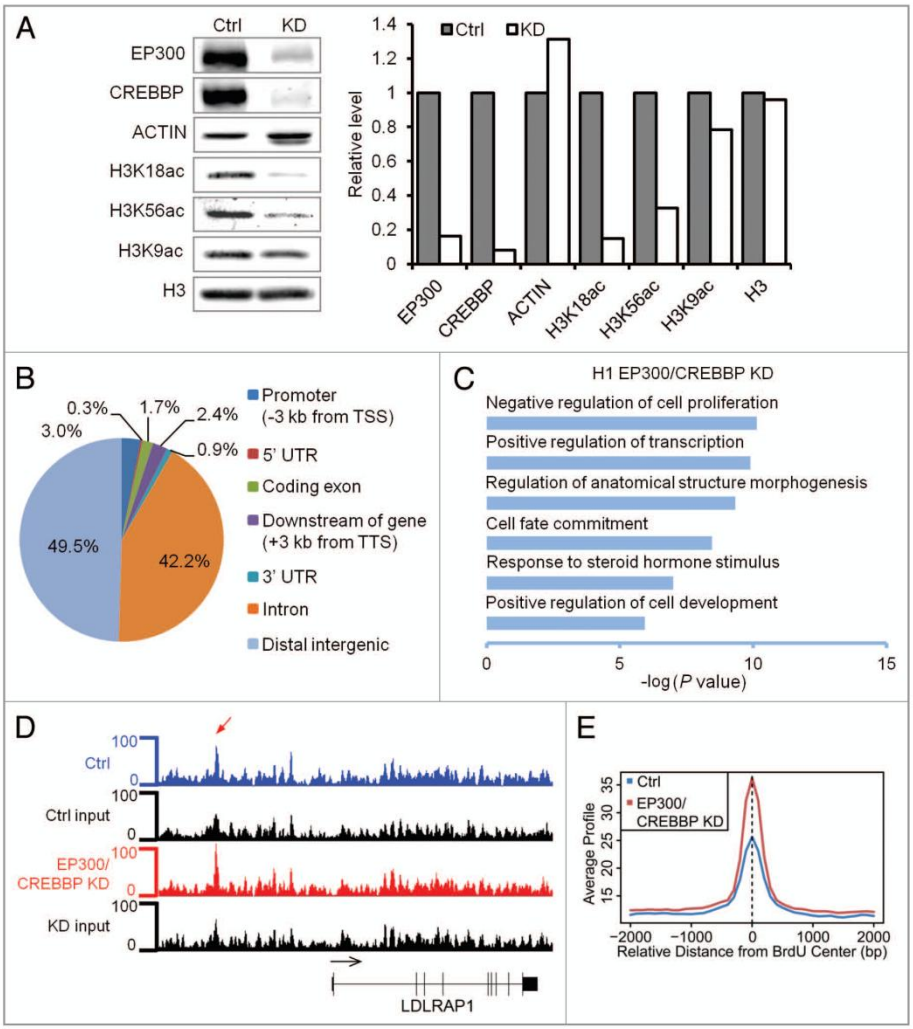


Figure 6. EP300 and CREBBP HATs are required for normal pattern of BrdU incorporation in hESCs. (**A**) Western blots and signal quantifications of the indicated factors are shown for control (ctrl) and KD of EP300/CREBBP in H1 hESCs. (**B**) Distribution of BrdU peaks in KD cells is shown as a pie chart. (**C**) GO analysis of BrdU peaks in KD cells. (**D**) Enhanced BrdU incorporation upstream of LDLRAP1 gene on chromosome 1 is detected in KD relative to ctrl H1 hESCs. Red arrow points to the BrdU peak. (**E**) Genome-wide average of BrdU incorporation (all tag counts) at regions with newly detected BrdU peaks in EP300/CREBBP KD relative to ctrl H1 cells.

replaced by other epigenetic modifications. It was shown recently that H3K56me1 regulates DNA replication in HeLa cells through direct interaction with PCNA.[33] Since methylation and acetylation of lysines are mutually exclusive, H3K56ac in hESCs may be replaced by H3K56me1 in more differentiated cells. Third, hESCs and fibroblasts may represent classes of cell types with different epigenetic requirements for active replication. Examination of additional cell types may reveal categories of cells that have similar replication-associated epigenomic profiles as hESCs. For instance, in U2OS osteosarcoma cell line, PR-Set7 mediates H4K20 monomethylation of histones at a limited number of examined replication origins at the onset of replication origin licensing.[29] This may also be the case in Saos-2 osteosarcoma cells but no genome-wide data set on H4K20me1 is available in either cancer cell line. Further genome-wide studies of replication origins and epigenetic marks will be required to resolve these possibilities.

Whether the histone marks play a direct role in DNA replication in hESCs remain to be established. Knockdown of EP300/CREBBP significantly decreased H3K18ac, H3K56ac, and BrdU incorporation but since EP300 and CREBBP have many cellular activities[34] and interact with and acetylate other replication machinery proteins,[35,36] our data cannot unequivocally link histone acetylation to regulation of DNA replication in hESCs. Knockdown of these enzymes may also result in loss of hESCs pluripotency with subsequent effects on DNA replication. Indeed, the KD cells showed morphological changes associated with differentiation (data not shown). Nonetheless, ectopic targeting of histone acetyltransferases to a replication origin stimulates the origin's activity[12,37] and H4 methylation can help recruit ORC to chromatin.[13,14] Whatever the case may be, our data reveal that the global association of DNA replication with epigenetic marks is dependent on the cell type. It will be of interest to determine whether the epigenetic profiles of replication origins in iPS cells are similar to ESCs or to the cell types from which iPS cells were derived.

Materials and Methods

Cell culture

H1 hESCs were obtained from UCLA embryonic stem cell bank and maintained with mTeSR1 Basal Medium (STEMCELL Technologies) at 37 °C in 5% CO_2. IMR90 and Saos-2 cells were purchased from ATCC. Cells were maintained at 37 °C in 5% CO_2 with DMEM containing 10% FBS and penicillin-streptomycin.

BrdU labeling and genomic DNA purification

Cells were pulse labeled with 30 μM BrdU for 30 min before being harvested. Genomic DNA was purified as previously described with some modification.[38] Briefly, BrdU labeled cells were suspended in lysis Buffer (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 25 mM EDTA, 1% SDS and 600 μg/ml proteinase K) and incubated at 37 °C overnight. Each sample was then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform. After extraction, DNA was precipitated with isopropanol and centrifuged for 30 min at 4 °C. The pellet was resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and treated with 10 μg/ml RNase A at 37 °C for 30 min. The DNA was further extracted with phenol:chloroform:isoamyl alcohol and precipitated with ethanol following standard procedures. Purified DNA was re-suspended in TE buffer.

Immunoprecipitation and sequencing library preparation

Chromatin immunoprecipitation was performed as previously described using rabbit anti-H3K18ac antibody.[39] The immunoprecipitation of BrdU labeled DNA was adapted from Azuara et al. with modifications.[40] Four micrograms of BrdU labeled DNA was heated in 0.1 M NaOH at 95 °C for 5 min to remove RNA primers from nascent strand DNA. HCl was then added to neutralize the pH. TE buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA) was added to bring the total volume 500 μl. DNA was fragmented by sonication and denatured by heating at 95 °C for 5 min to generate single-stranded DNA that can be identified by anti-BrdU antibody. An amount of 0.1 volume of adjusting buffer (110 mM sodium phosphate buffer, pH7.0, 1.54 M NaCl, 0.55% Triton X-100) was then added to DNA. DNA was subsequently incubated with 2 μg of anti-BrdU antibody and Dynabeads Protein G (Life Technologies). After immunoprecipitation, DNA was purified using phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation. Single-stranded DNA was quantified by Qubit Fluorometer (Life Technologies). Double-stranded DNA was generated from input and immunoprecipitated DNA by brief random priming with the Bioprime DNA Labeling System (Life Technologies) at 37 °C for 20 min. Sequencing libraries were prepared with ChIP-seq DNA Sample Prep Kit (Illumina) according to the manufacturer's instructions and sequenced using Illumina Genome Analyzer IIx or HiSeq 2000 sequencing systems at the High Throughput Sequencing Core of the UCLA Broad Stem Cell Research Center.

Gene ontology analysis

The analysis was performed with GREAT.[26] The gene regulatory domain was defined as the region that extends from TSS in both directions to the nearest gene's TSS but no more than 100 kb extension in one direction. The BrdU peaks were then overlapped with gene regulatory domains to identify associated genes.

Data access

Sequencing data have been submitted to GEO under accession number GSE43152.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Maria Vogelauer for critical reading of the manuscript and Matteo Pellegrini for help on bioinformatics analysis. We acknowledge the support of the UCLA Broad Stem Cell Research Center High-Throughput Sequencing Core Resource. Li JY was supported by a training grant of California Institute for Regenerative Medicine (CIRM) in stem cell research at UCLA. This work was funded by a CIRM grant to Kurdistani SK.

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/epigenetics/article/26870

References

1. Gilbert DM. Evaluating genome-scale approaches to eukaryotic DNA replication. Nat Rev Genet 2010; 11:673-84; PMID:20811343; http://dx.doi.org/10.1038/nrg2830
2. Stillman B. Origin recognition and the chromosome cycle. FEBS Lett 2005; 579:877-84; PMID:15680967; http://dx.doi.org/10.1016/j.febslet.2004.12.011
3. Remus D, Diffley JF. Eukaryotic DNA replication control: lock and load, then fire. Curr Opin Cell Biol 2009; 21:771-7; PMID:19767190; http://dx.doi.org/10.1016/j.ceb.2009.08.002
4. Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. Conserved nucleosome positioning defines replication origins. Genes Dev 2010; 24:748-53; PMID:20351051; http://dx.doi.org/10.1101/gad.1913210
5. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol 2012; 19:837-44; PMID:22751019; http://dx.doi.org/10.1038/nsmb.2339
6. Méchali M. DNA replication origins: from sequence specificity to epigenetics. Nat Rev Genet 2001; 2:640-5; PMID:11483989; http://dx.doi.org/10.1038/35084598
7. Miotto B, Struhl K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. Mol Cell 2010; 37:57-66; PMID:20129055; http://dx.doi.org/10.1016/j.molcel.2009.12.012
8. Suter B, Tong A, Chang M, Yu L, Brown GW, Boone C, Rine J. The origin recognition complex links replication, sister chromatid cohesion and transcriptional silencing in Saccharomyces cerevisiae. Genetics 2004; 167:579-91; PMID:15238513; http://dx.doi.org/10.1534/genetics.103.024851
9. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al. Global mapping of the yeast genetic interaction network. Science 2004; 303:808-13; PMID:14764870; http://dx.doi.org/10.1126/science.1091317
10. Han J, Zhou H, Horazdovsky B, Zhang K, Xu RM, Zhang Z. Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. Science 2007; 315:653-5; PMID:17272723; http://dx.doi.org/10.1126/science.1133234
11. Das C, Lucia MS, Hansen KC, Tyler JK. CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature 2009; 459:113-7; PMID:19270680; http://dx.doi.org/10.1038/nature07861
12. Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. Histone acetylation regulates the time of replication origin firing. Mol Cell 2002; 10:1223-33; PMID:12453428; http://dx.doi.org/10.1016/S1097-2765(02)00702-5
13. Beck DB, Burton A, Oda H, Ziegler-Birling C, Torres-Padilla ME, Reinberg D. The role of PR-Set7 in replication licensing depends on Suv4-20h. Genes Dev 2012; 26:2580-9; PMID:23152447; http://dx.doi.org/10.1101/gad.195636.112
14. Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature 2012; 484:115-9; PMID:22398447; http://dx.doi.org/10.1038/nature10956
15. Eaton ML, Prinz JA, MacAlpine HK, Tretyakov G, Kharchenko PV, MacAlpine DM. Chromatin signatures of the Drosophila replication program. Genome Res 2011; 21:164-74; PMID:21177973; http://dx.doi.org/10.1101/gr.116038.110
16. Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, Audit B, d'Aubenton-Carafa Y, Arneodo A, Hyrien O, et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res 2010; 20:447-57; PMID:20103589; http://dx.doi.org/10.1101/gr.098947.109
17. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22; PMID:19829295; http://dx.doi.org/10.1038/nature08514
18. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 2010; 6:479-91; PMID:20452322; http://dx.doi.org/10.1016/j.stem.2010.03.018
19. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 2013; 153:1134-48; PMID:23664764; http://dx.doi.org/10.1016/j.cell.2013.04.022
20. Shin H, Liu T, Manrai AK, Liu XS. CEAS: cis-regulatory element annotation system. Bioinformatics 2009; 25:2605-6; PMID:19689956; http://dx.doi.org/10.1093/bioinformatics/btp479
21. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet 2000; 16:418-20; PMID:10973072; http://dx.doi.org/10.1016/S0168-9525(00)02093-X
22. Eckert KA, Hile SE. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. Mol Carcinog 2009; 48:379-88; PMID:19306292; http://dx.doi.org/10.1002/mc.20499
23. Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, Sakaki Y, Ikemura T. Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. Hum Mol Genet 2002; 11:13-21; PMID:11772995; http://dx.doi.org/10.1093/hmg/11.1.13
24. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 2008; 453:948-51; PMID:18463634; http://dx.doi.org/10.1038/nature06947
25. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 2010; 107:139-44; PMID:19966280; http://dx.doi.org/10.1073/pnas.0912402107
26. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010; 28:495-501; PMID:20436461; http://dx.doi.org/10.1038/nbt.1630
27. Becchetti A. Ion channels and transporters in cancer. 1. Ion channels and cell proliferation in cancer. Am J Physiol Cell Physiol 2011; 301:C255-65; PMID:21430288; http://dx.doi.org/10.1152/ajpcell.00047.2011
28. Pedersen SF, Stock C. Ion channels and transporters in cancer: pathophysiology, regulation, and clinical potential. Cancer Res 2013; 73:1658-61; PMID:23302229; http://dx.doi.org/10.1158/0008-5472.CAN-12-4188
29. Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, Sardet C, Julien E. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nat Cell Biol 2010; 12:1086-93; PMID:20953199; http://dx.doi.org/10.1038/ncb2113
30. Horwitz GA, Zhang K, McBrian MA, Grunstein M, Kurdistani SK, Berk AJ. Adenovirus small e1a alters global patterns of histone modification. Science 2008; 321:1084-5; PMID:18719283; http://dx.doi.org/10.1126/science.1155544
31. Scheel C, Schaefer KL, Jauch A, Keller M, Wai D, Brinkschmidt C, van Valen F, Boecker W, Dockhorn-Dworniczak B, Poremba C. Alternative lengthening of telomeres is associated with chromosomal instability in osteosarcomas. Oncogene 2001; 20:3835-44; PMID:11439347; http://dx.doi.org/10.1038/sj.onc.1204493
32. Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res 2011; 21:1438-49; PMID:21750104; http://dx.doi.org/10.1101/gr.121830.111
33. Yu Y, Song C, Zhang Q, DiMaggio PA, Garcia BA, York A, Carey MF, Grunstein M. Histone H3 lysine 56 methylation regulates DNA replication through its interaction with PCNA. Mol Cell 2012; 46:7-17; PMID:22387026; http://dx.doi.org/10.1016/j.molcel.2012.01.019
34. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. Genes Dev 2000; 14:1553-77; PMID:10887150
35. Glozak MA, Seto E. Acetylation/deacetylation modulates the stability of DNA replication licensing factor Cdt1. J Biol Chem 2009; 284:11446-53; PMID:19276081; http://dx.doi.org/10.1074/jbc.M809394200
36. Naryzhny SN, Lee H. The post-translational modifications of proliferating cell nuclear antigen: acetylation, not phosphorylation, plays an important role in the regulation of its function. J Biol Chem 2004; 279:20194-9; PMID:14988403; http://dx.doi.org/10.1074/jbc.M312850200
37. Goren A, Tabib A, Hecht M, Cedar H. DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. Genes Dev 2008; 22:1319-24; PMID:18443145; http://dx.doi.org/10.1101/gad.468308
38. Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 2000; 287:2023-6; PMID:10720330; http://dx.doi.org/10.1126/science.287.5460.2023
39. Ferrari R, Su T, Li B, Bonora G, Oberai A, Chan Y, Sasidharan R, Berk AJ, Pellegrini M, Kurdistani SK. Reorganization of the host epigenome by a viral oncogene. Genome Res 2012; 22:1212-21; PMID:22499665; http://dx.doi.org/10.1101/gr.132308.111
40. Azuara V. Profiling of DNA replication timing in unsynchronized cell populations. Nat Protoc 2006; 1:2171-7; PMID:17487209; http://dx.doi.org/10.1038/nprot.2006.353

Supplemental Material to:

Bing Li, Trent Su, Roberto Ferrari, Jing-Yu Li,
and Siavash K Kurdistani

A unique epigenetic signature is associated with active DNA replication loci in human embryonic stem cells

Epigenetics 2014; 9(2)
http://dx.doi.org/10.4161/epi.26870

http://www.landesbioscience.com/journals/epigenetics/article/26870/

SUPPLEMENTARY METHODS
Li et al. *Epigenetics* 2013

Dot blot

BrdU dot blot was performed according to a previously described method.[1] Genomic DNA was quantified by Qubit Fluorometer (Life Technologies). DNA was denatured by incubation with 10 volume of 0.4 M NaOH for 30 min at room temperature. After incubation, an equal volume of 1 M Tris-HCl solution (pH 6.8) was added to neutralize the pH. 5 ng of DNA was dot blotted onto a nitrocellulose membrane (Schleicher & Schuell BioScience) and crosslinked by UV-light. The membranes were blocked with PBS containing 0.1% (v/v) Tween 20 and 5% (w/v) non-fat milk at room temperature before incubation with the anti-BrdU antibody (BD Biosciences, 347580) overnight at 4°C. The signal was detected by standard Western blotting.

Cell synchronization and flow cytometry analysis

Early passage IMR90 cells were arrested in G1 phase by contact inhibition. Cells were then replated at low density and cultured in DMEM containing 20% FBS and penicillin-streptomycin to stimulate cell growth. At 24 hour after splitting, cells were trypsinized, washed and fixed in ice-cold ethanol at 4°C overnight. The cells were then stained with propidium iodide for 15 min at 37°C, followed by flow cytometry analysis with FACSCan Flow cytometer (Becton Dickinson). Data were analyzed by the Flowjo software (Tree Star, Inc.) and the ModFit software (Verity Software House).

Knockdown and Western blotting

EP300 and CREBBP siRNA (Thermo Scientific) was mixed with DharmaFECT 1 (Thermo Scientific) in Opti-MEM (Life Technologies). After incubation for 20 min at room temperature, the siRNA-DharmaFECT 1 complex was added to 6-well plates that were pre-coated with Matrigel (BD Biosciences). $3x10^5$ of hESCs in single cell suspension were added to each well and mixed with transfection reagents. HA 1077 (EMD Millipore) was added to culture media to final concentration 10 μM. Same amount of non-targeting siRNA was used as control. Cells were collected at 72 hour post-transfection and subjected to further analysis. For Western blotting, cells were washed with PBS supplemented with protease inhibitors (Roche) and lysed with protein sample buffer. Cell lysate was subjected to SDS-PAGE and analyzed by Western blotting. EP300 (Millipore, 05-257), CREBBP (Santa Cruz Biotechnology, sc-369), histone H3 (Abcam, ab10799), ACTIN (Santa Cruz Biotechnology, sc-8432), H3K56ac (Active Motif, 39282), H3K18ac[2] and H3K9ac (Millipore, 07-352) antibodies were used for Western blotting.

ChIP-seq data analysis

Bowtie 0.12.7 was used to align sequenced reads to human genome (hg19). The data were processed as previously described.[2] Reads that aligned to more than one location in the human genome or have more than 2 mismatches were discarded. The total number of reads in the input and immunoprecipitated samples was normalized. To define peaks, human genome was divided into 100 bp windows and a *P* value for Poisson distribution was calculated of enriched immunoprecipitated DNA relative to input for each window. Significant peaks were defined as those windows with significant *P* value and with two neighboring windows at the

same significance. *P* value<10^{-4} cutoff was used for BrdU-seq and *P* value<10^{-3} cutoff was used for ChIP-seq of histone modifications and chromatin-binding proteins. We defined co-occupancy of BrdU peaks, NS DNA and histone modifications with BrdU peaks if they were within 1 kb from the any edge of the BrdU peaks. BrdU blocks and various genomic elements were considered to be associated with each other if their genome coverage overlapped. H1 hESC specific peaks were defined as those with no co-occupancy with BrdU peaks from other cell types.

BrdU blocks

To define BrdU blocks, the human genome was scanned with a 450 kb moving window and the windows that contained at least five BrdU peaks were retained and overlapping windows were combined to form blocks. The blocks that were smaller than 50 kb were filtered out.

External data source

Nascent strand DNA data for H9 ESCs and IMR90 cells were downloaded from the Gene Expression Omnibus (GEO; GSE37757) (Besnard et al., 2012). Raw data for histone H3K9ac, H3K27ac, H3K56ac, H4K5ac and H4K20me1 of hESCs and IMR90 cells was downloaded from the NCBI epigenome roadmap: (http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/).[3] Raw data for H3K56ac in adipocytes and H4K20me1 in HeLa S3 and K562 cells were downloaded from the GEO (GSE24326 and GSE29611).[4, 5] The Giemsa chromosome band annotation was downloaded from UCSC table browser website.[6] LADs were described by Guelen and co-

workers.[7] Early replication regions were defined previously.[8] DNA methylation sequencing data was retrieved from http://neomorph.salk.edu/human_methylome.[9] All downloaded data were processed or converted to UCSC hg19 build.

Visualization tools

Venn diagrams were generated using Venn Diagram Plotter from the Pacific Northwest National Laboratory (http://omics.pnl.gov/software/VennDiagramPlotter.php). Sequencing data were visualized with Integrated Genome Browser.[10]

SUPPLEMENTARY REFERENCES

1. Ueda J, Saito H, Watanabe H, Evers BM. Novel and quantitative DNA dot-blotting method for assessment of in vivo proliferation. Am J Physiol Gastrointest Liver Physiol 2005; 288:G842-7.
2. Ferrari R, Su T, Li B, Bonora G, Oberai A, Chan Y, et al. Reorganization of the host epigenome by a viral oncogene. Genome Res 2012; 22:1212-21.
3. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 2010; 28:1045-8.
4. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57-74.
5. Lo KA, Bauchmann MK, Baumann AP, Donahue CJ, Thiede MA, Hayes LS, et al. Genome-wide profiling of H3K56 acetylation and transcription factor binding sites in human adipocytes. PLoS One 2011; 6:e19778.
6. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002; 12:996-1006.
7. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 2008; 453:948-51.
8. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 2010; 107:139-44.
9. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22.
10. Nicol JW, Helt GA, Blanchard SG, Jr., Raja A, Loraine AE. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 2009; 25:2730-1.

Table S1. BrdU-seq alignment results using Bowtie 0.12.7 (hg19)

DNA	Total reads	Aligned reads	Percentage
H1 input	129325335	80306756	62.10%
H1 IP	146245314	96281267	65.84%
IMR90 input	173492285	103146696	59.45%
IMR90 IP	162159332	93376043	57.58%
Saos-2 input	47020762	22782589	48.45%
Saos-2 IP	45410425	20402965	44.93%
EP300/CREBBP KD input	119503688	76246481	63.80%
EP300/CREBBP KD IP	145591810	100038999	68.71%

Table S2. Peak calling with *P* value<1e-3

Cells	Total peaks	FDR
H1	13302	5.86%
IMR90	14958	7.72%
Saos-2	9405	9.21%

Table S3. Overlap of BrdU peaks with DNA repetitive elements

DNA Repeat Element	H1	EP300/CREBBP KD	IMR90	Saos-2
Microsatellite	2484 (56.48%)	384 (31.04%)	2301 (49.09%)	3804 (61.79%)
LINE	799 (18.17%)	309 (24.98%)	801 (17.09%)	787 (12.78%)
SINE	503 (11.44%)	260 (21.02%)	497 (10.6%)	110 (1.79%)
Low complexity	232 (5.28%)	45 (3.64%)	792 (16.9%)	1088 (17.67%)
LTR	279 (6.34%)	187 (15.12%)	205 (4.37%)	215 (3.49%)
DNA	83 (1.89%)	40 (3.23%)	80 (1.71%)	30 (0.49%)
Satellite	7 (0.16%)	6 (0.49%)	6 (0.13%)	122 (1.98%)
RC	2 (0.05%)	0 (0%)	0 (0%)	0 (0%)
snRNA	2 (0.05%)	0 (0%)	0 (0%)	0 (0%)
RNA	1 (0.02%)	0 (0%)	0 (0%)	0 (0%)
rRNA	1 (0.02%)	0 (0%)	0 (0%)	0 (0%)
Unknown	1 (0.02%)	0 (0%)	0 (0%)	0 (0%)
scRNA	0 (0%)	0 (0%)	0 (0%)	0 (0%)
srpRNA	0 (0%)	4 (0.32%)	0 (0%)	0 (0%)
tRNA	0 (0%)	1 (0.08%)	0 (0%)	0 (0%)
Other	4 (0.09%)	1 (0.08%)	5 (0.11%)	0 (0%)
Total	4398 (100%)	1237 (100%)	4687 (100%)	6156 (100%)

Figure S1

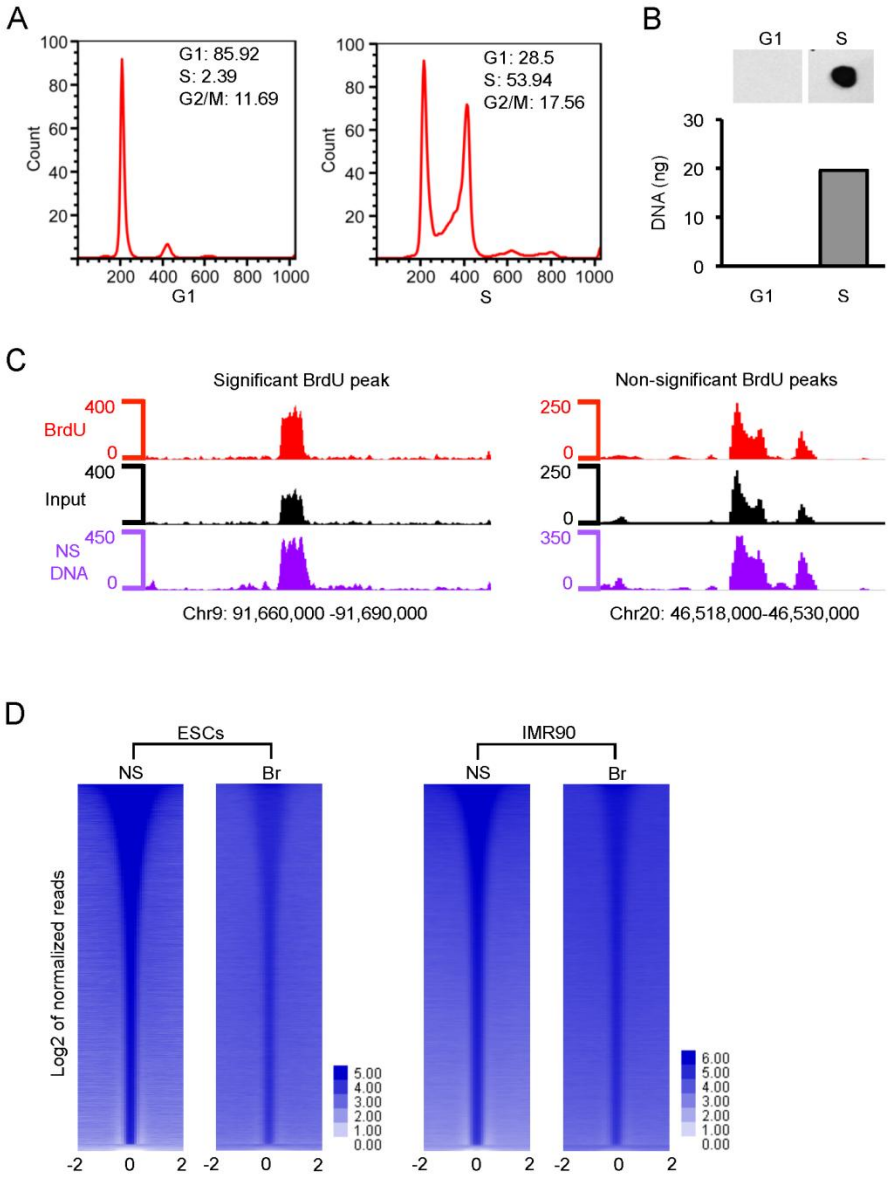

A
G1: 85.92
S: 2.39
G2/M: 11.69
G1: 28.5
S: 53.94
G2/M: 17.56
Count
G1
S
B
G1
S
DNA (ng)
C
Significant BrdU peak
Non-significant BrdU peaks
BrdU
Input
NS DNA
Chr9: 91,660,000 -91,690,000
Chr20: 46,518,000-46,530,000
D
ESCs
IMR90
NS
Br
Log2 of normalized reads

Figure S1. The anti-BrdU antibody specifically recognizes BrdU-labeled DNA in S phase

(**A**) Flow cytometry analysis of IMR90 cell cycle profile by propidium iodide staining. The percentage of cells in each phase is shown.

(**B**) Dot blot of BrdU-labeled DNA purified from cells in G1 and S phases of the cell cycle by the anti-BrdU antibody (upper panel) and quantification of immunoprecipitated DNA by the anti-BrdU antibody from BrdU-labeled cells in G1 and S phase (lower panel).

(**C**) Genome browser views of significant and non-significant BrdU peaks along representative regions of chromosomes 9 and 20. The chromosome coordinates of each track are indicated.

(**D**) Heat maps show the distribution of total sequencing reads from NS DNA and BrdU-seq at the replication origins defined by NS DNA as the $\log_2$ value of read number in ±2 kb region from the center of each NS DNA origin (pearson correlation coefficient r=0.45 and 0.6 for H9 and H1 ESCs and IMR90 cells respectively; *P* value=2.2e-16. Note that the slightly lower correlation for ESCs is likely due to comparison of the different cell types).

Figure S2

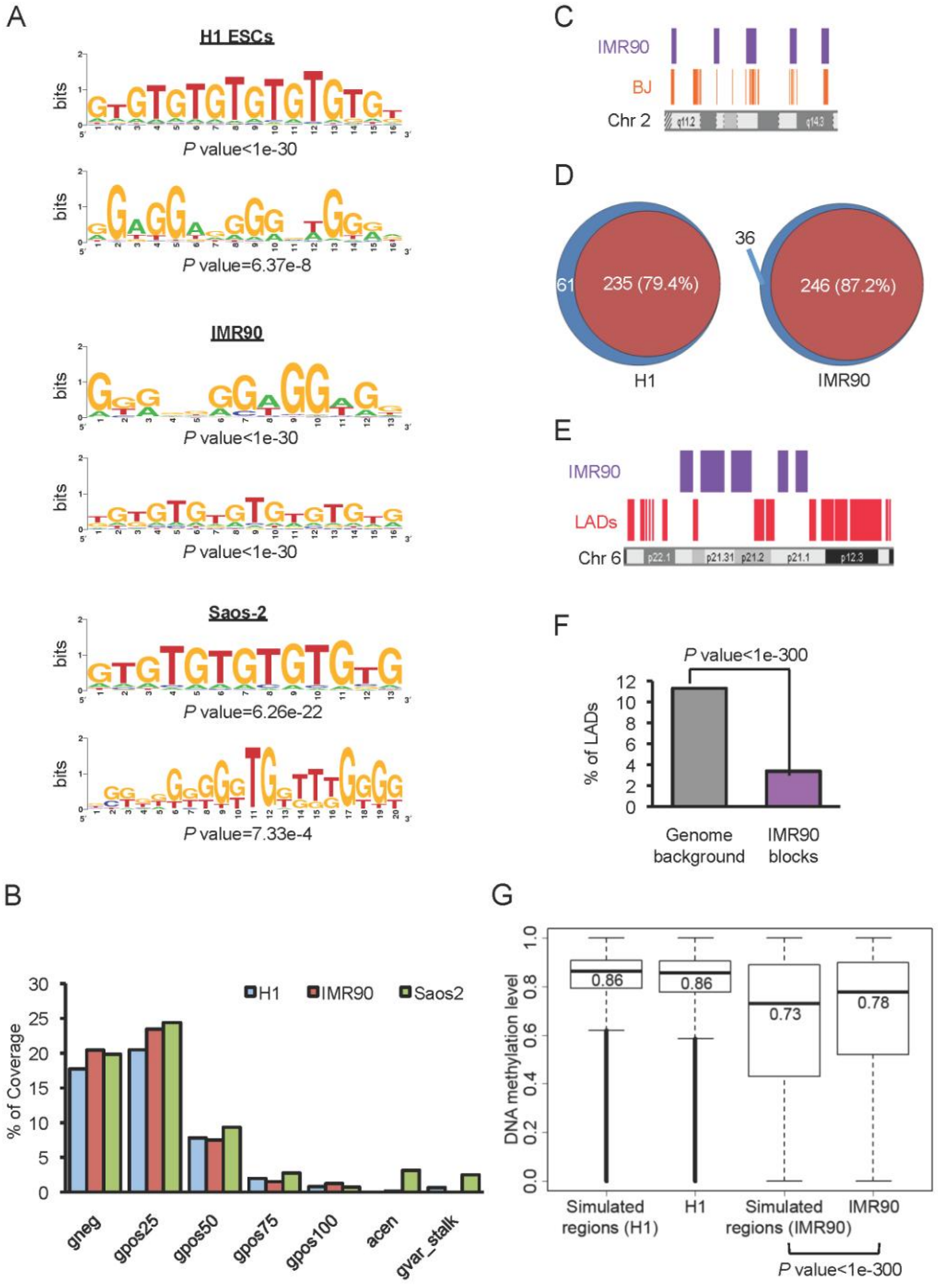

A
H1 ESCs
bits
P value<1e-30
bits
P value=6.37e-8
IMR90
bits
P value<1e-30
bits
P value<1e-30
Saos-2
bits
P value=6.26e-22
bits
P value=7.33e-4
B
H1
IMR90
Saos2
% of Coverage
gneg
gpos25
gpos50
gpos75
gpos100
acen
gvar
stalk
C
IMR90
BJ
Chr 2
D
61
235 (79.4%)
36
246 (87.2%)
H1
IMR90
E
IMR90
LADs
Chr 6
F
P value<1e-300
% of LADs
Genome background
IMR90 blocks
G
DNA methylation level
0.86
0.86
0.73
0.78
Simulated regions (H1)
H1
Simulated regions (IMR90)
IMR90
P value<1e-300

Figure S2. The majority of BrdU peaks occur in proximity of each other within early replicating euchromatin

(**A**) Significant DNA sequence motifs as defined by SeqPos (Cistrome) that are enriched around the BrdU peaks in the three cell types are shown as sequence logos.

(**B**) The distribution of BrdU blocks within Giemsa chromosome bands. The percentage of genome within each indicated region covered by BrdU block is represented as bar chart. The BrdU peaks were enriched mostly in gneg and gpos25 bands and were progressively less abundant in gpos100, acen, var and stalk regions.

(**C**) Locations of BrdU blocks in IMR90 and early replication regions in BJ fibroblasts along a region of chromosome 2 are shown.

(**D**) The Venn diagrams indicate the degree of overlap between BrdU blocks in hESCs or IMR90 cells and early replicating regions.

(**E**) Locations of BrdU blocks and lamina associated domains (LADs) along a region of chromosome 6 in human fibroblasts are shown.

(**F**) Percentage of LADs overlapping with BrdU blocks or with random blocks is shown as a bar chart.

(**G**) Box plots show the distribution of DNA methylation levels within BrdU blocks or within similarly-sized random blocks in each cell line. The thick horizontal line in each box represents the median value which is indicated.

Figure S3

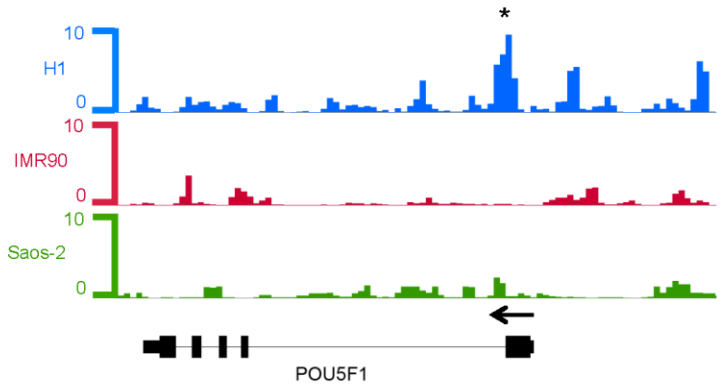


Figure S3. BrdU peaks at POU5F1 (OCT4) locus

The peak height corresponds to -$\log_{10}$ of *P* value for Poisson distribution for the enrichment of BrdU-seq reads. The asterisk indicates the significant peak that passed the peak calling criteria.

Figure S4

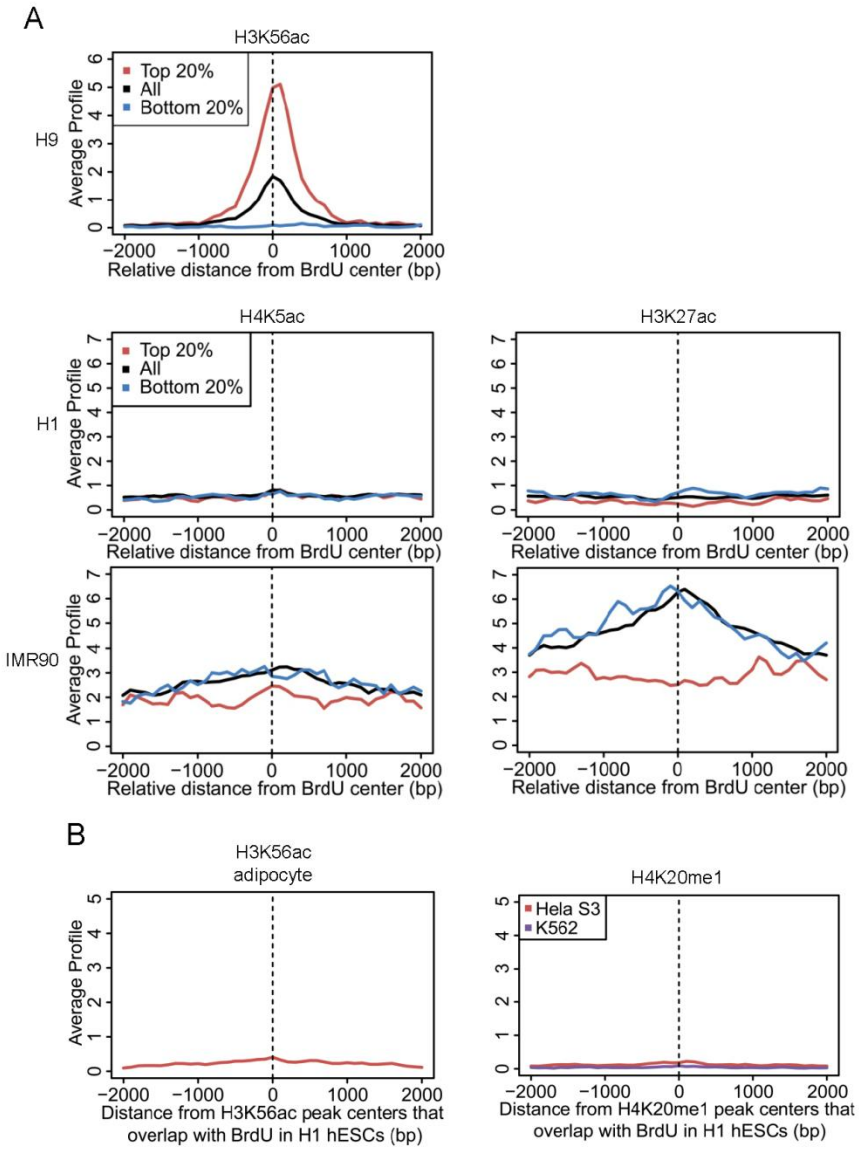


Figure S4. Co-occupancy of BrdU peaks with histone modifications is ESCs specific

(**A**) The average levels of the indicated histone modifications (significant reads counts) relative to the center of BrdU incorporating regions are shown for the top and bottom 20% according to the BrdU signal intensity in comparison to all BrdU incorporating regions for the specified cell types.

(**B**) H3K56ac and H4K20me1 are not enriched in BrdU and histone modifications associated regions in differentiated cells. Left panel, H3K56ac in adipocytes. Right panel, H4K20me1 in HeLa S3 and K562 cells.

Chapter 4

Histone acetylation

regulates intracellular pH.

This chapter was originally published in *Molecular Cell* 2013 Jan 24;49(2):310-21. with supplemental material. This chapter describes the relationship between intracellular pH and global histone acetylation. Low pH treatment in human hela cells induced global reduction of histone acetylation. The reduction in histone acetylation was coupled with proton excretion from cell membrane through monocarboxylate transporter (MCT) proteins. Global reduction in histone acetylation resulted in dramatic redistribution of histone acetylation marks. This study characterizes the first functional link betweenhistone deactylation and regulation of intracellular pH.

Molecular Cell
Article

Histone Acetylation Regulates Intracellular pH

Matthew A. McBrian,[1,2,7] Iman Saramipoor Behbahan,[1,7] Roberto Ferrari,[1] Trent Su,[1,3] Ta-Wei Huang,[1] Kunwu Li,[1] Candice S. Hong,[4] Heather R. Christofk,[4] Maria Vogelauer,[1] David B. Seligson,[5] and Siavash K. Kurdistani[1,2,5,6,*]
[1]Department of Biological Chemistry
[2]Molecular Biology Institute
[3]Division of Oral Biology and Medicine, School of Dentistry
[4]Department of Molecular and Medical Pharmacology
[5]Department of Pathology and Laboratory Medicine
[6]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[7]These authors contributed equally to this work
*Correspondence: skurdistani@mednet.ucla.edu
http://dx.doi.org/10.1016/j.molcel.2012.10.025

SUMMARY

Differences in global levels of histone acetylation occur in normal and cancer cells, although the reason why cells regulate these levels has been unclear. Here we demonstrate a role for histone acetylation in regulating intracellular pH (pH_i). As pH_i decreases, histones are globally deacetylated by histone deacetylases (HDACs), and the released acetate anions are coexported with protons out of the cell by monocarboxylate transporters (MCTs), preventing further reductions in pH_i. Conversely, global histone acetylation increases as pH_i rises, such as when resting cells are induced to proliferate. Inhibition of HDACs or MCTs decreases acetate export and lowers pH_i, particularly compromising pH_i maintenance in acidic environments. Global deacetylation at low pH is reflected at a genomic level by decreased abundance and extensive redistribution of acetylation throughout the genome. Thus, acetylation of chromatin functions as a rheostat to regulate pH_i with important implications for mechanism of action and therapeutic use of HDAC inhibitors.

INTRODUCTION

Targeted acetylation of lysine residues of histone proteins at distinct genomic loci is linked to regulation of essentially all DNA-templated processes, including transcription, replication, repair, recombination, and the formation of specialized chromatin structures such as heterochromatin (Kouzarides, 2007). For example, alterations in histone acetylation at select gene promoters—via recruitment of histone acetyltransferases (HATs) and histone deacetylases (HDACs) by sequence-specific DNA-binding transcription factors—regulate the transcriptional activity of the targeted genes (Ferrari et al., 2012). Histone acetylation regulates such DNA-templated processes by influencing the local chromatin structure and by regulating the binding or exclusion of bromo-domain-containing proteins to and from the chromatin (Shogren-Knaak et al., 2006; Taverna et al., 2007). The role of histone acetylation has largely been interpreted in this local, site-specific context (Margueron et al., 2005; Zhou et al., 2011). However, histone acetylation levels also differ at a cellular or global level (Horwitz et al., 2008; Vogelauer et al., 2000). Examination of acetylation by methods that assess total histone content—such as western blotting (WB) or immunohistochemistry (IHC)—has revealed heterogeneity in the levels of global histone acetylation in different tissues and cell types (Ferrari et al., 2012; Iwabata et al., 2005; Suzuki et al., 2009). IHC studies on a variety of primary cancer tissues have shown that an increased prevalence of cells with lower cellular levels of histone acetylation is associated with more aggressive cancers and poorer clinical outcome such as increased risk of tumor recurrence or decreased survival rates (Elsheikh et al., 2009; Fraga et al., 2005; Manuyakorn et al., 2010; Seligson et al., 2005, 2009). Such associations underscore the biological relevance of global differences in histone acetylation levels. However, very little is known about what function(s) the changes in global levels of histone acetylation serve for the cell. While a few studies have shown the necessity for a pool of acetyl coenzyme A (ac-CoA) to maintain global histone acetylation (Friis et al., 2009; Takahashi et al., 2006; Wellen et al., 2009), the biological factor(s) in response to which global histone acetylation levels change and what cellular processes are affected by this outcome have remained unknown (Friis and Schultz, 2009).

Cycles of histone acetylation and deacetylation occur continuously and rapidly throughout the genome, consuming ac-CoA and generating negatively charged acetate anions in the process. Since ac-CoA and acetate anions participate in many metabolic processes, we hypothesized that histone acetylation may be linked to certain metabolic or physiologic cues. We therefore systematically studied how global levels of histone acetylation change in response to alterations of various components of the standard tissue culture medium (Dulbecco's modified Eagle's medium, DMEM). Strikingly, we found that as intracellular pH (pH_i) is decreased, histones become globally hypoacetylated in an HDAC-dependent manner. The resulting free acetate anions are transported with protons by the proton (H^+)-coupled monocarboxylate transporters (MCTs) to the

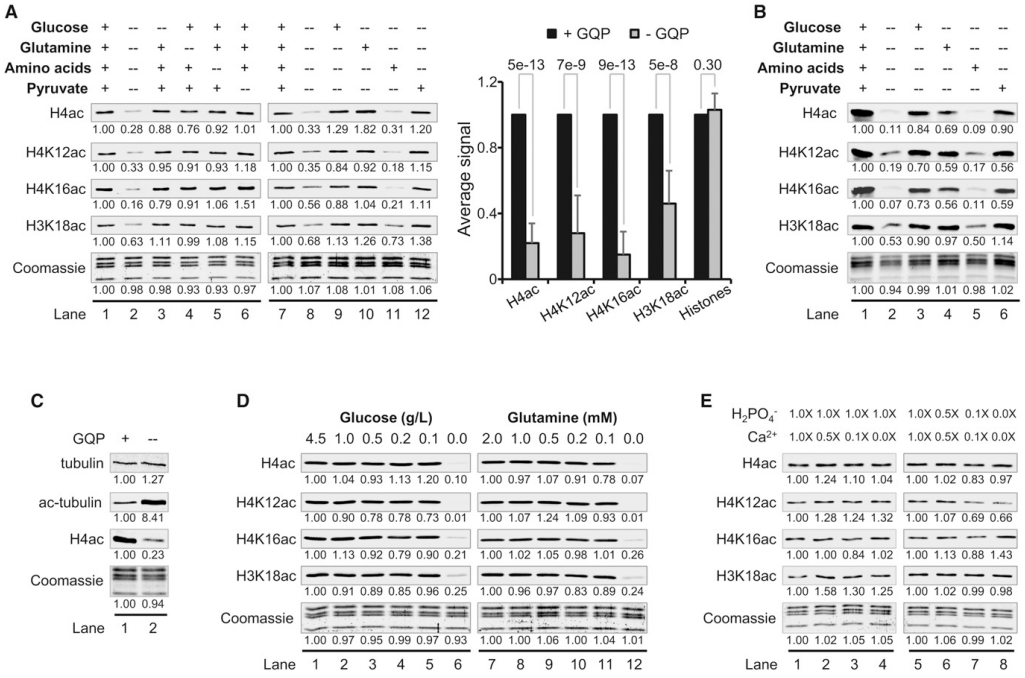


Figure 1. Minimal Levels of G, Q, or P Maintain Global Levels of Histone Acetylation

(A) WBs of histone acetylation in HeLa cells cultured for 16 hr in DMEM salts and vitamins with the indicated ac-CoA sources. The bar graph shows average values ± SD from 15 independent experiments and p value derived from the Student's t test.

(B) WBs of acetylation in normal IMR90 fibroblasts cultured as in (A).

(C) WBs of tubulin and histone H4 acetylation in HeLa cells under conditions of G, Q, and P deprivation.

(D) WBs of acetylation in HeLa cells cultured at the indicated concentrations of G or Q.

(E) WBs showing the effects of varying concentrations of Ca^{2+} and phosphate on histone acetylation in HeLa cells. Lane 1 in each panel (A–E) is the reference condition, the values of which are set to 1. See also Figure S1.

extracellular environment, thereby reducing the intracellular H^+ load and resisting further reductions in pH_i. As pH_i increases, the flow of acetate and protons is favored toward the inside of the cell leading to global histone hyperacetylation. Our data reveal that chromatin, through the basic chemistry of histone acetylation and deacetylation, coupled with MCTs, function as a system for rheostatic regulation of pH_i.

RESULTS

Glucose, Glutamine, or Pyruvate Is Required to Maintain Global Histone Acetylation

The metabolites in standard DMEM that are required to maintain a pool of ac-CoA for histone acetylation have not been systematically identified. Thus, we began by asking if any or all of the ac-CoA producing sources in DMEM are required to maintain steady-state levels of histones H3 and H4 acetylation. These sources potentially include glucose (G), glutamine (Q), pyruvate (P) and the 14 other amino acids (aa) present in DMEM. HeLa and MDA-MB-231 (231) cells were cultured for 16 hr in complete medium or in medium lacking all or one of the potential ac-CoA sources. Simultaneous removal of GQP and aa led to significant (~40%–99%) reduction in the acetylation of multiple lysine residues on histones H3 and H4 (Figures 1A and S1A, lane 2, available online). Elimination of G, Q, P, or aa individually had little or no effect on histone acetylation. These results suggest that the pool of ac-CoA that is used for histone acetylation derives from one or more of these carbon sources.

To determine which of these carbon sources are necessary to maintain histone acetylation, HeLa, 231, and IMR90 normal primary lung fibroblasts were cultured for 16 hr in medium with all ac-CoA sources, none, or a single ac-CoA source. Cells that were cultured in only G, Q, or P maintained normal levels of histone acetylation in cancer cells and substantial amounts in IMR90 fibroblasts (Figures 1A, 1B, and S1A). However, histone acetylation was significantly reduced in the presence of all other aa. We conclude that G, Q, or P—but not the remaining aa present in DMEM—are sufficient to maintain normal levels

of histone acetylation. It is interesting that, in the absence of GQP and aa, although histone acetylation was significantly decreased, we observed up to an 8-fold increase in tubulin acetylation (Figures 1C and S1B). Higher tubulin acetylation suggests that cellular pools of ac-CoA are not completely depleted when GQP are removed. Therefore, the loss of histone acetylation may be partly due to selective allocation of ac-CoA to other molecular processes under GQP starvation.

Minimal Amounts of G or Q Are Required to Maintain Normal Levels of Histone Acetylation

We next tested whether global levels of histone acetylation change with varying concentrations of G or Q as the sole ac-CoA source. As expected, at zero concentration of G or Q, histone acetylation was significantly reduced after 16 hr of culture (Figures 1D and S1C, lanes 6 and 12). It is surprising however, that global levels of histone acetylation remained largely unchanged over a 45-fold range of G concentration (0.1–4.5 g/l) and a 20-fold range of Q concentration (0.1–2 mM) (Figures 1D and S1C), which include concentrations insufficient to maintain cell growth and replication. These data indicate that global levels of histone acetylation do not change dose-dependently in response to variation in G and Q concentrations within the range tested.

Vitamins, Calcium, and Phosphate Do Not Affect the Global Levels of Histone Acetylation

We next asked whether global levels of acetylation change in response to the vitamins, cofactors, or salts present in DMEM (see Supplemental Experimental Procedures). Culturing HeLa cells for 16 hr in media with varying amounts of vitamins and cofactors in the presence or absence of carbon sources did not significantly affect the global levels of histone acetylation (Figure S1D). Calcium (Ca^{2+}) can enhance the permeability of the mitochondrial membrane to ac-CoA in the presence of phosphate (Benjamin et al., 1983). However, reducing the amount of Ca^{2+} in the presence or absence of phosphate did not lead to significant changes in the global levels of histone acetylation in HeLa or 231 cells after 16 hr of culture (Figures 1E and Figure S1E). We conclude that global levels of histone acetylation do not change significantly in the absence of vitamins and cofactors or Ca^{2+} and phosphate.

Global Levels of Histone Acetylation Change in Response to pH

Finally, we asked whether the buffering component and the resulting pH of the medium could affect histone acetylation. Changes in pH_e are known to induce corresponding changes in pH_i (Fellenz and Gerweck, 1988; Mizuno et al., 2002). We measured pH_i in our cell lines at different bicarbonate concentrations (i.e., different pH_e levels) using the pH-sensitive dye BCECF-AM (see Experimental Procedures). Consistent with previous reports (Wong et al., 2002), we found that decreasing pH_e led to decreasing pH_i, with the range of pH_i changes (6.3–7.3) being more restricted than that of pH_e changes (5.9–7.6) (see top graphs in Figures 2A and S2A). Varying bicarbonate concentration from 45 mM to 10 mM resulted in pH_e values that led to slight changes in pH_i and histone acetylation (Figures 2A and S2A, lanes 1–3). However, decreasing bicarbonate from 10 mM down to 1 mM resulted in pH_e values that resulted in pronounced effects on pH_i and significant reductions in histone acetylation as assessed by WB (Figures 2A and S2A, lanes 3–5) and immunofluorescence (Figure 2B). Acetylation of the cytoplasmic pool of histones was affected by pH in a similar manner (Figures 2C and S2B). The effect of pH_e on histone acetylation was not limited to cancer cells but was also observed in H1 human embryonic stem cells, IMR90 normal primary lung fibroblasts, and even in *Saccharomyces cerevisiae*, suggesting that changes in histone acetylation in response to varying pH is a fundamentally conserved process (Figure S2C). It is important to note that acetylation in yeast was reduced at levels of pH_e (~pH 4.0) that have the most pronounced effects on pH_i (Valli et al., 2005).

Histone acetylation was also significantly reduced at low pH when the culture medium was buffered with HEPES instead of bicarbonate (Figures 2D and S2D). This suggests the reduction in acetylation at low pH was in fact due to changes in pH and not to a reduction in bicarbonate concentration. The reduction in acetylation was also not due to changes in acetyl-CoA concentration (Figures 2E and S2E) but was dependent on the activity of class I and II HDACs. Treatment of cells with HDAC inhibitors Trichostatin A (TSA) and sodium butyrate, but not nicotinamide, prior to lowering the pH_e blocked the effects of pH on acetylation (Figures 2F and S2F). In addition, deletion of the major HDACs, RPD3 and HDA1, in yeast also greatly diminished the pH-induced hypoacetylation, confirming the requirement for HDAC activity (Figure S2C, lanes 13 and 14). In all pH-related experiments, histone methylation levels at multiple sites were largely unaffected, suggesting that the effect of pH on histone acetylation is not due to general impairment of the cellular posttranslational machinery (Figure 2G). It is interesting that tubulin acetylation was only minimally affected as pH_i decreased, indicating some degree of specificity for histone acetylation (Figures 2H and S2G). We conclude that global levels of histone acetylation change in response to pH_i, with increased deacetylation of histones occurring at lower values of pH_i.

The pH-Induced Changes in Histone Acetylation Levels Do Not Require Specific Carbon Sources or Salts

Lowering pH_e in the presence of only G, Q, or P resulted in reduction of histone acetylation (Figures 3A and S3A), indicating that the effect of pH on histone acetylation is independent of the source of ac-CoA. Next, we asked if specific salts are required for hypoacetylation of histones at low pH. Cells were cultured at normal and low pH_e in the presence and absence of sodium (Na^+) or chloride (Cl^-), using Q as the ac-CoA source. We found that global histone acetylation decreased in low pH_i regardless of the presence of either ion (Figures 3B, 3C, S3B, S3C). Ca^{2+} and phosphate were also not required for pH-mediated changes of histone acetylation levels, regardless of whether the cells were deprived of these salts for several days prior to lowering the pH (Figures 3D and S3D) or deprived at the same time the pH was decreased (Figure S3E). We conclude that the alteration of histone acetylation levels in response to pH does not depend on a specific carbon source or extracellular Na^+, Cl^-, Ca^{2+} or phosphate.

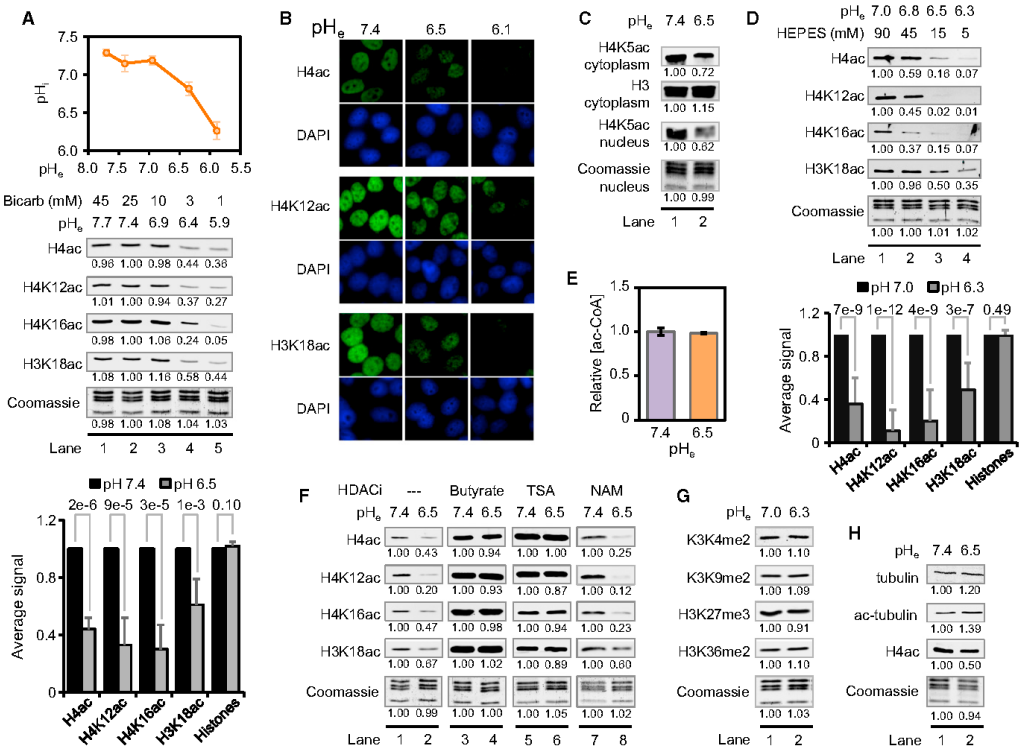


Figure 2. Global Levels of Histone Acetylation Decrease when pH_i Decreases

(A) Measurements of pH_i (mean ± SD) as a function of pH_e and effects on histone acetylation in HeLa cells cultured at the indicated pH_e for 16 hr in complete DMEM as assessed by WB. The reference condition is set to 25 mM bicarbonate (lane 2), which approximates the normal physiological concentration. The bar graph shows average values ± SD from seven independent experiments and p value derived from the Student's t test.

(B) Immunofluorescent analysis of histone acetylation of HeLa cells cultured as in (A).

(C) WB analysis of H4K5ac in HeLa cells in the indicated histone fractions and pH.

(D) WBs of histone acetylation in HeLa cells cultured in media buffered with HEPES. The bar graph shows average values ± SD from 17 independent experiments and p value derived from the Student's t test.

(E) Acetyl-CoA measurements (mean ± SD) in HeLa cells treated for 4 hr at the indicated pH_e.

(F) WBs of histone acetylation in HeLa cells that were treated with or without 5 mM sodium butyrate, 500 nM TSA, or 2 mM nicotinamide (NAM) for 6 hr in complete DMEM at pH_e 7.4, followed by incubation at the indicated pH_e for 4 hr.

(G) WBs of histone methylation in HeLa cells cultured at the indicated pH.

(H) WBs of tubulin and histone H4 acetylation in HeLa cells treated for 16 hr at the indicated pH values. See also Figure S2.

Acetylation of H4K16 Is Decreased and Drastically Redistributed at Low pH

To determine how the pH-induced changes in global levels of acetylation map to specific genomic loci, we treated HeLa cells for 4 hr at pH_e 7.4 or pH 6.5 and then performed chromatin immunoprecipitation sequencing (ChIP-seq) analysis of H4K16ac. (For each pH treatment, both input and immunoprecipitated DNA were sequenced.) Consistent with the global decrease in H4K16ac levels, there were significant reductions in both the number of base pairs covered as well as the total number of genomic loci significantly associated with H4K16ac at low pH (Figure 4A). Clustering of the significant peaks of H4K16ac revealed drastic redistribution and decreased overall intensity of this acetylation site at low pH. Figure 4B shows the clusters of H4K16ac peaks (heat maps), the average value of each cluster (line graphs), and the genomic distribution of the peaks (pie charts) with associated gene ontology. Cluster 1 contained 2,751 peaks of H4K16ac in high pH, which were essentially fully deacetylated at low pH. These peaks are mostly near transcription start sites (TSS) and introns of genes whose gene ontology (GO) terms include in protein synthesis and cell cycle control. Cluster 2 contained 595 peaks that were present at both pH

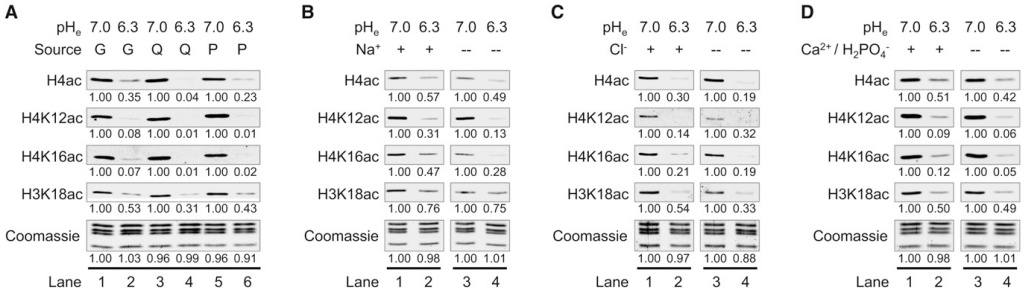


Figure 3. Changes in Histone Acetylation Levels in Response to pH Do Not Require Specific Carbon Sources or Salts

(A–D) WBs of histone acetylation in HeLa cells cultured for 16 hr in DMEM salts at pH_e 7.0 or 6.3 either (A) including the indicated carbon source or including Q but lacking (B) Na^+, (C) Cl^-, and (D) Ca^{2+} and phosphate. In (D), cells were starved of Ca^{2+} and phosphate for 3 days prior to pH_e treatment. See also Figure S3.

values but with much decreased intensity at low pH. More than 90% of cluster 2 peaks occur at TSS and introns of genes enriched in GO terms encoding ribonucleoproteins including mitochondrial and cellular ribosomal protein genes as well as, interestingly, the NuA4 HAT complex that acetylates histone H4 (Arnold et al., 2011). Cluster 3 contained 1,971 new peaks of H4K16ac at low pH, which are mostly in regions away from TSS and are associated with GO terms including spindle organization and replicative senescence.

We also performed mRNA sequencing (mRNA-seq) under the same conditions to correlate the redistribution of H4K16ac to changes in gene expression. Compared to high pH, 740 genes were upregulated greater than 2-fold at low pH, including histone genes and genes involved in regulation of transcription, signaling, and metabolism (Table S1). A total of 888 genes were downregulated greater than 2-fold at low pH with functions in plasma membrane and extracellular matrix biology (Table S1). It was surprising that there were no significant changes in the average expression of genes in the three acetylation clusters (Figure 4C). We conclude that treatment at low pH results in elimination of most H4K16ac peaks and re-establishment of less than half as many peaks with lower intensity at new genomic locations predominantly away from TSS regions. This redistribution does not significantly correlate with gene expression after 4 hr of treatment at low pH.

Changes in Histone Acetylation in Response to Nutrient Availability and pH Occur via Different Mechanisms

We next performed kinetic experiments to compare the histone acetylation changes that occur in response to the acidification of the intracellular milieu to those resulting from changes in nutrient availability. First, we performed time course experiments to examine the kinetics of the decrease in acetylation upon each type of treatment. HeLa and 231 cells were cultured for 1, 2, 4, 8, and 16 hr at normal or low pH in the presence or absence of GQP (Figures 5A and S4A). Removal of GQP at normal pH resulted in a gradual decrease in histone acetylation over 16 hr (compare solid and dashed green lines). In contrast, acidification of the medium in the presence of GQP resulted in an abrupt decrease in acetylation at 1 hr, which then remained relatively constant over time (compare solid green and orange lines). In addition, removal of GQP in the low pH medium resulted in a sharp decrease in acetylation during the first hour, followed by a gradual decline over 16 hr (compare solid and dashed orange lines). These data indicate that the kinetics of hypoacetylation in response to acidic pH_i and to GQP deprivation are markedly different.

Second, we performed time course experiments to investigate the kinetics of the recovery of histone acetylation in cells that were returned to steady-state culture conditions after being treated at low pH or being deprived GQP. HeLa and 231 cells were cultured for 16 hr at pH_e 7.4 or 6.5. A cohort group of cells were cultured for 16 hr in medium with or without GQP at pH_e 7.4. In each case, the cells were then returned to normal culture conditions for various amounts of time as shown in Figures 5B and S4B. The recovery of acetylation upon returning GQP was gradual and required ~24 hr to return to steady-state levels (compare purple and green lines). The recovery of acetylation upon return to a normal pH_e was much faster (compare purple and orange lines). These data show that the kinetics of acetylation recovery are also different for cells treated at low pH versus those that have been GQP starved. Together with the differential effects on tubulin acetylation, we conclude that medium acidification and nutrient deprivation affect histone acetylation by distinct mechanisms.

Decreased pH Results in an HDAC-Dependent Excretion of Acetate and Protons by MCTs

Since the levels of acetylation were regulated in response to pH, we hypothesized that histone acetylation may impact cellular control of pH_i. While deacetylation of lysines in and of itself has no net effect on pH_i, export of free, negatively charged acetate anions, which are membrane impermeable (Walter and Gutknecht, 1984), through the H^+-coupled MCTs would buffer against further acidification of pH_i by decreasing the intracellular H^+ load. As obligate H^+ symporters, MCTs 1–4 bidirectionally cotransport acetate and other small organic anions—such as lactate—with a H^+, thereby participating in pH_i regulation (Boron et al., 1988; Halestrap and Meredith, 2004).

To determine if HDAC-derived acetate anion is excreted increasingly as pH decreases, we incubated cells with ^{3}H-acetate

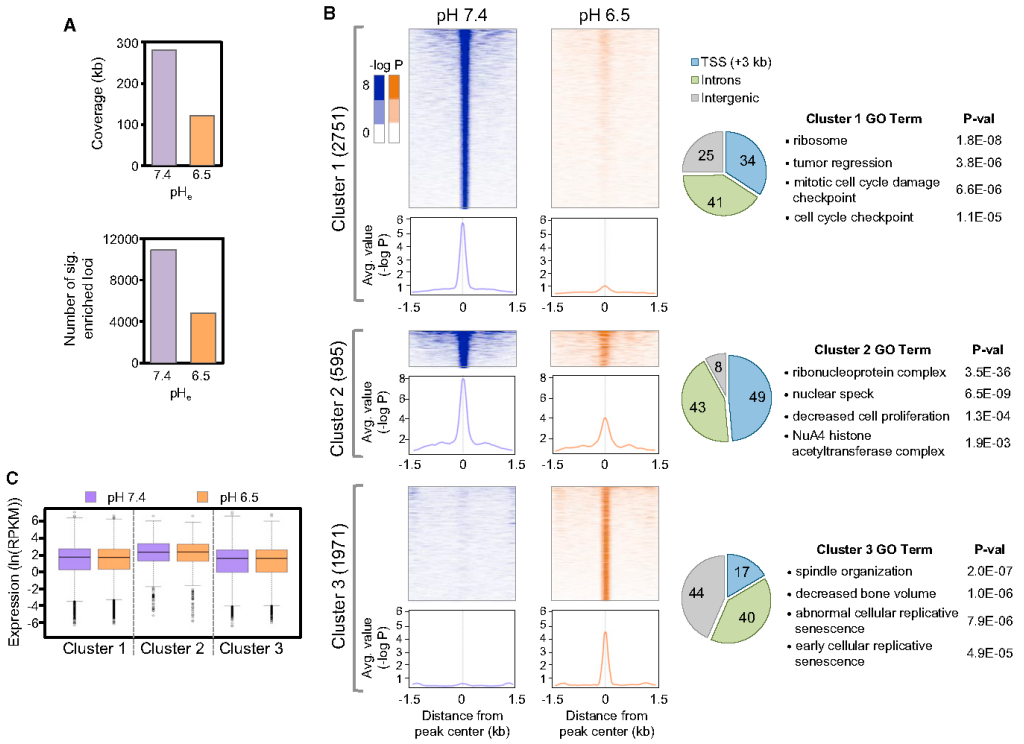


Figure 4. Low pH Causes Global Decrease and Redistribution of H4K16ac throughout the Genome

The distribution of H4K16ac in HeLa cells treated for 4 hr at pH 7.4 or 6.5 was determined by ChIP-seq.

(A) Number of DNA base pairs covered by significant ($p < 10^{-4}$) peaks of H4K16ac and number of genomic regions (25 bp windows) associated with significant H4K16ac peaks.

(B) Heat maps showing the distribution of H4K16ac peaks at the indicated pH values as the −log p value of enrichment in ±1.5 kb region from the center of each peak. All peaks were clustered in three groups based on their combinatorial presence in each pH condition. The number of peaks in each cluster is indicated. The average value of enrichment as −log p value for each cluster is shown as a line chart below each heat map. The distribution of peaks in each cluster is shown as a pie chart. The gene ontology (GO) of genes associated with the peaks of H4K16ac in each cluster is shown. The GO terms include biological process, cellular component, molecular function, and mouse phenotypes of homologous gene knockout.

(C) Box plots showing the distribution of expression of genes associated with the H4K16ac peaks in each cluster under the indicated pH conditions. See also Table S1.

in complete medium for 1 hr in order to label histone acetyl groups (Carmen et al., 1996). Following a 30 min chase, HeLa and 231 cells were treated in media of varying pH, which was assayed for released tritium (^3H) at the indicated time points in Figure 6A. (Note that at each time point, the entire volume of medium was removed for liquid scintillation counting and fresh medium was added.) Incubation in decreasing pH_e resulted in immediate, dose-dependent increases in the appearance of ^3H for up to 30 min, which we attribute to acetate anion released from the cell (Figure 6A). The fast nature of the response is consistent with other regulators of pH_i (Hulikova et al., 2011). To determine if global histone acetylation is affected in a similarly short time period, we performed kinetic studies analogous to those done in Figure 5 but on a shorter time scale. Loss of global acetylation at most sites examined was detected by WB at 30 min after lowering pH_e (Figure S5A). It is possible that histones are deacetylated at earlier time points but at levels below the sensitivity of the antibodies used in this study. Indeed, we detected recovery of acetylation as early as 5 min after return to normal pH_e (Figure S5B). We conclude that global deacetylation of histones in response to pH is associated with increased excretion of acetate anion from cells.

The appearance of ^3H in the media at both normal and low pH was significantly decreased by treating cells for 30 min with the MCT1/MCT4 inhibitor α-cyano-4-hydroxycinnamate (CNCn) after the labeling step but before addition of medium at

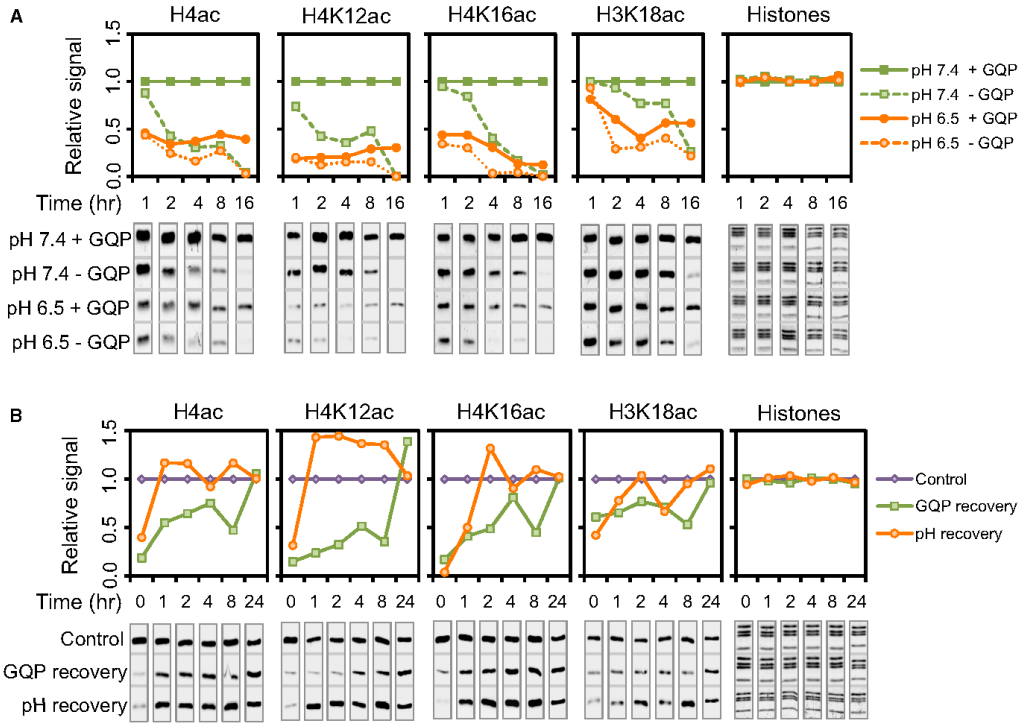


Figure 5. Nutrient Availability and pH Alter Histone Acetylation Differently

(A and B) WBs of histone acetylation in HeLa cells cultured in DMEM salts (A) for the indicated amount of time and pH_e in the presence (+) and absence (–) of GQP and (B) at pH_e 7.4 with GQP (control), pH_e 7.4 without GQP (for GQP recovery), or pH_e 6.5 with GQP (for pH recovery) for 16 hr followed by treatment for the indicated amount of time in medium at pH_e 7.4 with GQP. Membrane images were cropped in order to place them below the graph at the corresponding time points. See also Figure S4.

normal or low pH (Figure 6B). In addition, the class I HDAC inhibitor TSA inhibited the increase in the appearance of ^3H in the media at low pH_e (Figure 6C). These data suggest that histones are globally deacetylated at low pH in an HDAC-dependent manner with subsequent excretion of acetate anion and protons from the cell through the MCTs.

The potential of HDAC-dependent, H^+-coupled acetate excretion to regulate pH_i depends upon sufficient rates of acetate excretion. To determine the rate of acetate excretion from cells, we used an enzyme-based assay to measure the amount of acetate in culture media from 231 cells incubated in DMEM of varying pH_e for 30 min. Assuming an average cell volume of 500 fl (Kim et al., 2007), we found that, as the pH_e was decreased in a stepwise manner, the rate of acetate anion excretion increased from ~0.43 mM/min at pH_e 7.4 to ~0.74 mM/min at pH_e 6.7, reaching a rate of ~0.96 mM/min at pH_e 6.1 (Figure 6D). This rate of acetate and H^+ secretion is within the range of other known regulators of pH_i such as NHE1, which, for example, secretes protons from acid-loaded (pH_i of 6.8) HeLa and MDA-MB-468 cells at a rate ~0.5 mM/min and ~2.0 mM/min, respectively, at a pH_e of 6.8 (Hulikova et al., 2011). Since MCT-dependent efflux of acetate anions and protons has a 1:1 stoichiometry, the acetate efflux rate could significantly contribute to pH_i regulation by enabling efflux of protons at the same rate. It is important to note that levels of acetylation were increased in 231 cells that were treated with TSA after they had been cultured at low pH_e (Figure 6E). This suggests that HAT activity is maintained at low pH, thereby ensuring a steady supply of acetate anions.

Inhibition of HDACs or MCTs Decreases pH_i

Our data predict that HDAC inhibition should decrease pH_i by preventing the release of acetate from histones, thereby reducing the ability of MCTs to couple H^+ efflux to acetate anion efflux. Treatment with increasing concentrations of TSA in complete medium at normal pH_e resulted in a dose-dependent decrease of pH_i in both HeLa and 231 cells (Figures 7A and S6A, left panel). This effect was not due to passive acidification

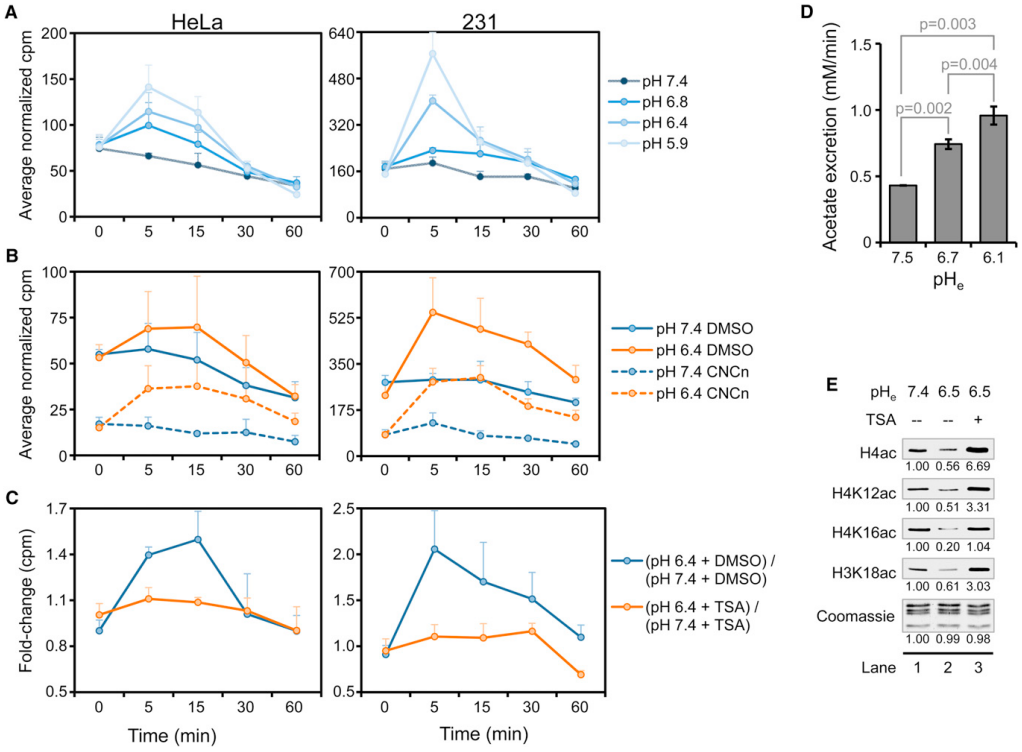


Figure 6. Efflux of Acetate through the MCTs at Low pH Is Dependent on HDAC Activity

(A–C) Liquid scintillation counts (mean ± SD) of media from HeLa and 231 cells labeled with ^{3}H-acetate and treated as indicated in 10% FBS-containing DMEM. (A) Cells were exposed to media of different pH_e for the indicated time. (B) Cells were treated with the MCT inhibitor CNCn (10 mM) or dimethyl sulfoxide during the chase and during subsequent treatment at the indicated pH_e. (C) Cells were treated with 500 nM TSA or DMSO prior to labeling and during subsequent treatment at varying pH_e.

(D) Acetate excretion rate (mean ± SD) in 231 cells incubated for 30 min in 10% FBS-containing DMEM without P at the indicated pH_e.

(E) WBs of histone acetylation in 231 cells treated for 16 hr in complete DMEM at the indicated pH_e followed by treatment at the same pH_e with or without 500 nM TSA. See also Figure S5.

of the intracellular environment by TSA—which is weakly acidic—as cotreatment with TSA and ITSA-1, an inhibitor of TSA (Koeller et al., 2003), blocked the TSA-induced acidification of pH_i (Figures 7A, middle panel and S6A, right panel). Since histone deacetylation and acetate anion excretion is increased as pH_i decreases, we reasoned that inhibition of HDACs may have more pronounced effects on pH_i in more acidic conditions. Indeed, this was the case. Treatment of cells with 100 nM TSA had little effect on pH_i at normal pH but disrupted the ability of cells to maintain their pH_i at lower values of pH_e (Figure 7A, right panel). The effect of HDAC inhibition on pH_i was not limited to TSA as the HDAC inhibitors Panobinostat (LBH589) and sodium butyrate (which are both weak acids) and Apicidin (a cyclic peptide antibiotic which is not an acid) all caused dose-dependent decreases in pH_i (Figure 7B). Nicotinamide, a sirtuin HDAC inhibitor, had essentially no effect on pH_i at any pH_e tested (Figures 7B and S6B).

Our data also predict that inhibition of MCTs should lead to decreases in pH_i by diminishing the rate of H^+-coupled acetate excretion. Indeed, treatment with CNCn in complete medium with decreasing pH_e resulted in decreased pH_i (Figures 7C, left panel and S6C, top left panel). It is important to note that the effects of CNCn on pH_i of HeLa cells were most prominent at lower values of pH_e, where global histone acetylation is decreased and rates of acetate anion excretion are increased. In addition, knockdown of MCT1 also resulted in reduced pH_i compared to cells treated with nontargeting siRNAs (Figure 7C, right panel).

Since the effect of MCT inhibition on lowering pH_i could potentially be explained by inhibition of lactate rather than acetate

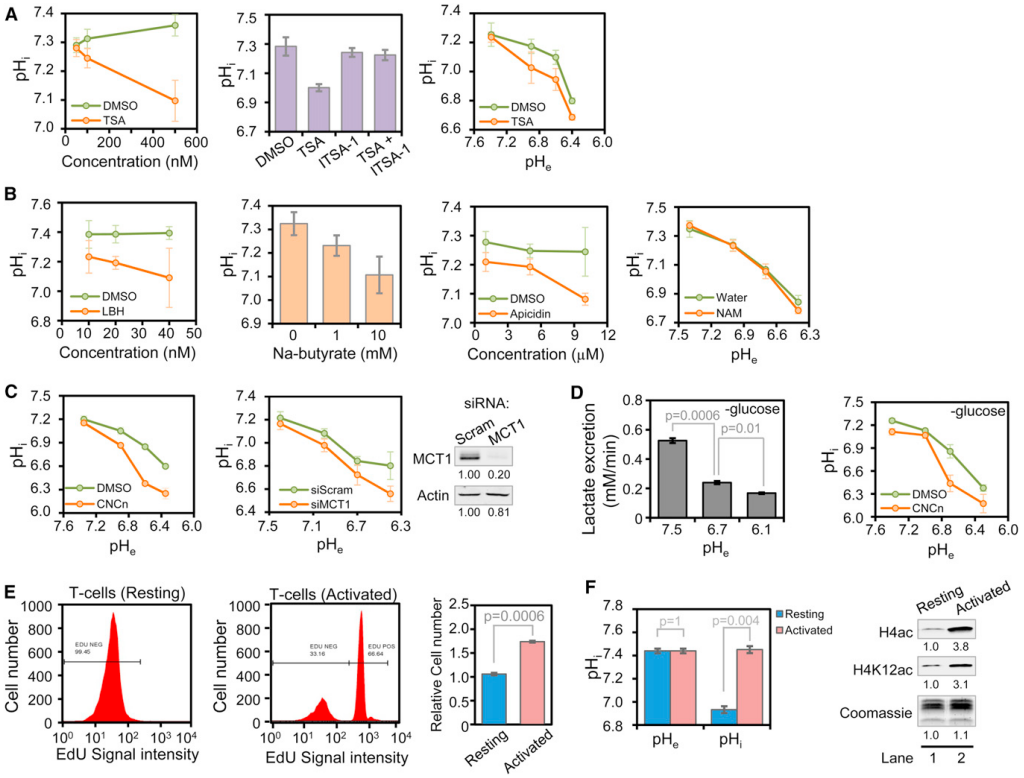


Figure 7. HDACs, MCTs, and Cellular Proliferation Affect pH_i and Global Histone Acetylation Levels

(A and B) pH_i of HeLa cells treated in standard DMEM overnight with (A) the indicated concentration of TSA (left panel); 250 nM TSA, 50 μM ITSA-1, or both (middle panel); 100 nM TSA followed by 6 hr of culture at the pH_e indicated (right panel); or (B) the listed HDAC inhibitors at the indicated concentrations (NAM = 2 mM nicotinamide).

(C) pH_i of HeLa cells in which MCT function was inhibited by treatment for 1 hr with 10 mM CNCn (left panel) or by siRNA-mediated knockdown of MCT1 for 96 hr (right panel) followed by treatment for 30 min at the pH_e shown.

(D) Rates of lactate excretion in HeLa cells after 30 min of treatment in DMEM without G and P at the indicated pH_e (left panel); pH_i measurements under identical conditions with 10 mM CNCn or DMSO (right panel).

(E) T cells were stimulated to proliferate by CD3/CD28-coated dynabeads and interleukin-2 treatment for 48 hr. EdU incorporation was determined by flow cytometry, and cell number was counted with and without 48 hr stimulation.

(F) pH_e and pH_i values, as well as histone acetylation, of resting and activated T cells after 48 hr. Note that pH_i changes independently of pH_e. All data (A–F) are presented as mean ± SD. See also Figure S6.

excretion, we examined the rates of lactate excretion at varying pH_e. We found that, as pH_e was decreased, the rate of lactate excretion was drastically reduced in both HeLa and 231 cells (Figures 7D and S6C, lower left panel) and became even slower than that of acetate excretion at lower values of pH_e in G-deprived 231 cells (Figure S6C, lower right panel). It is important to note that MCT inhibition under G-deprived conditions still resulted in decreased pH_i (Figures 7D, right panel and S6C, top right panel). It is interesting that MCT inhibition by CNCn did not affect pH-induced histone hypoacetylation, indicating that histone deacetylation is not dependent on acetate excretion (Figure S6D). Taken together, these results indicate that decreasing acetate release from histones by HDAC inhibition or efflux from the cell by MCT inhibition results in lower pH_i.

Rapid Cell Proliferation Is Associated with Increased pH_i and Histone Hyperacetylation

Although our pH experiments were all done at pH_e values that are known to occur in certain cancer tissue (Zhang et al., 2010), we wished to investigate a system in which pH_i changes independently of pH_e. Considering that rapidly dividing cells have higher pH_i than nondividing cells (Reshkin et al., 2000), our data predict

that changes in proliferation are associated with changes in histone acetylation. To test this prediction, we used an in vitro system that mimics the physiological stimulation of T cells to proliferate by antigen-presenting cells (Trickett and Kwan, 2003). Stimulation of proliferation of normal, primary peripheral blood T cells—as evidenced by the incorporation of 5-ethynyl-2′-deoxyuridine (EdU)—led to an increase in overall cell number after 48 hr (Figure 7E). In agreement with previous reports (Bental and Deutsch, 1994), the activation of T cells was associated with an increase in pH_i independent of changes in pH_e. As predicted, increased pH_i was associated with an increase in global histone acetylation (Figure 7F). These data indicate that in physiological conditions in which pH_i is affected independently of pH_e, global histone acetylation levels change correspondingly.

DISCUSSION

Cells have developed multiple mechanisms to maintain and regulate pH_i in response to changes in pH_e or increased acid production within the cell (Casey et al., 2010; Parks et al., 2011). These mechanisms include Na^+/H^+ and Cl^-/bicarbonate exchangers, H^+ pump ATPases, and bidirectional transport of organic acids by H^+-coupled MCTs (Cardone et al., 2005; Wahl et al., 2002). Notably, the MCTs contribute to pH_i regulation particularly at low pH_e values (Cardone et al., 2005) in tissues such as muscle during exercise (Messonnier et al., 2007) and in certain cancer cells (Wahl et al., 2002). Our data now reveal that the dynamic acetylation and deacetylation of histones together with flux of acetate anions and protons in and out of the cell through the MCTs provides an additional mechanism for cells to modulate their pH_i. Deacetylation of histones releases acetic acid from chromatin. Owing to its pKa of 4.74, acetic acid immediately releases a proton that is readily accepted by the ε-amino group (pKa 10.5) of the deprotonated lysine within physiologically relevant pH range. The impermeability of the acetate anion to the plasma membrane necessitates a transporter for the movement of acetate in and out of the cell. This need is met in part by the proton-coupled MCTs, imparting chromatin with the capability to serve as a buffer of pH_i. Whether the acetate molecules serve additional functions once excreted from the cell remains to be determined.

The mechanism(s) by which pH affects global histone acetylation levels remains to be determined, but our data suggest that HAT activity is not compromised and that the activity of HDACs is required. It is possible that, at low pH_i levels, the balance of histone acetylation and deacetylation is shifted toward the latter, resulting in continual generation of free acetate anions for H^+ transport by MCTs. This would result in the appearance of globally hypoacetylated histones but increased production of acetate anions due to continued activity of HATs (see Graphical Abstract). Different sites of histone acetylation may also respond differently to changes in pH. We consistently observed more reduction in global levels of histone H4 than H3K18ac, suggesting that some fraction of H3K18ac may be protected against or unresponsive to pH alterations.

At a molecular level, global hypoacetylation of H4K16 at low pH is associated with elimination of 82% of peaks that are normally present at high pH and with decreased intensity of the remaining 18% of the peaks. Conversely, novel peaks of H4K16ac are established at low pH, covering less than half as many regions of the genome as in a high pH condition. It is interesting that many more peaks of H4K16ac are associated with promoter regions at high pH compared to low pH. Despite the widespread decrease and redistribution of H4K16ac, there was essentially no correlation with the gene expression changes that occurred at low pH in the same time frame. It is possible that redistribution of H4K16ac precedes corresponding changes in gene expression that may occur at a later time point or that other sites of histone acetylation correlate better with transcriptional changes. Nonetheless, our data suggest that rapid changes in gene expression are decoupled from hypoacetylation and redistribution of H4K16ac in response to low pH.

Participation of histone acetylation in pH regulation is consistent with the long-standing phenomenon that acetyl groups of histones have a rapid turnover rate, on the order of a few minutes (Waterborg, 2002). Such rapid turnover would be required for cells to respond swiftly to pH fluctuations. Our data may also be of particular relevance to cancers with low levels of histone acetylation that display a poor clinical outcome (Manuyakorn et al., 2010; Seligson et al., 2009; Suzuki et al., 2009). Cancer tissues displaying low levels of histone acetylation may have enhanced acetate and proton efflux, thereby contributing to an alkaline pH_i and an acidic pH_e. Tumor microenvironments are commonly found to be acidic, and some with pH values below 6.5 have been reported (Zhang et al., 2010). In addition, tumors with an alkaline pH_i and/or acidic microenvironment exhibit more aggressive phenotypes (Harguindey et al., 2005; Lora-Michiels et al., 2006; Moellering et al., 2008; Perona and Serrano, 1988; Reshkin et al., 2000; Rofstad et al., 2006). Our data also suggest that the use of HDAC inhibitors as therapeutic agents in cancer treatment should also be considered from the additional perspective that HDAC inhibition may disrupt the pH_i-buffering capacity of cells. Indeed, HDAC inhibitors other than those used in this study also lower pH_i (Chung et al., 2008), suggesting an alternative mechanism of action for this class of drugs. Altogether, our findings uncover chromatin as an essential component of an alternative pH_i regulatory system involving histone acetylation and deacetylation and MCTs.

EXPERIMENTAL PROCEDURES

Cell Culture

HeLa, 231, and IMR90 cells were maintained in DMEM supplemented with G, Q, P, sodium bicarbonate, antibiotics, and 5% dialyzed fetal bovine serum (FBS). Experiments were performed in the absence of serum, unless otherwise noted. H1 cells were maintained in mTeSR1 medium (STEMCELL Technologies #05850) on Matrigel. T cells were purchased from AllCells (PB009-1F) and were maintained in RPMI-1640 (ATCC #30-2001) with 10% FBS (Hyclone #SV3001403) that was heat inactivated by treatment at 56°C for 1 hr.

WB Analysis

WBs were performed using the LI-COR Odyssey system. Blots were performed on acid-extracted histones from isolated nuclei or on whole-cell lysates in the case of tubulin, cytoplasmic histones, and MCT1. A representative blot of three or more independent biological replicates is shown. Statistical analysis was performed in cases where at least seven replicate data points were obtained.

Immunofluorescence

Cells were plated on chambered slides (Fisher #12-565-110N) in control medium and then treated for 16 hr at the indicated pH. Cells were fixed in methanol and stained with the indicated antibodies at 1:200 dilutions in PBS-Tween with 5% bovine serum albumin. Alexa Fluor 488 goat anti-rabbit secondary antibody was used at 1:1,000 dilution. Images were taken using an inverted fluorescent microscope and processed using Slidebook.

Measurements of ac-CoA

Cells grown on 10 cm dishes were treated for 4 hr in DMEM at pH 7.4 or 6.5. The plates containing cells were then directly submerged in liquid nitrogen and shipped on dry ice to the University of Michigan Metabolomics Resource Core for mass spectrometry measurements (Lorenz et al., 2011).

ChIP-Seq and mRNA-Seq

Cells were cultured for 4 hr at the indicated pH in complete DMEM. DNA and RNA were subjected to standard ChIP-seq and mRNA-seq as previously described (Ferrari et al., 2012). Briefly, sequenced reads were aligned to the human genome (Hg19), and only those that matched a unique location with up to two sequence mismatches were retained. To define peaks of enrichment, we segmented the human genome into 25 bp windows and compared the ChIP and normalized input DNA read counts in each window. Using the Poisson distribution, we calculated p values for the enrichment of ChIP reads in each window. A cutoff p value $<10^{-4}$ was used to maintain a false discovery rate <1%. Significant peaks were defined as those with a p value $<10^{-4}$ and with significant windows at the same p value in the two neighboring windows.

Total RNA was extracted from cells using QIAGEN RNeasy Mini kit and treated with Ambion's Turbo DNase. Two micrograms of total RNA was used to start the library preparation according to the manufacturer's instructions (Illumina TruSeq RNA Sample Preparation Kit). Libraries were sequenced using Illumina HiSeq system to obtain 50 bp-long reads. Alignment of mRNA-seq reads to Human genome (Hg19) was performed using default parameters of Tophat (Trapnell et al., 2009). SAMMate software (Xu et al., 2011) was used to determine the transcript RPKM (reads per kilobase of exon per million of reads).

Acetate and Lactate Excretion Assays

Cells grown in 10 cm dishes were incubated in 5 ml of medium at the indicated pH_e for 30 min. Acetate was measured using the K-ACET kit, and lactate was measured using the K-LATE kit (Megazyme) according to manufacturer's instructions. We confirmed the compatibility of all reagents used in experiments with this kit using the provided standards. Note that an unknown compound in HeLa cell culture medium interfered with the reactions of the K-ACET kit, precluding direct acetate measurement.

pH_i Measurements

Adherent cells were grown on 35 mM poly-lysine-coated glass bottom dishes in the presence or absence of pharmacological inhibitors or siRNA. Cells were loaded with 5 nmol/ml of the pH-sensitive dye BCECF-AM in EBSS for 25 min. We then determined pH_i by obtaining the ratio of emission at 535 nm wavelength for excitation wavelengths of 495 nm and 440 nm. The pH_i of non-adherent T cells was obtained by flow cytometry. Cells were resuspended in Earle's balanced salt solution (EBSS) and loaded with 5 nmol/ml of the pH-sensitive dye SNARF-1 for 30 min. The pH_i was determined by obtaining the ratio of emission wavelength of 580 and 640 nm for excitation wavelength of 488 nm. Ratios were converted to pH using in situ calibration curves as described elsewhere (Nehrke, 2006).

T Cell Stimulation

T cells were stimulated to proliferate using recombinant human IL-2 (Invitrogen #CTP0021) and CD3/CD28-coated Dynabeads (Invitrogen #111.31D) according to manufacturer's instructions. 5-ethynyl-2′-deoxyuridine (Invitrogen #C35002) incorporation was assessed by flow cytometry according to manufacturer's instructions.

Additional details are available in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

ChIP-seq and mRNA-seq data are available for download at the NCBI Gene Expression Omnibus under accession number GSE40114.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.10.025.

ACKNOWLEDGMENTS

We thank Klara Olofsdotter Otis and Kelsey Martin for the use of their fluorescent microscope; Yasutada Akiba and Jonathan Kaunitz for initial help with pH_i measurement; and Steven Clarke, Sohail Tavazoie, and Saeed Tavazoie for discussions. M.A.M. was supported in part by a University of California, Los Angeles, Genetics Training Grant fellowship. This work was supported by grants from the California Institute for Regenerative Medicine and the American Cancer Society and by Beckman Young Investigator, Howard Hughes Medical Institute Early Career, and National Institutes of Health Director's Innovator awards to S.K.K.

Received: March 27, 2012
Revised: August 7, 2012
Accepted: October 25, 2012
Published online: November 29, 2012

REFERENCES

Arnold, K.M., Lee, S., and Denu, J.M. (2011). Processing mechanism and substrate selectivity of the core NuA4 histone acetyltransferase complex. Biochemistry *50*, 727–737.

Benjamin, A.M., Murthy, C.R., and Quastel, J.H. (1983). Calcium-dependent release of acetyl-coenzyme A from liver mitochondria. Can. J. Physiol. Pharmacol. *61*, 154–158.

Bental, M., and Deutsch, C. (1994). 19F-NMR study of primary human T lymphocyte activation: effects of mitogen on intracellular pH. Am. J. Physiol. *266*, C541–C551.

Boron, W.F., Siebens, A.W., and Nakhoul, N.L. (1988). Role of monocarboxylate transport in the regulation of intracellular pH of renal proximal tubule cells. Ciba Found. Symp. *139*, 91–105.

Cardone, R.A., Casavola, V., and Reshkin, S.J. (2005). The role of disturbed pH dynamics and the Na+/H+ exchanger in metastasis. Nat. Rev. Cancer *5*, 786–795.

Carmen, A.A., Rundlett, S.E., and Grunstein, M. (1996). HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex. J. Biol. Chem. *271*, 15837–15844.

Casey, J.R., Grinstein, S., and Orlowski, J. (2010). Sensors and regulators of intracellular pH. Nat. Rev. Mol. Cell Biol. *11*, 50–61.

Chung, Y.L., Troy, H., Kristeleit, R., Aherne, W., Jackson, L.E., Atadja, P., Griffiths, J.R., Judson, I.R., Workman, P., Leach, M.O., and Belouecheb-Babari, M. (2008). Noninvasive magnetic resonance spectroscopic pharmacodynamic markers of a novel histone deacetylase inhibitor, LAQ824, in human colon carcinoma cells and xenografts. Neoplasia *10*, 303–313.

Elsheikh, S.E., Green, A.R., Rakha, E.A., Powe, D.G., Ahmed, R.A., Collins, H.M., Soria, D., Garibaldi, J.M., Paish, C.E., Ammar, A.A., et al. (2009). Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. Cancer Res. *69*, 3802–3809.

Fellenz, M.P., and Gerweck, L.E. (1988). Influence of extracellular pH on intracellular pH and cell energy status: relationship to hyperthermic sensitivity. Radiat. Res. *116*, 305–312.

Ferrari, R., Su, T., Li, B., Bonora, G., Oberai, A., Chan, Y., Sasidharan, R., Berk, A.J., Pellegrini, M., and Kurdistani, S.K. (2012). Reorganization of the host epigenome by a viral oncogene. Genome Res. *22*, 1212–1221.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat. Genet. *37*, 391–400.

Friis, R.M., and Schultz, M.C. (2009). Untargeted tail acetylation of histones in chromatin: lessons from yeast. Biochem. Cell Biol. *87*, 107–116.

Friis, R.M., Wu, B.P., Reinke, S.N., Hockman, D.J., Sykes, B.D., and Schultz, M.C. (2009). A glycolytic burst drives glucose induction of global histone acetylation by picNuA4 and SAGA. Nucleic Acids Res. *37*, 3969–3980.

Halestrap, A.P., and Meredith, D. (2004). The SLC16 gene family-from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. Pflugers Arch. *447*, 619–628.

Harguindey, S., Orive, G., Luis Pedraz, J., Paradiso, A., and Reshkin, S.J. (2005). The role of pH dynamics and the Na+/H+ antiporter in the etiopathogenesis and treatment of cancer. Two faces of the same coin—one single nature. Biochim. Biophys. Acta *1756*, 1–24.

Horwitz, G.A., Zhang, K., McBrian, M.A., Grunstein, M., Kurdistani, S.K., and Berk, A.J. (2008). Adenovirus small e1a alters global patterns of histone modification. Science *321*, 1084–1085.

Hulikova, A., Vaughan-Jones, R.D., and Swietach, P. (2011). Dual role of CO2/HCO3(-) buffer in the regulation of intracellular pH of three-dimensional tumor growths. J. Biol. Chem. *286*, 13815–13826.

Iwabata, H., Yoshida, M., and Komatsu, Y. (2005). Proteomic analysis of organ-specific post-translational lysine-acetylation and -methylation in mice by use of anti-acetyllysine and -methyllysine mouse monoclonal antibodies. Proteomics *5*, 4653–4664.

Kim, U., Shu, C.W., Dane, K.Y., Daugherty, P.S., Wang, J.Y., and Soh, H.T. (2007). Selection of mammalian cells based on their cell-cycle phase using dielectrophoresis. Proc. Natl. Acad. Sci. USA *104*, 20708–20712.

Koeller, K.M., Haggarty, S.J., Perkins, B.D., Leykin, I., Wong, J.C., Kao, M.C., and Schreiber, S.L. (2003). Chemical genetic modifier screens: small molecule trichostatin suppressors as probes of intracellular histone and tubulin acetylation. Chem. Biol. *10*, 397–410.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lora-Michiels, M., Yu, D., Sanders, L., Poulson, J.M., Azuma, C., Case, B., Vujaskovic, Z., Thrall, D.E., Charles, H.C., and Dewhirst, M.W. (2006). Extracellular pH and P-31 magnetic resonance spectroscopic variables are related to outcome in canine soft tissue sarcomas treated with thermoradiotherapy. Clin. Cancer Res. *12*, 5733–5740.

Lorenz, M.A., Burant, C.F., and Kennedy, R.T. (2011). Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. Anal. Chem. *83*, 3406–3414.

Manuyakorn, A., Paulus, R., Farrell, J., Dawson, N.A., Tze, S., Cheung-Lau, G., Hines, O.J., Reber, H., Seligson, D.B., Horvath, S., et al. (2010). Cellular histone modification patterns predict prognosis and treatment response in resectable pancreatic adenocarcinoma: results from RTOG 9704. J. Clin. Oncol. *28*, 1358–1365.

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr. Opin. Genet. Dev. *15*, 163–176.

Messonnier, L., Kristensen, M., Juel, C., and Denis, C. (2007). Importance of pH regulation and lactate/H+ transport capacity for work production during supramaximal exercise in humans. J. Appl. Physiol. *102*, 1936–1944.

Mizuno, S., Demura, Y., Ameshima, S., Okamura, S., Miyamori, I., and Ishizaki, T. (2002). Alkalosis stimulates endothelial nitric oxide synthase in cultured human pulmonary arterial endothelial cells. Am. J. Physiol. Lung Cell. Mol. Physiol. *283*, L113–L119.

Moellering, R.E., Black, K.C., Krishnamurty, C., Baggett, B.K., Stafford, P., Rain, M., Gatenby, R.A., and Gillies, R.J. (2008). Acid treatment of melanoma cells selects for invasive phenotypes. Clin. Exp. Metastasis *25*, 411–425.

Nehrke, K. (2006). Intracellular pH measurements in vivo using green fluorescent protein variants. Methods Mol. Biol. *351*, 223–239.

Parks, S.K., Chiche, J., and Pouyssegur, J. (2011). pH control mechanisms of tumor survival and growth. J. Cell. Physiol. *226*, 299–308.

Perona, R., and Serrano, R. (1988). Increased pH and tumorigenicity of fibroblasts expressing a yeast proton pump. Nature *334*, 438–440.

Reshkin, S.J., Bellizzi, A., Caldeira, S., Albarani, V., Malanchi, I., Poignee, M., Alunni-Fabbroni, M., Casavola, V., and Tommasino, M. (2000). Na+/H+ exchanger-dependent intracellular alkalinization is an early event in malignant transformation and plays an essential role in the development of subsequent transformation-associated phenotypes. FASEB J. *14*, 2185–2197.

Rofstad, E.K., Mathiesen, B., Kindem, K., and Galappathi, K. (2006). Acidic extracellular pH promotes experimental metastasis of human melanoma cells in athymic nude mice. Cancer Res. *66*, 6699–6707.

Seligson, D.B., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M., and Kurdistani, S.K. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. Nature *435*, 1262–1266.

Seligson, D.B., Horvath, S., McBrian, M.A., Mah, V., Yu, H., Tze, S., Wang, Q., Chia, D., Goodglick, L., and Kurdistani, S.K. (2009). Global levels of histone modifications predict prognosis in different cancers. Am. J. Pathol. *174*, 1619–1628.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

Suzuki, J., Chen, Y.Y., Scott, G.K., Devries, S., Chin, K., Benz, C.C., Waldman, F.M., and Hwang, E.S. (2009). Protein acetylation and histone deacetylase expression associated with malignant breast cancer progression. Clin. Cancer Res. *15*, 3163–3171.

Takahashi, H., McCaffery, J.M., Irizarry, R.A., and Boeke, J.D. (2006). Nucleocytosolic acetyl-coenzyme a synthetase is required for histone acetylation and global transcription. Mol. Cell *23*, 207–217.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

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

Trickett, A., and Kwan, Y.L. (2003). T cell stimulation and expansion using anti-CD3/CD28 beads. J. Immunol. Methods *275*, 251–255.

Valli, M., Sauer, M., Branduardi, P., Borth, N., Porro, D., and Mattanovich, D. (2005). Intracellular pH distribution in Saccharomyces cerevisiae cell populations, analyzed by flow cytometry. Appl. Environ. Microbiol. *71*, 1515–1521.

Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. (2000). Global histone acetylation and deacetylation in yeast. Nature *408*, 495–498.

Wahl, M.L., Owen, J.A., Burd, R., Herlands, R.A., Nogami, S.S., Rodeck, U., Berd, D., Leeper, D.B., and Owen, C.S. (2002). Regulation of intracellular pH in human melanoma: potential therapeutic implications. Mol. Cancer Ther. *1*, 617–628.

Walter, A., and Gutknecht, J. (1984). Monocarboxylic acid permeation through lipid bilayer membranes. J. Membr. Biol. *77*, 255–264.

Waterborg, J.H. (2002). Dynamics of histone acetylation in vivo. A function for acetylation turnover? Biochem. Cell Biol. *80*, 363–378.

Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R., and Thompson, C.B. (2009). ATP-citrate lyase links cellular metabolism to histone acetylation. Science *324*, 1076–1080.

Wong, P., Kleemann, H.W., and Tannock, I.F. (2002). Cytostatic potential of novel agents that inhibit the regulation of intracellular pH. Br. J. Cancer *87*, 238–245.

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

Zhang, X., Lin, Y., and Gillies, R.J. (2010). Tumor pH and its measurement. J. Nucl. Med. *51*, 1167–1170.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.

Molecular Cell, Volume 49

Supplemental Information

Histone Acetylation Regulates Intracellular pH

Matthew A McBrian, Iman Saramipoor Behbahan, Roberto Ferrari, Trent Su, Ta-Wei Huang, Kunwu Li, Candice S. Hong, Heather R. Christofk, Maria Vogelauer, David B. Seligson, and Siavash K. Kurdistani

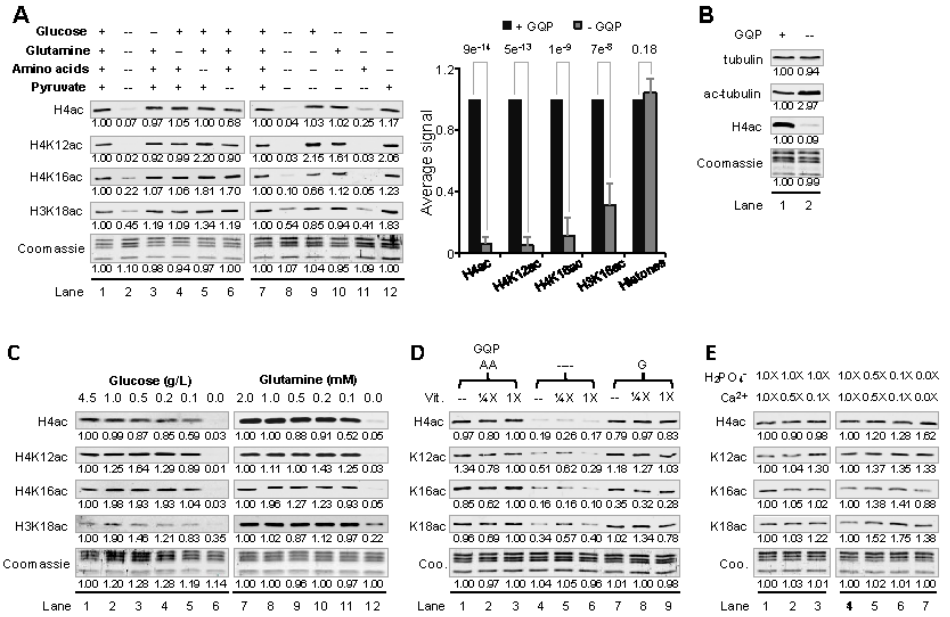


Figure S1, Related to Figure 1. Minimal Levels of Glucose, Glutamine or Pyruvate Maintain Global Levels of Histone Acetylation Independently of Vitamins and Salts. (**A**) Western blots (WBs) of histone acetylation in 231 cells cultured for 16 hrs in DMEM salts and vitamins with the indicated ac-CoA sources at concentrations found in standard DMEM. The bar graph shows average values ± standard deviation from ten independent experiments and p values derived from the Student's t-test. (**B**) WBs of tubulin and histone H4 acetylation from 231 cells

under conditions of glucose (G), glutamine (Q) and pyruvate (P) deprivation. (**C**) WBs of histone acetylation from 231 cells cultured at the indicated concentrations of G or Q in DMEM salts and vitamins. (**D**) WBs showing the effects of varying vitamins concentration on histone acetylation with the indicated ac-CoA source. (**E**) WBs showing the effects of varying concentrations of Ca^{2+} and phosphate on histone acetylation in 231 cells. Lane 1 in each panel – except for panel D – is the reference condition, the values of which are set to 1. In panel D, the reference condition is lane 3.

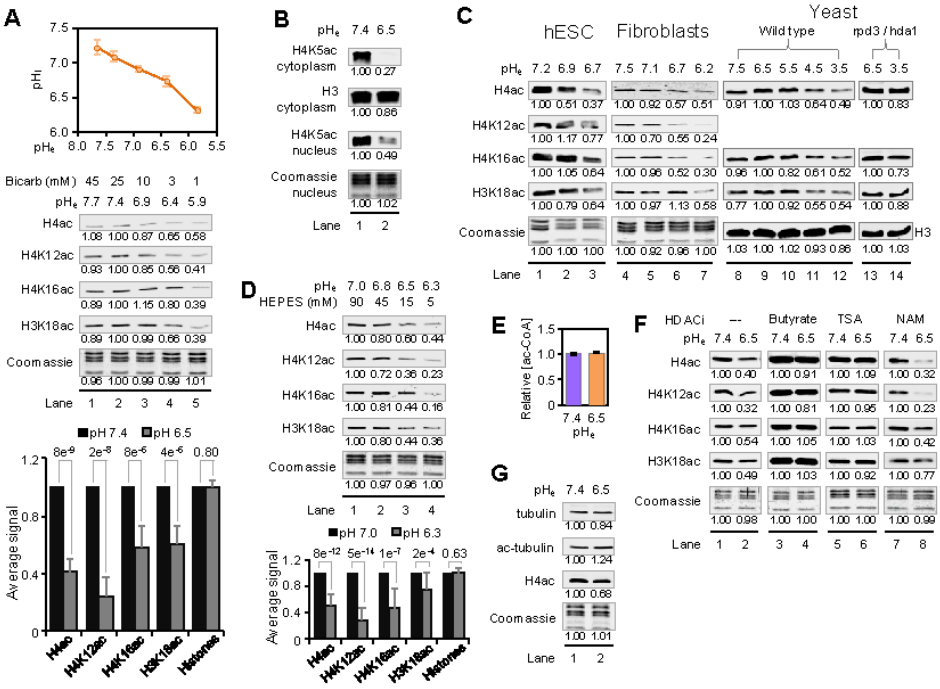


Figure S2, Related to Figure 2. Global Levels of Histone Acetylation Change in Response to pH Alterations. (**A**) Measurements of pH_i (mean ± standard deviation) as a function of pH_e and effects on histone acetylation in 231 cells cultured at the indicated pH_e for 16 hrs in complete DMEM. The reference condition is set to 25 mM bicarbonate (lane 2), as this approximates the

normal physiological concentration. The bar graph shows average values ± standard deviation from ten independent experiments and p values derived from the Student's t-test. (**B**) Western blot (WB) analysis of H4K5ac in 231 cells in the indicated histone fractions and pH. (**C**) WBs of acetylation in H1 human embryonic stem cells cultured for 3 hrs in m-TESR medium at the indicated pH_e, or IMR90 normal primary lung fibroblasts treated for 16 hrs in complete DMEM at the indicated pH_e or budding yeast (YDS2) grown for 4 hrs in YPD at the indicated pH. Histone H3 was used as loading control for the yeast experiments. Note that pH has little effect on histone acetylation in the strain with Rpd3 and Hda1 deleted. (**D**) WBs of histone acetylation in 231 cells cultured in media buffered with HEPES. The bar graph shows average values ± standard deviation from twenty one independent experiments and p values derived from the Student's t-test. (**E**) Ac-CoA measurements (mean ± standard deviation) of 231 cells treated for 4 hrs at the indicated pH_e. (**F**) WBs of histone acetylation in 231 cells that were treated with or without 5 mM sodium butyrate, 500 nM TSA or 2 mM nicotinamide (NAM) for 6 hrs in complete DMEM at pH_e 7.4 followed by incubation at the indicated pH_e for 4 hrs. (**G**) WBs of tubulin and histone H4 acetylation in 231 cells treated for 16 hrs at the indicated pH values.

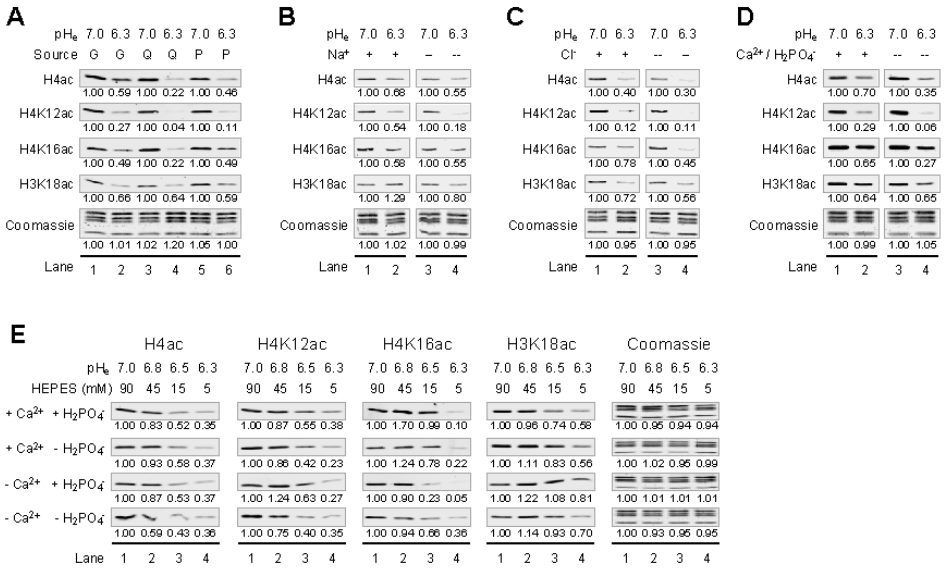


Figure S3, Related to Figure 3. Changes in Histone Acetylation Levels in Response to pH Do Not Require Specific Carbon Sources or Salts. Western blots of histone acetylation in 231 cells cultured for 16 hrs in DMEM salts at pH_e 7.0 or 6.3 (**A**) including the indicated carbon source or lacking (**B**) Na^+, (**C**) Cl^-, and (**D-E**) Ca^{2+} and phosphate. In panel **D**, cells were starved of Ca^{2+} and phosphate for 3 days prior to pH_e treatment, while in panel **E**, Ca^{2+} and phosphate were removed from the culture medium at the time of pH_e treatment.

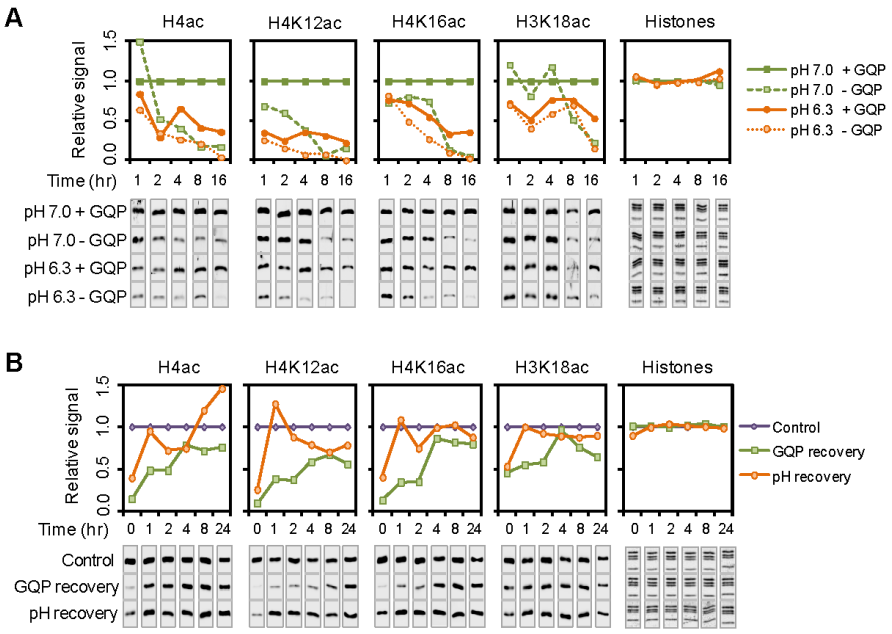


Figure S4, Related to Figure 5. Nutrient Availability and pH Alter Acetylation Dynamics Differently in 231 Cells. Western blots of histone acetylation in 231 cells cultured in DMEM salts (**A**) for the indicated amount of time and pH_e in the presence (+) and absence (-) of GQP and (**B**) at pH_e 7.4 with GQP (control), pH_e 7.4 without GQP (for GQP recovery) or pH_e 6.5 with GQP (for pH recovery) for 16 hrs followed by treatment for the indicated amount of time in medium at pH_e 7.4 with GQP. Membrane images were cropped in order to place them below the graph at the corresponding time points.

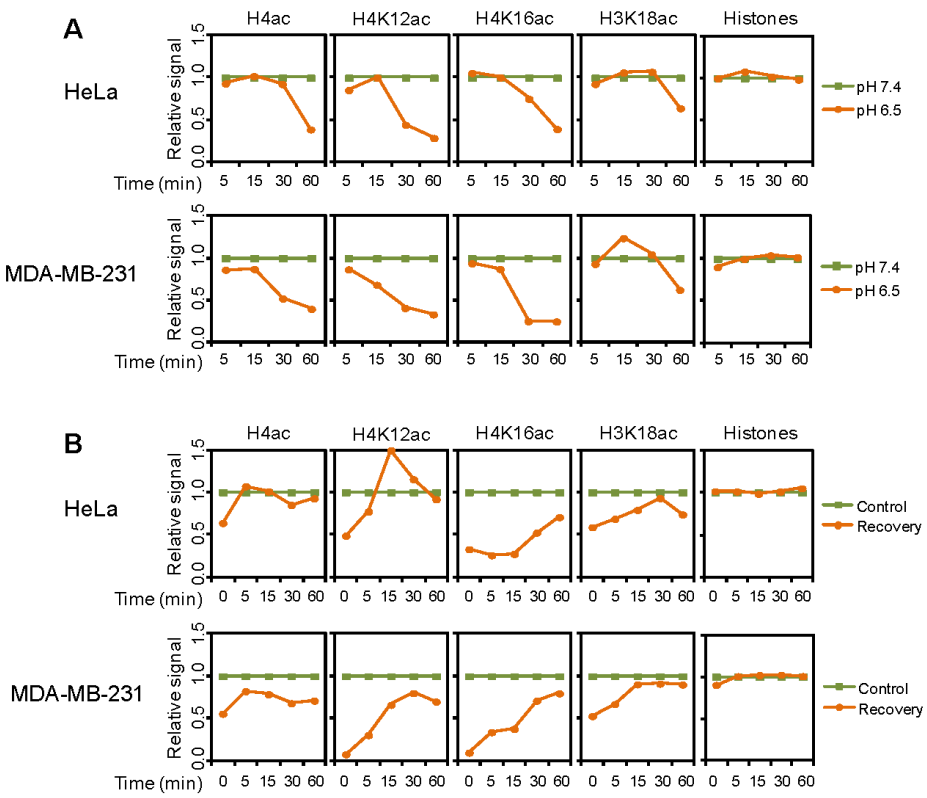


Figure S5, Related to Figure 6. Loss and Recovery of Acetylation by pH Treatment is Rapid. Western blots of histone acetylation from HeLa or 231 cells cultured in DMEM (**A**) for the indicated time and pH, or (**B**) for 16 hrs in DMEM at pH 7.4 (control) or 6.5 (recovery) followed by treatment at pH 7.4 for the indicated amount of time. Note that the recovery of acetylation begins within 5 min.

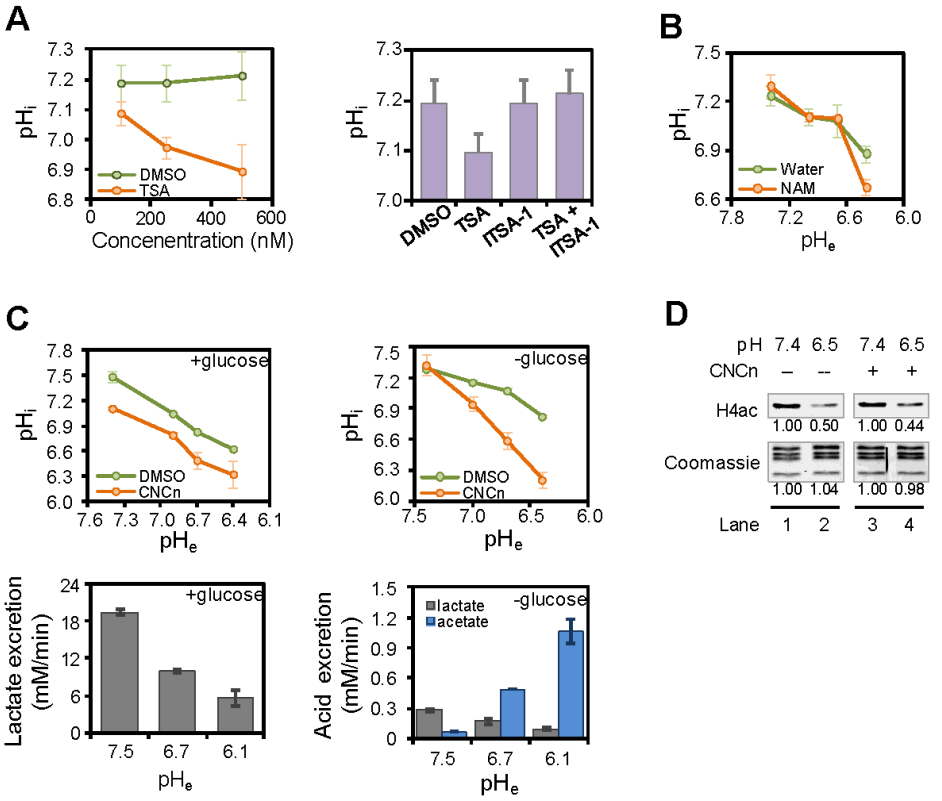


Figure S6, Related to Figure 7. Global Histone Acetylation and Deacetylation is Linked to pH_i. pH_i of 231 cells treated in standard DMEM overnight with (**A**) the indicated concentration of TSA (left panel); or 250 nM TSA, 50 μM ITSA-1 or both (right panel); or (**B**) 2 nM nicotinamide (NAM). (**C**) pH_i of 231 cells in which MCT function was inhibited by treatment for 1 hr with 10 mM CNCn in DMEM with (top left panel) and without (top right panel) glucose along with rates of acetate and lactate excretion (bottom panels) under identical conditions. Note that in absence of glucose, acetate excretion at high pH is reduced to minimal levels. In contrast, rate of acetate excretion at low pH is not dependent on the presence of glucose. (**D**) Western blots of histone acetylation from 231 cells treated for 4 hrs in DMEM at pH 7.4 with 10 mM CNCn followed by treatment with 10 mM CNCn at the indicated pH_e for an additional 2 hrs. Data in **A-C** is presented as mean ± standard deviation.

Table S1, Related to Figure 4. Gene Ontology Analysis of Gene Expression Changes in HeLa Cells in Response to pH.

Gene Ontology Terms	Benjamini p value
Up-regulated >2-fold at low pH	
Histone core	3.21E-07
Basic-leucine zipper (bZIP) transcription factor	5.78E-05
MAPK signaling pathway	1.63E-04
Regulation of transcription from RNA polymerase II promoter	2.77E-04
Positive regulation of nitrogen compound metabolic process	3.72E-04
Positive regulation of cellular biosynthetic process	8.12E-04
Response to endogenous stimulus	7.19E-03
Down-regulated >2-fold at low pH	
Plasma membrane	2.92E-06
Glycoprotein	1.75E-05
Extracellular matrix	8.22E-04

Supplemental Experimental Procedures

Cell culture

HeLa, MDA-MB-231 and IMR90 cells were generally cultured in DMEM without glucose, glutamine, phenol red, sodium pyruvate and sodium bicarbonate (Sigma #D5030) to which was added 1 g/L glucose, 2 mM glutamine, 1 mM sodium pyruvate 26 mM sodium bicarbonate and 1X antibiotic/antimycotic (Gibco #15240). Medium pH was adjusted to 7.2 by the addition of HCl before being filter sterilized. Dialyzed fetal bovine serum (FBS) (Gibco#26400) was re-dialyzed against 150 mM NaCl using 10,000 molecular weight cutoff dialysis tubing (SpectrumLabs#08670252) to remove trace nutrients (Wice et al., 1981). Dialysis was performed at 4°C for 2 hrs for a total of 3 times in 6 L dialysate each time. Redialyzed serum was added to media at a concentration of 5%. The use of dialyzed serum significantly improved the reproducibility of data.

Media preparation

Custom media were prepared where appropriate as indicated in the main text. DMEM consists of salts, vitamins, amino acids, glucose, glutamine, pyruvate and bicarbonate. For experiments in DMEM salts, we followed the recipe from Sigma (#E7510) with the exception that glucose and phenol red were omitted. Experiments in which sodium was omitted were done using the same salt recipe but substituting an equimolar mixture of lithium chloride and potassium chloride for sodium chloride, HEPES for sodium bicarbonate, and potassium phosphate for sodium phosphate. Experiments in which chloride was omitted were done by substituting sodium nitrate for sodium chloride, potassium nitrate for potassium chloride, and calcium nitrate for calcium chloride. Experiments in which calcium and phosphate were omitted were done by simply excluding them from the salt formulation since they contribute very little to the overall tonicity of the medium. For experiments in DMEM salts and vitamins, we added vitamins purchased from Sigma (#M6895) to our DMEM salts solution. For experiments in complete DMEM we used Sigma powder DMEM (#D5030) which was prepared as described above. The pH of the media was varied by altering the buffering component of the media.

Adding different amounts of bicarbonate resulted in the pH as shown in figures 2A and S2A when incubated at 37°C and 5% CO_2. Media pH was also adjusted by adding varying amounts of HEPES and then adjusting the pH to 7.2 by the addition of acid or base. This resulted in the pH as shown in figures 2D and S2D when incubated at 37°C and 5% CO_2. All pH_e measurements were taken by bubbling 5% CO_2 into the media for 30 minutes (min) prior to taking measurements using a standard pH meter. Media pH did not change more than 0.2 units during any of the experiments performed. For H1 hESCs, the pH of media was varied by adding adequate acid while bubbling 5% CO_2. The pH of yeast growth media (YPD) was adjusted by addition of acid or base.

Western blotting

Western blots were performed on acid-extracted histones and on whole cell lysates and results were quantified using the Odyssey Infrared Imaging System (LI-COR Biosciences). This system utilizes fluorescently-labeled secondary antibodies which allow for fluorescent intensity values to be recorded upon excitation with a fluorescent scanner. These values are reported as relative values below images of the fluorescently scanned membranes which were converted to grayscale. Westerns using the histone antibodies were performed with 15% polyacrylamide gels while those done on the whole cell lysates were run on 4-12% Bis-Tris gradient gels (Invitrogen #NPO321BOX). In both cases, samples were prepared to sufficient volume to load all gels for the experiment which were run in parallel. Transfer was done using the iBlot system with Immobilon-FL membrane (Millipore #IPFL00010) replacing the PVDF membrane provided (Invitrogen #IB4010-01). Primary antibodies used were H4ac (Millipore #07-352), H4K16ac (Active Motif #39167), H3K18ac (Active Motif #39587), H4K5ac (Millipore #07-327), H3K4me2 (Abcam #32356), K3K9me2 (Millipore #07-441), H3K27me3 (Millipore #07-449), H3K36me2 (Millipore #07-274), H3 (Abcam #1791), tubulin (BD Biosciences #556321), acetyl-tubulin (Sigma #T6793), MCT1 (Sigma #HPA003324); H4K12ac was kindly provided by Michael Grunstein.

Acid extracted histones were prepared from a 10-cm plate of cells at roughly 70% density. All reagents used were chilled to 4°C and spins were done using a chilled microfuge

maintained at 4°C. Briefly, nuclei were collected following hypotonic lysis of cells and then acid-extracted with H_2SO_4. The proteins in the extraction were collected by TCA precipitation and dissolved in 100 µl of 95°C H_2O. Protein concentration was estimated with the BCA assay (Thermo Scientific #23225) using the manufacturer's protocol. The histone solution was then diluted in an equal volume of 2X SDS protein buffer (SPB) and boiled for 5 min. Histone protein concentration was then normalized between samples by using the Odyssey system to quantify the four histone bands on SimplyBlue (Invitrogen #LC6060) stained gels loaded with 1 µg of protein (as determined by BCA). The loading amount was then adjusted to ensure even loading of histones. This method of loading normalization consistently resulted in an even loading of histone proteins much better than using various antibodies recognizing total histone H3 or H4. Even loading of histone proteins was a critical factor in accurately assessing levels of modification.

Whole cell lysates were prepared from a 10-cm plate by washing the cells with PBS and then adding 750 µl of 1X SPB. Samples were then boiled and concentration was approximated between samples by comparison of SimplyBlue stained gels using the Odyssey system. Histone bands as well as several other prominent bands were quantified and used to adjust the loading amounts for subsequent westerns. Preparation of yeast histones by TCA-precipitation was performed as described previously (Peng et al., 2008) and loading between samples was normalized using results from immunoblots against total H3.

^{3}H-Acetate excretion assay

Culture medium pH is typically established by the concentration of bicarbonate in the medium and the level of CO_2 in the atmosphere. Cells in culture are usually grown in medium containing 26 mM bicarb in a 37°C incubator at 5% CO_2 which results in a medium pH of around 7.4. In this particular assay, cells were taken in and out of a 37°C incubator frequently during the course of the experiment. This did not allow for a constant concentration of CO_2; therefore, the experiment was performed at 37°C at atmospheric CO_2 conditions. As such, in order to obtain media of varying pH, we added the amount of bicarbonate that would have resulted in varying pH at 5% CO2 as shown in figures 2A and S2A. We then adjusted the pH of

the solution by the addition of acid or base to equal the pH that would have been attained at 5% CO2 with the given amount of bicarbonate. Medium at pH 7.4 contained 25 mM bicarbonate, pH 6.8 had 7 mM bicarbonate, pH 6.4 had 3 mM bicarbonate and pH 5.9 had 1 mM bicarbonate.

To start the experiment, $2x10^5$ cells were plated per well of a 6-well plate in DMEM (pH 7.4) with 10% FBS. 48hrs later cells were labeled for 1 hour (hr) in 3 ml DMEM (pH 7.4) with 66 μCi/ml of ^{3}H-acetate (Moravek #634). Following the label, a 30 minute chase was performed in 3 ml DMEM (pH 7.4). Chase medium was then removed for counting purposes and 1 ml of DMEM (pH 7.4) was incubated for 10 min to assess the rate of label extrusion prior to changes in media pH. (This rate is indicated in graphs as time point 0.) Beginning at time point 0, cells were incubated in 1 ml media of varying pH for different time intervals. At each time point, 800 μl of media from each well was collected for counting purposes, the residual media was discarded and 1 ml fresh media was added. The appearance of label was monitored at each time point for each medium pH by a liquid scintillation counter. Counts were normalized between experiments using time point 0 to correct for differences between labeling efficiency.

For experiments with α-cyano-4-hydroxycinnamate (CNCn; Sigma #595667), the reagent was dissolved in DMSO as a 500 mM stock and added to media at a final concentration of 10 mM (Wahl et al., 2002). Control media were prepared by diluting DMSO into medium at a ratio of 1:50. The pH of CNCn and DMSO control media was adjusted after CNCn or DMSO addition and was then used during the chase and all subsequent steps. Experiments with TSA were performed by diluting a 5 mM TSA in DMSO solution (Sigma #T1952) to a final concentration of 500 nM. Control media were prepared by diluting DMSO into medium at a ratio of 1:10,000. Cells were incubated in TSA (or DMSO control) media for 24 hrs prior to labeling. All subsequent steps were done using TSA or DMSO control media.

Intracellular pH measurement

Adherent cells were grown on 35-mm poly lysine coated glass bottom culture dishes (MatTek, Ashland, MA) to 40% confluency in DMEM plus 10% FBS and 1% Anti-Anti (Gibco #15240). All pH_i measurement solutions were bubbled with 5% CO_2 and contained the

appropriate amount of bicarbonate to obtain the desired pH according to the Henderson-Hasselbach equation. Cells were washed with Earle's balanced salt solution (EBSS) and then loaded with 5 nmol/ml of the pH-sensitive dye 2′,7′-bis(2-carboxyethyl)-5-(and-6)carboxyfluorescein, acetoxymethylester (BCECF-AM; Molecular Probes #B1170) in EBSS for 25 min in the 37°C incubator with 5% CO_2. pH_i was measured immediately after loading by obtaining measures of fluorescent intensity at an emission wavelength of 535 nm, with alternating excitation wavelengths of 495 and 440. Images were obtained with an Axiovert 200M Zeiss florescent microscope equipped with a Chroma BCECF filter set (Chroma #71001a) and a high-resolution video camera (Axio CAM MRm, Zeiss, Germany). Ratio between 495/440 channels was calculated for a field of cells by AxioVision 4.8 and Slidebook 4.2 after subtracting background in each channel. During the experiment, cells were superfused with 37°C EBSS at with a chamber insert diamond medium bath (Warner Instruments, Hamden, CT) at a rate of 1 ml/min while pH was monitored continuously.

pH_i of non-adherent T-cells was measured using a four-laser digital BD LSRII flow cytometer as described previously with some modifications (Chow and Hedley, 2001). Cells were resuspended in loading solution (EBSS with 26 mM bicarbonate and 20 mM HEPES with the pH adjusted to 7.4 by the addition of sodium hydroxide) and loaded with 5 nmol/ml of 5-(and-6)-Carboxy SNARF®-1, Acetoxymethyl Ester, Acetate (SNARF-1; Molecular Probes #C1272) for 30 min at 37°C and 5% CO_2. After loading, cells were resuspended in loading solution or high potassium calibration solutions at various pH containing 10 μM nigericin. After 20 min of incubation at atmospheric conditions cells were loaded in the flow cytometer and fluorescent intensities were collected at emission wavelengths of 580 nm and 640 nm for an excitation wavelength of 488 nm using linear amplification. Fluorescent ratios of 640/580 nm were used for determination of pH_i. Data were analyzed using Flow Jo software.

Ratios from both types of experiments were converted to pH by in situ calibration for each set of experiments using nigericin and high potassium isotonic solution as described previously (Nehrke, 2006).

Supplementary References

Chow, S., and Hedley, D. (2001). Flow cytometric measurement of intracellular pH. Curr Protoc Cytom *Chapter 9*, Unit 9 3.

Nehrke, K. (2006). Intracellular pH measurements in vivo using green fluorescent protein variants. Methods Mol Biol *351*, 223-239.

Peng, W., Togawa, C., Zhang, K., and Kurdistani, S.K. (2008). Regulators of cellular levels of histone acetylation in Saccharomyces cerevisiae. Genetics *179*, 277-289.

Wahl, M.L., Owen, J.A., Burd, R., Herlands, R.A., Nogami, S.S., Rodeck, U., Berd, D., Leeper, D.B., and Owen, C.S. (2002). Regulation of intracellular pH in human melanoma: potential therapeutic implications. Mol Cancer Ther *1*, 617-628.

Wice, B.M., Reitzer, L.J., and Kennell, D. (1981). The continuous growth of vertebrate cells in the absence of sugar. J Biol Chem *256*, 7812-7819.

Chapter 5

Reorganization of the host epigenome

by a viral oncogene

This chapter was originally published in *Genome Research.* 2012 Jul;22(7):1212-21. with supplemental material. This chapter describes changes in epigenetic landscape of primary human lung fibroblast cells after infection with Adenovirus *dl1500*. Adenovirus *dl1500* forces expression of small e1a oncoprotein in primary cells. As previous shown by Ferrari etl al and Horwitz et al. (Ferrari, Pellegrini et al. 2008; Horwitz, Zhang et al. 2008), small e1a causes drastic reduction in global H3K18ac levels. Using ChIP-seq method, changes in global distribution of H3K18ac was revealed. Drastic reduction of H3K18ac was observed in the intergenic region and new peaks of H3K18ac were gained at Rb-E2F binding site at promoter of cell cycle genes. This study provided the first genome-wide observation of viral oncogene induced global redistribution of H3K18ac and H3K9ac in relationship to guardians of G1/S transition, Rb family of proteins, in normal contact-inhibited cells.

Research

Reorganization of the host epigenome by a viral oncogene

Roberto Ferrari,[1,2] Trent Su,[1,3] Bing Li,[1] Giancarlo Bonora,[1,4] Amit Oberai,[1] Yvonne Chan,[1] Rajkumar Sasidharan,[5] Arnold J. Berk,[6,7] Matteo Pellegrini,[2,5] and Siavash K. Kurdistani[1,2,7,8,9]

[1]*Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA;* [2]*Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, California 90095, USA;* [3]*Division of Oral Biology and Medicine, School of Dentistry, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA;* [4]*UCLA Bioinformatics Interdepartmental Degree Program, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA;* [5]*Department of Molecular, Cellular, and Developmental Biology, University of California, Los Angeles, California 90095, USA;* [6]*Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California 90095, USA;* [7]*Molecular Biology Institute, University of California, Los Angeles, California 90095, USA;* [8]*Department of Pathology and Laboratory of Medicine, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA*

Adenovirus small e1a oncoprotein causes ~70% reduction in cellular levels of histone H3 lysine 18 acetylation (H3K18ac). It is unclear, however, where this dramatic reduction occurs genome-wide. ChIP-sequencing revealed that by 24 h after expression, e1a erases 95% of H3K18ac peaks in normal, contact-inhibited fibroblasts and replaces them with one-third as many at new genomic locations. The H3K18ac peaks at promoters and intergenic regions of genes with fibroblast-related functions are eliminated after infection, and new H3K18ac peaks are established at promoters of highly induced genes that regulate cell cycling and at new putative enhancers. Strikingly, the regions bound by the retinoblastoma family of proteins in contact-inhibited fibroblasts gain new peaks of H3K18ac in the e1a-expressing cells, including 55% of RB1-bound loci. In contrast, over half of H3K9ac peaks are similarly distributed before and after infection, independently of RB1. The strategic redistribution of H3K18ac by e1a highlights the importance of this modification for transcriptional activation and cellular transformation as well as functional differences between the RB-family member proteins.

[Supplemental material is available for this article.]

Adenovirus 5 (Ad5), a DNA tumor virus, encodes a highly conserved 243-amino acid protein named small E1A protein (e1a) that drives quiescent mammalian cells into S-phase by overcoming cellular processes that normally inhibit inappropriate cell cycling (Berk 2005). The ability of e1a to transform cells depends on its interactions with several host cell proteins, particularly the retinoblastoma (RB)-family proteins (RB1, RBL1 [p107] and RBL2 [p130]) and the closely related EP300 and CREBBP lysine acetyltransferases (KATs) (Sherr and McCormick 2002; Berk 2005; Ferrari et al. 2008; Horwitz et al. 2008). We showed previously that during the initial 24 h of infection, e1a associates with a large fraction of human gene promoters in a temporally ordered manner, resulting in redistribution of host transcription cofactors such as RB1 and EP300 and reprogramming of gene expression for cell replication (Ferrari et al. 2008). The interaction of e1a with EP300/CREBBP KATs (Horwitz et al. 2008), the major enzymes responsible for H3K18ac in vivo, causes ~70% reduction in total cellular levels of H3K18ac but not of several other histone modifications (Horwitz et al. 2008; Jin et al. 2011). The dynamic binding of e1a to the genome was accompanied by relative reduction of H3K18ac at most gene promoters but increased levels at promoters of genes involved in cell cycling and DNA replication (Ferrari et al. 2008).

[9]**Corresponding author**
E-mail skurdistani@mednet.ucla.edu
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.132308.111.

The previous data were obtained using chromatin immunoprecipitation combined with DNA microarrays (ChIP-chip) that contained probes covering primarily gene promoters. Since promoter regions represent a small fraction of the genome, the ChIP-chip experiments did not provide a molecular explanation for the global decrease in the levels of H3K18ac or how it differs from H3K9ac, another histone acetylation site that is also associated with gene expression.

In this study, we used ChIP coupled with massive parallel sequencing (ChIP-seq) (Johnson et al. 2007) to determine the genome-wide distributions of H3K18ac and H3K9ac in human primary lung fibroblasts (IMR90) before and after infection with the Ad5 mutant *dl*1500 which expresses e1a with little or no expression of other viral genes (Montell et al. 1982). We found that 95.4% of peaks of H3K18ac that exist prior to expression of e1a were deacetylated by 24 h after infection, essentially erasing the normal pattern of H3K18ac. New peaks of H3K18ac are established primarily at the promoter regions of highly induced genes that function in cell replication. Approximately 25% of these new peaks are found in regions that were bound by RB1 and to some extent by RBL2 and RBL1 prior to e1a expression. Altogether, the e1a-expressing cells have 33% as many peaks of H3K18ac as in uninfected cells, accounting for the global decrease in H3K18ac. In contrast, the total number of peaks of H3K9ac increased by 28.2% after infection. Interestingly, we also found distinct peaks of nucleosomes acetylated on H3K9 and H3K18 associated with the viral genome.

Results

General experimental approach

IMR90 fibroblasts were grown to confluence and cultured an additional day before they were either mock-infected or infected with *dl*1500. Twenty-four h post-infection (p.i.), cells were harvested for mRNA-seq and ChIP-seq analyses. For all ChIP-seq experiments, the input DNA was also sequenced. Sequenced reads were aligned to the human genome (Hg19), and only those that matched a unique location with up to two sequence mismatches were retained (Supplemental Table S1). To define peaks of enrichment, we segmented the human genome into 100-bp windows and compared the ChIP and normalized input DNA read counts in each window. Using the Poisson distribution, we calculated *P*-values (*P*-val) for the enrichment of ChIP reads in each window. A cut-off *P*-val was determined for each experiment to maintain a false discovery rate (FDR) ≤ 1.5%. For histone acetylation and RB-family ChIPs, significant windows were defined as those with a *P*-val < 10^{-3} and *P*-val < 10^{-4}, respectively, and with significant peaks at the same *P*-vals in the two neighboring windows (see Supplemental Methods). To detect H3K18ac, we raised a polyclonal antibody (814) that showed specificity, not only in ELISA and peptide dot blots but also in ChIP experiments, with yeast strains harboring mutations in histone H3 lysines (Suka et al. 2001; Supplemental Fig. S1A,B). As expected, cells expressing e1a showed a significant decrease in global H3K18ac levels when using the 814 antibody (Supplemental Fig. S1C). All analyses of H3K18ac enrichment were done in parallel with H3K9ac, since this mark is also associated with transcriptional activation, but a global change in its level in response to e1a expression was not observed (Horwitz et al. 2008; Supplemental Fig. S2A).

ChIP-seq reveals genome-wide hypoacetylation of H3K18 caused by adenovirus e1a

To determine how the distributions of H3K18ac and H3K9ac were affected by e1a expression, we performed ChIP-seq from mock- and *dl*1500-infected cells 24 h p.i. We found 16,660 peaks of H3K18ac in mock- and 5500 in *dl*1500-infected cells, with only 772 peaks shared between the two conditions, indicating that 95.4% of the H3K18ac peaks in mock-infected cells were absent in the infected cells (Fig. 1A). In the e1a-expressing cells, the total

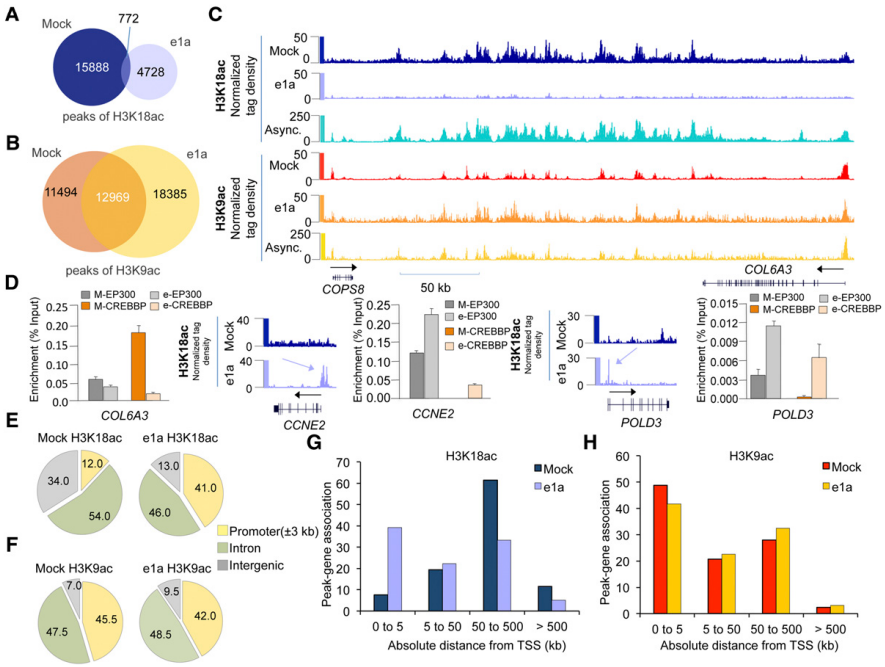


Figure 1. Small e1a causes global deacetylation and redistribution of H3K18ac. (*A*) Venn diagram showing the overlap between significant peaks of H3K18ac in mock- (dark blue) and *dl*1500-infected cells (light blue). (*B*) Venn diagram showing the overlap between significant peaks of H3K9ac in mock- (dark orange) and *dl*1500-infected cells (yellow). (*C*) Patterns of H3K18ac and H3K9ac in the intergenic region between *COPS8* and *COL6A3* genes in mock-, *dl*1500-infected, and asynchronous IMR90 cells. (*D*) Quantitative PCR (% of input) for EP300 and CREBBP in mock- and *dl*1500-infected cells for the *COL6A3* intergenic region, and *CCNE2* and *POLD3* promoters are shown as bar plots. Patterns of H3K18ac in mock- and *dl*1500-infected cells at *CCNE2* and *POLD3* loci are also shown. For each histone modification, the *y*-axis indicates the number of input-normalized ChIP-seq reads across the locus (*x*-axis). (Light blue arrows) New peaks of acetylation in e1a-expressing cells. (Dark arrows) Direction of transcription. (*E*,*F*) Overview of distribution of H3K18ac and H3K9ac peaks in mock- and *dl*1500-infected cells in relation to gene structure are shown as pie charts. (*G*) Distribution of significant peaks of H3K18ac with respect to TSS in mock- (dark blue) and *dl*1500-infected (light blue) cells. (*H*) Distribution of significant peaks of H3K9ac with respect to TSS in mock- (dark orange) and *dl*1500-infected (yellow) cells.

number of H3K18ac peaks was reduced to 33% of the number in mock-infected cells. The number of H3K9ac peaks in infected cells increased by 28.2% compared to the number in mock-infected cells, with 53% of the peaks in mock-infected cells also present in the *dl*1500-infected cells (Fig. 1B). When we compared the peak coverage area (i.e., number of kb covered by all peaks) for H3K18ac and H3K9ac, we found 76.6% lower coverage by H3K18ac but 12% more coverage by H3K9ac in the infected cells (Supplemental Table S1). Thus, the ChIP-seq of H3K18ac in *dl*1500- versus mock-infected cells recapitulated the observed loss of H3K18ac observed by immunofluorescence, Western blotting, and protein mass spectrometry (Horwitz et al. 2008). H3K18ac reduction by e1a is not simply due to cell cycling, since cellular levels of H3K18ac were equivalent in a replicating asynchronous IMR90 cell population (Supplemental Fig. S2A), which has ~35% of cells in S-phase compared to mock-infected contact-inhibited IMR90 cells with only 16% of cells in S-phase (Supplemental Fig. S2B).

Small e1a drastically alters the genomic distribution of H3K18ac

The global decrease of H3K18ac associated with e1a expression (Fig. 1A) was particularly evident at intergenic regions (IR), defined here as >3 kb upstream of the transcription start site (TSS) and >3 kb downstream from the transcription termination site (TTS) of all annotated genes. An example is the ~200-kb locus on chromosome 2 between the *COPS8* and *COL6A3* genes where substantial levels of H3K18ac in mock-infected cells were essentially erased upon infection but maintained in asynchronously growing IMR90 cells (Fig. 1C; Supplemental Fig. S2C). The same IR retained similar levels of H3K9ac in all three conditions (Fig. 1C). ChIP-qPCR showed that loss of H3K18ac in e1a-expressing cells at the *COPS8-COL6A3* locus was accompanied by a decrease in EP300 and CREBBP binding (Fig. 1D). Together with a substantial decrease in H3K18ac, we also observed the establishment of new peaks throughout the genome of e1a-expressing cells. Figure 1D shows two examples of cell cycle-regulated gene promoters, cyclin E2 (*CCNE2*) and DNA polymerase δ subunit 3 (*POLD3*), that gained H3K18ac in the infected cells. The gain of H3K18ac at these gene promoters was also confirmed by qPCR of ChIPed DNA (Fig. 1D; Supplemental Fig. S2D,E). ChIP-qPCR also revealed increased binding of EP300 to *CCNE2* and of EP300 and CREBBP to the *POLD3* promoters (Fig. 1D). Overall, we found that in mock-infected cells, H3K18ac was preferentially found in IRs (34%) and introns (54%), with only 12% of peaks at promoter regions (spanning 3 kb upstream of and downstream from the TSS) (Fig. 1E). The e1a-expressing cells showed a profoundly different distribution of H3K18ac, with IRs and promoter regions harboring 13% and 41% of the peaks, respectively (Fig. 1E). The H3K18ac peak distribution in introns was decreased to some extent in the infected compared to mock-infected cells (54% vs. 46%). The same analysis for H3K9ac showed that the distributions of the peaks in *dl*1500- vs. mock-infected cells were essentially unchanged (Fig. 1F). The redistribution of H3K18ac was also evident when we determined the absolute distance of significant peaks relative to the closest TSS. The majority of the H3K18ac peaks in mock-infected cells were positioned between 50 and 500 kb away from TSSs, with a small fraction within 5 kb (Fig. 1G, dark blue bars). In the *dl*1500-infected cells, however, the fraction of peaks mapping at regions closest to the TSS (5 kb) increased significantly, and the number of more distant peaks (50 to ≥500 kb) was decreased (Fig. 1G, light blue bars). The overall distribution of H3K9ac relative to the nearest TSS did not change substantially in the two conditions (Fig. 1H). We conclude that e1a induces a global decrease and considerable redistribution of H3K18ac by eliminating the majority of peaks in normal cells and establishing new, mostly promoter proximal peaks in the infected cells.

H3K18ac is increased at promoters of genes involved in cell cycling

To further characterize the changes in H3K18ac at promoters, we tiled the data from promoter regions of all genes from −5 to +5 kb of TSS into 100-bp windows. All promoters were grouped into three clusters that best captured the trend of the data based on the levels of H3K18ac in mock and infected cells using an unbiased clustering algorithm (*k*-means). The data are visualized as heat maps of enrichment values (*Z*-scores) in Figure 2A. The levels of H3K9ac for each promoter region are also shown. H3K18ac showed little correlation between mock and infected fibroblasts, while the H3K9ac pattern was more similar between the two conditions (Fig. 2A). The infected cells had 1154 promoters with at least one peak of H3K18ac (Fig. 2A, cluster 1), compared to 436 in mock cells (Fig. 2A, cluster 2). In contrast to H3K9ac, the peaks of H3K18ac in either condition were restricted to cluster 1 and 2 genes, indicating the specificity of H3K18ac distribution. The average change in expression (e1a-expressing vs. mock) of each cluster is shown in Figure 2B (see Methods).

Cluster 1 genes showed little acetylation of H3K18 in mock-infected cells but a significant increase just downstream from the TSS in the infected cells, with little change in levels or distribution of H3K9ac (Fig. 2A). Gene ontology (GO) analysis of this cluster showed enrichment for genes involved in cell cycling, DNA replication, and ribosomes (Fig. 2E). Expression of this cluster was significantly up-regulated in the e1a-expressing cells (Z-score = 6.8, $P < 0.0001$), consistent with induction of S-phase and replication. Cluster 2 genes were associated with fibroblast function (Fig. 2F), which in mock-infected cells had high levels of H3K18ac both upstream of and downstream from the TSS. Upon infection, cluster 2 genes were significantly deacetylated on H3K18 and H3K9 (Fig. 2A) and repressed (Z-score = −20.6, $P < 0.0001$) (Fig. 2B). An example from each of cluster 1 and 2 genes is shown in Figure 2C,D, respectively. Cluster 3 promoters were enriched in membrane protein coding genes (Fig. 2G), without significant changes in acetylation (Fig. 2A), and showed marginal changes in gene expression (Z-score = 2.1, $P = 0.011$) (Fig. 2B). Taken together, our data indicate that e1a induces a general redistribution of H3K18ac from cell type-specific gene promoters, which are repressed, to those involved in cell growth and replication, which are induced.

The RB-family network in contact-inhibited IMR90 cells

Since a major requirement for induction of cell cycling by e1a is to overcome the RB1-mediated repression of cell cycle-regulated genes (Ghosh and Harter 2003; Berk 2005; Ferrari et al. 2008), we determined how the redistribution of H3K18ac induced by e1a relates to the binding of RB proteins before e1a expression. We performed ChIP-seq of all three RB-family members in contact-inhibited IMR90 cells. Our RB1 binding data in the contact-inhibited fibroblasts were very similar to a previous report on RB1 binding in quiescent IMR90 fibroblasts (Chicas et al. 2010; Supplemental Fig. S3). We found 1979, 1983, and 2336 significant peaks for RB1, RBL2, and RBL1, respectively (Fig. 3A). All three members were found commonly at gene promoters, with RB1 having the highest fraction of peaks in IRs (Fig. 3A). The average

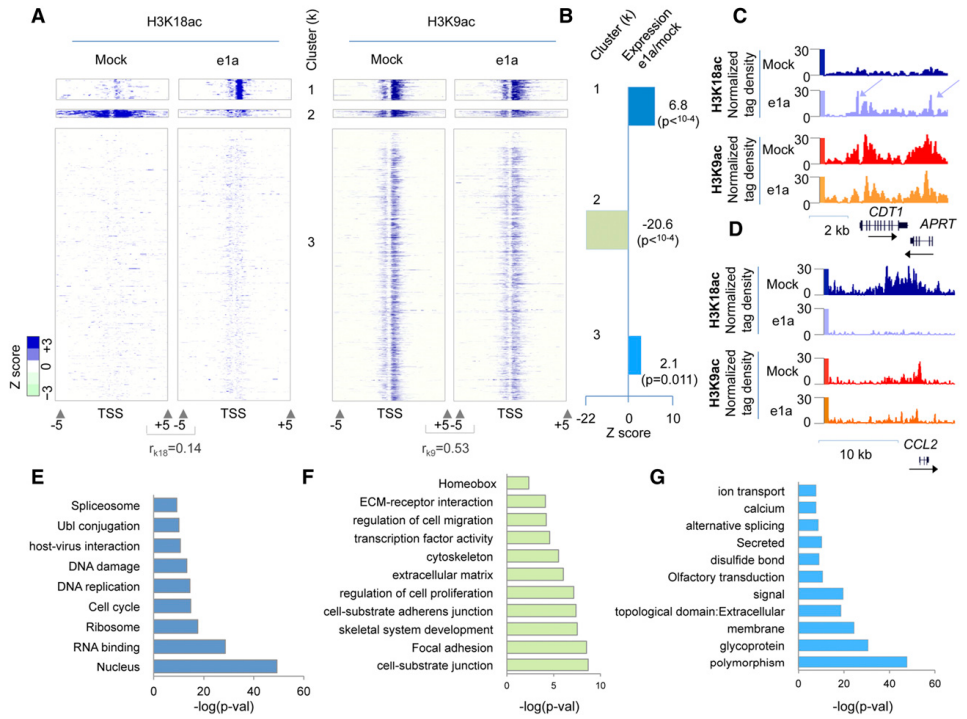


Figure 2. Redistribution of H3K18ac at gene promoters is associated with gene expression changes in the e1a-expressing cells. (*A*) The distributions of H3K18ac and H3K9ac across ±5 kb from TSSs for all annotated transcripts in mock- and *dl*1500-infected cells are shown as heat maps. Each row represents one gene promoter. All promoters were grouped into three clusters based on *k*-means clustering of acetylation data. (*B*) Relative gene expression changes of the three clusters after e1a expression at 24 h p.i. Note the scale. (*C*,*D*) Patterns of H3K18ac and H3K9ac at representative genes from cluster 1 (*CDT1* locus) and cluster 2 (*CCL2* locus). (Light blue arrows) New peaks of H3K18ac in e1a-expressing cells. (Dark arrows) Direction of transcription. (*E*–*G*) GO analyses of genes from promoter clusters 1, 2, and 3 are shown, respectively. Bars represent $-\log_{10}$ of the *P*-value for the selected GO terms.

binding of each protein for all its target gene promoters, spanning ±3 kb of the closest TSSs, is shown in Figure 3B. Interestingly, while the peak of RBL2 is centered at the TSS, the peaks of RBL1 and RB binding are just upstream of and downstream from the TSS, respectively. Motif analysis using Galaxy SeqPos showed a differential enrichment for transcription factors (TFs) binding signatures in RB1 peaks compared to RBL2 and RBL1 (Fig. 3C). The RB1 target promoters were predominantly associated with E2F transcription factor motifs, whereas RBL2 and RBL1 target promoters contained motifs for other transcription factors (Fig. 3C).

To determine whether the RB proteins have an overlapping set of target genes, we generated a tiling profile of their binding patterns across all promoters that had at least one of the RB-family proteins bound (Fig. 3D; Supplemental Fig. S4). We found 3942 genes with at least one RB-family member bound within ±5 kb of the TSS. These were grouped into seven clusters based on the combinatorial patterns of binding. Clusters 1–4 were bound by RB1, with clusters 1–3 bound by at least one of the RB1 paralogs. Clusters 5–7 were enriched for RBL1 and RBL2, RBL2 alone, and RBL1 alone, respectively (Fig. 3D; Supplemental Fig. S4). Notably, each of these clusters showed distinct, albeit overlapping, gene ontology enrichments (Supplemental Fig. S5), and their expression levels were differentially affected by e1a expression (Fig. 3E). Cluster 1, which contains 45 genes with all three RB-family proteins bound to their promoters, is highly enriched for DNA replication genes and was up-regulated by *dl*1500-infection more than any of the other clusters (Fig. 3E). Clusters 2–4, which are all bound by RB1 alone or with one other RB paralog, were also up-regulated after infection. However, clusters 5–7, which are RBL1 and RBL2 target genes, were less significantly induced as a group after infection (Fig. 3E). Altogether, the binding patterns of the RB-family proteins in uninfected cells are consistent with their known functions in control of the cell cycle and DNA replication, but each protein also has a unique set of target genes with differing expression changes in response to e1a expression.

Small e1a establishes new peaks of H3K18ac at and around regions bound by RBl in uninfected cells

Since e1a removes the RB proteins from the promoters of cell cycle-regulated genes (Nevins et al. 1997; Ghosh and Harter 2003; Berk 2005; Ferrari et al. 2008; Sha et al. 2010), we next asked whether

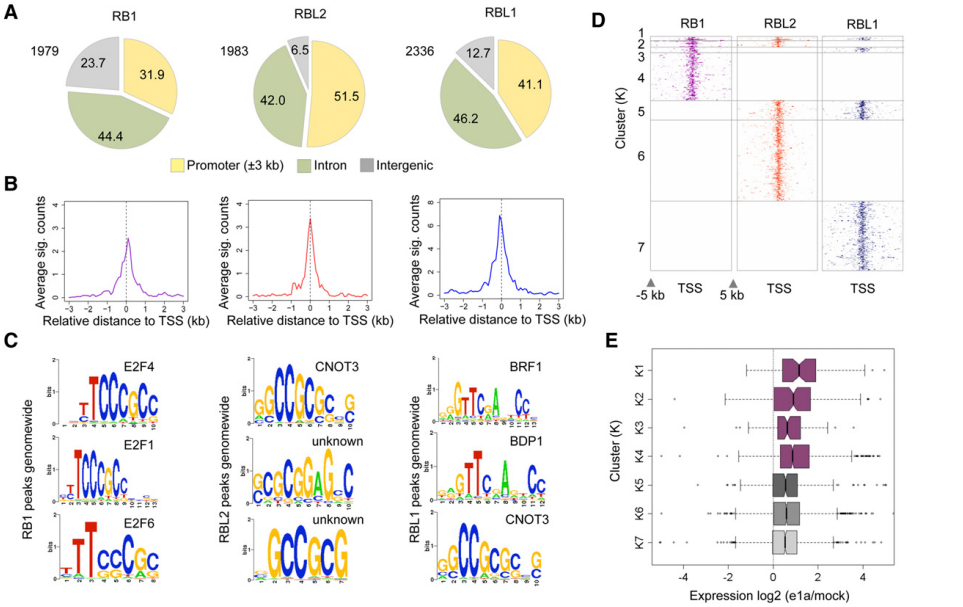


Figure 3. Analyses of RB-family protein genome-wide binding in contact-inhibited IMR90 fibroblasts. (*A*) Overview of RB1, RBL2, and RBL1 peak distributions in mock-infected cells in relation to gene structure are shown as pie charts. (*B*) Average binding profiles (significant counts) of the indicated RB-family members across the TSS regions of their respective target genes. (*C*) A 600-bp region around the peaks of RB1, RBL2, and RBL1 in mock-infected cells were analyzed for TF binding motifs using sitepro (Galaxy). Top three significant motifs for each family member are shown. All *P*-values are less than 1×10^{-10}. (*D*) The distributions of the RB-family proteins across ±5 kb of TSS for genes with at least one RB-family member bound are shown as heat maps. The seven clusters are based on combinatorial binding patterns of the three proteins. (*E*) Relative gene expression changes of each of the seven clusters after e1a expression at 24 h p.i. are shown as box plots.

changes in H3K18ac or H3K9ac at promoter regions following e1a expression are related to the binding of the RB-family proteins prior to infection. We determined the significantly increased peaks of H3K18ac or H3K9ac in *dl*1500- vs. mock-infected cells and calculated the average change in the levels of acetylation in 100-bp intervals relative to the TSS for each of the clusters in Figure 3D. We found increased levels of H3K18ac in the promoter regions of genes in clusters 1–4, which were all bound by RB1 (Fig. 4A). Clusters 5–7 bound by RBL2 and RBL1 but without RB1 showed little to no change in the levels of H3K18ac relative to the average H3K18ac levels across the genome (i.e., the genome background) (Fig. 4B). None of the clusters showed any significant change in H3K9ac upon infection (Fig. 4C,D).

To explore the RB1 and H3K18ac relationship further, we selected and centered all the peaks of RB1, RBL2, and RBL1 throughout the genome in mock-infected cells. We then plotted the average levels of H3K18ac and H3K9ac in mock- and *dl*1500-infected cells within ±5 kb of each RB-family member peak center (Fig. 5A–F). We also calculated the fraction of H3K18 and H3K9 acetylation peaks in uninfected and e1a-expressing cells in the same interval (Fig. 5G–L). The levels of H3K18ac in mock-infected fibroblasts at RB1-bound target regions were minimal (Fig. 5A, purple line) but increased significantly in the infected cells (Fig. 5A, red line). The levels of H3K18ac also increased significantly at RBL2-bound target regions (Fig. 5B) and, to a minor extent, at RBL1-bound regions (Fig. 5C). The levels of H3K18ac in e1a-expressing cells also increased at RBL2-bound regions that were 5 kb away from RB1 peaks (Supplemental Fig. S6), indicating that new peaks of H3K18 were also established at RBL2-bound regions independent of RB1. Consistently, the number of new peaks of H3K18ac increased substantially at and around RB1-bound regions (Fig. 5G) as well as in regions bound by RBL2 (Fig. 5H) and less by RBL1 (Fig. 5I). Strikingly, 8% of H3K18ac peaks in the e1a-expressing cells occurred precisely where RB1 was bound, compared to 0.7% in mock-infected cells (Supplemental Fig. S7A). A similar but much smaller trend was observed for H3K18ac peaks that occurred precisely at RBL2- and RBL1-bound loci (Supplemental Fig. S7A). Interestingly, a double-peak of H3K18ac was observed only at RBL1-bound regions (cf. Fig. 5C and Fig. 5A,B). Such double-peaks of acetylation are commonly observed for acetylation marks relative to their HATs (Rada-Iglesias et al. 2011).

When we examined the average H3K9ac levels at the RB protein peaks, we found a single peak of H3K9ac at the RB1-bound regions, with slightly lower levels of H3K9ac in the e1a-expressing cells (Fig. 5D). Remarkably, the H3K9ac profile showed a double-peak of acetylation around both the RBL2- and RBL1-bound regions in mock- and *dl*1500-infected cells, with lower levels in the latter (Fig. 5E,F). Consistently, the fraction of H3K9ac peaks colocalizing with any of the RB-family members did not change significantly between mock and e1a-expressing cells

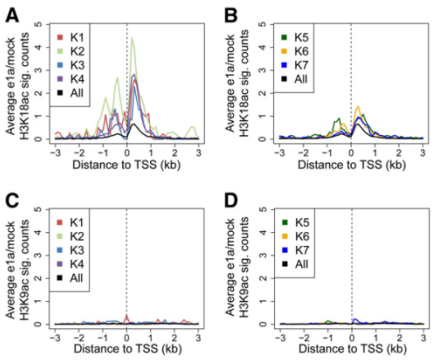


Figure 4. Relationship of the RB-binding clusters to histone acetylation change. (*A*,*B*) Average significant (sig.) counts of change in H3K18ac levels in *dl*1500- versus mock-infected cells are shown for the seven RB-clusters (see Fig. 3C). (*C*,*D*) Average significant (sig.) counts of change in H3K9ac levels in *dl*1500- versus mock-infected cells are shown for the seven RB-clusters (see Fig. 3C).

(Fig. 5J–L; Supplemental Fig. S7B). Taken together, these data suggest that a large fraction of the new H3K18ac peaks in the infected cells occur at or around regions that were bound by RB1 before e1a expression and, to a lesser extent, by RBL2 and RBL1. In addition, the different relationship between H3K18 and H3K9 acetylation and the RB-family members' binding sites suggests that H3K18 and H3K9 acetylation have different functions in gene regulation.

To gain further insight into the possible mechanism by which H3K18ac levels could be increased at clusters 1–4 (Fig. 4A), we analyzed published data sets of EP300 and CREBBP binding in the T98G glioblastoma line (Ramos et al. 2010), as well as EP300 and E2F1 binding in H9 human embryonic (hESCs) and MCF7 cells, respectively (Cao et al. 2011; Rada-Iglesias et al. 2011). In G1-arrested T98G cells, none of the seven RB-clusters showed significant enrichment for either EP300 or CREBBP (Supplemental Fig. S8A,C). However, upon serum stimulation, when cells are induced to enter S-phase as in e1a-expressing cells, clusters 1–4 showed greatly increased levels of EP300 binding compared to clusters 5–7 (Supplemental Fig. S8A–D). Upon serum stimulation, CREBBP showed increased binding to all clusters (Supplemental Fig. S8D). Also, in H9 hESCs, levels of EP300 binding were greater in clusters 1–4 compared to clusters 5–7 (Supplemental Fig. S8E). We also found that E2F1 binding in MCF7 cells is preferentially enriched in RB1-bound clusters 1–4 (Supplemental Fig. S8F). Finally, examination of e1a binding at a limited number of promoters showed that e1a may preferentially bind to RB1-bound gene promoters (*POLD3* and *CDC7*) compared to those without RB1 (*FOS* and *GAPDH*) (Supplemental Fig. S9). Altogether, these data suggest that, compared to promoters bound by RBL2 or RBL1 without RB1, promoters bound by RB1 might have differential recruitment of EP300 and E2F1 as well as e1a itself. Such a scenario may also account for the preferential acetylation of H3K18 in clusters 1–4 versus 5–7 in the e1a-expressing cells.

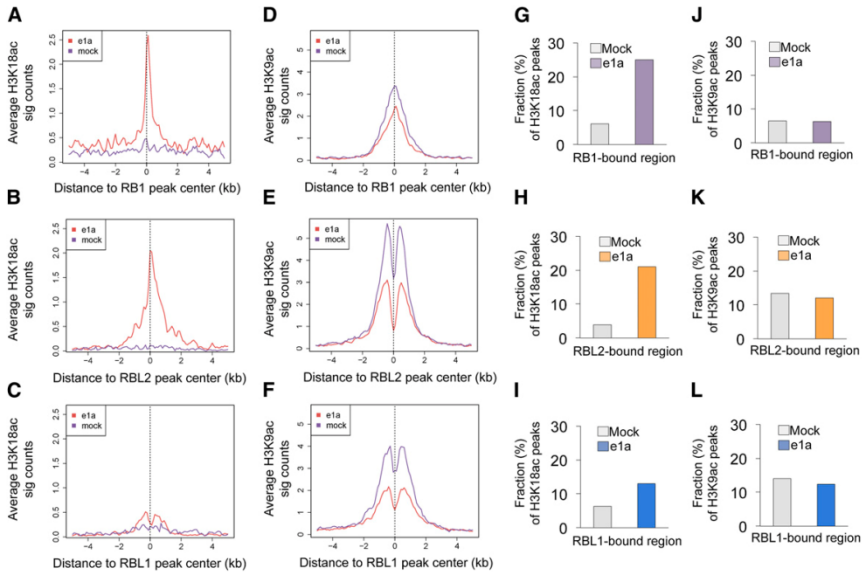


Figure 5. Genome-wide peaks of H3K18ac in the infected cells are established at regions that are bound by RB1, and to some extent RBL2, in uninfected cells. Average levels (sig. counts) of H3K18ac and H3K9ac in mock- and *dl*1500-infected (e1a) cells, respectively, are shown relative to (*A*,*D*) RB1-centered peaks, (*B*,*E*) RBL2-centered peaks, and (*C*,*F*) RBL1-centered peaks. (*G*–*I*) The fractions of H3K18ac peaks in mock- and *dl*1500-infected (e1a) cells that are within 5 kb of RB1, RBL2, and RBL1 peaks in mock-infected cells are shown as bar charts. (*J*–*L*) The same analysis as in *G*–*I* for H3K9ac.

We also found that 63% of intergenic RB1 peaks in uninfected cells were within 5 kb of H3K18ac peaks in e1a-expressing cells (Supplemental Fig. S10A). These RB1 peaks were enriched for TF binding motifs such as E4F1 and ATF2 (Supplemental Fig. S10B). Interestingly, E4F1 interacts with RB1, and E4F1 activity is regulated through the action of adenovirus e1a (Fajas et al. 2000). Surprisingly, the remaining IR RB1 peaks (390) that did not gain H3K18ac were enriched for E2F binding signatures (Supplemental Fig. S10C). These data suggest that e1a establishes new peaks of H3K18ac at a distinct subset of RB1-bound intergenic regions.

The new peaks of H3K18ac that are not within 5 kb of RB1 peaks were associated with 2533 genes and distributed between the TSS and mostly regions away from the TSS in e1a-expressing cells (Supplemental Fig. S11A). The majority of these genes were induced (Supplemental Fig. S11B) with significant enrichment for cell cycle genes, chromatin and RNA binding genes as well as TF motifs such as NFKB1, PLAGL1, TOPORS, and E2F4 (Supplemental Fig. S11C). The repressed genes were associated with fibroblast-related functions and TF motifs such as ZIC2, ZIC1, PLAG1, and AFFA4 (Supplemental Fig. S11D). These data suggest that e1a might exploit a distinct set of transcription factors or cofactors in addition to RB1 for acetylation of H3K18 and subsequent gene expression changes that favor the viral life cycle.

Specific intergenic regions are also targets of acetylation in infected cells

To characterize the H3K18ac redistribution at intergenic regions, we calculated the H3K18ac tag density per kb to determine a single enrichment value for H3K18ac at these locations (Fig. 6A). We found a small set of 406 IRs that gained new peaks of H3K18ac upon infection, compared to 1959 IRs that had drastically decreased levels of H3K18ac (Fig. 6A). The intergenic regions downstream from the ribosomal protein S6 kinase (*RPS6KA1*) and upstream of the Interleukin 6 (*IL6*) loci are examples of IRs where H3K18ac increased and decreased, respectively (Fig. 6B,C). Since H3K18ac at promoters is associated with active transcription, modulation of H3K18ac at IRs might also affect expression of the neighboring genes. To test this, we first performed gene ontology analysis of all the genes neighboring significant peaks of H3K18ac in IRs associated with mock- and *dl*1500-infected cells using the Genomic Region Annotation Tool (GREAT). We found that genes associated with H3K18ac in IRs of infected cells were enriched for genes activated by fibroblast growth factors and encoding transcriptional regulators (Supplemental Fig. S12A). In contrast, genes associated with IRs that decreased H3K18ac in infected cells are involved in differentiated fibroblast functions and immune response (Supplemental Fig. S12B). Genes associated with H3K18ac peaks in IRs of e1a-expressing cells were generally activated upon

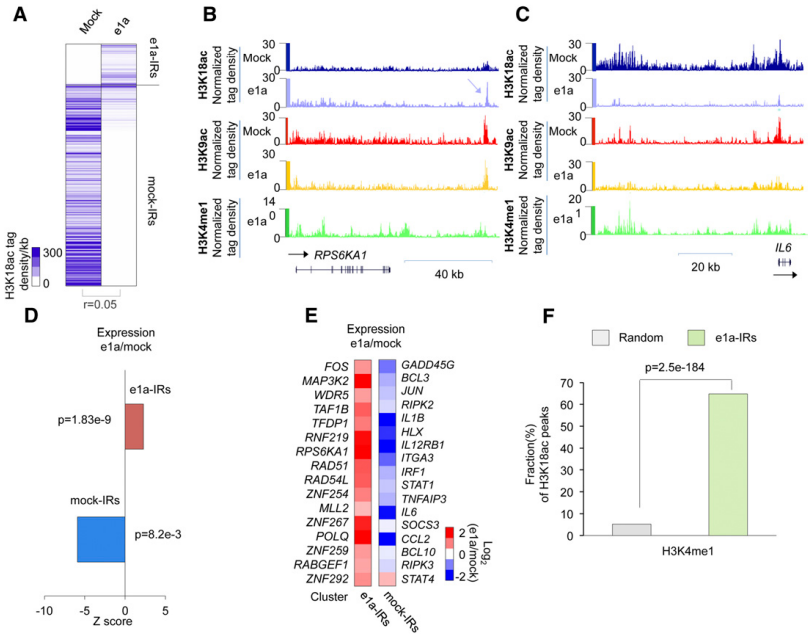


Figure 6. Intergenic peaks of H3K18ac coincide with potential distal regulatory elements to control gene expression. (*A*) The total H3K18ac tag density (number of sequence tags per kb) in intergenic regions (IRs) that are affected upon e1a expression is shown as a heat map. "Mock-IRs" and "e1a-IRs" are intergenic regions that harbor significant H3K18 acetylation in mock- or *dl*1500-infected cells, respectively. Distributions of H3K18ac, H3K9ac, and H3K4me1 in mock- and *dl*1500-infected cells around the *RPS6KA1* (*B*) and the *IL6* loci (*C*) are shown. (*D*) Relative gene expression changes of the e1a- and mock-IRs clusters after e1a expression at 24 h p.i. (*E*) Expression levels of candidate genes associated with e1a- and mock-IRs clusters are shown as heat maps. (*F*) Bar plot showing the fraction of H3K18ac peaks in the e1a-IRs that coincide with H3K4me1 in e1a-expressing cells.

infection, while genes linked to IRs with decreased H3K18ac were generally repressed (Fig. 6D). Examples of genes associated with H3K18 acetylated IRs in the e1a-expressing cells include *FOS* and *RPS6KA1*, which are involved in regulation of cell proliferation, *TAF1B*, a basal transcription factor, and TFDP1, which heterodimerizes with E2F proteins to enhance their DNA-binding activity and promote transcription of genes involved in cell cycle progression (Polager and Ginsberg 2008; Fig. 6E). The genes associated with IRs that were deacetylated in the infected cells include *IL1B*, *IRF1*, *IL6*, and *CCL2* which are all important mediators of the immune response (Akira and Kishimoto 1992; Feigenblum et al. 1998; Quinones et al. 2007). These results suggest that acetylation and deacetylation of H3K18 at IRs is also associated with induction and repression of transcription at a distance.

To provide evidence that the H3K18 acetylated IRs in *dl*1500-infected cells may act as distal regulatory elements (Bulger and Groudine 2011), we performed ChIP-seq for H3K4me1, a histone methylation mark generally associated with enhancer elements (Heintzman et al. 2009; Hawkins et al. 2010; Rada-Iglesias et al. 2011) in e1a-expressing cells, and calculated the fraction of H3K18ac peaks in IRs that overlap with H3K4me1. We found that 65% of e1a-induced H3K18ac IR peaks occur preferentially at genomic locations that are also enriched for H3K4me1 (Fig. 6B,C,F). Analysis of potential TF binding motifs within 600 bp of peaks of H3K18ac in IRs of e1a-expressing cells yielded significant enrichment for TFs involved in cytokine-independent proliferation (e.g., SP1, PIR, GTF3C2, ERG1, and E2F1) (Zellmer et al. 2010; Supplemental Fig. S13A). The H3K18ac peaks in IRs of mock-infected cells were also highly enriched for TF binding signatures such as ZEB1, MZF1, REST, and NFKB1 (Supplemental Fig. S13B), all of which are involved in immune response and differentiation (Baeuerle and Henkel 1994; Hutton et al. 2004; Wang et al. 2009; Roizman 2011). These data suggest that the new H3K18ac peaks in IRs coincide with potential distal regulatory elements that promote cell growth and division, and the loss of H3K18ac from IRs may modulate the cell's response to viral infection and its phenotypic identity.

Histone acetylation marks specific regions of the Ad5 genome

Several reports have indicated that the adenovirus genome becomes associated with nucleosomes following infection (Sergeant et al. 1979; Karen and Hearing 2011; Komatsu et al. 2011; Ross et al. 2011). To determine if acetylated nucleosomes are localized to specific regions of the viral genome, we aligned the reads from input and acetylation ChIPs to the Ad5 genome (see Supplemental Methods) and searched for regions of enrichment using the same algorithm described above. We found several significant peaks of H3K9 and H3K18 acetylation across the viral genome (Supplemental Fig. S14). Our data are consistent with packaging of the Ad5 genome into chromatin, although the functional significance of histone acetylation relative to the hallmarks of the tightly packed viral genome remains to be determined.

Discussion

The interaction of adenovirus small e1a with the EP300/CREBBP KATs decreases the cellular level of H3K18ac in primary human cells to ~30% of the level observed in uninfected cells (Horwitz et al. 2008). We now provide a molecular explanation for the global decrease in H3K18ac levels. Our data reveal that 95% of the H3K18ac peaks in uninfected cells are erased in the e1a-expressing cells, and one-third as many new peaks as before infection are established.

Loss of H3K18ac in the e1a-expressing cells was observed in both promoter and intergenic regions and was not a consequence of S-phase of the cell cycle per se. The affected promoter regions regulate genes with functions relevant to fibroblast biology and were repressed upon infection. Deacetylation of H3K18 in intergenic regions may also lead to deactivation of potential distant regulatory elements and to repression of associated genes with immune regulatory and fibroblast-related functions. This suggests that the e1a-induced transformation process involves suppressing the phenotypic identity of the infected cell and the cell's ability to mount a defense response by switching off the pertinent gene expression networks.

Small e1a establishes new peaks of H3K18ac at promoters of cell cycle-regulated genes that are highly induced and would eventually favor viral replication by driving the otherwise quiescent cell to proliferate. The new intergenic peaks of H3K18ac in *dl*1500-infected cells may also play a beneficial role for the virus by acting as enhancer elements for genes involved in cell proliferation. More intriguing is the fact that 55% of RB1-bound regions in the uninfected host cell gain significant levels of H3K18ac after infection. In fact, a subset of new H3K18ac occurs precisely where RB1 was bound prior to e1a expression. Although the genome-wide binding pattern of RB1 in the *dl*1500-infected cell is not known, numerous experimental results indicate that RB1 would be physically removed by e1a (Ghosh and Harter 2003; Berk 2005; Liu and Marmorstein 2007; Ferrari et al. 2008; DeCaprio 2009). So, the removal of RB1 by e1a is likely followed by acetylation of the underlying or surrounding nucleosomes on H3K18. This would be consistent with a model of e1a-targeted gene induction that not only requires removal of a transcriptional repressor (i.e., RB1) but also addition of an activating epigenetic mark (i.e., H3K18ac), probably through recruitment of EP300 and/or CREBBP (Ferrari et al. 2008; Horwitz et al. 2008). How H3K18ac effects are implemented by the cell is unknown, but e1a has highlighted a distinct function for H3K18ac among many histone acetylation sites including H3K9ac in regulating gene expression. Jin et al. (2011) also have recently reported that H3K18ac and H3K27ac, and not H3K9ac, are generated by EP300 and CREBBP and are required for activation of target genes by nuclear receptors.

The mechanism of genome-wide H3K18 deacetylation is not fully understood but requires interaction of e1a with EP300/CREBBP KATs (Horwitz et al. 2008), which are also relocalized upon infection (Ferrari et al. 2008). Whether one or more lysine deacetylases (KDACs), including certain sirtuins (SIRTs), which can deacetylate H3K18 in vitro (Black et al. 2008), are also involved remains to be determined. Interestingly, SIRT1 inhibition promotes differentiation (Pickard et al. 2010; Zhang et al. 2011), and its deletion induces inflammatory signaling in response to environmental stress (Schug et al. 2010; Zhang et al. 2010). This raises the possibility that H3K18 deacetylation could occur by sequestration of EP300/CREBBP and recruitment of these KATs to new but limited locations as well as active deacetylation by one or more KDACs to inhibit fibroblasts functions and antiviral responses and induce cell proliferation in the e1a-expressing cells.

Finally, our data have, for the first time, revealed the location of acetylated histones in the adenoviral genome. Together with previous reports that the adenoviral genome becomes chromatinized in the host cell (Sergeant et al. 1979; Karen and Hearing 2011; Komatsu et al. 2011; Ross et al. 2011), our data suggest that histone acetylation may also function to regulate the expression of adenovirus genes.

Methods

Cell culture and viruses

IMR90 human primary lung embryo fibroblasts (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 100 U/mL penicillin, 100 μg/mL streptomycin, and 10% fetal bovine serum (FBS) at 37°C in 5% CO_2. Propagation of the *dl*1500 virus was done as described in Horwitz et al. (2008). Briefly, 293 spinner cells were grown in Joklik's modified Eagle minimal essential medium (SMEM) plus 5% FBS with 100 U/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine. Propagation of viruses was in 293 spinner cells with 2% FBS. Ad *dl*1500 stock titers were determined by plaque assay in duplicate on 293 cells.

Western blot

Isolation of core histones was performed as previously described (Pilch et al. 2004). The dried pellets were dissolved in 6× Laemmli sample loading buffer for subsequent analysis by Western blot. Western blots were performed using LI-COR.

Chromatin immunoprecipitation and library preparation

Chromatin immunoprecipitation for ChIP-seq was performed essentially as described (Ferrari et al. 2008) with few modifications. Briefly, 1×10^8 IMR90 lung fibroblasts were grown to 80% confluency (asynchronous population) or to confluence in 15-cm dishes. After 24 h, the cells were incubated with mock- or the *dl*1500- adenovirus for 1 h in media with 2% serum. After 24 h post-infection (p.i.), formaldehyde was added for 10 min at 37°C. After PBS washing, cross-linked cells were scraped from the plates and washed with 1 mL of PBS containing protease inhibitors (Roche). 2×10^7 cells were resuspended in 450 μl of lysis buffer and incubated for 10 min on ice and immediately sonicated using Misonix cup-horn sonicator. 100 μl of the lysate (corresponding to 5×10^6 cells) were used for each immunoprecipitation with a given antibody (listed below); 10 μl of the lysate were used as input. After overnight reversal of cross-linking at 65°C, samples were treated with RNase A for 30 min at 37°C and proteinase K for 2 h at 56°C. DNA was subsequently purified using phenol/chloroform extraction and precipitation. DNA concentration was measured using Qubit (Invitrogen). At least 10 ng of dsDNA for both input and IP were used for library preparation according to the manufacturer's instructions (Illumina). Libraries were sequenced using Illumina Genome Analyser II to obtain 76-bp-long reads (H3K9ac and corresponding input libraries) or Illumina HIseq-2000 to obtain 50-bp-long reads (H3K18ac and correspondent input libraries). ChIP for RB1 was performed as described (Chicas et al. 2010). Each library was checked for mismatch distribution.

A list of the antibodies used in this study is reported in the Supplemental Methods.

RNA extraction and mRNA-seq library preparation

Total RNA was extracted from mock- and *dl*1500-infected cells using the Qiagen easy RNA kit. The maximum amount of RNA was used to start the library preparation according to the manufacturer's instructions (Illumina). Libraries were sequenced using Illumina Genome Analyser II to obtain 76-bp-long reads. Alignment of mRNA-seq reads was performed using default parameters of TopHat (Trapnell et al. 2009; Kleindienst et al. 2010). Aligned reads were converted to sam format, and SAMMate software (Xu et al. 2011) was used to determine the transcript RPKM (reads per kilobase of exon per million of reads).

FACS analysis

FACS analysis has been carried out using the two-step EdU-PI staining kit (Life Technology) following the manufacturer's instructions.

Generation of H3K18ac antibody

A detailed description is given in the Supplemental Methods.

ChIP-qPCR

The same amount of chromatin was used for all ChIPs. ChIPed DNA was subjected to quantitative PCR using the respective input and IgG ChIP as control. qPCRs were performed using FastStart Universal SYBR Green Master (Rox). The ChIP final product (5 μl), 0.2 μM primers concentration, and 1× SYBR green mix were used for each reaction. qPCR analysis was performed using MxPro (Stratagene) with triplicate technical and biological experiments. ChIP-qPCR signals were calculated as a percentage of input. Standard deviations were measured and represented as error bars. Primers used for qPCR are listed in the Supplemental Methods.

ChIP-seq analysis

Reads were mapped to the Human (hg19) genome using Bowtie software. Only reads that aligned to a unique position in the genome with no more than two sequence mismatches were retained for further analysis. Duplicate reads that mapped to the same exact location in the genome were counted only once to reduce clonal amplification effects. The genome was tiled into 100-bp windows. Each read was extended by 150 bases (we refer to tags as the extend read counts within a bin) and was counted as one read to each window to which it partially or fully matched. The total counts of the input and ChIP samples were normalized to each other. The input sample was used to estimate the expected counts in a window; the average value for all windows was assigned to windows with zero counts. Finally, we used the Poisson distribution to estimate the probability of observing the ChIP counts within a window given the expected counts in the input sample window. We considered all windows with *P*-values less than 1.0×10^{-3} to have significant peaks. A *P*-value $< 1.0 \times 10^{-3}$ was chosen to give a False Discovery Rate (FDR) of <5%. The FDR was calculated by applying the same statistic described above to the two halves of the same input library. We considered the total number of significant peaks obtained this way as an estimate of the number of false-positive peaks. To compare two ChIP samples to identify regions that are significantly different between the two samples (for instance, e1a vs. mock), we replaced the input sample with the sample used for comparison; all downstream analyses were identical. Our algorithm produced several files that were subsequently used for analysis: BED files containing the coordinates of the significant windows of enrichment; Wiggle (wig) files (chromosome tiling, fixed step) with normalized read counts for the significant windows (if a window was not significant, we placed zero tags); GR files of normalized raw counts for input, and ChIP samples were also created for genome browser visualization (e.g., Fig, 1C,D). Tiling profiles of promoter regions for the hg19 annotated human promoters were also generated (Figs. 2A, 3D). These represent the 100-bp tiling of a 10-Kb region spanning the transcription start site. We reported the number of reads falling into significant windows and zero for the nonsignificant ones. *Z*-score scaling was applied to acetylation tiling profiles in Figure 2A in order to compare the levels of the different histone marks.

Detailed statistical methods are described in the Supplemental Methods.

Data access

All data in this work are available for download at the NCBI Gene Expression Omnibus (GEO) under accession number GSE32340.

Acknowledgments

G.B. was supported by a USPHS National Research Service Award, GM07104. Work in the A.J.B. lab is supported by R37CA025235. This work was supported by grants from the American Cancer Society and Howard Hughes Medical Institute and by an NIH Director's Innovator award to S.K.K.

Author contributions: R.F., A.B., and S.K.K. conceived and designed the experiments. R.F. performed experiments and data analysis. T.S. and Y.C. performed RNA-seq; T.S., B.L., G.B., A.O., R.S., and M.P. performed script programming and helped with data analysis and interpretation. A.B. provided *dl*1500 virus. R.F., A.B., and S.K.K. wrote the manuscript.

References

Akira S, Kishimoto T. 1992. IL-6 and NF-IL6 in acute-phase response and viral infection. *Immunol Rev* **127:** 25–50.
Baeuerle PA, Henkel T. 1994. Function and activation of NF-κB in the immune system. *Annu Rev Immunol* **12:** 141–179.
Berk AJ. 2005. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **24:** 7673–7685.
Black JC, Mosley A, Kitada T, Washburn M, Carey M. 2008. The SIRT2 deacetylase regulates autoacetylation of p300. *Mol Cell* **32:** 449–455.
Bulger M, Groudine M. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144:** 327–339.
Cao AR, Rabinovich R, Xu M, Xu X, Jin VX, Farnham PJ. 2011. Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome. *J Biol Chem* **286:** 11985–11996.
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:** 376–387.
DeCaprio JA. 2009. How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. *Virology* **384:** 274–284.
Fajas L, Paul C, Zugasti O, Le Cam L, Polanowska J, Fabbrizio E, Medema R, Vignais ML, Sardet C. 2000. pRB binds to and modulates the transrepressing activity of the E1A-regulated transcription factor p120E4F. *Proc Natl Acad Sci* **97:** 7738–7743.
Feigenblum D, Walker R, Schneider RJ. 1998. Adenovirus induction of an interferon-regulatory factor during entry into the late phase of infection. *J Virol* **72:** 9257–9266.
Ferrari R, Pellegrini M, Horwitz GA, Xie W, Berk AJ, Kurdistani SK. 2008. Epigenetic reprogramming by adenovirus e1a. *Science* **321:** 1086–1088.
Ghosh MK, Harter ML. 2003. A viral mechanism for remodeling chromatin structure in G0 cells. *Mol Cell* **12:** 255–260.
Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6:** 479–491.
Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108–112.
Horwitz GA, Zhang K, McBrian MA, Grunstein M, Kurdistani SK, Berk AJ. 2008. Adenovirus small e1a alters global patterns of histone modification. *Science* **321:** 1084–1085.
Hutton JJ, Jegga AG, Kong S, Gupta A, Ebert C, Williams S, Katz JD, Aronow BJ. 2004. Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system. *BMC Genomics* **5:** 82. doi: 10.1186/1471-2164-5-82.
Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SY, Ge K. 2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30:** 249–262.
Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316:** 1497–1502.
Karen KA, Hearing P. 2011. Adenovirus core protein VII protects the viral genome from a DNA damage response at early times after infection. *J Virol* **85:** 4135–4142.
Kleindienst R, Moeller L, Sinzinger S. 2010. Highly efficient refractive Gaussian-to-tophat beam shaper for compact terahertz imager. *Appl Opt* **49:** 1757–1763.
Komatsu T, Haruki H, Nagata K. 2011. Cellular and viral chromatin proteins are positive factors in the regulation of adenovirus gene expression. *Nucleic Acids Res* **39:** 889–901.
Liu X, Marmorstein R. 2007. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev* **21:** 2711–2716.
Montell C, Fisher EF, Caruthers MH, Berk AJ. 1982. Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site. *Nature* **295:** 380–384.
Nevins JR, Leone G, DeGregori J, Jakoi L. 1997. Role of the Rb/E2F pathway in cell growth control. *J Cell Physiol* **173:** 233–236.
Pickard A, Wong PP, McCance DJ. 2010. Acetylation of Rb by PCAF is required for nuclear localization and keratinocyte differentiation. *J Cell Sci* **123:** 3718–3726.
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.
Polager S, Ginsberg D. 2008. E2F—at the crossroads of life and death. *Trends Cell Biol* **18:** 528–535.
Quinones MP, Estrada CA, Jimenez F, Martinez H, Willmon O, Kuziel WA, Ahuja SK, Ahuja SS. 2007. CCL2-independent role of CCR2 in immune responses against *Leishmania major*. *Parasite Immunol* **29:** 211–217.
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:** 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 p300 and CBP in transcription regulation. *Nucleic Acids Res* **38:** 5396–5408.
Roizman B. 2011. The checkpoints of viral gene expression in productive and latent infection: The role of the HDAC/CoREST/LSD1/REST repressor complex. *J Virol* **85:** 7474–7482.
Ross PJ, Kennedy MA, Christou C, Risco Quiroz M, Poulin KL, Parks RJ. 2011. Assembly of helper-dependent adenovirus DNA into chromatin promotes efficient gene expression. *J Virol* **85:** 3950–3958.
Schug TT, Xu Q, Gao H, Peres-da-Silva A, Draper DW, Fessler MB, Purushotham A, Li X. 2010. Myeloid deletion of SIRT1 induces inflammatory signaling in response to environmental stress. *Mol Cell Biol* **30:** 4712–4721.
Sergeant A, Tigges MA, Raskas HJ. 1979. Nucleosome-like structural subunits of intranuclear parental adenovirus type 2 DNA. *J Virol* **29:** 888–898.
Sha J, Ghosh MK, Zhang K, Harter ML. 2010. E1A interacts with two opposing transcriptional pathways to induce quiescent cells into S phase. *J Virol* **84:** 4050–4059.
Sherr CJ, McCormick F. 2002. The RB and p53 pathways in cancer. *Cancer Cell* **2:** 103–112.
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:** 473–479.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.
Wang J, Lee S, Teh CE, Bunting K, Ma L, Shannon MF. 2009. The transcription repressor, ZEB1, cooperates with CtBP2 and HDAC1 to suppress IL-2 gene activation in T cells. *Int Immunol* **21:** 227–235.
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:** 2. doi: 10.1186/1751-0473-6-2.
Zellmer S, Schmidt-Heck W, Godoy P, Weng H, Meyer C, Lehmann T, Sparna T, Schormann W, Hammad S, Kreutz C, et al. 2010. Transcription factors ETF, E2F, and SP-1 are involved in cytokine-independent proliferation of murine hepatocytes. *Hepatology* **52:** 2127–2136.
Zhang Z, Lowry SF, Guarente L, Haimovich B. 2010. Roles of SIRT1 in the acute and restorative phases following induction of inflammation. *J Biol Chem* **285:** 41391–41401.
Zhang Y, Wang J, Chen G, Fan D, Deng M. 2011. Inhibition of Sirt1 promotes neural progenitors toward motoneuron differentiation from human embryonic stem cells. *Biochem Biophys Res Commun* **404:** 610–614.

Received September 21, 2011; accepted in revised form April 5, 2012.

Supplement Information

Reorganization of the host epigenome by a viral oncogene

Roberto Ferrari[1,8], Trent Su[1,2], Bing Li[1], Giancarlo Bonora[1,3], Amit Oberai[1], Yvonne Chan[1], Rajkumar Sasidharan[7] Arnold J. Berk[5,6], Matteo Pellegrini[7,8] and Siavash K. Kurdistani[1,4,6,8]

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

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

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

[8]Eli 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'-AGATGCCCCTTTCTTACCAATC-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'-ATCTTTGTTCCCGGAGCTGT-3'
POLD3-F 5'-GAGCCTTGACCCTCTGTCTG-3'
POLD3-R 5'-ACTTCCGTGTGTGCTGGAG-3'
COL6A3-F 5'-CACTTCTGAGCAGCCAACTG-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'- CGTCTCCCTGACTAGTTTGACC-3'
*POLD3*_EP300/CREBBP-R 5'-CCTCTAACCCCTCTCTCAATCA-3'
*COL6A3*_EP300/CREBBP-F 5'-CATTTTACAGGACTGCCTGTGT-3'
*COL6A3*_EP300/CREBBP-R 5- GGCTGTAGAATTATGCCAGATG-3'
GAPDH-F 5'- GCCAATCTCAGTCCCTTCCC-3'
GAPDH-R 5'-CGCCCGTAAAACCGCTAGTAG-3'
CDC7-F 5'-GTGAGTTTCCGACGGTTTGT-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 publication[8]. 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.

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

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, McBrian 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. *dl*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

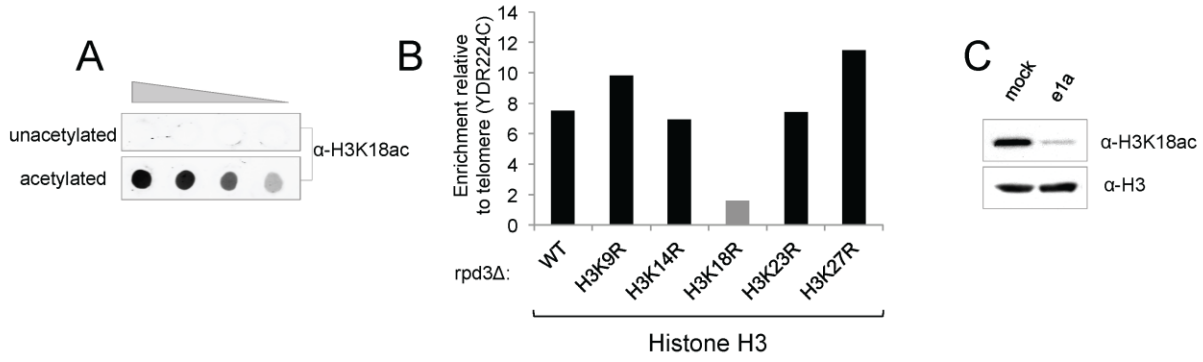


Fig. S1. Characterization of H3K18ac (custom 814) antibody. A) Peptide blot for unacetylated and acetylated H3K18 peptide. Decreasing amounts of peptide were blotted (grey triangle). B) Yeast strains with the indicated histone mutations were used to determine the specificity of H3K18ac antibody in ChIP 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 rpd3Δ-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.

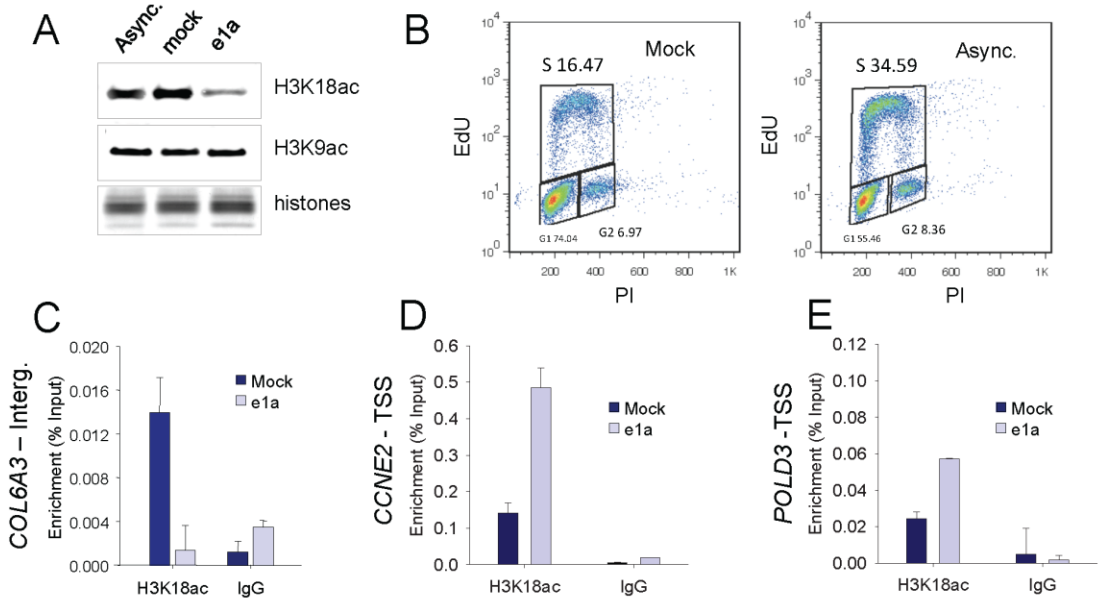


Fig. S2. Global level of H3K18ac and ChIP-quantitative PCR (ChIP-qPCR) validate ChIP-seq enrichments. A) Western blot for H3K18ac (814) and H3K9ac in asynchronous IMR90, mock- and *dl*1500-infected cells showing that a global decrease of H3K18ac is only observed upon e1a expression. Global levels of H3K9ac are comparable in all conditions. Total histones are used as loading control. **B)** FACS analysis (5-ethynyl-2'-deoxyuridin (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 *dl*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.

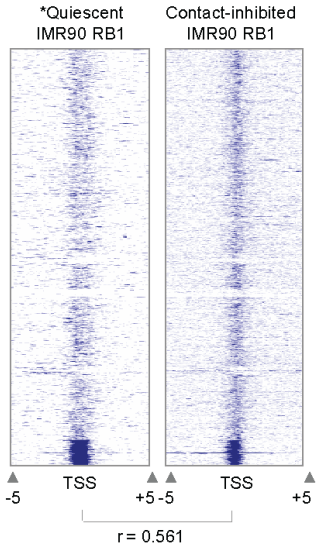


Fig. S3. RB1 binding in contact-inhibited (mock-infected) and quiescent* IMR90 fibroblast cells. Heat maps showing the binding of RB1 in the two conditions over a 10-kb region spanning the TSS of all annotated transcripts. The majority of the significant peaks (shades of dark blue) grouped in the same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

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

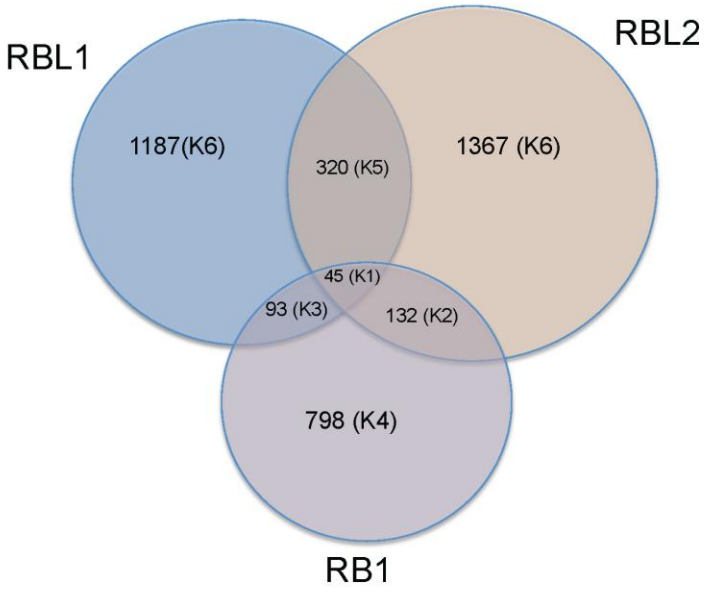


Fig. S4. RB family members share overlapping but distinct set of target promoter regions. Promoters having at least one significant peak of binding for any of the RB-family members RB1, RBL1 and RBL2 were selected and a Venn diagram representing the overlap among the bound promoters is shown. The numbers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

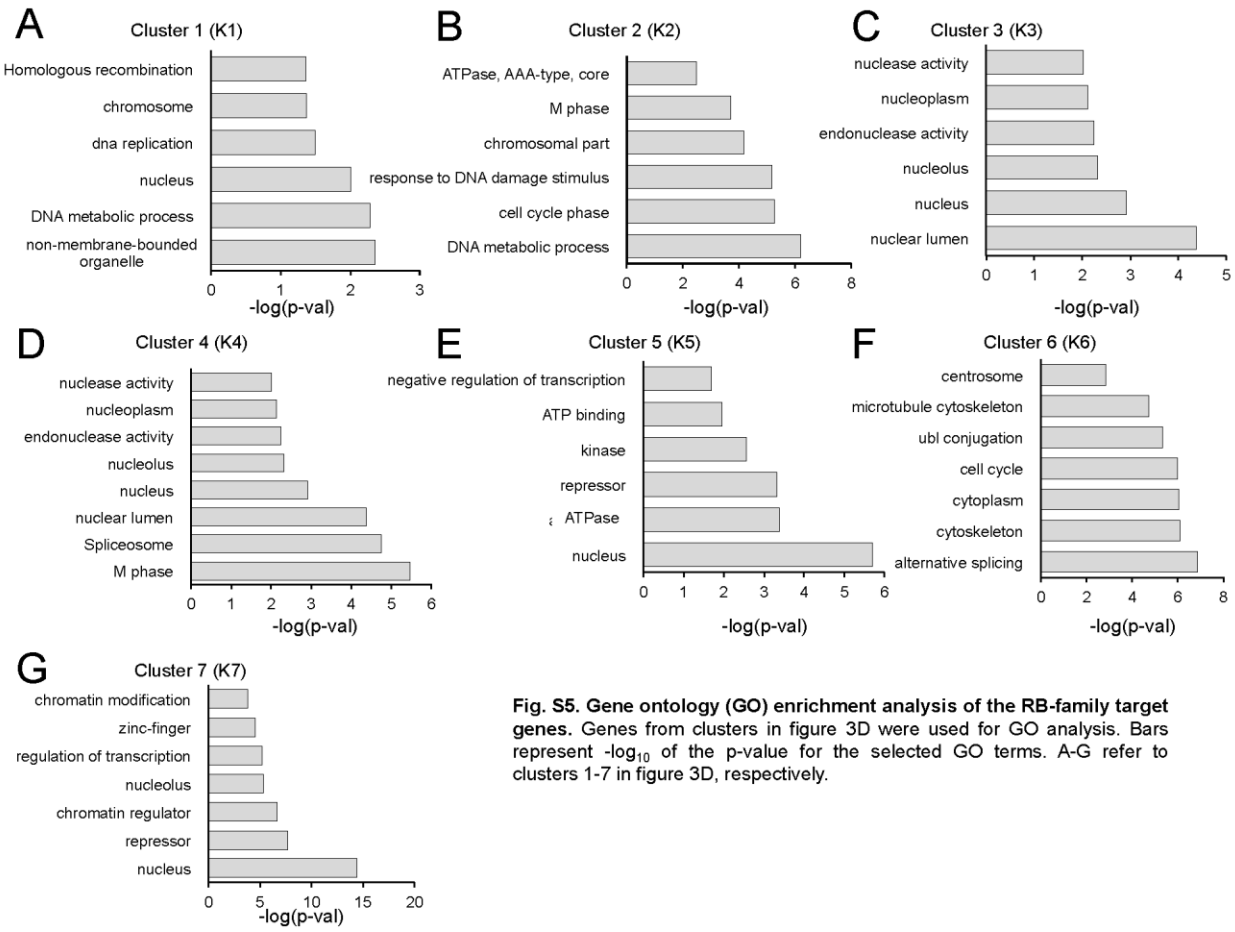


Fig. S5. Gene ontology (GO) enrichment analysis of the RB-family target genes. Genes from clusters in figure 3D were used for GO analysis. Bars represent $-\log_{10}$ of the p-value for the selected GO terms. A-G refer to clusters 1-7 in figure 3D, respectively.

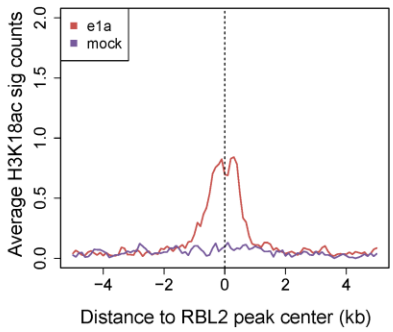


Fig. S6. The regions that are bound by RBL2 in contact-inhibited fibroblasts are targets of H3K18ac in e1a-expressing cells. Average levels (sig. counts) of H3K18ac in mock-infected and e1a-expressing cells are shown relative to RBL2-centered peaks that are 5 kb away from RB1 peaks.

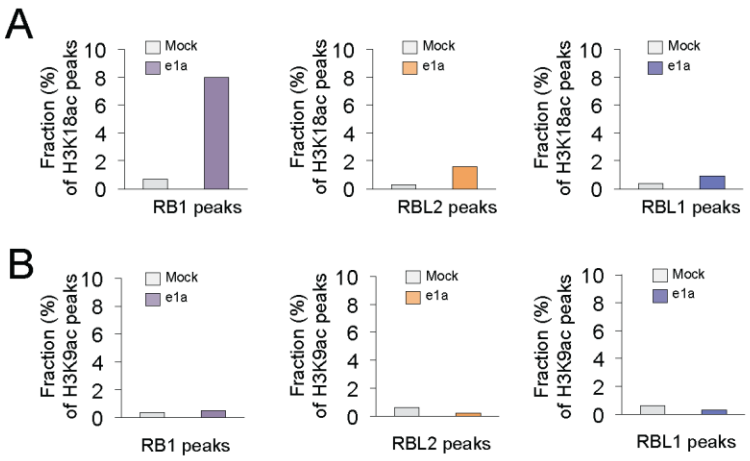


Fig. S7. H3K18ac in *dl*1500-infected cells is preferentially enriched precisely at regions bound by RB1 in uninfected cells. A) The fraction of H3K18ac peaks that precisely overlap with RB1, RBL2 and RBL1 peaks in mock- and *dl*1500-infected cells is shown. **B)** The same analysis as in A) for H3K9ac.

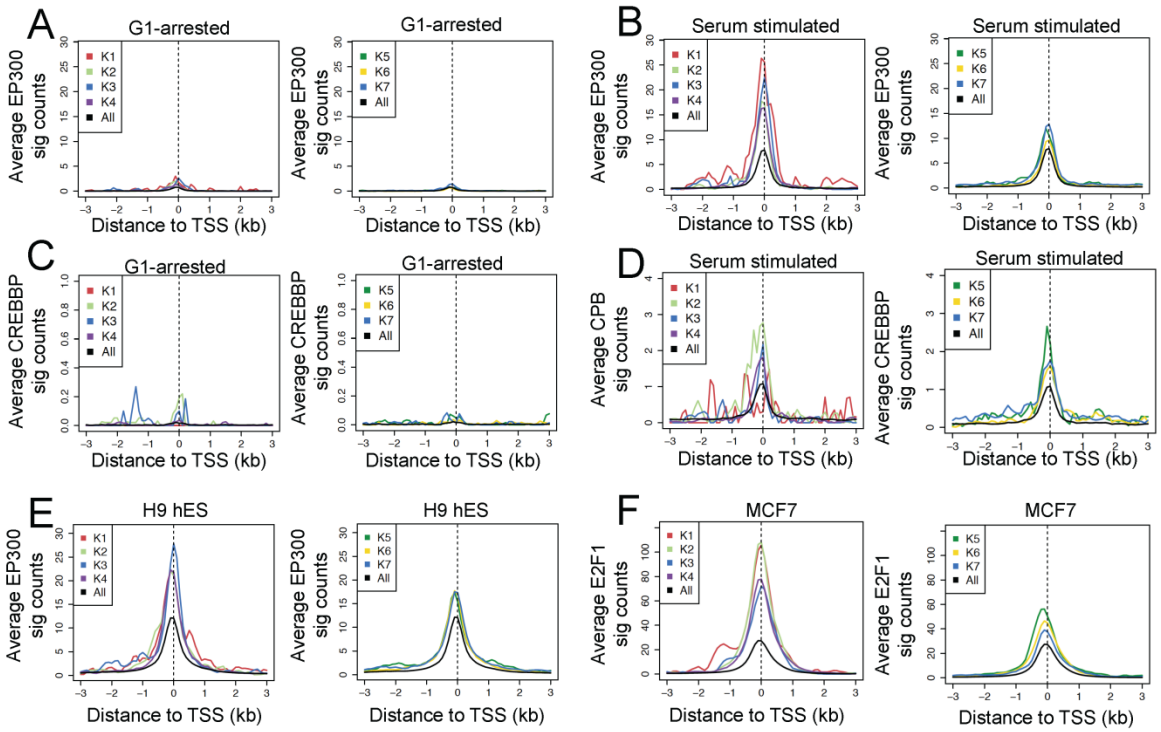


Fig. S8. Promoters bound by RB1 show increased levels of EP300 after serum stimulation in T98G glioblastoma and H9 hES cells. **A-B**) Average levels (significant (sig.) counts) of EP300 binding in the seven RB-clusters (see Fig. 3D) for G1-arrested and serum stimulated T98G cells. **C-D**) Average levels 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.

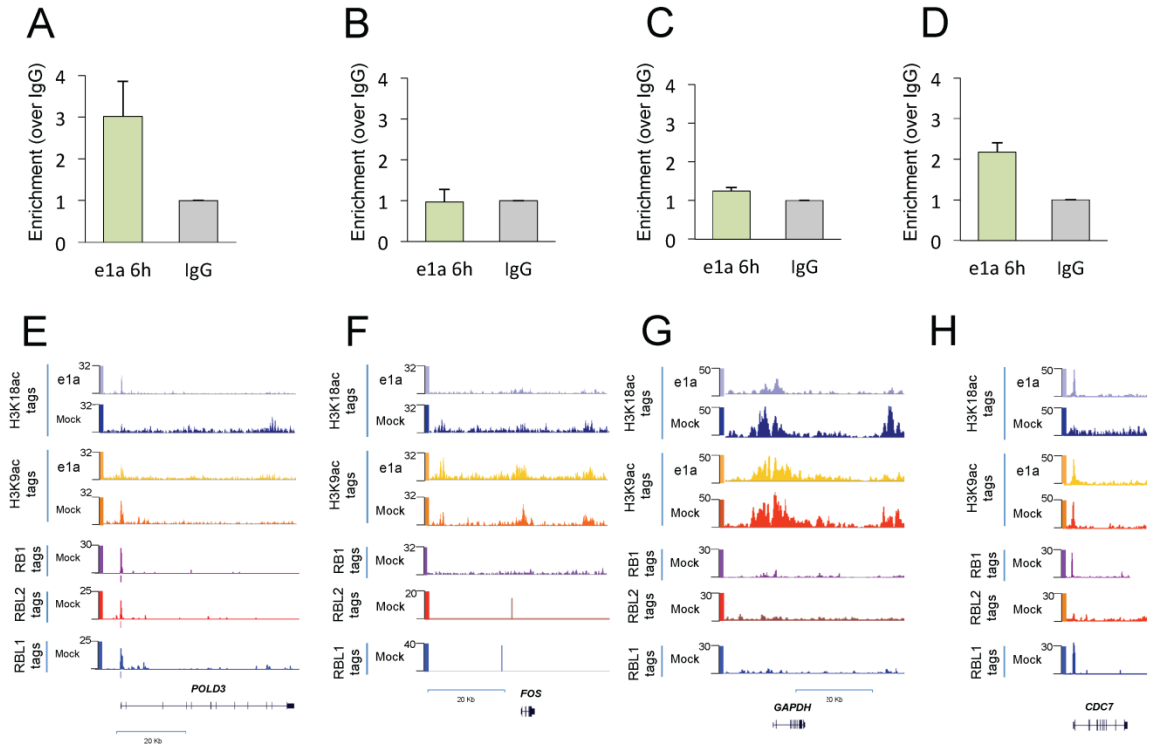


Fig. S9. Regions bound by RB1 in contact-inhibited fibroblasts are also targets of e1a binding. A-D) Small e1a binding at 6 h post infection is shown for *POLD3*, *FOS*, *GAPDG* and *CDC7* loci compared to IgG control. The 6 h time point was chosen based on transient and early binding of e1a to cell cycle regulated genes (Ferrari et al. 2008). **E-H)** Patterns of H3K18ac, H3K9ac in mock- and *dl*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).

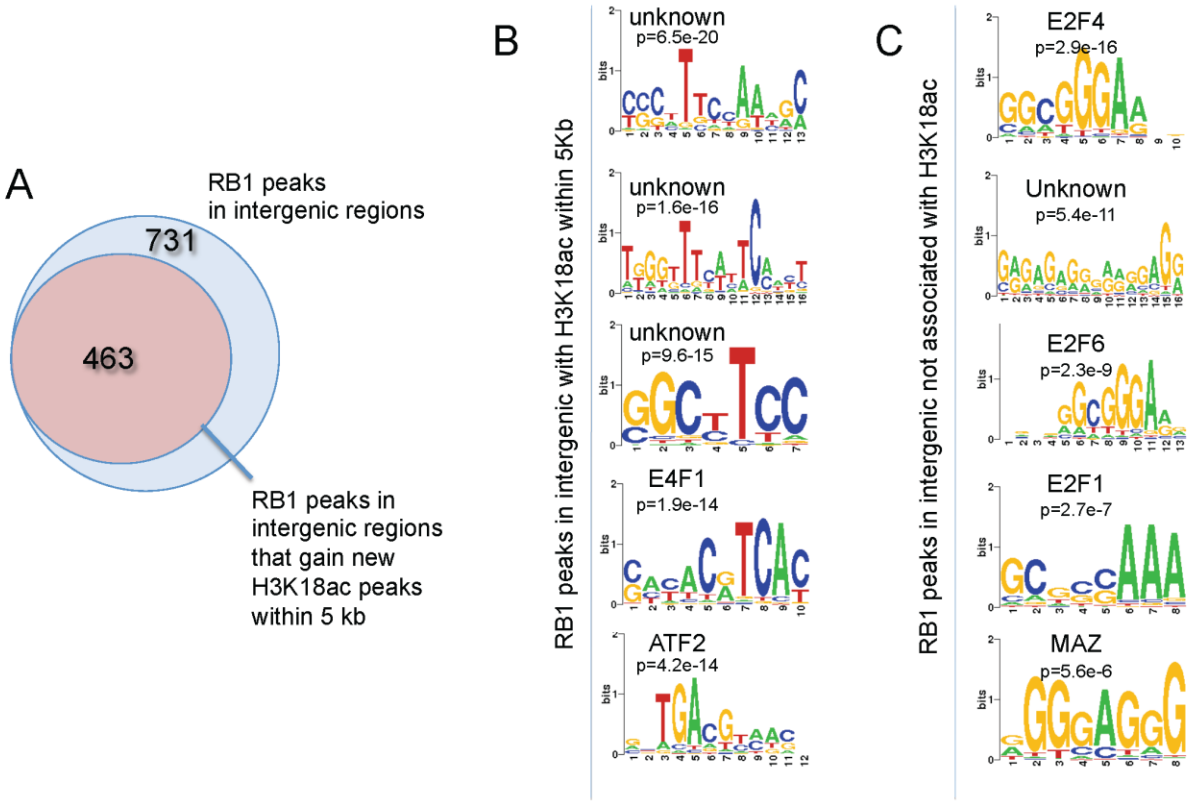


Fig. S10. H3K18ac peaks are associated with differential enrichment for transcription factor binding signatures in RB1-bound intergenic regions (IRs). A) ~63% of RB1 peaks in IRs are associated with H3K18ac. **B)** The 600-bp region around the peaks of IR RB1 associated with H3K18ac in e1a-expressing 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.

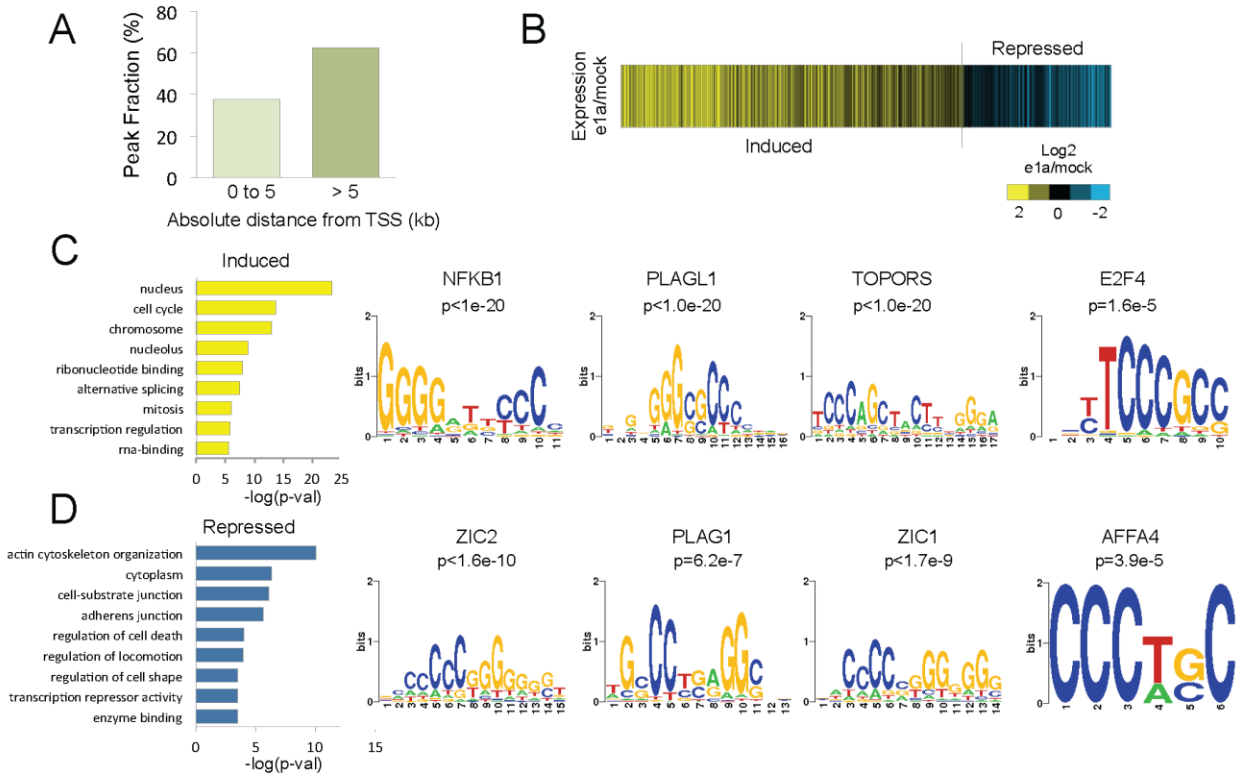


Fig. S11. H3K18ac peaks away from RB1 bound regions are relevant to the effect of e1a on the host cell. 3894 peaks of H3K18ac in e1a-expressing cells were found to be at least 5 kb away from any RB1 peak. **A)** GREAT analysis of peak locations. 60% of the 3894 peaks are located more than 5 kb from the 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).

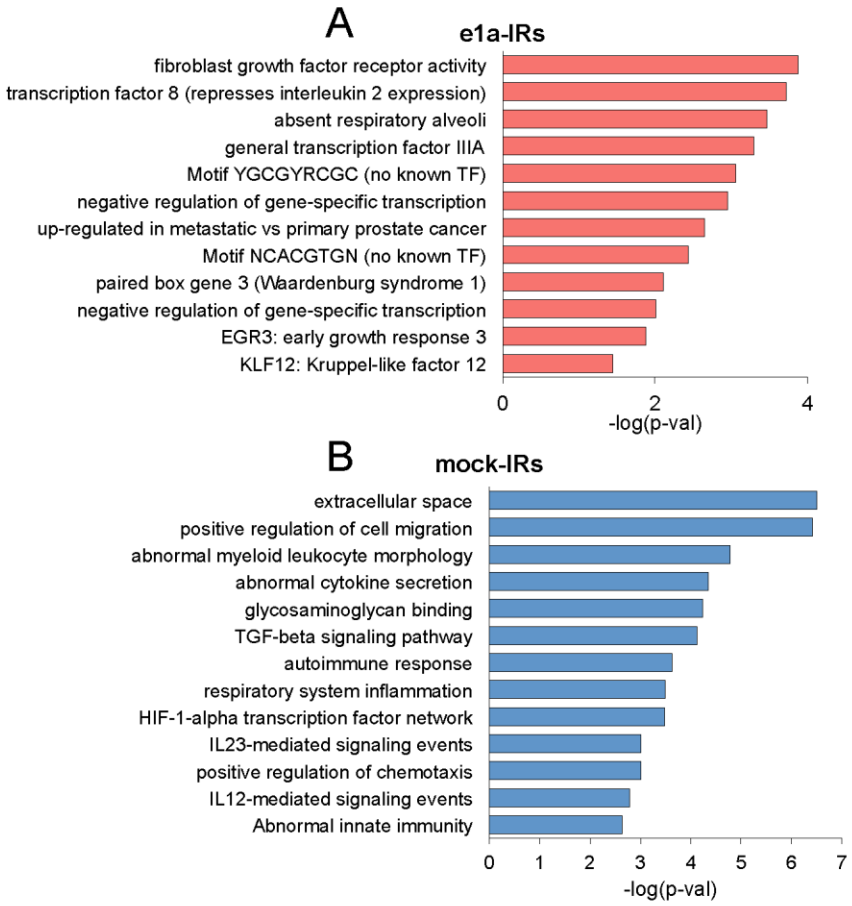


Fig. S12. Gene ontology (GO) enrichment analysis of genes associated with peaks of H3K18ac in intergenic regions. GREAT was used to find peak-gene associations for the peaks of H3K18ac in intergenic regions of mock-infected (mock-IRs) and e1a-expressing (e1a-IRs) cells. Bars represent $-\log_{10}$ of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

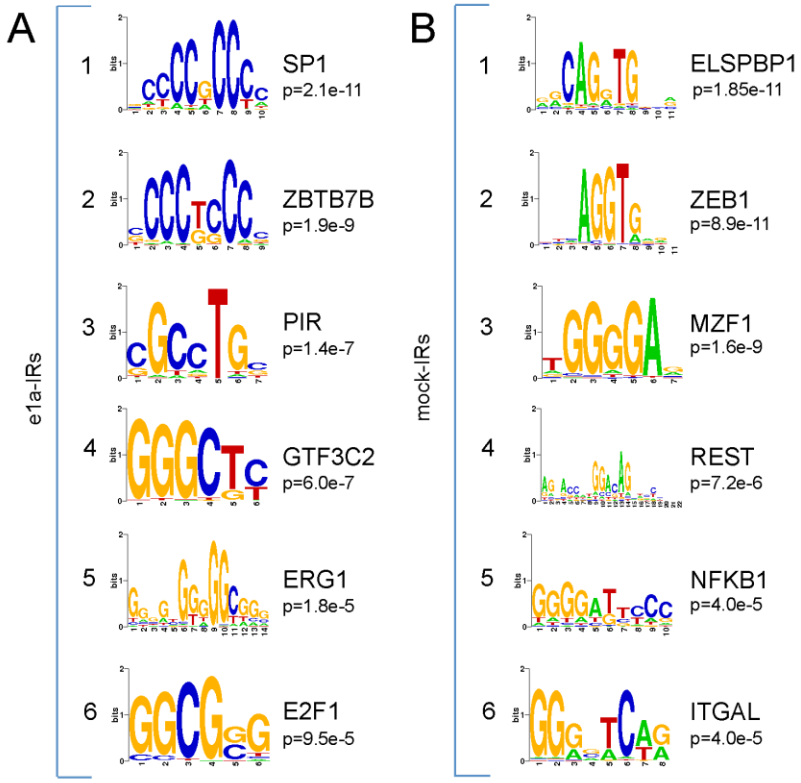


Fig. S13. Sequence logo representation of TFs binding motifs associated with intergenic peaks of H3K18ac. The 600-bp region around the peaks of H3K18ac in the intergenic regions of mock- (mock-IRs) and e1a-expressing (e1a-IRs) cells (Fig. 5A) were analyzed for TFs binding motifs using Seqpos (Galaxy).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

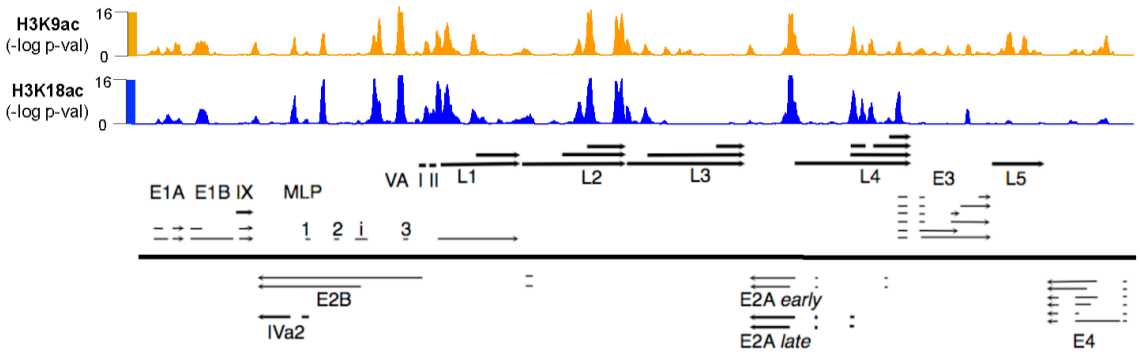


Fig. S14. Histones associated with the adenovirus genome are acetylated. Shown is the –log p-value of input-normalized tags across the entire adenovirus genome for H3K9ac (yellow track) and H3K18ac (dark blue track). Black arrows represent adenoviral transcripts and direction of the arrow corresponds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

Chapter 6

Genome-wide binding map

of

the HIV-1 Tat protein to the human genome

This chapter was originally published in *PLOS One.* 2011;6(11):e26894. with supplemental material. HIV-1 Tat protein functions in promotion of viral transcription. However, genome-wide distribution of HIV-1 Tat binding is still unclear. Using ChIP-seq in HIV-1 Tat expressing Jurkat cells, Tat protein was found to preferentially target genomic regions with Alu repeat. Additionally, gene promoters occupied by CBP in control Jurkat cells were also favored binding sites for HIV-1 Tat protein. This study provided the first observation of genome-wide binding pattern of HIV-1 Tat protein on chromatin of human cell line.

OPEN ACCESS Freely available online

PLoS one

Genome-Wide Binding Map of the HIV-1 Tat Protein to the Human Genome

Céline Marban[1,2◔], Trent Su[2,3◔], Roberto Ferrari[2], Bing Li[2], Dimitrios Vatakis[4], Matteo Pellegrini[5,6], Jerome A. Zack[4,6], Olivier Rohr[1,7]*, Siavash K. Kurdistani[2,6,8]*

1 Institut de Virologie, Université de Strasbourg, Strasbourg, France, **2** Department of Biological Chemistry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, United States of America, **3** Division of Oral Biology and Medicine, School of Dentistry, University of California Los Angeles, Los Angeles, California, United States of America, **4** Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, United States of America, **5** Department of Molecular, Cellular, and Developmental Biology, University of California Los Angeles, Los Angeles, California, United States of America, **6** Eli and Edythe Broad Centre of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, California, United States of America, **7** Institut Universitaire de France (IUF), Paris, France, **8** Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, United States of America

Abstract

The HIV-1 Trans-Activator of Transcription (Tat) protein binds to multiple host cellular factors and greatly enhances the level of transcription of the HIV genome. While Tat's control of viral transcription is well-studied, much less is known about the interaction of Tat with the human genome. Here, we report the genome-wide binding map of Tat to the human genome in Jurkat T cells using chromatin immunoprecipitation combined with next-generation sequencing. Surprisingly, we found that ~53% of the Tat target regions are within DNA repeat elements, greater than half of which are Alu sequences. The remaining target regions are located in introns and distal intergenic regions; only ~7% of Tat-bound regions are near transcription start sites (TSS) at gene promoters. Interestingly, Tat binds to promoters of genes that, in Jurkat cells, are bound by the ETS1 transcription factor, the CBP histone acetyltransferase and/or are enriched for histone H3 lysine 4 tri-methylation (H3K4me3) and H3K27me3. Tat binding is associated with genes enriched with functions in T cell biology and immune response. Our data reveal that Tat's interaction with the host genome is more extensive than previously thought, with potentially important implications for the viral life cycle.

Citation: Marban C, Su T, Ferrari R, Li B, Vatakis D, et al. (2011) Genome-Wide Binding Map of the HIV-1 Tat Protein to the Human Genome. PLoS ONE 6(11): e26894. doi:10.1371/journal.pone.0026894

Editor: John J. Rossi, Beckman Research Institute of the City of Hope, United States of America

Received July 12, 2011; **Accepted** September 22, 2011; **Published** November 4, 2011

Funding: Dr. Marban was supported in part by a fellowship from the Agence Nationale de Recherche sur le Sida. This work was also supported by a Howard Hughes Medical Institute award, and a National Institutes of Health (NIH) Innovator award as well as a California Institute for Regenerative Medicine grant to Dr. Kurdistani. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: skurdistani@mednet.ucla.edu (SKK); olivier.rohr@unistra.fr (OR)

◔ These authors contributed equally to this work.

Introduction

After gaining entry into host cell, the HIV-1 genome is reverse-transcribed and the proviral DNA is integrated into the host genome. Subsequently, the HIV-1 provirus is transcribed allowing assembly and release of new viral particles from the infected cell. HIV-1 Tat is essential for efficient viral gene expression and replication [1]. By recruiting the general RNA polymerase II elongation factor P-TEFb to Tat response element (TAR) that forms at the 5′ end of nascent viral transcripts, Tat promotes efficient elongation of viral transcription [2]. Moreover, Tat acetylation by cellular histone acetyltransferases (HATs) such as p300, CBP and PCAF is crucial for its transactivation activity [3]. While the role of Tat in viral gene expression has been well studied, much less is known about the interaction of Tat with the host genome. Previous studies that aimed to define the role of Tat at the host gene promoters found that Tat regulates transcription of the interleukin 6 [4], MHC class I [5], ß2 microglobulin [6] and mannose receptor [7] promoters. Tat also induces host cell apoptosis through association with promoters of PTEN and two PP2A subunits [8]. Therefore, Tat may have roles in regulation of gene expression from the viral as well as the host genome. However, a genome-wide map of Tat interaction with the human genome is still lacking. Such a binding map may reveal additional roles for Tat in creating the proper cellular environment for generating progeny virions.

To generate a genome-wide map of Tat binding to the human genome, we performed chromatin immunoprecipitation combined with next generation sequencing (ChIP-seq) of Tat in Jurkat T cells (Jurkat-Tat). We also utilized microarrays to compare global gene expression changes in Jurkat-Tat versus Jurkat T cells and related the expression differences to histone acetylation changes. We found that the bulk of Tat binding sites are outside the immediate promoter regions of genes. Intriguingly, Tat binds preferentially to specific DNA repetitive elements, especially the Alu repeat elements. Binding of Tat to the promoter regions did not correlate with gene expression. The majority of Tat binding sites at gene promoters in Jurkat-Tat cells are in close proximity with regions bound by the ETS1 transcription factor or CBP in Jurkat T cells. Our data provide the first comprehensive map of Tat binding to

Table 1. ChIP-seq alignment results using Bowtie 0.12.7 (Hg19).

DNA	Total reads	Aligned Reads (%)
Input	42249133	26416727 (62.5)
ChIPed	44360277	27831745 (62.7)

doi:10.1371/journal.pone.0026894.t001

the human genome, revealing an unexpected array of target regions.

Results

Genome-wide Tat binding locations defined by ChIP-seq

To determine whether Tat binds to specific regions in the host genome, we performed ChIP-seq to map Tat binding sites in Jurkat T cells that stably express Tat under G418 selection [9]. We first validated Tat expression in Jurkat-Tat cells with Western blotting (Figure S1). Subsequently, we sequenced both input and ChIPed Tat-bound DNA using the Illumina GAIIx Sequencer. The obtained sequences were aligned to the human genome (version Hg19) using the Bowtie software [10]. For both input and ChIPed samples, ~62% of all sequences were uniquely aligned to the human genome (Table 1). We segmented the human genome into 100 bp windows and calculated the ChIPed DNA read counts, which were compared to input DNA read counts in each window. Using the Poisson distribution, we calculated *P*-values for the enrichment of ChIPed reads in each window. Significant peaks were defined as those windows with a *P*-value$<10^{-4}$ and with two neighboring windows at the same significance *P*-value. Based on these criteria, we identified 2074 genomic regions occupied by Tat in Jurkat-Tat T cells.

Using the cis-regulatory element annotation system (CEAS) software [11] to determine the Tat binding distribution pattern across individual chromosomes, we observed that chromosomes 2, 9, 14, 15, 21 and 22 were significantly ($P<10^{-4}$) enriched for Tat binding (Figure 1A). The majority (82%) of Tat binding occurs within introns and intergenic regions in the genome (Figure 1B). Intergenic regions are defined as those regions that are at least 3 kilo bases (kb) away from any known gene. Only ~7% of Tat binding sites are located within the promoter regions. These data suggest that Tat binding to the genome is non-random, with preferential binding to certain chromosomes and intergenic regions.

Tat binding loci are enriched within repeat sequences

To determine whether Tat binding regions are associated with specific genomic features, we obtained the coordinates of various DNA elements from the UCSC table browser website [12]. To our surprise, we found that 53% of Tat bound regions lie within repeat-masker (rmsk) regions, which record repeat elements found by RepeatMasker [13,14] (Figure 2A). We systematically determined the enrichment of Tat binding in various repeat elements and found that, strikingly, 58 percent of all repeat elements bound

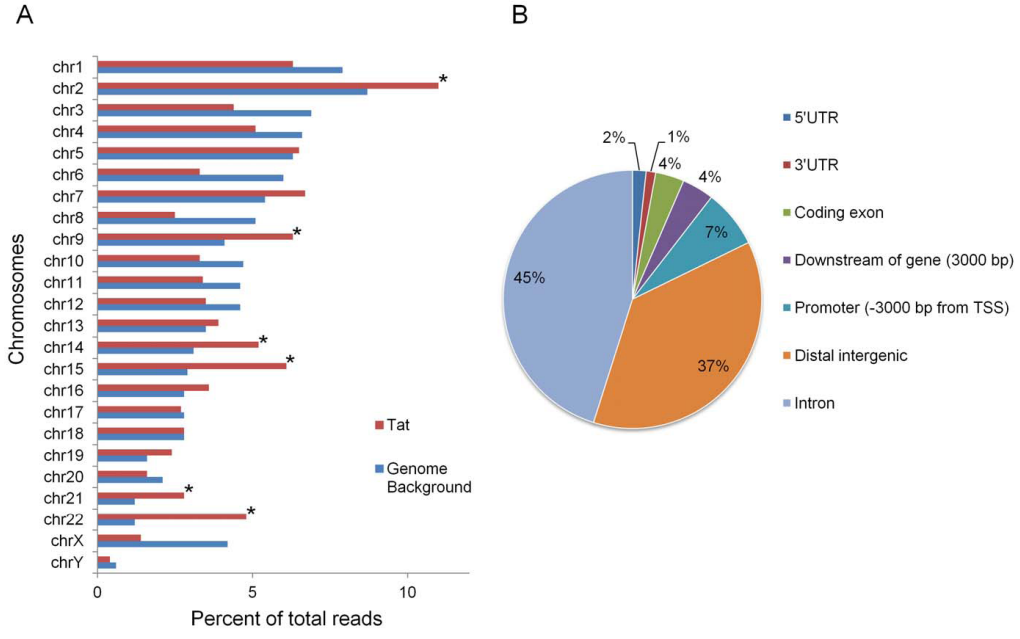


Figure 1. Chromosomal distribution and genomic location of Tat binding sites. (**A**) Enrichment pattern of Tat-bound regions among individual chromosomes is shown as a bar chart. Percent of total Tat-binding sites (red bars) and what would be expected by random chance (blue bars) for each chromosome is shown. The asterisks denotes enrichment *P*-value$<10^{-4}$. (**B**) Distribution of all Tat-binding peaks in relation to gene structure is shown as a pie chart. Intergenic regions are defined as at least 3 kb away from the start and end of any transcript.
doi:10.1371/journal.pone.0026894.g001

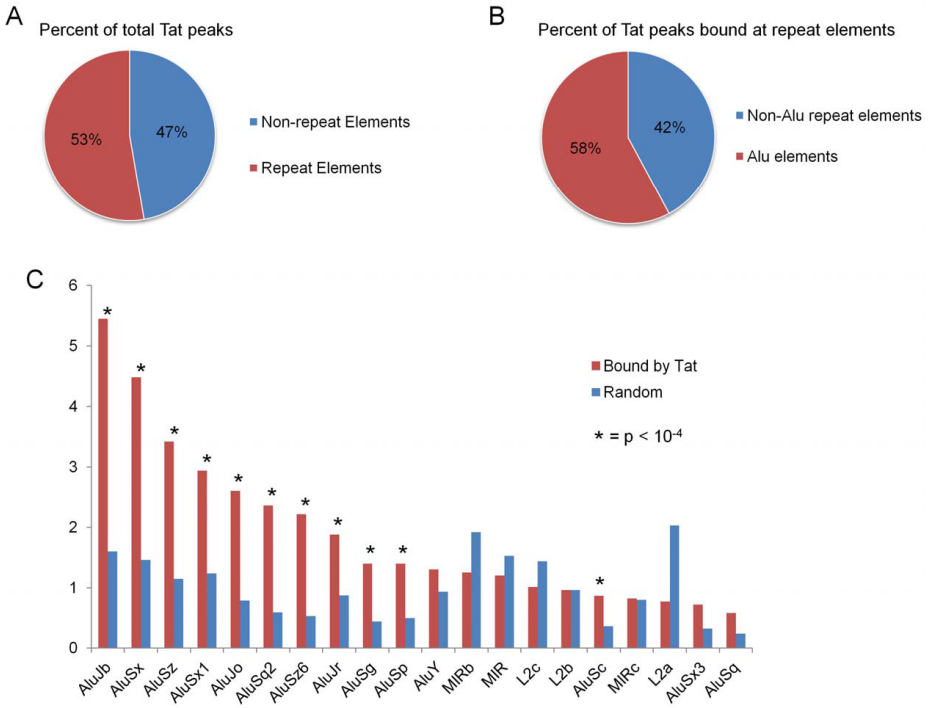


Figure 2. Tat binds mainly to DNA repeat elements. (**A**) Distribution of Tat-binding peaks in repeat versus non-repeat elements is shown as a pie chart. (**B**) Distribution of Tat-binding peaks within repeat elements in Alu versus non-Alu sequences is shown as a pie chart. (**C**) Percent of Tat-binding peaks within individual repeat element subtypes is shown for the top 20 enriched elements.
doi:10.1371/journal.pone.0026894.g002

by Tat are Alu repeats (Figure 2B). In fact, the top ten repeat element types bound by Tat all belong to the Alu family of DNA repeats (Figure 2C). These data reveal that a large fraction of Tat binding regions in the genome are within DNA repeat elements, especially the Alu sequences.

Tat binding is enriched in the middle to the 3′ end of Alu elements

Since more than one third of Tat-bound genomic regions contain Alu elements, we sought to determine the binding profile of Tat to Alu elements. As shown in Figure 3A, Alu elements have an average size of 300 base pairs (bp). We generated an average profile of Tat binding centered at the start of all Tat-bound Alu elements. The profile of Tat enrichment over Alu elements was generated by averaging *P*-values of Tat binding enrichment in 100-bp windows ±1 kb from the start of Alu elements. Tat binding on average peaked in the middle of Alu elements with a skewed enrichment toward the 3′ end and downstream regions, up to 200 bp past the average length of an Alu element (Figure 3A). Figure 3B shows an example of Tat binding at an Alu element in an intergenic region of chromosome 10 as indicated. We then asked whether Tat-bound Alu-elements are enriched at specific genomic regions using the CEAS software [11]. The Tat-bound Alu elements are primarily located in introns and intergenic regions of the genome (Figure 3C). In comparison to global Tat binding patterns, Alu elements bound by Tat are more enriched within introns (45% vs 53%). Essentially none of the Tat-bound Alu elements was located within coding exons. Altogether, our data reveals that Tat binds specifically to a fraction of Alu elements within the human genome. These Alu elements may be near or within the introns of genes with functions potentially related to the viral life cycle such as the Alu element shown in Figure 3B.

Genomic regions bound by Tat are associated with genes enriched in T cell-related functions

As the majority of genomic regions bound by Tat are at least 3 kb away from any TSS, we asked if the nearest genes to Tat binding sites are enriched for specific functions. To observe long range interactions between Tat binding sites and their target genes, we used the Genomic Regions Enrichment Annotations Tool (GREAT) [15] to determine whether genomic regions bound by Tat are located in potential *cis*-regulatory regions of genes important for HIV function. GREAT analysis of Tat-bound regions revealed that the genes potentially associated with distal Tat binding are mostly 5 kb or further away from Tat binding sites. This distribution implies that there may be long range interactions between Tat and its potential target genes (Figure 4A). In support of this, genes associated with regions bound by Tat are

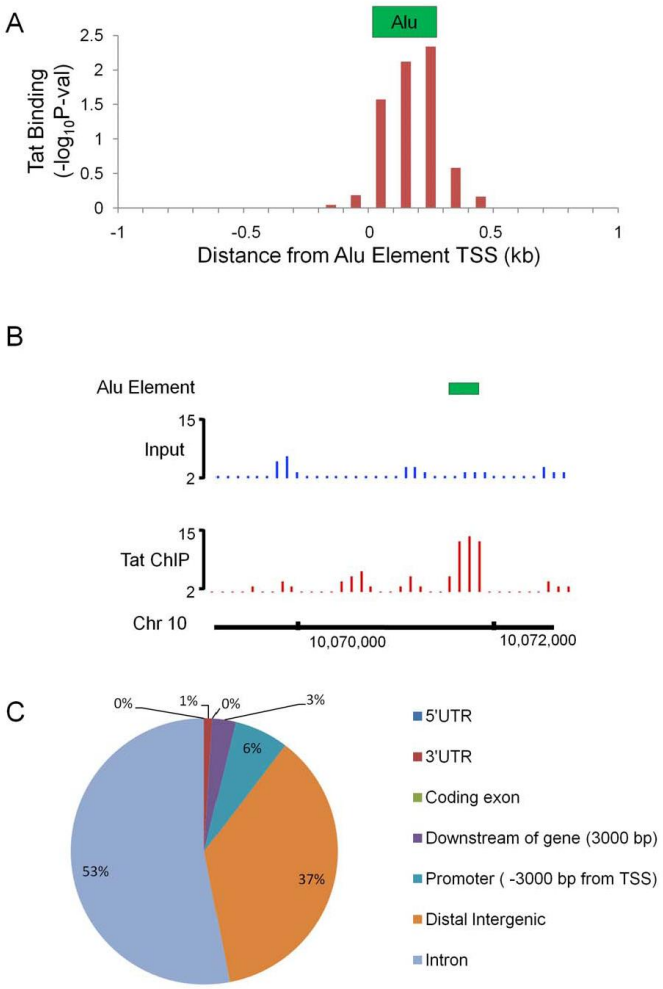


Figure 3. Analysis of Tat binding to Alu elements. (**A**) Average *P*-value of Tat-binding profile centered at the start of Alu elements for all Tat-bound Alu elements is shown. The x axis depicts ±1 kb away from start of Tat-bound Alu elements in 100-bp intervals. The y axis denotes $-\log_{10}$ of Tat-binding enrichment *P*-value of ChIPed Tat DNA count over input DNA count. (**B**) Genome browser representation of Tat enrichment profile compared to the input DNA at a representative Alu element is shown. The peak height corresponds to read counts. (**C**) Distribution of Tat-bound Alu elements with respect to gene structure is shown as a pie chart.
doi:10.1371/journal.pone.0026894.g003

significantly enriched in Mouse Genome Informatics (MGI) Phenotype ontology terms related to T cell function (Figures 4B and 4C). MGI analyzes the knockout phenotypes of mouse genes that are homologous to the queried human genes. The mouse homologues of the potentially Tat-regulated genes exhibit knockout phenotypes such as changes in T cell morphology and reduced number of CD4+ and CD8+ T cells (Figure 4B). Similar gene ontology terms were observed when only genes associated with Tat-bound Alu elements were used for MGI analysis (Figure 4C). These data suggest that Tat may exert its effects on the host genes by binding to distant *cis*-regulatory elements.

Global gene expression changes in Jurkat-Tat cells

To relate global gene expression changes to Tat binding, we used Agilent microarrays to compare global gene expression between Jurkat and Jurkat-Tat cells. Overall, 475 and 319 transcripts showed greater than two-fold increase and decrease in gene expression, respectively. To investigate the functions of the

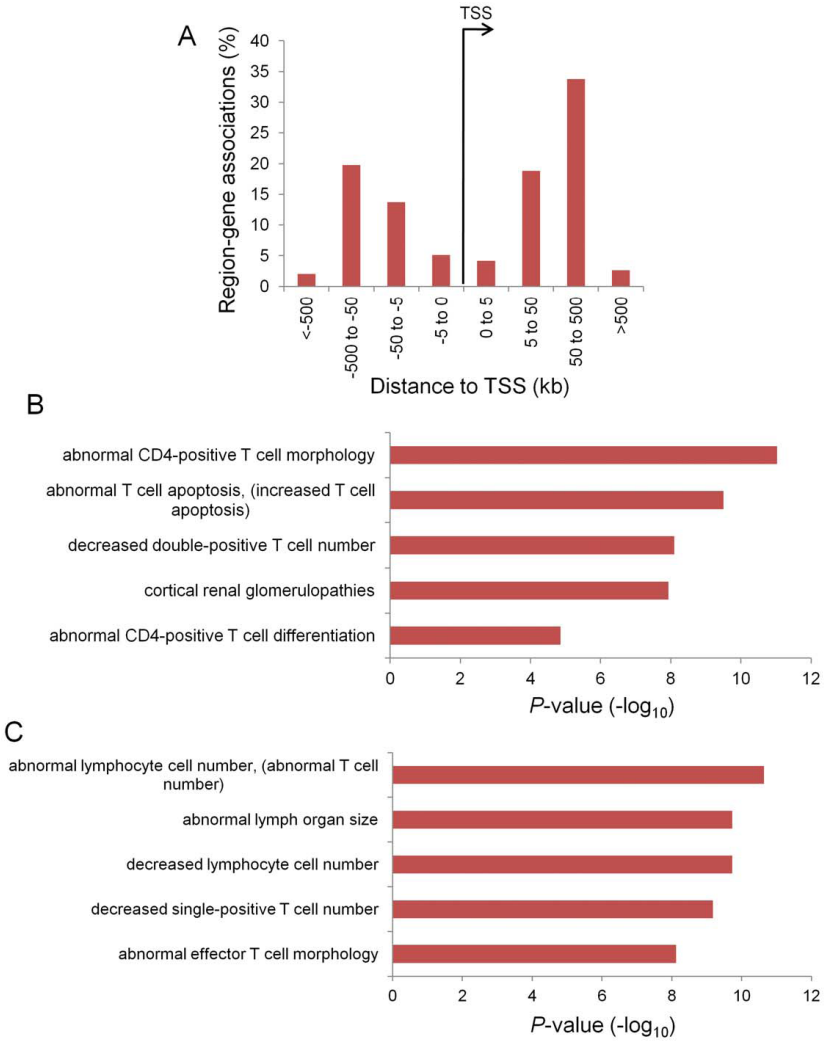


Figure 4. Functional annotation of Tat binding regions. (**A**) The bar chart shows the distribution of the distances between Tat binding peaks and transcriptional start sites (TSS). Note that a Tat peak may be assigned to more than one gene. Functional annotation of (**B**) Tat-bound regions and (**C**) Tat-bound Alu elements, using gene ontology terms generated by Mouse Genome Informatics (MGI) from mouse homologous gene knock out phenotypes. The x-axis values are $-\log_{10}$ of binomial raw *P*-values.
doi:10.1371/journal.pone.0026894.g004

deregulated genes and their relevance to Tat over expression, we performed Gene Ontology (GO) enrichment analysis using DAVID [16]. Genes overexpressed in Jurkat-Tat cells are enriched in cell immune response, cell adhesion, and regulation of cell death (Figure 5A). However, for the down-regulated genes, no significant GO enrichment was found. To determine whether Tat binding is associated with increased or decreased expression of its target genes, we used the GREAT assigned gene-peak associations to correlate Tat binding to gene expression. Surprisingly, we found no significant correlation. Additionally, we performed ChIP combined with Agilent promoter microarrays (ChIP-chip) to monitor changes in histone H3 lysine 9 acetylation (H3K9ac) a histone modification associated with gene activity in Jurkat-Tat versus Jurkat cells. We found that changes in H3K9ac correlated positively with gene expression changes; promoter regions of gene that are up- and down-regulated in

A

DAVID Gene Ontology of Up-regulated Genes in Tat-Jurkat cells

Term	P Value	Benjamini
immune response	8.35E-07	0.001931
biological adhesion, (cell adhesion), (cell-cell adhesion)	1.92E-05	0.014673
regulation of cell death	9.10E-05	0.034493
immune system development	1.51E-04	0.038056
positive regulation of cytokine biosynthetic process	1.50E-04	0.042562
cell activation, (leukocyte activation), (T cell differentiation)	8.14E-05	0.04601

DAVID Gene Ontology of Down-regulated Genes in Jurkat-Tat cells

Term	P Value	Benjamini
No significant enrichment (Benjamini<0.05)		

B

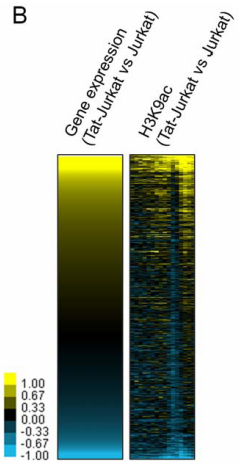


Figure 5. Gene expression profile of Jurkat-Tat cells. (**A**) Gene ontology annotation of genes with more than two-fold change in expression using DAVID. (**B**) Relationships of gene expression and H3K9ac changes to Tat-binding are shown. Genes are arranged in descending order based on their expression in Jurkat-Tat versus Jurkat cells. Each row represents one gene. For H3K9ac ChIP-chip analysis, each column represents a 500-bp window, spanning −5.5 to +2.5 kb of annotated TSS. Moving average of H3K9ac enrichment in three consecutive 500-bp windows is shown in each column (right panel).
doi:10.1371/journal.pone.0026894.g005

Jurkat-Tat cells have higher and lower levels of H3K9ac, respectively, compared to Jurkat cells (Figure 5B). These data indicate that changes in gene expression are associated with similar changes in histone acetylation but Tat may have more subtle effects on gene expression that is not detected by our microarray analysis.

Tat in Jurkat-Tat cells binds to locations occupied by CBP and ETS1

To determine if Tat binding sites are associated with specific chromatin marks, cellular transcription factors or co-factors, we searched the Gene Expression Omnibus (GEO)[17] for published ChIP-seq data in Jurkat or Jurkat-Tat cells. We found five datasets that examined global distributions of ETS1, CBP, RUNX1, H3K4 tri-methylation (H3K4me3) and H3K27me3 in Jurkat cells (GSE23080, GSE17954 [18]). No published dataset in Jurkat-Tat cells was found. ETS1 is a transcription factor that is highly expressed in lymphoid lineage cells and is important for regulating functions of immune cells [19]. In mouse models, inactivation of ETS1 leads to T cell apoptosis [20]. ETS1 and Tat were previously shown to bind at the same region upstream of IL-10 promoter and induce IL-10 transcription [21]. P300 and CBP HATs acetylate Tat and serve as co-activators of Tat-dependent HIV-1 gene expression [3]. RUNX1 is a member of Runt-related transcription factor (RUNX) family of genes that function in normal hematopoiesis. H3K4me3 and H3K27me3 are histone modifications that are generally associated with gene activity and repression, respectively [22].

We downloaded ChIP-seq raw data from GEO [17] and compared each ChIP channel to its corresponding input DNA from the same experimental set. Using the same peak finding algorithm described above with a *P*-value cutoff of $<10^{-4}$, we defined significant peaks of CBP, ETS1, H3K4me3, H3K27me3 in Jurkat T cells. For RUNX1, we found much less significant reads using *P*-value of 10^{-4}. To get more similar numbers of total significant peaks, *P*-value of 10^{-2} was used to determine significant reads for RUNX1. Significant peaks found in Jurkat T cells were then compared to Tat binding sites in Jurkat-Tat cells. We defined positive co-occupancy of each factor with Tat when there was at least one significant peak of binding within ±500 bp of the Tat-binding sites. Figure 6A shows the fraction of all (2074) Tat binding sites that are occupied by the indicated factors in Jurkat cells. Only ~12% of all Tat binding sites coincide with H3K4me3 and/or H3K27me3 and even less so with CBP or ETS1. However, when the analysis is limited to the Tat binding sites within ±1 kb of TSS regions (162 sites), 55% and 42% of Tat peaks are bound by ETS1 and CBP, respectively, in Jurkat cells; 35% are co-occupied by both factors (Figure 6B). Only 1% of Tat binding sites near TSS overlap with RUNX1 binding in Jurkat cells with *P*-value of 10^{-2} and 10^{-4}. Furthermore, 28% of Tat TSS binding sites are enriched for H3K4me3 and H3K27me3 in Jurkat cells, including 58% of binding sites with both ETS1 and CBP bound near the TSS (Figure 6B). Figure 6C shows an example of such genes. The promoter of ZNF143 gene is co-occupied by ETS1 and CBP and enriched for H3K4me3 and H3K27me3 in Jurkat cells. The same region is bound by Tat in Jurkat-Tat cells. In contrast, no significant co-occupancy of Tat in Jurkat-Tat and RUNX1 in Jurkat cells was found near TSS. Although the occupancy of these factors in Jurkat-Tat cells is not known, our analysis raises the possibility that specific cellular transcription factors may help recruit Tat or denote the genomic regions to which Tat binds.

Discussion

We have generated a global map of Tat binding to the human genome. Surprisingly, the majority of Tat target regions lie within DNA repeat elements. In fact, over 30% of all Tat target regions are located at or near Alu elements. Interestingly, Tat increases the transcription of Alu repeat elements by increasing the activity of

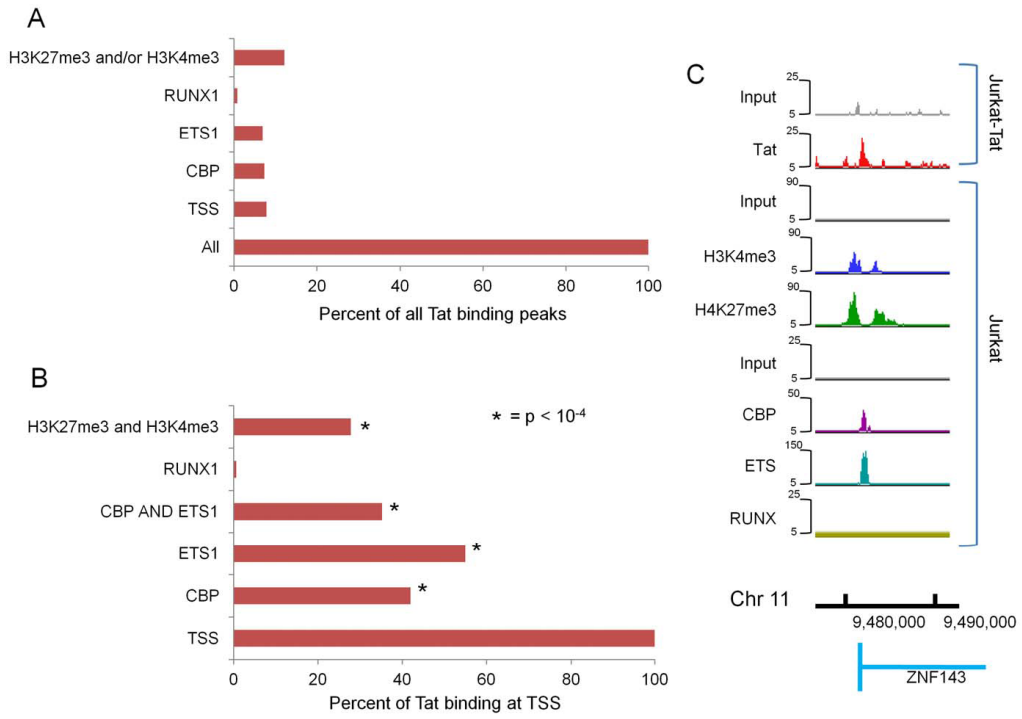


Figure 6. Tat binding peaks in Jurkat-Tat cells are associated with specific cellular factors and chromatin marks in Jurkat cells. (**A**) Percent of all Tat peaks in Jurkat-Tat cells that are bound by CBP, ETS1, RUNX1, H3K4me3 and/or H3K27me3 in Jurkat cells is shown. (**B**) Percent of Tat peaks at TSS in Jurkat-Tat cells that are bound by CBP, ETS1, RUNX1, H3K4me3 and H3K27me3 in Jurkat cells is shown. (**C**) Shown is the Tat enrichment profile at the ZNF143 gene promoter in Jurkat-Tat cells compared to CBP, ETS1, RUNX1, H3K4me3, H3K27me3 enrichment at the same genomic location in Jurkat cells. The peak heights (y-axis) correspond to read counts. The x-axis represents the genomic coordinates.
doi:10.1371/journal.pone.0026894.g006

cellular transcription factor TFIIIC in Jurkat cells [23]. Since Alu elements antagonize the interferon-induced protein kinase R (PKR) activation [24] and PKR is known to repress protein synthesis when cells are under stress, Tat binding at Alu elements may be important to enable efficient viral replication in the host cell. In addition, these Alu elements may affect regulation of genes with functions related to HIV-1 biology.

In relation to gene regulation, Tat binding sites are distal to genes with functions in T cell biology as determined by knockout models in mice. These data suggest that Tat may exert its effects on its target genes through distant regulatory elements. If so, then Tat binding, in and of itself, may highlight previously unknown *cis*-regulatory elements within the genome. We did not find a significant correlation between Tat binding and gene expression. The effects of Tat on gene expression may be too subtle for the microarrays to detect. Alternatively, Tat may affect the regulation of its target genes through effects on mRNA structure that is not evident by expression analysis.

We found only a few genes with Tat binding to the vicinity of their TSS. This is partly due to the stringent cutoff used in this study to maintain a False Discovery Rate (FDR) under 5%. For instance, gene promoters previously shown to be bound by Tat binding such as IL-6 [4] and ß2 microglobulin [6], showed some Tat binding above background but did not pass our criteria for statistical significance. Also, previous studies used different cell culture systems to determine Tat binding in vivo. These differences may also lead to identification of different sets of target genes.

Since Tat is not known to bind DNA directly, the mechanism by which Tat binds to specific regions of the genome may partly involve interactions with host cellular factors. By comparing our Tat binding data to the published datasets in Jurkat cells, we found that Tat binds to gene promoters that were bound by ETS1 and CBP, but not with RUNX1, in non-Tat expressing Jurkat cells. A similar positive relationship was found between Tat binding and two histone methylation marks, H3K4me3 and H3K27me3. These data raise the possibility that Tat may distinguish its target regions through specific host transcription co-factors or chromatin marks. However, our data does not include or exclude ETS1, CBP or the histone modifications as the mediators of Tat recruitment. Binding analyses of specific Tat mutants that disrupt its interaction with specific cellular factors are required to determine potential mechanisms of Tat recruitment. It is also possible that Tat binds to the genome through an RNA component. Nonetheless, our results demonstrate that in addition to known roles for Tat in enhancing

elongation of viral transcription, Tat also binds to the host genome at specific genomic locations with potentially important consequences for the viral life cycle.

Materials and Methods

Cell culture

Jurkat T cells were obtained from ATCC (TIB-15) and cultured under standard tissue culture conditions. Jurkat-Tat cells were obtained from the NIH AIDS Research & Reference Reagent Program and maintained in RPMI supplemented with 10% fetal bovine serum, 1% penicillin and streptomycin and 800 μg/ml of G418.

Western blotting

Cells were lysed with RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.1% SDS, 1% sodium deoxycholate) supplemented with protease inhibitors (Roche). Cell lysates were subjected to SDS-PAGE and analyzed by Western blot using standard procedures. The antibodies used for Western blotting were as follows: Anti-Tat (Abcam ab43014), Anti-Beta-actin (Abcam ab8224), Anti-H3 (Abcam ab10799).

Chromatin immunoprecipitation and ChIP-seq library preparation

Jurkat-Tat cells in exponential growth phase were fixed with 1% formaldehyde (v/v) for 10 min at 37°C. Fixation was stopped by addition of glycine to a final concentration of 140 mM. Cell were lysed and chromatin was digested with the micrococcal nuclease from *S. aureus* (Roche) for 90 min at 4°C according to the manufacturer's instructions, re-suspended in lysis buffer and sonicated with Misonix ultrasonic liquid processor. 1% of the lysate was used as an input control. Lysates were immunoprecipitated with 5 μg of anti-Tat antibodies (Abcam ab43014) using the standard ChIP protocol. Both purified input and Tat chromatin samples were used to prepare ChIP-seq libraries according to the manufacturer's instructions (Illumina). Libraries were sequenced using Illumina Genome Analyser II to obtain 76 bp-long reads.

Chromatin immunoprecipitation and microarray hybridization

Jurkat-Tat and Jurkat cells in exponential growth phase were fixed with 1% formaldehyde (v/v) for 10 min at 37°C. Fixation was stopped by addition of glycine to a final concentration of 140 mM. Histone H3 acetyl lysine 9 antibody (Upstate 07-352) was used for ChIP. The ChIP-chip and subsequent analysis were performed essentially as described previously [25].

ChIP-seq data analysis

Sequenced reads were base-called using the standard Illumina software. Bowtie 0.12.7 was used to align the reads to the Human genome (Hg19) allowing up to two mismatches; reads that aligned to more than one location in the genome were discarded. For unaligned reads, 5 bp from the 5′ end and 25 bp from the 3′ end of the reads were trimmed and re-aligned to the genome. The total number of reads in the input sample was normalized to the ChIP counts. Genomic regions with significant enrichment of Tat chromatin over input chromatin were calculated within 100 bp windows that tiled the genome. The *P*-value for enrichment of ChIP versus input reads was calculated using the cumulative Poisson distribution.

Expression profiling

Total RNA was isolated from Jurkat and Jurkat-Tat cells using the RNeasy Mini kit (Qiagen). cRNAs were generated from 250 ng of total RNA and labeled with Cy3 (Jurkat) or Cy5 (Jurkat-Tat) using the Low Input Quick Amp Labeling Kit (Agilent) according to manufacturer's instructions. Labeled cRNAs were hybridized to the Agilent Human whole-genome array (G2534-600110) according to Agilent protocol. Raw intensity data from resulting gene expression data were normalized by medium background-subtracted intensities between Cy5 and Cy3 channels followed by LOWESS normalization using Matlab. Normalized $\log_2$ ratios of Jurkat-Tat cRNAs (Cy5) over Jurkat cRNAs (Cy3) were calculated and results from replicates were averaged. Gene ontology enrichment analysis was performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 [16].

All data have been deposited in the Gene Expression Omnibus (GEO) [17] under accession number GSE30739.

Supporting Information

Figure S1 Tat is expressed in Jurkat-Tat cells. Western blots of the indicated factors in normal Jurkat T cells (Jurkat) and Jurkat cells stably expressing Tat (Jurkat-Tat).
(TIF)

Author Contributions

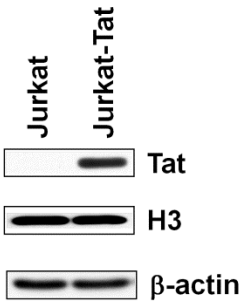
Conceived and designed the experiments: CM OR SKK. Performed the experiments: CM. Analyzed the data: TS CM RF BL. Contributed reagents/materials/analysis tools: MP DV JAZ. Wrote the paper: TS CM SKK.

References

1. Jeang KT, Chang Y, Berkhout B, Hammarskjold ML, Rekosh D (1991) Regulation of HIV expression: mechanisms of action of Tat and Rev. AIDS 5 Suppl 2: S3 14.
2. Parada CA, Roeder RG (1999) A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. The EMBO journal 18: 3688 3701.
3. Deng L, de la Fuente C, Fu P, Wang L, Donnelly R, et al. (2000) Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones. Virology 277: 278 295.
4. Scala G, Ruocco MR, Ambrosino C, Mallardo M, Giordano V, et al. (1994) The expression of the interleukin 6 gene is induced by the human immunodeficiency virus 1 TAT protein. The Journal of experimental medicine 179: 961 971.
5. Weissman JD, Brown JA, Howcroft TK, Hwang J, Chawla A, et al. (1998) HIV-1 tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes. Proceedings of the National Academy of Sciences of the United States of America 95: 11601 11606.
6. Carroll IR, Wang J, Howcroft TK, Singer DS (1998) HIV Tat represses transcription of the beta 2-microglobulin promoter. Molecular immunology 35: 1171 1178.
7. Caldwell RL, Egan BS, Shepherd VL (2000) HIV-1 Tat represses transcription from the mannose receptor promoter. Journal of immunology 165: 7035 7041.
8. Kim N, Kukkonen S, Gupta S, Aldovini A (2010) Association of Tat with promoters of PTEN and PP2A subunits is key to transcriptional activation of apoptotic pathways in HIV-infected CD4+ T cells. PLoS pathogens 6.
9. Caputo A, Sodroski JG, Haseltine WA (1990) Constitutive expression of HIV-1 tat protein in human Jurkat T cells using a BK virus vector. Journal of acquired immune deficiency syndromes 3: 372 379.
10. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.
11. Shin H, Liu T, Manrai AK, Liu XS (2009) CEAS: cis-regulatory element annotation system. Bioinformatics 25: 2605 2606.
12. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome research 12: 996 1006.
13. Jurka J (2000) Repbase update: a database and an electronic journal of repetitive elements. Trends in genetics : TIG 16: 418 420.

14. Smit AF (1996) The origin of interspersed repeats in the human genome. Current opinion in genetics & development 6: 743 748.
15. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology 28: 495 501.
16. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4: 44 57.
17. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207 210.
18. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, et al. (2009) DNA specificity determinants associate with distinct transcription factor functions. PLoS genetics 5: e1000778.
19. Russell L, Garrett-Sinha LA (2010) Transcription factor Ets-1 in cytokine and chemokine gene regulation. Cytokine 51: 217 226.
20. Bories JC, Willerford DM, Grevin D, Davidson L, Camus A, et al. (1995) Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. Nature 377: 635 638.
21. Li JC, Lau AS (2007) A role for mitogen-activated protein kinase and Ets-1 in the induction of interleukin-10 transcription by human immunodeficiency virus-1 Tat. Immunology 121: 337 348.
22. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707 719.
23. Jang KL, Collins MK, Latchman DS (1992) The human immunodeficiency virus tat protein increases the transcription of human Alu repeated sequences by increasing the activity of the cellular transcription factor TFIIIC. Journal of acquired immune deficiency syndromes 5: 1142 1147.
24. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Molecular and cellular biology 18: 58 68.
25. Ferrari R, Pellegrini M, Horwitz GA, Xie W, Berk AJ, et al. (2008) Epigenetic reprogramming by adenovirus e1a. Science 321: 1086 1088.

Fig S1-Marban et al.



Chapter 7

Dynamics of Host Cell Transcriptome

in DNA Tumor Virus Infection

This chapter includes my recent findings that further examined molecular mechanisms involved in e1a reprogramming of stationary host cell into proliferative state. To elucidate genome-wide distribution of viral oncoproteins, e1a, on host cell chromatin, we mapped temporal-based binding of e1a by carrying out ChIP-seq experiments in IMR90 cells infected with Adenovirus *dl1500*. To characterize the impact of e1a-binding on host cell gene expression, global RNA transcript levels were measured by RNA sequencing method. This study reveals the genome-wide mapping of e1a binding patterns at 6 and 24 hour time point after viral infection. We found that through interactions with P300/CBP, binding of e1a at active enhancers and gene bodies of anti-proliferative genes results in strong transcriptional repression. In contrast, interactions between e1a and Rb at promoter of cell cycle genes induce strong transcriptional activation. Most surprisingly, we observed that replication origins defined by nascent strand DNA synthesis sequencing in normal cycling IMR90 cells are bound by e1a in infected cells. Furthermore, these e1a-bound regions overlaps better with sites of nascent DNA synthesis in highly proliferative cell lines (H9 and hela) compare to overlap with IMR90 cells.

Summary

Viral oncogene small e1a, a 243 amino acid protein, is capable of forcing contact-inhibited G1 cells into S phase. Through binding with P300/CBP and Rb family of pocket proteins, Rb, p107, and P300, e1a modulates global gene expression to reprogram host cells into replicative state. To understand the full extent of e1a's influence on rewiring the chromatin state of infected host cells, ChIP-seq and RNAseq was performed at 6 and 24 hours post infection to follow the progress of e1a induced cellular changes. We found that e1a targeted active enhancer regions enriched with H3K18ac and H3K27ac and gene bodies of activator proteins in the TGF-β pathway early in infection for the purpose of transcriptional repression. Promoter region of cell cycle genes bound by Rb is jointly targeted by e1a at 6 and 24hours post infection to relieve Rb-mediate transcriptional repression. Potential role of e1a participating in regulation of host cell DNA replication was observed as genomic regions with binding of e1a was found to be in common with replication origin sites in normal cycling cells.

Introduction

Adenovirus oncoproteins small e1a, when expressed alone in contact-inhibited fibroblast cells, is sufficient in forcing stationary cells into S phase (Howe, Mymryk et al. 1990; Ghosh and Harter 2003). This oncogenic process is accomplished through interactions between e1a and multiple key cellular proteins in the infected host cell including Rb family of pocket proteins (Rb, p107, p130), P300/CBP, and P400. Previous study showed that within 24 hours after infection, e1a binds to gene promoter regions and modulate host cell gene expression toward its replicative advantage

through inducing re-distribution of P300/CBP and Rb (Ferrari, Pellegrini et al. 2008). However, gene promoters only accounts for a small percentage of the genome. To examine whether e1a binds and reorganizes chromatin landscape for the rest of the human genome, we performed ChIP-seq assay to map genome-wide binding regions of e1a in contact-inhibited IMR90 cells infected with Adenovirus mutant *dl1500*. Adenovirus *dl1500* only expresses small e1a as the major viral protein product since large e1a is responsible for activating expression of rest of the Adenoviral proteins (Montell, Fisher et al. 1982). Contact-inhibited IMR90 cells infected with *dl1500* is characterized with global hypoacetylation of histone modifications H3K18 and H3K27 at active enhancers (Ferrari, Su et al. 2012), repression of growth inhibition genes (Ferrari, Gou et al. 2014), and activation of cell cycle genes (Frisch and Mymryk 2002; Ferrari, Pellegrini et al. 2008). With e1a binding mapped genome-wide in the same cell model system, we will examine the consequence of direct e1a binding at these important genomic regions.

Results

Promoter region is a common target of e1a binding at both 6 and 24 hours post infection

Previous studies have shown that small e1a is capable of forcing G1 contact-inhibited normal human lung fibroblast IMR90 cells into S phase through multiple interacting partners including P300, P400 and Rb family of proteins. Through its interaction with proteins bound on chromatin of host cells, e1a was able to remodel the epigenetic landscape of host cells at both promoter regions and active enhancer regions. However, it is still unclear whether changes in histone modifications at key

functional regions were resulted from direct binding of e1a or through secondary pathways. To map genome-wide distribution of small e1a binding after viral infection, high throughput ChIP-seq method was utilized in mock infected IMR90 cells and IMR90 cells infected with Adenovirus *dl1500*, which only expresses small e1a at low MOI (Ferrari, Pellegrini et al. 2008; Horwitz, Zhang et al. 2008). At 6 hours and 24 hours after dl1500 infection, samples were collected using previously described methods (Ferrari, Pellegrini et al. 2008; Ferrari, Su et al. 2012). 18 million and 21 million reads aligned to human genome version hg19 were randomly sampled from input reads and ChIPed DNA reads for both 6 hours and 24 hours post infection samples to normalize total read counts between two different time points. Human genome was segmented into 50 bp windows and ChIP enrichment P-value between input and ChIPed DNA reads was calculated using Poisson distribution as previous described (Ferrari, Su et al. 2012). Significant peaks were defined in 50 bp windows with P-value < 10^{-3} and requiring both neighboring windows to have P-value < $10^{-2.5}$. Using cis-regulatory element annotation system (CEAS) software (Shin, Liu et al. 2009), distribution pattern of e1a binding was determined and represented in pie charts (Figure 1A). Over 70% of e1a binding regions were located inside the intron and intergenic regions at both 6 hours and 24 hours PI. Only 9% and 12% of e1a binding events occurred near transcription start site (TSS) at 6 and 24 hours PI. As shown in Figure 1B, 15% of regions bound by e1a at 6 hours PI overlaps with 21.5% of e1a bound regions at 24 hours PI. Interestingly, genomic regions targeted by e1a at both 6 and 24 hours PI has a higher portion of regions located near TSS and lower portion of reads located inside intergenic regions when compared to e1a's binding pattern at individual time points. As

much as 26% of the shared binding sites between 6 and 24 hour e1a are located near TSS (Figure 1C) which suggested that e1a maintained strong interaction with host proteins bound near TSS especially since e1a lacks the ability to bind DNA directly (Frisch and Mymryk 2002). The dynamic binding of e1a at different time points after infection is reflected in the heat map of e1a binding patterns at 6 and 24 hours PI.

Binding of e1a leads to global deacetylation at active enhancers

Enhancers regions are characterized with high levels of H3K4Me1 and low levels of H3K4Me3. Additionally, active enhancers have high abundance of H3K27ac or H3K18ac which are both acetylated by presence of P300/CBP (Heintzman, Hon et al. 2009). Previous studies have shown that e1a caused global decrease of H3K18ac (Horwitz, Zhang et al. 2008). Most of the reduction in H3K18ac levels took place in the intergenic region (Ferrari, Su et al. 2012). However, it is unclear whether e1a directly binds to active enhancers to mediate reduction of H3K18ac levels. To define active enhancer regions in IMR90 cells, we analyzed previously- published H3K4Me1, H3K18ac, H3K27ac, and P300 ChIP-seq data from contact-inhibited IMR90 cells and dl1500 infected IMR90 cells (Ferrari, Su et al. 2012; Ferrari, Gou et al. 2014). Heatmap of significant peak intensity centered at e1a-bound regions was generated for all downloaded ChIP-seq data. Consequence of e1a binding at active enhancers was examined. Consistent with previous findings, majority of active enhancers in mock infected cells showed strong reduction in H3K18ac and H3K27ac levels after dl1500 infection and subsequent e1a binding (Figure 2A). Additionally, we found that active enhancers were early targets of e1a with more than twice as many regions being bound

at 6 hours PI compared to 24 hours PI by e1a (Figure 2A, Figure S1A). Interestingly, different sets of active enhancers were targeted by e1a at 6 and 24 hours PI, as only 295 regions are bound by e1a at both 6 and 24 hours PI (Figure 2B). Small set of active enhancer regions overlap between e1a binding at 6 and 24 hours implies e1a targeted different sets of active enhancer regions at different time point. Figure S2 plots average significant read counts of e1a binding at active enhancers during 6 and 24 hour PI. Interestingly, with temporal difference in e1a binding, the degree of hypoacetylation for H3K18 and H3K27 was similar at active enhancers bound. Overall, binding of e1a at active enhancer region led to drastic reduction of H3K18ac and H3K27ac (Figure 1A, S1B, S1D) while the level of P300 binding remained unchanged in both mock and dl1500 infected cells (Figure S1C).

Active enhancers bound by e1a leads to reduction of neighboring gene expression

To determine functional consequence of e1a induced H3K18 and H3K27 hypoacetylation at active enhancer regions, stranded RNA-seq was performed to measure level of strand-specific transcripts at 6 and 24 hours PI in both mock-infected and dl1500 infected IMR90 cells. RNAseq data were aligned to human genome version hg19 using Tophat2 v2.0.8 (Kim, Pertea et al. 2013) and read counts were normalized among different samples using Cufflinks v2.0.2 (Trapnell, Williams et al. 2010). To associate e1a binding regions at active enhancers to single closest gene within 1000 kb, Genomic Region Annotation Tool (GREAT) (McLean, Bristor et al. 2010) was utilized for neighboring gene assignment. Figuer 2C shows boxplot of log2 of

normalized transcript levels for genes neighboring active enhancers in both mock and dl1500 infected cells at 6 and 24 hours PI. Genes that maintained H3K18ac and H3K27ac levels at active enhancers in mock infected cells did not have significant changes in transcript levels. However, e1a binding at active enhancers and subsequent H3K18 and H3K27 hypoacetylation caused significantly reduced transcription levels of neighboring genes when compare to same gene set in mock infected cells (Figure 2C). Gene ontology (GO) of genes neighboring e1a-bound active enhancers were determined using GREAT (McLean, Bristor et al. 2010). Gene ontology analysis indicated that genes associated with hypoacetylated enhancers were involved in fibroblastic cell type specific cellular processes such as production and maintenance of extracellular matrix and collagen (Figure 2D). An genome browser example of e1a binding at active enhancer leading to hypoacetylation of H3K18ac and H3K27ac and reduced transcription of neighboring gene, COL6A3, is shown in Figure S3.

Multiple binding of e1a throughout gene body mediates strong gene repression

Recent study by Ferrari et al. 2014 (Ferrari, Gou et al. 2014) revealed that in e1a expressing IMR90 cells, genes with high levels of P300 and Rb binding throughout the gene body resulted in strong repression of gene expression. Additionally, through confocal microscopy of e1a- LacI-mCherry DNA array, e1a was observed to induce chromatin condensation through interaction with Rb and P300. Furthermore, the chromatin condensation process required HDAC activity as HDAC inhibitor prevented e1a-mediated chromatin condensation. In the same study, a list of glycoprotein genes (cluster rc1) highly repressed by wild type e1a but not by P300 binding mutant of e1a

was found to contain high levels of Rb and P300 binding throughout the genic region. However, it was unclear whether e1a participated in the recruitment of Rb and P300 binding at these highly repressed genes through direct interaction with Rb and P300. From analyzing e1a ChIP-seq at 6 and 24 hours PI, Figure 3A shows that rc1 genes are enriched with very high frequency of e1a binding. The bar graph in Figure 3A also illustrated that e1a binding at 6 hours is the main driver of e1a's effort in orchestrating strong repression at rc1 genes as more binding peaks from 6 hour e1a were found at gene bodies of rc1 genes in comparison to 24 hour e1a binding. Genes in rc1 cluster play essential roles in TGF-β pathway to regulate G1/S transition. Gene bodies of TGFB2 and SMAD3 in the rc1 cluster contained over 10 e1a binding sites. Boxplot of RNA-seq measurement of transcript levels (log2 of FPKM) from RC1 genes bound by e1a (Figure 3B) showed that RC1 genes were expressed at significantly higher levels in mock infected cells compare to average genes (p-value = 0.000131). While transcription levels did not change significantly at 6 and 24 hours after mock infection (p-value=0.99). Expression of e1a in IMR90 cells and binding to rc1 cluster genes caused significant reduction in gene expression levels at 24 hours PI (p-value= 0.004445) (Figure 3B). Genome browser example of e1a binding throughout THBS1 gene from rc1 gene cluster is shown in Figure 3C. Binding of e1a on THBS1 gene was primarily detected at 6 hours PI suggesting THBS1 as an early target of e1a-mediated repression. High levels of P300 binding detected at 24 hours PI is associated with drastic reduction in Pol II binding throughout the transcribed region which correlated with radical decrease of transcript levels at 24 hours PI detected through RNAseq (Figure 3C).

Promoter binding of e1a at cell cycle genes leads to gain of H3K18ac

To examine e1a binding pattern near gene promoters, ChIP-seq binding of e1a at 6 and 24 hours PI was separated into 3 clusters according to levels of H3K18ac enrichment at +/- 3kb around the TSS. In the heat map (Figure 4A), cluster k1 was assigned with promoters having more than 50 percent gain in H3K18ac levels at TSS in infected versus mock cells while cluster k2 was assigned with genes deacetylated at least 50 percent of H3K18ac at TSS in infected vs mock cells. The rest of the genes bound by e1a around TSS were assigned to cluster k2. Clusters k1, k2, and k3 are genes with promoters bound by e1a at either 6 or 24 hours PI. ChIP-seq data of Rb, P107, P130 in mock infected IMR90 cells and H3K18ac in mock and dl1500 infected IMR90 cells were downloaded and analyzed from previous study (Ferrari, Su et al. 2012). ChIP-seq data of P300 and H3K27ac was downloaded and analyzed from recent publication on P300 mediated repression in e1a expressing IMR90 cells (Ferrari, Gou et al. 2014). Figure S4 displays the average significant counts profile of Rb, p107, p130 binding at k1, k2, and k3 gene promoters. The plots confirmed findings from previous study as gene clusters gaining H3K18ac (cluster k1) in infected cells contained highest level of Rb binding in mock cells (Ferrari, Su et al. 2012) while differential binding for p107 and p130 was not observed between different clusters. In accordance with clustering method, Figure S5 showed that H3K18ac level in k2 cluster was highest among all three clusters in mock cells. In infected cells, cluster k1 became the cluster with highest level of H3K18ac while H3K18ac in cluster k2 as highly deacetylated. As acetylation of H3K18 was known to correlate with activation of transcription, transcription of k2 gene clusters decreased significantly (p-value < 2.2e-16) compared

to k3 gene cluster while transcription of k1 gene clusters were up-regulated when comparing to k3 gene cluster as shown in Figure 4B where FPKM ratios at 24 hours PI between infected cells over mock infected cells were compared. Interestingly, while H3K18ac and H3K27ac were all known to be acetylated by P300/CBP, differential levels of H3K18ac and H3K27ac was observed at promoters bound by e1a.

Differential regulation of H3K18ac and H3K27ac at e1a-bound promoters

In Figure S6, all promoters bound by e1a had high levels of H3K27ac in mock cells. With highest level of H3K18ac in mock cells, cluster k2 also has highest level of H3K27ac in mock cells (Figures S6). However, after dl1500 infection, all three clusters were hypoacetylated at H3K27ac while differential levels of H3K18ac between k1, k2, and k3 was observed (Figure S5, Figure S6). To further investigate whether this differential regulation between H3K18ac and H3K27ac was caused by differential binding of P300, histone acetyltransferase for H3K18 and H3K27, average binding of P300 for k1, k2, and k3 was plotted (Figure S7). Interestingly, all three clusters gained P300 binding which was unable to explain the differential regulation of H3K18ac and H3K27ac at e1a-bound gene promoters. The molecular mechanism behind this differential regulation of H3K18ac and H3K27ac at e1a-bound gene promoters is currently unclear. Using GREAT software for gene ontology enrichment analysis, E2F motif and cell cycle function was observed for cluster k1(Figure S8) which gained H3K18ac and increased gene expression at 24 hours PI (Figure 4A, Figure 4B). For genes repressed at 24 hours, activators of TGF-beta pathway and genes highly expressed in fibroblast cell types were observed in cluster k2 (Figure S8). Genes in k3 cluster were expressed before and after infection. This gene cluster was enriched with

genes with roles in translation and metabolic process which was not inhibited by e1a binding. With requirement to translate newly-transcribed cell cycle genes, uninterrupted translation and metabolic process may be critical in promoting S phase entry.

Genome browser view of e1a binding peaks at TSS of cell cycle gene, CCNE2, co-localized with binding peaks of Rb, p107, and p130 in mock infected cells. At 24 hours PI, gain of P300 and Pol II was observed at CCNE2 promoter which aligned with Rb binding region in mock infected cells and in close proximity with e1a binding peaks at both 6 and 24 hours PI (Figure S9A).

Genomic regions bound by e1a are sites of nascent strand DNA synthesis in asynchronous IMR90 cells

By forcing contact-inhibited G1 cells into S phase entry, the primary function of e1a is to create the proper cellular environment for viral replication. By comparing e1a binding regions to recent nascent strand (NS) DNA synthesis sequencing data conducted in asynchronous IMR90 cells (Besnard, Babled et al. 2012), we found that e1a binding regions contained significant overlap with sites of NS DNA synthesis as shown in Figure 5A. Furthermore, e1a binding sites overlap with sites of higher levels of NS DNA synthesis activity as shown in boxplot of Figure 5A. Heatmap centered at genome-wide e1a binding sites in Figure 5B was generated through analysis from multiple ChIP-seq experiments from previous studies (Chicas, Wang et al. 2010; Besnard, Babled et al. 2012; Soufi, Donahue et al. 2012; Ferrari, Gou et al. 2014) and clustered by binding of e1a at 6 and 24 hours PI. The order within each cluster was sorted based on total sum of significant counts from NS DNA synthesis sequencing experiment in asynchronous IMR90 cells. Binding regions of e1a at both 6 and 24

hours PI are significantly enriched with sites of NS DNA synthesis when compared to genome background (Figure S10B, Figure S11B).

Myc binds at sites of NS DNA synthesis

The purpose of NS DNA synthesis sequencing study was to map sites of replication origin genome-wide (Besnard, Babled et al. 2012). Interestingly, Myc was found to bind at replication origin and over-expression of Myc increased the density of replication origins (Dominguez-Sola, Ying et al. 2007). Furthermore, e1a was found to stabilize Myc protein through binding with P400 (Tworkowski, Chakraborty et al. 2008). To examine whether e1a shares any global associations with Myc, ChIP-seq data of Myc and over-expressed Myc in BJ fibroblast cells was downloaded and analyzed (Soufi, Donahue et al. 2012). Myc and nascent strand DNA synthesis sites are both enriched at e1a binding sites in 6 and 24 hours PI (Figure 5A, Figure S10E, Figure S11E). Additionally, Myc binding levels are also significantly enriched over background at e1a binding sites for both normal and over expressed Myc in BJ fibroblast cells (Figure S10CD, Figure S11CD). Myc-bound replication origins are known to be early-firing replication origins (Srinivasan, Dominguez-Sola et al. 2013). By comparing binding patterns of NS DNA in IMR90 cells and Myc binding in BJ fibroblast cells, Figure S10 and Figure S11 suggests sites of NS DNA synthesis in e1a-bound regions are associated with early-firing replication origins. In contrast, since Lamin B1 is known to associate with replication in late S phase (Moir, Montag-Lowy et al. 1994) and Lamin B1 ChIP-seq data of Lamin B1 in IMR90 cells was analyzed (Sadaie, Salama et al. 2013). Figure 5B shows that very low levels of Lamin B1 binding in IMR90 cells overlap with

NS DNA synthesis sites in growing IMR90 cells and e1a binding sites in mock and infected cells, which further suggests e1a binding overlaps with early replication origins.

Binding of e1a overlap with sites of nascent DNA synthesis in highly proliferative cell lines

To compare e1a binding site to sites of NS DNA synthesis in other cell lines, NS DNA synthesis sequencing data for hela and H9 human stem cell line was downloaded and analyzed (Becker, Ghule et al. 2006). Unexpectedly, higher overlap between NS DNA synthesis sites from highly proliferative cell lines (hela, H9) and e1a binding sites in IMR90 was observed (Figure 5B). NS DNA synthesis sequencing results form hela and H9 cell lines also have higher percentage of overlap with e1a-bound regions compared to NS DNA synthesis experiment conducted in IMR90 cells (Figure S12). Since nascent DNA synthesis maps replication origins, our analysis suggests e1a binds to genomic regions that are potentially replication origins in multiple cell lines.

Host replication is inhibited in e1a expressing cells to promote viral replication

To determine the effect of e1a expression on host cell replication, BrdU-seq was performed by pulse-labeling newly synthesized DNA with BrdU and followed by high-throughput sequencing. In normal asynchronous IMR90 cells, all sequencing data from BrdU-seq aligned to human genome version hg19. In contrast, BrdU-seq data collected from e1a-expressing cells at 24 hours post infection, only 5% of the data was aligned to human genome version hg19 and 95% of the data only aligned to adenoviral Ad5 genome (Figure S13). Since e1a binds to replication origins of normal cycling cells, e1a binding potentially inhibits proper replication of host genome to promote replication of viral genome.

Discussion

Through binding with P300/CBP, e1a was known to induce reduction of global H3K18ac and H3K27ac genome-wide (Ferrari, Su et al. ; Ferrari, Pellegrini et al. 2008; Horwitz, Zhang et al. 2008; Ferrari, Gou et al. 2014). Using ChIP-seq method, the genome-wide binding patterns of e1a at 6 and 24 hours PI were examined for the first time. We now provide evidence of direct e1a binding at active enhancers, cell cycle gene promoters, gene body of highly transcribing genes, and replication origins in IMR90 cells. All of which are regions in the genome with essential and critical functions. With an ultimate goal of promoting viral replication, e1a reprograms the host cell environment to promote viral replication.

In contact-inhibited cells, highly-transcribed cell-type specific genes such as collagen and extracellular matrix genes in fibroblast cells maintained elevated levels of transcription through interaction between activator proteins at active enhancers and transcriptional machinery at gene promoters. Active enhancers contain histone modifications H3K18ac and H3K27ac which maintains open chromatin structure at active enhancers. Through binding to P300/CBP located in active enhancers, e1a aimed for inhibition of P300/CBP activity as early as 6 hours PI. At 24 hours PI, histone acetylation levels at H3K18 and H3K27 were greatly reduced along with strong reduction of transcription in neighboring genes. Since active enhancers are enriched with more e1a binding at 6 hours PI compared to 24 hours PI, e1a might have higher affinity toward P300/CBP with active histone acetyl-transferase activity.

Interestingly, genes highly repressed by e1a binding throughout the gene body did not initially have high levels of P300/CBP in the gene body. This set of highly repressed genes at 24 hours PI, gene set rc1 from (Ferrari, Gou et al. 2014), also attracted high frequency of e1a binding at 6 and 24 hours PI. Once again, more e1a binding was observed at 6 hours compared to 24 hours PI. Highly levels H3K18ac and H3K27ac were observed at promoter of this set of genes which suggest e1a at 6 hours target highly active P300/CBP for transcriptional repression through inducing hyopacetylation of H3K18ac and H3K27ac. Formation of P300-e1a-Rb was suggested to mediate repression at rc1 gene clusters as Rb was shown to associate with HDAC1 and condense chromatin into closed conformation (Ferrari, Gou et al. 2014). However, it is unclear as to what role does e1a binding at 24 hours PI plays in further recruitment of high levels of P300 proteins to the gene body of rc1 gene cluster at 24 hours PI as levels of e1a binding at rc1 cluster decreased at 24 hours PI in comparison to 6 hour e1a bindings.

To induce expression of cell cycle genes, k1 clusters in Figure 4A gained high levels of H3K18ac in infected cells at 24 hours PI. With high level of Rb binding at k1 gene clusters in mock cells and high affinity binding between LXCXE motif in CR2 region of e1a and Rb (Ikeda and Nevins 1993), e1a potentially was attracted to promoter region early in infection starting at 6 hours PI and continued at 24 hours PI. Interestingly, H3K18ac and H3K27ac were regulated differently at e1a-bound promoters. While all e1a-bound promoters gained p300 binding, only cluster k1 gained H3K18ac while H3K18ac at k2 promoters was deacetylated. Additionally, H3K27ac was highly deacetylated at all e1a-bound promoters. The underlying mechanism that

caused this differential regulation in histone acetylation at e1a-bound promoter remains unclear.

Interaction between e1a and P400 is essential for e1a mediated cell transformation (Fuchs, Gerber et al. 2001). Recent study discovered e1a is capable of stabilizing Myc protein through P400 binding and over expression of Myc was able to rescue P400 binding mutant of e1a (Tworkowski, Chakraborty et al. 2008). Furthermore, Myc binds at replication origin and Myc induction increases density of replication origins (Dominguez-Sola, Ying et al. 2007). By comparing to NS DNA synthesis sequencing data in growing IMR90 cells, we found that genomic regions bound by e1a at both 6 and 24 hours correlates with sites of NS synthesis and Myc binding in BJ fibroblast. Comparing e1a binding with NS DNA synthesis sequencing from hela and H9 cell lines suggest e1a-bound region overlap with sites of DNA synthesis in highly proliferative cell lines (H9 and Hela). Since e1a is not known to contain DNA-binding domain and interact directly with DNA, e1a potentially interact with protein complexes bound at pre-determined replication origins. Through binding at replication origin, e1a prevents replication of host genome and promotes viral replication in host cell S phase environment.

Figures

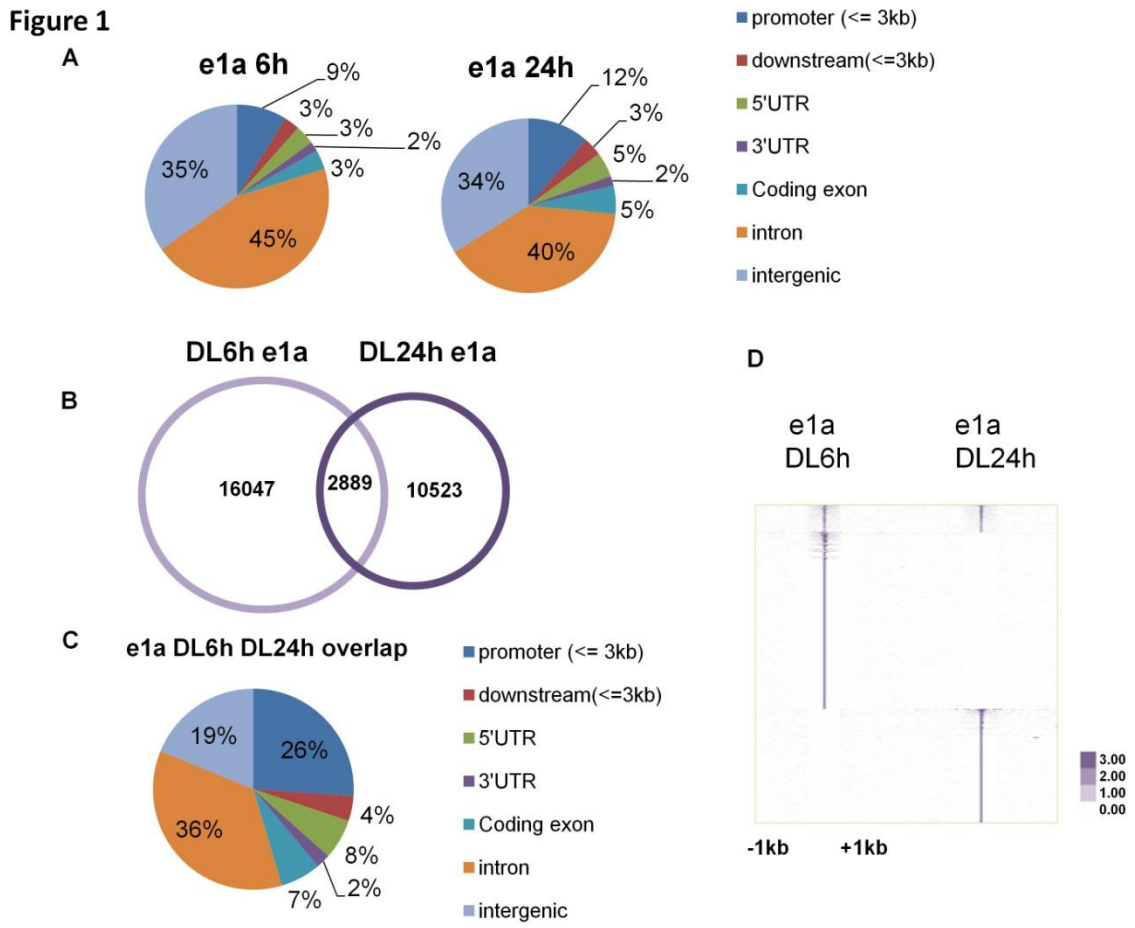


Figure 1. Gene promoters are persistent target regions for e1a at 6 and 24 hours PI. (A) Pie chart representation of genome-wide distributions of e1a peaks at 6h and 24h PI. Genomic regions at least 3 kb away from start and end of genes were defined as intergenic regions. (B) Venn diagram of overlapping regions between e1a bound at 6 and 24 hours PI (C) Genomic distribution of regions bound by e1a at both 6 and 24 hours PI (D) Heat map of e1a binding peak intensity centered at binding regions of e1a at 6 and 24 hours PI.

Figure 2

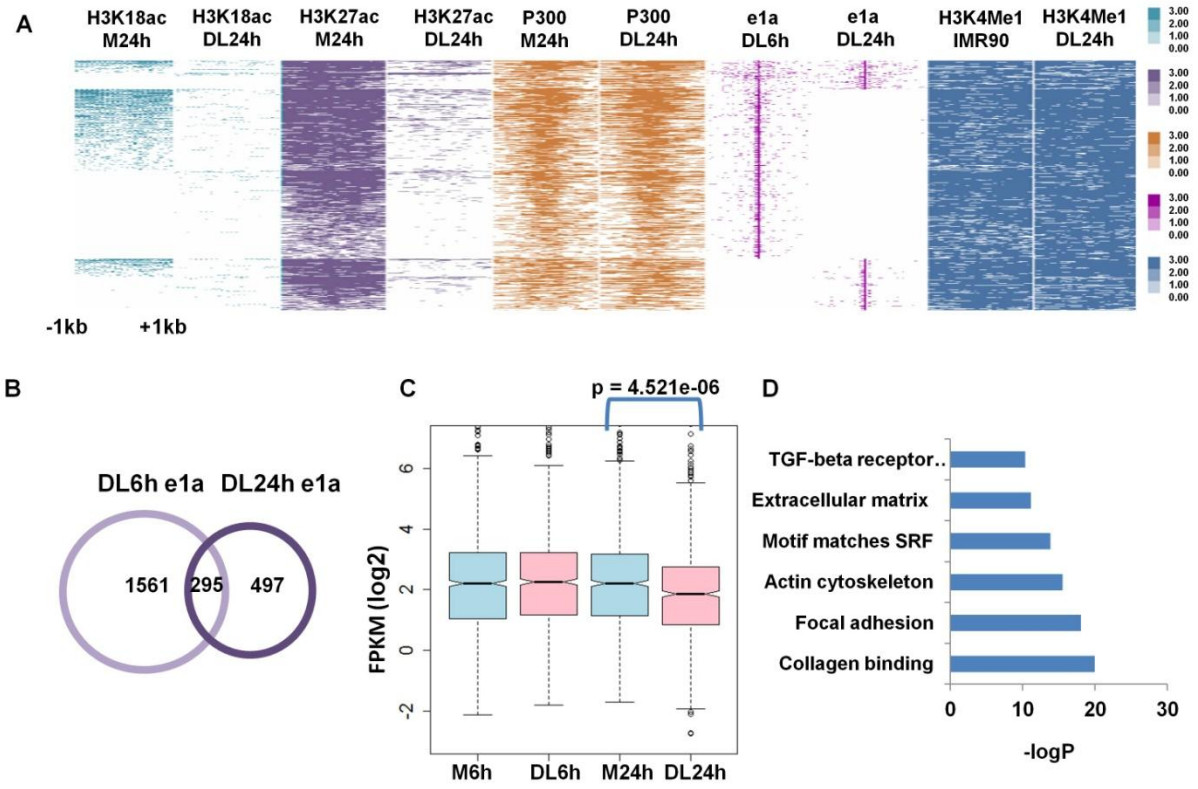


Figure 2. Active enhancers are bound by e1a at 6 hours PI. (A) Heatmap of P300, e1a binding distributions at active enhancer with histone modifictaions H3K4Me1, H3K18ac,and H3K27ac plotted across +/- 1kb from e1a binding regions. (B)Venn diagram overlap between e1a-bound active enhancer regions at 6 hours and 24 hours post infection (C) Boxplot of log2 of normalized transcript read counts (FPKM) of genes neighboring active enhancer regions bound by e1a. (D) Gene ontology from GREAT of genes neighboring active enhancers bound by e1a

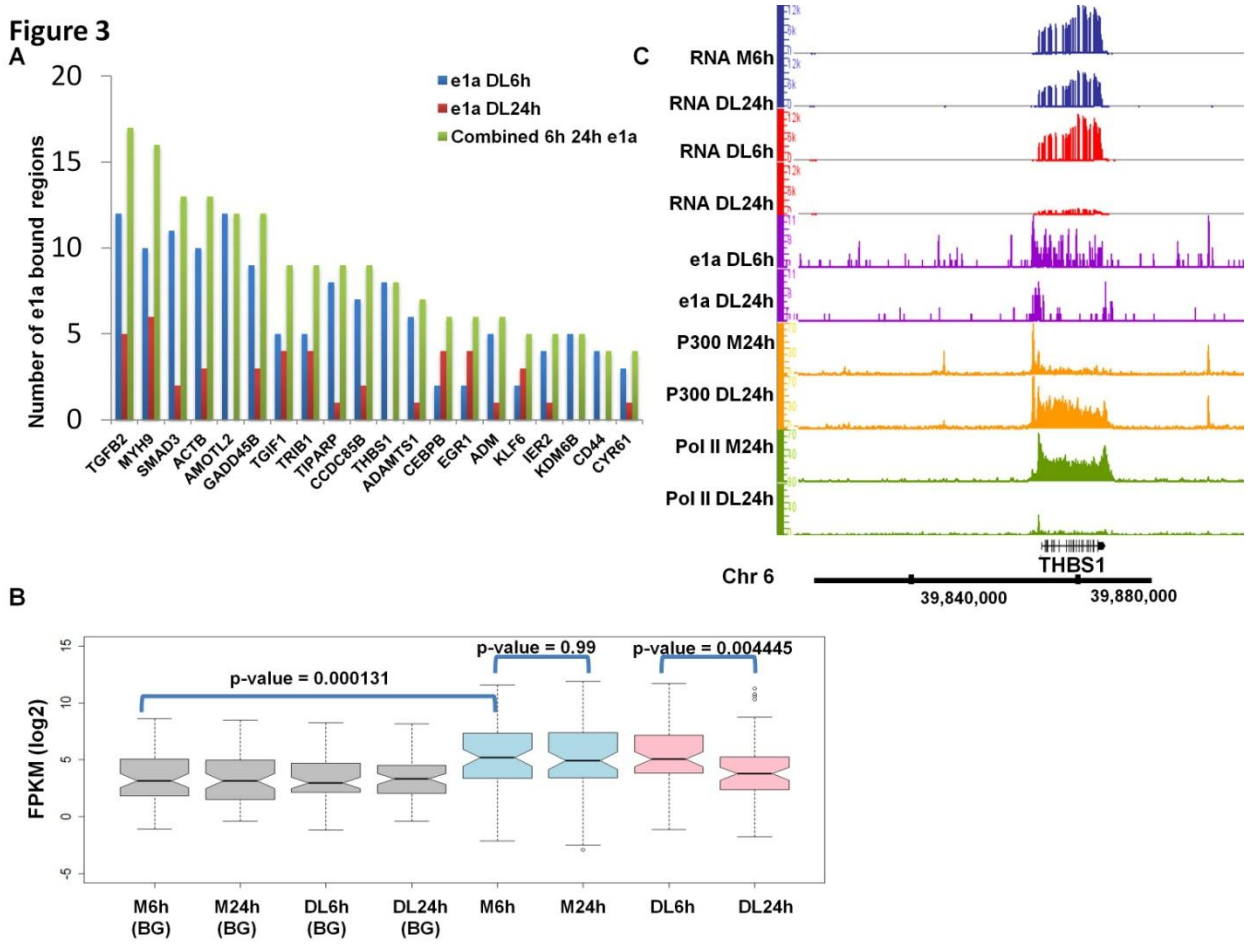


Figure 3. Highly represses genes are tiled by e1a binding. (A) Bar graph of top 20 highly repressed genes ranked by total number of regions bound by e1a at 6 and 24 hours PI. (B) Boxplot of (-log2 FPKM) of e1a-bound repressed gene set. (C) Genome browser view of ChIP-seq and RNAseq result showing Pol II, P300, and e1a binding at different time points in mock and infected IMR90 cells.

Figure 4

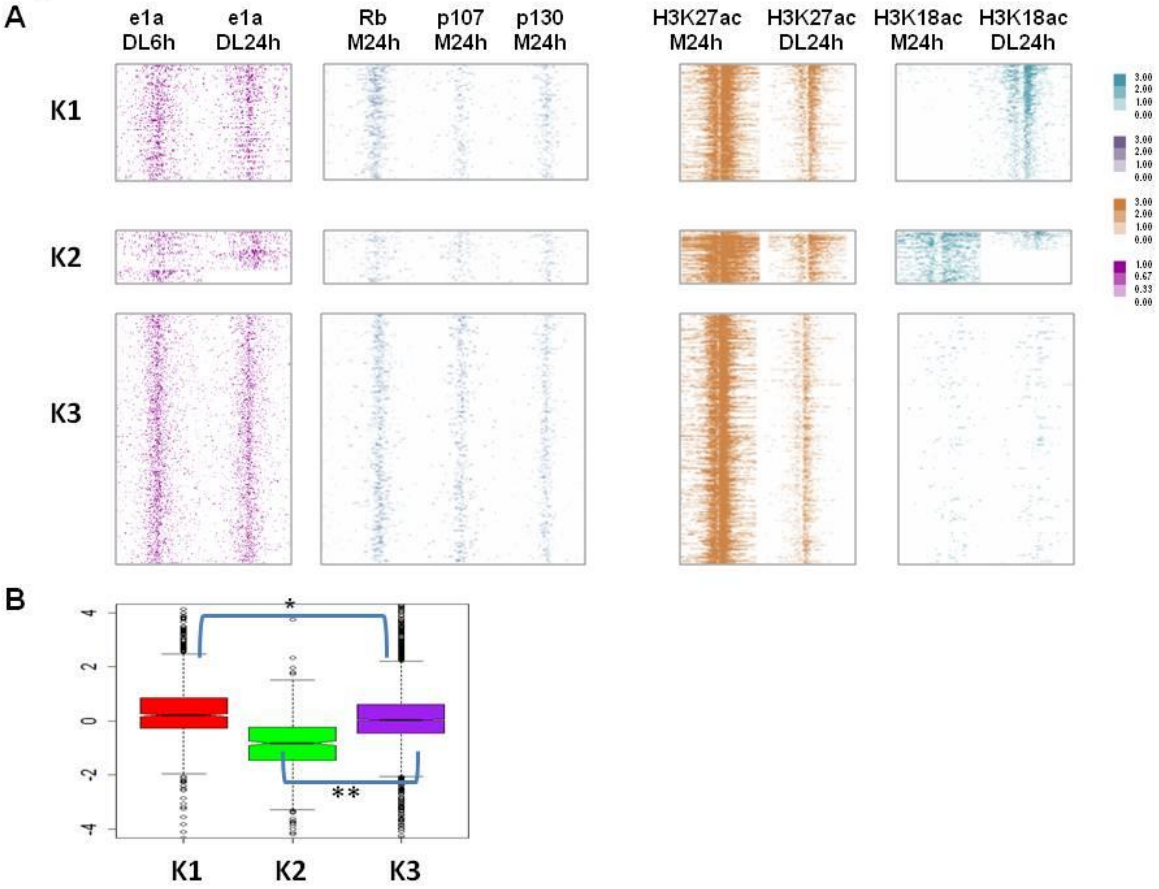


Figure 4. e1a binding at Rb-bound gene promoters mediates differential gain of H3K18ac and loss of H3K27ac at gene promoters. (A) Heat map of e1a, p300, H3k27ac, H3K18ac ChIP-seq results in mock and dl1500 infected cells and Rb, p107,p130 ChIP seq results from mock infected cells. Plot with significant counts on the y-axis is centered at transcription start site (TSS) with +/- 3KB tiled in 100 bp windows (B) Boxplot of log2 ratio of FPKM in infected versus mock cells at 24 hours PI for all gene clusters. * denotes P-value = 1.251e-08 and ** denotes P-value < 2.2e-16

Figure 5

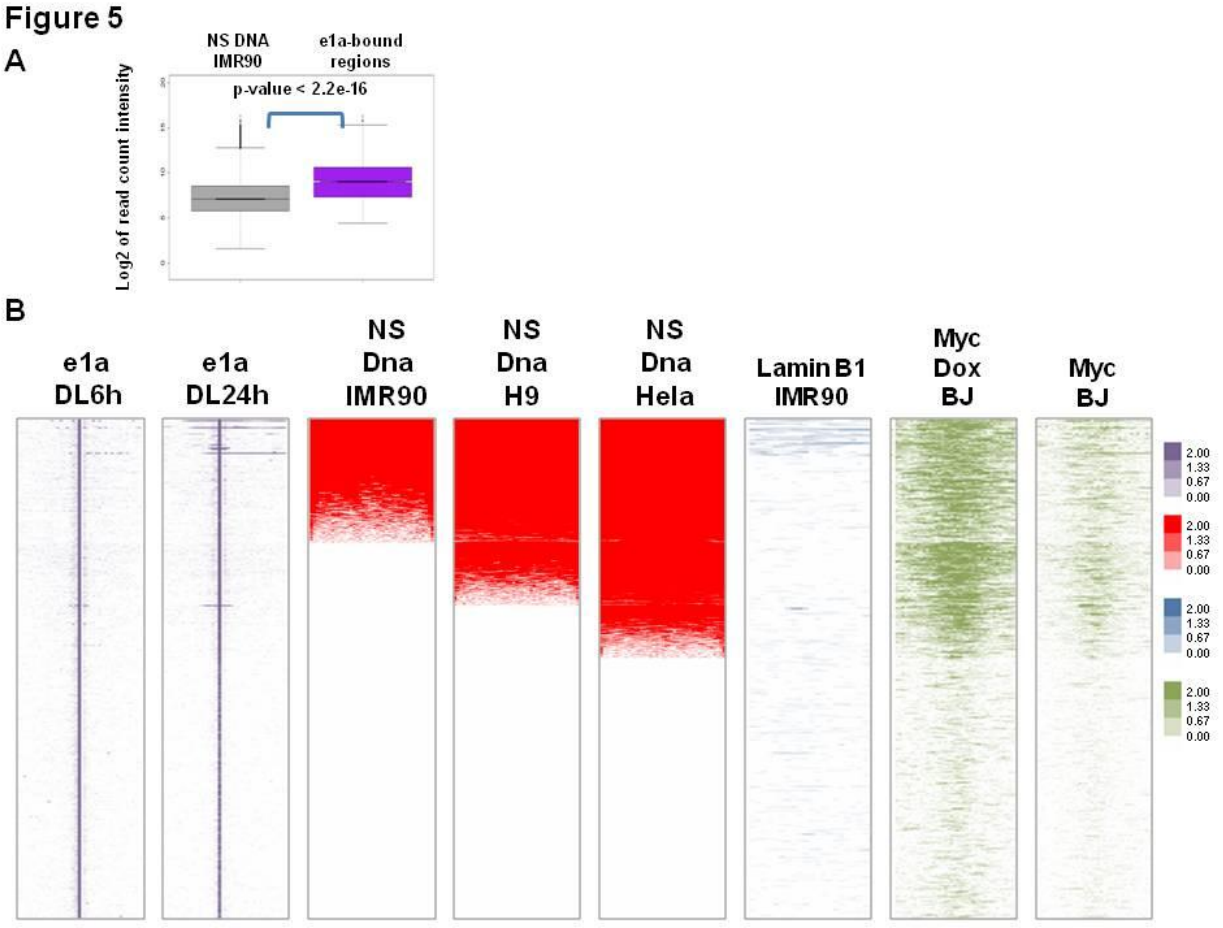


Figure 5. e1a-bound genomic regions are associated with sites of nascent strand (NS) DNA synthesis in asynchronous IMR90 cells. (A) Boxplot of log2 read counts of nascent strand sequencing in IMR90 versus log2 recounts from e1a-bound regions. (B) Heatmap of ChIP-seq data for e1a in *dl1500* infected IMR90 cells at 6 and 24 hours PI. Data from nascent strand DNA sequencing is from IMR90, hela, and H9 cells. ChIP-seq data for Myc is from BJ fibroblast cells under normal or overexpressed (Myc-dox) condition. ChIP-seq data for LaminB is from IMR90cells. Each row in the heat map represents +/- 1kb plotted around e1a binding region at the center.

Material and Methods

Cell culture

IMR90, human primary lung embryo fibroblasts from ATCC, was grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 100 U/mL penicillin, 100 µg/mL streptomycin, and 10% fetal bovine serum (FBS) at 37°C in 5% CO_2. Propagation of the *dl*1500 virus was done as described in previous study (Ferrari, Su et al. 2012).

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq)

Chromatin immunoprecipitation for ChIP-seq was performed as described in (Ferrari, Su et al. 2012) with modifications. In short, IMR90 cells were grown to confluence in 150 mm tissue culture plates and after 24 hours of contact inhibition, cells were infected with Adenovirus *dl1500* or mock infected (no virus infection) in DMEM with 100 U/mL penicillin, 100 µg/mL streptomycin, and 2% fetal bovine serum (FBS). Cells were collected at 6 and 24 hours after infection. Cells were fixed in plate with formaldehyde at 1% final concentration for 10 min at 37°C. Cell pellet was collected by cell scraper after PBS washing containing protease inhibitors (Roche). Cells were resuspended in 400 µl of lysis buffer and incubated for 10 min on ice. Pellet spun down at max speed was resuspended in 400 microL of lysis buffer and sonicated using Misonix cup-horn sonicator. 100 µl of the lysate was used for each immunoprecipitation with specific antibody (rabbit anti-e1a antibody, gift from Arnold Berk lab; rabbit anti-P400 antibody, gift from Mike Carey lab; Mouse anti-Pol II antibody, MMS-126R(Covance)). 10 µl of the lysate were used as input. Overnight incubation at 65°C was performed to reverse crosslinking. RNAseA incubation for 1 hour at 37°C followed by 2 hours of proteinase K digestion for 2 hours at 56 °C. Immunoprecipitated DNA was purified using

phenol/chloroform extraction and precipitated using EtOH and glycogen. Final DNA concentration was measured using Qubit (Invitrogen), and 1 ng of dsDNA for both input and IP were used for library construction according to the manufacturer's instructions for multiplex library preparation kit (Kapa biosystems) with index adapters from BIOO Scientific . Libraries were sequenced using Illumina HiSeq 2000 to obtain 50-bp-long reads.

Stranded total RNA-seq

Total RNA was extracted from mock- and *dl*1500-infected cells using Trizol (Invitrogen). 1 microgram of total RNA was used to start the library preparation according to the manufacturer's instructions (Illumina stranded total RNAseq library preparation kit). Libraries were sequenced using Illumina Hiseq 2000 to obtain 50-bp-long reads. Alignment of stranded totalRNA-seq reads was performed using TopHat 2 with –g 1 option (Kim, Pertea et al. 2013). Aligned reads were converted to sam format, and cufflinks software (Trapnell, Williams et al. 2010)was used to determine the transcript FPKM (fragments per kilo bases of exons for per million mapped reads).

ChIP-seq Analysis

All sequencing data from ChIP-seq and RNA-seq were mapped to human genome version hg19. ChIP-seq data was mapped to human genome using Bowtie software using parameters –m 1 to ensure each read only aligns to single location in the genome. Genome was split into 50 bp windows for e1a and Pol II ChIP-seq. Windows with P-value less than 10^{-3} was determined to be significant only if its two immediate neighboring windows were both with P-values less than 10^{-3}. P-value was calculated

using Poisson distribution and custom Matlab script. Two sample Kolmogorov–Smirnov test was used to calculate P-value in boxplots using R software.

Data source for ChIP-seq data

ChIP-seq data on Rb binding in growing or senescent IMR90 cells from previous study (Chicas, Wang et al. 2010) was downloaded from NCBI GEO (GSE19898). ChIP-seq data on Rb, p107, p130 in mock infected IMR90 cells and H3K18ac in mock and dl1500 infected IMR90 cells from previous study (Ferrari, Su et al. 2012) was downloaded from NCBI GEO (GSE32340). ChIP-seq data on H3K27ac, Rb and P300 in mock and dl1500 infected IMR90 cells from previous study (Ferrari, Gou et al. 2014) was downloaded from NCBI GEO (GSE59680). ChIP-seq data on Myc and Myc-dox from previous study (Soufi, Donahue et al. 2012) was downloaded from NCBI GEO (GSE36570). ChIP-seq data on H3K4Me3 and H3K4Me1 in IMR90 cells was downloaded from NCBI GEO (GSE16256). ChIP-seq data on LaminB1 in IMR90 cells from previous study (Sadaie, Salama et al. 2013) was downloaded from NCBI GEO (GSE49341).

Data source for nascent strand DNAseq

Nascent strand DNA sequencing data for IMR90, H9, and hela cells from previous study (Besnard, Babled et al. 2012) was download from NCBI GEO (GSE37757).

Figure S1

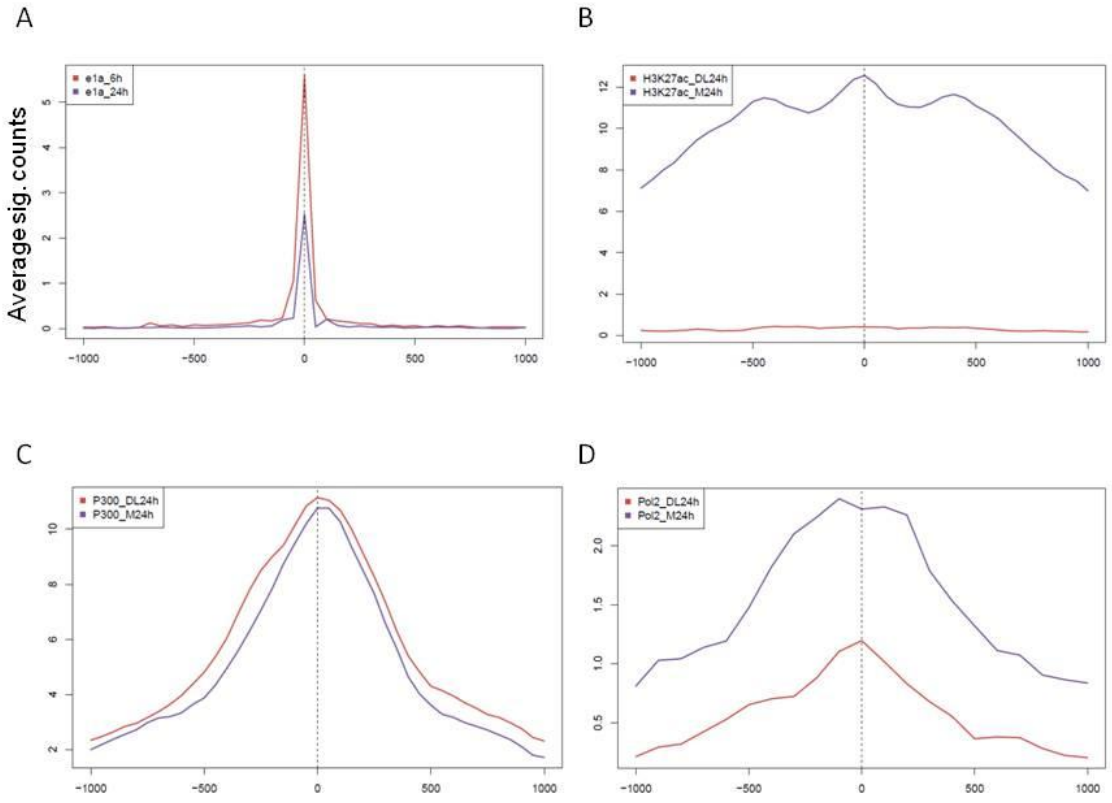


Figure S1. Active enhancer is targeted by e1a at 6 hours PI. (A) Plot of average significant counts from ChIP-seq of e1a binding at 6 and 24 hours PI centered at all e1a binding sites located within active enhancers. (B) Plot of average significant counts from H3K27ac ChIP-seq in mock and infected cells at 24 hours PI centered at all e1a-bound active enhancers. (C) Plot of average significant counts from P300 ChIP-seq in mock and infected cells at 24 hours PI centered at all e1a-bound active enhancers. (D) Plot of average significant counts from Polymerase II ChIP-seq in mock and infected cells at 24 hours PI centered at all e1a-bound active enhancers. Analysis plots were generated by CEAS software.

Figure S2

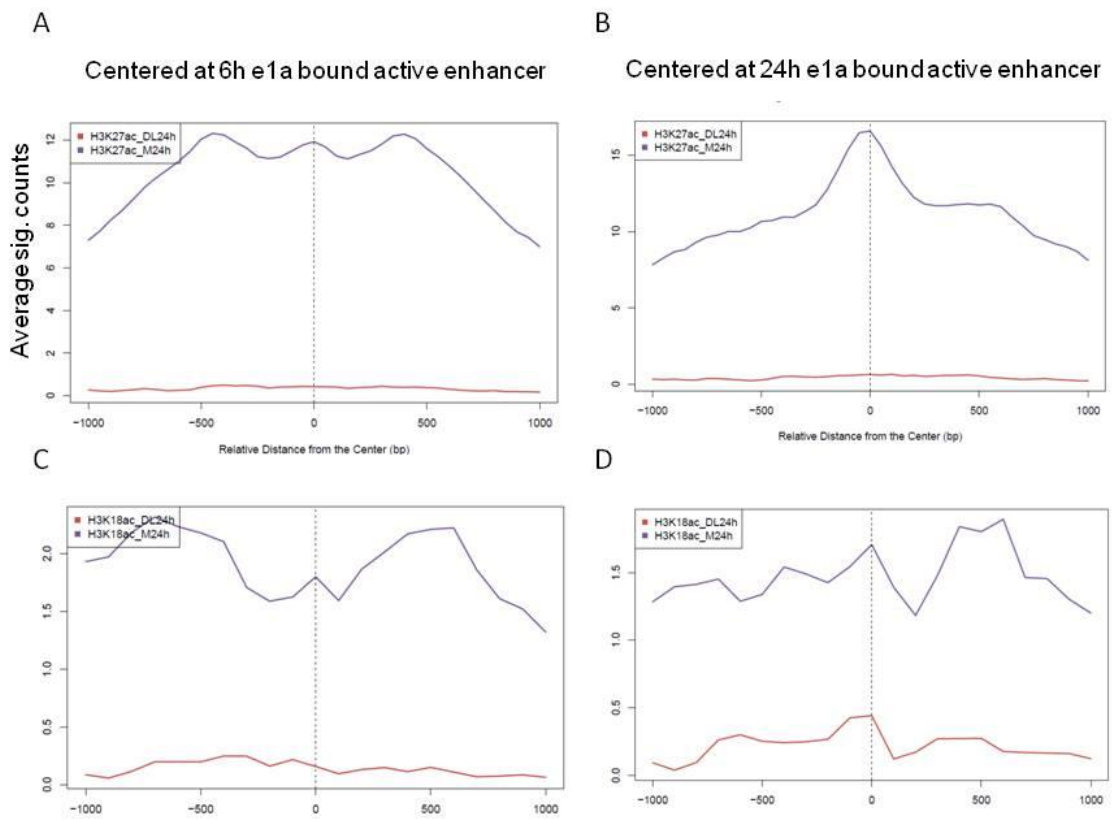


Figure S2 Deacetylation of H3K18ac and H3K27ac at e1a-bound active enhancers occurs at 24 hours PI. (A) Plot of average significant counts from H3K27ac ChIP-seq in mock and infected cells centered at regions bound by e1a at 6 hours. (B) Plot of average significant counts from H3K27ac ChIP-seq in mock and infected cells centered at regions bound by e1a at 24 hours PI. (C) Plot of average significant counts from H3K18ac ChIP-seq in mock and infected cells centered at regions bound by e1a at 6 hours PI. (D) Plot of average significant counts from H3K18ac ChIP-seq in mock and infected cells centered at regions bound by e1a at 24 hours PI. Analysis plots were generated by CEAS software.

Figure S3

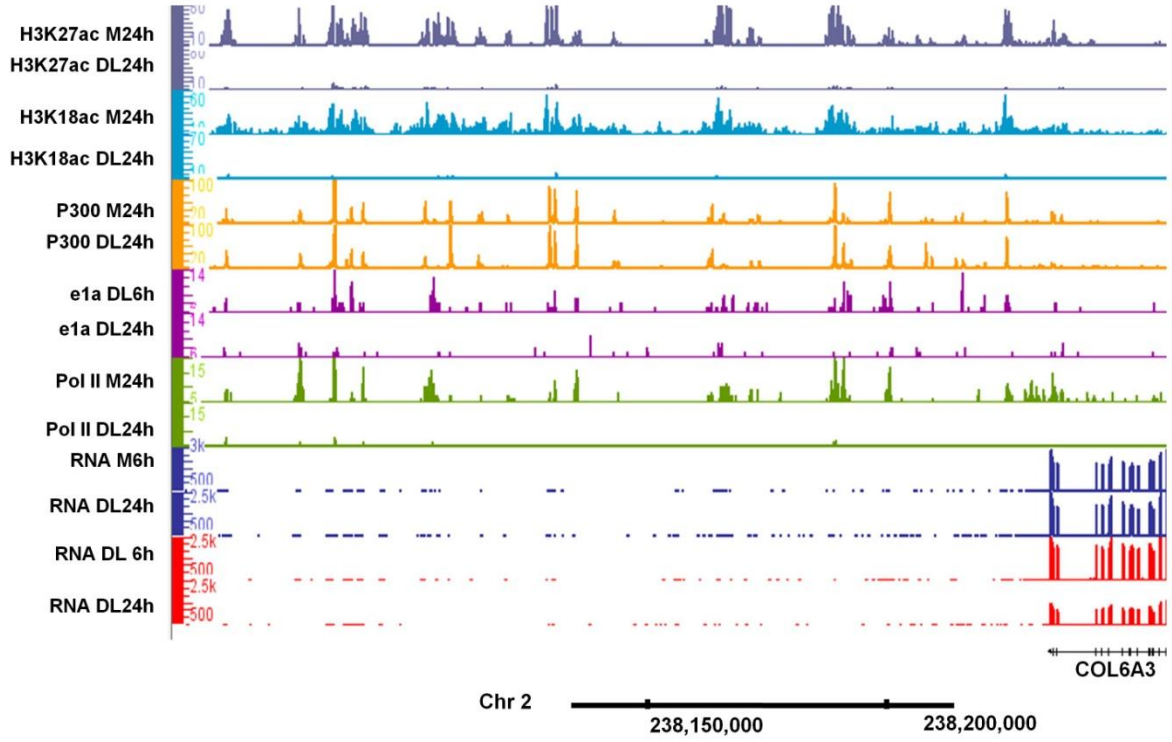


Figure S3. Active enhancers neighboring COL6A3 are bound by e1a at 6 hours PI. Intergenic regions with deacetylation of H3K18ac and H3K27ac coincide with e1a and P300 binding sites. Genome browser view of normalized read counts from ChIP-seq and stranded RNA-seq from mock and infected cells were plotted on the y-axis.

Figure S4

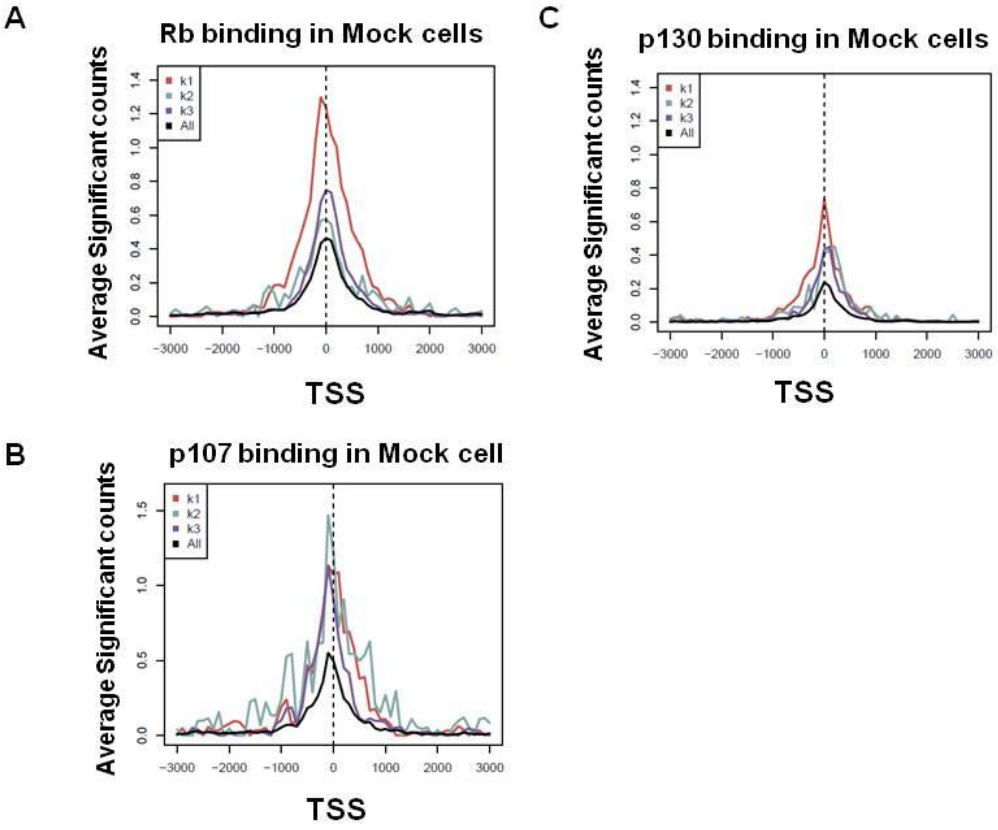


Figure S4 Promoters in k1 gene clusters have highest level of Rb binding in mock-infected cells. Binding levels of p107 and p130 are similar in k1, k2, and k3 gene clusters. (A) Plot of average significant counts from Rb ChIP-seq in mock infected cells for clusters k1, k2, k3 centered at merged e1a binding sites from 6 and 24 hours PI (B) Plot of average significant counts from p107 ChIP-seq in mock infected cells for clusters k1, k2, k3 centered at merged e1a binding sites from 6 and 24 hours PI (C) Plot of average significant counts from p130 ChIP-seq in mock infected cells for clusters k1, k2, k3 centered at merged e1a binding sites from 6 and 24 hours PI. Analysis plots were generated by CEAS software.

Figure S5

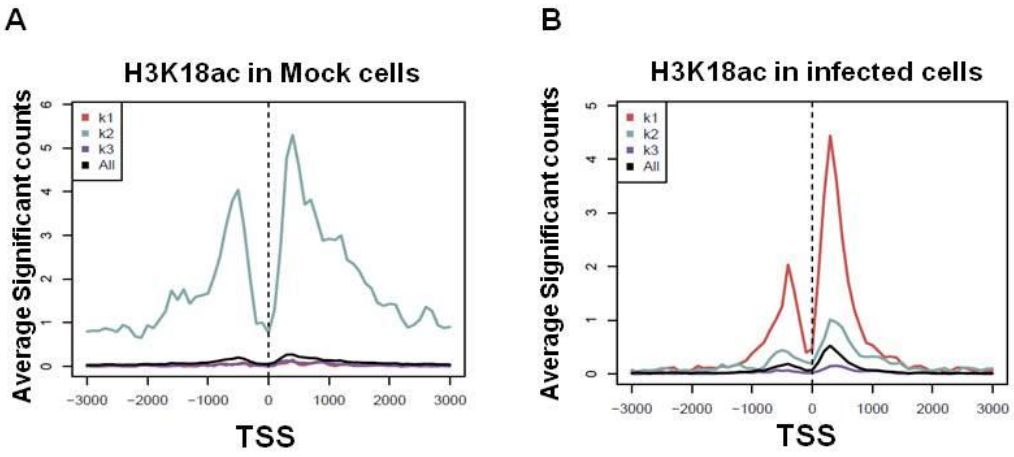


Figure S5. Cluster k2 has highest level of H3K18ac in mock infected cells but is deactylated at 24 hours PI. Cluster k1 gains H3K18ac at 24 hours PI. (A) Plot of average significant counts from H3K18ac ChIP-seq in mock-infected cells centered at TSS of k1,k2,k3 gene clusters. (B) Plot of average significant counts from H3K18ac ChIP-seq in infected cells at 24 hours PI centered at TSS of k1, k2, k3 gene clusters. Analysis plots were generated by CEAS software.

Figure S6

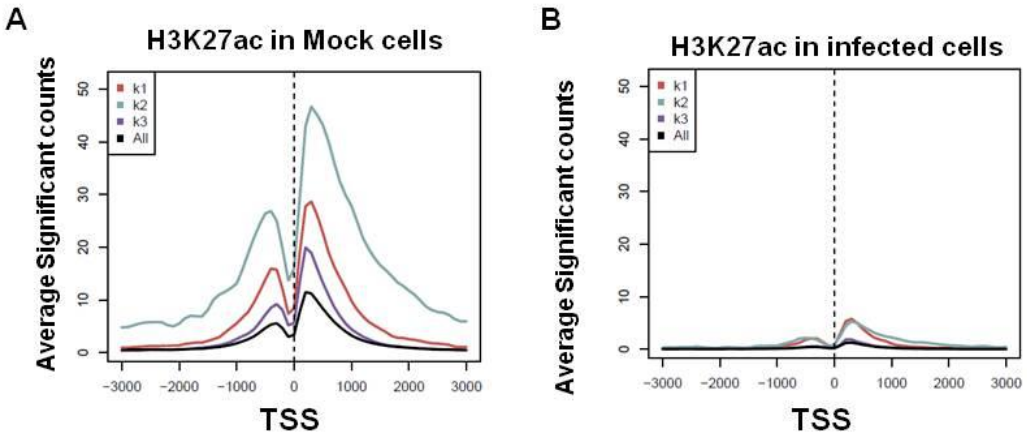


Figure S6. With e1a binding, H3K27ac at cluster k1, k2, k3 are deactylated at 24 hours PI. (A) Plot of average significant counts from H3K27ac ChIP-seq in mock-infected cells centered at TSS of k1,k2,k3 gene clusters. (B) Plot of average significant counts from H3K27ac ChIP-seq in infected cells at 24 hours PI centered at TSS of k1, k2, k3 gene clusters. Analysis plots were generated by CEAS software.

Figure S7

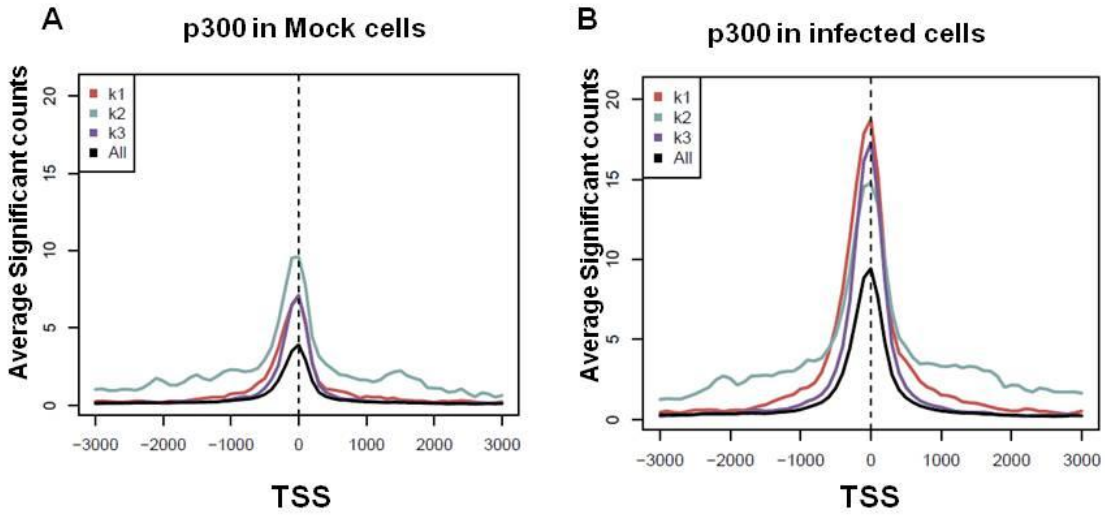


Figure S7. With e1a binding, cluster k1, k2, k3 gain p300 binding at 24 hours PI. (A) Plot of average significant counts from p300 ChIP-seq in mock-infected cells centered at TSS of k1,k2,k3 gene clusters. (B) Plot of average significant counts from p300 ChIP-seq in infected cells at 24 hours PI centered at TSS of k1, k2, k3 gene lusters. Analysis plots were generated by CEAS software

Figure S8

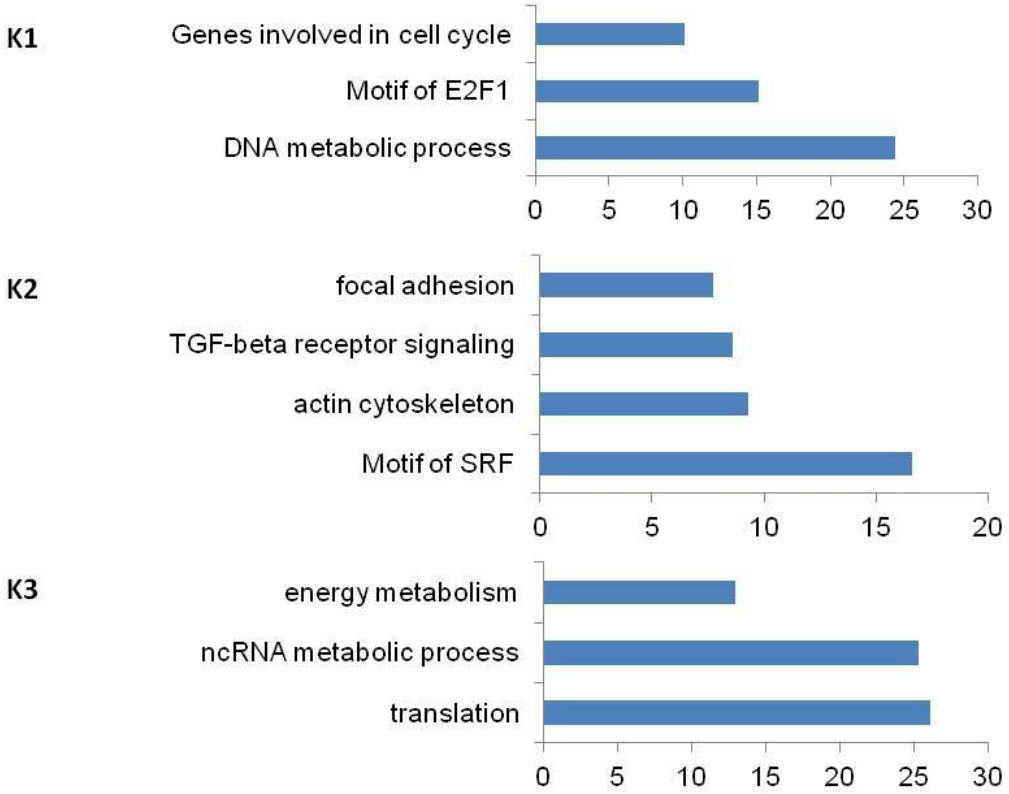


Figure S8. Cluster k1 is enriched with E2F regulated cell cycle genes. Cluster k2 is enriched with fibroblastic gene functions in cell matrix synthesis and maintenance. Cluster k3 is enriched with translation and cell metabolic processes. Bar graph of gene ontology and transcription factor binding motif enrichments of k1, k2, k3 gene clusters. X-axis represents –log of P-value calculated for gene ontology term enrichment. Y-axis shows the significantly-enriched gene ontology term. Gene ontology enrichment was analyzed with GREAT software. Cluster k1 is enriched with E2F-regulated cell cycle genes. Cluster k2 is enriched with fibroblast-specific genes responsible for matinenance and synthesis of cellular matrix. Cluster k3 represents general cell function in translation and energy metabolism.

Figure S9

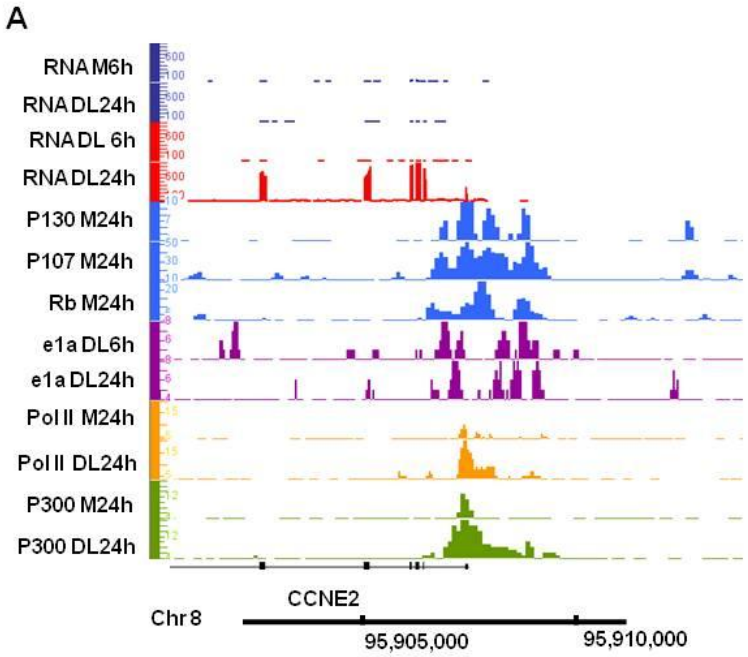


Figure S9. Promoter of CCNE2 is bound by e1a at both 6 and 24 hours PI. (A) Genome browser example of e1a binding at 6 and 24 hours PI at TSS of CCNE2. Y-axis represents normalized total counts from ChIP-seq of Rb, p107, and p130 in mock cells and ChIP-seq of e1a, RNA polymerase II, and P300 in mock and infected cells. RNAseq results from in mock and infected cells at 6 and 24 hours PI is shown at top with Y-axis representing normalized total counts.

Figure S10

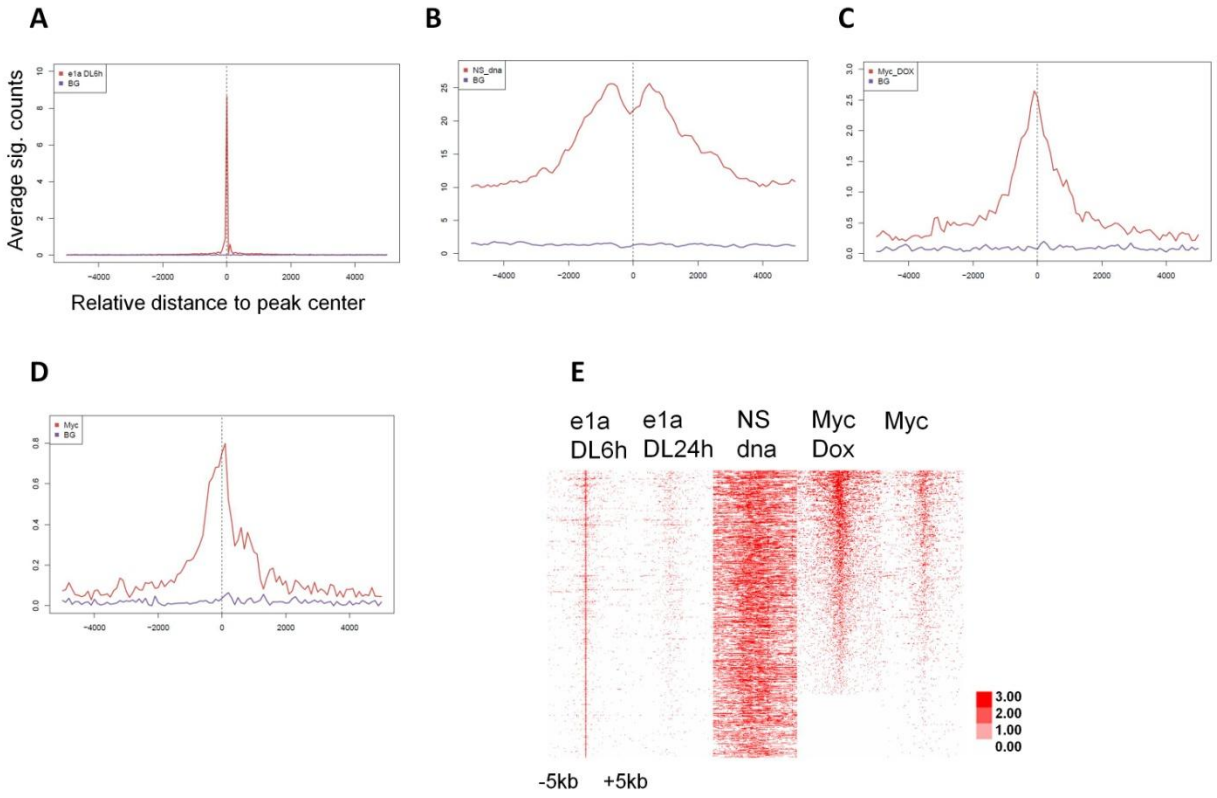


Figure S10. Binding of e1a at 6 hours PI overlaps with sites of NS DNA synthesis in growing cells and Myc binding in BJ fibroblasts. (A) Plot of average significant counts from 6 hour e1a ChIP-seq centered at 6 hour e1a binding sites versus expected genome background binding sites. (B) Plot of average significant counts from nascent strand DNA synthesis-seq in growing IMR90 cells centered at 6 hour e1a binding sites versus expected genome background binding sites. (C)(D) Plot of average significant counts from Myc ChIP-seq in BJ fibroblast overexpressing Myc (Myc-dox) or mock-induced cells. centered at 6 hour e1a binding sites versus expected genome background binding sites. (e) Heatmap of significant counts from multiple ChIP-seq datasets and nascent strand DNA synthesis sequencing in growing IMR90 cells.

Figure S11

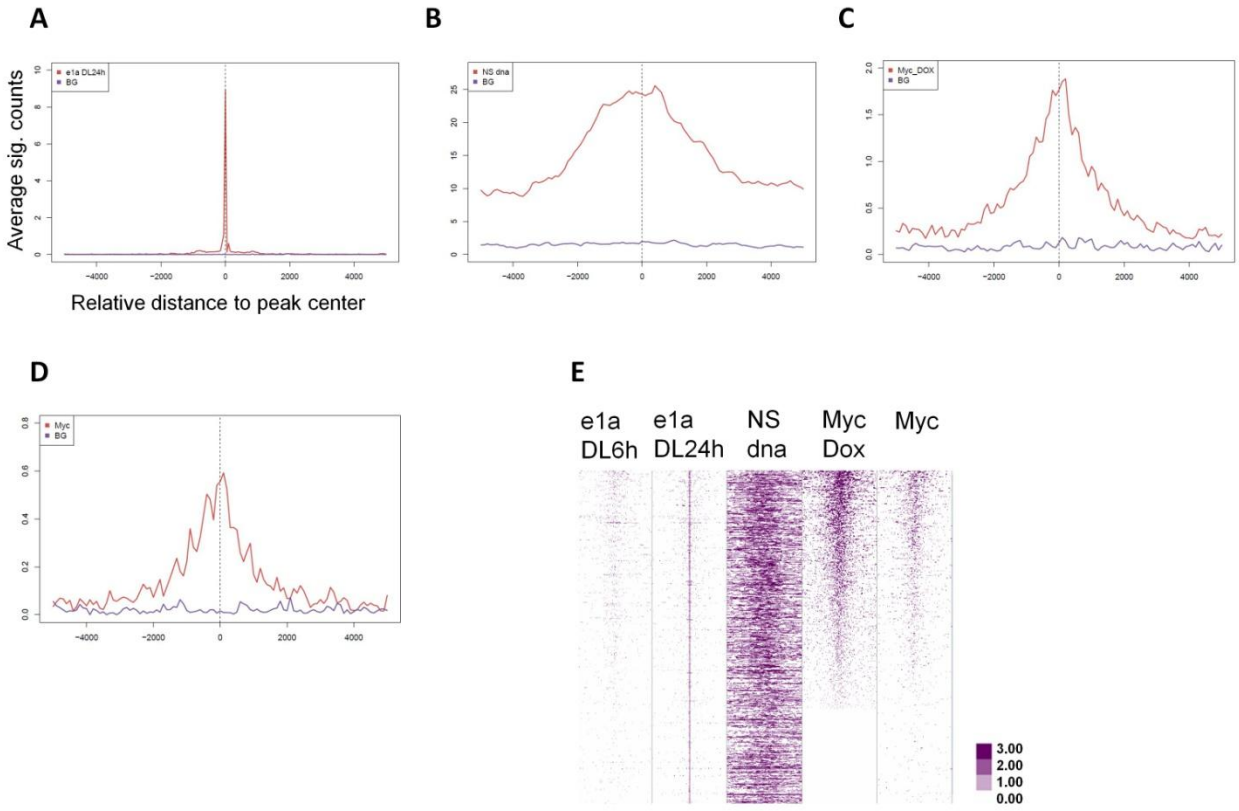


Figure S11. Binding of e1a at 24 hours PI overlaps with sites of NS DNA synthesis in growing cells and Myc binding in BJ fibroblasts. (A) Plot of average significant counts from 24 hour e1a ChIP-seq centered at 24 hour e1a binding sites versus expected genome background binding sites. (B) Plot of average significant counts from nascent strand DNA synthesis-seq in growing IMR90 cells centered at 24 hour e1a binding sites versus expected genome background binding sites. (C)(D) Plot of average significant counts from Myc ChIP-seq in BJ fibroblast overexpressing Myc (Myc-dox) or mock-induced cells. centered at 24 hour e1a binding sites versus expected genome background binding sites. (e) Heatmap of significant counts from multiple ChIP-seq datasets and nascent strand DNA synthesis sequencing in growing IMR90 cells.

Figure S12

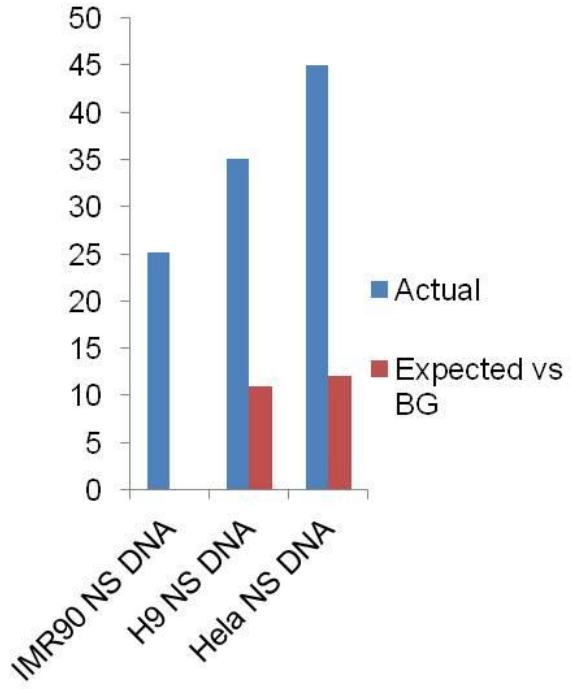


Figure S12. Regions bound by e1a in IMR90 cells has higher overlap with regions of DNA synthesis in H9 and Hela cells comparing to regions of DNA synthesis in IMR90 cells. Y-axis plots percentage of e1a-bound region in IMR90 cells overlapping with nascent strand DNA synthesis regions in IMR90, H9 and hela cells. The actual percent overlap was compared to expected background overlap percentage.

Figure S13

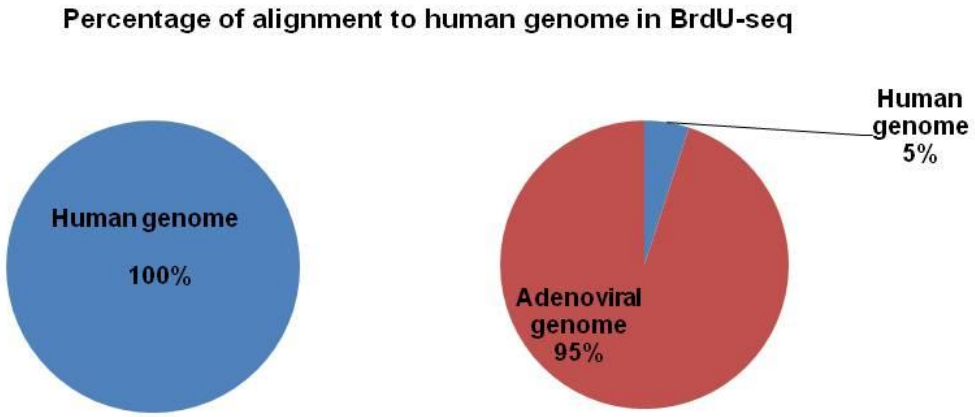


Figure S13. BrdU-seq reveals 95% of newly synthesized DNA in infected cells comes from adenoviral genome. Pie charts plot results from BrdU-seq in asynchronous IMR90 (left) or infected IMR90 cells (right). Percentage of BrdU labeled DNA aligned to human or adenoviral genome is shown in pie chart. BrdU-seq experiment was performed by Bing Li in Kurdistani Lab.

References

Becker, K. A., P. N. Ghule, et al. (2006). "Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase." J Cell Physiol **209**(3): 883-893.

Berk, A. J. (2005). "Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus." Oncogene **24**(52): 7673-7685.

Besnard, E., A. Babled, et al. (2012). "Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs." Nat Struct Mol Biol **19**(8): 837-844.

Branton, P. E., S. T. Bayley, et al. (1985). "Transformation by human adenoviruses." Biochim Biophys Acta **780**(1): 67-94.

Burkhart, D. L. and J. Sage (2008). "Cellular mechanisms of tumour suppression by the retinoblastoma gene." Nat Rev Cancer **8**(9): 671-682.

Chakravarti, D., V. Ogryzko, et al. (1999). "A viral mechanism for inhibition of p300 and PCAF acetyltransferase activity." Cell **96**(3): 393-403.

Chan, H. M., M. Krstic-Demonacos, et al. (2001). "Acetylation control of the retinoblastoma tumour-suppressor protein." Nat Cell Biol **3**(7): 667-674.

Chen, P. B., J. H. Hung, et al. (2013). "Hdac6 regulates Tip60-p400 function in stem cells." Elife **2**: e01557.

Chicas, A., X. Wang, et al. (2010). "Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence." Cancer Cell **17**(4): 376-387.

Choi, S., H. R. Kim, et al. (2014). "Genomic Alterations in the RB Pathway Indicate Prognostic Outcomes of Early-Stage Lung Adenocarcinoma." Clin Cancer Res.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Dahiya, A., M. R. Gavin, et al. (2000). "Role of the LXCXE binding site in Rb function." Mol Cell Biol **20**(18): 6799-6805.

Dick, F. A. (2007). "Structure-function analysis of the retinoblastoma tumor suppressor protein - is the whole a sum of its parts?" Cell Div **2**: 26.

Dominguez-Sola, D., C. Y. Ying, et al. (2007). "Non-transcriptional control of DNA replication by c-Myc." Nature **448**(7152): 445-451.

Eckner, R., M. E. Ewen, et al. (1994). "Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor." Genes Dev **8**(8): 869-884.

Fazzio, T. G., J. T. Huff, et al. (2008). "An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity." Cell **134**(1): 162-174.

Ferrari, R., A. J. Berk, et al. (2009). "Viral manipulation of the host epigenome for oncogenic transformation." Nat Rev Genet **10**(5): 290-294.

Ferrari, R., D. Gou, et al. (2014). "p300/CBP Function in Repression as well as Activation Regulated by Adenovirus Small E1A." Cell Host and Microbe **16**(5): 663-676.

Ferrari, R., M. Pellegrini, et al. (2008). "Epigenetic reprogramming by adenovirus e1a." Science **321**(5892): 1086-1088.

Ferrari, R., T. Su, et al. "Reorganization of the host epigenome by a viral oncogene." Genome Res **22**(7): 1212-1221.

Ferrari, R., T. Su, et al. (2012). "Reorganization of the host epigenome by a viral oncogene." Genome Res **22**(7): 1212-1221.

Frisch, S. M. and J. S. Mymryk (2002). "Adenovirus-5 E1A: paradox and paradigm." Nat Rev Mol Cell Biol **3**(6): 441-452.

Fuchs, M., J. Gerber, et al. (2001). "The p400 complex is an essential E1A transformation target." Cell **106**(3): 297-307.

Fujii, T., T. Ueda, et al. (2010). "Essential role of p400/mDomino chromatin-remodeling ATPase in bone marrow hematopoiesis and cell-cycle progression." J Biol Chem **285**(39): 30214-30223.

Gayther, S. A., S. J. Batley, et al. (2000). "Mutations truncating the EP300 acetylase in human cancers." Nat Genet **24**(3): 300-303.

Gevry, N., H. M. Chan, et al. (2007). "p21 transcription is regulated by differential localization of histone H2A.Z." Genes Dev **21**(15): 1869-1881.

Gevry, N., S. Hardy, et al. (2009). "Histone H2A.Z is essential for estrogen receptor signaling." Genes Dev **23**(13): 1522-1533.

Ghosh, M. K. and M. L. Harter (2003). "A viral mechanism for remodeling chromatin structure in G0 cells." Mol Cell **12**(1): 255-260.

Goodrich, D. W., N. P. Wang, et al. (1991). "The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle." Cell **67**(2): 293-302.

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Horwitz, G. A., K. Zhang, et al. (2008). "Adenovirus small e1a alters global patterns of histone modification." Science **321**(5892): 1084-1085.

Howe, J. A., J. S. Mymryk, et al. (1990). "Retinoblastoma growth suppressor and a 300-kDa protein appear to regulate cellular DNA synthesis." Proc Natl Acad Sci U S A **87**(15): 5883-5887.

Huebner, R. J., W. P. Rowe, et al. (1962). "Oncogenic effects in hamsters of human adenovirus types 12 and 18." Proc Natl Acad Sci U S A **48**: 2051-2058.

Ikeda, M. A. and J. R. Nevins (1993). "Identification of distinct roles for separate E1A domains in disruption of E2F complexes." Mol Cell Biol **13**(11): 7029-7035.

Jin, Q., L. R. Yu, et al. (2011). "Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation." EMBO J **30**(2): 249-262.

Kidder, B. L., G. Hu, et al. (2011). "ChIP-Seq: technical considerations for obtaining high-quality data." Nat Immunol **12**(10): 918-922.

Kim, D., G. Pertea, et al. (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol **14**(4): R36.

Kurdistani, S. K. and M. Grunstein (2003). "Histone acetylation and deacetylation in yeast." Nat Rev Mol Cell Biol **4**(4): 276-284.

Land, H., L. F. Parada, et al. (1983). "Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes." Nature **304**(5927): 596-602.

Liu, X. and R. Marmorstein (2007). "Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor." Genes Dev **21**(21): 2711-2716.

Macher-Goeppinger, S., J. L. Bermejo, et al. (2013). "Senescence-associated protein p400 is a prognostic marker in renal cell carcinoma." Oncol Rep **30**(5): 2245-2253.

McLean, C. Y., D. Bristor, et al. (2010). "GREAT improves functional interpretation of cis-regulatory regions." Nat Biotechnol **28**(5): 495-501.

Mechali, M. (2010). "Eukaryotic DNA replication origins: many choices for appropriate answers." Nat Rev Mol Cell Biol **11**(10): 728-738.

Mendoza-Maldonado, R., R. Paolinelli, et al. (2010). "Interaction of the retinoblastoma protein with Orc1 and its recruitment to human origins of DNA replication." PLoS One **5**(11): e13720.

Moir, R. D., M. Montag-Lowy, et al. (1994). "Dynamic properties of nuclear lamins: lamin B is associated with sites of DNA replication." J Cell Biol **125**(6): 1201-1212.

Montell, C., E. F. Fisher, et al. (1982). "Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site." Nature **295**(5848): 380-384.

Murphree, A. L. and W. F. Benedict (1984). "Retinoblastoma: clues to human oncogenesis." Science **223**(4640): 1028-1033.

Nevins, J. R. (2001). "The Rb/E2F pathway and cancer." Hum Mol Genet **10**(7): 699-703.
Rubin, S. M., A. L. Gall, et al. (2005). "Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release." Cell **123**(6): 1093-1106.

Ruley, H. E. (1983). "Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture." Nature **304**(5927): 602-606.

Ruley, H. E. (1990). "Transforming collaborations between ras and nuclear oncogenes." Cancer Cells **2**(8-9): 258-268.

Sadaie, M., R. Salama, et al. (2013). "Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence." Genes Dev **27**(16): 1800-1808.

Shin, H., T. Liu, et al. (2009). "CEAS: cis-regulatory element annotation system." Bioinformatics **25**(19): 2605-2606.

Soufi, A., G. Donahue, et al. (2012). "Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome." Cell **151**(5): 994-1004.

Srinivasan, S. V., D. Dominguez-Sola, et al. (2013). "Cdc45 is a critical effector of myc-dependent DNA replication stress." Cell Rep **3**(5): 1629-1639.

Trapnell, C., B. A. Williams, et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Tworkowski, K. A., A. A. Chakraborty, et al. (2008). "Adenovirus E1A targets p400 to induce the cellular oncoprotein Myc." Proc Natl Acad Sci U S A **105**(16): 6103-6108.

van den Elsen, P., S. de Pater, et al. (1982). "The relationship between region E1a and E1b of human adenoviruses in cell transformation." Gene **18**(2): 175-185.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Xu, Y., M. K. Ayrapetov, et al. (2012). "Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair." Mol Cell **48**(5): 723-733.

Zentner, G. E., P. J. Tesar, et al. (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions." Genome Res **21**(8): 1273-1283.