HSV-1 reprogramming of the host transcriptional environment

by

Sarah E. Dremel

B.S., University of Minnesota Twin Cities, 2015

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Sarah E. Dremel

It was defended on

March 20, 2020

and approved by

Jennifer Bomberger, Associate Professor, Department of Microbiology and Molecular Genetics

Fred Homa, Professor, Department of Microbiology and Molecular Genetics

Nara Lee, Assistant Professor, Department of Microbiology and Molecular Genetics

Martin Schmidt, Professor, Department of Microbiology and Molecular Genetics

Dissertation Director: Neal DeLuca, Professor, Department of Microbiology and Molecular Genetics

HSV-1 reprogramming of the host transcriptional environment

Sarah E. Dremel, PhD

University of Pittsburgh, 2020

Herpes Simplex Virus-1 (HSV-1) is a ubiquitous pathogen of the oral and genital mucosa. The 152 kilobase double stranded DNA virus employs a coordinated cascade of transcriptional events to efficiently generate progeny. Using Next Generation Sequencing (NGS) techniques we were able to determine a global, unbiased view of both the host and pathogen. We propose a model for how viral DNA replication results in the differential utilization of cellular factors that function in transcription initiation. Our work outlines the various cis- and trans- acting factors utilized by the virus for this complex transcriptional program. We further elucidated the critical role that the major viral transactivator, ICP4, plays throughout the life cycle. We found that ICP4 coated the viral genome early during infection, and that rather than occluding the genome, this dense binding was associated with recruitment of the Pol II machinery to viral promoters. ICP4 discriminately bound the viral genome due to the absence of cellular nucleosomes and high density of cognate binding sites. We posit that ICP4's ability to recruit components of the Pol II machinery to the viral genome creates a competitive transcriptional environment. These distinguishing characteristics ultimately result in a rapid and efficient reprogramming of the host's transcriptional machinery.

Table of Contents

1.0 Introduction 1

1.1 Pathogenesis and Human Health 1

1.1.1 Herpesvirales 1

1.1.2 Pathogenesis 2

1.1.3 Treatment 6

1.1.4 HSV Therapeutic Vectors 7

1.2 General Biology and HSV-1 Lifecycle 8

1.2.1 Virion Structure 8

1.2.2 Genome Structure 9

1.2.3 Productive Infection 11

1.2.3.1 Viral Entry and Trafficking 12

1.2.3.2 Evasion of Nuclear Restriction Factors 14

1.2.3.3 Overview of Pol II Transcription 15

1.2.3.4 Transcriptional Cascade 16

1.2.3.5 VP16-Mediated Transcription 17

1.2.3.6 ICP4-Dependent Transcription 18

1.2.3.7 DNA Replication-Dependent Transcription 18

1.2.3.8 Viral DNA Replication 19

1.2.3.9 Packaging and Assembly 22

1.3 Rationale 23

2.0 Genome replication affects transcription factor binding mediating the cascade of Herpes Simplex Virus transcription..... 25

2.1 Project Summary..... 25

2.2 Importance..... 26

2.3 Introduction..... 27

2.4 Results..... 28

2.4.1 Dependence of Late Viral Transcription on Genome Replication 28

2.4.2 Classification of Viral DNA Replication-Dependent Transcripts. 30

2.4.3 Viral Replication Facilitates Preinitiation Complex Formation on Previously Silent Promoters. 32

2.4.4 A Single Duplication Alters Genomic Accessibility. 34

2.4.5 SP1 Preferentially Binds to the Viral Genome Pre-replication. 36

2.4.6 Differential TFIID Promoter Context Preferred After Replication.....37

2.4.7 Continuous Requirement for ICP4 in Viral Transcription. 39

2.5 Supplemental Figures..... 42

2.6 Discussion..... 48

2.7 Materials and Methods..... 50

3.0 Herpes simplex viral nucleoprotein creates a competitive transcriptional environment facilitating robust viral transcription and host shut off..... 56

3.1 Project Summary..... 56

3.2 Importance..... 57

3.3 Introduction..... 57

3.4 Results..... 59

3.4.1 ICP4 binding is altered by viral genome replication. 59

3.4.2 ICP4 stabilizes GTF binding promoting cooperative preinitiation complex (PIC) assembly.....62

3.4.3 Genome-bound ICP4 does not affect accessibility. 65

3.4.4 ICP4 binds to cellular transcription start sites (TSS) early during infection. 68

3.4.5 ICP4 binding is restricted to accessible regions of the cellular genome.....70

3.4.6 ICP4 mediates depletion of Pol II on cellular promoters 73

3.5 Supplemental Figures..... 76

3.6 Discussion 85

3.7 Materials and Methods 90

4.0 Summary and Discussion 97

4.1 Summary of Thesis 97

4.2 General Discussion 98

5.0 Future Directions 104

Bibliography 107

List of Tables

Table 1 Classification of viral genes bases on DNA replication dependence 32

List of Figures

Figure 1 Species tree of human herpesviruses 1

Figure 2 Complications associated with HSV-1 infection and dissemination 2

Figure 3 HSV-1 antivirals 6

Figure 4 HSV-1 virion structure..... 8

Figure 5 HSV-1 genome structure..... 9

Figure 6 HSV-1 productive life cycle 12

Figure 7 Overview of Pol II transcription 15

Figure 8 Model of HSV-1 DNA replication 20

Figure 9 Initial rounds of replication are sufficient to alter late and early gene transcription 29

Figure 10 Replication dependence of viral transcripts..... 31

Figure 11 Initiation complex formation as a function of genome replication 33

Figure 12 A single round of genome replication immediately alters cellular initiation factor binding 35

Figure 13 Genome replication alters the binding of SP1 to viral promoters 36

Figure 14 Analysis of TFIID-binding sites in E and L promoters..... 38

Figure 15 ICP4 is continuously required for transcription after the onset of genome replication 40

Figure 16 ICP4 binding at key points in the viral life cycle..... 60

Figure 17 ICP4 recruitment of host Pol II machinery to viral promoters 64

Figure 18 ICP4 dependence of viral genome accessibility and Histone H3 binding..... 67

Figure 19 ICP4 binds cellular promoters during early infection 69

Figure 20 Association between cellular ICP4 binding and chromatin..... 71

Figure 21 The role of ICP4 in Pol II loss on host promoters 74

Figure 22 Model for ICP4 function. 89

List of Supplemental Tables

Supplemental Table 1 Motif Discovery..... 46

List of Supplemental Figures

Supplemental Figure 1 Viral transcription kinetics as a function of genome replication inhibition 42

Supplemental Figure 2 General transcription factors bind their respective binding elements 43

Supplemental Figure 3 Transcription factor binding in presence and absence of viral genome replication 44

Supplemental Figure 4 Transcription factor binding at different stages of genome replication 45

Supplemental Figure 5 Synchronicity of productive viral infection 48

Supplemental Figure 6 GTF recruitment to the viral genome in the presence and absence of ICP4 77

Supplemental Figure 7 Analysis of Viral ChIP-Seq data quality 78

Supplemental Figure 8 Analysis of Cellular ChIP-Seq data quality 79

Supplemental Figure 9 Epigenetic markers on the viral genome 81

Supplemental Figure 10 Epigenetic markers on the host genome during early infection ... 83

Supplemental Figure 11 Binding location of ICP4 relative to epigenetic markers and binding motifs 85

1.0 Introduction

1.1 Pathogenesis and Human Health

1.1.1 Herpesvirales

In 2008 the International Committee on Taxonomy of Viruses established a new order of double stranded (ds) DNA viruses called Herpesvirales. This order contains three families differentiated by species of infection, these include: Alloherpesviridae, Herpesviridae, and Malacoherpesviridae. Our work focuses on the family Herpesviridae which infects mammals, birds and reptiles (1). The defining characteristic of all viruses classified as herpesviridae is the virion morphology (2). There are nine species of herpesviruses known to infect humans (Fig. 1).

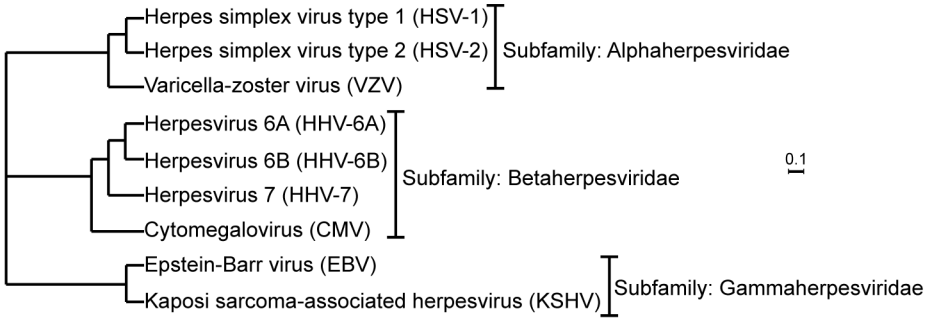


Figure 1 Species tree of human herpesviruses

Phylogenetic tree is based on data from (3). Scale bar represents species divergence.

All of these viruses share the following characteristics: 1. Encode enzymes for DNA replication and nucleotide metabolism, 2. Replicate and assemble within the host nucleus, 3. Cause cell death at the end of productive infection, and 4. Establish latency and persist through the life of the host (4). There are 44 genes conserved within herpesviridae, these largely include structural components and the DNA replication machinery (5). From here we will focus on herpes simplex virus type 1 (HSV-1), an alphaherpesvirus so defined by its ability to infect a wide range of hosts, replicate rapidly, and establish latency in sensory ganglia (4).

1.1.2 Pathogenesis

HSV-1 infection has been associated with a wide range of pathologies and symptoms, representative of the large range of cells permissive to replication (Fig. 2).

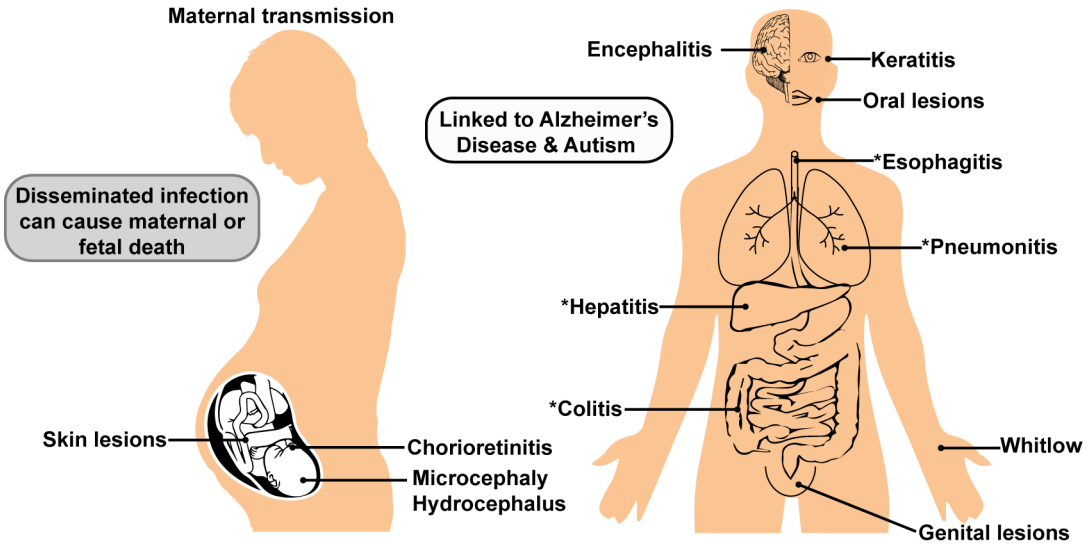


Figure 2 Complications associated with HSV-1 infection and dissemination

***Complications are documented in immunocompromised individuals, largely within transplant recipients.**

Individuals are typically exposed to HSV-1 before the age of 5, although the age of first exposure appears to be increasing. A recent analysis of data from the National Health and Nutrition Examination Survey determined that seroprevalance of HSV-1 ranged from 47-52% in Americans age 20-29, and 27-33% in Americans age 14-19. Across all age groups there was a decrease in HSV-1 seroprevalence from 1976 to 2010, however this trend was the most drastic in the 14-19 age group (6).

In immunocompetent individuals HSV-1 infection is generally asymptomatic and associated with lesions at the primary site of infection. Primary infection generally occurs at the oral or genital mucosa, with latency being established in the innervating sensory nerve—namely the trigeminal or sacral ganglia (7, 8). HSV-1 can then reactivate causing recurrent outbreaks at the site of primary infection. HSV-1 reactivation has been demonstrated in response to ultraviolet light, hypothermia, emotional stress, menstruation, iontophoresis, or epinephrine (4). Historically HSV-1 was the causative agent of oral lesions, while HSV-2 was the primary cause of genital lesions. In recent years, there has been a shift, in which nearly 60% of new genital herpes infection are caused by HSV-1 (9, 10). This may be due to an increasing age of acquiring HSV-1, leading to an absence of protective antibodies once the individual becomes sexually active.

Infection can occur—although less commonly—at other locations, resulting in whitlow and keratitis. Herpetic whitlow is a lesion of the hands, and was quite common in dentists prior to the usage of gloves during examinations (11). Herpes keratitis is HSV-1 corneal infection, and an estimated 50,000 cases recur or emerge each year in the U.S (12). Worldwide the incidence is estimated to be about 10 million individuals, with 40,000 cases annually of visual impairment or blindness as a result (13).

In unique cases HSV-1 can be neurovirulent and cause severe pathogenesis. In the United States encephalitis occurs at an incidence of 7 out of 100,000. Herpes simplex is the largest cause of encephalitis, accounting for 50-75% of cases when the cause was determined (14). Herpes simplex encephalitis primarily occurs in children and adolescents, and is linked to genetic abnormalities including mutations in the genes for TLR3, TRIF, UNC93B1, and TAF3 (4). Early patient treatment with acyclovir drastically improves patient outcomes, although encephalitis remains a largely incurable disease with lasting effects (14).

Herpes simplex possesses the ability to pass the placental barrier and thus can be passed from mother to fetus or newborn perinatally. If a pregnant mother acquires a new genital herpes infection in the last trimester the risk of transmission to the fetus is between 30-50%. Primary genital infection in rare cases can cause severe disease and viral dissemination, resulting in maternal death. In cases of recurrent genital herpes the risk of transmission to the fetus is significantly lower, approaching 3% (15). HSV-1 transmission to the fetus (congenital herpes) can result in spontaneous abortion, microcephaly, hydrocephaly and chorioretinitis (16). Herpes simplex transmission to the newborn (neonatal herpes) occurs in about 1 out of 3000 live births in the U.S. Neonatal herpes can result in encephalitis, multi-organ viral dissemination, and skin lesions (17, 18).

In immunocompromised individuals HSV-1 can disseminate throughout the body. This is perhaps best studied in the case of transplant patients, in which herpes reactivation can cause encephalitis (19), esophagitis (20), upper and lower respiratory infection (21, 22), hepatitis (23), and colitis (24). Before prophylactic use of the antiviral, acyclovir, HSV-1 reactivated commonly in transplant patients with the highest observed rate (82%) in hematopoietic stem cell transplant

recipients (25). Resistance to acyclovir has been observed in transplant patients, and current estimates range from 1-14% (26, 27).

Due to herpes simplex' ability to infect the central nervous system (CNS) recent studies have investigated a possible link between the virus and development of neurologic disorders. In 2017 a paper was published proposing a link between maternal HSV-1 and autism (28). Autism is a development disorder with symptoms ranging on a spectrum from severe to mild defined by deficits in interactions and communication, typically presenting with repetitive behaviors. Autism generally presents during childhood, however individuals have developed autism at ages 11, 14, and 31 following herpes encephalitis (29-31). It is of note that all of these documented cases occurred prior to 1991, and no additional cases have been published. This may be due to the use of acyclovir in treatment of herpes encephalitis or better diagnostic tools for autism. The cause of autism is unknown, although it has been proposed that early infection by a viral or microbial pathogen may affect development. Again this supports the possibility of HSV-1 as a cause, as the virus is both neurotropic and can cross the placental barrier. To date there is no work done in animal models linking early HSV-1 infection to behavioral changes, although these studies have been done with other pathogens. Additional work is required for any conclusion regarding HSV-1 and autism.

Perhaps more convincing, is the link between Alzheimer's disease (AD) and recurrent HSV-1 infections. Research has demonstrated that HSV-1 infection can induce accumulation of amyloid beta deposits and hyperphosphorylated tau, both of which contribute to AD development (32). Current data suggests that HSV-1 is not the sole causative agents, but rather a contributor to the development of AD. More research is required to determine why select individuals develop AD, when HSV-1 is resident in more than 60% of individuals over the age of 50 (6).

1.1.3 Treatment

There are multiple antiviral drugs approved by the FDA for treatment of HSV-1, these include acyclovir, valacyclovir, and famciclovir (Fig. 3).

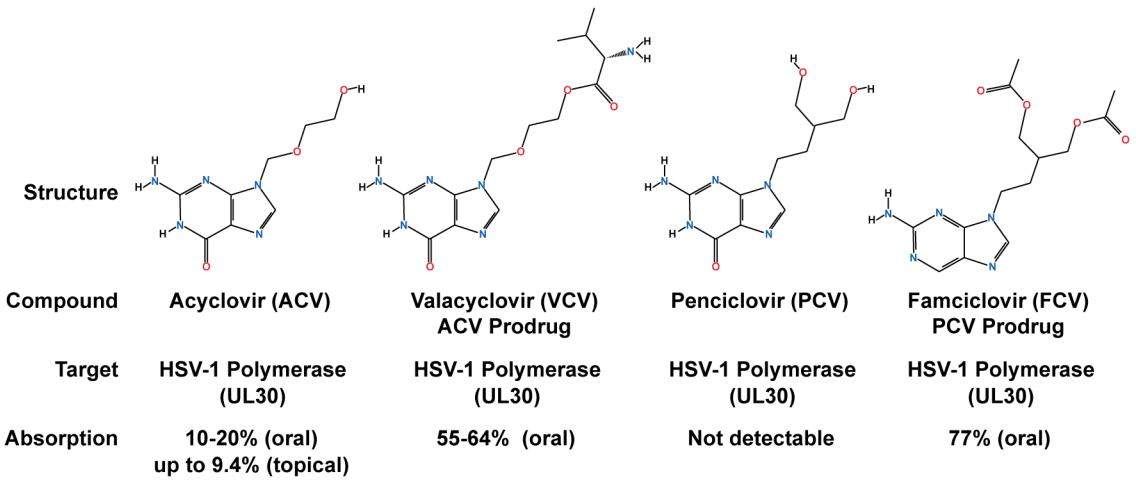


Figure 3 HSV-1 antivirals

Acyclovir (ACV) was approved by the FDA in 1982 as Zovirax. ACV is a nucleotide analog with an acyclic side chain lacking a 3' hydroxyl group (33). After uptake into infected cells ACV is phosphorylated by the viral thymidine kinase (tk) (34). After this event ACV is di- and tri-phosphorylated by host cell kinases (35, 36). In its active form ACV is incorporated into actively replicating DNA resulting in chain termination (37, 38). ACV is a particularly effective antiviral for two reasons: 1. The initial phosphorylation can be performed in the absence of the virus, however cellular kinases are 1 million-fold less capable of the enzymatic reaction. 2. The triphosphorylated ACV is a preferred substrate for the viral polymerase. ACV has low oral bioavailability, so valaciclovir (VACV) was synthesized which is the L-valyl ester of acyclovir

(39). VACV has similar efficacy, but a higher bioavailability (40). Another derivative with higher bioavailability is famciclovir (FCV). FCV is converted in the liver from its prodrug form to 6-deoxypencicloviv, or penciclovir (PCV) (41). PCV is an acyclic guanine derivative with a similar mechanism of action to ACV (42). FCV has the highest bioavailability of HSV-1 antivirals (43). ACV, FCV, and VCV are used to treat most cases of HSV-1. In cases of drug resistance additional antivirals including cidofovir, foscarnet, ganciclovir and valganciclovir are used. These antivirals share one major problem—they target the viral polymerase and can only inhibit HSV-1 productive infection. To date there are no antivirals effective at targeting the latent viral reservoir. Furthermore there is a rising incidence of resistance to these drugs in immunocompromised patients (44), indicating a need to develop a new class of antivirals.

1.1.4 HSV Therapeutic Vectors

In 2015, the first viral-based therapy drug was approved by the FDA. T-vec (talimogene laherparepvec) is an oncolytic HSV-1 based vector expressing GM-CSF. In phase III clinical trials, T-Vec was approved as an intralesional treatment for advanced malignant melanoma and demonstrated to suppress tumor growth and prolong patient survival. Additional herpes based vectors are currently undergoing clinical trials for treatment of glioblastoma, and have had success in animal models for the treatment of various cancers, painful diabetic neuropathy, and sensory neuropathy. HSV-1 is an appealing viral vector for therapy treatment due to its broad cell tropism, including the peripheral nervous system (PNS) and CNS, low toxicity, large insert capacity, and long-term expression (45).

1.2 General Biology and HSV-1 Lifecycle

1.2.1 Virion Structure

Structural morphology is actually what defines the vast number of viruses contained within the order herpesviridae. All herpesviruses, from those infecting humans to mollusks, possess a linear dsDNA genome contained within a T-16 capsid surrounded by a proteinaceous tegument contained within a lipid envelope studded with glycoproteins (1, 2) (Fig. 4).

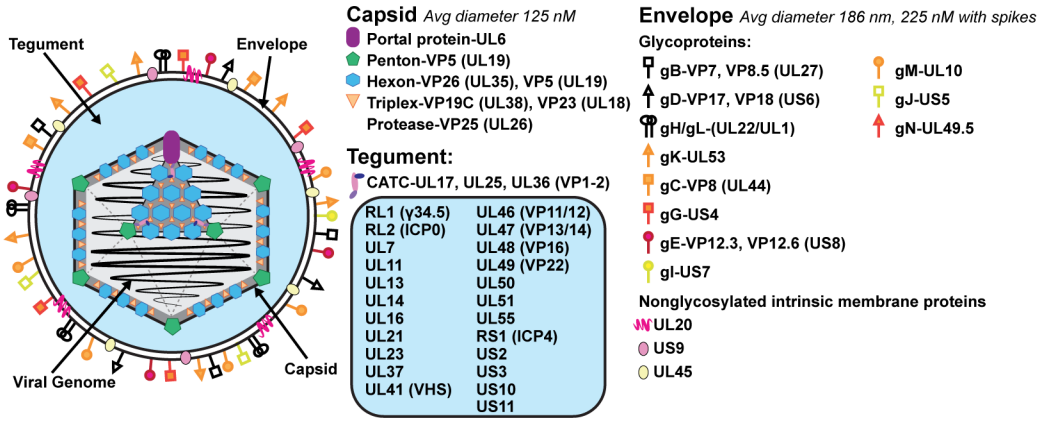


Figure 4 HSV-1 virion structure

Schematic for the capsid and interaction with the capsid-associated tegument complex (CATC), was based on (46, 47).

The viral envelope is a lipid bilayer derived from the host's trans-Golgi network (TGN) (48) or endosomes (49). The envelope is studded with up to 12 different glycoproteins—gB, gC, gD, gE, gG, gH, gI, gJ, gK, gL, gM, gN—and three nonglycosylated intrinsic membrane proteins—UL20, US9, and UL45. Additional proteins UL24, UL43, and UL34 may also be present

in the envelope (4). Inside the envelope is the tegument composed of up to 26 different proteins (50), notably the capsid-associated tegument complex (CATC). Five copies of the CATC surround and stabilize the major capsid penton at each vertex (46). Other tegument proteins do not have a structural role but instead exert immediate activity to promote a pro-viral host environment, these include ICP0, ICP4, VP16, γ34.5, and VHS (51). Located asymmetrically within the tegument is the capsid (52). The capsid is made up of 150 hexons (1 UL35: 6 UL19), 11 pentons (5 UL19), 320 triplexes (1 UL38: 2 UL18), one portal protein (UL6) and the protease VP24 (UL26) (46, 47). Coiled tightly within the capsid is approximately one copy of the linear viral genome. The highly negative charge of the viral genome is believed to be at least partially neutralized by the presence of the basic polyamines spermidine and spermine contained within the capsid (53).

1.2.2 Genome Structure

HSV-1 has a 152 kilobase linear double stranded (ds) genome of high GC-content (approaching 70%) (54, 55) with nicks and gaps (56, 57). The genome consists of a unique long (UL) and unique short (US) component which are linked by a joint containing inverted repeat regions (Fig. 5).

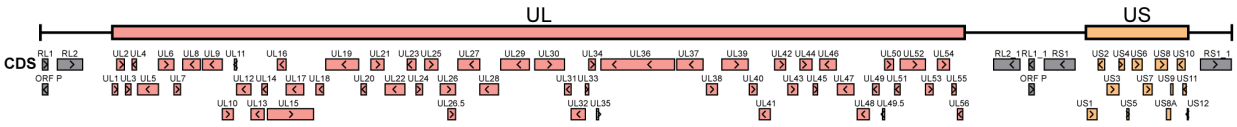


Figure 5 HSV-1 genome structure

Genome structure is based on KT899744.1 assembly; Unique Long (UL), Unique Short (US), coding region (CDS).

The UL and US components can invert relative to each other, resulting in four different genomic isomers found at equal proportions in mature virions (58, 59). The inverted repeats or 'a' sequence is some variation of CGCTCCTCCCCC, with differences observed in sequence and repeat number between different lab adapted strains (60). With the exception of the genes for γ34.5 (RL1), ORF P, ICP0 (RL2), and ICP4 (RS1) which flank the UL or US regions and have two copies within the genome, all other viral genes are only present in one copy. Current genome assemblies contain around 84 protein-coding open reading frames (ORFs), although more recent studies have proposed that there are upwards of 200 transcripts and ORFs (61). The viral genome contains a 5 kilobase (kb) noncoding (nc) transcript primarily made during latency, or the latency-associated transcript (LAT). Studies have also indicated at least 16 miRNA's expressed by HSV-1 (62). To put this in reference, the cellular genome is 3.2 billion base pairs containing at least 60,000 genes that make about 200,000 different transcripts, or $1.9x10^{-2}$ genes/kbp (Gencode v33). While the total size and complexity of the cellular genome is much greater, the viral genome is significantly more compact, with at least 0.6 genes/kbp.

Another structural feature of the viral genome is the lack of chromatin structure. The cellular genome is compacted into chromatin which is facilitated by nucleosomes composed of two copies of each histone protein: H3, H4, H2A and H2B. 146-147 bp of the cellular genome is bound and wrapped around the histone octamer, which is facilitated by the basic charge of the histone complex (63). Nucleosomes space evenly along the genome with the linker histone H1 bound in between (64). These structures facilitate compaction of the cellular genome into 30 nm fibers (65). Euchromatin refers to transcriptionally active chromatin whereas heterochromatin refers to transcriptional silent chromatin. The chromatin states are regulated by different post-translational modifications on the histones composing the octamer. Generally, acetylation reduces

the basic charge of the histone tail altering the attraction between the nucleosome and genomic DNA resulting in reduced compaction. Alternatively, methylation of the core histones commonly on lysine or arginine residues can make the histone tail more rigid—effectively binding and compacting the genomic DNA. These are simplified principles and the reality is that both acetylation and methylation have been associated with eu- or hetero-chromatin depending on the residue of the histone octamer modified. Chromatin remodeling plays a major role in regulation of transcription, replication and DNA repair for the cellular genome (66).

During productive infection, HSV-1 has a mechanism by which it prevents assembly and binding of cellular nucleosomes on the viral genome, this is indicated by the following evidence: i. HSV-1 virions do not contain histones (67-69), ii. viral genomes are not bound by nucleosomes after nuclear entry (69-73), and iii. cellular histones are excluded from replication compartments (74-76). This may be due to the short time from when the viral genome enters the nucleus prior to host cell death (12-24 hours), and the amount of time required for histone turnover and nascent nucleosome assembly. Or this may be due to the direct actions of viral tegument proteins, ICP0, VP16, and ICP4, to combat chromatinization (77). The exact contribution of these factors remains to be explored. Compared to the host in which 75-90% of the genome is packaged in nucleosomes, the viral genome serves as a unique model to study the other 10-25% and examine cellular processes in the absence of chromatin structural dynamics. Our studies discussed in Chapter 2 and 3 reinforce this idea.

1.2.3 Productive Infection

HSV-1 undergoes lytic, or productive, infection at the site of primary infection which is typically the oral or genital mucosa. During productive infection, the virus undergoes a tightly

ordered life cycle and within 12 to 24 hours produces thousands of viral progeny and results in death of the host cell (Fig. 6).

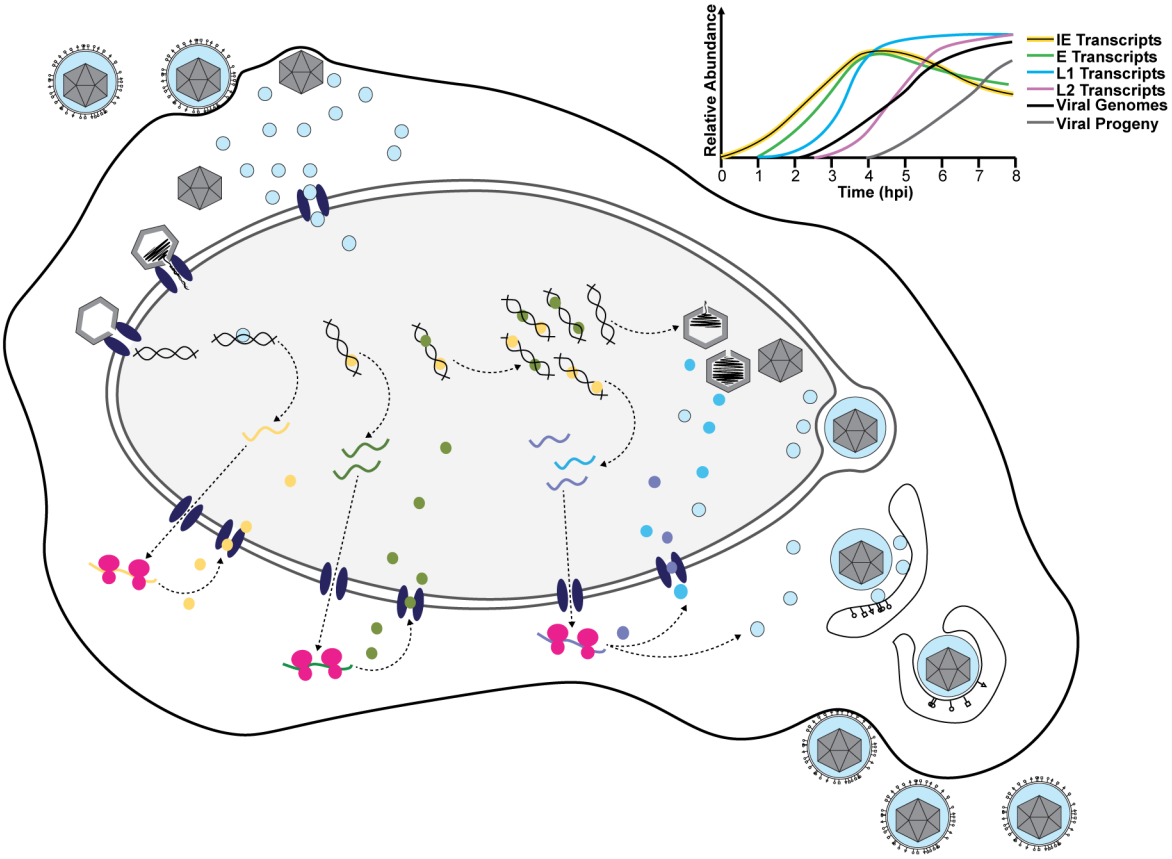


Figure 6 HSV-1 productive life cycle

Schematic of the viral life cycle in fibroblasts; Kinetic classes are colored with immediate early (IE) in yellow, early (E) in green, leaky late (L1) in teal, and true late (L2) in purple.

1.2.3.1 Viral Entry and Trafficking

HSV-1 has a wide host and tissue tropism and has been shown in lab settings to infect mice, rabbits, guinea pigs, and canine kidney cells. This is likely due to the virus' ability to utilize a number of different entry receptors and mechanisms of entry. HSV-1 expresses up to 12 different

glycoproteins and of these only gD, gH, gL, and gB are essential for viral entry in cell culture and animal models of infection (78, 79). These glycoproteins are key in coordinating the steps of attachment: 1. gB and gC attach to host surface heparan sulfate. 2. gD binds to host nectin-1, herpes virus entry mediator, or 3-O-sulfated heparan sulfate. This binding event alters the conformation of gD so that it now binds the heterodimeric glycoprotein, gH/gL. 3. In complex with gH/gL, gB mediates entry to the host cell. The host attachment molecule that gD binds to appears to decide the method of entry, whether it be endocytosis or fusion (80). Studies have identified more specialized role for some of the other "nonessential" glycoproteins, for instance gK promotes membrane fusion in neurons (81), and gC and gG promote apical entry of polarized epithelial cells (82, 83).

After either membrane fusion or endocytosis, viral tegument proteins are released into the cytoplasm. Some tegument proteins remain associated with the capsid and traffic to the nucleus, while other tegument proteins traffic independently. Capsid trafficking is mediated by the microtubule network using the motor adaptor, dynein (84, 85). Capsids travel from the periphery to the centrosome by minus-end directed transport, switching to plus-end directed transport to arrive at the nucleus (86, 87). The viral genome is ejected through the nuclear pore complex, which requires host importin-beta (88-90) and cleavage of the CATC component, UL36 (91). A recent study found that the ionic and temperature conditions present in native host cells facilitates efficient ejection of the HSV-1 genome into the nucleus (92). The US end of the genome is ejected first from the capsid (93). Unlike other viruses, the HSV-1 capsid does not dissolve during delivery of the viral genome as empty capsids have been visualized docked at nuclear pore complexes (94, 95).

1.2.3.2 Evasion of Nuclear Restriction Factors

After the viral genome enters the nucleus, ICP0 acts within the first hour of infection to overcome host cell intrinsic defense mechanisms. Successful host restriction results in quiescent infection where the viral genome is maintained and transcriptionally silent. This prevents production of multiple virus encoded proteins which act to directly inhibit innate immune sensing, host translation, and transcription. Thus, quiescently infected cells can signal to neighboring cells and prevent HSV-1 spread. Host restriction is largely conferred by components of nuclear domain 10 (ND10) or promyelocytic leukemia nuclear (PML) bodies, including PML, SP100, DAXX, ATRX, and HIRA (96). These host proteins are shown to respond within an hour of viral nuclear entry. In the absence of ICP0 these structures sequester the viral genome and severely reduce the incidence of productively replicating HSV-1 genomes (97).

ICP0 is an IE protein that encodes a U3 ubiquitin ligase (98), and is essential for viral replication at low but not high multiplicity of infection (MOI). At a low MOI with an ICP0 null mutant, many infected cells enter into a quiescent state in which there is a significantly reduced viral gene expression (96). These viral genomes can be de-repressed by delivery in trans of ICP0 (99), and to a lesser extent ICP4, VP16 (100-102) or histone deacetylase (HDAC) inhibitors (103). Addition of the proteasome inhibitor MG132 prevents ICP0 from reactivating the quiescent viral genome, suggesting that the E3 ubiquitin ligase activity is critical to ICP0 function (104). Many cellular factors have been identified as targets of ICP0's E3 ubiquitin ligase activity including PML, Sp100, DNA protein kinase, CENP-C, CENP-A, ATRX, and IFI-16 (105-111).

As discussed above in section 1.2.2, during productive infection the HSV-1 genome lacks nucleosomes or characteristics of cellular chromatin. The ability of HDAC inhibitors to rescue ICP0 null quiescent viral genomes suggests that ICP0 plays a direct role in preventing nucleosome

assembly (103, 112, 113). There are currently two proposed mechanisms by which ICP0 ameliorates heterochromatization of the viral genome. First ICP0 may induces degradation of a host protein in this pathway via it's U3 ubiquitin ligase activity. Second ICP0 may associate and inhibit activity of various host chromatin remodelers. ICP0 has been shown to complex with HDAC5, HDAC6, HDAC7 (114) and the REST/CoREST complex (115, 116).

1.2.3.3 Overview of Pol II Transcription

HSV-1 utilizes the host RNA Polymerase II (Pol II) machinery for viral transcription (117). We will briefly discuss the basic steps in Pol II transcription (Fig. 7).

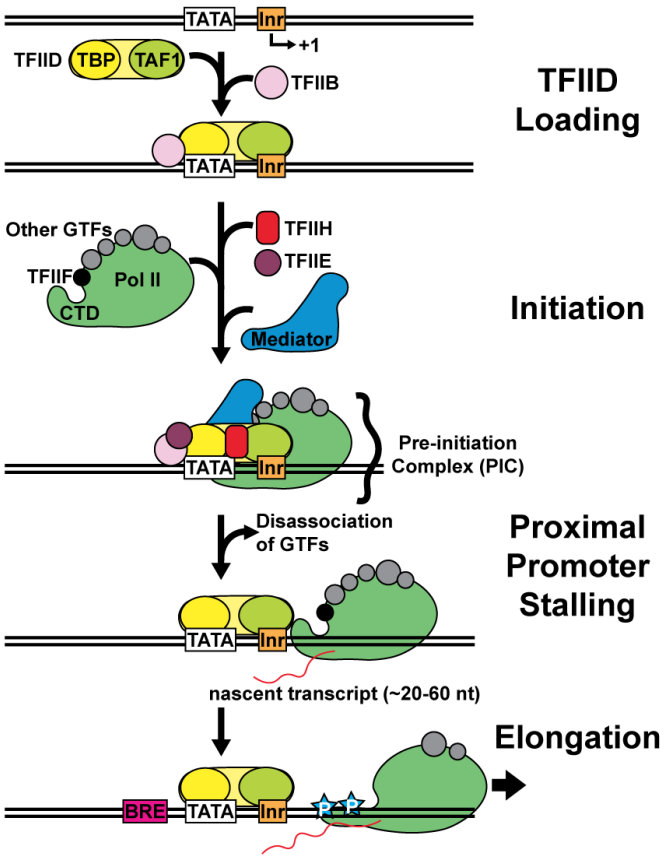


Figure 7 Overview of Pol II transcription

First the TFIID complex is recruited to specific cis-elements in the core promoter. TFIID is comprised of the TATA-binding protein (TBP) and up to 11 different TAFs. The TBP component binds to TATA boxes and the TAFs bind to other motifs including the initiator element (Inr). Binding of TFIID distorts the DNA helix near the TATA box, which aids in binding and assembly of additional Pol II transcription factors, including TFIIA and TFIIB. Pol II is recruited through protein-protein interactions and positioned relative to the TSS by DNA-bound GTFs, namely TFIIB and TFIID. This complex is termed the pre-initiation complex (PIC) and it is composed of TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIJ, TFIIH, and Pol II. Additional activators and enhancers such as the Mediator complex are also critical components of the PIC. After the PIC is formed, Pol II begins transcription. It is believed that most GTF's absent DNA binding ability dissociate from the promoter at this point. Pol II synthesizes approximately 20-60 nucleotides (nt) before pausing. This is termed proximal promoter stalling, and is not relieved until the C-terminal domain (CTD) of Pol II is phosphorylated by elongation factors including TFIIH, CDK9/P-TEFb. Now that the CTD has been phosphorylated Pol II proceeds with transcription (118-120). This is a general overview of Pol II transcription, however most of these core GTF's have been identified as critical for viral transcription—with the exception of TFIIA, which appears to be function redundantly to ICP4 (121, 122).

1.2.3.4 Transcriptional Cascade

Like other DNA viruses HSV-1 coordinates a transcriptional cascade ensuring coordinated and efficient transcription of at least 80 viral proteins (123-125). Briefly the gene classes are: immediate early (IE or α), early (E or β), and late (L or γ).

IE proteins were classified as those expressed in the absence of de novo viral protein synthesis. There are five IE proteins and their expression peaks 2-4 hpi (126). ICP0 plays a role in

counteracting host intrinsic defenses and chromatin formation as discussed above. ICP4 has a critical role in E and L gene transcription as discussed in section 1.2.3.6 (127). ICP22 plays an essential role in promoting Pol II elongation at later times during infection (128). ICP27 promotes viral mRNA 3' processing (129), and is reported to also function in mRNA export and inhibition of host splicing. ICP47 localizes to the cytoplasm where it binds host TAP1/TAP2 and inhibits antigen presentation on the cell surface (130).

E proteins were classified as those expressed in the presence of functional alpha proteins before DNA replication. E protein expression peaks 6-12 hpi. E proteins include all components of the DNA replication machinery, classically this includes: UL2, UL5, UL8, UL9, UL12, UL23, UL29, UL30, UL39, UL40, UL42, UL50, and UL52 (4).

L proteins were classified as those expressed in the presence of functional alpha proteins after DNA replication. Late proteins were further delineated as leaky late (γ1 or L1) and true late (γ2 or L2) based on whether genome replication increases expression or is absolutely required. L protein expression peaks 10-15 hpi. L proteins include the structural, packaging, and assembly components of the virion.

1.2.3.5 VP16-Mediated Transcription

IE genes are transcribed immediately after delivery of the viral genome to the nucleus and this is largely due to the activating role of VP16. VP16 (UL48) is a L1 gene that encodes a transcriptional activator (131, 132). VP16 is delivered from the viral tegument to the cytosol where it complexes with HCF-1 and then travels to the nucleus (133). HCF-1 is a member of the host cell factor family and was first identified in complex with VP16 (134-136). Work since then has found that in uninfected cells HCF-1 plays a critical role in control of the cell cycle (137, 138). In the nucleus, the VP16-HCF-1 heterodimer binds OCT1 (also called POU2F1). OCT1 is a member of

the POU transcription factor family and binds the octamer motif ATTTGCAT to activate transcription (139). In the context of HSV-1 infection, OCT1 mediates binding of the complex to promoter elements containing TAATGARAT elements; where R is A or G (140, 141). These TAATGARAT elements are contained exclusively within the viral IE promoters, where VP16 robustly recruits the cellular GTFs, SP1, TFIIB, TFIIH, and TFIID (142). VP16 may also play a role in preventing chromatinization of the incoming viral genome, as VP16 has been shown to recruit chromatin remodelers including CBP, p300, and the SWI-SNF complex (BRG-1, BRM) (77).

1.2.3.6 ICP4-Dependent Transcription

ICP4 is the main viral transcriptional activator, and is essential for transcription of E and L genes (143-145). ICP4 is known to form dimers or multimers on DNA, each form has a different affinity and specificity for binding (146, 147). Namely at high concentrations ICP4 multimerizes on dsDNA sequence-nonspecifically (148). At low concentrations ICP4 binds sequence-specifically to the motif 5′ATCGTCNNNNYCGRC-3’ (149). ICP4 interacts with a number of cellular general transcription factors (GTFs), including components of TFIID, TFIIH, Mediator, CPSF, and the chromatin remodelers SWI-SNF, NURD, and Ino80 (150-152), facilitating their recruitment to the viral genome through its DNA binding activity (151, 153, 154). Recruitment of these transcription factors facilitates PIC formation on viral promoters, activating robust transcription (122, 155).

1.2.3.7 DNA Replication-Dependent Transcription

While both E and L genes are activated in an ICP4-dependent manner, L genes are not efficiently transcribed or expressed until after viral DNA replication has started. The exact reason

for this requirement is still not known. Unlike other viruses, HSV-1 does not encode a late gene specific trans activating factor. Additionally, there is still no evidence of a repressive factor bound to the viral genome prior to replication.

One model proposes that a component of the viral DNA replication machinery serves a dual function and is continuously required for late gene transcription. Numerous interactions have been demonstrated between the core DNA synthesis components and cellular factors. For instance, the viral polymerase holoenzyme was reported to interact with cdc2, topoisomerase IIa, HCF-1, Asf1b, hsp90 and Fen-1 (156-160). These factors may aid in recruitment of the Pol II machinery to the viral genome. Whether recruitment of any of these factors is essential to late gene transcription remains to be tested.

Prior studies identified promoter differences between E and L genes, however they focused on select examples from each gene class, which have not been broadly applied. The canonical E gene, thymidine kinase (UL23 or tk), possesses upstream promoter elements (UPEs), namely, two GC-boxes and a CAAT box, which are critical for robust transcription in the absence of DNA replication. Deletion of these promoter elements resulted in tk being expressed with a L1 gene phenotype (161, 162). Other UPE's may be responsible for activation of other E genes, although this remains to be studied.

1.2.3.8 Viral DNA Replication

There are seven essential viral proteins required for HSV-1 DNA replication: DNA polymerase holoenzyme (UL30/UL42), helicase-primase complex (UL5/UL8/UL52), single stranded DNA (ssDNA) binding protein (ICP8), and origin binding protein (UL9). Excepting UL9 these six components are conserved in herpesviridae. The production of concatemeric viral genomes supports a rolling circle model of synthesis (163), however a recombination based

mechanism may also be viable (164-167). We will describe what is currently known about the mechanism of viral DNA replication (Fig. 8).

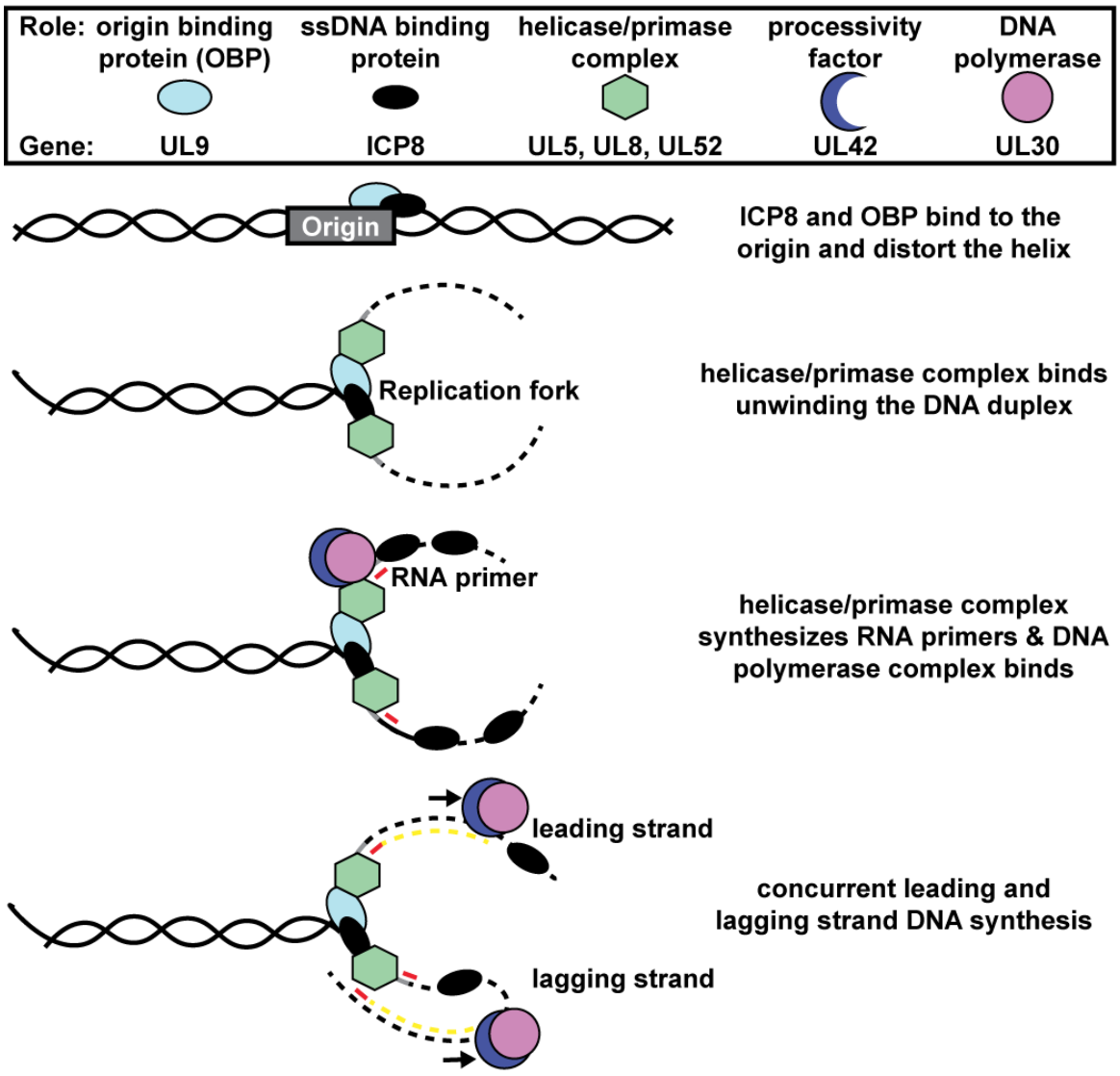


Figure 8 Model of HSV-1 DNA replication

First, the origin binding protein (OBP) and ssDNA binding protein bind to the origin of replication and distort the double helix (168, 169). Current lines of evidence suggest that based on

the rate of DNA replication, multiple origins must be firing simultaneously during replication (164). The HSV-1 genome contains three origins of replication (ori): OriS which is present in two copies and located in the repeat regions adjacent to ICP4, and OriL which is located in the UL region between UL29 and UL30 (170, 171). Both origins contain an AT-rich region flanked by binding sites for the origin binding protein (OBP). The OBP has been demonstrated in vitro to function as a nucleoside triphosphatase, DNA helicase, and nonspecific ssDNA binding protein (166, 167). The OBP also possesses sequence specific binding for repeated GTTCGCAC sites, contained within the viral origins (172). ICP8 has been shown to preferentially bind ssDNA, and to a lesser degree duplex DNA and single stranded oligos (173, 174). ICP8 has also been shown to destabilize double helixes (175, 176). Experiments using temperature sensitive (ts) mutants demonstrated that the OBP is not continuously required for ongoing DNA replication. This indicates that at some point DNA replication likely becomes origin independent (177).

Second, the helicase-primase complex is recruited to the distorted helix of the origin. Studies determined that the OBP interacts with UL8, and ICP8 contacts UL5, UL8, and UL52. These contacts facilitate recruitment of the helicase-primase complex to the distorted DNA helix. Once bound to the origin, the helicase-primase complex further unwinds the duplex DNA and synthesizes RNA primers at the nascent replication fork (166, 167). In vitro assays suggest that the helicase primase complex must be present as a dimer or higher-order structure for optimal primase activity (178).

Third, the viral polymerase holoenzyme binds to the replication fork where it catalyzes concurrent leading and lagging strand synthesis. Recruitment of the DNA polymerase requires that the helicase primase complex have actual primase activity, suggesting that either the active conformation of the helicase primase complex or the actual RNA primer itself are critical for

efficient recruitment. Recent work has demonstrated that the DNA polymerase is capable of catalyzing concurrent leading and lagging strand synthesis (179).

There are additional viral factors that function in DNA synthesis, including a thymidine kinase (UL23), ribonucleotide reductase (UL39/UL40), deoxyuridine triphosphatase (UL50), alkaline nuclease (UL12), and uracil-DNA glycosylase (UL2). However, none of these genes are essential for either DNA replication or late gene transcription in cell culture (167).

1.2.3.9 Packaging and Assembly

Production of infectious progeny is detectable starting around 6 hpi. Capsid assembly and genome packaging occur within the nucleus. The capsid is composed of the six proteins UL6, UL18, UL19, UL26, UL35, and UL38 as described in 1.2.1. DNA packaging occurs in pre-assembled viral capsids. The viral genes UL6, UL15, UL17, UL25, UL28, UL32, and UL33 are all essential for DNA packaging (180). Among these genes is the portal protein found at one vertex of the capsid, which is composed of 12 copies of UL6. The terminase complex, UL15, UL28, and UL33, is responsible for cleaving the viral genome sequence specifically. UL17 and UL25 are components of the CATC complex and are critical for capsid assembly and structure during packaging (46). Data suggests that UL32 may play a role in ensuring capsids localize to the appropriate nuclear region for packaging (181).

Viral DNA replication produces conjoined genomes that need to be cleaved during packaging so that mature virions contain one copy of the linear genome. Two specific cleavage signals, pac1 and pac2 are contained within the tandem repeat regions, or ‘a’ sequence (182). DNA encapsidation likely starts at the UL end of the genome and finishes at the US end (93). The terminase complex is responsible for recognizing the cleavage signals and ensuring that a single linear copy of the genome is within each capsid.

Recent work found the DNA packaging and capsid assembly components UL17, UL18, UL19, UL25, UL26, and UL38 enriched away from nascent replication forks in viral genomes. This is consistent with the observation that active replication is not required for DNA packaging (183). In contrast UL6, the portal protein, was one of the most abundant proteins found at the replication fork of nascent viral genomes (164). This result may be indicative of a role for UL6 in directing the viral genome to the capsid for packaging.

Envelopment is still controversial, as the virion must exit the nucleus, acquire a glycoprotein studded envelope and finally be released from host cells. We will discuss the most likely model. First the mature capsid buds through the nuclear membranes where it is enveloped and then de-enveloped. This is supported by the envelope composition of mature virions, which best matches the lipid composition of the TGN or endosomes (48, 49). After exiting the nucleus the virus buds through the TGN network to acquire its final envelope studded with viral glycoproteins. Tegument proteins are likely acquired in the cytosol prior to envelopment in the TGN. Finally, enveloped virions exit the cell via exocytosis, mimicking how the TGN membrane is recycled and fused with the plasma membrane (184).

1.3 Rationale

Viruses have evolved unique mechanisms to invade hosts, alter cellular pathways, and redirect host factors for viral processes. These mechanisms are only beginning to emerge as recent technical advancements provide unbiased, comprehensive tools to investigate viral genome-protein interactions. We investigated how herpes simplex virus type 1 (HSV-1) alters the nuclear environment to manipulate host processes, namely the Pol II transcription machinery. We found

that HSV-1 utilizes specific promoter elements and trans-acting viral factors to promote robust viral transcription, at the expense of the host. These virus-host interactions are essential to facilitate the rapid pace of HSV-1 productive infection.

In Chapter 2 we will discuss how viral DNA replication drastically alters the transcriptional landscape and kinetically delineates viral transcription.

In Chapter 3 we will discuss how ICP4 interacts with host GTF's and contributes to a shift in favor of viral transcription at the expense of the host.

2.0 Genome replication affects transcription factor binding mediating the cascade of Herpes Simplex Virus transcription

2.1 Project Summary

In herpes simplex virus type 1 (HSV-1) infection, the coupling of genome replication and transcription regulation has been known for many years; however, the underlying mechanism has not been elucidated. We performed a comprehensive transcriptomic assessment and factor-binding analysis for Pol II, TBP, TAF1, and SP1 to assess the effect genome replication has on viral transcription initiation and elongation. The onset of genome replication resulted in the binding of TBP, TAF1, and Pol II to previously silent late promoters. The viral transcription factor, ICP4, was continuously needed in addition to DNA replication for activation of late gene transcription

initiation. Furthermore, late promoters contain a motif that closely matches the consensus initiator element (Inr), which robustly bound TAF1 post-replication. Continued DNA replication resulted in reduced binding of SP1, TBP, and Pol II to early promoters. Therefore, the initiation of early gene transcription is attenuated following DNA replication. Herein, we propose a model for how viral DNA replication results in the differential utilization of cellular factors that function in transcription initiation, leading to the delineation of kinetic class in HSV-productive infection.

2.2 Importance

Nuclear-replicating DNA viruses employ a common strategy of coupling late gene transcription to genome replication. The mechanism by which genome replication facilitates nascent transcription is unknown. We performed RNA-Seq and ChIP-Seq for Pol II, TBP, TAF1, and SP1 to quantitatively assess transcription initiation and elongation across the viral genome. The onset of replication permanently altered genomic accessibility and, with the aid of ICP4, facilitated binding of initiation factors to previously silenced promoters. Continued replication precluded binding of transcription factors to promoters active prior to DNA replication. The results provide a model for the mechanism of the transcription switch in HSV infection, which may be applicable to other nuclear-replicating DNA viruses.

2.3 Introduction

The genome of Herpes simplex virus I (HSV-1) is 152 kb of linear double-stranded DNA, containing three origins of replication and over 90 unique ORFs (170, 171, 185, 186). Genome replication and transcription occur in the host cell nucleus. HSV-1 encodes its own genome replication machinery (187), and viral gene products augment the RNA Polymerase II (Pol II) machinery of the cell for the transcription of the genome (117). Viral transcription is incredibly robust, allowing a single infecting virus to produce progeny between 4 and 6 h postinfection, culminating in ~1,000 infectious progeny per cell within 18 h.

The HSV-1 transcriptional cascade was originally defined by adding cyclohexamide (123), canavanine (124), and phosphonoacetic acid (125) to infected cells and assessing peptide and, later, RNA accumulation. These experiments defined immediate early (IE or α), early (E or β), and late (L or γ) genes as those expressed in the absence of de novo viral protein synthesis, in the presence of functional alpha proteins before DNA replication, and after genome replication, respectively. Late genes were further delineated as leaky late (L1) and true late (L2) based on whether genome replication increases the rate of synthesis or initiates synthesis. α gene promoters possess promoter elements, which bind the viral tegument protein, VP16, activating their transcription (131, 132, 140, 141). ICP4 is the product of an α gene that is required for the transcription of viral E and L genes (143, 145, 188). A canonical E gene, thymidine kinase (UL23 or tk), contains upstream promoter elements (UPEs), namely, two GC-boxes and a CAAT box (161), which bind the cellular transcription factors SP1 and NF1, respectively (161, 162). A prototypic L2 gene, glycoprotein C (UL44 or gC), consists of a TATA box and an initiator (Inr) sequence located near the start site of transcription, the latter of which is important for robust

transcription and ICP4 activation (189-192). How this gene architecture is coordinated with genome replication to produce the observed cascade of gene expression is unknown.

We wanted to determine the transcriptional changes associated with viral genome replication. We report a comprehensive assessment of how genome replication alters the viral transcriptional landscape and facilitates a shift in transcript production. Our results lead us to propose a model of the viral transcriptional cascade which provides insight into a mechanism used by DNA viruses to coordinate efficient virion production.

2.4 Results

2.4.1 Dependence of Late Viral Transcription on Genome Replication

To investigate the coupling of viral transcription to replication, we carried out RNA-Seq in Vero cells infected with wild-type HSV-1 (strain KOS). An inhibitor of viral genome replication, acyclovir (ACV), was added at 0, 2, 3, 4, or 6 hpi. DNA and RNA was harvested at 12 hours post infection (hpi) to compare the relative abundance of viral genomes and transcripts as a function of time of inhibition of viral DNA synthesis. The control was harvested at 12 hpi without the addition of ACV.

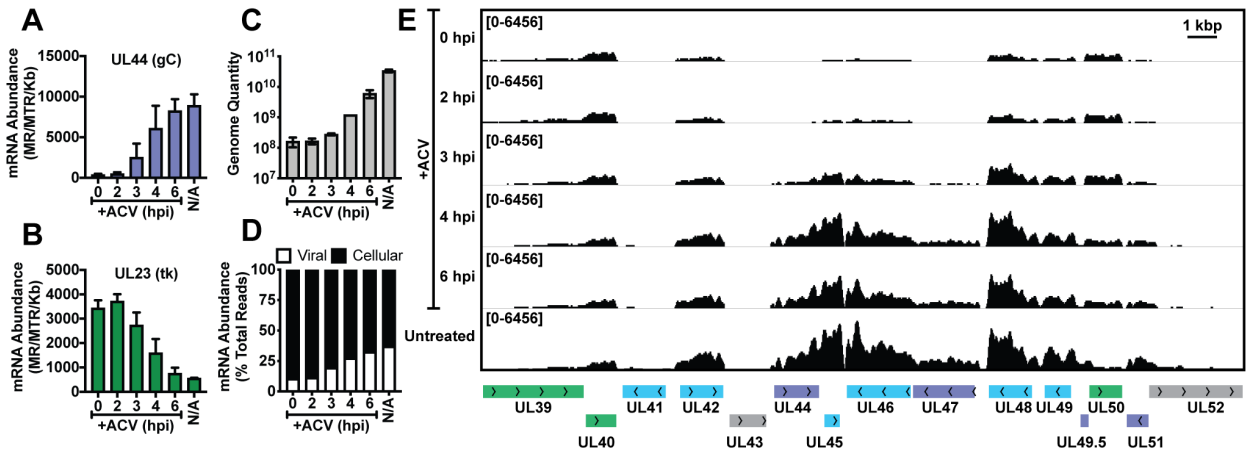


Figure 9 Initial rounds of replication are sufficient to alter late and early gene transcription

Vero cells infected with HSV-1 and treated with acyclovir (ACV) at 0, 2, 3, 4, or 6 hpi. Error bars are the standard deviation of biological duplicates. A-B, D-E. PolyA-selected RNA-Seq. A-B. mRNA abundance of UL44 and UL23, y-axes values are mapped reads per million total reads per kilobase (MR/MTR/Kb). D. Reads mapped to the viral or host genome as a percentage of total reads. C. Viral genomes produced by 2 million infected cells, quantified by qPCR. E. Reads mapped to a region of the HSV-1 genome, y-axes values are MR/MTR. Viral coding sequence (CDS) are indicated, color coded by gene class with E as green, leaky late (L1) as blue, and true late (L2) as purple.

When ACV was added prior to the onset of genome replication, 0 and 2 hpi, we observed no accumulation of the true late gene UL44 (glycoprotein C, gC) (Fig. 9A). By comparison the canonical early gene UL23 (thymidine kinase, tk) reached the greatest abundance in these samples (Fig. 9B). When ACV was added at 3 hpi we measured approximately one round of genome replication (Fig. 9C), which coincided with a doubling in the abundance of viral transcripts, from 11% to 20% of total mRNA (Fig. 9D). Consistent with these data, we observed transcription of gC (Fig. 9A) and other true late genes (Fig. 9E, Supplemental Fig. 1) in samples treated with ACV at 3 hpi. Additional genome replication resulted in increased accumulation of leaky and true late

transcripts (Supplemental Fig. 1), consistent with Jones and Roizman, 1979. These data show that continuing rounds of genome replication are not required for transcription of viral true late genes. Thus the initial rounds of viral DNA replication are sufficient to alter the transcriptional landscape of HSV-1 genomes enough to activate late gene and alter early gene transcription.

2.4.2 Classification of Viral DNA Replication-Dependent Transcripts.

We performed a global analysis to obtain a data set of genes in each class as a prelude to determining what molecular features distinguish early and late genes. We infected human fibroblast (MRC5) cells with wild-type HSV-1 (KOS) and a UL30 C-terminal truncation mutant, ΔC1216. UL30 encodes the viral DNA-dependent DNA polymerase (193). The C-terminal tail of UL30 was shown to be essential for interaction with the processivity factor, UL42 (194). In the absence of this interaction, HSV genome replication is inhibited (195). We assessed genome replication and transcriptional activity for this mutant and wild-type virus.

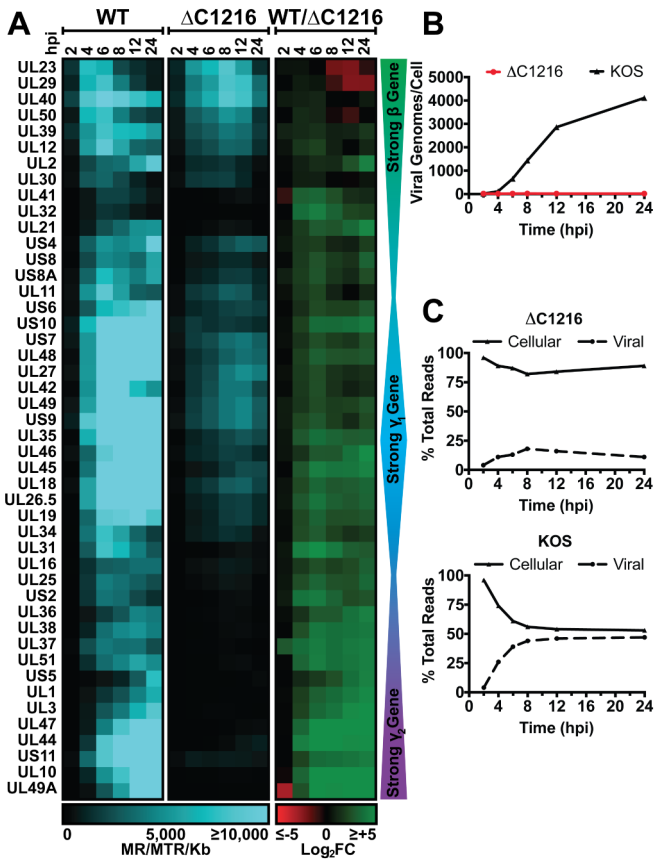


Figure 10 Replication dependence of viral transcripts.

MRC5 cells were infected with WT HSV-1 and ΔC1216. (A) PolyA-selected RNA-Seq data as mapped reads per million total reads per kilobase (MR/MTR/KB) or log_2 fold-change of WT over ΔC1216. (B) Number of viral genomes per cell. (C) RNA-Seq reads mapping to the viral or cellular genome as a percentage of total reads.

As expected, there was no genome replication in ΔC1216-infected cells (Fig. 10B). Consistent with previous studies, thymidine kinase and single-stranded DNA-binding protein (UL29) accumulated to higher levels in ΔC1216-infected cells. By the end of the time course almost all other viral transcripts were produced more robustly in wild-type infection, reaching 50% of total RNA sampled, compared with only ~12% in ΔC1216 (Fig. 10C). We clustered viral

transcripts based on their accumulation in the presence and absence of genome replication (Fig. 10A). Twenty-three genes were excluded from classification, as their reads were present at very low levels (<2,000 MR/MTR/KB) in all samples analyzed. Forty-six viral genes were classified strictly as either E, L1, or L2. As expected, the E gene class encompassed proteins with a role in genome replication and maintenance. All other transcripts were classified as dependent on genome replication. These more closely defined gene classes (Table 1) were used for all further analyses.

Table 1 Classification of viral genes bases on DNA replication dependence

Gene Class	Genes	# Genes
α	RL2, UL54, RL2_1, RS1, US1, US12, RS1_1	7
β	UL2, UL12, UL23, UL29, UL30, UL39, UL40, UL50	8
γ_1	UL11, UL18, UL19, UL21, UL26.5, UL27, UL32, UL34, UL35, UL41, UL42, UL45, UL46, UL48, UL49, US4, US6, US7, US8, US8A, US9, US10	22
γ_2	UL1, UL3, UL10, UL16, UL25, UL31, UL36, UL37, UL38, UL44, UL47, UL49A, UL51, US2, US5, US11	16
Weak transcription (no class)	RL1, UL4, UL5, UL6, UL7, UL8, UL9, UL13, UL14, UL15, UL17, UL20, UL22, UL24, UL26, UL28, UL33, UL43, UL52, UL53, UL55, UL56, US3	23

2.4.3 Viral Replication Facilitates Preinitiation Complex Formation on Previously Silent Promoters.

We infected MRC5 cells with wild-type HSV-1 in the presence or absence of an inhibitor of viral replication, ACV. Infected cells were harvested at 4 hpi, and ChIP-Seq was performed for Pol II, TATA-binding protein (TBP), and transcription initiation factor TFIID subunit 1 (TAF1). Using these data, we were able to assess transcription factor dynamics in conditions optimal for either E or L gene transcription. ChIP-Seq reads mapped to previously characterized viral promoters demonstrated that each factor was positioned relative to its respective binding element (Supplemental Fig. 2).

To quantitatively compare inhibited (+ACV) and uninhibited (−ACV) samples we had to account for viral genome replication. Quantification of input samples provided the relative amounts of viral genomes present in each condition (~30-fold higher in −ACV) (Fig. 11A).

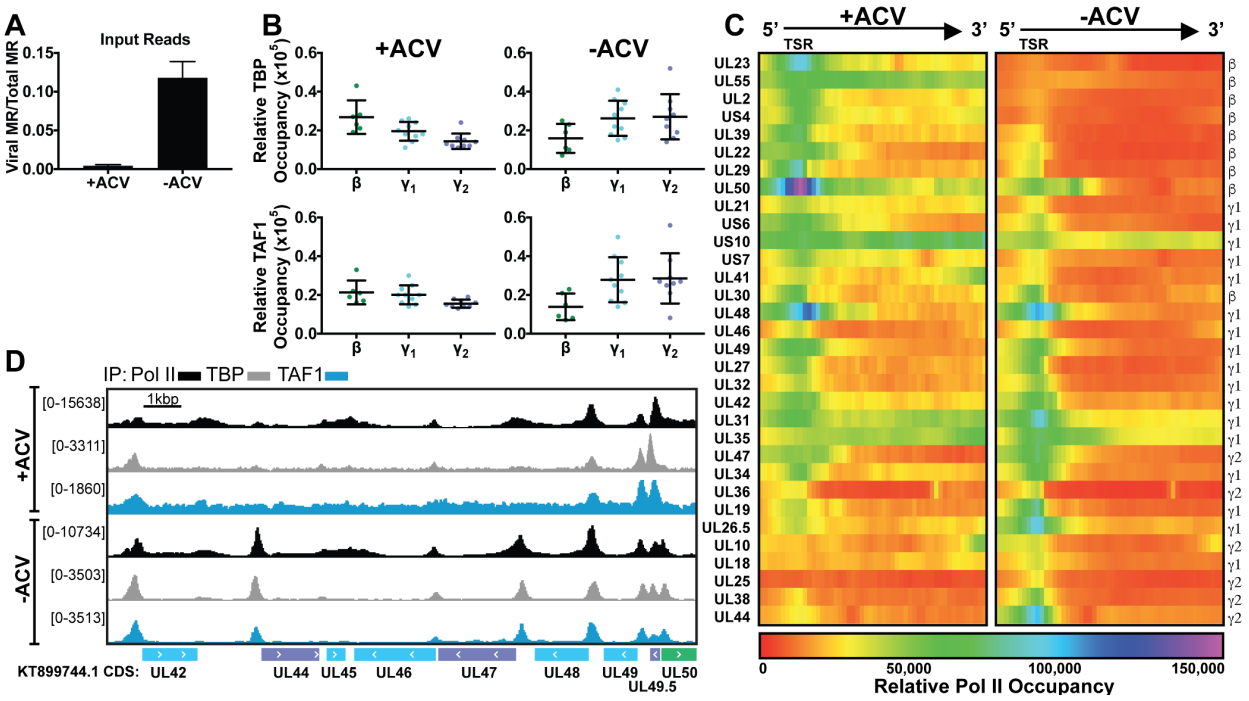


Figure 11 Initiation complex formation as a function of genome replication

MRC5 cells were infected with HSV-1 in the presence (+ACV) or absence (−ACV) of acyclovir and ChIP-Seq for TBP, TAF1, and Pol II was performed. All data are normalized to the amount of factor per viral genome ("Relative Occupancy"). (A) Viral genomes quantified as input reads mapped to the viral genome per total cellular and viral mapped reads (Viral MR/Total MR). Error bars represent SD. (B) TBP and TAF1 promoter occupancy. Each point is a distinct promoter. (C) Pol II occupancy heat maps for viral loci. Transcription start region (TSR) indicates the main Pol II promoter peak. (D) Overlay of Pol II (black), TBP (gray), and TAF1 (blue) binding mapped to viral UL42-UL50 loci.

We used this ratio to normalize each immunoprecipitated (IP) sample for the amount of factor per genome. We call this measurement relative factor occupancy and quantified all ChIP-Seq data in the same way. To assess preinitiation complex formation, we quantified relative factor occupancy on viral promoters. In the absence of DNA replication (+ACV) there was a complete absence of TBP, TAF1, and Pol II on L2 gene promoters. By comparison, in the presence of DNA replication (−ACV) there was a severe reduction in TBP, TAF1, and Pol II on E gene promoters (Supplemental Fig. 3). As expected, L1 genes displayed an intermediate phenotype, with initiation factors bound in both conditions but a greater amount in the presence of replication (Fig. 11B-C). This trend is quite visible when looking at UL50, UL48, and UL44 as strong examples of E, L1, and L2 genes, respectively (Fig. 11D). These results suggest DNA replication alters promoter accessibility to cellular transcription factors involved in the initiation of transcription.

2.4.4 A Single Duplication Alters Genomic Accessibility.

Next, we assessed the extent of genome replication required to alter viral transcription dynamics. This was to determine whether the transcriptional shift was due to an immediate cis-effect of DNA replication or an increase in genome number titrating out a repressive factor. MRC5 cells were infected with wild-type HSV-1 for 2, 3, and 4 hpi, and ChIP-Seq for Pol II, TBP, and TAF1 was performed (Supplemental Fig. 4). HSV-1 genome replication starts between 2.5 and 3 hpi, so each time point represents a different replication state: 2 hpi (pre-replication), 3 hpi (1–2 genome duplications), and 4 hpi (3–4 genome duplications) (Fig. 12A).

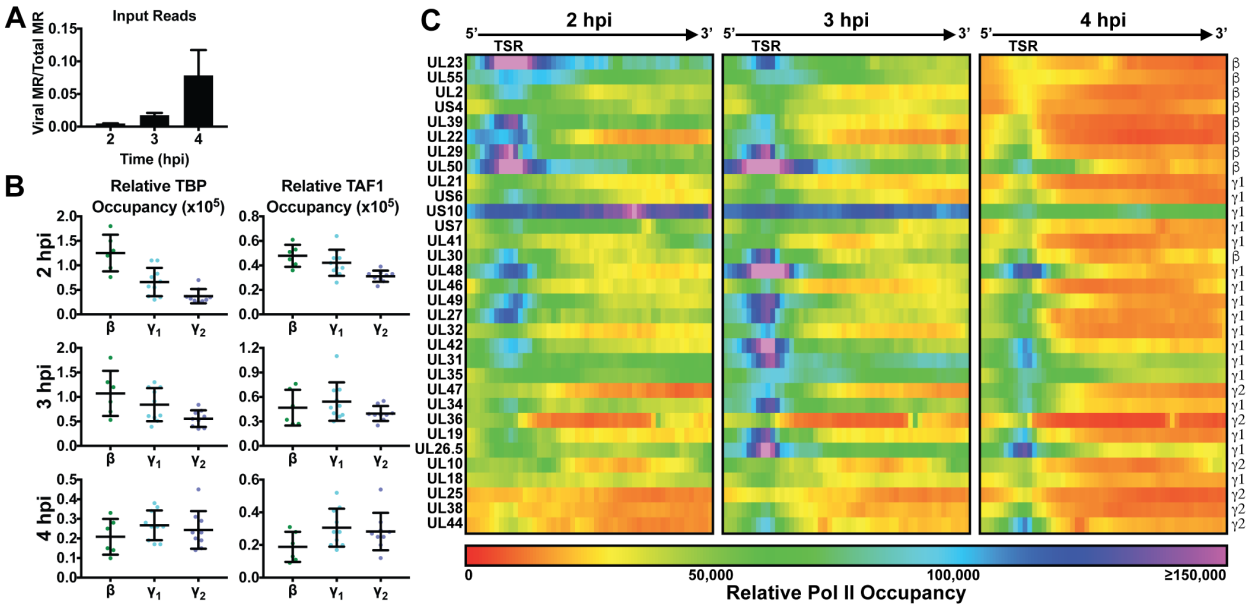


Figure 12 A single round of genome replication immediately alters cellular initiation factor binding

MRC5 cells were infected with HSV-1 for 2, 3, or 4 h, and ChIP-Seq for TBP, TAF1, and Pol II was performed. All data are normalized to the amount of factor per viral genome ("Relative Occupancy"). (A) Viral genomes quantified as input reads mapped to the viral genome per total cellular and viral mapped reads (Viral MR/Total MR). Error bars represent SD. (B) TBP and TAF1 promoter occupancy. Each point is a distinct promoter. (C) Pol II occupancy heat maps for viral loci. Transcription start region (TSR) indicates the main Pol II promoter peak.

Consistent with the results of Fig. 11, there was a complete absence of transcription factors on L2 promoters prereplication and a severe reduction of transcription factors on E promoters post replication (Fig. 12B-C). L2 promoter occupancy initiated between 2 and 3 hpi. These data suggest genome replication had an immediate cis-effect on the accessibility of L2 gene promoters to transcription factors. Conversely, diminished E gene promoter occupancy was not observed until 4 hpi. Thus, genome replication must immediately alter the form or state of the viral genome. This alteration resulted in a global increase in viral promoter accessibility to cellular factors.

2.4.5 SP1 Preferentially Binds to the Viral Genome Pre-replication.

SP1 has been previously shown to bind and activate the viral E gene, thymidine kinase (161, 162). To investigate if SP1 is globally responsible for activation of E genes we performed ChIP-Seq for SP1 on MRC5 cells infected with wild-type HSV-1 for 2 and 4 h.

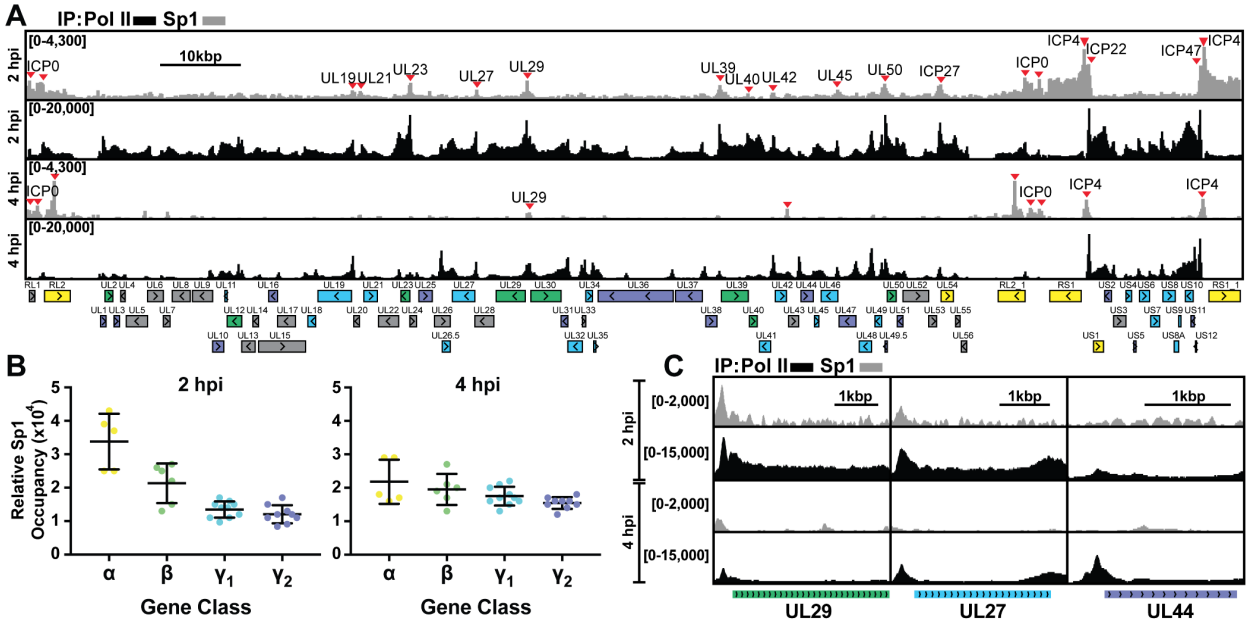


Figure 13 Genome replication alters the binding of SP1 to viral promoters

MRC5 cells were infected with HSV-1 for 2 or 4 h, and ChIP-Seq for Pol II and SP1 was performed. All data are normalized to the amount of factor per viral genome ("Relative Occupancy"). (A) Pol II (black) and SP1 (gray) binding mapped to the viral genome. Distinct SP1 peaks were highlighted with red arrows; if the peak was located within a viral promoter the gene is annotated above. (B) SP1 promoter occupancy. Each point is a distinct promoter. (C) Pol II (black) and SP1 (gray) binding mapped to UL29 (E gene), UL27 (L1 gene), and UL44 (L2 gene).

Pre-replication we observed numerous distinct SP1 promoter peaks binding to 100% of IE genes, 71% of E genes, 22% of L1 genes, and 0% of L2 genes (Fig. 13A-B). We noted a severe reduction in SP1 binding at 4 hpi, with only the IE genes ICP4 and ICP0 retaining strong binding (Fig. 13A-B). We noted that promoters bound by SP1 had decreased Pol II occupancy between 2 and 4 hpi, while promoters always lacking an SP1 peak did not experience this decrease (Fig. 13C). This global decrease in SP1 binding at 4 hpi may be due to a previously identified phosphorylation event of SP1 occurring at the same time during infection (196). Alternatively, the amount of SP1 in the cell may be limiting, or the very act of genome replication precludes SP1 binding.

2.4.6 Differential TFIID Promoter Context Preferred After Replication.

We observed variation in the relative amplitude and shape of TAF1 peaks, with early genes exhibiting weak, broad binding. Genes with a weak TAF1 peak include UL23, UL39, UL50, UL2, and UL12. These promoters still possessed distinct, strong TBP and Pol II binding. To assess whether this binding phenotype was sequence-specific we performed motif discovery to predict TBP- and TAF1-binding motifs. We then determined consensus TATA boxes and Inr elements for each gene class (Supplemental Table 1). We noted little difference in TATA box motifs between the different gene classes (Fig. 14A).

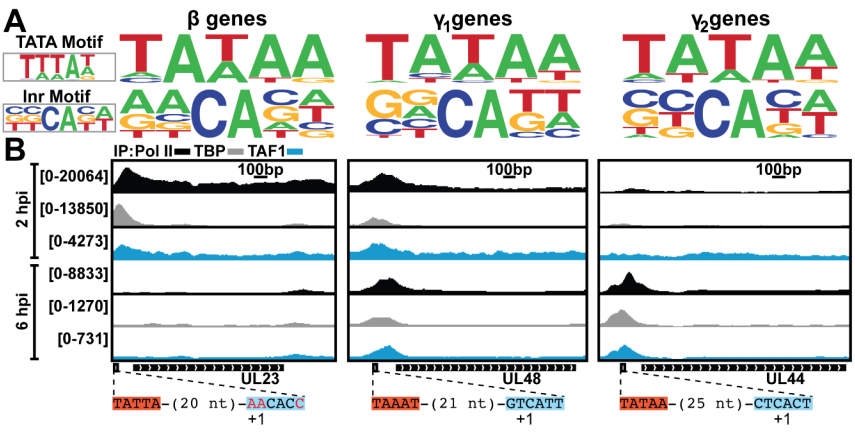


Figure 14 Analysis of TFIID-binding sites in E and L promoters

MRC5 cells were infected with HSV-1 for 2 or 6 h, and ChIP-Seq for Pol II, TBP, and TAF1 was performed. All data are normalized to the amount of factor per viral genome. Probability plot of (A) TATA box and Initiator elements delineated by gene class. Human consensus is given in gray box. (B) Overlay of Pol II (black), TBP (gray), and TAF1 (blue) binding mapped to viral loci. Viral core promoters are outlined, with the TATA box in orange and Inr in blue. Deviations from motif consensus are highlighted in red. Experimentally determined transcription start sites are indicated as "+1" (197-199).

On an individual gene basis, TATA boxes matching the consensus, TATAW, were more robustly transcribed. There was a distinct deviation in Inr motifs between the gene classes, with L genes more closely matching the consensus sequence BBCA$_{+1}$BW (200) (Fig. 14A). Unlike the consensus motif, E genes favored adenosine at positions 1 and 2. Promoters containing strong Inr, such as UL44 and UL48, were more robustly transcribed post-replication (Fig. 14B). Genes without a canonical Inr sequence had weak and broad TAF1 peaks. We posit that TAF1 is present in TFIID for all viral promoters; however, the stronger the Inr element, the increased affinity of TAF1 for the Inr, the stronger the cross-linking, and subsequent peak.

The TAF1 ChIP-Seq data in pre-replication conditions are indicative of a high level of nonspecific binding (Supplemental Fig. 3-4). As this was consistently present between biological duplicates and only present in 2 hpi or +ACV samples we suspect that this is not an artifact. Because this was not seen with TBP, this signal may represent TAF1 binding independent of TFIID at this time. This remains to be studied.

2.4.7 Continuous Requirement for ICP4 in Viral Transcription.

ICP4 is the major viral transcriptional protein and is required for production of E genes (145). We wanted to determine whether ICP4 is always critical for viral transcription, even after L transcription has been licensed by genome replication. We utilized a temperature-sensitive mutant of ICP4 (tsKos), in which a single point mutation makes the protein unstable and incapable of DNA binding at elevated temperature. MRC5 cells were infected with tsKos and incubated either at the permissive temperature (P), shifted from permissive to nonpermissive temperature at 4 hpi (S), or nonpermissive temperature (N).

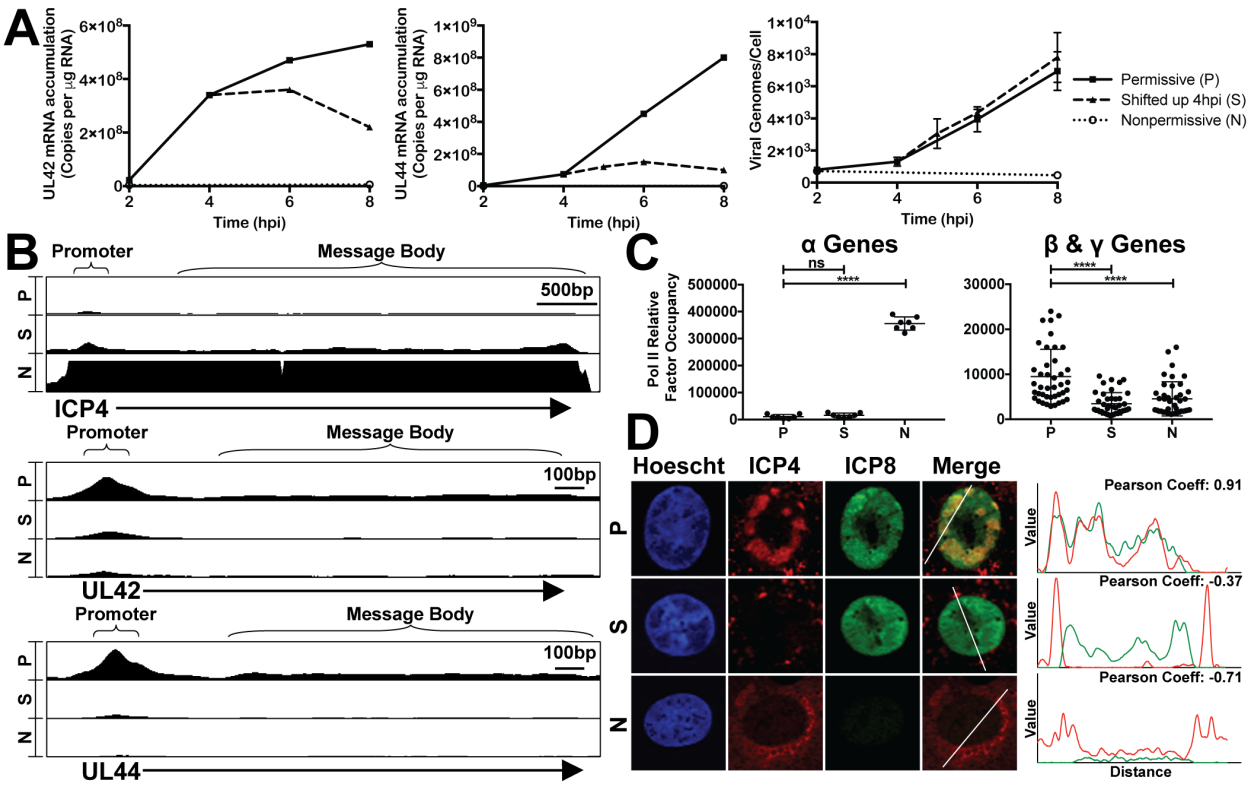


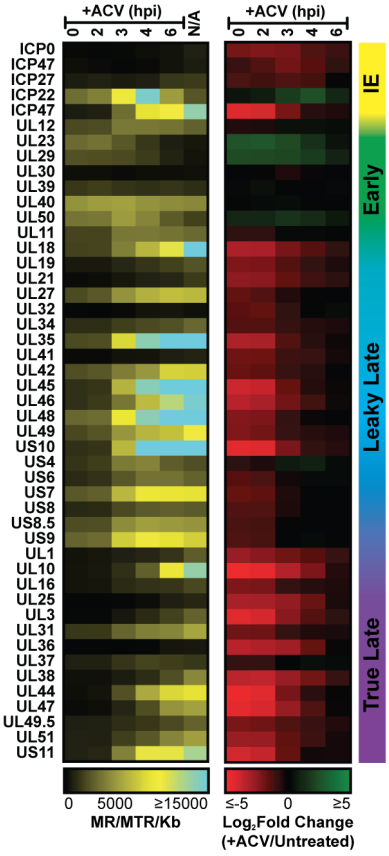
Figure 15 ICP4 is continuously required for transcription after the onset of genome replication

MRC5 cells were infected with tsKos and grown at permissive conditions (P), shifted up from permissive to nonpermissive conditions at 4 hpi (S), or nonpermissive conditions (N). (A) UL42 (L1 gene) and UL44 (L2 gene) mRNA copies per microgram total cellular RNA. Number of viral genomes per cell. Error bars represent the SD of biological duplicates. (B and C) ChIP-Seq for Pol II was performed on P, S, and N samples at 6 hpi. All data are normalized to the amount of factor per viral genome. (B) Pol II traces mapped to ICP4 (IE gene), UL42 (L1 gene), and UL44 (L2 gene) loci. y axes are 0–7500. (C) Pol II promoter occupancy. Each point is a distinct promoter. (D) Images of infected Vero cultures grown for 6 hpi. Pearson correlation test was performed on the red and green intensity profiles.

Immunofluorescence of infected samples at 6 hpi demonstrates an absence of nuclear ICP4 in shifted and nonpermissive conditions (Fig. 15D). We measured viral genome replication and mRNA accumulation for two genes robustly transcribed from 4 to 8 hpi (Fig. 15A). Infected cells

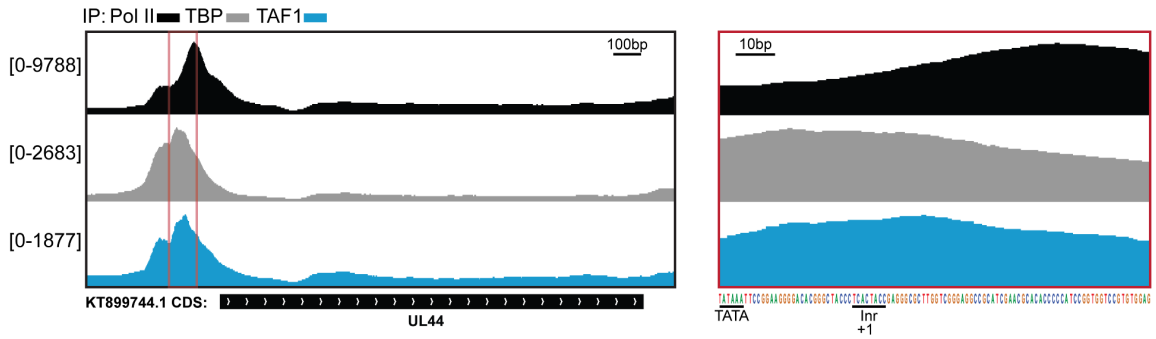
grown entirely at nonpermissive temperature lacked genome replication and transcription. Infected cells grown entirely at permissive temperature underwent genome replication and transcription similar to wild-type kinetics. Infected cells shifted from permissive to nonpermissive temperature at 4 hpi immediately halted transcription after the shift. By comparison, genome replication continued after shift at a similar rate to samples grown at permissive temperature. Thus, even after the onset of DNA replication, ICP4 was essential for continued transcription. To determine which transcriptional stage ICP4 controls we performed Pol II ChIP-Seq on permissive samples, shifted up at 4 hpi, and nonpermissive samples harvested at 6 hpi (Fig. 15B-C). In shifted-up cells there was a complete absence of Pol II bound to viral promoters of E and L genes. Similarly, in nonpermissive conditions, Pol II was absent from viral promoters, with the exception of IE genes. Thus, ICP4 is critical for formation of a preinitiation complex and is continuously required for robust transcription of E and L genes.

2.5 Supplemental Figures



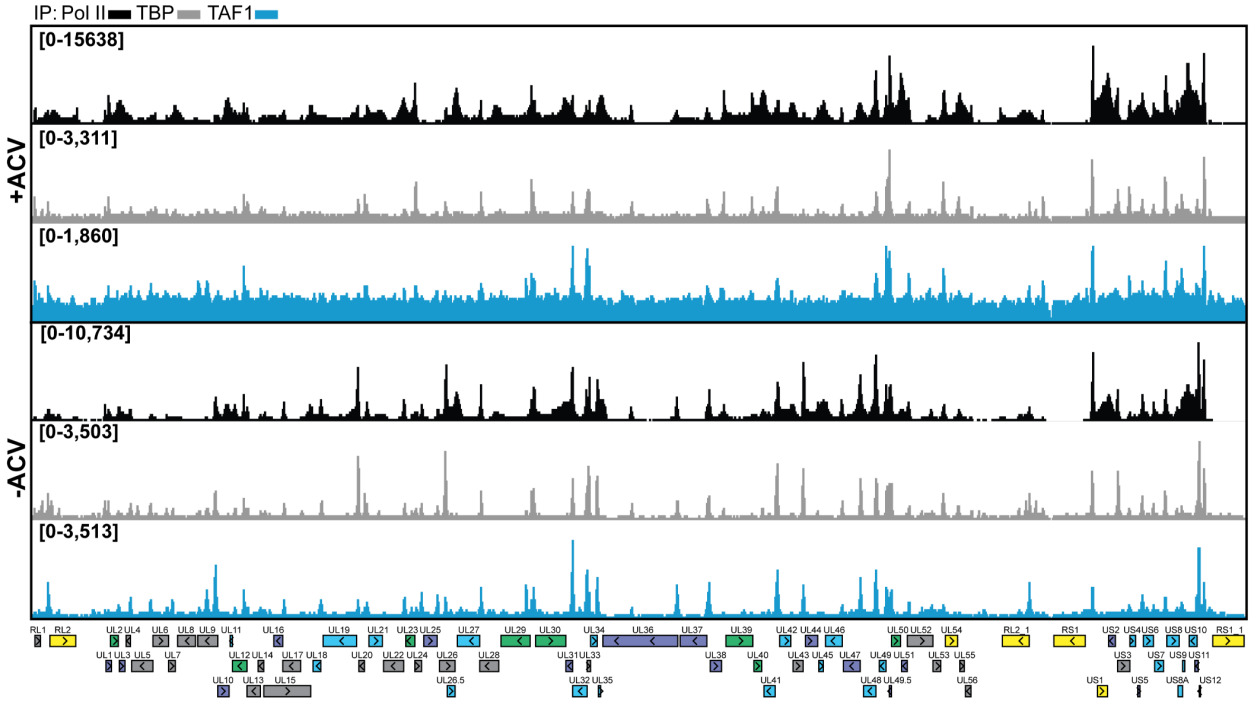
Supplemental Figure 1 Viral transcription kinetics as a function of genome replication inhibition

Vero cells were infected with wild-type HSV-1 at an MOI of 10 PFU/cell and acyclovir (ACV) was added at 0, 2, 3, 4, or 6 hpi. Total RNA was collected at 12 hpi and polyA-selected RNA-Seq was performed. Data is the average of biological duplicates. Values plotted are mRNA abundance, mapped reads per million total reads per kilobase (MR/MTR/Kb), or the $\log_2$ fold change of samples treated with ACV over the control untreated sample.



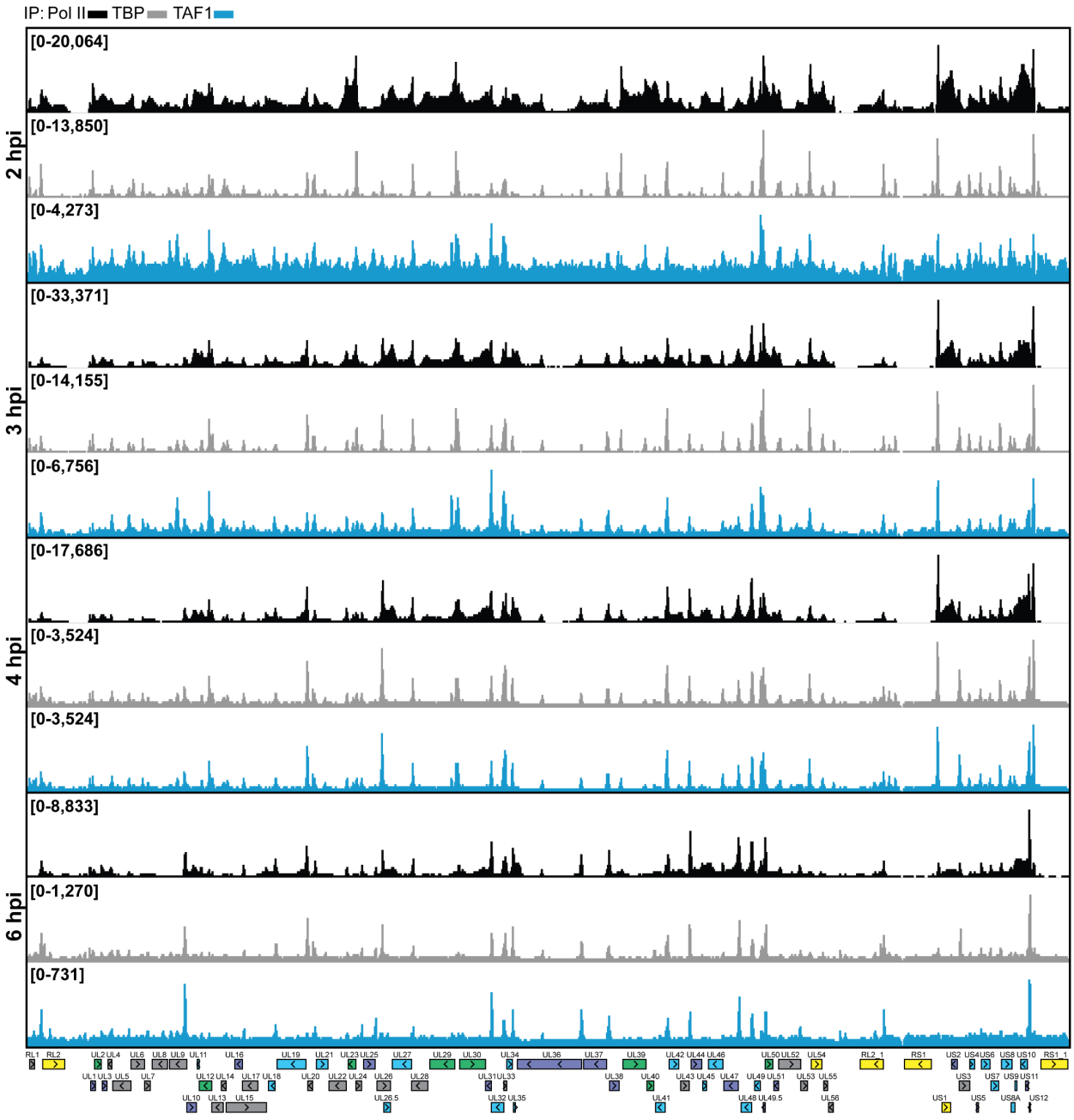
Supplemental Figure 2 General transcription factors bind their respective binding elements

MRC5 cells were infected with HSV-1 for 4 hours and ChIP-Seq for Pol II, TBP, and TAF1 was performed. All data is normalized to the amount of factor per viral genome, represented as relative factor occupancy. Data is the average of two biological replicates. Traces mapped to the UL44 locus. On the right is a zoomed in promoter view with the TATA box and Initiator element (Inr) underlined, the transcript starts at the first C of the Inr element.



Supplemental Figure 3 Transcription factor binding in presence and absence of viral genome replication

MRC5 cells were infected with HSV-1 in the presence (+ACV) or absence (-ACV) of acyclovir for four hours and ChIP-Seq for TBP, TAF1, and Pol II was performed. All data is normalized to the amount of factor per viral genome, represented as relative factor occupancy. Data is the average of two biological replicates. Viral genes are annotated below, color coded for gene classes with IE (yellow), E (green), L1 (blue), and L2 (purple).



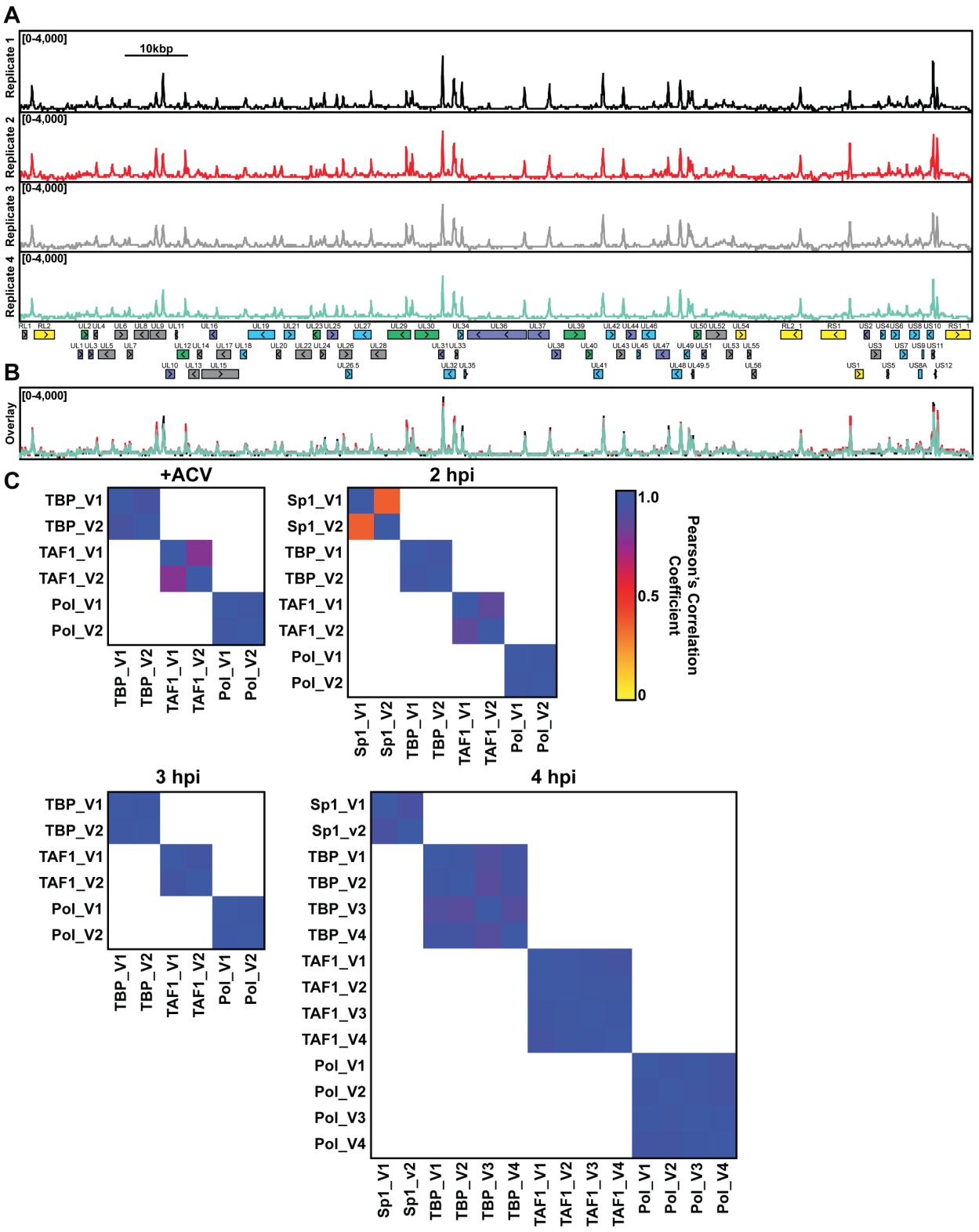
Supplemental Figure 4 Transcription factor binding at different stages of genome replication

MRC5 cells were infected with HSV-1 for 2, 3, 4 or 6 hours and ChIP-Seq for TBP, TAF1, and Pol II was performed. All data is normalized to the amount of factor per viral genome, represented as relative factor occupancy. Data is the average of two biological replicates. Viral genes are annotated below, color coded for gene classes with IE (yellow), E (green), L1 (blue), and L2 (purple).

Supplemental Table 1 Motif Discovery

Annotated positions and sequences of TATA boxes and Initiator elements (relative to KT899744.1 reference).

			TATA Box			Inr Element		
Gene	**Gene Class**	**Strand**	**Sequence**	**Start**	**End**	**Sequence**	**Start**	**End**
UL1	Gamma 2	Sense	tatat	9064	9068			
UL2	Beta	Sense	tataa	9562	9566	gacagt	9586	9591
UL3	Gamma 2	Sense	aataa	10793	10797	ctcata	10824	10829
UL4	N/A	Antisense	tatta	12334	12338	tccacc	12308	12313
UL7	N/A	Sense	tttaa	16777	16781	gacaga	16806	16811
UL10	Gamma 2	Sense	ttttt	22849	22853	accaca	22875	22880
UL12	Beta	Antisense	tataa	26914	26918			
UL16	Gamma 2	Antisense	taaat	31481	31485	tccaca	31452	31457
UL18	Gamma 1	Antisense	taaaa	36118	36122			
UL19	Gamma 1	Antisense	tatat	40638	40642			
UL21	Gamma 1	Sense	tacta	41695	41699			
UL22	N/A	Antisense	taaaa	46449	46453	ttcata	46421	46426
UL23	Beta	Antisense	tatta	47775	47779	aacacc	47749	47754
UL25	Gamma 2	Sense	ttaag	48442	48446	ctcatt	48469	48474
UL26.5	Gamma 1	Sense	tataa	51458	51462	gacata	51485	51490
UL27	Gamma 1	Antisense	tatat	55963	55967	cacact	55935	55940
UL29	Beta	Antisense	taaag	62208	62212	atcaga	62182	62187
UL30	Beta	Sense	cataa	62438	62442			
UL31	Gamma 2	Antisense	tattt	67345	67349	gccaca	67318	67323
UL32	Gamma 1	Antisense	tctag	69110	69114	cgcagt	69078	69083
UL34	Gamma 1	Sense	tataa	69274	69278	cgcaga	69303	69308
UL35	Gamma 1	Sense	tataa	70318	70322	gccagt	70348	70353
UL36	Gamma 2	Antisense	cataa	80347	80351	ggcaca	80317	80322
UL37	Gamma 2	Antisense	tataa	84081	84085	cgcaac	84051	84056
UL38	Gamma 2	Sense	tttaa	84203	84207	gccagt	84231	84236
UL39	Beta	Sense	taaaa	86022	86026	accaca	86050	86055
UL40	Beta	Sense	tataa	89579	89583	tgcaag	89612	89617
UL41	Gamma 1	Antisense	tataa	92614	92618			
UL42	Gamma 1	Sense	tattt	92737	92741			
UL44	Gamma 2	Sense	tataa	95969	95973	ctcact	95999	96004
UL45	Gamma 1	Sense	tataa	97750	97754	gtcatc	97779	97784
UL46	Gamma 1	Antisense	taaaa	100851	100855	ggcata	100822	100827
UL47	Gamma 2	Antisense	tataa	103167	103171	cgcatt	103138	103143
UL48	Gamma 1	Antisense	taaat	105115	105119	gtcatt	105088	105093
UL49	Gamma 1	Antisense	tataa	106401	106405	gccact	106370	106375
UL50	Beta	Sense	taaaa	106637	106641	gacatt	106664	106669
UL49.5	Gamma 2	Antisense	taaaa	106984	106988	ttcagt	106954	106959
UL51	Gamma 2	Antisense	aataa	109164	109168	gccatt	109134	109139
UL55	N/A	Sense	tataa	115249	115253	gtcata	115278	115283
US3	N/A	Sense	tatat	134760	134764	ggcact	134787	134792
US2	Gamma 2	Antisense	tataa	135155	135159			
US4	Beta	Sense	aaaaa	136531	136535	ggcatc	136564	136569
US6	Gamma 1	Sense	taaaa	138139	138143	gtcata	138167	138172
US7	Gamma 1	Sense	tataa	139499	139503	gacagt	139530	139535
US8	Gamma 1	Sense	tttaa	140953	140957			
US8A	Gamma 1	Sense	ttaaa	142409	142413	tacatt	142441	142446
US9	Gamma 1	Sense	acaaa	143032	143036	cgcagc	143057	143062
US10	Gamma 1	Antisense	taaaa	144964	144968	cccaga	144934	144939
US11	Gamma 2	Antisense	taaaa	145261	145265	gccaga	145232	145237


A
Replicate 1
Replicate 2
Replicate 3
Replicate 4
[0-4,000]
10kbp
B
Overlay
C
+ACV
2 hpi
3 hpi
4 hpi
TBP_V1
TBP_V2
TAF1_V1
TAF1_V2
Pol_V1
Pol_V2
Sp1_V1
Sp1_V2
Sp1_v2
TBP_V3
TBP_V4
TAF1_V3
TAF1_V4
Pol_V3
Pol_V4
1.0
0.5
0
Pearson's Correlation Coefficient

Supplemental Figure 5 Synchronicity of productive viral infection

A) Example replicate TAF1 ChIP-Seq data for KOS at 4 hpi. B) Overlay of traces from part A. C) Normalized bigwig files were analyzed in 50 bp bins using multiBigwigSummary. These bins were compared between experimental replicates using a Pearson's correlation analysis.

2.6 Discussion

Our findings provide a global mechanism by which HSV-1 genome replication controls transcription. Due to the sensitivity and specificity of the approaches used, we were able to make conclusions regarding the transcriptional activity or priming of individual viral promoters. Specifically, it should be noted that the majority of transcriptional events take less time and occur earlier in infection than previously assumed. Transcription for most viral genes has at least been initiated by 3 hpi and decreases by 4 hpi. This trend continues and is ever more drastic at 6 hpi. Since data were quantified as transcription factor occupancy per genome, this suggests a segregation in genome function. Recent work from the N.A.D. laboratory following prelabeled viral genomes found the genome to be associated with transcription factors at 3 hpi and that by 6 hpi the genome was predominantly associated with packaging and assembly factors (153). Furthermore, replication forks were more enriched for transcription factors than previously replicated DNA (164). These data lead us to conclude that before genome replication there is little segregation of function, most genomes are actively transcribed. We propose that after two rounds of genome replication there is a functional coupling, in which newly synthesized genomes are actively transcribed, whereas "older" genomes begin the assembly and packaging process. This functional coupling results in efficient virion production from 5 to 18 hpi.

Our data have allowed us to propose how promoter architecture and genome replication determine transcriptional kinetics. Before replication the genome exists in a state that is not accessible to general transcription factors (GTF) on L2 promoters. What is not clear at present is why the viral chromatin at this time does not allow for TFIID and, hence, Pol II binding on promoters just containing TBP/TAF1-binding sites. Perhaps the restrictive chromatin is due to histone presence, or a specific distribution of viral and cellular genome-binding proteins. Before replication, initiation complexes form only on promoters containing UPEs, i.e., TAATGARAT sites, GC-boxes, and CAAT-boxes, and thus they are robustly transcribed. As expected, we observed SP1 binding to the promoters of most IE and E genes before replication. Select L1 genes were also transcribed at this time, likely due to upstream promoter elements. In these conditions, robust transcription does not require an Inr element. We posit that early during infection, the relatively high density of ICP4 on the viral genome results in the recruitment of TFIID to viral promoters lacking a strong initiator element (122, 150, 152), which have been rendered accessible by the function of upstream activators. We believe this allows for stable TFIID binding to promoters with weak or nonexistent Inr elements, facilitating robust E gene transcription and some leaky L1 gene transcription.

At the onset of genome replication there is an immediate alteration to the structure of the viral genome, such that promoters from all genes classes had an increase in Pol II promoter occupancy. This alteration and the presence of ICP4 was critical for the shift to robust viral transcription. As the number of viral genomes increases, promoters possessing initiator elements that make strong TAF1 contacts are favored. Most L genes robustly recruited TAF1 and possessed strong Inr elements matching the consensus motif, BBCABW. We propose that the increase in viral genomes reduces the relative concentrations of GTFs, such as SP1, TFIIA, TFIID, and ICP4.

Ultimately, the relative decrease in host GTFs and absence of Inr elements resulted in attenuation of E gene transcription. Viral genome numbers continued to increase, resulting in prolonged robust L gene transcription, despite the average transcriptional activity per genome being decreased.

We believe our findings elucidate the major mechanisms by which HSV-1 controls transcription. Our study found that a single round of genome replication permanently altered the transcriptional landscape of HSV-1. The alteration facilitated an increase in genome accessibility to RNA Pol II, TBP, and TAF1. Our results suggest that genome replication was itself responsible for promoting this shift, rather than titration of a factor. This mechanism acted as the switch necessary to promote a global increase in viral transcription and initiate synthesis of previously silent promoters. In this way, synthesis of genes required for later stages of the life cycle, i.e., capsid assembly and egress, is not initiated until sufficient production of earlier viral gene products and recruitment of essential cellular factors. This general mechanism could explain the coupling of genome replication and nascent transcription for other viruses which undergo nuclear DNA replication. This allows for an initial "colonization" stage of infection before a massive wave of proliferation, also rendering the infection less detectable to host defense mechanisms before virion production.

2.7 Materials and Methods

Cells and Viruses. MRC5 (human fetal lung) or Vero (African green monkey kidney) cells were obtained from and propagated as recommended by ATCC. Viruses used in this study include mutants ΔC1216 and tsKos and wild-type KOS. Mutants were made using BAC two-step red-mediated recombination (201-203). To generate the UL30 C-terminal truncation mutant (ΔC1216),

the cassette TAGTTAACTAG, containing a premature termination codon in all three reading frames and an HpaI restriction site, was inserted after the codon for alanine 1216 of the UL30 ORF. This UL30 truncation mutant has previously been characterized *in vitro* (194, 195). The tsK mutant (204) was generated in KOS background as previously described (205) with the following alteration to the ICP4 ORF: alanine to valine at position 475 by mutating the C to T and G to A.

Viral Infection. MRC5 or Vero cells were infected with 10 PFU per cell. Virus was adsorbed in TBS for one hour at room temperature. Viral inoculum was removed, and cells were washed quickly with TBS before adding 2% FBS media. Infected samples were incubated at 37ºC. If designated 100 μM acyclovir was added to infected cultures after adsorption (0 hpi). For tsKos, infected samples were incubated either at permissive (33.5ºC) or nonpermissive (39.6ºC) temperature.

RNA-Sequencing. $2x10^6$ MRC5 cells were infected as described above. RNA was harvested from cultures at 2, 3, 4, 6, 8, 12, or 24 hpi using the Ambion RNAqueous-4PCR kit. Total RNA was quantified using the Agilent RNA 6000 Nano Kit. RNA-Seq libraries were generated from 2 μg RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB #E7490 and #E7420). Libraries were quantified using the Agilent DNA 7500 Kit, and samples were mixed together at equimolar concentration. Illumina HiSeq 2500 was carried out at the Tufts University Core Facility.

ChIP-Sequencing. Confluent MRC5 cells were infected as described above. ChIP-Seq was performed as described previously (128), with the exception of performing seven LiCl washes on immunoprecipitation mixtures. For TBP, TAF1, and Pol II IP approximately $5x10^7$ infected cells were applied to each antibody. For SP1 IP $7x10^7$ infected cells were applied. Samples were immunoprecipitated with 25 µg of the following antibodies: Pol II 4H8 (AbCam #ab5408) Pol II 8WG16 (AbCam #ab817), TBP (AbCam #ab51841), TAF1 (SantaCruz #sc-735), or SP1 (SantaCruz #sc-17824). 2-20 ng of each sample was used to create sequencing libraries using the NEBNext Ultra II DNA Library preparation kit (NEB #E7103S). Libraries were quantified using the Agilent DNA 7500 Kit, and samples were mixed together at equimolar concentration. Illumina HiSeq 2500 was carried out at the Tufts University Core Facility.

Immunofluorescence. 1.7×10^5 Vero cells were infected as described above. Coverslips were fixed at 6 hpi with 3.7% paraformaldehyde. Cellular DNA was stained with 1:2000 Hoescht, and immunofluorescence was carried out using rabbit anti-ICP4 (n15, 1:500), mouse anti-ICP8 (ab20194, 1:200), Goat anti-mouse and anti-rabbit Alexa Fluor 488 and 594-conjugated secondary antibodies (Santa Cruz, 1:500). Images were taken using an Olympus Fluoview FV1000 confocal microscope.

Data Analysis. Data was uploaded to the Galaxy web platform, and we used the public server at usegalaxy.org to analyze the data (206). RNA-Seq Data was aligned using HISAT2 to the HSV-1 strain KOS genome (KT899744.1) or human genome (hg38) (207). FeatureCounts was performed using the KT89974.1 CDS as the reference GFF (208). Raw counts were normalized as mapped reads per million total reads per kilobase (MR/MTR/Kb). Bam files were visualized using

DeepTools bamcoverage to generate bigwig files (209). Transcripts were classified using the maximum RNA abundance value from the time course, with the following values used for delineation:

E Genes: KOS Max ≥1500; ΔC1216 Max ≥1500; Log_2 Fold Change (KOS/C1216) Max≤2.5

L1 Genes: KOS Max ≥1500; ΔC1216 Max ≥150; Log_2 Fold Change (KOS/C1216) Max≥1

L2 Genes: KOS Max ≥1500; ΔC1216 Max ≤900; Log_2 Fold Change (KOS/C1216) Max≥2

ChIP-Seq Data was aligned using Bowtie2 to the HSV-1 strain KOS genome (KT899744.1) or human genome (hg38) (210). Bam files were visualized using DeepTools bamcoverage with a bin size of 1 to generate bigwig files. Data was viewed in IGV viewer and exported as EPS files. Bigwig files were related as the amount of factor bound per viral genome, called "relative factor occupancy" which was calculated as $\frac{IP\ Mapped\ Reads / IP\ Viral\ Reads+Cellular\ Reads}{Input\ Mapped\ Reads / Input\ Viral\ Reads+Cellular\ Reads}$. This normalization method was performed for all ChIP-Seq data and then biological duplicates were averaged. Reproducibility of biological replicates was assessed using a Pearson's correlation analysis (Supplemental Fig. 5). Promoter occupancy was quantified using MapBed (211) with a reference bed file of viral promoters (Table S3). Viral genes without visible Pol II peaks in any conditions were excluded from analysis, these include: UL14, UL15, UL28, UL33, UL43, and US5. Viral promoter regions were defined manually using Pol II ChIP-Seq data. The center of each transcript's Pol II peak was annotated, and the promoter region was designated as +125 to -125 bp from this peak. Robust, distinct promoters from each class were used as the plotted examples. α includes ICP4, ICP0, ICP27, ICP22, ICP47. E includes UL23, UL29, UL39, UL40, UL50. L_1 includes UL18, UL19, UL27, UL31, UL35, UL42, UL48, US6, US7, US10. L2 includes UL10, UL25, UL36, UL38, UL44,

UL47, UL49A, US2, US11. Pol II occupancy heat maps were generated from normalized bigwig files using computeMatrix and plotHeatmap (209). To define transcripts we set the 5' and 3' bounds as 100 bp upstream of the Pol II promoter peak and 10 bp downstream of the KT899744.1 CDS end locus, with a bin size of 10 bp.

TATA boxes and Initiator elements were predicted based on sequence and ChIP-Seq positioning constraints. TATA boxes were first determined by looking for AT-rich elements approximately 35 bp upstream of the center of the TBP ChIP-Seq peak. This positioning requirement was determined based on mutagenesis confirmed TATA elements, e.g. UL44 and UL23. Initiator elements were determined by scanning for the best Inr consensus (BBCABW) match (200) downstream of the previously determined TATA box. If no "CA" was present 20-30 bp downstream of the TATA box, no Inr was annotated. See summary of data in Table S2. Probability plots were generated using WebLogo (212).

qPCR Quantification of viral DNA and mRNA. $2x10^6$ MRC5 cells were infected as described above. RNA and DNA was harvested from replicate cultures at 2, 4, 6, and 8 hours post infection. DNA was harvested by removing supernatant and adding 0.2 mL DNA extraction buffer [0.5% SDS, 400µg/mL proteinase K, 100 mM NaCl]. Samples were incubated at 37ºC for 12-18 hours and heat inactivated for 30 minutes at 65ºC. DNA was serial diluted 1:1000 and measured using qPCR with primers specific to viral gC and cellular GAPDH. Standard curves were generated using purified KOS DNA and purified human DNA. Viral genomes per cell was calculated as $\frac{\boldsymbol{gC\ copy\ \#}}{\boldsymbol{GAPDH\ copy\ \#/2}}$. RNA was isolated using the Ambion RNAqueous-4PCR kit. cDNA was generated from 500 ng total RNA, as quantified using the Agilent RNA 6000 Nano Kit. RNA was reverse transcribed with 20 units Riboguard RNase inhibitor, 0.5 pmols oligo(dT) primer, 100

units MMLV-HP reverse transcriptase, 10 mM dithiothreitol, 2.5 mM dNTPs, and 1x reaction buffer (Epicentre cat no. RT80125K and RG90910K). RNA and oligo(dT) were first incubated at 85ºC for 3 minutes and then incubated on ice for 2 minutes during which remaining reaction components were added. The entire reaction was incubated at 65ºC for 2 minutes and then 37ºC for 1 hour. To heat inactivate components the cDNA was incubated at 85ºC for 5 minutes. cDNA was diluted 1:100 and quantified using qPCR with primers specific to viral gC and UL42. Standard curves were generated using purified KOS. Viral mRNA accumulation was calculated as $\frac{\boldsymbol{mRNA\ copy\ \# \times dilution}}{\boldsymbol{0.5\mu g\ total\ RNA}}$. Primers used in this study include: UL42 ds F (GTCCCGCCGCTCCAGAC), UL42 ds R (CTTGCTTCTCCGGTCGGG), GAPDH F (CAGAACATCATCCCTGCCTCTACT), GAPDH R (GCCAGTGAGCTTCCCGTTCA), gC ds 3 F (GTGACGTTTGCCTGGTTCCTGG), and gC ds 3 R (GCACGACTCCTGGGCCGTAACG).

Data Availability.

All data are publicly accessible in the SRA database (SRP172751, SRP172780, SRP172782, and SRP172783).

3.0 Herpes simplex viral nucleoprotein creates a competitive transcriptional environment facilitating robust viral transcription and host shut off

3.1 Project Summary

Herpes simplex virus-1 (HSV-1) replicates within the nucleus coopting the host’s RNA Polymerase II (Pol II) machinery for production of viral mRNAs culminating in host transcriptional shut off. The mechanism behind this rapid reprogramming of the host transcriptional environment is largely unknown. We identified ICP4 as responsible for preferential recruitment of the Pol II machinery to the viral genome. ICP4 is a viral nucleoprotein which binds double-stranded DNA. We determined ICP4 discriminately binds the viral genome due to the absence of cellular nucleosomes and high density of cognate binding sites. We posit that ICP4’s ability to recruit not just Pol II, but also more limiting essential components, such as TBP and Mediator, create a competitive transcriptional environment. These distinguishing characteristics ultimately result in a rapid and efficient reprogramming of the host’s transcriptional machinery, which does not occur in the absence of ICP4.

3.2 Importance

ICP4 is the major viral transactivator of HSV-1, it is required for efficient transcription of early (E) and late (L) genes and production of viral progeny. We demonstrate that the virus relies on ICP4's scaffolding ability to prioritize recruitment of the Pol II machinery at the expense of the host. This facilitates the rapidly progressing infection while limiting the extent to which the host can respond to viral challenge. We also found that ICP4 at high concentrations bound sequence nonspecifically to naked or accessible host and viral DNA. What implications or effects this may have on host transcription require further investigation. We propose that ICP4 functions in place of traditional cellular chromatin, and allows for the robust recruitment of cellular transcription factors specifically to the viral genome.

3.3 Introduction

Like most DNA viruses, the genome of Herpes simplex virus-1 (HSV-1) is transcribed by RNA Polymerase II (Pol II) (117). Its approximately 85 genes (170, 185, 186) are transcribed in a temporally coordinated sequence, such that their protein products are expressed at the appropriate time in the life cycle of the virus (123, 124). Immediate early (IE) gene products enable the efficient expression of early (E) and late (L) genes. The protein products of E genes are mostly involved in DNA replication. DNA replication and IE proteins enable the efficient transcription of L genes, which encode the structural components of the virus. DNA replication licenses L promoters, enabling the binding of core Pol II transcription factors, thus activating the initiation of L transcription (213). This entire transcriptional cascade is observed within 3 hour (h) post entry

(153, 213), culminating in production of the first viral progeny between 4 and 6 h post-infection (hpi). To accomplish this robust and rapidly changing program of transcription, the viral genome must compete with the vastly larger cellular genome for numerous Pol II transcription factors, in addition to mediating the possible constraints of cellular histones.

A major component of this cascade is the IE protein Infected Cell Polypeptide 4 (ICP4) (214). ICP4 is essential for viral growth because it promotes efficient transcription of viral E and L genes (143-145). Thus, in the absence of ICP4, E and L proteins are poorly expressed, IE proteins are overproduced, DNA replication does not occur, and there is no detectable viral yield (215). ICP4 was first shown to bind to DNA cellulose made from salmon sperm DNA (148). Faber and Wilcox (1986) later showed ICP4 has sequence-specific DNA binding activity. ICP4 interacts with a number of cellular general transcription factors (GTFs), predominantly components of TFIID and the Mediator complex (150-152), facilitating their recruitment to the viral genome through its DNA binding activity (151, 153, 154). ICP4 is synthesized early in infection, binds to the viral genome located at ND10 structures (216), and remains associated with the genome throughout all phases of infection (153). Therefore, ICP4 has the potential to influence events occurring on the viral genome from a time when genome number is at a minimum, and ICP4 expression is peaking, through a time when genome numbers are greatly elevated by replication.

Studies have also shown that epigenetic modulation of histones associated with the viral genome early in infection can affect productive viral infection (77, 217). However, we have shown that the abundance of histones is relatively low or absent, and that ICP4 is one of the most abundant proteins on viral genomes during productive infection (75). In this study, we set out to determine the relationship between ICP4 and histones binding to the viral and cellular genomes, and the consequences for viral and cellular transcription. We propose that ICP4 is a major component of

viral nucleoprotein, which functions in place of traditional cellular chromatin, and allows for the robust recruitment of cellular transcription factors specifically to the viral genome.

3.4 Results

3.4.1 ICP4 binding is altered by viral genome replication.

Given the central role of ICP4 in viral gene transcription at all stages of infection, we were interested in how ICP4 interacts with the virus genome as the number exponentially increases, as a consequence of replication. We infected human fibroblast (MRC5) cells with wild-type HSV-1 (KOS) for 2, 4, and 6 hpi and performed ChIP-Seq for ICP4. Each time point represents a different replication state: 2 hpi (prereplication), 4 hpi (3-4 genome duplications), 6 hpi (5-6 genome duplications) (Fig. 16A).

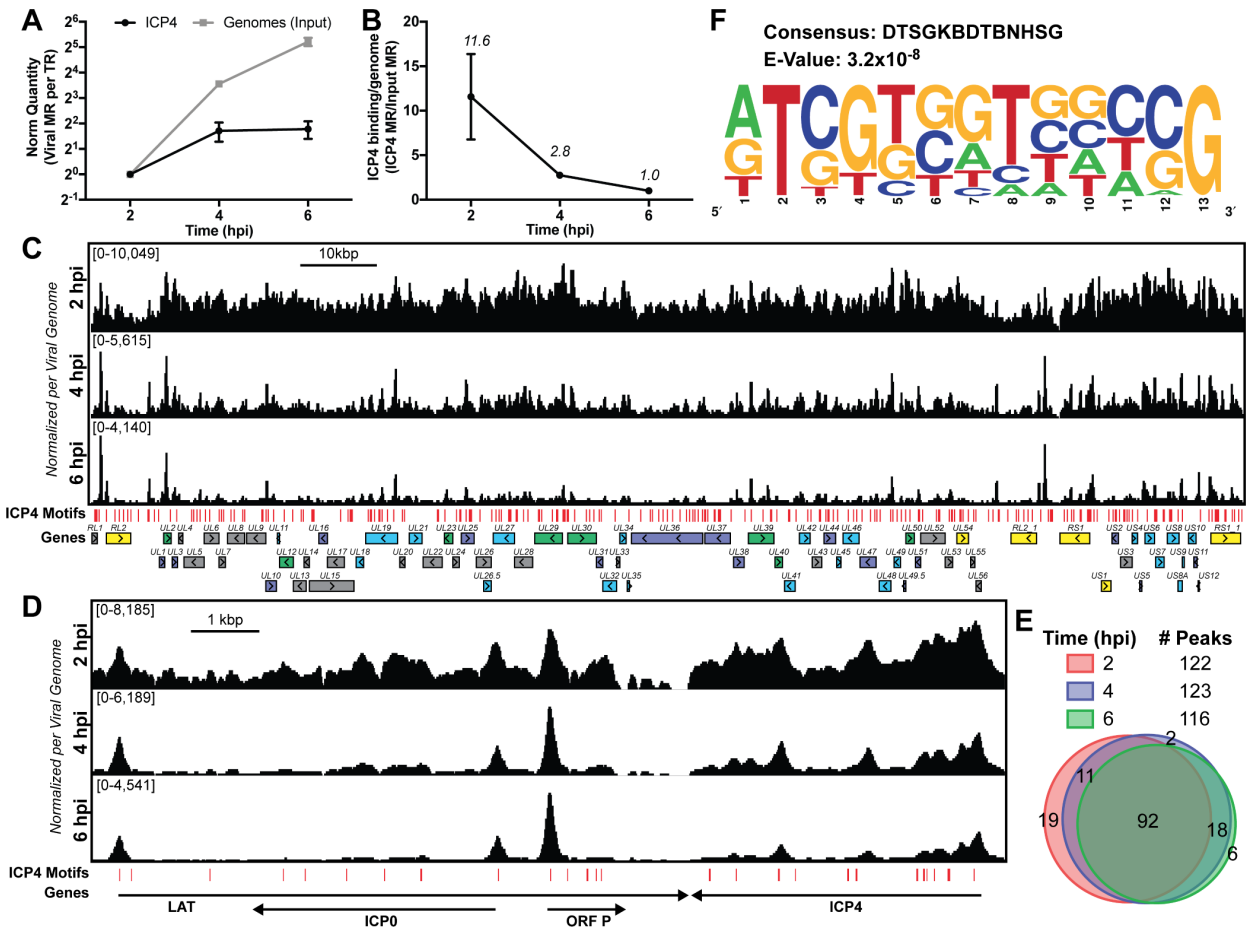


Figure 16 ICP4 binding at key points in the viral life cycle

MRC5 cells were infected with HSV-1 for 2, 4, or 6 h, and ChIP-Seq for ICP4 was performed. (A) Quantification of viral genomes, measured as Input sample viral mapped reads (MR), normalized for sequencing depth or Total Reads (TR). Samples were normalized to two hpi, which was set as 1. (B) Quantification of ICP4 binding per viral genome measured as viral MR from ICP4 immunoprecipitation (IP) per viral MR from Input. (C–D) All data was normalized for sequencing depth and viral genome number using input ChIP-Seq reads. Viral ORFs are indicated, color coded by gene class with IE as yellow, E as green, leaky late (L1) as blue, and true late (L2) as purple. Find Individual Motif Occurrences (FIMO) identified genome sequences matching the consensus motif in F are indicated in red. (E) Intersection of MACS2 identified ICP4 occupied regions. (F) ICP4 consensus binding motif.

To quantitatively compare samples, we had to account for viral genome replication. Input samples provided the relative number of viral genomes present at each time point. We used this ratio to normalize immunoprecipitated (IP) sample for the amount of factor per genome. Early during infection (2h) ICP4 densely coated the viral genome (Fig. 16B-C). Viral genome replication decreased the amount of ICP4 bound per genome (Fig. 16A-B) resulting in a pattern containing sharper peaks. By 6 hpi ICP4 binding was retained exclusively on strong ICP4 binding motifs (Fig. 16C). Some of the retained binding sites were those previously established as having an inhibitory effect on the gene promoter bound, including ORF P, ICP4, and LAT (Fig. 16D). A closer analysis of ICP4 peaks demonstrated that the location and number of high confidence occupied sites did not alter significantly throughout infection (Fig. 16E). Instead the amount of ICP4 bound between distinct peaks decreased as genome number increased (Fig. 16C-D).

Although ICP4 exhibited a dense binding pattern at early times (2 hpi), with relatively broad, overlapping peaks we were able to determine high confidence binding sites. The final sites were consistent between biological replicates. We analyzed 100 bp extensions from the summits of the peaks seen at all 3 times (Fig. 16E) for motif discovery. DTSGKBDTBNHSG was the only motif discovered (Fig. 16F), where D is A, G, T; S is C or G; K is G or T; B is C, G, T; H is A, C, T. This binding motif is similar to that previously discovered using in vitro techniques, RTCGTCNNYNYSG (218).

These results demonstrate that ICP4 binds to specific sites, but also coats the genome early in infection forming a type of nucleoprotein. Due to mass action, ICP4-nucleoprotein changes as infection proceeds, limiting binding to predominantly strong cognate binding sites as the number of genomes increase due to replication.

3.4.2 ICP4 stabilizes GTF binding promoting cooperative preinitiation complex (PIC) assembly.

We wanted to investigate how the formation of ICP4-nucleoprotein affects the transcription factor landscape across the viral genome. We compared the binding of ICP4, Pol II, TATA-binding protein (TBP), SP1, Med1, and Med23 in ICP4 null (n12) and wild-type (WT) HSV-1 infected human fibroblasts at 2.5 hpi by ChIP-Seq (Supplemental Fig. 6). In all Pol II IP's we used an antibody which preferentially binds to phospho S5 of the C-terminal domain repeat region. This post-translational modification is associated with Pol II found on mRNA promoters and splice sites.

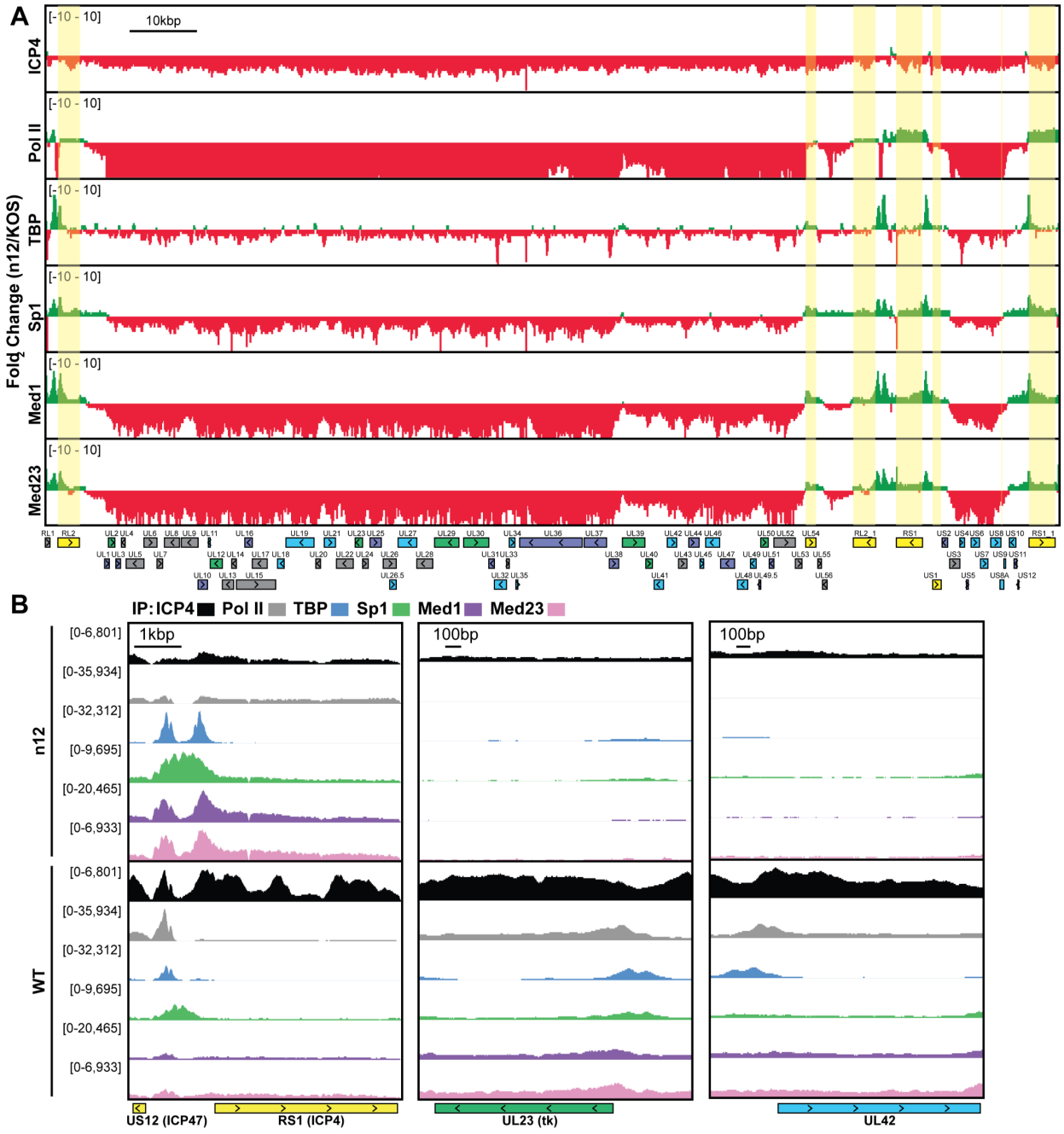

A
[-10 - 10]
10kbp
ICP4
Pol II
TBP
Sp1
Med1
Med23
Fold Change (n12/KOS)
B
IP: ICP4 Pol II TBP Sp1 Med1 Med23
n12
WT
[0-6,801]
[0-35,934]
[0-32,312]
[0-9,695]
[0-20,465]
[0-6,933]
1kbp
100bp
100bp
US12 (ICP47)
RS1 (ICP4)
UL23 (tk)
UL42

Figure 17 ICP4 recruitment of host Pol II machinery to viral promoters

MRC5 cells were infected with an ICP4 null mutant (n12) or HSV-1 (WT) for 2.5 h and ChIP-Seq for ICP4, Pol II, TBP, SP1, Med1, and Med23 was performed. All data was normalized for sequencing depth and viral genome number using input ChIP-Seq reads. Viral ORFs are indicated, color coded by gene class with IE as yellow, E as green, leaky late (L1) as blue, and true late (L2) as purple. (A) Fold change of n12 over WT aligned to the viral genome. Loci with greater binding in n12 or WT are colored in green or red, respectively. (B) ChIP-Seq reads normalized per viral genome and aligned to canonical IE, E, and L1 genes.

In n12 infection, we observed a decrease in binding for all the factors to most viral promoters, with the exception of IE genes (Fig. 17A-B), where there was an increase in binding. There were detectable, although highly reduced, peaks of TBP and SP1 binding to the UL23 promoter in the absence of ICP4. It has been previously shown that these sites are functional in the n12 background reflecting the basal binding activity of TBP and SP1 (219) (Fig. 17B). Similar to UL23 we observed TBP and SP1 bound to selective E promoters in the absence of ICP4, namely UL23, UL29, UL39, and UL50 (Supplemental Fig. 6).

There were a relatively small number of reads in the ICP4 ChIP of n12 (Fig. 17A, Supplemental Fig. 6). As ICP4 is an essential viral protein, we prepared stocks of n12 in an ICP4 complementing cell line. ICP4 is packaged into the tegument of virions (220), which results in packaging of wild type ICP4 into n12 virions (153). The reads in the ICP4 ChIP of n12 are most likely due to ICP4 packaged in the virion. While the ultimate source of these reads is not clear at present, the amount of binding of ICP4 from the virion to DNA is not sufficient to promote transcription complexes on viral early and late genes.

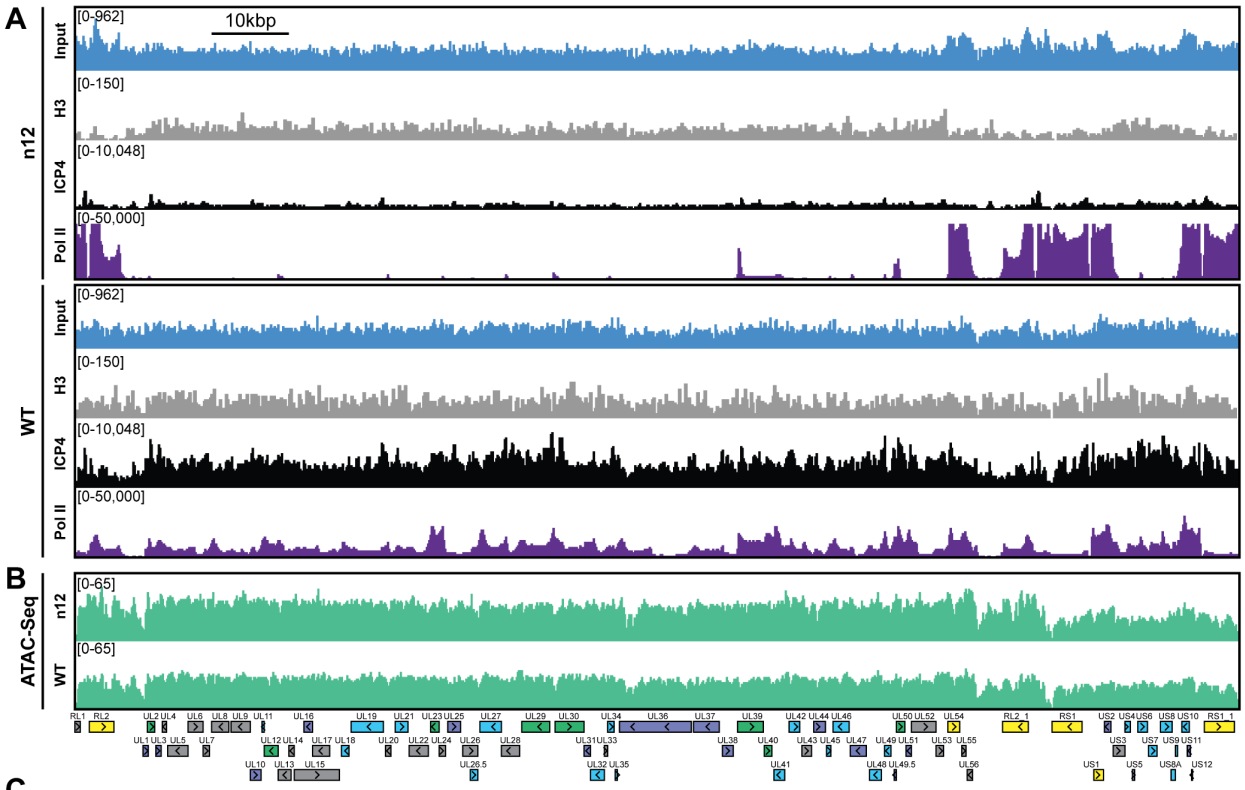
Med1 and Med23 bound the viral genome with an almost identical pattern (Supplemental Fig. 6, Fig. 17), indicating they are parts of the Mediator complex bound early during viral infection. In WT infected cells, the binding of Mediator concentrated near the starts sites of ICP4-

induced viral genes. However, the Mediator complex also densely coated the viral genome, resembling the ICP4 binding pattern. This dense coating was completely absent on early and late genes in n12 infection, demonstrating this phenotype is not an artifact of the IP. We suspect this reflects the fact that ICP4 and Mediator interact.

In the absence of ICP4, the binding of Pol II was reduced the most compared to the other transcription factors (Fig. 17A). This magnified difference is likely a result of the cooperative nature of Pol II recruitment requiring multiple protein-protein interactions. In summary ICP4 was required for robust recruitment of all GTF's tested, cooperatively recruiting Pol II to E and L promoters (Fig. 17B). The difference between n12 and WT shows the extent by which ICP4 mediated recruitment and bolstered the frequency of PIC assembly. IE promoters retained robust GTF recruitment via an independent mechanism involving a complex consisting of Oct-1, HCF and VP16 binding to TAATGARAT promoter elements (221-223).

3.4.3 Genome-bound ICP4 does not affect accessibility.

Part of the mechanism of ICP4 action in the recruitment of GTFs to the genome may involve a role in the exclusion of repressive chromatin. To address this hypothesis, we investigated the relationship between presence of ICP4, the abundance of histones, and the accessibility of the genome. We used ChIP-Seq to compare the binding of ICP4, Pol II, and histone H3 in n12 and WT HSV-1 infected human fibroblasts at 2 hpi (Fig. 18).



C

Sample	Genome	Genome Size (bp)	Genome Copy #	Total DNA (bp)	% of Total DNA*	%ATAC-Seq Read‡	Fold enrichment over expected**
n12	**Host**	3.2E+09	2	6.4E+09	99.4 (±0.4)	98.7 (±0.2)	1.0
	Viral	1.5E+05	254 (±169)	3.9 (±2.6) x10^7	0.6 (±0.4)	1.3 (±0.2)	2.2
	Rel Diff	*2.1E+04*		*377*			
WT 2h	**Host**	3.20E+09	2	6.40E+09	99.6 (±0.01)	99.2 (±0.5)	1.0
	Viral	1.52E+05	169 (±4)	2.6 (±0.06) x10^7	0.4 (±0.01)	0.8 (±0.5)	2.0
	Rel Diff	*2.13E+04*		*721*			

**Based on percent of total reads from ChIP-Seq Input reads, values are the mean (± SD) of four biological replicates

‡Based on percent of total reads from ATAC-Seq, values are the mean (± SD) of three biological replicates

**Fold enrichment=ATAC-Seq read percentage/percentage of Total DNA

D

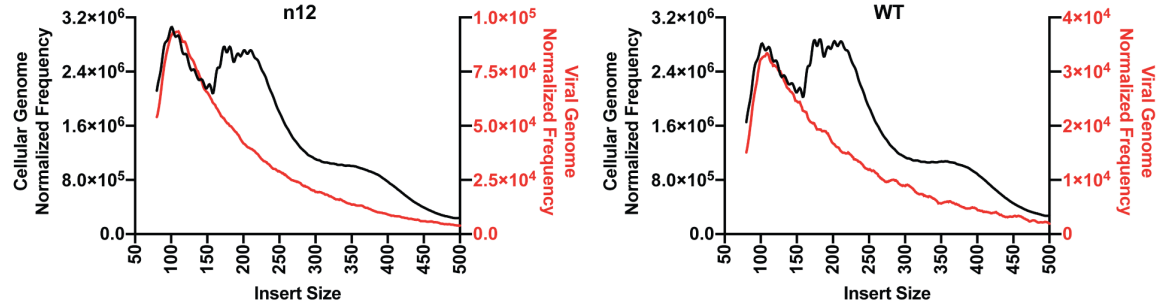


Figure 18 ICP4 dependence of viral genome accessibility and Histone H3 binding

MRC5 cells were infected with an ICP4 null mutant (n12) or HSV-1 (WT) and harvested prior to genome replication. (A) ChIP-Seq for ICP4, Pol II, H3 and B–D) ATAC-Seq was performed. All data was normalized for sequencing depth and viral genome number using input ChIP-Seq reads. (C) Quantitative analysis of ATAC-Seq data, measuring the relative tagmentation enrichment for the virus or host as compared to expected. (D) Histogram plot of ATAC-Seq fragment size for reads mapped to the viral (red) or cellular (black) genome. Mononucleosome protected fragments are approximately 180–250 bp.

We found in both WT and n12 infection that the number of H3 reads mapped to the viral genome was 100-fold less than ICP4, and the pattern was nearly identical to input reads (Fig. 18A), with R^2 correlations of 0.0004 and 0.02 (Supplemental Fig. 7). These data demonstrated H3 binding to the viral genome was minimal and not reproducible. Furthermore, H3 binding was still minimal in the absence of ICP4 (n12). This was not due to technical issues as the number and quality of H3 reads mapped to the cellular genome for the same samples was approximately 10 million with R^2 correlations of ≥0.97 (Supplemental Fig. 8). We saw a similar trend with H3K4me3, H3K27ac, H3K9me3, and H3K27me3 reads mapped to the viral genome (Fig. Supplemental Fig. 7-9).

Although H3 binding to the viral genome was similar in WT and n12 infection, we could not rule out the role of an alternative protein occluding the genome. To investigate genomic accessibility, we performed ATAC-Seq. Human fibroblasts were infected with WT and n12 HSV-1 at an MOI of 10 pfu/cell and collected prior to the onset of genome replication. Quantification of ChIP-Seq input reads allowed us to determine that the approximate number of genomes per cell in WT and n12 infection was 169 and 254, respectively (Fig. 18C). This value is consistent with infecting at an MOI of 10 pfu/cell and an approximately particle to pfu ratio of 20-30. We normalized ATAC-Seq traces to adjust for sequencing depth and input genome number. We

observed even tagmentation in both conditions (Fig. 18B) absent the nucleosomal laddering visible on the cellular genome (Fig. 18D). Quantification of ATAC-Seq reads determined that the viral genome in n12 and WT was 2.2 and 2-fold more accessible than the cellular genome (Fig. 18C). As we harvested samples pre-replication, we expect that a significant portion of viral genomes are defective and will not undergo replication. Our ATAC-Seq data is thus an average of tagmentation for defective and active viral genomes. For this reason, we expect our accessibility calculation is an underestimate. We conclude that the viral genome was much more accessible than the cellular genome, and this increased accessibility was not ICP4-dependent. ICP4 binding and GTF recruitment, not viral genome accessibility, was responsible for robust GTF binding.

3.4.4 ICP4 binds to cellular transcription start sites (TSS) early during infection.

Immunofluorescence (IF) studies of HSV-1 infection depict colocalization of ICP4 with EdC-labeled viral genomes and exclusion from dense areas of cellular chromatin (75). This phenomenon is so well established that ICP4 is largely used in IF studies as a proxy for HSV-1 genomes. To ascertain if ICP4 also binds to the cellular genome, we aligned our ICP4 ChIP-Seq data from 2, 4, and 6 hpi to the cellular genome. ICP4 bound to the cellular genome in a manner quite distinctive from the pattern observed on the viral genome.

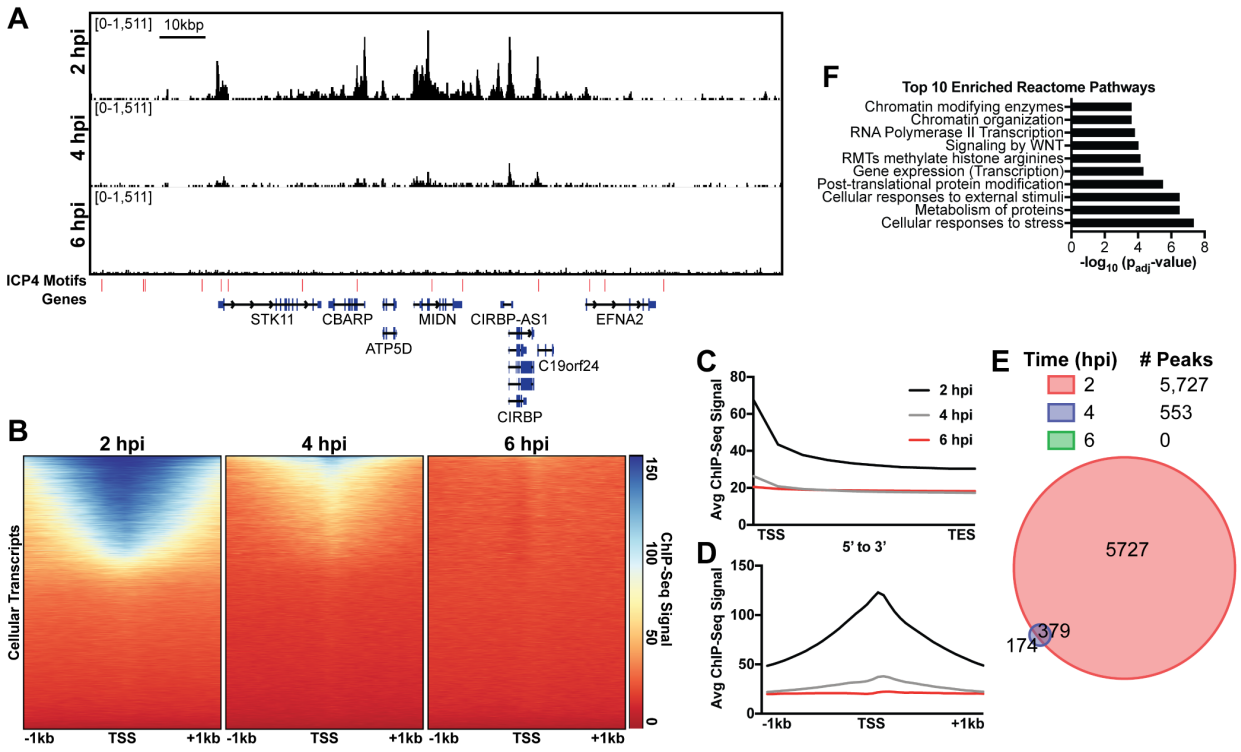


Figure 19 ICP4 binds cellular promoters during early infection

MRC5 cells were infected with HSV-1 for 2, 4, or 6 h, and ChIP-Seq for ICP4 was performed. Data was aligned to the human genome (hg38) and normalized for cellular genome sampling. (A) Representative region of the cellular genome. ICP4 binding motifs are indicated in red. (B, D) Sequencing data centered + /- 1 kilobase from the transcription start site (TSS) of cellular mRNAs. (C) Sequencing data scaled from mRNA TSS to the transcription end site (TES). (E) Intersection of MACS2 identified ICP4 occupied regions. (F) Top 10 enriched Reactome groups for genes (n = 2190) with ICP4 bound at the promoter at 2 hr.

ICP4 only bound in distinct peaks around cellular transcription start sites (TSS) of a subset of cellular genes (Fig. 19A-D). These genes grouped ontologically to common housekeeping functions including pathways related to chromatin, transcription, and metabolism (Fig. 19F). This binding reduced from 2 to 4 hpi, and became negligible at 6 hpi (Fig. 19A-B). At 2 hpi ICP4 bound to the cellular genome at 5,727 sites (Fig. 19E) or 0.002 peaks per kbp, whereas ICP4 bound to the

viral genome at 122 sites (Fig. 16E) or 0.8 peaks per kbp. Similar we found a much greater density of ICP4 binding motifs present in the viral genome (2 motifs/kbp) than the cellular genome (0.02 motifs/kbp). We observed ICP4 binding peaks that did not localize at an ICP4 binding consensus (Fig. 19A) suggesting that ICP4 may associate with the cellular genome by an alternative mechanism. We conclude that ICP4 bound to the cellular genome early during infection, when the relative concentration of ICP4 to viral genomes is still quite high. The amount of ICP4 on the cellular genome quickly dropped off as viral genome number increased and ICP4 preferentially bound to the viral genome.

3.4.5 ICP4 binding is restricted to accessible regions of the cellular genome.

Since ICP4 bound to a subset of cellular genes near mRNA start sites (Fig. 19), We hypothesized that ICP4 only bound to accessible regions of the cellular genome. To test this hypothesis, we performed ChIP-Seq for ICP4, Pol II, Histone H3 (H3), euchromatic markers H3K4-trimethyl (H3K4me3) and H3K27-acetyl (H3K27ac), and heterochromatic marker H3K9-trimethyl (H3K9me3) and H3K27-trimethyl (H3K27me3) on MRC5 cells that were infected with HSV for 2 h.

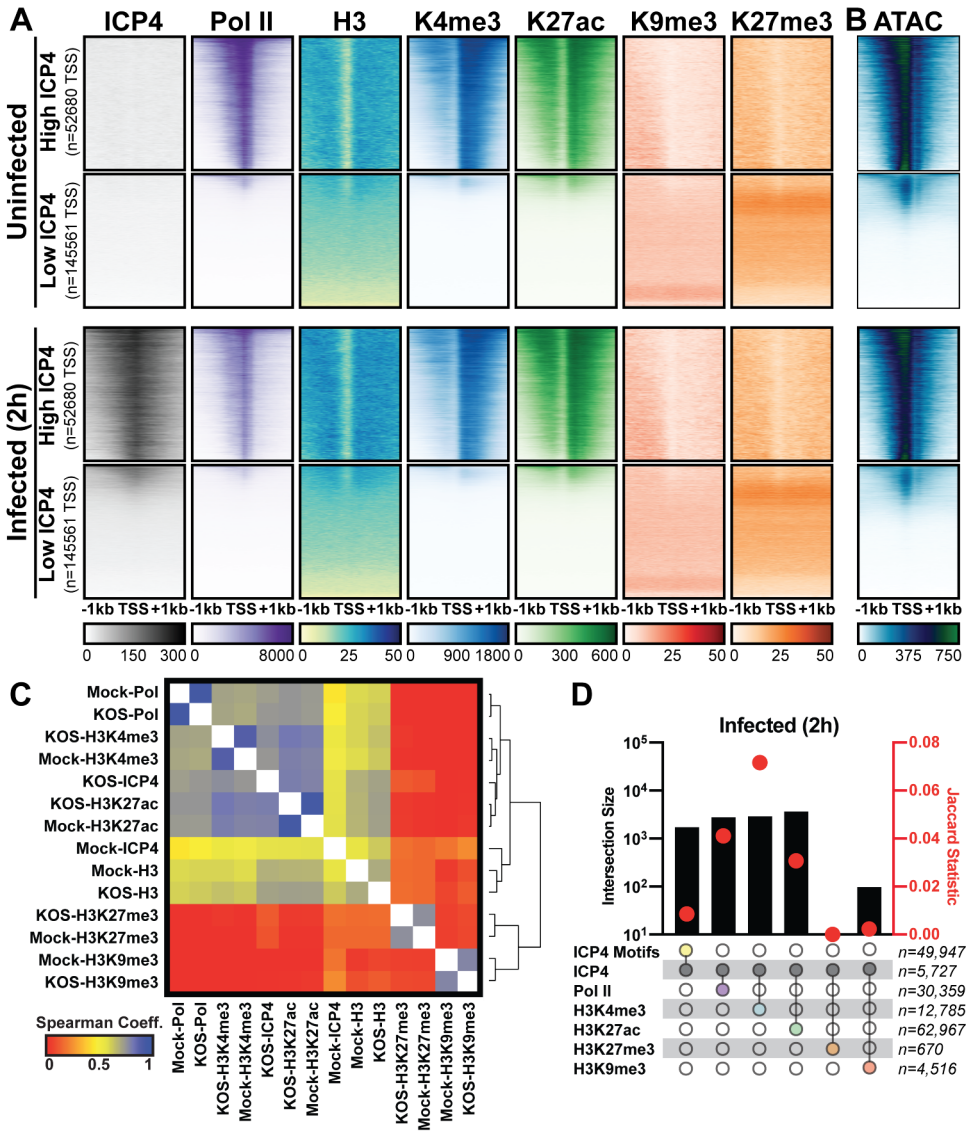


Figure 20 Association between cellular ICP4 binding and chromatin

MRC5 cells were uninfected or infected with HSV-1 for 2 h. All data were aligned to the human genome (hg38) and normalized for sequencing depth. (A, C–D) ChIP-Seq data for ICP4, Pol II, H3, H3K4me3, H3K27acetyl, H3K9me3, and H3K27me3. (B) ATAC-Seq data. (A–B) Sequencing data centered + /- 1 kilobase from the TSS of cellular mRNAs. Data was stratified for ICP4 binding using K-means clustering. (C) Spearman correlation analysis, limited to cellular transcripts. (D) Intersection of MACS2 peaks, analyzed as number of intersecting peaks or Jaccard statistic.

Cellular TSS were stratified using k-means clustering as high and low ICP4 binding (Fig. 20A). TSS with high ICP4 binding were also bound by Pol II and adjacent to euchromatic markers. TSS with low ICP4 binding were associated with only heterochromatic markers. Furthermore, genes clustered as high ICP4 binding had higher tagmentation frequency when assessed using ATAC-Seq (Fig. 20B). The data was mapped for representative cellular genes in Supplemental Fig. 10. We quantified the relationship between ICP4 and cellular chromatin in Fig. 20C-D. We found that the binding pattern of ICP4 was directly related (Spearman coefficient ≥0.5) to Pol II and cellular euchromatin, clustering as most similar to H3K27ac and H3K4me3 (Fig. 20C). The heterochromatic markers, H3K9me3 and H3K27me3, clustered together, and were not correlated (Spearman coefficient ~0) to ICP4 or cellular chromatin. These results were corroborated by analysis of distinct peaks called using MACS (Fig. 20D). Interestingly, ICP4 bound regions had little overlap with their cognate binding motifs (Fig. 20D). A closer analysis of the actual genomic region where each factor bound, revealed that 82% of ICP4 bound regions were within 1 kb of a promoter (Supplemental Fig. 11). By comparison only 10% of ICP4 predicted binding motifs were within 1 kb of a promoter. Furthermore, the euchromatic regions of the cellular genome that were occupied by ICP4 in infected cells were also euchromatic in uninfected cells, indicating that ICP4 does not globally promote open chromatin in these regions of the genome (Fig. 20A-B). These data support a model in which ICP4 is able to bind nonspecifically to accessible regions of the cellular genome, namely active promoters, early in infection when the relative concentration of ICP4 is high.

3.4.6 ICP4 mediates depletion of Pol II on cellular promoters

We observed depletion in Pol II binding to cellular promoters with infection (Fig. 20A). This observation is consistent with prior studies, which assessed HSV-1 infection post-replication at 3, 4, or 6 hpi (73, 224, 225). As we harvested samples prior to the onset of genome replication (2 hpi) we hypothesized that ICP4, which is produced immediately upon viral infection was responsible. First, we determined the effect of ICP4 on cellular promoters before the onset of genome replication. We mock-infected or infected fibroblasts at 10 pfu/cell with WT or ICP4-null (n12) HSV-1 for 2 h.

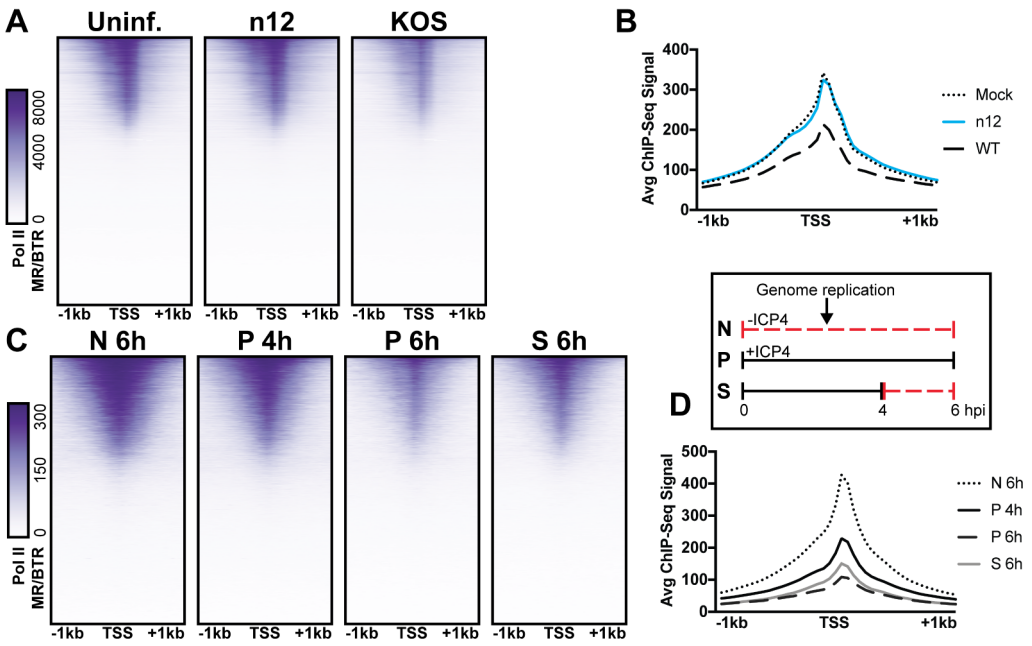


Figure 21 The role of ICP4 in Pol II loss on host promoters

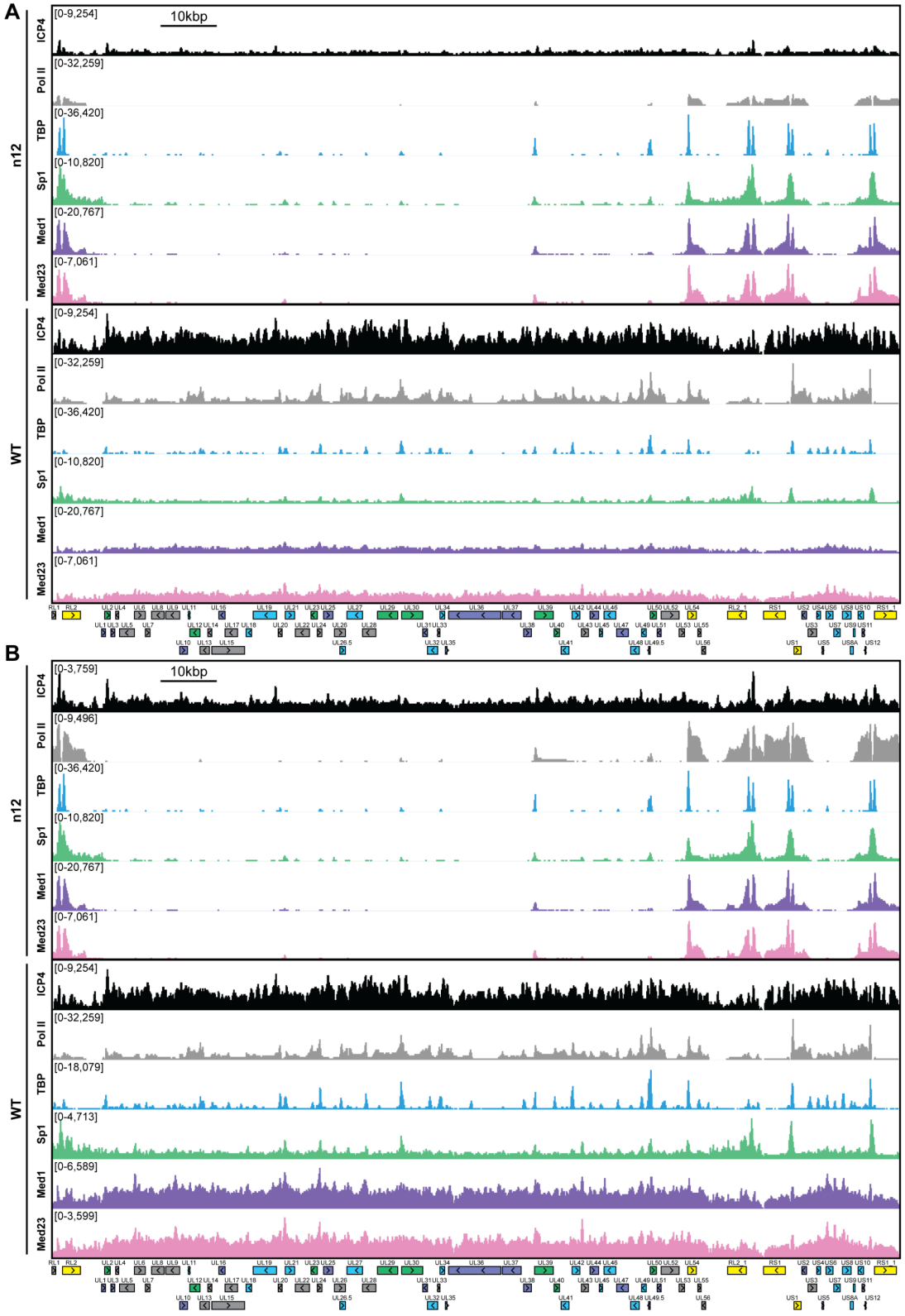
MRC5 cells were (A-B) uninfected or infected with n12 or WT HSV-1 and harvested prior to genome replication or C-D) infected with tsKos and grown at permissive conditions (P), shifted up from permissive to nonopermissive conditions at four hpi (S), or nonpermissive conditions (N). (A–D) ChIP-Seq for Pol II was performed. Data was aligned to the human genome (hg38) and normalized for cellular genome sampling. ChIP-Seq data was aligned to cellular promoters + /- 1 kilobase from TSS. The average signal for each condition plotted as a line graph.

We observed depletion of Pol II occupancy on cellular mRNA promoters only in WT infection (Fig. 21A-B). Thus we concluded that ICP4 was required for depletion of Pol II from host mRNA promoters, and this effect was independent of viral genome copy number.

We then assessed whether ICP4 was continuously required for cellular Pol II depletion, namely if ICP4 was still essential even after the onset of genome replication. We used a temperature sensitive ICP4 mutant (tsKos), in which growth at nonpermissive temperature

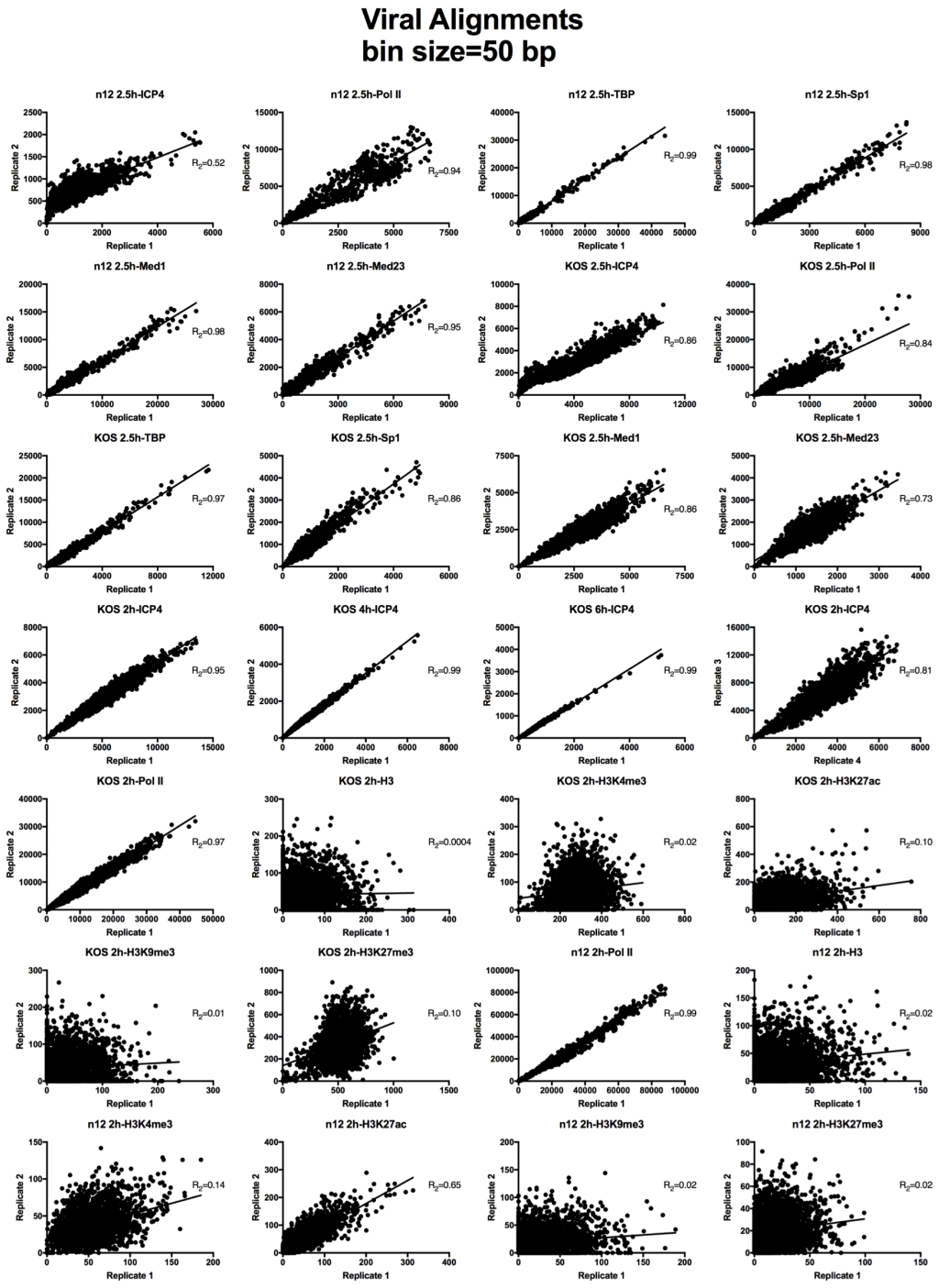
(39.6ºC) results in loss of ICP4 in the nucleus (213). We infected fibroblasts with tsKos grown at permissive conditions (P), shifted up from permissive to nonpermissive conditions at 4 hpi (S), or nonpermissive conditions (N). In this system we can separate the role of ICP4 in Pol II depletion, from ICP4's requirement in E and L transcription and viral genome replication. Infected cells were harvested at 4 or 6 hpi and Pol II ChIP-Seq was performed. We used nonpermissive conditions as a surrogate to mock-infection, as we just established that Pol II depletion does not occur in n12 infection (Fig. 21A-B). We observed significant depletion of Pol II from cellular promoters in permissive and shifted samples (Fig. 21C-D). Pol II depletion was not directly related to viral genome copy number. tsKos shifted up had the highest number of viral genomes present, but did not reach the same level of cellular Pol II depletion as cells grown at permissive temperature for the same length of time. These data suggest that the viral genome is not solely responsible for preferential recruitment of cellular Pol II. Instead ICP4 bound to the viral genome is required for depletion of Pol II from cellular promoters. These results suggest a model in which genome replication facilitates host Pol II depletion when the relative number of ICP4 to viral genomes is high (2h). As the number of ICP4 bound viral genomes increased, we observed a corresponding decrease in Pol II on host promoters.

3.5 Supplemental Figures



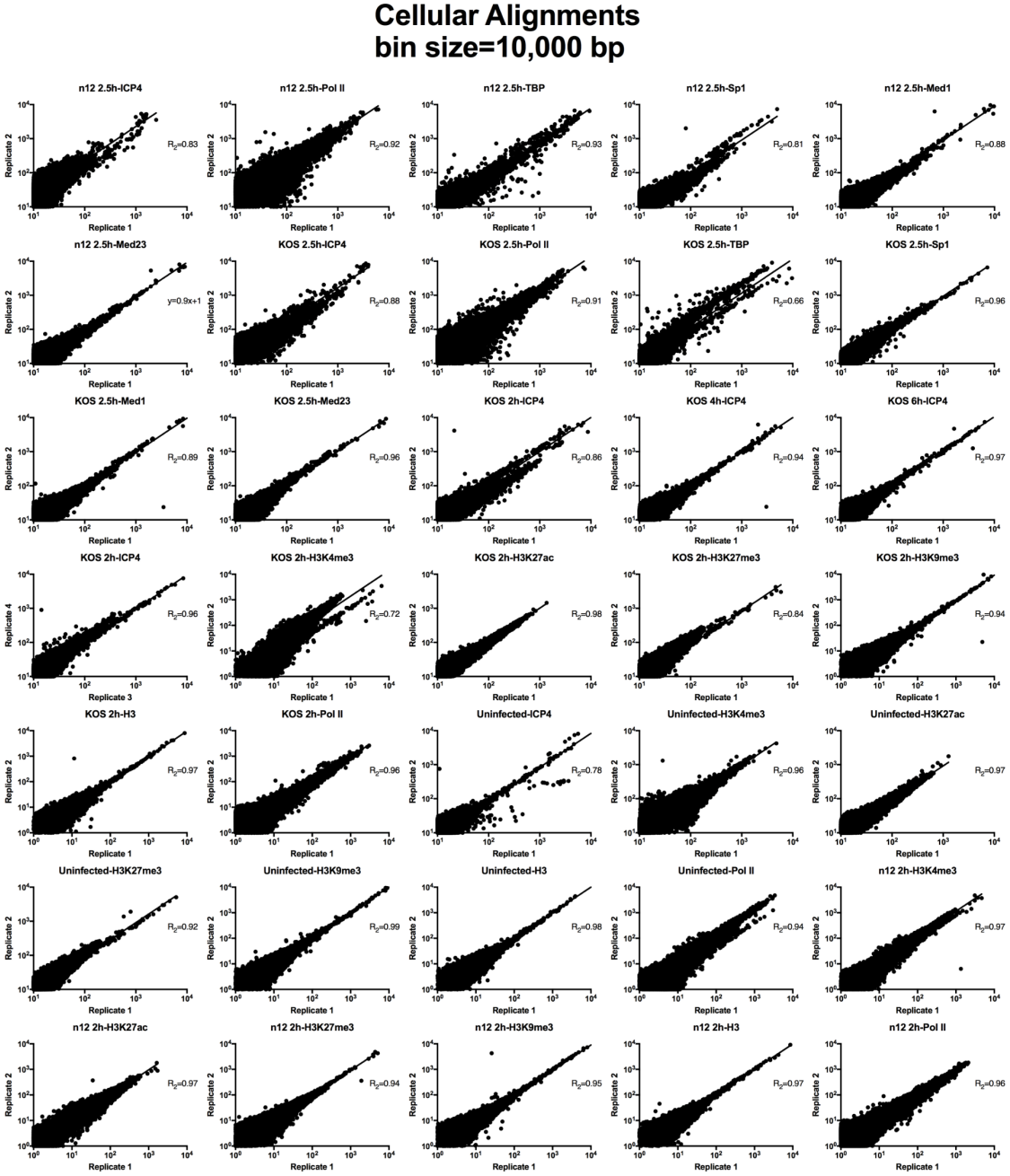
Supplemental Figure 6 GTF recruitment to the viral genome in the presence and absence of ICP4

MRC5 cells were infected with an ICP4 null mutant (n12) or HSV-1 (WT) for 2.5 h and ChIP-Seq for ICP4, Pol II, TBP, SP1, Med1, and Med23 was performed. All data was normalized for sequencing depth and viral genome number using input ChIP-Seq reads. Viral ORFs are indicated, color coded by gene class with IE as yellow, E as green, L1 as blue, and L2 as purple. y-axes were A) maintained the same between IP's or B) maximized for each trace to aid in visualization.



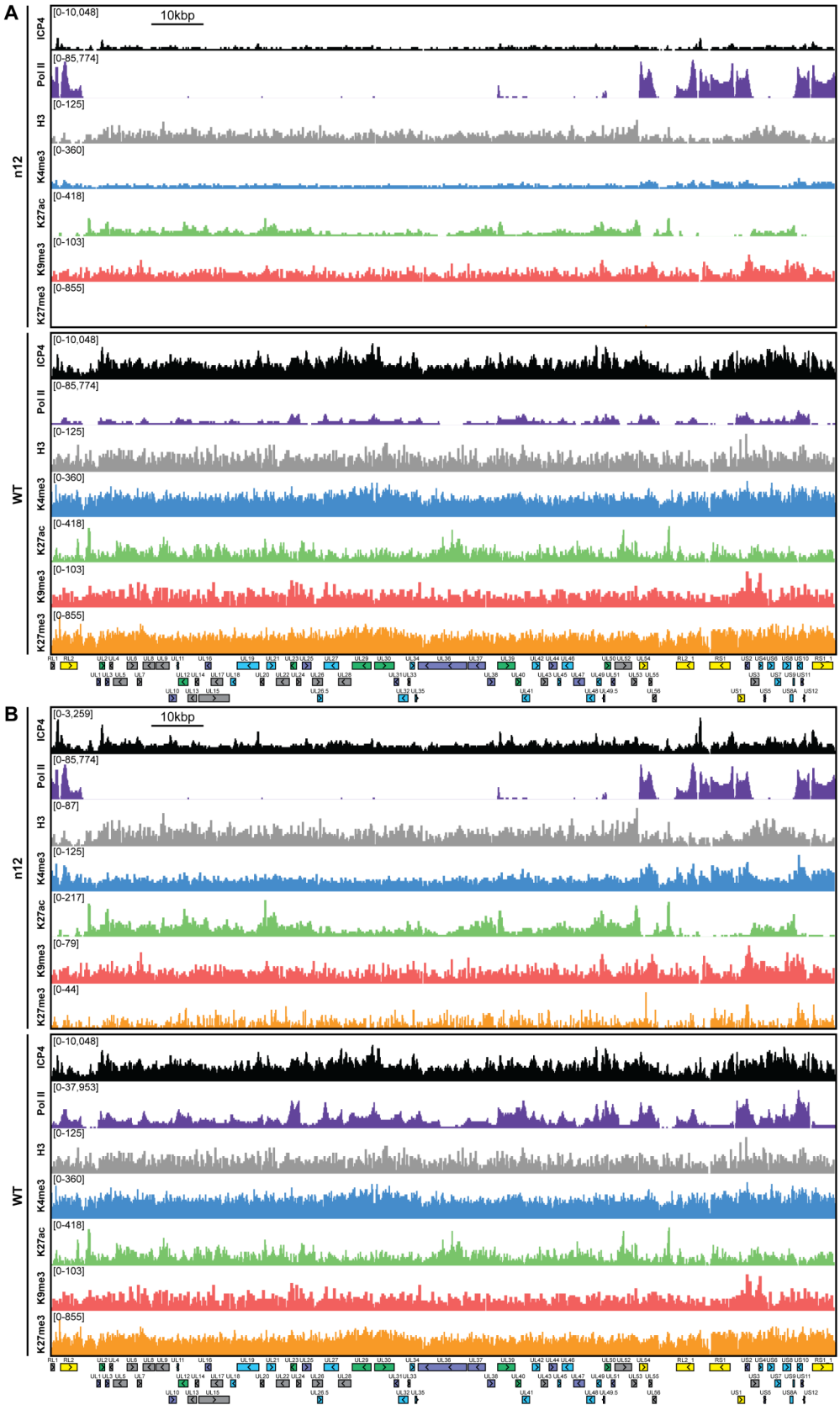
Supplemental Figure 7 Analysis of Viral ChIP-Seq data quality

Normalized viral aligned bigwig files were assessed using MultiBigwigSummary in 50 bp bins. The values within these bins were plotted for biological replicates and a linear regression analysis was performed.



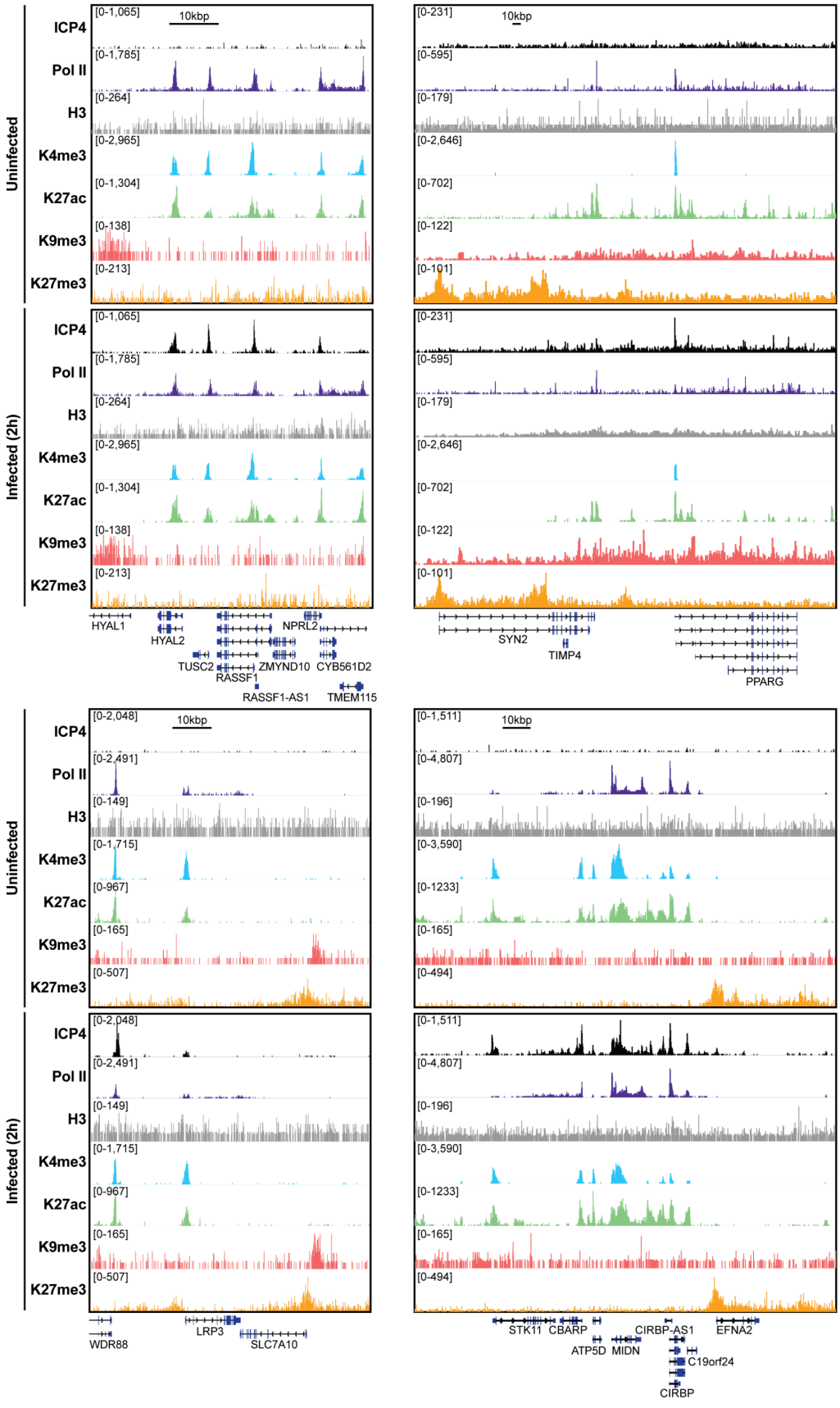
Supplemental Figure 8 Analysis of Cellular ChIP-Seq data quality

Normalized cellular aligned bigwig files were assessed using MultiBigwigSummary in 10,000 bp bins. The values within these bins were plotted for biological replicates and a linear regression analysis was performed.


A
n12
WT
ICP4
Pol II
H3
K4me3
K27ac
K9me3
K27me3
[0-10,048]
10kbp
[0-85,774]
[0-125]
[0-360]
[0-418]
[0-103]
[0-855]
B
[0-3,259]
[0-87]
[0-217]
[0-79]
[0-44]
[0-37,953]

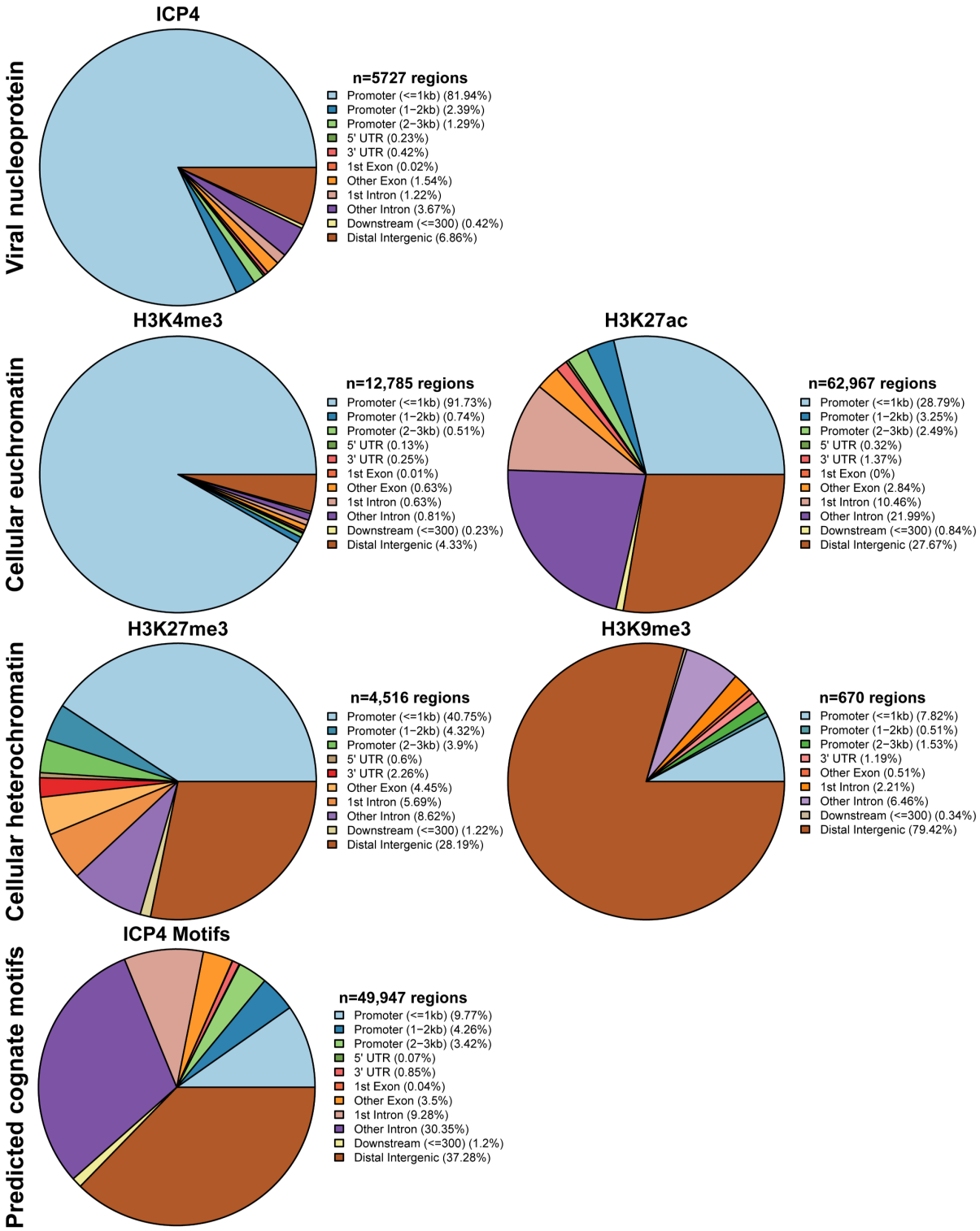
Supplemental Figure 9 Epigenetic markers on the viral genome

MRC5 cells were infected with an ICP4 null mutant (n12) or HSV-1 (WT) for 2 hr and ChIP-Seq for ICP4, Pol II, H3, H3K4me3, H3K27acetyl, H3K9me3, and H3K27me3 was performed. All data was normalized for sequencing depth and viral genome number using input ChIP-Seq reads. Viral ORFs are indicated, color coded by gene class with IE as yellow, E as green, L1 as blue, and L2 as purple. y-axes were (A) maintained the same between IP's or (B) maximized for each trace to aid in visualization.


Uninfected
Infected (2h)
ICP4
Pol II
H3
K4me3
K27ac
K9me3
K27me3
10kbp
[0-1,065]
[0-1,785]
[0-264]
[0-2,965]
[0-1,304]
[0-138]
[0-213]
[0-231]
[0-595]
[0-179]
[0-2,646]
[0-702]
[0-122]
[0-101]
[0-2,048]
[0-2,491]
[0-149]
[0-1,715]
[0-967]
[0-165]
[0-507]
[0-1,511]
[0-4,807]
[0-196]
[0-3,590]
[0-1233]
[0-494]
HYAL1
HYAL2
NPRL2
TUSC2
ZMYND10
CYB561D2
RASSF1
RASSF1-AS1
TMEM115
SYN2
TIMP4
PPARG
WDR88
LRP3
SLC7A10
STK11
CBARP
ATP5D
MIDN
CIRBP-AS1
EFNA2
C19orf24
CIRBP

Supplemental Figure 10 Epigenetic markers on the host genome during early infection

MRC5 cells were uninfected or infected with HSV-1 for 2 h, and ChIP-Seq for ICP4, Pol II, H3, H3K4me3, H3K27acetyl, H3K9me3, and H3K27me3 was performed. Data was aligned to the human genome (hg38) and normalized for sequencing depth.


Viral nucleoprotein
ICP4
n=5727 regions
Promoter (<=1kb) (81.94%)
Promoter (1−2kb) (2.39%)
Promoter (2−3kb) (1.29%)
5' UTR (0.23%)
3' UTR (0.42%)
1st Exon (0.02%)
Other Exon (1.54%)
1st Intron (1.22%)
Other Intron (3.67%)
Downstream (<=300) (0.42%)
Distal Intergenic (6.86%)
Cellular euchromatin
H3K4me3
n=12,785 regions
Promoter (<=1kb) (91.73%)
Promoter (1−2kb) (0.74%)
Promoter (2−3kb) (0.51%)
5' UTR (0.13%)
3' UTR (0.25%)
1st Exon (0.01%)
Other Exon (0.63%)
1st Intron (0.63%)
Other Intron (0.81%)
Downstream (<=300) (0.23%)
Distal Intergenic (4.33%)
H3K27ac
n=62,967 regions
Promoter (<=1kb) (28.79%)
Promoter (1−2kb) (3.25%)
Promoter (2−3kb) (2.49%)
5' UTR (0.32%)
3' UTR (1.37%)
1st Exon (0%)
Other Exon (2.84%)
1st Intron (10.46%)
Other Intron (21.99%)
Downstream (<=300) (0.84%)
Distal Intergenic (27.67%)
Cellular heterochromatin
H3K27me3
n=4,516 regions
Promoter (<=1kb) (40.75%)
Promoter (1−2kb) (4.32%)
Promoter (2−3kb) (3.9%)
5' UTR (0.6%)
3' UTR (2.26%)
Other Exon (4.45%)
1st Intron (5.69%)
Other Intron (8.62%)
Downstream (<=300) (1.22%)
Distal Intergenic (28.19%)
H3K9me3
n=670 regions
Promoter (<=1kb) (7.82%)
Promoter (1−2kb) (0.51%)
Promoter (2−3kb) (1.53%)
3' UTR (1.19%)
Other Exon (0.51%)
1st Intron (2.21%)
Other Intron (6.46%)
Downstream (<=300) (0.34%)
Distal Intergenic (79.42%)
Predicted cognate motifs
ICP4 Motifs
n=49,947 regions
Promoter (<=1kb) (9.77%)
Promoter (1−2kb) (4.26%)
Promoter (2−3kb) (3.42%)
5' UTR (0.07%)
3' UTR (0.85%)
1st Exon (0.04%)
Other Exon (3.5%)
1st Intron (9.28%)
Other Intron (30.35%)
Downstream (<=300) (1.2%)
Distal Intergenic (37.28%)

Supplemental Figure 11 Binding location of ICP4 relative to epigenetic markers and binding motifs

MRC5 cells were infected with HSV-1 for 2 h, and ChIP-Seq for ICP4, Pol II, H3, H3K4me3, H3K27acetyl, H3K9me3, and H3K27me3 was performed. Data was aligned to the human genome (hg38). IP peaks consistent between biological duplicate experiments were determined using MACS2. ChIPSeeker assessment of bound regions for each set of IP peaks.

3.6 Discussion

ICP4 as a sink for general transcription factors

ICP4 is synthesized shortly after the viral genome enters the nucleus and remains associated with the genome through all phases of infection. Our data demonstrated that ICP4 bound promiscuously to the viral genome prior to DNA replication. At this time point, the relative concentration of ICP4 to viral genomes was relatively high which likely promoted multimerization on DNA through ICP4-ICP4 interactions (147). We observed a similar phenotype for ICP4's interaction partner, Mediator (151, 152). Components of Mediator bound generally to the viral genome, concentrating near viral TATA boxes. Additional protein-Mediator interactions likely contribute to this distribution. This is a unique recruitment phenotype for Mediator which binds exclusively at cellular TSS via multiple protein-protein interactions. In the absence of ICP4, these interactions were not sufficient to support Mediator binding to the viral genome. With the exception of Mediator recruitment to IE promoters which does not require ICP4 and reflects the activity of VP16 (131, 132). Similarly, we observed a 2 to 10-fold decrease in recruitment of Pol II, TBP, and SP1 to viral E and L promoters without ICP4. This minimal level of recruitment is insufficient to support transcription, which explains why only IE transcripts are efficiently transcribed in the absence of ICP4.

ICP4-dependent GTF recruitment was not due to a global accessibility change. In the absence of ICP4 the viral genome remained absent of histones and had little change in tagmentation frequency. This is most likely due to the action of ICP0, which is an IE protein expressed in the absence of ICP4 and has been shown to preclude histones from the genome (112, 113). We posit that ICP4's ability to interact and recruit Mediator and TFIID generally to the viral genome creates a local concentration gradient. Ultimately this increases the incidence of Pol II transcription machinery recruitment to the viral genome, which is stabilized by contact with additional protein-DNA, protein-protein interactions. Our observation that ICP4 coats the viral genome, a unique recruitment phenotype for a protein that functions in GTF recruitment likely reflects the architecture of the viral genome. The condensed organization of the HSV genome results in overlap of viral promoters and ORFS, with little to no unused coding space. For these reasons, it may be advantageous for ICP4 to generally and dynamically bind to the viral genome. It is also possible that the relatively high density of ICP4 on the viral genome prior to DNA replication may serve to repress transcription of the true late promoters, which only contain binding sites for the core GTFs. It has been shown that ICP4 binding can impose an increased dependence on DNA replication for expression from relatively simple promoters (226, 227). These data demonstrated the critical role ICP4 serves as a general viral transcription factor, essential for activation and continued transcription of E and L genes.

ICP4 differentiates between the viral and cellular genome

ICP4 possesses the ability to bind to double stranded DNA independent of sequence, an ability that is facilitated by ICP4 oligomerization on the genome. At early time points, when the relative concentration of ICP4 to the viral genome was high, we observed promiscuous ICP4

binding. This coating phenotype provides an explanation for why no specific binding sites on the genome affect the ability of ICP4 to activate transcription (161, 228). Instead the high density of ICP4 binding motifs on the viral genome aggregate to create a global affinity for ICP4.

Early during infection, we also observed binding of ICP4 to cellular promoters, a novel observation. ICP4 bound promoters of genes which grouped ontologically to common housekeeping functions including pathways related to chromatin, transcription, and metabolism. Furthermore, we found that ICP4 specifically bound where there was an absence of histones, adjacent to euchromatic markers. ICP4 did not alter the chromatic markers of the promoters it bound to, rather ICP4 bound regions which were also accessible in the absence of infection. This likely reflects ICP4's ability to bind naked double-stranded DNA in a sequence independent manner. This ability is promoted by ICP4 multimerization (147), which is likely why we see the binding early during infection. The GTF's present on these cellular promoters, namely TFIID and Mediator complex, may further stabilize ICP4 binding at these promoters via their well-characterized interactions (150-152).

The consequences, if any, of this binding for the transcription of specific cellular genes remains to be determined. The binding of ICP4 to the cellular genome was greatly diminished by four hpi, which corresponded to 3–4 viral genome duplications. At this time point we also observed a decrease in ICP4 coating the viral genome. However, ICP4 still bound abundantly, concentrating adjacent to strong cognate binding sites. This is most likely due to replication of the viral genome producing more ICP4 binding targets. Simple mass action results in binding to predominantly higher affinity sites. We propose that the binding preference of ICP4 for the viral genome is due to the 100-fold higher density of cognate binding sites and absence of cellular histones.

ICP4 as both viral transcription factor and chromatin

HSV-1 productive infection generates 1,000–10,000 viral progeny per infected cell within a 24 h window. To facilitate this rampant transcriptional shift HSV-1 manipulates host Pol II machinery to prioritize viral mRNAs. By six hpi viral mRNA's comprise almost 50% of the total mRNA present in the host nucleus (213). Furthermore, binding of Pol II to cellular promoters dramatically decreases upon HSV-1 infection (224, 225). A recent study concluded that viral replication compartments efficiently enrich Pol II into membraneless domains (73). Herein we identified the viral factor responsible for coopting the host Pol II machinery.

McSwiggen et al. proposed this phenomenon was dependent on the absence of nucleosomes which made the viral genome 100-fold more accessible than the cellular genome. While we agree that this accessibility is critical for viral infection, we believe it is essential for ICP4 binding. Similar to cellular chromatin, ICP4 coats the viral genome throughout productive infection (Fig. 22).

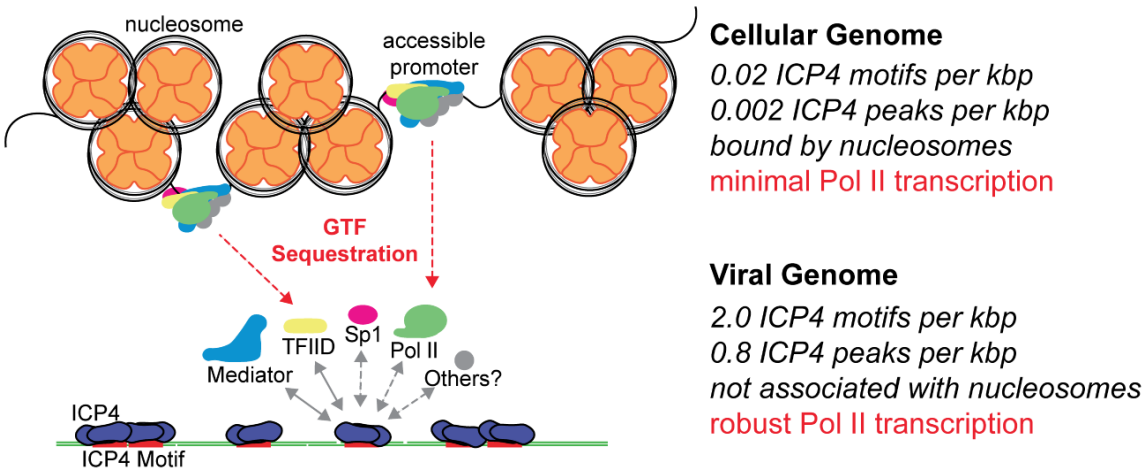


Figure 22 Model for ICP4 function.

ICP4 preferentially binds to the more accessible viral genome recruiting cellular transcription factors preferentially to the viral genome, thus activating the virus and inhibiting cellular transcription.

However, ICP4 also functions to scaffold Pol II transcription machinery to the viral genome. We demonstrated that Pol II depletion from cellular promoters was dependent on the number of ICP4 bound viral genomes. We propose that one or more components of the PIC, such as the ICP4 binding partners TFIID and Mediator, are limiting and ICP4 recruits these factors to the viral genome. As the number of viral genomes bound by ICP4 increases, the limiting PIC components no longer contacts cellular promoters. Ultimately this results in decreased Pol II occupancy on host promoters, preventing cellular transcription. This mechanism is essential to facilitate the rapidly progressing infection while limiting the extent to which the host can respond to viral challenge.

3.7 Materials and Methods

Cells and viruses

Vero (African green monkey kidney) and MRC5 (human fetal lung) cells were obtained from and propagated as recommended by ATCC. Viruses used in this study include n12 (229), tsKos (213) and KOS (230). n12 virus stocks were prepared and titered in a Vero-based ICP4 complementing cell line, E5. E5 cells were generated by stable transfection of Vero cells with ICP4 gene encoded on pK1-2 (GenBank Nucleotide JQ407535.1). E5 cells originate from a single colony, confirmed for its ability to complement the n12 virus and Western Blot assessed ICP4 expression (229). KOS virus stocks were prepared and titered in Vero cells. tsKos virus stocks were prepared and tittered in Vero cells at permissive temperature (33.5°C). All cells are regularly tested for the presence of mycoplasma contamination. Cells used in this study were mycoplasma free.

Antibodies

The following antibodies were used: Pol II 4H8 specific to phospho S5 of the C-terminal domain repeat region (AbCam #ab5408), TBP (AbCam #ab51841), SP1 (SantaCruz #sc-17824), Med23 (BD Pharmingen #550429), Med1 (Bethyl #A300-793A), H3K4me3 (Abcam #ab12209), H3K27me3 (AbCam #ab6002), H3K27acetyl (AbCam #ab4729), H3K9me3 (AbCam #ab176916), H3 (AbCam #1791), and ICP4 58S (derived from hybridomas-ATCC HB8183).

Viral infection

MRC5 cells were infected with 10 PFU per cell. Virus was adsorbed in tricine-buffered saline (TBS) for 1 hr at room temperature. Viral inoculum was removed, and cells were washed quickly

with TBS before adding 2% FBS media. Infected samples were incubated at 37°C unless otherwise specified.

ChIP-Sequencing

Infected cells were treated with 5 mL of 25% formaldehyde for 15 min at room temperature, followed by 5 mL of 2.5 M glycine. All following steps were performed at 4°C unless otherwise stated. Cultures were washed with TBS and scraped into 50 mL of FLB [5 mM 1,4-Piperazinediethanesulfonic acid (PIPES) pH 8, 85 mM KCl, 0.5% Igepal CA-630, 1x Roche protease inhibitor cocktail]. Cells were pelleted by low-speed centrifugation, resuspended in 1.1 mL RIPA buffer [1x phosphate-buffered saline (PBS), 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 1x Roche protease inhibitor cocktail]. Sample was sonicated for 6 intervals of 30 s with a Sonics Vibra-Cell VCX 130 sonicator equipped with a 3 mm microprobe and pelleted at 2000 x g for 15 min. 50 µl was stored as an input control, and the remainder was divided equally to use in immunoprecipitations (IP). $2–4 \times 10^7$ MRC5 cells were applied per IP. Samples were immunoprecipitated with 25 µg (TBP, SP1), or 10 µg (Pol II, H3K4me3, H3K27me3, H3K27ac, H3K9me3, H3) antibody. Antibody was previously bound to 50 µL of Dynabeads M280 sheep anti-mouse IgG beads, or Dynabeads M280 sheep anti-rabbit IgG beads in 5% bovine serum albumin (BSA) 1x PBS overnight. DNA samples were bound to the antibody-bead complex overnight rotating. The IP mixtures were washed seven times with LiCl wash buffer [100 mM Tris-HCl buffer pH 7.5, 500 mM LiCl, 1% Igepal CA-630, 1% sodium deoxycholate] and once with Tris-EDTA (TE) buffer. Beads were resuspended in IP elution buffer [1% SDS, 0.1M $NaHCO_3$] and incubated at 65°C for 2 hr 900 rpm. Input aliquot was suspended in IP elution buffer. Input and IP samples were incubated at 65°C 900 rpm overnight. The samples were extracted with

phenol-chloroform-isoamyl alcohol (25:24:1) and with chloroform-isoamyl alcohol (24:1) and then purified using Qiagen PCR cleanup columns. Each sample was quantified using a Qubit 2.0 fluorometer (Invitrogen) and 2–20 ng was used to create sequencing libraries using the NEBNext Ultra II DNA Library preparation kit (NEB #E7103S). Libraries were quantified using the Agilent DNA 7500 Kit, and samples were mixed together at equimolar concentration. Illumina HiSeq 2500 single-end 50 bp sequencing was carried out at the Tufts University Core Facility.

ATAC-Sequencing

We adapted the protocol from Buenrostro et al. (2013). Briefly, 2 million MRC5 cells were plated into 60 mm dishes and allowed to grow overnight. Cells were infected as described above. Uninfected and n12 infected cells were harvested at four hpi. WT HSV-1 infected cells were harvested pre-replication at two hpi. Infected samples were washed once with chilled TBS and lysis-1 buffer [10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$]. Samples were incubated with 2 mL lysis-2 buffer [10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA-630] for 3 min on ice. Cells were gently resuspended and dounced until nuclei were visible via trypan blue staining. Nuclei were spun at 500 g for 10 min at 4°C and resuspended in lysis-1 buffer. 250 µL (5×10^5 cells) was transferred to an epindorf tube and spun at 500 g for 10 min at 4°C. Nuclei were resuspended in 22 µL buffer TD (Illumina Catalog No. 15027866) 2.5 µL TDE1 (Illumina Catalog No. 15027865) and 22.5 µL water and incubated at 37°C for 30 min gently shaking. DNA was purified using the MinElute PCR purification kit (Qiagen Cat No./ID: 28004). PCR amplification was performed for 8–14 total cycles. Libraries were quantified using the Agilent DNA 7500 Kit, and samples were mixed together at equimolar concentration. Sequencing was carried out at the Tufts University Core Facility, Illumina HiSeq 2500 single-end 50 bp

sequencing was carried out for replicate 1, Illumina Mid-Output NextSeq was carried out for replicates 2 and 3.

Data analysis

ChIP-Seq

Data was uploaded to the Galaxy web platform, and we used the public server at usegalaxy.org to analyze the data (206). Data was first aligned using Bowtie2 (210) to the human genome (hg38), and then unaligned reads were mapped to the HSV-1 strain KOS genome (KT899744.1). Bam files were visualized using DeepTools bamcoverage (231) with a bin size of 1 to generate bigwig files. Data was viewed in IGV viewer and exported as EPS files. Bigwig files were normalized for sequencing depth and genome quantity. Mapped reads were multiplied by the 'norm factor' which was calculated as the inverse of $\frac{Input\ cellular\ reads}{Input\ cellular+viral\ mapped\ reads\ (TMR)} \times Billion\ sample\ TMR$ or $\frac{Input\ viral\ reads}{TMR} \times Million\ sample\ TMR$. ChIP-Seq experiments were repeated for a total of 2 to 4 biological replicates. The normalized bigwig files were averaged between replicates. Heatmaps and gene profiles were generated using MultiBigwigSummary (231) on normalized cellular bigwig files to all UCSC annotated mRNAs. Gene profiles and heatmaps were plotted using plotProfile and plotHeatmap (231). Spearman correlation analysis was performed using deeptools plotCorrelation on multiBigwigSummary limited to cellular transcripts (231).

Peak calling

Viral peaks were called using MACS2 call peak (232), pooling treatment and control files for each condition. Due to the small size of the viral genome (151974 bp) we could not use the shifting model option (--nomodel). To offset the dense binding of ICP4 we used a fixed background lambda as local lambda for every peak region and a more sophisticated signal processing approach to find subpeak summits in each enriched peak region (--call-summits).

Cellular peaks were called using MACS2 (232). We first removed non-uniquely mapped sequences with SAMtools, filter SAM or BAM for a minimum MAPQ quality score of 20 (233). We determined the approximate extension size for each IP using MACS2 predictd, and averaging the size estimate between replicates. We ran MACS2 call peak for individual replicates and pooled samples with no shifting model (--nomodel). To determine high confidence peaks present in each MACS2 output we used Galaxy Operate on Genomic Intervals, Join. Peak intersection was analyzed for intersection size and jaccard statistic using JaccardBed (231). ChIPseeker was run on MACS2 outputs to assess the cellular regions bound in each condition (234).

Motif discovery

Bedtools Multiple Intersect (211) was used to compare the MACS2 output for ICP4 IP at 2, 4, and 6 hr. A BED file was generated for regions + /- 100 bp from the summits of each identified peak. Peaks in common between all three experimental conditions were used to generate a fasta file using GetFastaBed (211) Peaks present in all three time points were submitted to MEME v.4.11.1.0 for motif analysis (235). The consensus sequence in Figure 16 had the most significant E-value, and was the only motif found in more than five peaks.

Correlation analysis

To assess quality and reproducibility of data we assessed normalized bigwig files for each IP replicate. For cellular and viral alignments we ran MultiBigwigSummary (231) with a bin size of 10,000 and 50 bp, respectively. Raw bin counts were plotted and a linear regression analysis was performed (Supplemental Fig. 7-8).

ATAC-Seq

Data was first aligned using Bowtie2 (210) to the human genome (hg38), and then unaligned reads were mapped to the HSV-1 strain KOS genome (KT899744.1) with the following parameters: –no-unal –local –very-sensitive-local –nodiscordant –no-mixed –contain –overlap –dovetail –phred33. Approximate fragment size for paired end read data was determined from mapped bam files using CollectInsertSizeMetrics (Broad Institute, Picard). Bam files were visualized using DeepTools bamcoverage (231) with a bin size of 1 to generate bigwig files. Data was viewed in IGV viewer and exported as EPS files. Cellular bigwig files were normalized for sequencing depth, the y-axes values are mapped reads per billion total reads. Viral bigwig files were further normalized per viral genome copy number, see Table 3C. The y-axes values are mapped reads per billion total reads per viral genome copy number. The normalized bigwig files were averaged between three biological replicates. Heatmaps and gene profiles were generated using MultiBigwigSummary (231) on normalized cellular bigwig files to all UCSC annotated mRNAs. Gene profiles and heatmaps were plotted using plotProfile and plotHeatmap (231). To calculate the percentage of total DNA corresponding to the virus or host in n12 and WT HSV-1 infection, we utilized ChIP-Seq input reads. We calculated the average percentage of total reads which mapped to either the virus or host in four biological replication ChIP-Seq samples. We used this

value to calculate the number of viral genomes contained within each nucleus. This value was used to determine the tagmentation enrichment observed relative to the actual amount of genome content present.

4.0 Summary and Discussion

4.1 Summary of Thesis

The transcriptional cascade of HSV-1 has been known since the 1970's, however studies investigating transcriptional dynamics have focused on select transcripts that typify each of the kinetic classes, i.e. thymidine kinase, glycoprotein C, glycoprotein D. Additional *in vitro* work has elucidated some of the core differences in Pol II transcription of viral and cellular genes. Recent NGS advancements allowed us to investigate the transcriptional environment in primary cells throughout HSV-1 infection. Thus we were able to determine a global unbiased view of both the host and pathogen.

We propose a model for how viral DNA replication results in the differential utilization of cellular factors that function in transcription initiation, delineating the kinetics of HSV-1 transcription. SP1 bound to robustly transcribed viral IE and E genes prior to viral genome replication. Other viral promoters were silenced, absent PIC recruitment. After a single genome duplication, the viral genome possessed enhanced recruitment of the PIC to all viral promoters. SP1 binding to the viral genome decreased after 3-4 rounds or replication, coinciding with a decreased rate of early gene transcription. We determined that genes highly transcribed later during infection possessed a canonical Inr with stable TAF1 binding and robust PIC recruitment.

Our work elucidated the critical role that ICP4 plays throughout the viral life cycle. We determined ICP4 discriminately binds the viral genome due to the absence of cellular nucleosomes and high density of cognate binding sites. We posit that ICP4's ability to recruit not just Pol II, but also more limiting essential components, such as TBP and Mediator, create a competitive

transcriptional environment. These distinguishing characteristics ultimately result in a rapid and efficient reprogramming of the host's transcriptional machinery, which does not occur in the absence of ICP4.

4.2 General Discussion

Transcription dynamics associated with viral genome replication.

We determined that the act of viral genome replication is not permanently coupled to late gene transcription. Instead initial rounds of genome replication are sufficient to license late gene transcription. We then used ChIP-Seq to assess Pol II PIC formation on the viral genome in the presence and absence of genome replication. We found that the viral genome appeared in a repressive state, prior to the onset of replication as evidenced by the limited number of PIC's on viral promoters. After a single round of replication, we found formation of PIC's on previously silenced viral promoters (L2 genes). We also found that genome replication enhanced formation of PIC's on promoters which were previously active (E and L1 genes). Our data demonstrated that the viral genome achieved peak transcriptional activity at 3 hpi, or after a single genome duplication. Later during infection, 6 hpi, the transcriptional activity per genome sharply decreased. We believe this demonstrates that late during infection only a subset of viral genomes are transcriptionally active, whereas others undergo replication and packaging. This model is supported by additional work from our lab, in which nascent viral DNA adjacent to replication forks was greatly enriched for cellular GTF's including Mediator, Integrator, Pol II, and TFIID. In contrast nascent DNA farther away from replication forks was enriched for viral packaging components including the portal protein, UL6 (164). This may be a mechanism by which the virus

partitions different genomes to efficiently coordinate their use as templates for transcription, or replication, or substrate for packaging.

Transcriptional silencing of early genes late during infection.

HSV-1 early gene transcription begins pre-replication, with transcription rates peaking around 4 hpi. This is distinct from L1 gene transcription which begins pre-replication, with transcription rates peaking around 6-8 hpi. E gene peptides actually reach higher levels of synthesis in infected cells where replication has been inhibited by addition of ACV or PAA. Our findings suggest two reasons for the decrease in E gene transcription.

First, we found that the cellular enhancer, SP1, which activates many E genes was no longer bound at these promoters at 4 hpi, or after 3-4 genome duplications. Our lab has previously determined that SP1 is phosphorylated around 4 hpi during HSV-1 infection (196), we posit that this phosphorylation event may be responsible for the decrease in SP1 binding to the viral genome. Alternatively the amount of SP1 may be limiting and additional rounds of genome replication titrate out the amount of SP1 to below the amount required to enhance E gene promoters. We would like to note that there were E and L1 genes, which were not bound by SP1, but still had robust PIC recruitment. These include UL2, UL12, UL30, UL34, US4, US6, US7, US8. We would expect that PIC recruitment to these gene promoters is facilitated prior to genome replication by other cellular enhancers, such as NF1 (162). This requires further investigation.

Second, we observed a significant deviation in the Inr of E genes from the canonical BBCA$_{+1}$BW. Using our ChIP-Seq data for TBP, and TAF1, the respective binding partners of TATA boxes and Initiator elements, we were able to call motifs with greater confidence as they were limited by the relative distance from their binding partner. We noted that viral transcripts

produced late during infection possessed canonical Inr elements and enriched TAF1 binding. We propose that strong contact between TAF1 and the viral promoter becomes critical later during infection to stabilize PIC complex and for this reason E gene transcription decreases later during infection.

Early during infection TAF1 did not form strong distinct binding peaks on the viral genome. This may represent a dynamic recruitment of TAF1 to viral promoters stabilized by protein-protein interactions rather than direct interaction with cis-binding elements. ICP4 has been shown to interact directly with TAF1 recruiting the TFIID complex (150-152). We showed that ICP4 densely coats the genome also early during infection. Thus we hypothesize that ICP4 may be responsible for the sequence nonspecific interaction of TAF1 with the viral genome. We showed that the amount of ICP4 bound to the genome decreases later during infection with ICP4 predominantly bound to cognate binding motifs. This may be the reason why canonical Inr elements are now required for TAF1 recruitment. Further studies using an ICP4 C-terminal truncation mutant (n208), which lacks the ability to multimerize (147), are necessary. If our hypothesis is correct we would expect TAF1 binding to be limited to initiator elements or ICP4 motifs, as ICP4 can no longer multimerize and coat the viral genome.

ICP4-dependent activation of E and L gene transcription

Using ChIP-Seq we assessed ICP4 binding dynamics on the viral genome. We found that ICP4 coated the viral genome early during infection, and that rather than occluding the genome, this coating phenotype was associated with recruitment of the Pol II machinery to viral E and L promoters. ICP4 binding was enriched at the following motif: DTSGKBDTBNHSG (Fig. 16F),

where D is A, G, T; S is C or G; K is G or T; B is C, G, T; H is A, C, T. The ICP4 binding motif was littered densely throughout the viral genome with 2 motifs per Kbp.

We queried whether the act of genome replication altered requirements for ICP4. Using a temperature sensitive mutant, we were able to selectively deplete ICP4 after the start of genome replication. We found that in the absence of ICP4, even with continued genome replication, there was no recruitment of Pol II to E or L gene promoters. Our studies demonstrated that ICP4 is always essential for E and L gene transcription and that this is due to ICP4's role in recruiting and stabilizing the PIC. This makes sense in light of our additional work demonstrating that in the absence of ICP4, Pol II, TBP, Sp1, and Mediator were not recruited to the viral genome. These results are consistent with previous work from our lab showing that ICP4 interacts with and recruits various components of the Pol II machinery including: Mediator, TFIID, TFIIH, elongation factors (CDK9, SUPT5H, SUPT6h), CPSF, SWI-SNF, Nurd, Ino80 (152, 153). Although ICP4 interacts with various chromatin remodelers, we did not observe chromatinization of the viral genome in the absence of ICP4. Meaning ICP4 acts to recruit and scaffold, not to affect genomic accessibility or chromatinization

Mechanism of ICP4 binding to the viral genome

ICP4 possesses the ability to bind nonspecifically to double stranded DNA However in immunofluorescence experiments ICP4 colocalizes almost exclusively with viral pre- and replication compartments. In our ChIP-Seq data we found that ICP4 bound to the promoters of select host housekeeping genes early during infection. This binding occurred independent of ICP4 binding motifs and was correlated with euchromatic markers, namely H3K4me3 and H3K27acetyl. Using ATAC-Seq we determined that the viral genome was not bound by host nucleosomes.

Furthermore we found no evidence of histone H3 binding. We concluded that early during infection ICP4 bound indiscriminately to "naked" or accessible DNA, whether that is the viral genome or open regions of host chromatin. Later during infection, ICP4 is at a relatively lower concentration and can no longer multimerize. This caused ICP4 to preferentially bind the viral genome due to the higher incidence of ICP4 binding motifs, ~100-fold more motifs/kbp than the cellular genome. Reduced ICP4 binding to the cellular genome late during infection may also be a result of host heterochromatization, this remains to be studied.

ICP4 mediates HSV-1 host transcriptional shut off

Recent work has highlighted HSV-1's ability to coopt the host Pol II machinery at the expense of cellular transcription (73, 224, 225). Due to it's role in viral Pol II recruitment, we expected that ICP4 would be required for reduction of Pol II on host promoters. We found that early during infection with an ICP4 null mutant there was no decrease in host Pol II recruitment. This was in contrast to a 1.5-fold reduction in Pol II recruitment to host promoters after wild-type infection. We again used a temperature sensitive mutant to selectively deplete ICP4 later during infection after the onset of genome replication. We found that ICP4 was always required for reduction of Pol II on host promoters, independent of the number of viral genomes present. Thus we propose a model in which ICP4 binds to the viral genome and promotes recruitment of Pol II and associated transcription factors to the viral genome. As the amount of ICP4 is in excess to viral genomes early during infection we observe a dependence upon replication, simply because it increases the number of ICP4 bound genomes. However later during infection when ICP4 becomes limited relative to the viral genome, additional rounds of genome replication do not affect Pol II recruitment to cellular promoters. Here we investigated Pol II depletion from host promoters.

However due to ICP4's role in recruiting a variety of Pol II GTF's including TFIID, TFIIH, and Mediator, we expect that these factors are also depleted from host promoters during infection.

5.0 Future Directions

We present an updated understanding of the basic transcriptional requirements and mechanisms that occur during HSV-1 productive infection. Our findings elucidate current gaps in knowledge that require further work to address:

1. How does the viral genome change after replication?

Our findings clearly indicate a shift in the form of the viral genome from repressive to permissive after a single round of genome replication. As only one genome duplication is required for this change, it is highly unlikely that genome replication titrates out a repressive factor. Furthermore, we found that the viral genome was not bound by host nucleosomes or histone H3. IF studies from our lab have also demonstrated exclusion of Histone H1 from replication compartments (75). This leads us to propose epigenetic modification of the viral genome itself. The host genome has been shown to be epigenetically modified via methylation of cytosine or adenine, these modified bases include 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), 5-carboxylcytosine (5-caC), and N^6-methyladenine (6-mA) (236). The genome of another herpesvirus, EBV, was shown to be heavily methylated inside mature virions, approaching 15% (237). Early studies found that the HSV-1 genome contained 5-mC at 4-9 hpi, but that mature virions did not contain methylated DNA (238). This study was performed over 40 years ago, and use of recent techniques such as affinity capture or bisulfite conversion followed by NGS may be more sensitive.

2. Does the DNA replication machinery directly enhance late gene transcription?

There are seven essential viral proteins required for HSV-1 DNA replication: DNA polymerase holoenzyme (UL30/UL42), helicase-primase complex (UL5/UL8/UL52), single stranded DNA binding protein (ICP8), and origin binding protein (UL9). Excepting UL9, these six components are conserved in all known herpesviruses (166, 167). It is possible that a component of the viral DNA replication machinery serves a dual function and is continuously required for late gene transcription. These factors may aid in recruitment of the Pol II machinery to the viral genome. For instance, the viral polymerase holoenzyme was reported to interact with cdc2, topoisomerase IIa, HCF-1, Asf1b, hsp90 and Fen-1 (156-160).

We partially addressed this hypothesis by inhibiting the viral polymerase with ACV and assessing late gene transcription. We found that ACV addition after the onset of replication did not inhibit continued late gene transcription. ACV is incorporated into nascent DNA and causes chain termination along with stalled replication forks (38). Although genome replication is stalled in this circumstance, the replication machinery may still associate with the viral genome sufficiently to serve as a scaffold for factors essential to late gene transcription. To address this hypothesis one would need to specifically deplete each component of the DNA replication machinery after the onset of genome replication and assess late gene transcription. We attempted to perform these experiments using preexisting temperature sensitive mutants. We found using tsA1 that ICP8 was not required for late gene transcription after the onset of replication. We were unable to make conclusions regarding the other replication components due to limitations of the temperature sensitive mutants.

3. What effect does ICP4 binding the cellular genome have?

We observed binding of ICP4 to euchromatic host promoters early during infection. We do not know what impact this binding event has on cellular transcription. Due to ICP4's role in recruiting various GTF's we hypothesize that ICP4 binding may transiently increase the transcription rate of these genes. When looking at RNA-Seq data sets which are a measure of mRNA abundance, we were not able to see any impact on the levels of ICP4 targeted transcripts. This is likely due to the short length of time which ICP4 interacts with host transcripts (~2 hours). Any impact ICP4 has on transcription is drowned out, considering the bulk of genes that ICP4 targeted are housekeeping genes which have constitutively high levels of abundance. This would need to be assessed using rate-determining techniques such as 4SU-Seq. Further investigation should also be done to determine the direct impact of ICP4 on host promoters outside the context of viral infection.

Bibliography

1. A. Davison *et al.*, The Order Herpesvirales. *Arch Virol* **154**, 171-177 (2008).

2. A. Davison *et al.*, The Order Herpesvirales. *Arch Virol* **154**, 171-177 (2009).

3. C. Zmasek, D. Knipe, P. Pellett, R. Scheuermann, Classification of human Herpesviridae proteins using Domain-architecture Aware Inference of Orthologs (DAIO). *Virology* **529**, 29-42 (2019).

4. Anonymous, *Fields Virology* (Lippincott Williams & Wilkins, Philadelphia, ed. 6th, 2014).

5. A. Davison, Herpesvirus systematics. *Vet Microbiol* **143**, 52-69 (2010).

6. H. Bradley, L. Markowitz, T. Gibson, G. McQuillan, Seroprevalence of Herpes Simplex Virus Types 1 and 2—United States, 1999–2010. *J Infect Dis* **209**, 325-333 (2014).

7. F. Bastian, A. Rabson, C. Yee, T. Tralka, Herpesvirus hominis: isolation from human trigeminal ganglion. *Science* **178**, 306-307 (1972).

8. J. Baringer, P. Swoveland, Recovery of herpes-simplex virus from human trigeminal ganglions. *N Engl J Med* **288**, 648-650 (1973).

9. R. Belshe *et al.*, Efficacy results of a trial of a herpes simplex vaccine. *N Engl J Med* **366**, 34-43 (2012).

10. D. Bernstein *et al.*, Epidemiology, clinical presentation, and antibody response to primary infection with herpes simplex virus type 1 and type 2 in young women. *Clin Infect Dis* **56**, 344-351 (2013).

11. N. Rowe, C. Heine, C. Kowalski, Herpetic whitlow: an occupational disease of practicing dentists. *J Am Dent Assoc* **105**, 471-473 (1982).

12. A. Tullo, Pathogenesis and management of herpes simplex virus keratitis. *Eye* **17**, 919-922 (2003).

13. A. Farooq, D. Shukla, Herpes Simplex Epithelial and Stromal Keratitis: An Epidemiologic Update. *Surv Opthamol* **57**, 448-462 (2012).

14. K. Tyler, Acute viral encephalitis. *N Engl J Med* **379**, 557-566 (2018).

15. S. James, J. Sheffield, D. Kimberlin, Mother-to-Child Transmission of Herpes Simplex Virus. *J Pediatric Infect Dis Soc* **3**, S19-S23 (2014).

16. C. Jones, Herpes simplex virus infection in the neonate: clinical presentation and management. *Neonatal Netw* **15**, 11-15 (1996).

17. R. Whitley, E. Davis, N. Suppapanya, Incidence of neonatal herpes simplex virus infections in a managed-care population. *Sex Transm Dis* **34**, 704-708 (2007).

18. M. Jenkins, S. Kohl, New aspects of neonatal herpes. *Infect Dis Clin North Am* **6**, 57-74 (1992).

19. F. Raschilas *et al.*, Outcome of and prognostic factors for herpes simplex encephalitis in adult patients: results of a multicenter study. *Clin Infect Dis* **35**, 254-260 (2002).

20. D. Buss, M. Scharyj, Herpesvirus infection of the esophagus and other visceral organs in adults. Incidence and clinical significance. *Am J Med* **66**, 457-462 (1979).

21. P. Ramsey, K. Fife, R. Hackman, J. Meyers, L. Corey, Herpes simplex virus pneumonia: clinical, virologic, and pathologic features in 20 patients. *Ann Intern Med* **97**, 813-820 (1982).

22. P. Bruynseels *et al.*, Herpes simplex virus in the respiratory tract of critical care patients: a prospective study. *Lancet* **362**, 1536-1541 (2003).

23. B. Kaufman, S. Gandhi, E. Louie, R. Rizzi, P. Illei, Herpes simplex virus hepatitis: case report and review. *Clin Infect Dis* **24**, 334-338 (1997).

24. S. Delis *et al.*, Herpes simplex colitis in a child with combined liver and small bowel transplant. *Pediatr Transplant* **5**, 374-377 (2001).

25. J. Meyers, N. Flournoy, E. Thomas, Infection with herpes simplex virus and cell-mediated immunity after marrow transplant. *J Infect Dis* **142**, 338-346 (1980).

26. F. Morfin, D. Thouvenot, Herpes simplex virus resistance to antiviral drugs. *J Clin Virol* **26**, 29-37 (2003).

27. V. Erard, A. Wald, L. Corey, W. Leisenring, M. Boeckh, Use of long-term suppressive acyclovir after hematopoietic stem-cell transplantation: impact on herpes simplex virus (HSV) disease and drug-resistant HSV disease. *J Infect Dis* **196**, 266-270 (2007).

28. M. Mahic *et al.*, Maternal Immunoreactivity to Herpes Simplex Virus 2 and Risk of Autism Spectrum Disorder in Male Offspring. *mSphere* **2**, e00016-00017 (2017).

29. C. Gillberg, Onset at age 14 of a typical autistic syndrome. A case report of a girl with herpes simplex encephalitis. *J Autism Dev Disord* **16**, 369-375 (1986).

30. G. DeLong, S. Bean, F. r. Brown, Acquired reversible autistic syndrome in acute encephalopathic illness in children. *Arch Neurol* **38**, 191-194 (1981).

31. I. Gillberg, Autistic syndrome with onset at age 31 years: herpes encephalitis as a possible model for childhood autism. *Dev Med Child Neurol* **33**, 920-924 (1991).

32. S. Harris, E. Harris, Molecular Mechanisms for Herpes Simplex Virus Type 1 Pathogenesis in Alzheimer's Disease. *Front Aging Neurosci* **10**, 48 (2018).

33. G. Elion *et al.*, The selectivity of action of an antiherpetic agent, 9-(2-hydroxyethoxymethyl) guanine. *Proc Natl Acad Sci USA* **74**, 5716-5720 (1978).

34. J. Fyfe, P. Keller, P. Furman, R. Miller, G. Elion, Thymidine kinase from herpes simplex virus phosphorylates the new antiviral compound, 9-(2-hydroxyethoxymethyl)guanine. *J Biol Chem* **253** (1978).

35. W. Miller, R. Miller, Phosphorylation of acyclovir (acyclo- guanosine) monophosphate by GMP kinase. *J Biol Chem* **255**, 7204-7207 (1980).

36. W. Miller, R. Miller, Phosphorylation of acyclovir diphos- phate by cellular enzymes. *Biochem Pharmacol* **31**, 3879-3884 (1982).

37. D. Derse, Y.-C. Cheng, P. Furman, M. St Clair, G. Elion, Inhibition of purified human and herpes simplex virus-induced DNA polymerases by 9-(2-hydroxyethoxymethyl)guanine triphosphate: effects on primer-template function. *J Biol Chem* **256**, 11447-11451 (1981).

38. P. Furman, M. St. Clair, T. Spector, Acyclovir triphosphate is a suicide inactivator of the herpes simplex virus DNA polymerase. *J Biol Chem* **259**, 9575–9579 (1984).

39. L. Beauchamp, G. Orr, P. de Miranda, T. Burnette, T. Krenitsky, Amino acid ester prodrugs of acyclovir. *Antiviral Chem Chemother* **3**, 157-164 (1992).

40. S. Weller *et al.*, Pharmacokinetics of the acyclovir pro-drug valaciclovir after escalating single- and multiple-dose administration to normal volunteers. *Clin Pharmacol Ther* **54**, 595-605 (1993).

41. R. Vere Hodge, Famciclovir and Penciclovir. The Mode of Action of Famciclovir Including Its Conversion to Penciclovir. *Antiviral Chem Chemother* **4**, 67-84 (1993).

42. D. Earnshaw *et al.*, Mode of antiviral action of penciclovir in MRC-5 cells infected with herpes simplex virus type 1 (HSV-1), HSV-2, and varicella-zoster virus. *Antimicrob Agents Chemother* **36**, 2747–2757 (1992).

43. R. Vere Hodge, D. Sutton, M. Boyd, M. Harnden, R. Jarvest, Selection of an oral prodrug (BRL 42810; famciclovir) for the antiherpesvirus agent BRL 39123 [9-(4-hydroxy-3-hydroxymethylbut-l-yl) guanine; penciclovir]. *Antimicrob Agents Chemother* **33**, 1765-1773 (1989).

44. J. Piret, G. Boivin, Resistance of Herpes Simplex Viruses to Nucleoside Analogues: Mechanisms, Prevalence, and Management. *Antimicrob Agents Chemother* **55**, 459-472 (2011).

45. K. Lundstrom, Viral Vectors in Gene Therapy. *Diseases* **6**, 42 (2018).

46. X. Dai, Z. Hong Zhou, Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. *Science* **360**, eaao7298 (2018).

47. J. Heming, J. Conway, F. Homa, Herpesvirus capsid assembly and DNA packaging. *Adv Anat Embryol Cell Biol* **223**, 119-142 (2017).

48. H. Granzow *et al.*, Egress of alphaherpesviruses: comparative ultrastructural study. *J Virol* **75**, 3675-3684 (2001).

49. M. Hollinshead *et al.*, Endocytic tubules regulated by Rab GTPases 5 and 11 are used for envelopment of herpes simplex virus. *EMBO J* **31**, 4204-4220 (2012).

50. S. Loret, G. Guay, R. Lippé, Comprehensive Characterization of Extracellular Herpes Simplex Virus Type 1 Virions. *J Virol* **82**, 8605–8618 (2008).

51. B. Kelly, C. Fraefel, A. Cunningham, R. Diefenbach, Functional roles of the tegument proteins of herpes simplex type 1. *Virus Res* **145**, 173-186 (2009).

52. K. Grünewald *et al.*, Three-Dimensional Structure of Herpes Simplex Virus from Cryo-Electron Tomography. *Science* **302**, 1396-1398 (2003).

53. W. Gibson, B. Roizman, Compartmentalization of spermine and spermidine in the herpes simplex virion. *Proc Natl Acad Sci USA* **68**, 2818–2821 (1971).

54. E. Kieff, S. Bachenheimer, B. Roizman, Size, composition, and structure of the deoxyribonucleic acid of herpes simplex virus subtypes 1 and 2. *J Virol* **8**, 125-132 (1971).

55. R. Colgrove *et al.*, History and genomic sequence analysis of the herpes simplex virus 1 KOS and KOS1.1 sub-strains. *Virology* **487**, 215-221 (2016).

56. N. Biswal, B. Murray, M. Benyesh-Melnick, Ribonucleotides in newly synthesized DNA of herpes simplex virus. *Virology* **61**, 87-99 (1974).

57. N. Frenkel, B. Roizman, Separation of the herpesvirus deoxyribonucleic acid duplex into unique fragments and intact strand on sedimentation in alkaline gradients. *J Virol* **10**, 565-572 (1972).

58. H. Delius, J. Clements, A partial denaturation map of herpes simplex virus type 1 DNA: evidence for inversions of the unique DNA regions. *J Gen Virol* **33**, 125-133 (1976).

59. G. Hayward, R. Jacob, S. Wadsworth, B. Roizman, Anatomy of herpes simplex virus DNA: evidence for four populations of molecules that differ in the relative orientations of their long and short components. *Proc Natl Acad Sci USA* **72**, 4243-4247 (1975).

60. G. Watson *et al.*, Sequence and comparative analysis of the genome of HSV-1 strain McKrae. *Virology* **433**, 528-537 (2012).

61. A. W. Whisnant *et al.*, Integrative functional genomics decodes herpes simplex virus 1. *bioRxiv*, 603654 (2019).

62. I. Jurak *et al.*, Numerous Conserved and Divergent MicroRNAs Expressed by Herpes Simplex Viruses 1 and 2. *Journal of Virology* **84**, 4659-4672 (2010).

63. K. Luger, A. Mader, R. Richmond, D. Sargent, T. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).

64. T. Caterino, J. Hayes, Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem Cell Biol* **89**, 35-44 (2011).

65. D. Tremethick, Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**, 651–654 (2007).

66. C. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487–500 (2016).

67. J. Oh, N. Fraser, Temporal Association of the Herpes Simplex Virus Genome with Histone Proteins during a Lytic Infection. *J Virol* **82**, 3530–3537 (2007).

68. G. Cohen *et al.*, Structural analysis of the capsid polypeptides of herpes simplex virus types 1 and 2. *J Virol* **34**, 521–531 (1980).

69. P. Pignatti, E. Cassai, Analysis of herpes simplex virus nucleoprotein complexes extracted from infected cells. *J Virol* **36**, 816–828 (1980).

70. S. Leinbach, W. Summers, The structure of herpes simplex virus type 1 DNA as probed by micrococcal nuclease digestion. *J Gen Virol* **51**, 45-59 (1980).

71. A. Lentine, S. Bachenheimer, Intracellular organization of herpes simplex virus type 1 DNA assayed by staphylococcal nuclease sensitivity. *Virus Res* **16**, 275-292 (1990).

72. M. Hu, D. Depledge, E. Flores Cortes, J. Breuer, L. Schang, Chromatin dynamics and the transcriptional competence of HSV-1 genomes during lytic infections. *PLoS Pathog* **15**, e1008076 (2019).

73. D. McSwiggen *et al.*, Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* **8**, e47098 (2019).

74. R. Gibeaul, K. Conn, M. Bildersheim, L. Schang, An Essential Viral Transcription Activator Modulates Chromatin Dynamics. *PLoS Pathog* **12**, e1005842 (2016).

75. J. Dembowski, N. DeLuca, Selective recruitment of nuclear factors to productively replicating herpes simplex virus genomes. *PLoS Pathog* **11**, e1004939 (2015).

76. K. Monier, J. Armas, S. Etteldorf, P. Ghazal, K. Sullivan, Annexation of the interchromosomal space during viral infection. *Nat Cell Biol* **2**, 661–665 (2000).

77. D. Knipe, A. Cliffe, Chromatin control of herpes simplex virus lytic and latent infection. *Nat Rev* **6**, 211-221 (2008).

78. A. Turner, B. Bruun, T. Minson, H. Browne, Glycoproteins gB, gD, and gHgL of herpes simplex virus type 1 are necessary and sufficient to mediate membrane fusion in a Cos cell transfection system. *J Virol* **72**, 873-875 (1998).

79. E. Heldwein, C. Krummenacher, Entry of herpesviruses into mammalian cells. *Cell Mol Life Sci* **65**, 1653–1668 (108).

80. A. Hilterbrand, E. Heldwein, Go go gadget glycoprotein!; HSV-1 draws on its sizeable glycoprotein tool kit to customize its diverse entry routes. *PLoS Pathog* **15**, e1007660 (2019).

81. F. Musarrat, N. Jambunathan, P. Rider, V. Chouljenko, K. Kousoulas, The Amino Terminus of Herpes Simplex Virus 1 Glycoprotein K (gK) Is Required for gB Binding to Akt, Release of Intracellular Calcium, and Fusion of the Viral Envelope with Plasma Membranes. *J Virol* **92**, e01842-01817 (2018).

82. A. Sears, B. McGwire, B. Roizman, Infection of polarized MDCK cells with herpes simplex virus 1: two asymmetrically distributed cell receptors interact with different viral proteins. *Proc Natl Acad Sci USA* **88**, 5087-5091 (1991).

83. L. Tran, J. Kissner, L. Westerman, A. Sears, A herpes simplex virus 1 recombinant lacking the glycoprotein G coding sequences is defective in entry through apical surfaces of polarized epithelial cells in culture and in vivo. *Proc Natl Acad Sci USA* **97**, 1818–1822 (2000).

84. B. Sodeik, M. Ebersold, A. Helenius, Microtubule-mediated transport of incoming herpes simplex virus 1 capsids to the nucleus. *J Cell Biol* **136**, 1007–1021 (1997).

85. K. Döhner *et al.*, Function of Dynein and Dynactin in Herpes Simplex Virus Capsid Transport. *Mol Biol Cell* **13**, 2795–2809 (2002).

86. M. McElwee, F. Beilstein, M. Labetoulle, F. Rixon, D. Pasdeloup, Dystonin/BPAG1 Promotes Plus-End-Directed Transport of Herpes Simplex Virus 1 Capsids on Microtubules during Entry. *J Virol* **87**, 11008 –11018 (2013).

87. V. Jovasevic, M. Naghavi, D. Walsh, Microtubule plus end–associated CLIP-170 initiates HSV-1 retrograde transport in primary human cells. *J Cell Biol* **211**, 323-337 (2015).

88. A. Copeland, W. Newcomb, J. Brown, Herpes simplex virus replication: roles of viral proteins and nucleoporins in capsid-nucleus attachment. *J Virol* **83**, 1660-1668 (2009).

89. P. Ojala, B. Sodeik, M. Ebersold, U. Kutay, A. Helenius, Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. *Mol Biol Cell* **20**, 4922-4931 (2000).

90. D. Pasdeloup, D. Blondel, A. Isidro, F. Rixon, Herpesvirus Capsid Association with the Nuclear Pore Complex and Viral DNA Release Involve the Nucleoporin CAN/Nup214 and the Capsid Protein pUL25. *J Virol* **83**, 6610-6623 (2009).

91. V. Jovasevic, L. Liang, B. Roizman, Proteolytic cleavage of VP1-2 is required for release of herpes simplex virus 1 DNA into the nucleus. *J Virol* **82**, 3311-3319 (2008).

92. U. Sae-Ueng *et al.*, Solid-to-fluid DNA transition inside HSV-1 capsid close to the temperature of infection. *Nat Chem Biol* **10**, 861-867 (2014).

93. W. Newcomb, S. Cockrell, F. Homa, J. Brown, Polarized DNA Ejection from the Herpesvirus Capsid. *J Mol Bio* **392**, 885-894 (2009).

94. C. Morgan, H. Rose, B. Mednis, Electron Microscopy of Herpes Simplex Virus: I. Entry. *J Virol* **2**, 507-516 (1968).

95. K. Miyamoto, C. Morgan, Structure and development of viruses as observed in the electron microscope. XI. Entry and uncoating of herpes simplex virus. *J Virol* **8**, 910-918 (1971).

96. R. Everett, "The Role of ICP0 in Counteracting Intrinsic Cellular Resistance to Virus Infection" in Alphaherpesviruses: Molecular Virology, S. Weller, Ed. (Caister Academic Press, Norfolk, UK, 2011), chap. 4.

97. R. Everett, Dynamic Response of IFI16 and Promyelocytic Leukemia Nuclear Body Components to Herpes Simplex Virus 1 Infection. *J Virol* **90**, 167-179 (2016).

98. C. Boutell, S. Sadis, R. Everett, Herpes Simplex Virus Type 1 Immediate-Early Protein ICP0 and Its Isolated RING Finger Domain Act as Ubiquitin E3 Ligases In Vitro. *J Virol* **76**, 841-850 (2002).

99. R. Everett, M.-L. Parsy, A. Orr, Analysis of the Functions of Herpes Simplex Virus Type 1 Regulatory Protein ICP0 That Are Critical for Lytic Infection and Derepression of Quiescent Viral Genomes. *J Virol* **83**, 4963–4977 (2009).

100. W. Halford, C. Kemp, J. Isler, D. Davido, P. Schaffer, ICP0, ICP4, or VP16 Expressed from Adenovirus Vectors Induces Reactivation of Latent Herpes Simplex Virus Type 1 in Primary Cultures of Latently Infected Trigeminal Ganglion Cells. *J Virol* **75**, 6143-6153 (2001).

101. C. Miller, R. Danaher, R. Jacob, ICP0 Is Not Required for Efficient Stress-Induced Reactivation of Herpes Simplex Virus Type 1 from Cultured Quiescently Infected Neuronal Cells. *J Virol* **80**, 3360-3368 (2005).

102. C. Preston, Reactivation of Expression from Quiescent Herpes Simplex Virus Type 1 Genomes in the Absence of Immediate-Early Protein ICP0. *J Virol* **81**, 11781-11789 (2007).

103. J. Arthur *et al.*, Herpes Simplex Virus Type 1 Promoter Activity during Latency Establishment, Maintenance, and Reactivation in Primary Dorsal Root Neurons In Vitro. *J Virol* **75**, 3885–3895 (2001).

104. R. Everett, A. Orr, C. Preston, A viral activator of gene expression functions via the ubiquitin-proteasome pathway. *EMBO J* **17**, 7161–7169 (1998).

105. M. Chelbi-Alix, H. de The, Herpes virus induced proteasome dependent degradation of the nuclear bodies-associated PML and Sp100 proteins. *Oncogene* **18**, 935–941 (1999).

106. J. Parkinson, R. Everett, Alphaherpesvirus Proteins Related to Herpes Simplex Virus Type 1 ICP0 Affect Cellular Structures and Proteins. *J Virol* **74**, 10006–10017 (2000).

107. J. Parkinson, S. Lees-Miller, R. Everett, Herpes simplex virus type 1 immediate-early protein vmw110 induces the proteasome-dependent degradation of the catalytic subunit of DNA-dependent protein kinase. *J Virol* **73**, 650-657 (1999).

108. R. Everett, W. Earnshaw, J. Findlay, P. Lomonte, Specific destruction of kinetochore protein CENP-C and disruption of cell division by herpes simplex virus immediate-early protein Vmw110. *EMBO J* **18**, 1526–1538 (1999).

109. P. Lomonte, K. Sullivan, R. Everett, Degradation of nucleosome-associated centromeric histone H3-like protein CENP-A induced by herpes simplex virus type 1 protein ICP0. *J Biol Chem* **276**, 5829–5835 (2001).

110. I. Jurak, L. Silverstein, M. Sharma, D. Coen, Herpes simplex virus is equipped with RNA- and protein-based mechanisms to repress expression of ATRX, an effector of intrinsic immunity. *J Virol* **86**, 10093–10102 (2012).

111. M. Orzalli, N. DeLuca, D. Knipe, Nuclear IFI16 induction of IRF-3 signaling during herpesviral infection and degradation of IFI16 by the viral ICP0 protein. *Proc Natl Acad Sci USA* **109**, E3008–E3017 (2012).

112. A. Cliffe, D. Knipe, Herpes simplex virus ICP0 promotes both histone removal and acetylation on viral DNA during lytic infection. *J Virol* **82**, 12030-12038 (2008).

113. M. Ferenczy, N. DeLuca, Epigenetic modulation of gene expression from quiescent herpes simplex virus genomes. *J Virol* **83**, 8514-8524 (2009).

114. P. Lomonte *et al.*, Functional interaction between class II histone deacetylases and ICP0 of herpes simplex virus type 1. *J Virol* **78**, 6744-6757 (2004).

115. H. Gu, Y. Liang, G. Mandel, B. Roizman, Components of the REST/CoREST/histone deacetylase repressor complex are disrupted, modified, and translocated in HSV-1-infected cells. *Proc Natl Acad Sci USA* **102**, 7571-7576 (2005).

116. H. Gu, B. Roizman, Herpes simplex virus-infected cell protein 0 blocks the silencing of viral DNA by dissociating histone deacetylases from the CoREST-REST complex. *Proc Natl Acad Sci USA* **104**, 17134-17139 (2007).

117. J. Alwine, W. Steinhart, C. Hill, Transcription of herpes simplex type 1 DNA in nuclei isolated from infected HEp-2 and KB cells. *Virology* **60**, 302-307 (1974).

118. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-143 (2015).

119. K. Adelman, J. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731 (2012).

120. B. Roeder, 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat Struct* **26**, 783–791 (2019).

121. S. Zabierowski, N. DeLuca, Differential Cellular Requirements for Activation of Herpes Simplex Virus Type 1 Early (tk) and Late (gC) Promoters by ICP4. *J Virol* **78**, 6162–6170 (2004).

122. S. Zabierowski, N. DeLuca, Stabilized binding of TBP to the TATA box of herpes simpex virus type 1 early (tk) and late (gC) promoters by TFIIA and ICP4. *J Virol* **82**, 3546-3554 (2008).

123. R. Honess, B. Roizman, Regulation of herpesvirus macromolecular synthesis, I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol* **14**, 8-19 (1974).

124. R. Honess, B. Roizman, Regulation of herpesvirus macromolecular synthesis: Sequential transition of polypeptide synthesis requires functional viral polypeptides. *Proc Natl Acad Sci USA* **72**, 1276-1280 (1975).

125. P. Jones, B. Roizman, Regulation of herpesvirus macromolecular synthesis. VIII. The transcription program consists of three phases during which both extent of transcriptoin and accumulation of RNA in the cytoplasm are regulated. *J Virol* **31**, 299-314 (1979).

126. J. Heine, R. Honess, E. Cassai, B. Roizman, Proteins specified by herpes simplex virus. XII. The virion polypeptides of type 1 strains. *J Virol* **14**, 640-651 (1974).

127. R. Sandri-Goldin, "The Functions and Activities of HSV-1 ICP27, a Multifunctional Regulator of Gene Expression" in Alphaherpesviruses: Molecular virology, S. Weller, Ed. (Caister Academic Press, Norfolk, UK, 2011), chap. 3, pp. 39-49.

128. H. Fox, J. Dembowski, N. DeLuca, A herpesviral immediate early protein promotes transcription elongation of viral transcripts. *MBio* **8**, e00745-00717 (2017).

129. X. Wang *et al.*, Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nature Communications* **11**, 293 (2020).

130. A. Hill *et al.*, Herpes simplex virus turns off the TAP to evade host immunity. *Nature* **375**, 411-415 (1995).

131. W. Batterson, B. Roizman, Characterization of the herpes simplex viron-associated factor responsible for the induction of alpha genes. *J Virol* **46**, 371-377 (1983).

132. M. Campbell, J. Palfreyman, C. Preston, Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. *J Mol Biol* **180**, 1-19 (1984).

133. S. La Boissière, T. Hughes, P. O'Hare, HCF-dependent nuclear import of VP16. *EMBO J* **18**, 480-489 (1999).

134. A. Wilson, K. LaMarco, M. Peterson, W. Herr, The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74**, 115-125 (1993).

135. T. Kristie, P. Sharp, Purification of the cellular C1 factor required for the stable recognition of the Oct-1 homeodomain by the herpes simplex virus alpha-trans-induction factor (VP16). *J Biol Chem* **268**, 6525-6534 (1993).

136. T. Kristie, J. Pomerantz, T. Twomey, S. Parent, P. Sharp, The Cellular C1 Factor of the Herpes Simplex Virus Enhancer Complex Is a Family of Polypeptides. *J Biol Chem* **270**, 4387-4394 (1995).

137. B. Khurana, T. Kristie, A Protein Sequestering System Reveals Control of Cellular Programs by the Transcriptional Coactivator HCF-1. *J Biol Chem* **279**, 33673-33683 (2004).

138. H. Goto *et al.*, A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. *Genes Dev* **11**, 726-737 (1997).

139. F.-Q. Zhao, Octamer-binding transcription factors: Genomics and functions. *Front Biosci* **18**, 1051-1071 (2013).

140. S. Mackem, B. Roizman, Structural features of the herpes simplex virus alpha gene 4, 0, and 27 promoter-regulatory sequences which confer alpha regulation on chimeric thymidine kinase genes. *J Virol* **44**, 939-949 (1982).

141. L. Post, S. Mackem, B. Roizman, Regulation of alpha genes of herpes simplex virus: Expression of chimeric genes produced by fusion of thymidine kinase with alpha gene promoters. *Cell* **24**, 555-565 (1981).

142. T. Kristie, "Early events pre-initiation of alphaherpes viral gene expression" in Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis, A. Arvin *et al.*, Eds. (Cambridge University Press, Cambridge, 2007), chap. 8.

143. R. Dixon, P. Schaffer, Fine-structure mapping and functional analysis of temperature-sensitive mutants in the gene encoding the herpes simplex virus type 1 immediate early protein VP175. *J Virol* **36**, 189-203 (1980).

144. C. Preston, Control of herpes simplex virus type 1 mRNA synthesis in cells infected with wild-type virus or the temperature sensitive mutant tsK. *J Virol* **29**, 275-284 (1979).

145. R. Watson, J. Clements, A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. *Nature* **285**, 329-330 (1980).

146. D. Metzler, K. Wilcox, Isolation of herpes simplex virus regulatory protein ICP4 as a homodimeric complex. *J Virol* **55**, 329-337 (1985).

147. R. Kuddus, N. DeLuca, DNA-dependent oligomerization of herpes simplex virus type 1 regulatory protein ICP4. *J Virol* **81**, 9230-9237 (2007).

148. K. Powell, D. Purifoy, DNA-binding proteins of cells infected by herpes simplex virus type 1 and type 2. *Intervirology* **7**, 225-239 (1976).

149. S. Faber, K. Wilcox, Association of herpes simplex virus type 1 transcriptional regulatory protein ICP4 with specific nucleotide sequences in DNA. *Nucleic Acids Res* **14**, 6067-6083 (1986).

150. M. Carrozza, N. DeLuca, Interaction of the viral activator protein ICP4 with TFIID through TAF250. *Mol Cell Biol* **16**, 3085-3093 (1996).

151. J. Lester, N. DeLuca, Herpes simplex virus 1 ICP4 forms complexes with TFIID and mediator in virus-infected cells. *J Virol* **85**, 5733-5744 (2011).

152. L. Wagner, N. DeLuca, Temporal association of herpes simplex virus ICP4 with cellular complexes functioning at multiple steps in Pol II transcription. *PLos One* **8**, e78242 (2013).

153. J. Dembowski, N. DeLuca, Temporal viral genome-protein interactions define distinct stages of productive herpesviral infection. *MBio* **9**, e01182-01118 (2018).

154. P. Sampath, N. DeLuca, Binding of ICP4, TATA-binding protein, and RNA polymerase II to herpes simplex virus type 1 immediate-early, early, and late promoters in virus-infected cells. *J Virol* **82**, 2339-2349 (2008).

155. B. Grondin, N. DeLuca, Herpes simplex virus type 1 ICP4 promotes transcription preinitiation complex formation by enhancing the binding of TFIID to DNA. *J Virol* **74**, 11504–11510 (2000).

156. S. Advani, R. Weichselbaum, B. Roizman, Cdc2 cyclin-dependent kinase binds and phosphorylates herpes simplex virus 1 UL42 DNA synthesis processivity factor. *J Virol* **75**, 10326–10333 (2001).

157. S. Advani, R. Weichselbaum, B. Roizman, Herpes simplex virus 1 activates cdc2 to recruit topoisomerase II alpha for post-DNA synthesis expression of late genes. *Proc Natl Acad Sci USA* **100**, 4825–4830 (2003).

158. H. Peng, M. Nogueira, J. Vogel, T. Kristie, Transcriptional coactivator HCF-1 couples the histone chaperone Asf1b to HSV-1 DNA replication components. *Proc Natl Acad Sci USA* **107**, 2461–2466 (2010).

159. A. Burch, S. Weller, Herpes simplex virus type 1 DNA polymerase requires the mammalian chaperone Hsp90 for proper localization to the nucleus. *J Virol* **79**, 10740–10749 (2005).

160. Y. Zhu, Z. Wu, M. Cardoso, D. Parris, Processing of lagging-strand intermediates in vitro by herpes simplex virus type 1 DNA polymerase. *J Virol* **84**, 7459–7472 (2010).

161. D. Coen, S. Weinheimer, S. McKnight, A genetic approach to promoter recognition during trans induction of viral gene expression. *Science* **234**, 53-59 (1986).

162. K. Jones, K. Tamamoto, R. Tijan, Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. *Cell* **42**, 559-572 (1985).

163. R. Skaliter, I. Lehman, Rolling circle DNA replication in vitro by a complex of herpes simplex virus type 1-encoded enzymes. *Proc Natl Acad Sci USA* **91**, 10665-10669 (1994).

164. J. Dembowski, S. Dremel, N. DeLuca, Replication-coupled recruitment of viral and cellular factors to herpes simplex virus type 1 replication forks for the maintenance and expression of viral genomes. *PLoS Pathog* **13**, e1006166 (2017).

165. D. Wilkinson, S. Weller, The Role of DNA Recombination in Herpes Simplex Virus DNA Replication. *IUBMB Life* **55**, 451-458 (2008).

166. S. Ward, S. Weller, "HSV-1 DNA replication" in Alphaherpesviruses: Molecular virology, S. Weller, Ed. (Caister Academic Press, Norfolk, UK, 2011), chap. 15, pp. 89–112.

167. S. Weller, D. Coen, Herpes simplex viruses: Mechanisms of DNA replication. *Cold Spring Harb. Perspect.* **4**, a013011 (2012).

168. A. Aslani, M. Olsson, P. Elias, ATP-dependent unwinding of a minimal origin of DNA replication by the origin-binding protein and the single-strand DNA-binding protein ICP8 from herpes simplex virus type I. *J Biol Chem* **277**, 41204-41212 (2002).

169. H. Weir, J. Calder, N. Stow, Binding of the herpes simplex virus type 1 UL9 gene product to an origin of viral DNA replication. *Nucleic Acids Res* **17**, 1409-1425 (1989).

170. D. McGeoch, A. Dolan, S. Donald, D. Brauer, Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res* **14**, 1727–1745 (1986).

171. N. Stow, Localization of an origin of DNA replication within the TRS/IRS repeated region of the herpes simplex virus type 1. *EMBO J* **1**, 863-867 (1982).

172. S. Simonsson, T. Samuelsson, P. Elias, The Herpes Simplex Virus Type 1 Origin Binding Protein: SPECIFIC RECOGNITION OF PHOSPHATES AND METHYL GROUPS DEFINES THE INTERACTING SURFACE FOR A MONOMERIC DNA BINDING DOMAIN IN THE MAJOR GROOVE OF DNA*. *J Biol Chem* **273**, 24633–24639 (1998).

173. C. Lee, D. Knipe, An immunoassay for the study of DNA-binding activities of herpes simplex virus protein ICP8. *J Virol* **54**, 731-738 (1985).

174. W. Ruyechan, A. Chytil, C. Fisher, In vitro characterization of a thermolabile herpes simplex virus DNA-binding protein. *J Virol* **59**, 31-36 (1986).

175. P. Boehmer, I. Lehman, Physical interaction between the herpes simplex virus 1 origin-binding protein and single-stranded DNA-binding protein ICP8. *Proc Natl Acad Sci USA* **90**, 8444-8448 (1993).

176. P. Boehmer, M. Dodson, I. Lehman, The herpes simplex virus type-1 origin binding protein DNA helicase activity. *J Biol Chem* **268**, 1220-1225 (1993).

177. J. Blümel, B. Matz, Thermosensitive UL9 gene function is required for early stages of herpes simplex virus type 1 DNA synthesis. *J Gen Virol* **76**, 3119-3124 (1995).

178. Y. Chen *et al.*, Herpes simplex virus type 1 helicase-primase: DNA binding and consequent protein oligomerization and primase activation. *J Virol* **85**, 968-978 (2011).

179. G. Stengel, R. Kuchta, Coordinated Leading and Lagging Strand DNA Synthesis by Using the Herpes Simplex Virus 1 Replication Complex and Minicircle DNA Templates. *J Virol* **85**, 957-967 (2011).

180. J. Conway, F. Homa, "Nucleocapsid Structure, Assembly and DNA Packaging of Herpes Simplex Virus" in Alphaherpesviruses: Molecular virology, S. Weller, Ed. (Caister Academic Press, Norfolk, UK, 2011), chap. 10, pp. 175-193.

181. C. Lamberti, S. Weller, The Herpes Simplex Virus Type 1 Cleavage/Packaging Protein, UL32, Is Involved in Efficient Localization of Capsids to Replication Compartments. *J Virol* **72**, 2463–2473 (1998).

182. L. Deiss, J. Chou, N. Frenkel, Functional domains within the a sequence involved in the cleavage-packaging of herpes simplex virus DNA. *J Virol* **59**, 605-618 (1986).

183. G. Church, A. Dasgupta, D. Wilson, Herpes Simplex Virus DNA Packaging without Measurable DNA Synthesis. *J Virol* **72**, 2745–2751 (1998).

184. D. Johnson, J. Baines, Herpesviruses remodel host membranes for virus egress. *Nat Rev Microbiol* **9**, 382-394 (2011).

185. D. McGeoch *et al.*, The complete DNA Sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol* **69**, 1531-1574 (1988).

186. D. McGeoch, A. Dolan, S. Donald, F. Rixon, Sequence determination and genetic content of the short unique region of the herpes simplex virus type 1. *J Mol Biol* **181**, 1-13 (1985).

187. C. Wu, N. Nelson, D. McGeoch, M. Challberg, Identification of herpes simplex virus type 1 genes required for origin-dependent DNA synthesis. *J Virol* **62**, 435-443 (1988).

188. N. DeLuca, P. Schaffer, Activation of immediate-early, early, and late promoters by temperature-sensitive and wild-type forms of herpes simplex virus type 1 protein ICP4. *Mol Cell Biol* **5**, 1997-2008 (1985).

189. P. Mavromara-Nazos, B. Roizman, Delineation of regulatory domains of early (beta) and late (gamma 2) genes by construction of chimeric genes expressed in herpes simplex virus 1 genomes. *Proc Natl Acad Sci USA* **86**, 4071-4075 (1989).

190. D.-B. Kim, S. Zabierowski, N. DeLuca, The initiator element in a herpes simplex virus type 1 late-gene promoter enhances activation by ICP4, resulting in abundant late-gene expression. *J Virol* **76**, 1548-1558 (2002).

191. F. Homa, J. Glorioso, M. Levine, A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. *Genes Dev* **2**, 40-53 (1988).

192. B. Gu, N. DeLuca, Requirements for activation of the herpes simplex virus glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *J Virol* **68**, 7953-7965 (1994).

193. D. Purifoy, R. Lewis, K. Powell, Identification of the herpes simplex virus DNA polymerase gene. *Nature* **269**, 621-623 (1977).

194. P. Digard, W. Bebrin, K. Weisshart, D. Coen, The extreme C terminus of herpes simplex virus DNA polymerase is crucial for functional interaction with processivity factor UL42 and for viral replication. *J Virol* **67**, 398-406 (1993).

195. S. Terrell, D. Coen, The pre-NH(2)-terminal domain of the herpes simplex virus 1 DNA polymerase catalytic subunit is required for efficient viral replication. *J Virol* **86**, 11057-11065 (2012).

196. D.-B. Kim, N. DeLuca, Phosphorylation of transcription factor Sp1 during herpes simplex virus type 1 infection. *J Virol* **76**, 6473-6479 (2002).

197. S. McKnight, The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nucleic Acids Res* **8**, 5959-5964 (1980).

198. M. Dalrymple, D. McGeoch, A. Davison, C. Preston, DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for the transcriptional activation of immediate early promoters. *Nucleic Acids Res* **13**, 7865-7879 (1985).

199. R. Frink, R. Eisenberg, G. Cohen, E. Wagner, Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *J Virol* **45**, 634-647 (1983).

200. L. Vo Ngoc, C. Cassidy, C. Huang, S. Duttke, J. Kadonaga, The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* **31**, 6-11 (2017).

201. B. Tischer, G. Smith, N. Osterrieder, "En passant mutagenesis: a two step markerless red recombination system" in In Vitro Mutagenesis Protocols, J. Braman, Ed. (Humana Press, Totowa, NJ, 2010), pp. 421-430.

202. B. Tischer, J. Einem, B. Kaufer, N. Osterrieder, Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in Escherichia coli. *BioTechniques* **40**, 191-197 (2006).

203. W. Gierasch *et al.*, Construction and characterization of bacterial artificial chromosomes containing HSV-1 strains 17 and KOS. *J Virol Methods* **135**, 197-206 (2006).

204. C. Preston, Abnormal properties of an immediate early polypeptide in cells infected with the herpes simplex virus type 1 mutant tsK. *J Virol* **32**, 357-369 (1979).

205. L. Wagner, J. Lester, F. Sivrich, N. DeLuca, The N-terminus and C-terminus of HSV-1 ICP4 cooperate to activate viral gene expression. *J Virol* **86**, 6862-6874 (2012).

206. E. Afgan *et al.*, The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544 (2018).

207. D. Kim, B. Langmead, S. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357-360 (2015).

208. Y. Liao, G. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

209. F. Ramírez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-W165.

210. B. Langmead, S. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

211. A. Quinlan, I. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

212. T. Doerks, R. Copley, J. Schultz, C. Ponting, P. Bork, Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* **12**, 47-56 (2002).

213. S. Dremel, N. DeLuca, Genome replication affects transcription factor binding mediating the cascade of herpes simplex virus transcription. *Proc Natl Acad Sci USA* **116**, 3734-3739 (2019).

214. R. Courtney, M. Benyesh-Melnick, Isolation and characterization of a large molecular-weight polypeptide of herpes simplex virus type 1. *Virology* **62** (1974).

215. N. DeLuca, A. McCarthy, P. Schaffer, Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. *J Virol* **56**, 558-570 (1985).

216. R. Everett, G. Sourvinos, A. Orr, Recruitment of herpes simplex virus type 1 transcriptional regulatory protein ICP4 into foci juxtaposed to ND10 in life, infected cells. *J Virol* **77**, 3680-3689 (2003).

217. Y. Liang, J. Vogel, A. Narayanan, H. Peng, T. Kristie, Inhibition of the histone demethylase LSD1 blocks alpha-herpesvirus lytic replication and reactivation from latency. *Nat Med* **15**, 1312-1317 (2009).

218. J. DiDonato, J. Spitzner, M. Muller, A predictive model for DNA recognition by the herpes simplex virus protein ICP4. *J Mol Bio* **219**, 451-470 (1991).

219. A. Imbalzano, D. Coen, N. DeLuca, Herpes simplex virus transactivator ICP4 operationally substitutes for the cellular transcription factor Sp1 for efficient expression of the viral thymidine kinase gene. *J Virol* **65**, 565-574 (1991).

220. F. Yao, R. Courtney, A major transcriptional regulatory protein (ICP4) of herpes simplex virus type 1 is associated with purified virions. *J Virol* **63**, 3338-3344 (1989).

221. C. Preston, M. Frame, M. Campbell, A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. *Cell* **52**, 425-434 (1988).

222. S. Stern, M. Tanaka, W. Herr, The Oct-1 homeodomain directs formation fo a multi-protein-DNA complex with the HSV transactivator VP16. *Nature* **341**, 624-630 (1989).

223. S. Stern, W. Herr, The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeodomain: evidence for a homeodomain recognition subdomain. *Genes Dev* **5**, 2555-2566 (1991).

224. R. Abrisch, T. Eidem, P. Yakovchuk, J. Kugel, J. Goodrich, Infection by herpes simplex virus 1 causes near-complete loss of RNA polymerase II occupancy on the host cell genome. *J Virol* **90**, 2503-2513 (2015).

225. C. Birkenheuer, C. Danko, J. Baines, Herpes simplex virus 1 dramatically alters loading and positioning of RNA polymerase II on host genes early in infection. *J Virol* **92**, e02184-02117 (2018).

226. K. Koop, J. Duncan, J. Smiley, Binding sites for the herpes simplex virus immediate-early protein ICP4 impose an increased dependence on viral DNA replication on simple model promoters located in the viral genome. *J Virol* **67**, 7254-7263 (1993).

227. R. Rivera-Gonzalez, A. Imbalzano, B. Gu, N. DeLuca, The role of ICP4 repressor activity in temporal expression of the IE-3 and latency-associated transcript promoters during HSV-1 infection. *Virology* **202**, 550-564 (1994).

228. J. Smiley, D. Johnson, L. Pizer, R. Everett, The ICP4 binding sites in the herpes simplex virus type 1 glycoprotein D (gD) promoter are not essential for efficient gD transcription during virus infection. *J Virol* **66**, 623-631 (1992).

229. N. DeLuca, P. Schaffer, Physical and functional domains of the herpes simplex virus transcriptional regulatory protein ICP4. *J Virol* **62**, 732-743 (1988).

230. K. Smith, Relationship between the envelope and the infectivity of herpes simplex virus. *Exp Biol Med* **115**, 814-815 (1964).

231. F. Ramírez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-W165 (2016).

232. J. Feng, T. Liu, B. Qin, Y. Zhang, X. Liu, Identifying ChIP-Seq enrichment using MACS. *Nat Protoc* **7**, 1728-1740 (2012).

233. H. Li *et al.*, The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

234. G. Yu, L. Wang, Q. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).

235. T. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-W208 (2008).

236. N. Plongthongkum, D. Diep, K. Zhang, Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet* **15**, 647-661 (2014).

237. E. Diala, R. Hoffman, Epstein-Barr HR-1 Virion DNA Is Very Highly Methylated. *J Virol* **45**, 482-483 (1983).

238. S. Sharma, N. Biswal, Studies on the in vivo methylation of replicating herpes simplex virus type 1 DNA. *Virology* **82**, 265-274 (1977).